Doubly robust proximal synthetic controls

Hongxiang Qiu, Xu Shi, Wang Miao, Edgar Dobriban, Eric Tchetgen Tchetgen

1Department of Epidemiology and Biostatistics, Michigan State University, East Lansing, MI 48824, United States; 2Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, United States; 3Department of Probability and Statistics, Peking University, Beijing 100871, China; 4Department of Statistics and Data Science, University of Pennsylvania, Philadelphia, PA 19104, United States

*Corresponding author: Hongxiang Qiu, Department of Epidemiology and Biostatistics, Michigan State University, East Lansing, MI 48824, United States (qiu.hong@msu.edu).

ABSTRACT

To infer the treatment effect for a single treated unit using panel data, synthetic control (SC) methods construct a linear combination of control units’ outcomes that mimics the treated unit’s pre-treatment outcome trajectory. This linear combination is subsequently used to impute the counterfactual outcomes of the treated unit had it not been treated in the post-treatment period, and used to estimate the treatment effect. Existing SC methods rely on correctly modeling certain aspects of the counterfactual outcome generating mechanism and may require near-perfect matching of the pre-treatment trajectory. Inspired by proximal causal inference, we obtain two novel nonparametric identifying formulas for the average treatment effect for the treated unit: one is based on weighting, and the other combines models for the counterfactual outcome and the weighting function. We introduce the concept of covariate shift to SCs to obtain these identification results conditional on the treatment assignment. We also develop two treatment effect estimators based on these two formulas and generalized method of moments. One new estimator is doubly robust: it is consistent and asymptotically normal if at least one of the outcome and weighting models is correctly specified. We demonstrate the performance of the methods via simulations and apply them to evaluate the effectiveness of a pneumococcal conjugate vaccine on the risk of all-cause pneumonia in Brazil.

KEYWORDS: doubly robust estimation; panel data; proximal causal inference; synthetic control.

1 INTRODUCTION

1.1 Background

Interventions such as policies are often implemented in a single unit such as a state, a city, or a school. Causal inference in these cases is challenging due to the small number of treated units, and due to the lack of randomization and independence. In various fields, including economics, public health, and biometry, synthetic control (SC) methods (Abadie and Gardeazabal, 2003; Abadie et al., 2010, 2015; Doudchenko and Imbens, 2016) are a common tool to estimate the intervention (or treatment) effect for the treated unit in time series from a single treated unit and multiple untreated units in both pre- and post-treatment periods. For example, SC methods have been used to estimate the effects of terrorist conflicts on GDP (Abadie and Gardeazabal, 2003), tobacco control program on tobacco consumption (Abadie et al., 2010), Kansas’s tax cut on GDP (Ben-Michael et al., 2021b; Rickman and Wang, 2018), Florida’s “stand your ground” law on homicide rates (Bonander et al., 2021), and pneumococcal conjugate vaccines on pneumonia (Bruhn et al., 2017a).

Classical SCs are linear combinations of control units that mimic the treated unit before the treatment. Outcome differences between the treated unit and the SC in the post-treatment period are used to make inferences about the treatment effect for the treated unit. In Abadie and Gardeazabal (2003) and Abadie et al. (2010), a SC is a weighted average of a pool of control units, called the donors. The weights are obtained by minimizing a distance between the SC and the treated unit in the pre-treatment period, under the constraint that the weights are non-negative and sum to unity. Many extensions have been proposed. For example, Abadie and L’Hour (2021) proposed methods for multiple treated units, and Ben-Michael et al. (2022) further considered the case where these treated units initiate treatment at different time points; Doudchenko and Imbens (2016) and Ben-Michael et al. (2022, 2021b) introduced penalization to improve performance; Athey et al. (2021) and Bai and Ng (2021) used techniques from matrix completion; Li (2020) studied statistical inference for SC methods; and Chernozhukov et al. (2021) and Cattaneo et al. (2021) considered prediction intervals for treatment effects. Among these extensions, some also incorporate the idea that, similarly to the control units’ outcomes in the post-treatment period, the treated unit’s outcomes in the pre-treatment period can be used to impute the counterfactual outcome had it not been treated (Arkhangelsky et al., 2021; Ben-Michael et al., 2021b).

Existing methods often rely on assuming linear models and on the existence of near-perfectly matching weights in the observed data. Under such assumptions, valid SCs are linear combinations, often weighted averages, of donors. However, if such assumptions do not hold, these methods may not produce a valid SC. This may happen if the outcomes in the donors have a
different measuring scale from the treated unit, or if the treated unit’s and the donors’ outcomes have a nonlinear relationship.

To relax these assumptions, Shi et al. (2023) viewed SCs from the proximal causal inference perspective. For independent and identically distributed (i.i.d.) observations, Miao et al. (2018), Deaneer (2018, 2021), Cui et al. (2020), and Tchetgen Tchetgen et al. (2020) derived nonparametric identification using proxies, variables capturing the effect of the unmeasured confounders. Shi et al. (2023) viewed control units’ outcomes as proxies and obtained nonparametric identification results for the potential outcome of the treated unit had it not been treated as well as the treatment effect in a general setting, beyond the common linear factor model (e.g., Abadie et al., 2010). They assumed the existence of a function of these proxies, termed confounding bridge function, that captures the (possibly nonlinear) effects of unobserved confounders. With this function, they imputed the expected counterfactual outcomes for the treated unit. Estimation of, and inference about, the average treatment effect for the treated unit (ATT) followed from this identification result. Instrumental variables (IVs) have also been used. For example, Holtz-Eakin et al. (1988) considered a linear model with interactive fixed effects and showed how to identify it using appropriate instruments. The solution to this problem relies on a particular differencing strategy, which may be viewed as an application of a confounding bridge function. Cunha et al. (2010), Freyberger (2018), and references therein considered general nonparametric models with interactive effects, showing how to identify it using appropriate instruments. While treatment confounding proxies in proximal causal inference are sometimes described as instruments, it is crucial to note that they are more general than IVs, in the sense that valid IVs are valid treatment confounding proxies, but invalid IVs dependent on hidden confounders are also valid treatment confounding proxies (Tchetgen Tchetgen et al., 2020). In addition, while IVs require a form of homogeneity condition for nonparametric identification (e.g., separable errors or monotonicity), proxies do not require such a condition.

1.2 Our contribution

Existing methods rely on correctly specifying an outcome model, based on which one can impute the counterfactual outcome trajectory of the treated unit, had it not been treated, after treatment. This outcome bridge function model may be difficult to specify correctly, or may not exist. In this paper, we relax this requirement by leveraging the proximal causal inference framework as in Shi et al. (2023). We develop two novel methods to estimate the ATT. One method relies on weighting and is a building block to a second method which we rigorously prove is doubly robust (Bang and Robins, 2005; Scharfstein et al., 1999). It is consistent and asymptotically normal if either the outcome model or the weighting function is correctly specified, without requiring that both are. An advantage of the doubly robust method compared to existing methods is that it allows for misspecifying one of the two models, without the user necessarily knowing which might be mis-specified.

We observed that our estimand of interest, the ATT, is closely related to the average treatment effect on the treated for i.i.d. data (e.g., Hahn, 1998; Imbens, 2004; Chen et al., 2008; Shu and Tan, 2018). The method in Shi et al. (2023) corresponds to using an outcome confounding bridge function (Miao et al., 2018), which is the proximal causal inference counterpart of G-computation, or an outcome regression-based approach in causal inference under unconfoundedness (Robins, 1986). Our proposed methods are motivated by the existing identification results in proximal causal inference in the i.i.d. setting (Cui et al., 2020): one result is based on weighting and the other is based on the influence function.

Despite these similarities, it remains challenging to adapt these ideas from the i.i.d. setting to panel data. Since treatment assignment is often viewed as fixed in SC problems, a key concept from the i.i.d. setting, the propensity score (Rosenbaum and Rubin, 1983), is undefined. Thus, existing results for the i.i.d. setting cannot be directly applied to SC problems. We leverage the notion of covariate shift (e.g., Quinonero-Candela et al., 2009) to circumvent this issue. We also find a relaxed version of the i.i.d. assumption to allow for serial correlation, while still obtaining identification via weighting. We illustrate our proposed methods in simulations and three empirical examples: Two examples concern public health outcomes, one studying the effect of the PCV10 vaccine in Brazil on pneumonia (Bruhn et al., 2017a), and the other studying the effect of Florida’s “stand your ground” law on homicide rates (Bonander et al., 2021); the third example concerns economic outcomes, studying the effect of Kansas’s tax cut on GDP (Rickman and Wang, 2018).

Both our doubly robust method and the method in Ben-Michael et al. (2021b) take the form of augmented weighted moment equations, but the known robustness properties of these methods differ. In Ben-Michael et al. (2021b), the treated unit’s counterfactual outcome had it not been treated after treatment can be imputed with two approaches. One is a weighted average of control units’ outcomes, identical to classical SC methods (Abadie et al., 2010); the other is a prediction model obtained using the treated unit’s outcomes before treatment. By combining these two approaches, Ben-Michael et al. (2021b) developed a SC method with improved performance. Arkhangelsky et al. (2021) proposed a method combining two imputation approaches based on similar ideas. Nevertheless, to date, neither method has formally been shown to be doubly robust. In contrast, we formally establish double robustness, inherited from the influence function of the ATT in i.i.d. cases.

2 PROBLEM SETUP

We observe data over $T$ time periods. The first $T_0$ time periods are the pre-treatment periods, and the last $T - T_0$ time periods are the post-treatment periods. For the treated unit, at each time period $t = 1, \ldots, T$, let $Y_t(0), Y_t(1) \in \mathbb{R}$ be the counterfactual outcome corresponding to no treatment and treatment, respectively, and $Y_t = \mathbb{1}(t \leq T_0)Y_t(0) + \mathbb{1}(t > T_0)Y_t(1)$ be the observed outcome. At each time period $t$, other variables such as other control units’ outcomes are observed. We provide more details about these variables below. We treat $T$, $T_0$ and the treated unit as deterministic, and treat other variables such as the treated unit’s potential outcomes $Y_t(1)$ and $Y_t(0)$ as random. In other words, our proposed methods are conditional on the study
We study the ATT causal estimand, that is,

$$\phi^*(t) = E[Y_t(1) - Y_t(0)]$$

in a post-treatment period $t > T_0$. We treat times, namely $T$ and $T_0$, and all units as deterministic, and treat the potential outcomes as stochastic (Greenland, 1987; Robins and Greenland, 1989, 2000; VanderWeele and Robins, 2012); that is, $Y_t(0)$ and $Y_t(1)$ are both stochastic processes indexed by $t$ that are randomly generated over time, rather than fixed unknown scalar sequences. In the frequentist interpretation, under repeated sampling, the times and units are all fixed and hence identical for all samples, but the outcomes are randomly generated from a fixed unknown joint distribution and hence may differ across samples. Stochastic counterfactuals are commonly assumed in the SC literature, implicitly in the random noise or residuals of a linear latent factor model or an autoregressive model (e.g. Abadie et al., 2010; Abadie and L’Hour, 2021; Athey et al., 2021; Ben-Michael et al., (2021b, 2022)). This notion of stochastic counterfactuals $Y_t(0)$ and $Y_t(1)$ as time series is required in our paper because the expectation in $\phi^*(t)$ is taken over the joint distribution of $(Y_t(0), Y_t(1))$.

In the main text, we focus on the case without covariates, and discuss using covariates in Web Appendix S5. We assume that all unmeasured confounding is captured by a latent factor $U_t$, with assumptions stated in later sections. We use $t_-$ and $t_+$ to denote generic times before and after treatment, respectively; that is, $t_- \leq T_0$ and $t_+ > T_0$. When stating asymptotic results, we consider the asymptotic regime where $T \rightarrow \infty$ with $T_0/T \rightarrow \gamma \in (0, 1)$. This asymptotic regime may be interpreted as the number of observations in both pre- and post-treatment periods growing to infinity, collecting more data before and after the treatment time. Beyond this asymptotic regime, finite-sample results concerning the error in pre-treatment fitting or treatment effect estimation have been established in previous works (e.g. Abadie and L’Hour, 2021; Athey et al., 2021; Ben-Michael et al., (2022, 2021b)).

3 REVIEW OF IDENTIFICATION VIA OUTCOME MODELING

In classical SC methods, a weighted average of a pool of control units called donors forms the SC (Abadie et al., 2010). The motivation for using control units’ outcomes to learn about $Y_t(0)$, despite the presence of potential unmeasured confounder $U_t$, is that these control units may be affected by, and thus contains information about, $U_t$. This characteristic resembles that of proxies in proximal causal inference. In an i.i.d. setting, proxies capture the effect of the unmeasured confounder on the outcome or treatment assignment. They may also be viewed as noisy observations of the unmeasured confounder. In a panel data setting, for each time $t$, we use $W_t \in \mathcal{W}$ to denote a proxy, the vector of the donors’ outcomes in classical SC settings. Under commonly assumed data-generating assumptions such as the linear factor model (Abadie et al., 2010), donors’ outcomes are ideal proxies of $U_t$: any variation in $U_t$ induces some variation in $W_t$. We use $Z_t \in \mathcal{Z}$ to denote a general supplemental proxy. We next state the causal conditions required and discuss the role of and the choice of $Z_t$. 

![Causal graphs satisfying Condition 1 at each time period t. The variable U_t is the unobserved confounder. In Figure 1a, additional unmeasured confounding between proxy W_t and the treated unit's outcome Y_t may be present. In Figure 1b, Z_t, W_t, and Y_t are mutually independent conditional on U_t, which is often sensible when (W_t, Z_t) are control units' outcomes.](attachment:image.png)

**Condition 1** For all pre-treatment time points $t$, the supplemental proxies are independent of the outcomes and the proxies, conditional on the confounders: $Z_t \perp \! \! \! \perp (Y_t, W_t) \mid U_t$.

Conditions similar to Condition 1 are common in proximal causal inference literature (e.g., Miao and Tchetgen Tchetgen, 2016; Miao et al., 2018; Cui et al., 2020; Tchetgen Tchetgen et al., 2020). As we will show later, $W_t$ is used to model the outcome $Y_t$ with $Z_t$ supplementing for identification, while $Z_t$ is used to model the weighting process introduced in the following section with $W_t$ supplementing for identification. The conditional independence of Condition 1 is implied by the factor model in classical SC (e.g., Abadie et al., 2010), so such assumptions are commonly made implicitly. Causal graphs of Condition 1 are in Figure 1.

**Condition 2** There exists a function $h^* : \mathcal{W} \rightarrow \mathbb{R}$, $h^*(W_t)$ that captures the conditional mean of $Y_t(0)$ given $U_t$: $E[h^*(W_t) \mid U_t] = E[Y_t(0) \mid U_t]$, for all time points $t$.

The function $h^*$ is called an outcome confounding bridge function in proximal causal inference, a terminology we adopt. Condition 2 states that the transformation $h^*(W_t)$ of the observable proxy $W_t$ matches the unobservable $Y_t(0)$ in expectation conditional on $U_t$. Sufficient conditions for the existence of $h^*$ and examples of $h^*$ can be found in Shi et al. (2023). The most popular model for $h^*$ is the linear model in classical SC methods (Abadie et al., 2010)—that is, the average of $Y_t(0)$ may be imputed by a linear combination of the donors’ outcomes $W_t$—but $h^*$ may also be nonlinear. This assumption substantially generalizes the form of SCs by allowing more flexible models. In this condition, we implicitly assume that $W_t$ is not causally impacted by the treatment because of the constant relationship over time.

Throughout, proxy $W_t$ will consist of donors’ outcomes a priori known not to be causally impacted by the treatment, to make Conditions 1–2 plausible. When many donors are viable, a subset may be selected based on, for example, the similarity of their outcome trajectories to the treated unit’s trajectory before treatment. Typical choices of $Z_t$ that may satisfy Condition 1 include (i) outcomes of control units that are not valid donors, (ii) outcomes of donors excluded from the model $h^*$ to impute $Y_t(0)$, and (iii) covariates of donors that are contemporaneous with $(Y_t, W_t)$. The proxy $Z_t$ may be impacted by the treatment. Shi
et al. (2023) Section 2.2 contains more discussion about how to choose proxies \( W_t \) and \( Z_t \).

Shi et al. (2023) showed that, under Condition 2, for all post-treatment time points \( t_+ > T_0 \),

\[ E[Y_{t_+}(0)] = E[h^*(W_{t_+})] \]

and thus the ATT \( \phi^*(t_+) = E[Y_{t_+} - h^*(W_{t_+})] \). Additionally under Condition 1, \( h^* \) solves

\[ E[Y_t - h(W_t)] | Z_t = 0 \quad \text{for all pre-treatment time points} \]

\[ t_0 \leq T_0 \] (2)

in \( h : \mathcal{W} \rightarrow \mathbb{R} \). Further, under Condition S1 in Web Appendix S1, (2) has a unique solution.

**Remark 1** In principle, it may be possible to allow \( h^* \) to depend on \( t \), but it may be challenging to obtain a consistent estimator because only one observation is available at each time point; assumptions such as smoothness may be necessary. Our subsequent confounding bridge functions can also depend on \( t \), but we do not pursue this direction.

### 4 WEIGHTED AND DOUBLY ROBUST IDENTIFICATION OF ATT

The method reviewed above is solely based on the treated unit’s outcome process model, similar to G-computation under unconfoundedness (Robins, 1986) and proximal G-computation (Tchetgen Tchetgen et al., 2020). Now, we introduce our first novel SC method, which is based on weighting and similar to inverse probability weighting under unconfoundedness (Robins et al., 1994) and proximal inverse probability weighting (Cui et al., 2020).

To illustrate the idea, for the moment, suppose that the observations are i.i.d. across time. Then, the estimand \( \phi^*(t_+) = E[Y_{t_+}(1) - Y_{t_+}(0)] \) corresponds to an average treatment effect on the treated, for which several identification formulas exist (eg., Hahn, 1998; Imbens, 2004; Chen et al., 2008), including those based on outcome regression and weighting. Therefore, one may identify \( E[Y_{t_+}(1) - Y_{t_+}(0)] \) in a SC setting via weighting. To avoid the issue that propensity scores are undefined conditional on the design, we use the concept of covariate shift (eg., Quinonero-Candela et al., 2009) and likelihood ratio weighting. We next describe our needed causal conditions to identify the ATT via weighting for panel data.

**Condition 3** The joint conditional distribution of the counterfactual outcome \( Y_t(0) \) and proxy \( W_t \) given \( U_t \) is identical for all time points \( t \).

Intuitively, this condition states that once the process \( U_t \) is generated, \( (Y_t(0), W_t) \) are then generated in the same way for all \( t \). The invariance of the conditional distribution of \( W_t \) given \( U_t \) rules out a causal effect of the treatment on the proxies \( W_t \), analogously to the exclusion restriction property of negative control outcomes (Lipsitch et al., 2010; Miao et al., 2018).

The next condition states that the marginal distribution of the confounder \( U_t \) is identical for all \( t_0 \), which is implied by stationarity of \( U_{t_0} \). Moreover, this condition ensures that only one approach is needed to impute post-treatment counterfactual outcomes \( Y_{t_+}(0) \).

**Condition 4** The distribution of the unobserved confounder \( U_{t_0} \) is identical for all post-treatment time points \( t_+ > T_0 \).

Conditions 3 and 4 together imply that the distribution of \( (Y_{t_0}(0), W_{t_0}, U_{t_0}) \) is identical for all \( t_0 \). This implication holds under stationarity after treatment, without necessarily requiring stationarity before treatment. Even if stationarity also holds before treatment, these two conditions hold if the distributions in these two periods differ. Stationarity or similar are often used in time series analysis and other areas, including SCs (eg., Hsiao et al., 2012; Hahn and Shi, 2017; Li, 2020; Cattaneo et al., 2021; Chernozhukov et al., 2021; Fermand Pinto, 2021). Conditions 3–4 may thus be plausible in certain applications. Stationarity in Condition 4 might be implausible when the number \( T - T_0 \) of post-treatment time periods is large. We assume Condition 4 to facilitate the presentation and present an identification approach applicable to non-stationary cases in Web Appendix S 3.1. These two conditions allow an instantaneous distributional shift in the unobserved confounder \( U_t \) at treatment \( T_0 \). Such an instantaneous shift is a source of confounding. Therefore, in general, pre-treatment outcomes \( Y_{t_0} \) cannot be directly used to impute post-treatment counterfactual outcomes \( Y_{t_+}(0) \). We also note that, although these two conditions appear to require instantaneous distributional shift, in practice one may specify a window around the treatment in which the transition may not be instantaneous, and restrict the pre- and post-treatment periods before and after the specified window, respectively, so that Conditions 3 and 4 are plausible.

The next condition states the existence of a treatment confounding bridge function (Cui et al., 2020), which models the likelihood ratio, namely a Radon–Nikodym derivative, \( dP_{U_{t_+}} / dP_{U_{t_0}} \) via a regression of the supplemental proxy \( Z_{t_0} \) on the unobserved confounder \( U_{t_0} \).

**Condition 5** For all times \( t_0 \leq T_0 \) and \( t_+ > T_0 \) the distribution \( P_{U_{t_0}} \) of the unobserved confounder \( U_{t_0} \) after treatment is dominated by that before treatment, namely \( P_{U_{t_0}} \), and there exists a function \( g^* : \mathcal{Z} \rightarrow \mathbb{R} \) capturing the distributional shift in \( U_{t_0} \); that is,

\[ E[q^*(Z_{t_0}) | U_{t_0} = u] = \frac{dP_{U_{t_0}}}{dP_{U_{t_0}}} (u) \] (3)

The left-hand side of (3) does not depend on \( t_+ \), due to Condition 4. We call \( q^* \) the treatment confounding bridge function, encoding information on how pre-treatment outcomes \( Y_{t_0} \) can be used to impute post-treatment counterfactual outcomes \( Y_{t_+}(0) \). The function \( q^* \) is closely related to that for estimating the ATT in i.i.d settings, such as in Theorem S.1 of Cui et al. (2020). One consequence of Condition 5 is that the distribution of \( U_{t_0} \) must be dominated by that of \( U_{t_0} \), for all \( t_0 \leq T_0 \) and \( t_+ > T_0 \). As shown in Theorem 1 Equation 6 below, under Conditions 1 and 3–5, this further implies that the distribution of \( W_{t_0} \) is dominated by that of \( W_{t_0} \) for all \( t_0 \leq T_0 \) and \( t_+ > T_0 \) which is a testable condition. Thus, for time series data subject to a significant secular trend, especially a monotone trend, we recommend
a pre-processing step to remove, to the extent possible, any significant secular trends to make Condition S as plausible as possible. We provide concrete approaches for removing time trends in Web Appendix S8 and the corresponding simulation results in Web Appendix S7.6. Although detrending with our approach may lead to slightly to moderately anti-conservative inference, failing to correct for such a trend will likely compromise one’s ability to implement the proposed weighted approach successfully. It is still an open question how to account for time trends appropriately to obtain asymptotically valid inference in our proposed methods. Sufficient conditions for the existence of \( q^* \) can be found in Web Appendix S9. We list a few examples of treatment confounding bridge functions below.

**Example 1** Suppose that, for all \( t_0 \leq t \leq T \) and \( t, U_{i.} \sim N(0, \sigma^2), U_{i.t} \sim N(0, \sigma^2), Z_{i}|U_{i} \sim N(\alpha U_{i}, \sigma^2) \) for some \( a \neq 0 \). If \( \sigma^{-2} a^2 - \sigma^{-2} > 0, \) then \( q^* : z \mapsto \exp(\alpha + \beta z^2) \) for \( \beta = \sigma^{-2}(-\sigma^{-2} - \sigma^{-2})/12(\sigma^{-2} a^2 - \sigma^{-2} a^2 - \sigma^{-2}) \) and some \( \alpha \in \mathbb{R} \) satisfies Condition S. In this example, we only specify the marginal distributions of \( (U_{i.}, Z_{i}) \) for each \( t \) and allow for serial correlation. For instance, \( U_{i.t} \) can be generated from an autoregressive model that is stationary before and after treatment, respectively. This might occur if the treatment is implemented at a random time point \( T_0 \) due to an abrupt change in \( U_{i.t} \) in a time window around \( T_0 \); with the observations in this window excluded from analysis, \( U_{i.t} \) can be stationary before and after treatment with possibly different distributions. Moreover, if \( U_{i.t} \) is stationary, Condition S holds with the constant bridge function \( q^* : z \mapsto 1 \) even if \( T_0 \) is random. Similar results hold for data marginally distributed as multivariate normal at each time \( t \).

**Example 2** Suppose that \( U_{i} = (U_{i.1}, \ldots, U_{i.K}) \) has all coordinates mutually independent at any time \( t \), while there may be serial correlations for each element over time. Let \( U_{i.t.K} \sim \text{Exponential}(\lambda_{-}) \) and \( U_{i.t.K} \sim \text{Exponential}(\lambda_{+}), \) for \( k = 1, \ldots, K \), with mutually independent entries distributed as, conditional on \( U_{i.}, U_{i.K} \sim \text{Poisson}(\alpha + \beta U_{i.1}, U_{i.K}) \) for some scalars \( \alpha \geq 0 \) and \( \beta > 0 \). Then, direct calculation shows that \( q^* : z \mapsto \prod_{k=1}^{K} \exp(a + bz_k) \) satisfies Condition S for some scalars \( a \) and \( b \).

**Remark 2** Condition S may fail when \( T_0 \) is random and depends on the unobserved confounder \( U_{i.t} \), even when we condition on \( T_0 \) to mimic a fixed treatment time design. Consider the following counterexample. Suppose that \( U_{i.t} \geq 0 \) are i.i.d. across \( t \), with support being \( \mathbb{R}^+ \), and \( T_0 = \max\{ t : U_{i.t} < a \} \) for an unknown fixed number \( a \). This corresponds to the case where the treatment initiates immediately when \( U_{i.t} \) crosses the threshold \( a \). We assume i.i.d. and unbounded \( U_{i.t} \) only for simplicity. Conditional on \( T_0 \), since \( \text{Pr}(U_{i.t} < a) = 1 \) but \( \text{Pr}(U_{i.t} \geq a) > 0 \), the Radon-Nikodym derivative \( dP_{U_{i.t}}/dP_{U_{i.t}} \) in Condition S does not exist. Thus, this condition fails. Therefore, a fixed treatment time is crucial to the weighted approach to ATT, and conditioning on a random \( T_0 \) might not suffice.

The above conditions lead to our first formal identification result.

**Theorem 1** (Identification of ATT with \( q^* \)) Let \( f : \mathbb{R} \rightarrow \mathbb{R} \) be any square-integrable function. Under Conditions 1 and 3–5, for all \( t_0 \leq t_0 \),

\[
E[f(Y_{i.}(0))] = E[q^*(Z_{i.}) f(Y_{i.})].
\]

(4)

In particular, taking \( f \) to be the identity function, it holds that

\[
E(Y_{i.}(0)) = E(q^*(Z_{i.}) Y_{i.})
\]

(5)

and thus the ATT is identified as \( \phi^*(t_+) = E(Y_{i.} - q^*(Z_{i.}) Y_{i.}) \) for all \( t_0 \). In addition, \( \delta_{W_{i.+,j}} \) is dominated by \( \delta_{W_{i.+,j}} \), the treatment confounding bridge function \( q^* \) solves

\[
E[q(Z_{i.}) | W_{i.} = w] = \frac{dP_{W_{i.}}(w)}{dP_{W_{i.}}} \text{ for all } t_0.
\]

(6)

in \( q : \mathcal{Z} \rightarrow \mathbb{R} \). Further, under Condition S2 in Web Appendix S1, (6) has a unique solution.

In contrast to SC methods based on outcome modeling, the identifying expression (5) cannot be directly interpreted as an outcome trajectory. Indeed, given a treatment confounding bridge function \( q^* \), the right-hand side of (5) only depends on observations before treatment but not after treatment. We also obtain a novel doubly robust identification result, which is motivated by doubly robust estimation of the average treatment effect on the treated (Cui et al., 2020) and is the basis of doubly robust inference.

**Theorem 2** (Doubly robust identification) Let \( h : \mathcal{W} \rightarrow \mathbb{R} \) and \( q : \mathcal{Z} \rightarrow \mathbb{R} \) be square-integrable. Under Conditions 1, 3, and 4, for all \( t_0 \leq T_0 \), we have

\[
\phi^*(t_+) = E\left( Y_{i.} - q(Z_{i.}) (Y_{i.} - h(W_{i.})) - h(W_{i.}) \right),
\]

(7)

if (i) Condition 2 holds and \( h = h^* \), or (ii) Condition 5 holds and \( q = q^* \).

Theorem 2 states that the ATT \( \phi^*(t_+) \) is identified by a single formula if at least one of the two nuisance functions \( h^* \) or \( q^* \) is known, and thus doubly robust estimation (Robins and Rotnitzky, 1995; Robins et al., 1995; Scharfstein et al., 1999; Bang and Robins, 2005) is possible. This result holds even if one of \( h^* \) and \( q^* \) exists but the other does not; that is, Conditions 2 and 5 need not hold simultaneously.

### 5 Doubly Robust Inference About ATT

We assume that one specifies parametric models for the confounding bridge functions \( h^* \) and \( q^* \). We use \( \alpha \) and \( \beta \) to denote the models for confounding bridge functions parameterized by \( \alpha \in \Lambda \subseteq \mathbb{R}^{d_\alpha} \) and \( \beta \in \mathbb{B} \subseteq \mathbb{R}^{d_\beta} \), respectively. For example, in classical SC methods, it is typically assumed that \( h^*(w) = w^\top \alpha_0 \) for a vector \( \alpha_0 \) of non-negative numbers that sum up to unity (Abadie et al., 2010). In this case, we may take \( h_\alpha \) to be \( w \mapsto w^\top \alpha \). In some cases, for example, in Example 2, we may take \( q_\beta(z) = \exp(\beta_0 z + \beta_1 z^2) \) where \( \beta = (\beta_0, \beta_1) \).

We assume that the function \( f(t^+) \mapsto \phi^*(t^+) \) encoding the potentially time-varying ATT is correctly parameterized by \( \lambda \in \Lambda \subseteq \mathbb{R}^{d_\lambda} \). We use \( \beta_0 \) to denote this model. For example, the ATT is commonly assumed to be constant overtime, which holds under stationarity of \( (Y_{i.}, t^+) \) and Conditions 3–4. In this case, we may set \( \beta_0 \) to be constant \( \lambda \in \mathbb{R} \).
By (2) and (6), respectively, we have that
\[
E[\{Y_t - h^s(W_{t-})\}g_x(Z_t)] = 0 \quad \text{and} \quad E(q^s(Z_t)g_y(W_{t-})) - g_y(W_{t-}) = 0
\]
(8)
for any functions $g_x : \mathbb{Z} \to \mathbb{R}$ and $g_y : \mathbb{W} \to \mathbb{R}$. We propose to use the generalized method of moments (GMM) to estimate $h^s$, $q^s$, and the ATT (eg, Hansen, 1982; Wooldridge, 1994; Hall, 2007). This method involves a parameter vector $\theta = (\alpha, \beta, \lambda, \psi, \psi_\cdot)$ in $\Theta$ including nuisance parameters $\psi$ and $\psi_\cdot$, a moment equation $G_t$ for each time $t$, and a weight matrix $\Omega_t$ that may depend on sample size $T$. Details of this method are presented in Web Appendix S2. We need some additional conditions to obtain valid inferences about $\phi^s$.

**Condition 6** (i) There is a unique parameter value $\theta_\infty = (\alpha_\infty, \beta_\infty, \lambda_\infty, \psi_\infty, \psi_\cdot) \in \Theta$ such that $E(G_t(\theta_\infty)) = 0$ for all time points $t$. (ii) The function $h^s = h_{\theta_\infty}$ is a valid outcome confounding bridge function satisfying Condition 2, or the function $q^s = q_{\theta_\infty}$ is a valid treatment confounding bridge function satisfying Condition 5.

Part (i) of Condition 6 is an identifying condition to ensure a unique solution to the population moment equation, which is standard for GMM. If both $h^s$ and $q^s$ are correctly specified, part (i) would hold if they are both unique. Part (ii) requires correct parametric specification of at least one of, but not necessarily both of, $h^s$ or $q^s$, without necessarily knowing a priori which model might be incorrect. Under Condition 6, Theorem 2 implies that $\phi_{\theta_\infty}$ equals the target estimand, ATT $\phi^s$. By standard asymptotic theory for GMM (eg, Theorems 7.1 and 7.2 in Wooldridge, 1994) along with Theorem 2, we have that $\hat{\theta}_T$ is consistent for $\theta_\infty$ and asymptotically normally distributed under conditions, as stated in Theorem 3 below. If the data are i.i.d., as we argued in Section 4, the estimand reduces to the average treatment effect on the treated and our estimator is locally efficient (Cui et al., 2020).

**Theorem 3** Under Conditions 1, 3, 4, and 6, as well as S3–S5 and S7 in Web Appendix S2, with the estimator $\hat{\theta}_T$ from (S1) and $\theta_\infty$ in Condition 6, it holds that, as $T \to \infty$, $\hat{\theta}_T$ is consistent for $\theta_\infty$. Additionally, under Conditions S6, S8, and S9, as $T \to \infty$,
\[
\sqrt{T}(\hat{\theta}_T - \theta_\infty) \overset{d}{\to} N(0, A^{-1}B^{-1}),
\]
where $A = R^T \Omega R$, $\Omega$ is from Condition S3,
\[
R = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} E\{\nabla_\theta G_t(\theta)\}_{\theta = \theta_\infty},
\]
\[
B = R^T \Omega \left[ \lim_{T \to \infty} \operatorname{var} \left\{ T^{-1/2} \sum_{t=1}^{T} G_t(\theta_\infty) \right\} \right] \Omega R.
\]
Suppose that the ATT function $\phi_\lambda$ is differentiable with respect to $\lambda$ and let $\phi_\lambda$ denote this partial derivative. Thus, with $\Pi := (0_{d_\psi \times (d_\psi + d_\theta)}, I_{d_\psi \times (d_\psi + d_\theta)})$ being the matrix consisting of zeros and ones with dimensions denoted in the subscript, for every $t_\psi \geq T_0$, it holds that $\phi_{\theta_\infty}$ is the ATT at time $t_\psi$ and $\sqrt{T}(\phi_{\lambda}(t_\psi) - \phi_{\lambda}(t_\psi)) \overset{d}{\to} N(0, \phi_{\theta_\infty}(t_\psi) \Pi^T A^{-1}B^{-1} \Pi \phi_{\theta_\infty}(t_\psi)).$

One can in principle use any GMM implementation, and we use the standard R package gmm for our simulation and data analyses. Confidence intervals of the ATT follows from standard outputs of gmm. In particular, when $\phi_\lambda$ is a constant function, since $\lambda_\infty$ is a component of $\theta_\infty$, a Wald test or confidence interval about the ATT $\phi_{\theta_\infty}(t_\psi) = \lambda_\infty$ follows immediately from inference about $\theta_\infty$. We have noted some numerical instabilities with this implementation in our simulations, and provide our empirical suggestions to alleviate numerical issues in Web Appendix S6. We expect these issues to be alleviated by using an improved GMM software implementation with, for example, more numerically stable optimization algorithms that are more capable to handle nonconvex problems, potentially under constraints.

Our asymptotic results in Theorem 3 rely on the total number $T$ of time periods tending to infinity. Thus, our proposed doubly robust method is applicable to cases with a large number of time periods and a relatively small number of model parameters. With many control units, expertise might be required to reduce the number of parameters before analysis using our method. Methods with different theoretical results that allow for a large number of control units include Athey et al. (2021), Ben-Michael et al. (2021b, 2022), among others.

The method and results corresponding to the identification results in Theorem 1 based on weighting alone are similar and can be found in Web Appendix S4. This is largely based on the outcome modeling-based approach developed by Shi et al. (2023). Compared with these two methods, the doubly robust method has the advantage that it only requires correct specification of one of $h^s$ and $q^s$ in a parametric model, but not necessarily both. One potential drawback of the doubly robust method, however, is that more parameters need to be estimated in GMMs compared with the other two methods. This issue of dimensionality might limit the usage of complicated models for $h^s$ and $q^s$ when the time series is short. In particular, if the number of parameters is comparable to the total number $T$ of time periods, the doubly robust method might lead to numerical instability due to too many parameters being estimated and might be impractical. Our methods have no guarantee in such scenarios. However, the number of parameters is much smaller than $T$ in many applications, for example, in our data analyses in Section 7 and Web Appendix S8. The number of control units is often highly related to the number of parameters, and the number of control units is commonly small compared to the number of time periods in SC applications, including the motivating ones (eg, Abadie and Gardeazabal, 2003; Abadie et al., 2010). Models with many variables might possibly be used by utilizing methods in Deane (2021) under linearity and sparsity. Chamberlain (1992) and Hansen (1982, 1985) showed that GMM attains the efficiency bound under conditional moment restrictions with a suitably chosen weighting matrix $\Omega_T$ in i.i.d settings as well as with panel data settings where asymptotics are in the number of units. We do not pursue high-dimensional models or efficiency further in this work and restrict ourselves to the setting of a bounded number of units (a single treated unit and a finite number of control units) while relying on large $T$ asymptotics; to the best of our knowledge, none of these prior results can directly apply to our more challenging setting and existing semiparametric efficiency.
theory does not appear to directly apply without additional restrictions (eg, Markov restrictions).

6 SIMULATIONS
We investigate the performance of our methods to estimate the constant ATT $\phi^* \equiv \phi^*(t_*)$ in several simulations. Here, we present the first simulation where the moment equations in the GMM are just identified. We compare the following methods: correct.DR, the doubly robust method with correctly specified parametric $h^*$ and $q^*$; correct.h, the outcome confounding bridge method from Shi et al. (2023) with correctly specified $h^*$; correct.q, the treatment confounding bridge method described in Web Appendix S4 with correctly specified $q^*$; mis.h.DR, the doubly robust method with misspecified $h^*$ and correctly specified $q^*$; mis.q.DR, the doubly robust method with misspecified $q^*$ and correctly specified $h^*$; mis.h, the outcome confounding bridge method with misspecified $h^*$; mis.q, the treatment confounding bridge method with misspecified $q^*$; OLS (ordinary least squares), the method based on unconstrained ordinary least squares and similar to the method from Abadie et al. (2010). OLS finds the linear combination of donors’ outcomes that best fits the treated unit’s outcome before treatment, and uses this combination as the SC. All methods except OLS are based on the proximal causal inference perspective. We let the number of latent confounders, the number of donors and of the other control units, all equal to $K$, range in $\{2,3,4,5\}$. Details of the data-generating mechanism are presented in Web Appendix S7.1.

We next present the simulation results. We have run 16,000 GMM involving weighting via a treatment confounding bridge function. Among them, only one run had numerical errors. The sampling distributions of the estimated ATT is presented in Figure 2. In all settings, the OLS estimator is biased. When at least one of $h^*$ or $q^*$ is correctly specified, our proposed doubly robust method appears consistent and asymptotically normal, aligning with Theorem 3. The other two methods, based on only one of $h^*, q^*$, are not doubly robust. They are consistent and asymptotically normal only when the bridge function they rely on is correctly specified, but are biased otherwise. The 95%-Wald confidence interval coverage of all above methods is presented in Figure 3. In large samples, confidence intervals based on consistent and asymptotically normal estimators have coverage close to the nominal level; otherwise, the confidence interval coverage is much lower than the nominal level.

7 ANALYSIS OF BRAZIL ALL-CAUSE PNEUMONIA HOSPITALIZATION
We study the effect of the introduction of the PCV10 vaccine in Brazil on the number of hospitalizations due to all-cause pneumonia (Bruhn et al., 2017a). Monthly hospitalizations and their causes were collected from 2003 to 2013. We focus on the subpopulation of children less than 12 months old in this analysis. PCV10 was introduced to Brazil in January 2010. Following the analysis in Bradley and Tone (2017), we allow two years for the introduction of the vaccine to take effect and set the evaluation period to be 2012–2013.

We view each group of causes of hospitalization as a unit and dismiss units with missing data. The time series data does not have a clear monotone trend; thus, Condition 5 may be plausible. To alleviate numerical issues due to nonlinearity in GMM and to reduce differences in scaling between units, we scale the numbers of hospitalizations due to each group of causes to the unit

---

**FIGURE 2** Sampling distribution of estimated ATT. The horizontal dotted line is the true ATT.
We use all other groups of causes as control units in the supplemental proxies $Z_i$. We consider three parametrizations of the treatment confounding bridge function with increasing sets of units included in the model: $q_β(Z_i) = \exp(β_0 + \sum_{j=1}^3 β_j Z_{ij})$ for $j = 1, 2, 3$, where $Z_{ij}$ are the numbers of hospitalizations due to (i) certain infectious and parasitic diseases, except intestinal, (ii) diseases of blood and blood-forming organs and certain disorders involving the immune mechanism, and (iii) premature delivery and low birth weight. The motivation for choosing these supplemental proxies $Z_i$ is to capture the effect of unmeasured confounders $U_i$ on general health issues related to infections and the immune system, which are both associated with the outcome of interest, all-cause pneumonia hospitalizations, among children less than 12 months old. We refer to the corresponding doubly robust (resp., weighted) estimators as DR, DR2, and DR3 (resp., treatment bridge, treatment bridge2, and treatment bridge3). As shown in the simulation in Web Appendix S7.7 choosing $W$ and $Z$ completely based on the data might lead to estimates with increased uncertainty due to the possibility of model misspecification. Such estimates may not be informative. When selecting units empirically, we recommend, when possible, using prior knowledge and not solely relying on the data, in which case this issue might be avoided. Ideally, proxies should be selected a priori to avoid potential post-selection inference issues.

The GMM estimator outlined in Section 5 is implemented with the user-specified functions being $g_ν: z \mapsto (1, z^T)^T$, where we recall that $z$ is the collection of all supplemental proxies, namely hospitalizations due to non-donor causes, and $g_ν: w \mapsto (1, w^T)^T$. We set the weight matrix to equal the identity matrix.

Besides proximal SC estimators, we also report results for the standard OLS estimator described in Section 6. We also consider the regression-based SC method (Abadie’s SC) proposed by Abadie et al. (2010) with the same three donors forming the proxy $W_i$ as above. The point estimates and 95% confidence intervals for the ATT are presented in Table 1. The trajectories of SCs and the actual number of hospitalizations are presented in Figure 4a.

Although we have scaled all outcomes to fall in the unit interval, Abadie’s SC does not output a good pre-treatment fit and its ATT estimate appears unreliable. Because the original scales of the outcomes across units differ substantially, the constraints in Abadie’s SC (i) that the intercept vanishes, and (ii) that the weights are non-negative and sum to one, might not be appropriate in this application (Doudchenko and Imbens, 2016); scaling to unit interval might still fail to justify the adequacy of these constraints. All other methods conclude a significant decrease in hospitalizations due to all-cause pneumonia.
TABLE 1 Estimate of the ATT of PCV10 and placebo treatment on the number of hospitalizations due to all-cause pneumonia among children less than 12 months old in Brazil for various methods with 95%-Wald confidence intervals.

| Method                | PCV10          | placebo        |
|-----------------------|----------------|----------------|
| Abadie’s SC           | 409            | 3092           |
| OLS                   | −3533 (−4137, −2930) | 253 (−287, 794)  |
| DR                    | −2745 (−3559, −1931) | 1192 (501, 1884)  |
| DR2                   | −3527 (−4663, −2392) | 317 (−407, 1042)  |
| DR3                   | −3548 (−6036, −1061) | 260 (−246, 767)   |
| Outcome bridge        | −3646 (−4693, −2598) | 565 (−224, 1355)  |
| Treatment bridge      | −3989 (−4373, −3605) | −532 (−1638, 574) |
| Treatment bridge2     | −3814 (−4941, −2688) | −205 (−1542, 1133) |
| Treatment bridge3     | −3895 (−6401, −1388) | 97 (−502, 695)    |

Abadie’s SC (Abadie et al., 2010) does not readily provide confidence intervals.

FIGURE 4 Trajectories of SCs (green dashed) and the number of hospitalizations due to all-cause pneumonia (red solid) among children less than 12 months old in Brazil for various methods. The vertical line is the last time point (month) before implementing the treatment. Methods based on weighting only (treatment bridge, treatment bridge2, and treatment bridge3) do not have SC trajectories.

after the introduction of PCV10, as expected. The estimate from DR is somewhat different from other proximal methods involving a treatment confounding bridge function with at least two units, suggesting a model mis-specification affecting DR in finite samples. Though the theory suggests that DR should be consistent, when at least one nuisance function is correctly specified, a larger sample size T than this data set might be needed for DR to be close to the truth.

We also conduct a falsification analysis. We consider a hypothetical placebo treatment in January 2009 and estimate its effect in the year 2009. The analysis results are presented in Table 1 and Figure 4b. Similarly to the main analysis, the pre-treatment fit from Abadie’s SC appears unreliable in this setting; therefore confirming the recommendation that SCs might perform poorly when the pre-treatment fit is suboptimal (Abadie et al., 2010). Our proposed methods perform better and offer alternative justification for SC methods in such settings. In fact, 95% confidence intervals from most proximal SC methods cover zero, correctly indicating a non-significant effect due to the placebo treatment; so does OLS in this case. The only exception DR echoes the poor performance from the main analysis. We conclude that, in this application, when the nuisance functions are approximately correctly specified, the proximal methods are among the best.

We also study the effect of Florida’s “stand your ground” law on homicide rates and the effect of a tax cut in Kansas on economic outcomes (see Web Appendix S8). In these analyses, our proposed doubly robust method outperforms Abadie’s SC or OLS in several cases.

8 Discussion

To estimate the ATT in panel data settings, classical SC methods often require correct specification of nuisance functions. Our proposed doubly robust methods involve estimating two nuisance functions but allow for mis-specification of one, without knowledge about which is mis-specified. Our identification results may enable the development of new SC methods, especially nonparametric or semiparametric ones that rely on weaker conditions. Such methods can extend the idea of SCs and proximal causal inference to more general applications.

Acknowledgments

We are grateful to an anonymous reviewer for a counterexample of Condition 5.
SUPPLEMENTARY MATERIALS

Supplementary material is available at Biometrics online. Web Appendices and Figures referenced in Sections 2, 4, 5, and 6 are available with this paper at the Biometrics website on Oxford Academic. Code for our simulations and data analyses can be found at https://github.com/QIU-Hongxiang-David/DR_Proximal_SC and online with this article.

FUNDING

This work was supported in part by the U.S. National Science Foundation and National Institutes of Health, and by Analytics at Wharton.

CONFLICT OF INTEREST

None declared.

DATA AVAILABILITY

The data that support the findings in this paper are openly available in Github at https://github.com/weinbergerlab/InterventionEvaluatR (Bruhn et al., 2017b) and at https://github.com/ebenmichael/augsynth (Ben-Michael et al., 2021a), and in Open Science Framework (OSF) at https://osf.io/6udsq/ (Esposito and Bonander 2020).

REFERENCES

Abadie, A., Diamond, A. and Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California’s Tobacco control program. Journal of the American Statistical Association, 105, 493–505.

Abadie, A., Diamond, A. and Hainmueller, J. (2015). Comparative politics and the synthetic control method. American Journal of Political Science, 59, 495–510.

Abadie, A. and Gardeazabal, J. (2003). The economic costs of conflict: A case study of the Basque country. American Economic Review, 93, 113–132.

Abadie, A. and L’Houle, J. (2021). A penalized synthetic control estimator for disaggregated data. Journal of the American Statistical Association, 116, 1817–1834.

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

Athey, S., Bayati, M., Doudchenko, N., Imbens, G. and Khosravi, K. (2021). Matrix completion methods for causal panel data models. Journal of the American Statistical Association, 116, 1716–1730.

Bai, J. and Ng, S. (2021). Matrix completion, counterfactuals, and factor analysis of missing data. Journal of the American Statistical Association, 116, 1746–1763.

Bang, H. and Robins, J. M. (2005). Doubly robust estimation in missing data and causal inference models. Biometrics, 61, 962–973.

Ben-Michael, E., Feller, A. and Rothstein, J. (2021a). augsynth: augmented synthetic control method. https://github.com/ebenmichael/augsynth. [Accessed 22 February 2023].

Ben-Michael, E., Feller, A. and Rothstein, J. (2021b). The augmented synthetic control method. Journal of the American Statistical Association, 116, 1789–1803.

Ben-Michael, E., Feller, A. and Rothstein, J. (2022). Synthetic controls with staggered adoption. Journal of the Royal Statistical Society. Series B: Statistical Methodology, 84, 351–381.

Bonander, C., Humphreys, D. and Degli Esposti, M. (2021). Synthetic control methods for the evaluation of single-unit interventions in epidemiology: a tutorial. American Journal of Epidemiology, 190, 2700–2711.

Bradley, R. C. and Tone, C. (2017). A central limit theorem for non-stationary strongly mixing random fields. Journal of Theoretical Probability, 30, 655–674.

Bruhn, C. A., Hetterich, S., Schuck-Paim, C., Kürtüm, E., Taylor, R. J., Lustig, R. et al. (2017a). Estimating the population-level impact of vaccines using synthetic controls. Proceedings of the National Academy of Sciences of the United States of America, 114, 1524–1529.

Bruhn, C. A., Hetterich, S., Schuck-Paim, C., Kürtüm, E., Taylor, R. J., Lustig, R. et al. (2017b). InterventionEvaluatR. https://github.com/weinbergerlab/InterventionEvaluatR [Accessed 22 February 2023].

Cattaneo, M. D., Feng, Y. and Titiunik, R. (2021). Prediction intervals for synthetic control methods. Journal of the American Statistical Association, 116, 1865–1880.

Chamberlain, G. (1992). Efficiency bounds for semiparametric regression. Econometrica, 60, 567.

Chen, X., Hong, H. and Tassozi, A. (2008). Semiparametric efficiency in GMM models with auxiliary data. Annals of Statistics, 36, 808–843.

Chernozhukov, V., Wuthrich, K. and Zhu, Y. (2021). An exact and robust conformal inference method for counterfactual and synthetic controls. Journal of the American Statistical Association, 116, 1849–1864.

Cui, Y., Pu, H., Shi, X., Miao, W. and Tchetgen, E. T. (2023). Semiparametric proximal causal inference. Journal of the American Statistical Association, 1–12.

Cunha, F., Heckman, J. J. and Schennach, S. M. (2010). Estimating the technology of cognitive and noncognitive skill formation. Econometrica, 78, 883–931.

Deane, B. (2018). Proxy controls and panel data. arXiv preprint arXiv:1810.00283v7.

Deane, B. (2021). Many proxy controls. arXiv preprint arXiv:2110.03973v1.

Doudchenko, N. and Imbens, G. W. (2016). Balancing, regression, difference-in-differences and synthetic control methods: a synthesis. Technical Report, Cambridge, MA: National Bureau of Economic Research.

Esposito, M. D. and Bonander, C. (2020). Data and replication code for: “Synthetic control methods for the evaluation of single-unit interventions in epidemiology: a tutorial”. https://osf.io/6udsq/ [Accessed 22 February 2023].

Ferman, B. and Pinto, C. (2021). Synthetic controls with imperfect pre-treatment fit. Quantitative Economics, 12, 1197–1221.

Freyberger, J. (2018). Non-parametric panel data models with interactive fixed effects. Review of Economic Studies, 85, 1824–1851.

Greenland, S. (1987). Interpretation and choice of effect measures in epidemiologic analyses. American Journal of Epidemiology, 125, 761–768.

Hahn, J. (1998). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. Econometrica, 66, 315.

Hahn, J. and Shi, R. (2017). Synthetic control and inference. Econometrics, 5, 52.

Hall, A. R. (2007). Generalized Method of Moments. Oxford: Oxford University Press.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. Econometrica, 50, 1029.

Hansen, L. P. (1985). A method for calculating bounds on the asymptotic covariance matrices of generalized method of moments estimators. Journal of Econometrics, 30, 203–238.

Holtz-Eakin, D., Newey, W. and Rosen, H. S. (1988). Estimating Vector Autoregressions with Panel Data. Econometrica, 56, 1371.

Hsiao, C., Steve Ching, H. and Ji Wen, S. (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, 705–740.

Imbens, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: a review. Review of Economics and Statistics, 86, 4–29.
Li, K. T. (2020). Statistical inference for average treatment effects estimated by synthetic control methods. *Journal of the American Statistical Association*, 115, 2068–2083.

Lipsitch, M., Tchetgen Tchetgen, E. and Cohen, T. (2010). Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology*, 21, 383–388.

Miao, W., Geng, Z. and Tchetgen Tchetgen, E. J. (2018). Identifying causal effects with proxy variables of an unmeasured confounder. *Biometrika*, 105, 987–993.

Miao, W. and Tchetgen Tchetgen, E. J. (2016). On varieties of doubly robust estimators under missingness not at random with a shadow variable. *Biometrika*, 103, 475–482.

Quiñonero-Candela, J., Sugiyama, M., Lawrence, N. D. and Schwaighofer, A. (2009). *Dataset Shift in Machine Learning*. MIT Press.

Rickman, D. S. and Wang, H. (2018). Two tales of two U.S. states: Regional fiscal austerity and economic performance. *Regional Science and Urban Economics*, 68, 46–55.

Robins, J. (1986). A new approach to causal inference in mortality studies with a sustained exposure period-application to control of the healthy worker survivor effect. *Mathematical Modelling*, 7, 1393–1512.

Robins, J. and Greenland, S. (1989). The probability of causation under a stochastic model for individual risk. *Biometrics*, 45, 1125.

Robins, J. M. and Greenland, S. (2000). Discussion of “Causal Inference without Counterfactuals” by A.P. Dawid. *Journal of the American Statistical Association*, 95, 477–482.

Robins, J. M. and Rotnitzky, A. (1995). Semiparametric efficiency in multivariate regression models with missing data. *Journal of the American Statistical Association*, 90, 122–129.

Robins, J. M., Rotnitzky, A. and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89, 846–866.

Robins, J. M., Rotnitzky, A. and Zhao, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, 90, 106–121.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41–55.

Scharfstein, D. O., Rotnitzky, A. and Robins, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association*, 94, 1096–1120.

Shi, X., Li, K., Miao, W., Hu, M. and Tchetgen, E. T. (2023). Theory for identification and inference with synthetic controls: a proximal causal inference framework. arXiv preprint arXiv:2108.13935v4.

Shu, H. and Tan, Z. (2018). Improved estimation of average treatment effects on the treated: local efficiency, double robustness, and beyond. arXiv preprint arXiv:1808.01408v1.

Tchetgen Tchetgen, E. J., Ying, A., Cui, Y., Shi, X. and Miao, W. (2020). An introduction to proximal causal learning. arXiv preprint arXiv:2009.10982v1.

VanderWeele, T. J. and Robins, J. M. (2012). Stochastic counterfactuals and stochastic sufficient causes. *Statistica Sinica*, 22, 379–392.

Wooldridge, J. M. (1994). Estimation and inference for dependent processes. In: *Handbook of Econometrics*. vol. 4, 2639–2738. Oxford, UK: Elsevier.