Journal of Economic Literature 2021, 59(2), 391–425
https://doi.org/10.1257/jel.20191450

Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects[†]

Alberto Abadie*

Probably because of their interpretability and transparent nature, synthetic controls have become widely applied in empirical research in economics and the social sciences. This article aims to provide practical guidance to researchers employing synthetic control methods. The article starts with an overview and an introduction to synthetic control estimation. The main sections discuss the advantages of the synthetic control framework as a research design, and describe the settings where synthetic controls provide reliable estimates and those where they may fail. The article closes with a discussion of recent extensions, related methods, and avenues for future research. (*JEL* B41, C32, C54, E23, F15, O47)

1. *Introduction*

Synthetic control methods (Abadie and Gardeazabal 2003; Abadie, Diamond, and Hainmueller 2010) have become widely applied in empirical research in economics and other disciplines. Under appropriate conditions, synthetic controls provide substantial advantages as a research design method in the social sciences. These advantages, I believe, explain the increasing popularity of synthetic control methods in empirical research. At the same time, the validity of synthetic control estimators depends on important practical requirements. Perfunctory applications that ignore the context of the empirical investigation and the characteristics of the data may miss the mark, producing misleading estimates.

My goal with this article is to provide guidance for empirical practice to researchers employing synthetic control methods. With this goal in mind, I put special emphasis on feasibility, data requirements, contextual requirements, and methodological issues related to the empirical application of synthetic controls. Particularly important is

* Department of Economics, Massachusetts Institute of Technology. I would like to thank Matias Cattaneo, Brigham Frandsen, Bruno Ferman, Javier Gardeazabal, Jens Hainmueller, Jérémy L'Hour, Guido Imbens, Devavrat Shah, Kaspar Wüthrich, participants at the Conference on Synthetic Controls and Related Methods at MIT/IDSS, NeurIPS 2019, the Chamberlain Online Seminar, Steven Durlauf (the editor), and five referees for insightful comments, and Kevin Li for expert research assistance. NSF support through grant SES-1756692 is gratefully acknowledged. Some fragments of this article build on an internal document previously commissioned by the Inter-American Development Bank.

† Go to https://doi.org/10.1257/jel.20191450 to visit the article page and view author disclosure statement(s).

characterizing the practical settings where synthetic controls may be useful and those where they may fail.

Section 2 briefly introduces the ideas behind the synthetic control methodology in the context of comparative case studies. Section 3 discusses some of the formal aspects of the synthetic control methodology that are of particular interest for empirical applications. Readers who are already familiar with the synthetic control methodology may only need to read subsections 3.3 to 3.5 in detail, and skim through section 2 and the rest of section 3 in order to acquaint themselves with terms and notation that will be employed in later sections. Sections 4 through 6 comprise the core of the article. Section 4 discusses the practical advantages of synthetic control estimators. Sections 5 and 6 discuss contextual and data requirements for synthetic control empirical studies. I discuss the validity of these requirements in applied settings and potential ways to adapt the research design when the requirements do not hold in practice. Section 7 describes robustness and diagnostic checks to evaluate the credibility of a synthetic control counterfactual and to measure the extent to which results are sensitive to changes in the study design. Section 8 discusses extensions and recent proposals. The final section contains conclusions and describes open areas for research on synthetic controls.

2. *A Primer on Synthetic Control Estimators*

In a recent *Journal of Economic Perspectives* survey on the econometrics of policy evaluation, Susan Athey and Guido Imbens describe synthetic controls as "arguably the most important innovation in the policy evaluation literature in the last 15 years" (Athey and Imbens 2017). In the last few years, synthetic controls have been applied to study the effects of right-to-carry laws (Donohue, Aneja, and Weber 2019), legalized prostitution (Cunningham and Shah 2018), immigration policy (Bohn, Lofstrom, and Raphael 2014), corporate political connections (Acemoglu et al. 2016), taxation (Kleven, Landais, and Saez 2013), organized crime (Pinotti 2015), and many other key policy issues. They have also been adopted as the main tool for data analysis across different sides of the issues in recent prominent debates on the effects of immigration (Borjas 2017, Peri and Yasenov 2019) and minimum wages (Allegretto et al. 2017, Jardim et al. 2017, Neumark and Wascher 2017, Reich et al. 2017). Synthetic controls are also applied outside economics: in the social sciences, biomedical disciplines, engineering, etc. (see, e.g., Heersink, Peterson, and Jenkins 2017; Pieters et al. 2017). Outside academia, synthetic controls have found considerable coverage in the popular press (see, e.g., Guo 2015, Douglas 2018) and have been widely adopted by multilateral organizations, think tanks, business analytics units, governmental agencies, and consulting firms. For example, the synthetic control method plays a prominent role in the official evaluation of the effects of the massive Bill & Melinda Gates Foundation's Intensive Partnerships for Effective Teaching program (Gutierrez, Weinberger, and Engberg 2016).

Synthetic control methods were originally proposed in Abadie and Gardeazabal (2003) and Abadie, Diamond, and Hainmueller (2010) with the aim to estimate the effects of aggregate interventions, that is, interventions that are implemented at an aggregate level affecting a small number of large units (such as a cities, regions, or countries), on some aggregate outcome of interest. More recently, synthetic control methods have been applied to settings with a large number of units.[1] We will discuss this and other extensions in section 8.

[1] See, for example, Acemoglu et al. (2016), Kreif et al. (2016), Abadie and L'Hour (2019), and Dube and Zipperer (2015).

Consider a setting where one aggregate unit, such as a state or a school district, is exposed to an event or intervention of interest. For example, Abadie, Diamond, and Hainmueller (2010) study the effect of a large tobacco control program adopted in California in 1988, and Bifulco, Rubenstein, and Sohn (2017) evaluate the effects of an educational program adopted in the Syracuse, New York, school district in 2008. In accordance with the program evaluation literature in economics, the terms "treated" and "untreated" will refer to units exposed and not exposed to the event or intervention of interest, respectively. I will use the terms "event," "intervention," and "treatment" interchangeably. Traditional regression analysis techniques require large samples and many observed instances of the event or intervention of interest and, as a result, they are often ill-suited to estimate the effects of infrequent events, such as policy interventions, on aggregate units. Economists have approached the estimation of the effects of large-scale but infrequent interventions using time-series analysis and comparative case studies. Single-unit time-series analysis is an effective tool to study the short-term effects of policy interventions in cases when we expect short-term effects to be of a substantial magnitude.[2] However, the use of time-series techniques to estimate medium and long-term effects of policy intervention is complicated by the presence of shocks to the outcome of interest, aside from the effect of the intervention. Comparative case studies are based on the idea that the effect of an intervention can be inferred by comparing the evolution of the outcome variables of interest between the unit exposed to treatment and a group of units that are similar to the exposed unit but were not affected by the treatment. This can be achieved whenever the evolution of the outcomes for the unit affected by the intervention and the comparison units is driven by common factors that induce a substantial amount of co-movement.

Comparative case studies have long been applied to the evaluation of large-scale events or aggregate interventions. For example, to estimate the effects of the massive arrival of Cuban expatriates to Miami during the 1980 Mariel boatlift on native unemployment in Miami, Card (1990) compares the evolution of native unemployment in Miami at the time of the boatlift to the average evolution of native unemployment in four other cities in the United States. Similarly, Card and Krueger (1994) use Pennsylvania as a comparison to estimate the effects of an increase in the New Jersey minimum wage on employment in fast food restaurants in New Jersey. A drawback of comparative case studies of this type is that the selection of the comparison units is not formalized and often relies on informal statements of affinity between the units affected by the event or intervention of interest and a set of comparison units. Moreover, when the units of observation are a small number of aggregate entities, like countries or regions, no single unit alone may provide a good comparison for the unit affected by the intervention.

The synthetic control method is based on the idea that, when the units of observation are a small number of aggregate entities, a combination of unaffected units often provides a more appropriate comparison than any single unaffected unit alone. The synthetic control methodology formalizes the selection of the comparison units using a data driven procedure. As we will discuss later, this formalization also opens the door to a mode of quantitative inference for comparative case studies.

[2] The literature on "interrupted time-series" is particularly relevant in the context of policy evaluation. See, for example, Cook and Campbell (1979), which discusses the limitations of this methodology if interventions are gradual rather than abrupt and/or if the causal effect of an intervention is delayed in time. Interrupted time-series methods are closely related to regression-discontinuity design techniques (see, e.g., Thistlethwaite and Campbell 1960).

3. *Formal Aspects of the Synthetic Control Method*

3.1 *The Setting*

Suppose that we obtain data for $J+1$ units: $j=1,2,\ldots,J+1$. Without loss of generality, we assume that the first unit ($j=1$) is the treated unit, that is, the unit affected by the policy intervention of interest.[3] The "donor pool," that is, the set of potential comparisons, $j=2,\ldots,J+1$ is a collection of untreated units not affected by the intervention. We assume also that our data span T periods and that the first T_0 periods are before the intervention. For each unit, j, and time, t, we observe the outcome of interest, Y_{jt}. For each unit, j, we also observe a set of k predictors of the outcome, $X_{1j},\ldots,X_{kj}$, which may include pre-intervention values of Y_{jt} and which are themselves unaffected by the intervention. The $k \times 1$ vectors $\mathbf{X}_1,\ldots,\mathbf{X}_{J+1}$ contain the values of the predictors for units $j=1,\ldots,J+1$, respectively. The $k \times J$ matrix, $\mathbf{X}_0 = \left[\mathbf{X}_2 \cdots \mathbf{X}_{J+1}\right]$, collects the values of the predictors for the J untreated units. For each unit, j, and time period, t, we will define Y_{jt}^N to be the potential response without intervention. For the unit affected by the intervention, $j=1$, and a post-intervention period, $t>T_0$, we will define Y_{1t}^I to be the potential response under the intervention.[4] Then, the effect of the intervention of interest for the affected unit in period t (with $t>T_0$) is:

$$\tau_{1t} = Y_{1t}^I - Y_{1t}^N. \tag{1}$$

Because unit "one" is exposed to the intervention after period T_0, it follows that for $t>T_0$ we have $Y_{1t}=Y_{1t}^I$. Simply put, for the unit affected by the intervention and a post-intervention period we observe the potential outcome under the intervention. The great policy evaluation challenge is to estimate Y_{1t}^N for $t>T_0$: how the outcome of interest would have evolved for the affected unit *in the absence of the intervention*. This is a counterfactual outcome, as the affected unit was, by definition, exposed to the intervention of interest after $t=T_0$. As equation (1) makes clear, given that Y_{1t}^I is observed, the problem of estimating the effect of a policy intervention is equivalent to the problem of estimating Y_{1t}^N. Notice also that equation (1) allows the effect of the intervention to change over time. This is crucial because intervention effects may not be instantaneous and may accumulate or dissipate as time after the intervention passes.

3.2 *Estimation*

Comparative case studies aim to reproduce Y_{1t}^N—that is, the value of the outcome variable that would have been observed for the affected unit in the absence of the intervention—using one unaffected unit or a small number of unaffected units that have similar characteristics as the affected unit at the time of the intervention. When the data consist of a few aggregate entities, such as regions or countries, it is often difficult to find a single unaffected unit that provides a suitable comparison for the unit affected by the policy intervention of interest. As mentioned above, the synthetic control method is based on the observation that a combination of units in the donor pool may approximate the characteristics of the

[3] The synthetic control framework can easily accommodate estimation with multiple treated units by fitting separate synthetic controls for each of the treated units. In practice, however, estimation with several treated units may carry some practical complications that are discussed in section 8.

[4] Y_{1t}^I and Y_{jt}^N are the potential outcomes of Rubin's model for causal inference (see, e.g., Rubin 1974, Holland 1986). To simplify notation, I exclude the start time of the intervention from the notation for Y_{1t}^I. Notice, however, that the value of Y_{1t}^I depends in general not only on when the intervention starts, but also other features of the intervention that are fixed in our analysis and, therefore, excluded from the notation.

affected unit substantially better than any unaffected unit alone. A synthetic control is defined as a weighted average of the units in the donor pool. Formally, a synthetic control can be represented by a $J \times 1$ vector of weights, $\boldsymbol{W} = \left(w_2, \ldots, w_{J+1}\right)'$. Given a set of weights, $\boldsymbol{W}$, the synthetic control estimators of Y_{1t}^N and τ_{1t} are, respectively:

$$\hat{Y}_{1t}^N = \sum_{j=2}^{J+1} w_j Y_{jt}, \tag{2}$$

and

$$\hat{\tau}_{1t} = Y_{1t} - \hat{Y}_{1t}^N. \tag{3}$$

To avoid extrapolation, the weights are restricted to be nonnegative and to sum to one, so synthetic controls are weighted averages of the units in the donor pool. The requirement that weights should be nonnegative and no greater than one can be relaxed at the cost of allowing extrapolation. For example, Abadie, Diamond, and Hainmueller (2015) show that, in the context of estimating the effect of a policy intervention, there is a regression estimator that can be represented as a synthetic control with weights that are unrestricted except for that the sum of the weights is equal to one. By not restricting the weights to be in $\left[0,1\right]$, regression allows extrapolation.[5] Restricting synthetic control weights to be nonnegative and sum to one generates synthetic controls that are weighted averages of the outcomes of units in the donor pool, with weights that are typically sparse (see section 4). That is, only a small number of units in the donor pool contribute to the estimate of the counterfactual of interest, $\hat{Y}_{1t}^N$, and the contribution of each unit is represented by its synthetic control weight. Because synthetic control weights define a weighted average and because they are sparse, the specific nature of a synthetic control counterfactual estimate is particularly transparent, relative to competing methods. Notice also that considering synthetic controls with weights that sum to one may be warranted only if the variables in the data are rescaled to correct for differences in size between units (e.g., per capita income) or if such correction is not needed because the variables in the data do not scale with size (e.g., prices).

As an example, a synthetic control that assigns equal weights, $w_j = 1/J$, to each of the units in the control group results in the following estimator for τ_{1t}:

$$\hat{\tau}_{1t} = Y_{1t} - \frac{1}{J}\sum_{j=2}^{J+1} Y_{jt}. \tag{4}$$

In this case, the synthetic control is the simple average of all the units in the donor pool. A population-weighted version is

$$\hat{\tau}_{1t} = Y_{1t} - \sum_{j=2}^{J+1} w_j^{pop} Y_{jt}, \tag{5}$$

where w_j^{pop} is the population in unit j (e.g., at the time of the intervention) as a fraction of the total population in the donor pool. If, however, a single unit, m, in the donor pool is used as a comparison, then $w_m = 1, w_j = 0$ for $j \neq m$, and

$$\hat{\tau}_{1t} = Y_{1t} - Y_{mt}. \tag{6}$$

For nearest-neighbor estimators, m is the index value that minimizes $\|\boldsymbol{X}_1 - \boldsymbol{X}_j\|$ over j for some norm $\|\cdot\|$.

Expressing the comparison unit as a synthetic control motivates the question of

[5] See section 4 for details. Doudchenko and Imbens (2016), Ferman (2019), and Li (2020) discuss the role of weight restrictions as regularization devices. Doudchenko and Imbens (2016) and Chernozhukov, Wüthrich, and Zhu (2019a) propose alternative regularization procedures for synthetic controls based on the elastic net and the lasso, respectively.

how the weights, $w_2, \ldots, w_{J+1}$, should be chosen in practice. Abadie and Gardeazabal (2003) and Abadie, Diamond, and Hainmueller (2010) propose to choose $w_2, \ldots, w_{J+1}$ so that the resulting synthetic control best resembles the pre-intervention values for the treated unit of predictors of the outcome variable. That is, given a set of nonnegative constants, $v_1, \ldots, v_k$, Abadie and Gardeazabal (2003) and Abadie, Diamond, and Hainmueller (2010) propose to choose the synthetic control, $\boldsymbol{W}^* = \left(w_2^*, \ldots, w_{J+1}^*\right)'$ that minimizes

$$(7)\quad \|\mathbf{X}_1 - \mathbf{X}_0 \boldsymbol{W}\| = \left(\sum_{h=1}^{k} v_h \left(X_{h1} - w_2 X_{h2} - \cdots - w_{J+1} X_{hJ+1}\right)^2\right)^{1/2}$$

subject to the restriction that $w_2, \ldots, w_{J+1}$ are nonnegative and sum to one.[6] Then, the estimated treatment effect for the treated unit at time $t = T_0 + 1, \ldots, T$ is

$$(8)\qquad \hat{\tau}_{1t} = Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt}.$$

The positive constants $v_1, \ldots, v_k$ in (7) reflect the relative importance of the synthetic control reproducing the values of each of the k predictors for the treated unit, $X_{11}, \ldots, X_{k1}$. For a given set of weights, $v_1, \ldots, v_k$, minimizing equation (7) can be easily accomplished using constrained quadratic optimization. That is, each potential choice of $\boldsymbol{V} = \left(v_1, \ldots, v_k\right)$ produces a synthetic control, $\boldsymbol{W}(\boldsymbol{V}) = \left(w_2(\boldsymbol{V}), \ldots, w_{J+1}(\boldsymbol{V})\right)'$, which can be determined by minimizing equation (7), subject to the restriction that the weights in $\boldsymbol{W}(\boldsymbol{V})$ are positive and sum to one.

Of course, a question remains about how to choose $\boldsymbol{V}$. A simple selector of v_h is the inverse of the variance of $X_{h1}, \ldots, X_{hJ+1}$, which in effect rescales all rows of $[\mathbf{X}_1 : \mathbf{X}_0]$ to have unit variance. Alternatively, Abadie and Gardeazabal (2003) and Abadie, Diamond, and Hainmueller (2010) choose $\boldsymbol{V}$, such that the synthetic control $\boldsymbol{W}(\boldsymbol{V})$ minimizes the mean squared prediction error (MSPE) of this synthetic control with respect to Y_{1t}^N:

$$\sum_{t \in \mathcal{T}_0} \left(Y_{1t} - w_2(\boldsymbol{V}) Y_{2t} - \cdots - w_{J+1}(\boldsymbol{V}) Y_{J+1t}\right)^2,$$

for some set $\mathcal{T}_0 \subseteq \{1, 2, \ldots, T_0\}$ of pre-intervention periods. Abadie, Diamond, and Hainmueller (2015) propose a related method to choose $v_1, \ldots, v_k$ via out-of-sample validation. The ideas behind out-of-sample validation selection of $v_1, \ldots, v_k$ are described next. The goal of the synthetic control is to approximate the trajectory that would have been observed for Y_{1t} and $t > T_0$ in the absence of the intervention. For that purpose, the synthetic control method selects a set of weights $\boldsymbol{W}$ such that the resulting synthetic control resembles the affected unit before the intervention along the values of the variables $X_{11}, \ldots, X_{k1}$. The question of choosing $\boldsymbol{V} = \left(v_1, \ldots, v_k\right)$ boils down to assessing the relative importance of each of $X_{11}, \ldots, X_{k1}$ as a predictor of Y_{1t}^N. That is, the value v_h aims to reflect the relative importance of approximating the value of X_{h1} for predicting Y_{1t}^N in the post-intervention period, $t = T_0 + 1, \ldots, T$. Because Y_{1t}^N is not observed for $t = T_0 + 1, \ldots, T$, we cannot directly evaluate the relative importance of fitting each predictor to approximate Y_{1t}^N in the post-intervention period. However, Y_{1t}^N is observed for the pre-intervention periods $t = 1, 2, \ldots, T_0$, so it is possible to use pre-intervention data to assess the predictive

[6] For the sake of expositional simplicity, I discuss only the normalized Euclidean norm in equation (7). Of course, other norms are possible. Also, to avoid notational clutter, dependence of the norm in equation (7) from the weights $v_1, \ldots, v_k$ is left implicit in the notation.

power on Y_{1t}^N of the variables $X_{1j}, \ldots, X_{kj}$. This can be accomplished in the following manner.

1. Divide the pre-intervention periods into a initial *training* period and a subsequent *validation* period. For simplicity and concreteness, we will assume that T_0 is even and the training and validation periods span $t = 1, \ldots, t_0$ and $t = t_0 + 1, \ldots, T_0$, respectively, with $t_0 = T_0/2$. In practice, the lengths of the training and validation periods may depend on application-specific factors, such as the extent of data availability on outcomes in the pre-intervention and post-intervention periods, and the specific times when the predictors are measured in the data.

2. For every value $\boldsymbol{V}$, let $\tilde{w}_2(\boldsymbol{V}), \ldots, \tilde{w}_{J+1}(\boldsymbol{V})$ be the synthetic control weights computed with training period data on the predictors. The MSPE of this synthetic control with respect to Y_{1t}^N in the validation period is

$$\sum_{t=t_0+1}^{T_0} \Big(Y_{1t} - \tilde{w}_2(\boldsymbol{V}) Y_{2t} - \cdots - \tilde{w}_{J+1}(\boldsymbol{V}) Y_{J+1t} \Big)^2. \tag{9}$$

3. Select a value $\boldsymbol{V}^* \in \mathcal{V}$ such that the MSPE in equation (9) is small, where $\mathcal{V}$ is a set of potential values for $\boldsymbol{V}$.

4. Use the resulting $\boldsymbol{V}^*$ and data on the predictors for the last t_0 periods before in the intervention, $t = T_0 - t_0 + 1, \ldots, T_0$, to calculate $\boldsymbol{W}^* = \boldsymbol{W}(\boldsymbol{V}^*)$.[7]

This is a heuristic procedure, and one that is useful only as long as it produces $\boldsymbol{V}^*$, such that $Y_{1t} \approx \tilde{w}_2(\boldsymbol{V}^*) Y_{2t} + \cdots + \tilde{w}_{J+1}(\boldsymbol{V}^*) Y_{J+1t}$ for $t = t_0 + 1, \ldots, T_0$, and $\mathbf{X}_1 \approx \mathbf{X}_0 \boldsymbol{W}^*$ for the set of predictors used to calculate $\boldsymbol{W}^*$.

To give sharpness to the discussion of the properties and practical implementation of synthetic control estimators I will refer, as a running example, to an application in Abadie, Diamond, and Hainmueller (2015), which estimates the effect of the 1990 German reunification on per capita GDP in West Germany. In this application, the intervention is the 1990 German reunification and the treated unit is the former West Germany. The donor pool consists a set of industrialized countries, and $\mathbf{X}_1$ and $\mathbf{X}_0$ collect prereunification values of predictors of economic growth. Figure 1, panel A, compares the trajectory of per capita GDP before and after the reunification for West Germany and a simple average of the countries in the donor pool, for the years 1960–2003. This is the comparison in equation (4). Average per capita GDP among the countries in the donor pool fails to reproduce the trajectory of per capita GDP for West Germany even before the reunification takes place in 1990. Moreover, the restriction of parallel trends required for difference-in-differences models (see, e.g., Abadie 2005, Angrist and Pischke 2009) fails to hold in the pre-intervention data. Figure 1, panel B, reports the trajectory of per capita GDP for West Germany and for a synthetic control calculated in the manner explained in this section. This figure shows that a weighted average of the countries in the donor pool is able to closely approximate the trajectory of per capita GDP for West Germany before the German reunification.

Moreover, the synthetic control of figure 1, panel B, closely reproduces the

[7] As discussed in Klößner et al. (2018), cross-validation weights are not always unique. That is, minimization of equation (9) may not have a unique solution. In principle, this could be dealt with via penalization (e.g., adding a term $\gamma \sum_{h=1}^{k} v_m^2$ for some $\gamma > 0$ to equation (9), which favors dense sets of weights). In practice, however, researchers should aim to demonstrate that their results are not overly sensitive to particular choices of $\boldsymbol{V}$.

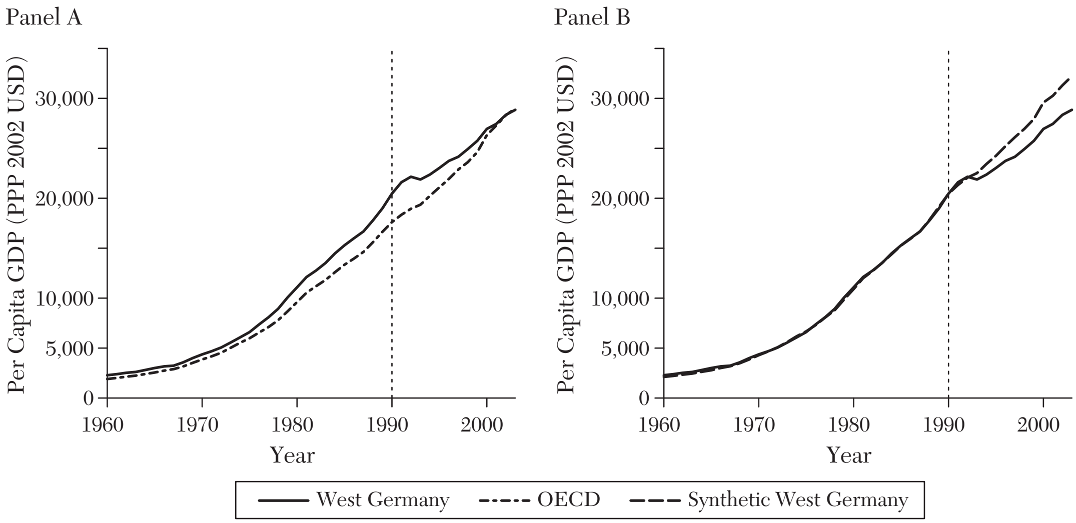

Figure 1. Synthetic Control Estimation in the German Reunification Example

Notes: Panel A compares the evolution of per capita GDP in West Germany to the evolution of per capita GDP for a simple average of OECD countries. In panel B the comparison is with a synthetic control calculated in the manner explained in subsection 3.2. See Abadie, Diamond, and Hainmueller (2015) for details.

prereunification values of economic growth predictors for West Germany. In columns 1 to 4, table 1 reports the value of economic growth predictors for West Germany, $\boldsymbol{X}_1$, for the synthetic control, $\boldsymbol{X}_0\boldsymbol{W}^*$, for the simple average of the units in the donor pool as in equation (4), and for the single comparison estimator of equation (6), where m is the index of the nearest neighbor in terms of the values of the predictors in $\boldsymbol{X}_1$ (see table 1 note for details). The results in table 1 illustrate the potential benefit in terms of the fit of the covariates from using synthetic control methods in studies with a few comparison units in the donor pool. While the simple average of the countries in the OECD sample and the nearest neighbor both fail to reproduce the economic growth predictors for West Germany prior to the reunification, a synthetic control provides a rather accurate approximation to the value of the predictors for West Germany.

Table 2 relays the identities and contributions of each of the units in the donor pool to the synthetic control for West Germany. Austria carries the largest weight, with the United States, Japan, Switzerland, and the Netherlands also contributing to the synthetic control with weights in decreasing order. The rest of the countries in the donor pool do not contribute to the synthetic control for West Germany. As we will see later, the sparsity of the weights in table 2 is typical of synthetic control estimators, and is a consequence of the geometric characteristics of the solution to the optimization problem that generates synthetic controls.

3.3 *Bias Bound*

Abadie, Diamond, and Hainmueller (2010) study the bias properties of synthetic

TABLE 1
ECONOMIC GROWTH PREDICTOR MEANS BEFORE THE GERMAN REUNIFICATION

	West Germany (1)	Synthetic West Germany (2)	OECD average (3)	Austria (nearest neighbor) (4)
GDP per capita	15,808.9	15,802.2	13,669.4	14,817.0
Trade openness	56.8	56.9	59.8	74.6
Inflation rate	2.6	3.5	7.6	3.5
Industry share	34.5	34.4	33.8	35.5
Schooling	55.5	55.2	38.7	60.9
Investment rate	27.0	27.0	25.9	26.6

Note: The first column reports $\mathbf{X}_1$, the second column reports $\mathbf{X}_0\mathbf{W}^*$, the third column reports a simple average of $\mathbf{X}_j$ for the 16 OECD countries in the donor pool, and the last column reports the value of $\mathbf{X}_j$ for the nearest neighbor of West Germany in terms of predictors values. GDP per capita, inflation rate, and trade openness are averages for the 1981–90 period. Industry share (of value added) is the average for 1981–89. Schooling is the average for 1980 and 1985. Investment rate is averaged over 1980–84. See Abadie, Diamond, and Hainmueller (2015) for variable definitions and sources. The nearest neighbor in column 4 minimizes the Euclidean norm of the pairwise differences between the values of the predictors for West Germany and for each of the countries in the donor pool, after rescaling the predictors to have unit variance.

TABLE 2
SYNTHETIC CONTROL WEIGHTS FOR WEST GERMANY

Australia	—
Austria	0.42
Belgium	—
Denmark	—
France	—
Greece	—
Italy	—
Japan	0.16
Netherlands	0.09
New Zealand	—
Norway	—
Portugal	—
Spain	—
Switzerland	0.11
United Kingdom	—
United States	0.22

controls estimators for the cases when Y_{1t}^N is generated by (i) a linear factor model, or (ii) a vector autoregressive model.[8] They show that, under some conditions, the synthetic control estimator is unbiased for a vector autoregressive model, and provide a bias bound for a linear factor model. Here, I will restrict the exposition to the linear factor

[8] Notice that the assumptions on the data-generating process involve Y_{jt}^N, but not Y_{1t}^I. Since $Y_{1t}^I = Y_{1t}$ is observed, estimation of τ_{1t} for $t > T_0$ requires no assumptions on the process that generates Y_{1t}^I.

model, which can be seen as a generalization of difference in differences. Consider the following linear factor model for Y_{jt}^N,

$$(10) \qquad Y_{jt}^N = \delta_t + \theta_t \mathbf{Z}_j + \lambda_t \mu_j + \varepsilon_{jt},$$

where δ_t is a time trend, $\mathbf{Z}_j$ and μ_j are vectors of observed and unobserved predictors of Y_{jt}^N, respectively, with coefficients θ_t and λ_t, and ε_{jt} is zero mean individual transitory shocks. In the time-series literature in econometrics, θ_t and λ_t are referred to as common factors, and $\mathbf{Z}_j$ and μ_j as factor loadings. The term δ_t is a common factor with constant loadings across units, while λ_t represents a set of common factors with varying loadings across units. A difference-in-differences/fixed effects panel model can be obtained from equation (10) by restricting λ_t to be time invariant, so $\lambda_t = \lambda$ (see Bai 2009). This has the effect of restricting the mean outcomes of units with the same values for the observed predictors, $\mathbf{Z}_j = \boldsymbol{z}$, to follow parallel trends, $\delta_t + \theta_t \boldsymbol{z} + \lambda\mu_j$. A linear factor model provides a useful extension to the difference-in-differences/fixed effects panel data models by allowing Y_{jt}^N to depend on multiple unobserved components, μ_j, with coefficients, λ_t, that change in time. In contrast to difference in differences, the linear factor model does not impose parallel mean outcome trends for units with the same values for $\mathbf{Z}_j$.

Abadie, Diamond, and Hainmueller (2010) provide a characterization of the bias of the synthetic control estimator for the case when the synthetic control reproduces the characteristics of the treated unit. Let $\mathbf{X}_1$ be the vector that includes $\mathbf{Z}_1$ and the pre-intervention outcomes for the treated unit, and let $\mathbf{X}_0$ be the matrix that collects the same variables for the untreated units. Suppose that $\mathbf{X}_1 = \mathbf{X}_0 \boldsymbol{W}^*$, that is, the synthetic control represented by $\boldsymbol{W}^*$, is able to reproduce the characteristics of the treated unit (including the values of the pre-intervention outcomes). Then the bias of $\hat{\tau}_{it}$ is controlled by the ratio between the scale of the individual transitory shocks, ε_{it}, and the number of pre-intervention periods, T_0. The intuition behind this result is rather immediate. Under the factor model in equation (10), a synthetic control that reproduces the values $\mathbf{Z}_1$ and μ_1 would provide an unbiased estimator of the treatment effect for the treated. If $\mathbf{X}_1 = \mathbf{X}_0 \boldsymbol{W}^*$, then the synthetic control matches the value of $\mathbf{Z}_1$. On the other hand, μ_1 is not observed, so it cannot be matched directly in the data. However, a synthetic control that reproduces the values of $\mathbf{Z}_1$ but fails to reproduce the values of μ_1 can only provide a close match for the pretreatment outcomes if differences in the values of the individual transitory shocks between the treated and the synthetic controls compensate for the differences in unobserved factor loadings. This is unlikely to happen when the scale of the transitory shocks, ε_{it}, is small or the number of pretreatment periods, T_0, is large. In contrast, a small number of pre-intervention periods combined with enough variation in the unobserved transitory shocks may result in a close match for pretreatment outcomes even if the synthetic control does not closely match the values of μ_1. This is a form of over-fitting and a potential source of bias.

In practice, the condition $\mathbf{X}_1 = \mathbf{X}_0 \boldsymbol{W}^*$ is replaced by the approximate version $\mathbf{X}_1 \approx \mathbf{X}_0 \boldsymbol{W}^*$. It is important to notice, however, that for any particular data set there are not ex ante guarantees on the size of the difference $\mathbf{X}_1 - \mathbf{X}_0 \boldsymbol{W}^*$. When this difference is large, Abadie, Diamond, and Hainmueller (2010) recommend against the use of synthetic controls because of the potential for substantial biases. For the factor model in equation (10), obtaining a good fit $\mathbf{X}_1 \approx \mathbf{X}_0 \boldsymbol{W}^*$ when $\mathbf{X}_1$ and $\mathbf{X}_0$ include pre-intervention outcomes typically requires that the variance of the transitory shock is small (see Ferman and Pinto 2019). Moreover, because the bias bound depends inversely on T_0, one could erroneously conclude that under the factor model in equation (10), the synthetic control estimator is unbiased as T_0 goes to infinity. However, the bias bound in Abadie, Diamond, and Hainmueller (2010) is derived under $\mathbf{X}_1 = \mathbf{X}_0 \boldsymbol{W}^*$, and its practical relevance depends on the ability of the synthetic control to reproduce the trajectory of the outcome for the treated unit. Sizable biases may persist as $T_0 \to \infty$, unless the quality of the fit, $\mathbf{X}_1 - \mathbf{X}_0 \boldsymbol{W}^*$, is good. That is, the ability of a synthetic control to reproduce the trajectory of the outcome variable for the treated unit over an extended period of time, as in figure 1 panel B, provides an indication of low bias. However, a large T_0 cannot drive down the bias if the fit is bad. In practice, synthetic controls may not perfectly fit the characteristics of the treated units. Section 7 discusses a backdating exercise that can often be used to obtain an indication of the size and direction of the bias arising from imperfect fit.

The risk of over-fitting may also increase with the size of the donor pool, especially when T_0 is small. For any fixed T_0, a larger J makes it easier to fit pretreatment outcomes even when there are substantial discrepancies in factor loadings between the treated unit and the synthetic control. Consistent with this argument, the bias bound for $\hat{\tau}_{1t}$ derived in Abadie, Diamond, and Hainmueller (2010) depends positively on J. Under a factor model for Y_{it}^N, a large number of units in the donor pool may create or exacerbate the bias of the synthetic control estimator, especially if the values of μ_j in the donor pool greatly differ from μ_1.[9] Moreover, the factor model in equation (10) should be interpreted only as an approximation to a more general (nonlinear) process for Y_{it}^N. If the process that determines Y_{it}^N is nonlinear in the attributes of the units, even a close fit by a synthetic control, which is a weighted average, could potentially result in large interpolation biases.

A practical implication of the discussion in the previous paragraph is that each of the units in the donor pool have to be chosen judiciously to provide a reasonable control for the treated unit. Including in the donor pool units that are regarded by the analyst to be unsuitable controls (because of large discrepancies in the values of their observed attributes $\mathbf{Z}_j$ or because of suspected large differences in the values of the unobserved attributes μ_j relative to the treated unit) is a recipe for bias.

There are other factors that contribute to the bias bound in Abadie, Diamond, and Hainmueller (2010). In particular, the value of the bound increases with the number on unobserved factors, that is, the number of components in μ_j. The dependence of the bias bound on the number of unobserved factors is relevant for the discussion on the choice of predictors for the synthetic control method in the next subsection.

3.4 *Variable Selection*

A synthetic control provides a predictor of Y_{1t}^N for $t > T_0$, the potential outcome without the intervention for the treated units in a post-intervention period. Like for any other prediction procedure, the choice of predictors (in $\mathbf{X}_1$ and $\mathbf{X}_0$ for synthetic control estimators) is a fundamental part of the estimation task. This subsection discusses variable selection in the synthetic control method. To aid the discussion of the different issues involved in variable selection for synthetic controls, I will employ the concepts and notation of the linear factor model framework of subsection 3.3. Predictor variables in $\mathbf{X}_1$ and $\mathbf{X}_0$ typically include both pre-intervention values of the outcome variable as well as other predictors, $\mathbf{Z}_j$.

[9] A large J may be beneficial in high-dimensional settings, as demonstrated in Ferman (2019), who shows that under certain conditions synthetic control estimators may asymptotically unbiased as $T_0 \to \infty$ and $J \to \infty$.

Pre-intervention values of the outcome variable, which are naturally available in panel data settings, play a crucial role in reproducing the unobserved factor loadings in μ_j in the linear factor model of subsection 3.3. They also arise organically as predictors for the synthetic control estimators under a vector autoregression model for the process that generates the data (see Abadie, Diamond, and Hainmueller 2010). The credibility of a synthetic control estimator depends on its ability to track the trajectory of the outcome variable for the treated unit for an extended pre-intervention period. Provided that a good fit for pre-intervention outcomes is attained, the researcher has some flexibility in the way pre-intervention outcomes are incorporated in $\mathbf{X}_1$ and $\mathbf{X}_0$. Consider the German reunification application of subsection 3.2. As reported in table 1, the set of predictors in $\mathbf{X}_1$ and $\mathbf{X}_0$ includes average per capita GDP in 1981–90, and no other pre-intervention outcome. Notice, however, that the resulting synthetic control is able to track the trajectory of per capita GDP for West Germany for the entire 1960–90 pre-intervention period. This happens because per capita GDP figures for OECD countries strongly co-move in time across countries. This co-movement of the outcome variable of interest across the different units in the data is exactly what synthetic controls are designed to exploit. It makes it possible to match the entire trajectory of GDP per capita for West Germany by fitting only the average level of GDP per capita in the 1981–90 period. Given this premise, one potential advantage from using a summary measure of prereunification GDP per capita to calculate the synthetic control for West Germany (as opposed to, say, including all ten different annual values of GDP per capita for 1981–90 as predictors) resides in a higher sparsity of the resulting synthetic control. As will be discussed in section 4 below, the number of units in the donor pool that carry positive weights in a synthetic control estimator is controlled by the number of predictors in $\mathbf{X}_1$ and $\mathbf{X}_0$, and sparse synthetic controls (that is, synthetic controls made of a small number of comparison units) are easy to interpret and evaluate.

Part of the literature on synthetic controls emphasizes estimators that depend only on pre-intervention outcomes and ignore the information of other predictors, $\mathbf{Z}_j$. This reliance on pre-intervention outcomes only, while adopted in many cases for technical or expositional convenience, may create the mistaken impression that other predictors play a minor role in synthetic control estimators. Notice, however, that in equation (10), covariates excluded from $\mathbf{Z}_j$ are mechanically absorbed into μ_j, which increases the number of components of μ_j and, therefore, the bound on the bias as discussed in the previous subsection. By excluding $\mathbf{Z}_j$ from the set of predictors in $\mathbf{X}_1$ and $\mathbf{X}_0$, not only do we aim to implicitly match the values of μ_j through their effects on the pre-intervention outcomes, but also the values of $\mathbf{Z}_j$.

Data-driven methods for variable selection evaluate the predictive power of alternative sets of predictors. This can be done in the synthetic control method framework by measuring the predictive power of alternative sets of variables. The procedure divides the pre-intervention periods into an initial training period and a subsequent validation period. Synthetic control weights are computed using data from the training period only. The validation period can then be used to evaluate the predictive power of the resulting synthetic control. This procedure can be used to select predictors or to evaluate the predictive power of a given set of predictors as in section 7. Section 7 discusses how to assess the robustness of the results to alternative sets of predictors.

Finally, it is worth noting that post-intervention outcomes are not used in the calculation of synthetic control weights. This property of synthetic control methods can be exploited to provide guarantees against the use of results to guide specification searches. This is because synthetic control weights can be calculated using pre-intervention data only in the design phase of the study, before post-intervention outcomes are observed or realized. Section 4 discusses this issue in more detail.

3.5 *Inference*

Abadie, Diamond, and Hainmueller (2010) propose a mode of inference for the synthetic control framework that is based on permutation methods. In its simpler version, the effect on the intervention is estimated separately for each of the units in the sample. Consider the case with a single treated unit, as in subsection 3.1. A permutation distribution can be obtained by iteratively reassigning the treatment to the units in the donor pool and estimating "placebo effects" in each iteration. Then, the permutation distribution is constructed by pooling the effect estimated for the treated unit together with placebo effects estimated for the units in the donor pool. The effect of the treatment on the unit affected by the intervention is deemed significant when its magnitude is extreme relative to the permutation distribution.

One potential complication with this procedure is that, even if a synthetic control is able to closely fit the trajectory of the outcome variable for the treated unit before the intervention, the same may not be true for all the units in the donor pool. For this reason, Abadie, Diamond, and Hainmueller (2010) propose a test statistic that measures the ratio of the post-intervention fit relative to the pre-intervention fit. For $0 \leq t_1 \leq t_2 \leq T$ and $j = \{1, \ldots, J+1\}$, let

$$(11) \quad R_j(t_1, t_2) = \left(\frac{1}{t_2 - t_1 + 1} \sum_{t=t_1}^{t_2} \left(Y_{jt} - \hat{Y}_{jt}^N \right)^2 \right)^{1/2},$$

where $\hat{Y}_{jt}^N$ is the outcome on period t produced by a synthetic control when unit j is coded as treated and using all other J units to construct the donor pool. This is the root mean squared prediction error (RMSPE) of the synthetic control estimator for unit j and time periods $t_1, \ldots, t_2$. The ratio between the post-intervention RMSPE and pre-intervention RMSPE for unit j is

$$(12) \quad r_j = \frac{R_j(T_0 + 1, T)}{R_j(1, T_0)}.$$

That is, r_j measures the quality of the fit of a synthetic control for unit j in the posttreatment period, relative to the quality of the fit in the pretreatment period. Abadie, Diamond, and Hainmueller (2010) use the permutation distribution of r_j for inference. An alternative solution to the problem of poor pretreatment fit in the donor pool is to base inference on the distribution $R_j(T_0 + 1, T)$ after discarding those placebo runs with $R_j(1, T_0)$ substantially larger than $R_1(1, T_0)$ (see Abadie, Diamond, and Hainmueller 2010).

A p-value for the inferential procedure based on the permutation distribution of r_j, as described above, is given by

$$p = \frac{1}{J+1} \sum_{j=1}^{J+1} I_+ \left(r_j - r_1 \right),$$

where $I_+(\cdot)$ is an indicator function that returns one for nonnegative arguments and zero otherwise. While p-values are often used to summarize the results of testing procedures, the permutation distribution of the test statistics, r_j, or of the placebo gaps, $Y_{jt} - \hat{Y}_{jt}^N$, are easy to report/visualize and provide additional information on (e.g., on the magnitude of the differences between the estimated treatment effect on the treated unit and the placebo gaps in the donor pool). Confidence intervals can be constructed by test inversion (see, e.g., Firpo and Possebom 2018).

Replacing $Y_{jt} - \hat{Y}_{jt}^N$ in $R(T_0+1,T)$ with their positive or negative parts, $\left(Y_{jt} - \hat{Y}_{jt}^N\right)^+$ or $\left(Y_{jt} - \hat{Y}_{jt}^N\right)^-$, leads to one-sided inference. One-sided inference may result in a substantial of gain of power.[10] This is an important consideration in many comparative case study settings, where samples are considerably small. Alternative test statistics (see, e.g., Firpo and Possebom 2018) could potentially be used to direct power to specific sets of alternatives.

As discussed in Abadie, Diamond, and Hainmueller (2010) this mode of inference reduces to classical randomization inference (Fisher 1935) when the intervention is randomly assigned, a rather improbable setting, especially in contexts with aggregate units. More generally, this mode of inference evaluates significance relative to a benchmark distribution for the assignment process, one that is implemented directly in the data. Abadie, Diamond, and Hainmueller (2010) use a uniform benchmark, but one could easily depart from the uniform case. Firpo and Possebom (2018) propose a sensitivity analysis procedure that considers deviations from the uniform benchmark.

Because in most observational settings assignment to the intervention is not randomized, one could, in principle, adopt permutation schemes that incorporate information in the data on the assignment probabilities for the different units in the sample (as in, e.g., Rosenbaum 1984). However, in many comparative case studies it is difficult to articulate the nature of a plausible assignment mechanism or even the specific nature of a placebo intervention. Consider, for example, the 1990 German reunification application in Abadie, Diamond, and Hainmueller (2015). In that context, it would be difficult to articulate the nature of the assignment mechanism or even describe placebo interventions. (France would reunify with whom?) Moreover, even if a plausible assignment mechanism exists, estimation of the assignment mechanism is often hopeless because many comparative case studies feature a single or a small number of treated units.

It is important to note that the availability of a well-defined procedure to select the comparison unit, like the one provided by the synthetic control method, makes the estimation of the effects of placebo interventions feasible. Without a formal description of the procedure used to choose the comparison for the treated unit, it would be difficult to reapply the same estimation procedure to the units in the donor pool. In this sense, the formalization of the choice of the comparison unit provided by the synthetic control method opens the door to a mode of quantitative inference in the context of comparative case studies.

Another important point to notice is that the permutation method described in this subsection does not attempt to approximate the sampling distributions of test statistics. Sampling-based statistical tests employ restrictions on the sampling mechanism (data-generating process) to derive a distribution of a test statistic in a thought experiment where alternative samples could have been obtained from the sampling mechanism that generated the data. In a comparative case study framework, however, sampling-based inference is complicated—sometimes because of the absence of a well-defined sampling mechanism or data-generating process, and sometimes because the sample is the same as the population. For example, in their study of the effect of terrorism on

[10] Notice that, in the presence of a treatment effect on the treated unit, permutations in which the treated unit contributes to the placebo synthetic control will tend to produce effects of the opposite sign to the effect on the treated unit, increasing the power of the one-sided test.

economic outcomes in Spain, Abadie and Gardeazabal (2003) employ a sample consisting of all Spanish regions. Here, sampling is not done at random from a well-defined super-population. As in classical randomization tests (Fisher 1935), design-based inference takes care of these complications by conditioning on the sample and considering only the variation in the test statistic that is induced by the assignment mechanism (see, e.g., Abadie et al. 2020).[11]

4. *Why Use Synthetic Controls?*

In this section, I will describe some advantages of synthetic control estimators relative to alternative methods. For the sake of concreteness and because linear regression is arguably the most widely applied tool in empirical research in economics, I emphasize the differences between synthetic control estimators and linear regression estimators. However, much of the discussion applies more generally to other estimators of treatment effects.

A linear regression estimator of the effect of the treatment can easily be constructed using the panel data structure described in subsection 3.1. Let $\boldsymbol{Y}_0$ be the $(T-T_0)\times J$ matrix of post-intervention outcomes for the units in the donor pool with (t,j)-element equal to $Y_{T_0+t,J+1}$. Let $\bar{\boldsymbol{X}}_1$ and $\bar{\boldsymbol{X}}_0$ be the result of augmenting $\boldsymbol{X}_1$ and $\boldsymbol{X}_0$, respectively, with a row of ones. For non-singular $\bar{\boldsymbol{X}}_0\bar{\boldsymbol{X}}_0'$, a regression-based estimator of the counterfactual Y_{1t}^N for $t>T_0$ is $\hat{\boldsymbol{B}}'\bar{\boldsymbol{X}}_1$, where $\hat{\boldsymbol{B}}=\left(\bar{\boldsymbol{X}}_0\bar{\boldsymbol{X}}_0'\right)^{-1}\bar{\boldsymbol{X}}_0\boldsymbol{Y}_0'$. That is, the regression-based estimator is akin to a synthetic control, as it uses a linear combination, $\boldsymbol{Y}_0\boldsymbol{W}^{reg}$, of the outcomes in the donor pool, with $\boldsymbol{W}^{reg}=\bar{\boldsymbol{X}}_0'\left(\bar{\boldsymbol{X}}_0\bar{\boldsymbol{X}}_0'\right)^{-1}\bar{\boldsymbol{X}}_1$, to reproduce the outcome of the treated unit in the absence of the intervention. Some advantages of synthetic controls relative to regression-based counterfactual are listed next.

No Extrapolation. Synthetic control estimators preclude extrapolation, because synthetic control weights are nonnegative and sum to one. It is easy to check that, like their synthetic control counterparts, the regression weights in $\boldsymbol{W}^{reg}$ sum to one. Unlike the synthetic control weights, however, regression weights may be outside the $[0,1]$ interval, allowing extrapolation outside of the support of the data (see Abadie, Diamond, and Hainmueller 2015 for details).[12] Table 3 reports regression weights for the German reunification example. In this application, the regression counterfactual utilizes negative values for four countries.

Transparency of the Fit. Linear regression uses extrapolation to guarantee a perfect fit of the characteristics of the treated unit, $\bar{\boldsymbol{X}}_0\boldsymbol{W}^{reg}=\bar{\boldsymbol{X}}_1$ (and, therefore, $\boldsymbol{X}_0\boldsymbol{W}^{reg}=\boldsymbol{X}_1$) even when the untreated units are completely dissimilar in their characteristics to the treated unit. In contrast, synthetic controls make transparent the actual discrepancy between the treated unit and the convex combination of untreated units that provides the counterfactual of interest, $\boldsymbol{X}_1-\boldsymbol{X}_0\boldsymbol{W}^*$. This discrepancy is equal to the difference between columns 1 and 2 in table 1. In addition figure 1, panel B, brings to light the fit of a synthetic control in terms of pre-intervention outcomes. That is, the information in table 1 and figure 1 makes clear the extent to which the observations in the donor pool can approximate the characteristics of the treated units by interpolation only. In some applications, comparisons

[11] In particular, this is in contrast to the bias bound calculations in Abadie, Diamond, and Hainmueller (2010), which are performed over the distribution of the individual transitory shocks, ε_{it}.

[12] See King and Zeng (2006) on the dangers of relying on extrapolation to estimate counterfactuals.

TABLE 3
Regression Weights for West Germany

Australia	0.12
Austria	0.26
Belgium	0.00
Denmark	0.08
France	0.04
Greece	−0.09
Italy	−0.05
Japan	0.19
Netherlands	0.14
New Zealand	0.12
Norway	0.04
Portugal	−0.08
Spain	−0.01
Switzerland	0.05
United Kingdom	0.06
United States	0.13

like that of columns 1 and 2 of table 1 may reveal that it is not possible to approximate the characteristics of the treated unit(s) using a weighted average of the units in the donor pool. In that case, Abadie, Diamond, and Hainmueller (2010, 2015) advise against using synthetic controls.

Safeguard against Specification Searches. In contrast to regression but similar to classical matching methods, synthetic controls do not require access to posttreatment outcomes in the design phase of the study, when synthetic controls are calculated. This implies that all the data analysis on design decisions like the identity of the units in the donor pool or the predictors in $\mathbf{X}_1$ and $\mathbf{X}_0$ can be made without knowing how they affect the conclusions of the study (see Rubin 2007 for a related discussion regarding matching estimators). Moreover, synthetic control weights can be calculated and preregistered/publicized before the posttreatment outcomes are realized, or before the actual intervention takes place. That is, preregistration of synthetic control weights can play a role similar to pre-analysis plans in randomized control trials (see, e.g., Olken 2015), providing a safeguard against specification searches and p-hacking.

Transparency of the Counterfactual. Synthetic controls make explicit the contribution of each comparison unit to the counterfactual of interest. Moreover, because the synthetic control coefficients are proper weights and are sparse (more on sparsity below), they allow a simple and precise interpretation of the nature of the estimate of the counterfactual of interest. For the application to the effects of the German reunification in table 2, the counterfactual for West Germany is given by a weighted average of Austria (0.42), Japan (0.16), the Netherlands (0.09), Switzerland (0.11), and the United States (0.22) with weights in parentheses. Simplicity and transparency of the counterfactual allows the use of the expert knowledge to evaluate the validity of a synthetic control and the directions of

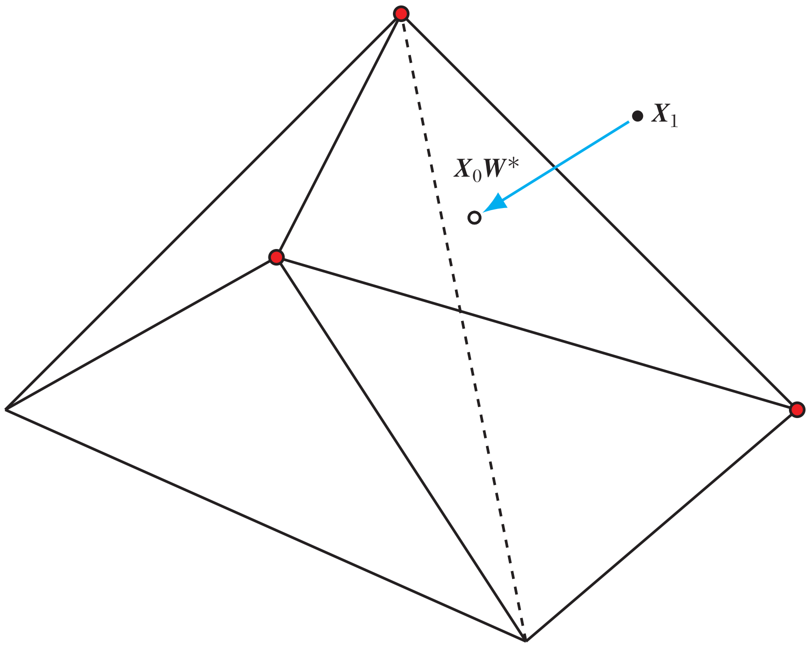

Figure 2. Projecting $\mathbf{X}_1$ on the Convex Hull of $\mathbf{X}_0$

potential biases. For instance, smaller neighboring countries to West Germany, such as Austria, the Netherlands, and Switzerland, have a substantial weight on the composition of the synthetic control of table 2. If economic growth in these countries were negatively affected by the German reunification during the 1990–2003 period (perhaps because West Germany diverted demand and investment from these countries to East Germany), this would imply that figure 1, panel B, estimates a lower bound on the magnitude (absolute value) of the negative effect of the German reunification on per capita GDP in West Germany.

Sparsity. As evidenced in the results of tables 2 and 3, synthetic controls are sparse, but regression weights are not. As discussed above, sparsity plays an important role for the interpretation and evaluation of the estimated counterfactual. The sparsity of synthetic control weights has an immediate geometric interpretation. Assume, for now, that $\mathbf{X}_1$ falls outside the convex hull of the columns of $\mathbf{X}_0$. This is typical in empirical practice and a consequence of the curse of dimensionality. Assume also that the columns of $\mathbf{X}_0$ are in general position (that is, there is no set of m columns, with $2 \leq m \leq k+1$, that fall into an $(m-2)$-dimensional hyperplane). Then, the synthetic control is unique and sparse—with the number of nonzero weights bounded by k—as it is the projection of $\mathbf{X}_1$ on the convex hull of the columns of $\mathbf{X}_0$. Figure 2 provides a visual representation of the geometric interpretation of the sparsity property of synthetic control estimators.

Only the control observations marked in red contribute to the synthetic control.

Notice that table 1 indicates that the synthetic control for West Germany falls close to but outside the convex hull of the values of economic growth predictors in the donor pool (otherwise, columns 1 and 2 would be identical). As a result, the number of nonzero weights in table 2 is not larger than the number of variables in table 1. If desired, sparsity can be increased by imposing a bound on the density (number of nonzero weights) of $\mathbf{W}^*$ in the calculation of synthetic controls (see Abadie, Diamond, and Hainmueller 2015).

In some cases, especially in applications with many treated units, the values of the predictors for some of the treated units may fall in the convex hull of the columns of $\mathbf{X}_0$. Then, synthetic controls are not necessarily unique or sparse. That is, a minimizer of equation (7) may not be unique or sparse, although sparse solutions with no more than $k+1$ nonzero weights always exist. A question is then how to choose among the typically infinite number of solutions to the minimization of equation (7). A modification of the synthetic control estimator in Abadie and L'Hour (2019) discussed in section 8 addresses this problem and produces synthetic controls that are unique and sparse (provided that untreated observations are in general quadratic position, see Abadie and L'Hour 2019 for details). In contrast, as shown in table 3, regression estimators are typically not sparse.

It is important to notice that the role of sparsity in the context of synthetic control methods differs from the usual role that sparsity plays in other statistical methods like the lasso, where a sparsity-inducing regularization is employed to prevent over-fitting, and where the interpretation of the lasso coefficients is often not at issue. Like for the lasso, the goal of synthetic controls is out-of-sample prediction; in particular, prediction of Y^N_{1t} for $t > T_0$. In contrast to the lasso, however, the identity and magnitude of nonzero coefficients constitute important information to interpret the nature of the estimate and evaluate its validity and the potential for biases.

One of the greatest appeals of the synthetic control method resides, in my opinion, in the interpretability of the estimated counterfactuals, which results from the weighted average nature of synthetic control estimators and from the sparsity of the weights.

Despite the practical advantages of synthetic control methods, successful application of synthetic control estimators crucially depends on important contextual and data requirements, which are discussed in the next two sections.

5. *Contextual Requirements*

This section will discuss contextual requirements, that is, the conditions on the context of the investigation under which synthetic controls are appropriate tools for policy evaluation, as well as suitable ways to modify the analysis when these conditions do not perfectly hold. It is important, however, to point out that most of the requirements listed in this section pertain not only to synthetic control methods, but also to any other type of comparative case study research design.

Size of the Effect and Volatility of the Outcome. As previously discussed, the goal of comparative case studies is to estimate the effect of a policy intervention on the unit (e.g., state or region) exposed to an intervention of interest. That is, comparative case studies typically estimate the effect of an intervention on a single treated unit or on a small number of treated units. The nature of this exercise indicates that small effects will be indistinguishable from other shocks to the outcome of the affected unit, especially if the outcome variable of interest is highly volatile.[13] As a result, the impact of "small" interventions with effects of a magnitude similar to the volatility of the outcome are difficult to detect. Even a large effect may be difficult to detect if the volatility of the outcome is also large. Outcome variables that include substantial random noise elevate the risk of over-fitting, as explained in subsection 3.3. In cases where substantial volatility is present in the outcome of interest it is advisable to remove it via filtering, in both the exposed unit as well as in the units in the donor pool, before applying synthetic control techniques.[14] Notice, however, that the challenge posed by volatility comes only from the fraction of it that is generated by unit-specific factors (e.g., the individual-specific transitory shocks, ε_{jt}, in equation (10)). Volatility generated by common factors affecting other units (e.g., the common factors λ_t in equation (10)) can be differentiated out by choosing an appropriate synthetic control.

Availability of a Comparison Group. The very nature of comparative case studies implies that inference based on these methods will be faulty in the absence of a suitable comparison group. First and foremost, in order to have units available for the donor pool, it is important that not all units adopt interventions similar to the one under investigation during the period of the study. Units that adopt an intervention similar to the one adopted by the unit of interest should not be included in the donor pool because they are affected by the intervention, very much like the unit of interest. It is also important to eliminate from the donor pool any units that may have suffered large idiosyncratic shocks to the outcome of interest during the study period, if it is judged that such shocks would not have affected the outcome of the unit of interest in the absence of the intervention.[15] Moreover, it is important to restrict the donor pool to units with characteristics that are similar to the affected unit. The reason is that, while the restrictions placed on the weights, $\boldsymbol{W}$, do not allow extrapolation, interpolation biases may still be important if the synthetic control matches the characteristics of the affected unit by averaging away large discrepancies between the characteristics of the affected unit and the characteristics of the units in the synthetic control. For the German reunification example, Abadie, Diamond, and Hainmueller (2015) restrict the donor pool to a set of OECD economies. Related to this point, Abadie and L'Hour (2019) propose adding to the objective function in equation (7) a set of penalty terms that depend on the discrepancies between the characteristics of the affected unit and the characteristics of the individual units included in the synthetic control (see section 8 for details).

No Anticipation. As in any research design that exploits time variation in the outcome variable to estimate the effect of an intervention, synthetic control estimators may be biased if forward-looking economic agents react in advance of the policy intervention under investigation, or if certain components of the intervention are put in place in advance of the formal implementation/enactment of the intervention. If there

[13] In studies that seek to estimate the average effect of an intervention that is observed in a large number of instances, the volatility of the outcome variable can often be reduced by averaging. In contrast, as explained above, comparative case studies often focus on the effect of a single event or intervention.

[14] For example, Amjad, Shah, and Shen (2018) propose singular value thresholding to de-noise data for synthetic controls.

[15] As an example, in their study of the effect of California's tobacco control legislation, Abadie, Diamond, and Hainmueller (2010) discard from the donor pool several states that adopted large-scale tobacco programs or substantially increased taxes on tobacco during the sample period of the study.

are signs of anticipation, it is advisable to backdate the intervention in the data set to a period before any anticipation effect can be expected, so the full extent of the effect of the intervention can be estimated. Notice that backdating the intervention in the data does not mechanically bias the estimator of the effect of the intervention even if some periods before the intervention are mistakenly recorded as post-intervention periods. The reason is that, as shown in equations (2) and (3), the synthetic control estimator does not restrict the time variation in the effect of the intervention. Therefore, periods barely affected by the intervention may show small or zero effects, while subsequent periods may produce a large estimated effect. This is in contrast with much of the practice using panel data models, where in many instances the effect of an intervention is restricted to be constant across post-intervention periods.

No Interference. In the setup of subsection 3.1, we defined the potential outcomes Y_{1t}^I and Y_{it}^N only in terms of the treatment status for unit 1 and unit i, respectively, at time t. This is the stable unit treatment value assumption in Rubin (1980), which implies that there is no interference across units. That is, units' outcomes are invariant to other units' treatments. In some instances, however, an intervention may have spillover effects on units that are not directly targeted by it. Assuming that such spillover effects do not exist is a strong restriction that must often be enforced in the design of the study or accounted for in the analysis of the results.

The assumption of no interference can be enforced in the design of a study by discarding from the donor pool those units with outcomes possibly affected by the intervention on the treated unit. Notice that there is a potential tension between this practice and the issues discussed in *Availability of a Comparison Group*. On the one hand, it is advisable to select for the donor pool units that are affected by the same regional economic shocks as the unit where the intervention happens. On the other hand, if spillover effects are substantial and affect units in close geographical proximity, those units may provide a biased estimate of the counterfactual outcome without intervention for the unit affected by the intervention. In cases when units potentially affected by spillover effects are discarded from the donor pool, the transparency of the fit of synthetic controls allows researchers to evaluate the reduction in the quality of the match between the characteristics of the treated unit and the characteristics of the synthetic control.

Potential spillover effects can also be accounted for in the analysis phase of a synthetic control study. If units affected by spillover effects are included in the synthetic control, the researcher should be aware of the potential direction of the bias of the resulting estimator. For example, Abadie, Diamond, and Hainmueller (2015) estimate the economic impact of the 1990 German reunification using a synthetic control of other OECD countries to approximate the trajectory of the counterfactual per capita GDP for West Germany in the absence of the unification. As explained above, if countries that compose the synthetic control for West Germany, like Austria, suffered from the negative effects of the German reunification, then we would expect the synthetic control estimator to be attenuated. That is, in this case, the synthetic control estimate would provide a lower bound on the magnitude of the causal effect of the German reunification on GDP per capita in West Germany. Notice that it is the transparency of the counterfactual and sparsity of the synthetic control counterfactual estimate that makes this exercise possible. In regression settings, like the one in section 4, the weight of each unit in the counterfactual estimate is rarely computed in empirical practice, and non-sparse weight vectors often including negative components complicate the evaluation of potential biases.

Convex Hull Condition. Synthetic control estimates are predicated on the idea that a combination of unaffected units can approximate the pre-intervention characteristics of the affected unit. Once the synthetic control is constructed, it should be checked that the differences in the characteristics of the affected unit and the synthetic control are small, that is:

$$X_{11} - w_2 X_{12} - \cdots - w_J X_{1J+1} \approx 0, \cdots,$$

$$X_{k1} - w_2 X_{k2} - \cdots - w_J X_{kJ+1} \approx 0.$$

In mathematical parlance, we need $(X_{11}, X_{21}, \ldots, X_{k1})$ to fall close to the convex hull of the set of points $\left\{(X_{12}, X_{22}, \ldots, X_{k2}), \ldots, \left(X_{1J+1}, X_{2J+1}, \ldots, X_{kJ+1}\right)\right\}$. If the unit affected by the intervention of interest is "extreme" in the value of a particular variable, such a value may not be closely approximated by a synthetic control.[16]

The fact that the value of a particular predictor for the treated unit cannot be closely approximated by the synthetic control may be less of a concern if the synthetic control closely tracks the trajectory of the outcome variable for the unit affected by the intervention during a hold-out validation period. In some cases, however, the unit affected by the intervention of interest may be extreme in the values of the outcome variable before the intervention and, as a result, there will not be a weighted average of untreated units that can approximate the trajectory of the outcome variable for the treated unit before the intervention. A potential way to proceed in those cases is to transform the outcome to time differences, $\Delta Y_{jt} = Y_{jt} - Y_{jt-1}$, or growth rates, $100 \times \Delta Y_{jt}/Y_{jt-1}$. Similarly, one could measure outcomes in differences with respect to pre-intervention means, $\tilde{Y}_{jt} = Y_{jt} - (1/T_0)\sum_{h=1}^{T_0} Y_{jh}$ (Ferman and Pinto 2019). Consider the particular case where a synthetic control is calculated on the basis of all pre-intervention outcomes (equally weighted). That is, each of the T_0 rows of $\boldsymbol{X}_1$ and $\boldsymbol{X}_0$ contains the outcome values in one of the pre-intervention periods for the treated unit and the donor pool, respectively, and all pre-intervention periods carry the same weight in the calculation of the synthetic control. For this particular case, measuring the outcomes in $\boldsymbol{X}_1$ and $\boldsymbol{X}_0$ in deviations with respect to the units' pre-intervention means is equivalent to a proposal in Doudchenko and Imbens (2016), who measure outcomes in levels but allow for a constant shift in the synthetic control fit, $\boldsymbol{X}_1 - \alpha \boldsymbol{I}_{k\times 1} - \boldsymbol{X}_0 \boldsymbol{W}$ (with tth row equal to $Y_{1t} - \alpha - w_2 Y_{2t} - \cdots - w_{J+1} Y_{J+1t}$), where $\boldsymbol{I}_{k\times 1}$ is a vector of ones of the same dimension as $\boldsymbol{X}_1$. As explained in Doudchenko and Imbens (2016), allowing for a constant shift between $\boldsymbol{X}_1$ and $\boldsymbol{X}_0 \boldsymbol{W}$ makes little sense in more general settings, when multiple covariates of different scales, instead of pre-intervention outcomes only, are included in $\boldsymbol{X}_1$ and $\boldsymbol{X}_0$. Notice, however, that measuring outcomes in deviations with respect to their pre-intervention means instead of allowing for a constant shift between $\boldsymbol{X}_1$ and $\boldsymbol{X}_0$, $\boldsymbol{W}$ may still be useful to account for differences in the level of the outcomes across units, even if $\boldsymbol{X}_1$ and $\boldsymbol{X}_0$ include other predictors aside from pre-intervention outcomes.

Transformations of the outcome variable like those in the previous paragraph may be useful in some cases because, as evidenced in the vast difference-in-differences literature, there are instances when a comparison

[16]For example, Abadie, Diamond, and Hainmueller (2015) find that because inflation levels were particularly low for West Germany before the reunification, the value of this variable cannot be closely reproduced by a synthetic control composed by other OECD countries. See table 1.

group can reproduce the changes in the outcome variable for the unit of interest even if the level of the outcome variable cannot be reproduced. In other cases, however, credible counterfactuals require reproducing not only the trend of the outcome variable for the treated but also the level. For example, some formulations of the convergence hypothesis in economic growth imply that countries with different levels of per capita GDP will tend to experience different growth rates on average, in the absence of an intervention. Similarly, nonlinearities in labor earnings profiles over the life cycle imply that differences in the age distribution across populations will typically result in differences in the growth of labor earnings.[17] More generally, there may not exist a combination of untreated units that provide a credible approximation to the treated units, and the conventional synthetic control estimator should not be used in that case.

It should also be noted that differencing the dependent variable may result in a substantial increase in the part of the variance of the outcome that is attributable to noise, potentially inducing an increase in bias. As an example, consider the linear factor model in equation (10). Differencing equation (10) we obtain

$$\Delta Y_{jt}^N = \Delta\delta_t + \Delta\boldsymbol{\theta}_t \mathbf{Z}_j + \Delta\boldsymbol{\lambda}_t \boldsymbol{\mu}_j + \Delta\varepsilon_{jt},$$

where $\Delta Y_{jt}^N = Y_{jt}^N - Y_{jt-1}^N$ with analogous expressions for $\Delta\delta_t$, $\Delta\boldsymbol{\theta}_t$, $\Delta\boldsymbol{\lambda}_t$, and $\Delta\varepsilon_{jt}$. Notice first that the differenced equation retains the linear factor structure. Notice also that differencing the outcome may help control the bias when the vectors of common factors $\boldsymbol{\theta}_t$ and $\boldsymbol{\lambda}_t$, or at least some of their components, vary little in time. In that case, the magnitudes of $\Delta\boldsymbol{\theta}_t$ and $\Delta\boldsymbol{\lambda}_t$ may be small even if the magnitudes of $\boldsymbol{\theta}_t$ and $\boldsymbol{\lambda}_t$ are large. This is the usual rationale for working with differenced outcomes and the basis for difference-in-differences estimators. There may be opposing forces at play, however. Suppose, in particular, that the idiosyncratic shocks, ε_{jt}, are independent or roughly independent in time. Then, the variance of $\Delta\varepsilon_{jt}$ is larger than the variance of ε_{jt}. Now, following the characterization of the bias in subsection 3.3, a larger residual variance may result in a higher risk of over-fitting and an increase in the bias of the synthetic control estimator.

Time Horizon. The effect of some interventions may take time to emerge or to be of sufficient magnitude to be quantitatively detected in the data. An obvious but unsatisfying approach to this problem is to wait until the effects of the intervention run their course. A more proactive approach is to use surrogate outcomes or leading indicators of the outcome variable of interest.

6. *Data Requirements*

This section discusses data requirements for credible applications of synthetic controls. Like many of the contextual requirements in the previous section, the data requirements discussed here apply not only to synthetic control estimation but, more generally, to comparative case study methods.

Aggregate Data on Predictors and Outcomes. From the previous discussion, it can be seen that the synthetic control method requires the availability of data on outcomes and predictors of the outcome for the unit or units exposed to the intervention of interest and a set of comparison units. Predictors and outcomes are often series reported by

[17] Notice that for nonlinearities in the process that generates Y_{it}^N may require that, for each unit j contributing to the synthetic control, $\mathbf{X}_j$ is reasonably close to $\mathbf{X}_1$. Section 8 describes a synthetic control estimator with weights that penalize $\|\mathbf{X}_1 - \mathbf{X}_j\|$.

government agencies, multilateral organizations, and private entities. Examples of these types of outcomes are state-level crime rates in the United States (Donohue, Aneja, and Weber 2019), country-level per capita GDP (Abadie, Diamond, and Hainmueller 2015), and state-level cigarette consumption statistics in the United States (Abadie, Diamond, and Hainmueller 2010), which are routinely reported in publications produced or commissioned by the Federal Bureau of Investigation, the World Bank, and tobacco industry groups, respectively. Sometimes, when aggregate data do not exist, aggregates of micro-data are employed in comparative case studies. For example, in his study of the labor market effects of the Mariel Boatlift in Miami, Card (1990) uses micro-data from the Current Population Survey (CPS) to estimate aggregate values for wage rates and unemployment for workers in Miami and a set of four comparison cities before and after the Mariel Boatlift. Similarly, Bohn, Lofstrom, and Raphael (2014) use data from the CPS to estimate the fraction of the population composed of Hispanic noncitizens by state in the United States.

Sufficient Pre-intervention Information. The credibility of a synthetic control estimator depends in great part on its ability to steadily track the trajectory of the outcome variable for the affected unit before the intervention. As discussed in subsection 3.3, Abadie, Diamond, and Hainmueller (2010) show that if the data-generating process follows a linear factor model, then the bias of the synthetic control estimator is bounded by a function that is inversely proportional to the number of pre-intervention periods (provided that the synthetic control closely tracks the trajectory of the outcome variable for the affected unit during the pre-intervention periods). Therefore, when designing a synthetic control study, it is of crucial importance to collect information on the affected unit and the donor pool for a large pre-intervention window.

A caveat to the preference for a large number of pre-intervention periods is given by the possibility of structural breaks. Consider the linear factor model of equation (10). In this model, structural stability is represented by the restriction of constant factor loadings. Even if the model is a good representation of the distribution of the data at a relatively short time scale, its accuracy may suffer once we allow the number of periods to be large enough. Choosing $v_1, \ldots, v_k$ to up-weight the most recent measures (relative to the prediction window) included in $\mathbf{X}_1$ and $\mathbf{X}_0$ helps alleviate structural instability concerns.

With a small number of pre-intervention periods, close or even perfect fit of the predictor values for the treated unit may be spuriously attained, in which case the resulting synthetic control may fail to reproduce the trajectory of the outcome for the treated unit in the absence of the intervention. The severity of this problem can be diminished if powerful predictors of post-intervention values of Y_{jt}^N, aside from pre-intervention values of the outcome, are included in $\mathbf{X}_j$, reducing the residual variance and, as a result, the risk of over-fitting.

Sufficient Post-intervention Information. This data requirement derives partly from the *Time Horizon* contextual requirement in section 5. The evaluation data must include outcome measures that are possibly affected by the intervention and are relevant for the policy decision or scientific inquiry that is the object of the study. This may be problematic if the effect of an intervention is expected to arise gradually over time and if no forward-looking measures of the outcome are available. Conversely, in some practical instances, the effect of an intervention may dissipate rapidly after showing substantial effectiveness for a few initial periods.

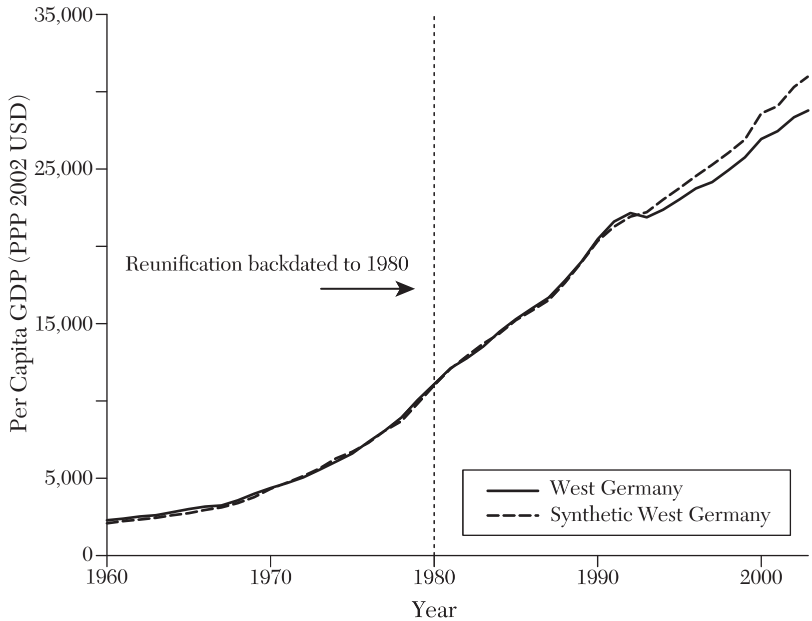

Figure 3. Backdating the 1990 German Reunification Application

Extensive post-intervention information allows a more complete picture of the effects of the intervention, in time and across the various outcomes of interest.

7. *Robustness and Diagnostic Checks*

The credibility of a synthetic control estimator depends on its ability to reproduce outcomes for the treated unit in the absence of the intervention. This section presents diagnostic checks that can be used to evaluate the credibility of synthetic control counterfactuals in actual applications, as well as robustness exercises to assess sensitivity of results to changes in the design of the study.

Backdating. The possibility of backdating was discussed in section 5 as a way to address anticipation effects on the outcome variable before an intervention occurs. In the absence of anticipation effects, the same idea can be applied to assess the credibility of a synthetic control in concrete empirical applications. Figure 3 shows the result of estimating the effect of the 1990 German reunification with the intervention backdated to 1980. Two important features of the results are as follows. First, as one would hope, the synthetic control estimator closely tracks per capita GDP for West Germany in the 1981–90 period, before the start of the actual intervention. This is the"in-time placebo test" in Abadie, Diamond, and Hainmueller (2015) and similar to the "preprogram test" in Heckman and Hotz (1989). The absence of estimated effects prior to the intervention provides credibility of the synthetic control estimator, as it demonstrates that the synthetic control is able to reproduce the trajectory of the outcome variable for the treated unit before the intervention occurs. Second, a gap between per capita GDP for West Germany and its synthetic control counterpart appears around the time of the German reunification, as in figure 1, panel B. This is the case even when the intervention is ten-year backdated in the data and the procedure uses no information on the timing of the actual intervention. The shape and direction of the gap in figure 3 is similar to that of figure 1, panel B, albeit of a somewhat smaller magnitude. The fact that the estimated effect of the German reunification appears shortly after 1990 even when the intervention is artificially ten-year backdated in the data provides credibility to the synthetic control estimator of the 1990 German reunification.

Robustness Tests. Regardless of the estimation method employed in the analysis, the main conclusions of an empirical study should display some level of robustness with respect to changes in the study design. In the context of synthetic controls, two important ways the design of a study may influence results are (i) the choice of units in the donor pool, and (ii) the choice of predictors of the outcome variable. The first choice corresponds to the columns in $\mathbf{X}_0$, and the second one corresponds to the rows in $[\mathbf{X}_1 : \mathbf{X}_0]$.

As a example of a robustness test, figure 4 reports the results of a leave-one-out re-analysis of the German reunification data in Abadie, Diamond, and Hainmueller (2015), taking from the sample one-at-a-time each of the countries that contribute to the synthetic control in table 2. All leave-one-out estimates closely track the per capita GDP series for West Germany before 1990. The resulting estimates for the years after the reunification are all negative and centered around the result produced using the entire donor pool. The main conclusion of a negative estimate of the German reunification on per capita GDP is robust to the exclusion of any particular country. In other examples, however, results may not be as robust as those in figure 4, and the scientific significance of the estimates should be evaluated with that information in mind. If the exclusion of a unit from the donor pool has a large effect on results without a discernible change in pre-intervention fit, this may warrant investigating if the change in the magnitude of the estimate is caused by the effects of other interventions or by particularly large idiosyncratic shocks on the outcome of the excluded untreated unit.

8. *Extensions and Related Methods*

As the literature on synthetic control methods and related methods has greatly expanded in recent years, it has become increasingly difficult for researchers interested in applying these methods to figure out what is available where. In this section, I provide a brief guide to the recent contributions in the area. This represents only an incomplete snapshot of a literature that is rapidly evolving.

Multiple Treated Units. Several recent articles consider estimation and inference with synthetic controls for the case where there are multiple treated units. Notice that the presence of multiple treated units does not give rise to additional conceptual challenges for the estimation of synthetic controls. Treatment effects can be estimated for each treated unit separately and aggregated in a second step if desired. However, the presence of multiple treated units creates some practical problems for estimation, as well as new challenges and opportunities for inference.

A potential complication with synthetic control estimation is that the minimizer of equation (7) subject to the weight constraints may not be unique, especially if the values of the predictors for a treated unit fall

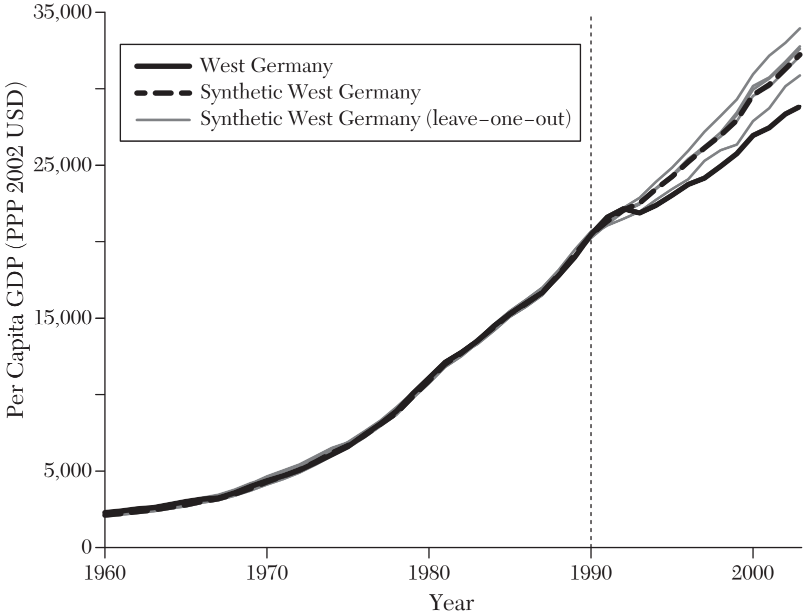

Figure 4. Leave-one-out Estimates of the Effect of the 1990 German Reunification

inside the convex hull of the values of the predictors for the donor pool. Suppose, for now, that only the first unit is treated. If $\mathbf{X}_1$ belongs to the convex hull of the columns of $\mathbf{X}_0$, this implies that we can find $\mathbf{W}^*$ such that $\mathbf{X}_1 = \mathbf{X}_0 \mathbf{W}^*$. Moreover, the number of minimizers to equation (7) may be (and will typically be) infinite.[18] That is, there may exist an infinite number of solutions to the problem of minimizing of equation (7), subject to the weight constraints, that perfectly reproduce $\mathbf{X}_1$. An algorithm minimizing equation (7) subject to the weight constrains may select a solution, $\mathbf{W}^*$, with positive entries for units that are far away from the treated unit in the space of the predictors, even when an alternative solution exists based only on units with predictor values similar to $\mathbf{X}_1$. This may, in turn, lead to large interpolation biases that remain unchecked under the illusion of perfect fit, $\mathbf{X}_1 = \mathbf{X}_0 \mathbf{W}^*$.[19]

Even in moderate dimensions, k, the curse of dimensionality works to keep treated observations outside of the convex hull of the units in the donor pool. Therefore, in settings with one treated unit, multiplicity of solutions is rarely an issue, and if it arises it can often be easily addressed by restricting the donor pool to units with predictor values most similar to the values of the predictor for the unit exposed to the treatment.

[18] If $\mathbf{W}_1^*$ and $\mathbf{W}_2^*$ are both minimizers of equation (7) subject to the weight constraints, then so is $a\mathbf{W}_1^* + (1-a)\mathbf{W}_2^*$, for $a \in (0,1)$. This implies that the number of solutions to the minimization equation (7) given the weight constraints can only be one or infinity.

[19] Notice, however, that perfect fit, $\mathbf{X}_1 = \mathbf{X}_0 \mathbf{W}^*$, where $\mathbf{W}^*$ minimizes equation (7) subject to the weight constraints, is indicative of the possibility of infinite solutions.

However, in settings with many treated and untreated units, multiplicity of solutions and how to choose among them become important issues for estimation. Moreover, large interpolation biases may also arise in settings where the predictor values for treated units fall outside the convex hull of the predictor values for the units in the donor pool, especially when the units contributing to synthetic controls are far away from the treated units in the space of predictors. That is, there may be cases such that $\mathbf{X}_1 \approx \mathbf{X}_0 \boldsymbol{W}^*$ but where $\mathbf{X}_j$ greatly differs from $\mathbf{X}_1$, for some unit j contributing to the synthetic control.

To address these challenges, Abadie and L'Hour (2019) propose a synthetic control estimator that incorporates a penalty for pairwise matching discrepancies between the treated units and each of the units that contribute to their synthetic controls. Consider a setting with I treated units and J untreated units. We will index observations so that the treated units come first. That is, units $j = 1, \ldots, I$ are treated and units $j = I+1, \ldots, I+J$ are untreated, with $I+J$ units in total. As in previous sections, $\mathbf{X}_j$ is the vector of predictor values for unit j, and $\mathbf{X}_0$ is the matrix of the predictor values for the units in the donor pool. For $\lambda > 0$, the estimator in Abadie and L'Hour (2019) minimizes

$$(13) \quad \|\mathbf{X}_i - \mathbf{X}_0 \boldsymbol{W}\|^2 + \lambda \sum_{j=I+1}^{I+J} w_i \|\mathbf{X}_i - \mathbf{X}_j\|^2$$

with respect to $\boldsymbol{W} = \left(w_{I+1}, \ldots, w_{I+J}\right)'$, for each treated unit, $i = 1, \ldots, I$, subject to the constraints that the weights $w_{I+1}, \ldots, w_{I+J}$ are nonnegative and sum to one.[20] The first term in equation (13) is the aggregate discrepancy between the predictor values for treated unit i and its synthetic control. The second term in equation (13) penalizes pairwise matching discrepancies between the predictor values for unit i and each of the units that contribute to its synthetic control, weighted by the magnitudes of their contributions. The penalty term is added to equation (13) with the aim of reducing interpolation biases. As $\lambda \to \infty$, the penalized estimator converges to one-to-one matching. As $\lambda \to 0$, the estimator uses an aggregate of pairwise matching discrepancies weighted by $\boldsymbol{W}$ to select among all synthetic controls that attain the minimal value for $\|\mathbf{X}_i - \mathbf{X}_0 \boldsymbol{W}\|$. Values of λ between zero and infinity trade off aggregate fit of the synthetic control and pairwise fit of each of the units that contribute to it.

Abadie and L'Hour (2019) show that if $\lambda > 0$, then the minimizer of equation (13) is unique and sparse (provided that the columns of $\mathbf{X}_0$ are in general quadratic position, see Abadie and L'Hour 2019 for details). They also provide cross-validation techniques to select λ.

Let $\boldsymbol{W}_i^* = \left(w_{iI+1}^*, \ldots, w_{iI+J}^*\right)'$ be the solution to the minimization problem in equation (13). Then, the estimated treatment effect for $i = 1, \ldots, I$ and $t = T_0 + 1, \ldots, T$ is as in (8),

$$(14) \qquad \hat{\tau}_{it} = Y_{it} - \sum_{j=I+1}^{I+J} w_{ij}^* Y_{jt},$$

with average treatment effect given by

$$\hat{\tau}_t = \frac{1}{I} \sum_{i=1}^{I} \hat{\tau}_{it}.$$

In many instances, especially where the sample units are aggregates like regions or countries, a weighted average (e.g., population-weighted, or GDP-weighted) treatment effect may be most relevant.

Dube and Zipperer (2015) and Abadie and L'Hour (2019) propose extensions of the permutation methods in Abadie, Diamond, and Hainmueller (2010) to the case with multiple

[20] Although this is not reflected in the notation in equation (13), λ may depend on i.

treated units. They employ rank-based statistics on the permutation distribution of treatment effects, where the identity of the treated units is permuted at random in the data. In particular, Abadie and L'Hour (2019) propose the following simple generalization of the permutation test in Abadie, Diamond, and Hainmueller (2010). They consider a setting with I treated units and J untreated units. In each permutation $b = 1, \ldots, B$, the identities of the I treated units are reassigned in the data among the $I + J$ units in the sample, and statistics $r_{b,1}, \ldots, r_{b,I}$ are calculated for the units coded as treated in the permutation. These statistics could be (bias-corrected) synthetic control estimates of treatment effects, or rescaled versions that take into account the pre-intervention fit as in equation (12), or their absolute values, positive parts, or negative parts, depending on the context. Notice that when I is small relative to $I + J$, it may be possible to consider all possible treatment reassignments, in which case B is equal to $(I + J)$-choose-I. If considering all possible treatment reassignments is computationally expensive, inference can be based on B random draws from all subsets of I units in the sample. Let $r_{0,1}, \ldots, r_{0,I}$ be the same statistics calculated for the actual treated units. Then, B permutation repetitions, in addition to the original sample values for treatment, produce $I \times (B + 1)$ statistics, $r_{0,1}, \ldots, r_{0,I}, \ldots, r_{B,1}, \ldots, r_{B,I}$. Now, for each $b = 0, \ldots, B$, one can calculate t_b equal to the sum of the ranks of $r_{b,1}, \ldots, r_{b,I}$ within $r_{0,1}, \ldots, r_{0,I}, \ldots, r_{B,1}, \ldots, r_{B,I}$. The permutation inference in Abadie and L'Hour (2019) is based on the "extremeness" of the statistic t_0 within the permutation distribution $t_0, t_1, \ldots, t_B$. Notice that, for $I = 1$ this mode of inference amounts to the permutation test in Abadie, Diamond, and Hainmueller (2010).

Hainmueller (2012) and Robbins, Saunders, and Kilmer (2017) consider also settings with multiple treated units. Instead of producing a separate synthetic control for each treated unit, they calculate a single synthetic control to match aggregate values of the predictors between the treated and nontreated samples. As in the usual synthetic control estimator, the weights in Hainmueller (2012) and Robbins, Saunders, and Kilmer (2017) are nonnegative and sum to a predetermined constant (typically equal to one, or to the number of treated units, depending on the scaling of the variables in the data set). These estimators require that there is at least a convex combination of units in the donor pool that exactly matches a prespecified set of moments of the predictors for the treated units. Among the sets of weights that perfectly reproduce the moments for the treated sample, Hainmueller (2012) and Robbins, Saunders, and Kilmer (2017) choose the one that minimizes a measure of discrepancy with respect to constant weights.

Bias Correction. Another practical complication in settings with many treated units is that, even with a moderate k, the predictor values for some of the treated units may not be closely reproduced by a synthetic control, or may be closely reproduced only by combinations of units with large pairwise matching discrepancies in predictor values with respect to the treated unit. At the same time, including those ill-fitted units in the calculation of the aggregate effect may be important for the desired interpretation of the estimate (e.g., as an estimate of the average effect of the treatment on the treated). In that case, one could be concerned about the potential biases produced by matching discrepancies between the values of the predictors for the treated units and those for the respective synthetic controls.[21] Bias corrections

[21] Related to this problem, Ferman and Pinto (2019) and Botosaru and Ferman (2019) study the properties of synthetic control estimators for cases where the value of the predictors for a treated unit cannot be closely matched by a synthetic control.

play also an important role in reducing regularization biases in inferential methods for regression-based variants of synthetic controls (see, e.g., Arkhangelsky et al. 2019, Chernozhukov, Wüthrich, and Zhu 2019b).

Abadie and L'Hour (2019) and Ben-Michael et al. (2020) propose modifications of the synthetic control estimator along the lines of the bias-correction techniques of Rubin (1973), Quade (1982), and Abadie and Imbens (2011). They use regression adjustments to attenuate the bias of synthetic control estimators in settings where the synthetic control counterfactual is constructed using untreated units with values of the predictors that do not closely reproduce the predictor values for the treated unit or units. For $t = T_0 + 1, \ldots, T$, let $\hat{\mu}_{0t}$ be a sample regression function (parametric or nonparametric) estimated by regressing the untreated outcomes, $Y_{I+1,t}, \ldots, Y_{I+J,t}$, on the values of the predictors for the untreated units, $\mathbf{X}_{I+1}, \ldots, \mathbf{X}_{I+J}$. The bias-corrected synthetic control estimator for unit i is

$$(15)\quad \hat{\tau}_{it} = \left(Y_{it} - \sum_{j=I+1}^{I+J} w_{ij}^{*} Y_{jt} \right) - \sum_{j=I+1}^{I+J} w_{ij}^{*} \Big(\hat{\mu}_{0t}(\mathbf{X}_i) - \hat{\mu}_{0t}\big(\mathbf{X}_j\big) \Big).$$

The first term on the right-hand side of (15) is the synthetic control estimator in (14). The second term uses a regression adjustment to correct for discrepancies between the predictor values for the treated unit and the predictor values for the units that contribute to the synthetic control. Alternatively, the estimator in (15) can be expressed as

$$(16)\quad \hat{\tau}_{it} = \big(Y_{it} - \hat{\mu}_{0t}(\mathbf{X}_i) \big) - \sum_{j=I+1}^{I+J} w_{ij}^{*} \Big(Y_{jt} - \hat{\mu}_{0t}\big(\mathbf{X}_j\big) \Big).$$

Equation (16) provides an interpretation of the bias-corrected synthetic control estimator as a synthetic control estimator applied to regression residuals. The bias correction in equation (16) is related to the proposal in Doudchenko and Imbens (2016) to residualize the outcomes with respect to covariates before calculating synthetic controls.

A different avenue to evaluate the bias of synthetic control estimators, which was discussed in section 7, is given by the availability of pre-intervention periods, when the effect of the treatment is not yet realized. In the absence of anticipation effects, estimates of treatment effects before the intervention are reflective of estimation biases. If biases are stable in time, estimates of those biases could be used to correct synthetic control estimates. Bias adjustments of this type, which are closely related to difference-in-differences methods, are proposed in Arkhangelsky et al. (2019) and Chernozhukov, Wüthrich, and Zhu (2019b).

Regression-Based Methods and Extrapolation. Several articles have contributed regression-based estimators for synthetic controls. These procedures allow extrapolation by considering synthetic controls that are not convex combinations of the units in the donor pool. Doudchenko and Imbens (2016) consider an estimator that fits all pretreatment outcomes for the treated, with weights that may be negative and may not sum to one, and allow for a constant shift in the level of the synthetic control estimator. They propose to use an elastic net—that is, a combination of lasso (L_1) and ridge (L_2) penalties—to regularize the weights. The counterfactual estimates for $t = T_0 + 1, \ldots, T$ in Doudchenko and Imbens (2016) are

$$(17)\quad \hat{Y}_{1t}^{N} = \hat{\alpha} + \sum_{j=2}^{J+1} \hat{w}_j Y_{jt},$$

where $\hat{\alpha}, \hat{w}_2, \ldots, \hat{w}_{J+1}$ minimize

$$(18) \quad \sum_{t=1}^{T_0}\left(Y_{1t} - \alpha - \sum_{j=2}^{J+1} w_j Y_{jt}\right)^2 + \lambda_1\left(\frac{1-\lambda_2}{2}\sum_{j=2}^{J+1} w_j^2 + \lambda_2 \sum_{j=2}^{J+1} |w_j|\right),$$

with respect to $(\alpha, w_2, \ldots, w_{J+1}) \in \mathbb{R}^{J+1}$, where $\lambda_1 \geq 0$ and $0 \leq \lambda_2 \leq 1$ are regularization parameters selected by cross-validation. To incorporate additional predictors in their estimation procedure, Doudchenko and Imbens (2016) propose to use least squares in a first step to residualize the outcomes Y_{jt} for $j = 1, \ldots, J+1$ in equation (18) with respect to any other covariates. Chernozhukov, Wüthrich, and Zhu (2019a) consider different penalty terms, including lasso regularization (i.e., $\lambda_2 = 1$), in the context of an inferential procedure for synthetic controls. Regression estimators of this type are also related to the panel data approach to the program evaluation estimator in Hsiao, Ching, and Wan (2012), where $\lambda_1 = 0$ and the parameters in equation (17) are estimated by unpenalized least squares. Li (2019) considers the same estimator as in Hsiao, Ching, and Wan (2012), but regularizes the weights $\hat{w}_2, \ldots, \hat{w}_{J+1}$ to be nonnegative.

Arkhangelsky et al. (2019) introduce a synthetic control estimator that weights not only the units in the control group, but also the pre-intervention time periods, to approximate the counterfactual of interest. The time weights in Arkhangelsky et al. (2019) play a similar role as the predictor weights, $v_1, \ldots, v_k$, of subsection 3.2. They reflect the importance of each of the individual predictors, which in the leading version of the estimator of Arkhangelsky et al. (2019) are past outcome values.

Matrix Completion/Estimation Methods. Amjad, Shah, and Shen (2018); Amjad et al. (2019); and Athey et al. (2020) propose related methods that use tools from the matrix completion/matrix estimation literature. Suppose, as before, that unit $j = 1$ is the treated unit, and units $j = 2, \ldots, J+1$ are not treated. Amjad, Shah, and Shen (2018) posit a nonlinear factor-structure model for the untreated, $Y_{jt}^N = f(\mu_j, \lambda_t) + \varepsilon_{jt}$, with $j = 2, \ldots, J+1$ and $t = 1, \ldots, T$, where ε_{jt} is random noise. Their framework allows for the presence of missing values in the matrix $\left\{Y_{jt}^N\right\}$, with $j = 1, \ldots, J+1$, and $t = 1, \ldots, T$. Using matrix estimation methods, in particular singular value thresholding (see Chatterjee 2015), Amjad, Shah, and Shen (2018) estimate a low-rank approximation, $\left\{\hat{M}_{jt}\right\}$, to the matrix $\left\{M_{jt}\right\} = \left\{f(\mu_j, \lambda_t)\right\}$. The objects $\hat{M}_{jt}$ are used to de-noise the outcomes Y_{jt}^N and to impute missing values, if any. Then, synthetic controls are obtained as linear combinations of $\hat{M}_{jt}$, with coefficients estimated by ridge regression of Y_{1t} on $\hat{M}_{2t}, \ldots \hat{M}_{J+1t}$ in the pre-intervention periods. The estimator in Amjad, Shah, and Shen (2018) does not incorporate covariates, using data on outcomes, Y_{jt}, only. Amjad et al. (2019) modify the estimator in Amjad, Shah, and Shen (2018) to incorporate additional variables aside from the outcome of interest, under the assumption that all variables depend on common latent factors. Athey et al. (2020) postulate the model $Y_{jt}^N = M_{jt} + \varepsilon_{jt}$, for $j = 1, \ldots, J+1$ and $t = 1, \ldots, T$, where ε_{jt} is again random noise. In their framework, missing entries in the matrix $\left\{Y_{jt}^N\right\}$, with $j = 1, \ldots, J+1$ and $t = 1, \ldots, T$, arise naturally for the treated observation (or treated observations, if multiple units are treated) in the posttreatment periods. Athey et al. (2020) assume that the matrix $\left\{M_{jt}\right\}$, with $j = 1, \ldots, J+1$ and $t = 1, \ldots, T$, is low-rank, which allows them to obtain an estimate, $\left\{\hat{M}_{jt}\right\}$, via matrix completion techniques. The estimated counterfactual outcomes without the treatment for the treated are the values of $\hat{M}_{jt}$ such that Y_{jt}^N is missing. Extensions allow for models with covariates and the inclusion of time fixed effects and unit fixed effects (separate from the low-rank matrix, M_{jt}).

Inference. Several studies have proposed inferential tools for synthetic controls as alternatives to the permutation test in subsection 3.5. Firpo and Possebom (2018) propose several generalizations of the permutation test in subsection 3.5 and contribute confidence sets based on inverting the results of these tests. In a repeated sampling framework for stationary data and large T_0, Hahn and Shi (2017) propose to apply the end-of-sample instability test of Andrews (2003) to obtain an inferential procedure for synthetic control estimators. In the context of synthetic control estimators, the end-of-sample instability test of Andrews (2003) is related to the backdating ideas of section 7. It compares the values of treatment effects computed for the $T - T_0$ post-intervention periods to the distribution of the of same values computed for every subset of $T - T_0$ consecutive pre-intervention periods. Related also to Andrews's end-of-sample instability test, Chernozhukov, Wüthrich, and Zhu (2019a) devise a sampling-based inferential procedure for synthetic controls and related methods that employs permutations of regression residuals in the time dimension. In particular, Chernozhukov, Wüthrich, and Zhu (2019a) assume $Y_{1t}^N = P_t^N + u_t$, where $u_1, \ldots, u_T$ are stationary and weakly dependent with mean zero. Let $\tau_{T_0+1}, \ldots, \tau_T$ be the effects of the treatment on the treated unit (unit one) at times $t = T_0 + 1, \ldots, T$.[22] The potential outcome under the intervention is $Y_{1t}^I = P_t^N + \tau_t + u_t$ for $t > T_0$. To simplify the exposition, assume that $u_1, \ldots, u_T$ are i.i.d. Then, the distribution of a function, $S\left(u_{T_0+1}, \ldots, u_T\right)$, of the post-intervention values of u_t should be the same as the distribution of $S\left(u_{\pi(T_0+1)}, \ldots, u_{\pi(T)}\right)$, where $\pi(1), \ldots, \pi(T)$ is a random permutation of $1, \ldots, T$. Suppose for now that P_t^N is known. Then, under a null hypothesis, $\tau_{T_0+1} = a_{T_0+1}, \ldots, \tau_T = a_T$, we can compute $u_t = Y_{1t} - P_t^N - a_t$, where $a_t = 0$ for $1 \leq t \leq T_0$. As a result, we can test the null hypothesis by comparing the value of $S\left(u_{T_0+1}, \ldots, u_T\right)$ to its permutation distribution, that is, the distribution of $S\left(u_{\pi(T_0+1)}, \ldots, u_{\pi(T)}\right)$, which can be directly computed in the data. A feasible implementation of the test requires estimation of the residuals, $u_1, \ldots, u_T$. In the context of the synthetic control method, Chernozhukov, Wüthrich, and Zhu (2019a) adopt the model $P_t^N = \sum_{j=2}^{J+1} w_j Y_{jt}$ with nonnegative weights that sum to one, and $E\left[u_t Y_{jt}\right] = 0$ for $j = 2, \ldots, J+1$, and implement their test on constrained least squares residuals, $\hat{u}_1, \ldots, \hat{u}_T$. The proposal in Chernozhukov, Wüthrich, and Zhu (2019a) differs from other synthetic control procedures in two important respects. First, while much of the literature on synthetic controls has adopted the linear factor model of subsection 3.3 as a working model to understand the properties of synthetic control estimators, Chernozhukov, Wüthrich, and Zhu (2019a) adopt instead the restriction $E\left[u_t Y_{jt}\right] = 0$ for $j = 2, \ldots, J+1$ to estimate P_t^N.[23] They

[22] To be consistent with the notation for P_t^N and u_t and because only unit one is treated, here I drop the subscript indicating that identity of the treated unit from the notation for treatment effect.

[23] To understand the differences between these two frameworks, notice that when the data are generated by the linear factor model of subsection 3.3, and there is an unbiased synthetic control—that is, a synthetic control with weights, $w_2, \ldots, w_{J+1}$ that exactly reproduces $\mathbf{Z}_1$ and μ_1—then, the restriction $E\left[u_t Y_{jt}\right] = 0$ for $j = 2, \ldots, J+1$ and $u_t = Y_{1t}^N - \sum_{j=2}^{J+1} w_j Y_{jt}$ does not hold in general (see Ferman and Pinto 2019). One exception is given by the results in Ferman (2019), which imply that $E\left[u_t Y_{jt}\right] = 0$ for $j = 2, \ldots, J+1$ will approximately hold as

show, however, that regardless of the validity of the model, their testing procedure remains valid as long as the estimated residuals, $\hat{u}_1, \ldots, \hat{u}_T$ are exchangeable under the null hypothesis. Second, in contrast to other synthetic control procedures that compute the weights $w_2, \ldots, w_{J+1}$ using pre-intervention data only, in the inferential procedure of Chernozhukov, Wüthrich, and Zhu (2019a) the synthetic control weights are estimated under the null hypothesis, $\tau_{T_0+1} = a_{T_0+1}, \ldots, \tau_T = a_T$, using data on $Y_{it}^N = Y_{it} - a_t$ for all periods, including the periods after the intervention. For a similar set of models, Chernozhukov, Wüthrich, and Zhu (2019b) propose bias-corrected synthetic control estimation and confidence intervals for the mean value of the treatment effect over the post-intervention period, $\left(\tau_{T_0+1} + \cdots + \tau_T\right)/(T - T_0)$ in settings when both T_0 and $T - T_0$ are large. Similar to difference in differences, the bias-correction procedure of Chernozhukov, Wüthrich, and Zhu (2019b) adjusts for differences in pre-intervention outcomes between the treated unit and the synthetic control. Confidence intervals are based on an asymptotically pivotal t-statistic and centered on the average of K-fold cross-fitted versions of the bias-corrected synthetic control estimate. Cattaneo, Feng, and Titiunik (2021) propose predictive intervals for synthetic control estimators and related methods. They adopt a predictive model $Y_{1t}^N = P_t^N + u_t$, where P_t^N depends on observed predictors and unknown parameters, and u_t is an unobserved random error. Their predictive intervals for $\hat{\tau}_{1t} = Y_{1t} - Y_t^N$ (with $t > T_0$) take into account estimation uncertainty about the values of the parameters in P_t^N as well as irreducible uncertainty about the value of u_t.

$J \to \infty$ if there are weights, $w_2, \ldots, w_{J+1}$ that asymptotically recover Z_1 and μ_1 and are increasingly diluted among the units in the donor pool.

Other Contributions. In this article, I have provided a brief description of selected strands of the literature on synthetic controls and related methods, starting with the canonical estimator in sections 2 and 3, and describing some extensions and related methods in the current section. The literature is vast in its totality, however, and there are many noteworthy contributions I did not cover. They include Bai and Ng (2019); Brodersen et al. (2015); Gobillon and Magnac (2016); Gunsilius (2020); Kennedy-Shaffe, de Gruttola, and Lipsitch (2020); Viviano and Bradic (2019); and Xu (2017), among many others. Samartsidis et al. (2019) study the performance of the canonical synthetic control estimator and related methods in the context of the German reunification example of subsection 3.2. As the set of methods on synthetic controls keeps expanding and enriching the applied econometrics toolkit, this is still a young literature and much remains to be done. I mention some open areas in the final section of this article.

9. *Conclusions*

Synthetic controls provide many practical advantages for the estimation of the effects of policy interventions and other events of interest. However, like for any other statistical procedure (and especially for those aimed at estimating causal effects), the credibility of the results depends crucially on the level of diligence exerted in the application of the method and on whether contextual and data requirements are met in the empirical application at hand. In this article, I emphasize the notion that mechanical applications of synthetic controls that do not take into account the context of the investigation or the nature of the data are risky enterprises. To this end, the article discusses the methodological underpinnings of synthetic control estimators and the conditions under which they provide suitable estimates of causal effects. It also describes how the analysis may be modified in the cases when those conditions do not hold. Finally, the article discusses some recent extensions that widen the applicability, robustness, and flexibility of the method.

Open areas of related research abound, both methodological and empirical. Results on sampling-based inference, external validity, sensitivity to model restrictions, estimation with multiple interventions, and the identification of the channels though which the effect of an event or intervention operates, to mention a few, are scant or absent in the synthetic controls literature. An area of recent heightened interest regarding the use of synthetic controls is the design of experimental interventions in settings where the intervention of interest can only be applied to one or a small number of aggregate units. In addition, existing results on robust and efficient computation of synthetic controls are scarce, and more research is needed on the computational aspects of this methodology. On the empirical side, many of the events and policy interventions economists care about take place at an aggregate level, affecting entire aggregate units like school districts, cities, regions, or countries. This is exactly the setting synthetic controls were designed for, and potential applications of synthetic controls in economics are many.

References

Abadie, Alberto. 2005. "Semiparametric Difference-in-Differences." *Review of Economic Studies* 72 (1): 1–19.

Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey M. Wooldridge. 2020. "Sampling-Based versus Design-Based Uncertainty in Regression Analysis." *Econometrica* 88 (1): 265–96.

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association* 105 (490): 493–505.

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

Abadie, Alberto, and Javier Gardeazabal. 2003. "The Economic Costs of Conflict: A Case Study of the Basque Country." *American Economic Review* 93 (1): 113–32.

Abadie, Alberto, and Guido W. Imbens. 2011. "Bias-Corrected Matching Estimators for Average Treatment Effects." *Journal of Business & Economic Statistics* 29 (1): 1–11.

Abadie, Alberto, and Jérémy L'Hour. 2019. "A Penalized Synthetic Control Estimator for Dissagregated Data." Unpublished. https://sites.google.com/site/jeremylhour/research.

Acemoglu, Daron, Simon Johnson, Amir Kermani, James Kwak, and Todd Mitton. 2016. "The Value of Connections in Turbulent Times: Evidence from the United States." *Journal of Financial Economics* 121 (2): 368–91.

Allegretto, Sylvia, Arindrajig Dube, Michael Reich, and Ben Zipperer. 2017. "Credible Research Designs for Minimum Wage Studies: A Response to Neumark, Salas, and Wascher." *ILR Review* 70 (3): 559–92.

Amjad, Muhammad, Devavrat Shah, and Dennis Shen. 2018. "Robust Synthetic Control." *Journal of Machine Learning Research* 19 (22): 1–51.

Amjad, Muhammad J., Vishal Misra, Devavrat Shah, and Dennis Shen. 2019. "mRSC: Multidimensional Robust Synthetic Control." *Proceedings of the ACM on Measurement and Analysis of Computing Systems* 3 (2). https://doi.org/10.1145/3341617.3326152.

Andrews, D. W. K. 2003. "End-of-Sample Instability Tests." *Econometrica* 71 (6): 1661–94.

Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton and Oxford: Princeton University Press.

Arkhangelsky, Dmitry, Susan Athey, David A. Hirshberg, Guido W. Imbens, and Stefan Wager. 2019. "Synthetic Difference in Differences." NBER Working Paper 25532.

Athey, Susan, Mohsen Bayati, Nikolay Doudchenko, Guido Imbens, and Khashayar Khosravi. 2020. "Matrix Completion Methods for Causal Panel Data Models." Available on arXiv at 1710.10251.

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

Bai, Jushan. 2009. "Panel Data Models with Interactive Fixed Effects." *Econometrica* 77 (4): 1229–79.

Bai, Jushan, and Serena Ng. 2019. "Matrix Completion, Counterfactuals, and Factor Analysis of Missing

Data." Available on arXiv at 1910.06677.

Ben-Michael, Eli, Avi Feller, and Jesse Rothstein. 2020. "The Augmented Synthetic Control Method." Available on arXiv at 1811.04170.

Bifulco, Robert, Ross Rubenstein, and Hosung Sohn. 2017. "Using Synthetic Controls to Evaluate the Effect of Unique Interventions: The Case of Say Yes to Education." *Evaluation Review* 41 (6): 593–619.

Bohn, Sarah, Magnus Lofstrom, and Steven Raphael. 2014. "Did the 2007 Legal Arizona Workers Act Reduce the State's Unauthorized Immigrant Population?" *Review of Economics and Statistics* 96 (2): 258–69.

Borjas, George J. 2017. "The Wage Impact of the Marielitos: A Reappraisal." *ILR Review* 70 (5): 1077–1110.

Botosaru, Irene, and Bruno Ferman. 2019. "On the Role of Covariates in the Synthetic Control Method." *Econometrics Journal* 22 (2): 117–30.

Brodersen, Kay H., Fabian Gallusser, Jim Koehler, Nicolas Remy, and Steven L. Scott. 2015. "Inferring Causal Impact Using Bayesian Structural Time-Series Models." *Annals of Applied Statistics* 9 (1): 247–74.

Card, David. 1990. "The Impact of the Mariel Boatlift on the Miami Labor Market." *ILR Review* 43 (2): 245–57.

Card, David, and Alan B. Krueger. 1994. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania." *American Economic Review* 84 (4): 772–93.

Cattaneo, Mattias D., Yinjie Feng, and Rocio Titiunik. 2021. "Prediction Intervals for Synthetic Control Methods." Available on arXiv at 1912.07120v2.

Chatterjee, Sourav. 2015. "Matrix Estimation by Universal Singular Value Thresholding." *Annals of Statistics* 43 (1): 177–214.

Chernozhukov, Victor, Kaspar Wüthrich, and Yinchu Zhu. 2019a. "An Exact and Robust Conformal Inference Method for Counterfactual and Synthetic Controls." Available on arXiv at 1712.09089v6.

Chernozhukov, Victor, Kaspar Wüthrich, and Yinchu Zhu. 2019b. "Practical and Robust *t*-Test Based Inference for Synthetic Control and Related Methods." Available on arXiv at 1812.10820v2.

Cook, Thomas D., and Donald T. Campbell. 1979. *Quasi-Experimentation: Design & Analysis Issues for Field Settings*. Boston: Houghton Mifflin.

Cunningham, Scott, and Manisha Shah. 2018. "Decriminalizing Indoor Prostitution: Implications for Sexual Violence and Public Health." *Review of Economic Studies* 85 (3): 1683–1715.

Donohue, John J., Abhay Aneja, and Kyle D. Weber. 2019. "Right-to-Carry Laws and Violent Crime: A Comprehensive Assessment Using Panel Data and a State-Level Synthetic Control Analysis." *Journal of Empirical Legal Studies* 16 (2): 198–247.

Doudchenko, Nikolay, and Guido W. Imbens. 2016. "Balancing, Regression, Difference-in-Differences and Synthetic Control Methods: A Synthesis." NBER Working Paper 22791.

Douglas, Jason. 2018. "How an Analysis of Basque Terrorism Helps Economists Understand Brexit." *Wall Street Journal*, November 7.

Dube, Arindrajit, and Ben Zipperer. 2015. "Pooling Multiple Case Studies Using Synthetic Controls: An Application to Minimum Wage Policies." IZA Discussion Paper 8944.

Ferman, Bruno. 2019. "On the Properties of the Synthetic Control Estimator with Many Periods and Many Controls." Available on arXiv at 1906.06665.

Ferman, Bruno, and Cristine Pinto. 2019. "Synthetic Controls with Imperfect Pre-treatment t." Available on arXiv at 1911.08521.

Firpo, Sergio, and Vitor Possebom. 2018. "Synthetic Control Method: Inference, Sensitivity Analysis and Confidence Sets." *Journal of Causal Inference* 6 (2): 20160026.

Fisher, Ronald. 1935. *The Design of Experiments*. Edinburgh and London: Oliver and Boyd.

Gobillon, Laurent, and Thierry Magnac. 2016. "Regional Policy Evaluation: Interactive Fixed Effects and Synthetic Controls." *Review of Economics and Statistics* 98 (3): 535–51.

Gunsilius, Florian. 2020. "Distributional Synthetic Controls." Available on arXiv at 2001.06118v3.

Guo, Jeff. 2015. "Seriously, Here's One Amazing Math Trick to Learn What Can't Be Known." *Washington Post*, October 30.

Gutierrez, Italo A., Gabriel Weinberger, and John Engberg. 2016. *Improving Teaching Effectiveness: Impact on Student Outcomes: The Intensive Partnerships for Effective Teaching Through 2013–2014*. Santa Monica, CA: RAND Corporation.

Hahn, Jinyong, and Ruoyao Shi. 2017. "Synthetic Control and Inference." *Econometrics* 5 (4): Article 52.

Hainmueller, Jens. 2012. "Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies." *Political Analysis* 20 (1): 25–46.

Heckman, James J., and V. Joseph Hotz. 1989. "Choosing among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training." *Journal of the American Statistical Association* 84 (408): 862–74.

Heersink, Boris, Brenton D. Peterson, and Jeffrey A. Jenkins. 2017. "Disasters and Elections: Estimating the Net Effect of Damage and Relief in Historical Perspective." *Political Analysis* 25 (2): 260–68.

Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical*

Association 81: 945–60.

Hsiao, Cheng, H. Steve Ching, and Shui Ki Wan. 2012. "A Panel Data Approach for Program Evaluation: Measuring the Benefits of Political and Economic Integration of Hong Kong with Mainland China." *Journal of Applied Econometrics* 27 (5): 705–40.

Jardim, Ekaterina, Mark C. Long, Robert Plotnick, Emma van Inwegen, Jacob Vigdor, and Hilary Wething. 2017. "Minimum Wage Increases, Wages, and Low-Wage Employment: Evidence from Seattle. NBER Working Paper 23532.

Kennedy-Shaffer, Lee, Victor de Gruttola, and Marc Lipsitch. 2020. "Novel Methods for the Analysis of Stepped Wedge Cluster Randomized Trials." *Statistics in Medicine* 39 (7): 815–44.

King, Gary, and Langche Zeng. 2006. "The Dangers of Extreme Counterfactuals." *Political Analysis* 14 (2): 131–59.

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

Klößner, Stefan, Ashok Kaul, Gregor Pfeifer, and Manuel Schieler. 2018. "Comparative Politics and the Synthetic Control Method Revisited: A Note on Abadie et al. (2015)." *Swiss Journal of Economics and Statistics* 154 (1): Article 11.

Kreif, Noémi, Richard Grieve, Dominik Hangartner, Alex James Turner, Silviya Nikolova, and Matt Sutton. 2016. "Examination of the Synthetic Control Method for Evaluating Health Policies with Multiple Treated Units." *Health Economics* 25 (12): 1514–28.

Li, Kathleen T. 2020. "Statistical Inference for Average Treatment Effects Estimated by Synthetic Control Methods." *Journal of the American Statistical Association* 115 (532): 2068–83.

Neumark, David, and William Wascher. 2017. "Reply to 'Credible Research Designs for Minimum Wage Studies'." *ILR Review* 70 (3): 593–609.

Olken, Benjamin A. 2015. "Promises and Perils of Pre-analysis Plans." *Journal of Economic Perspectives* 29 (3): 61–80.

Peri, Giovanni, and Vasil Yasenov. 2019. "The Labor Market Effects of a Refugee Wave: Synthetic Control Method Meets the Mariel Boatlift." *Journal of Human Resources* 54 (2): 267–309.

Pieters, Hannah, Daniele Curzi, Alessandro Olper, and Johan Swinnen. 2017. "Effect of Democratic Reforms on Child Mortality: A Synthetic Control Analysis." *Lancet Global Health* 4 (9): e627–32.

Pinotti, Paolo. 2015. "The Economic Costs of Organised Crime: Evidence from Southern Italy." *Economic Journal* 125 (586): F203–32.

Quade, Dana. 1982. "Nonparametric Analysis of Covariance by Matching." Special issue, *Biometrics* 38 (3): 597–611.

Reich, Michael, Sylvia A. Allegretto, and Anna Godoey. 2017. "Seattle's Minimum Wage Experience 2015–16." CWED Policy Brief, University of California, Berkeley.

Robbins, Michael W., Jessica Saunders, and Beau Kilmer. 2017. "A Framework for Synthetic Control Methods with High-Dimensional, Micro-Level Data: Evaluating a Neighborhood-Specific Crime Intervention." *Journal of the American Statistical Association* 112 (517): 109–26.

Rosenbaum, Paul R. 1984. "Conditional Permutation Tests and the Propensity Score in Observational Studies." *Journal of the American Statistical Association* 79 (387): 565–74.

Rubin, Donald B. 1973. "The Use of Matched Sampling and Regression Adjustments to Remove Bias in Observational Studies." *Biometrics* 29 (1): 185–203.

Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66 (5): 688–701.

Rubin, Donald B. 1980. "Comment." *Journal of the American Statistical Association* 75 (371): 591–93.

Rubin, Donald B. 2007. "The Design *versus* the Analysis of Observational Studies for Causal Effects: Parallels with the Design of Randomized Trials." *Statistics in Medicine* 26 (1): 20–36.

Samartsidis, Pantelis, Shaun R. Seaman, Anne M. Presanis, Matthew Hickman, and Daniela De Angelis. 2019. "Assessing the Causal Effect of Binary Interventions from Observational Panel Data with Few Treated Units." *Statistical Science* 34 (3): 486–503.

Thistlethwaite, Donald L., and Donald T. Campbell. 1960. "Regression-Discontinuity Analysis: An Alternative to the Ex Post Facto Experiment." *Journal of Educational Psychology* 51 (6): 309–17.

Viviano, Davide, and Jelena Bradic. 2019. "Synthetic Learner: Model-Free Inference on Treatments over Time." Available on arXiv at 1904.01490.

Xu, Yiqing. 2017. "Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models." *Political Analysis* 25 (1): 57–76.