A Penalized Synthetic Control Estimator for Disaggregated Data

Alberto Abadie   Jérémy L'Hour
MIT   CREST

October 24, 2020

Abstract

Synthetic control methods are commonly applied in empirical research to estimate the effects of treatments or interventions on aggregate outcomes. A synthetic control estimator compares the outcome of a treated unit to the outcome of a weighted average of untreated units that best resembles the characteristics of the treated unit before the intervention. When disaggregated data are available, constructing separate synthetic controls for each treated unit may help avoid interpolation biases. However, the problem of finding a synthetic control that best reproduces the characteristics of a treated unit may not have a unique solution. Multiplicity of solutions is a particularly daunting challenge when the data includes many treated and untreated units. To address this challenge, we propose a synthetic control estimator that penalizes the pairwise discrepancies between the characteristics of the treated units and the characteristics of the units that contribute to their synthetic controls. The penalization parameter trades off pairwise matching discrepancies with respect to the characteristics of each unit in the synthetic control against matching discrepancies with respect to the characteristics of the synthetic control unit as a whole. We study the properties of this estimator and propose data-driven choices of the penalization parameter.

Alberto Abadie, Department of Economics, Massachusetts Institute of Technology and NBER, abadie@mit.edu. Jérémy L'Hour, CREST, ENSAE Paris and Insee, jeremy.l.hour@ensae.fr. We benefited from detailed comments by Guido Imbens, and from discussions with Victor-Emmanuel Brunel, Xavier D'Haultfœuille, Laurent Davezies, Yannick Guyonvarch, Clément de Chaisemartin, David Margolis and participants at the 14th IZA Labor Market Policy Evaluation, the 10th French Econometrics Conference, the MIT/NBER Conference on Synthetic Control and Related Methods and CREST and Université de Montréal internal seminars. Charlie Rafkin and Jaume Vives provided expert research assistance. NSF support through grant SES-1756692 is gratefully acknowledged. R and Matlab scripts are available at https://github.com/jlhourENSAE/pensynth.

1. Introduction

Synthetic control methods (Abadie and Gardeazabal, 2003; Abadie et al., 2010, 2015; Doudchenko and Imbens, 2016) are often applied to estimate the treatment effects of aggregate interventions (see, e.g., Kleven et al., 2013; Bohn et al., 2014; Hackmann et al., 2015; Cunningham and Shah, 2018). Suppose we observe data for a unit that is affected by the treatment or intervention of interest, as well as data on a donor pool, that is, a set of untreated units that are available to approximate the outcome that would have been observed for the treated unit in the absence of the intervention. The idea behind synthetic controls is to match each unit exposed to the intervention or treatment of interest to the weighted average of the units in the donor pool that most closely resembles the characteristics of the treated unit before the intervention. Once a suitable synthetic control is selected, differences in outcomes between the treated unit and the synthetic control are taken as estimates of the effect of the treatment on the unit exposed to the intervention of interest.

The synthetic control method is akin to nearest neighbor matching estimators (Dehejia and Wahba, 2002; Abadie and Imbens, 2006; Imbens and Rubin, 2015) but departs from nearest neighbor matching methods in two important aspects. First, the synthetic control method does not impose a fixed number of matches for every treated unit. Second, instead of using a simple average of the matched units with equal weights, the synthetic control method matches each treated unit to a weighted average of untreated units with weights calculated to minimize the discrepancies between the treated unit and the synthetic control in the values of the matching variables. Synthetic control estimators retain, however, appealing properties of nearest neighbor matching estimators. In particular, like nearest neighbor matching estimators, synthetic control estimators use weights that are non-negative and sum to one. In addition, synthetic control weights are often sparse. That is, like nearest neighbor matching estimators, they only assign positive weights to a relatively small number of untreated units. Sparsity and non-negativity of the weights, along with the fact that synthetic control weights sum to one and define a weighted average, are important features that allow the use of expert knowledge to evaluate and interpret the estimated counterfactuals (see Abadie et al., 2015). As shown in Abadie et al. (2015), similar to the synthetic control estimator, a regression-based estimator of the counterfactual of interest–i.e., the outcome for the treated in the absence of an intervention–implicitly uses a linear combination of outcomes for the untreated with weights that sum to one. However, unlike synthetic control weights, regression weights are not explicit in the outcome in the procedure, they are not sparse, and they can be negative or greater than one, allowing unchecked extrapolation outside the support of the data and complicating the interpretation of the estimate and the nature of the implicit comparison. While many applications of the synthetic control framework have focused on cases where only one aggregate unit is exposed to the intervention of interest, the method has found recent applications in contexts with disaggregated data, where data sets contain multiple treated units. In some cases, especially in cases with a small number of treated units, the interest may lie on the treament effects for each of the treated. In other cases, especially in settings with a large number of treated units, the interest may lie on the average effect of the treatment among the treated (see, e.g., Acemoglu et al., 2016; Gobillon and Magnac, 2016; Kreif et al., 2016). In such settings, one could simply construct a synthetic control for an aggregate of all treated units. However, interpolation biases may be much smaller if the estimator of the aggregate outcome that would have been observed for the treated in the absence of the treatment is based on the aggregation of multiple synthetic controls, one for each treated unit.

Using synthetic controls to estimate treatment effects with disaggregated data creates some practical challenges. In particular, when the values of the matching variables for a treated unit fall in the convex hull of the corresponding values for the donor pool, it may be possible to find multiple convex combinations of untreated units that perfectly reproduce the values of the matching variables for the treated observation. That is, the best synthetic control may not be unique. One practical consequence of the curse of dimensionality is that, even for a moderate number of matching variables, each particular treated unit is unlikely to fall in the convex hull of the untreated units, especially if the number of untreated units is not very large. As a result, lack of uniqueness is rarely an issue in settings with one or a small number of treated units and, if it arises, it can typically be solved by ad-hoc methods, like increasing the number of covariates or by restricting the donor pool to units that are similar to the treated units. In settings with many treated and many untreated units, non-uniqueness may be an important consideration and a problem which is harder to solve.

More generally, in contrast to common aggregate data settings with a small donor pool (see, e.g., Abadie and Gardeazabal, 2003; Abadie et al., 2010), in settings with a large number of units in the donor pool, single untreated units may provide close matches to the treated units in the data. Therefore, in such settings, the researcher faces a trade-off between minimizing the covariate discrepancy between each treated unit and its synthetic control as a whole (synthetic control case) and minimizing the covariate discrepancy between each treated unit and each unit that contributes to its synthetic control (matching case).

This article provides a generalized synthetic control framework for estimation and inference. We introduce a penalization parameter that trades off pairwise matching discrepancies with respect to the characteristics of each unit in the synthetic control against matching discrepancies with respect to the characteristics of the synthetic control unit as a whole. This type of penalization is aimed to reduce interpolation biases by prioritizing inclusion in the synthetic control of units that are close to the treated in the space of the matching variables. Moreover, we show that as long as the penalization parameter is positive, the generalized synthetic control estimator is unique and sparse. If the value of the penalization parameter is close to zero, our procedure selects the synthetic control that minimizes the sum of pairwise matching discrepancies (among the synthetic controls that best reproduce the characteristic of the treated units). If the value of the penalization parameter is large, our estimator coincides with the pair-matching estimator. We study formal properties of the penalized synthetic control estimator and propose data-driven choices of the penalization parameter. We propose, in addition, a bias-corrected version of the penalized synthetic control estimator, which is analogous to the one applied to matching estimators in Rubin (1973) and Abadie and Imbens (2011). We show that the bias-correction substantially improves the properties of penalized synthetic control estimators.

Doudchenko and Imbens (2016), Athey et al. (2017), Amjad et al. (2018), Arkhangelsky et al. (2018) and Chernozhukov et al. (2019) have also proposed penalization schemes for synthetic controls and related methods. Doudchenko and Imbens (2016), Arkhangelsky et al. (2018) and Chernozhukov et al. (2019) use an L_1 penalty term (lasso), an L_2 penalty term (ridge), or a combination of both (elastic net) to regularize synthetic control weights. This is different from our penalization scheme, which depends on the matching discrepancy between the treated unit and the units in the synthetic control. Athey et al. (2017) assume an underlying sparse factor structure for the outcome under no treatment and adapt matrix completion techniques to estimate a counterfactual. Their estimator penalizes the complexity of the factor structure. The estimator in Amjad et al. (2018) uses low-rank approximation techniques to de-noise the outcomes for the units in the donor pool. Then, potential outcomes without the treatment for the treated are estimated as linear combinations of de-noised outcomes for the units in the donor pool, with ridge-regularized coefficients. Bias-corrected synthetic control estimators have been independently studied in Ben-Michael et al. (2019) and Arkhangelsky et al. (2018).

The rest of the article is organized as follows. Section 2 presents the penalized synthetic control estimator and discusses several of its geometric properties. Section 3 discusses permutation inference. Section 4 presents ways to choose the penalization term. Section 5 illustrates the properties of the estimator through simulations. Section 6 contains an application. Section 7 contains a summary of the article and conclusions. The appendix gathers the proofs.

2. Penalized Synthetic Control

2.1. Synthetic Control for Disaggregated Data

Assume we observe n units, some of which are exposed to the treatment or intervention of interest. We code the treatment status of unit i using the binary variable D_i, so $D_i = 1$ if i is treated and $D_i = 0$ otherwise. To define treatment effects we adopt a potential outcomes framework, as in Rubin (1974). Let Y_{1i} and Y_{0i} be random variables representing potential outcomes under treatment and under no treatment, respectively, for unit i. The effect of the treatment for unit i is $Y_{1i} - Y_{0i}$. Realized outcomes are defined as

$$Y_i = \begin{cases} Y_{1i} & \text{if } D_i = 1, \\ Y_{0i} & \text{if } D_i = 0. \end{cases}$$

Let X_i be a $(p \times 1)$-vector of pre-treatment predictors of Y_{0i}. We assume that we observe $(Y_i, X_i) = (Y_{1i}, X_i)$ for n_1 treated observations and $(Y_i, X_i) = (Y_{0i}, X_i)$ for n_0 untreated observations. Combining data for treated and nontreated we obtain the pooled data set, $\{(Y_i, D_i, X_i)\}_{i=1}^{n}$, with $n = n_0 + n_1$. To simplify notation, we reorder the observations in the data so that the n_1 treated observations come first. The quantities of interest are the treatment effects on the treated units, $\tau_i = Y_{1i} - Y_{0i}$ for $i = 1, \dots, n_1$, and/or the average treatment effect on the treated (ATET):

$$\tau = \frac{1}{n_1} \sum_{i=1}^{n_1} \big(Y_{1i} - Y_{0i}\big). \tag{1}$$

Many estimators of τ, are of the form,

$$\frac{1}{n_1} \sum_{i=1}^{n} Y_i D_i - \frac{1}{n_0} \sum_{i=1}^{n} Y_i (1 - D_i) V_i. \tag{2}$$

Popular estimators of this type in micro-econometrics include most notably regression (Angrist and Pischke, 2008; Abadie et al., 2015), matching estimators (Rosenbaum and Rubin, 1983; Dehejia and Wahba, 2002; Abadie and Imbens, 2006), and propensity score weighting estimators (Hirano et al., 2003). For example, in the case of the pair-matching estimator, the weight V_i given to control unit i is equal to the number of times control unit i is the nearest neighbor of a treated unit, rescaled by n_0/n_1. The synthetic control method (Abadie and Gardeazabal, 2003; Abadie et al., 2010, 2015; Doudchenko and Imbens, 2016) also belongs to this class of estimators. It matches each treated unit to a "synthetic control", that is, a weighted average of untreated units with weights chosen to make the values of the predictors of the outcome variable of each synthetic control closely match the values of the same predictors for the corresponding treated units.

For any $(p \times 1)$ real vector X and any $(p \times p)$ real symmetric positive-definite matrix Γ, define the norm $\|X\| = (X'\Gamma X)^{1/2}$. Because Γ is diagonalizable with strictly positive eigenvalues, we can always transform the vector X so that the matrix Γ becomes the $(p \times p)$ identity matrix. As a result, without loss of generality, we will consider only $\Gamma = I$. In the synthetic control framework, model selection–that is, the choice of the variables included in X–is operationalized through the choice Γ, which rescales or weights each predictor in X according to its predictive power on the outcome (see Abadie et al., 2010). In a setting with many treated and untreated units, the standard synthetic control estimation procedure is as follows:

1. For each treated unit, $i = 1, \ldots, n_1$, compute the n_0-vector of weights $W_i^* = (W_{i,n_1+1}^*, \ldots, W_{i,n}^*)$ that solves

$$\min_{W_i \in \mathbb{R}^{n_0}} \left\| X_i - \sum_{j=n_1+1}^{n} W_{i,j} X_j \right\|^2 \tag{3}$$
$$\text{s.t.} \quad W_{i,n_1+1} \geq 0, \ldots, W_{i,n} \geq 0,$$
$$\sum_{j=n_1+1}^{n} W_{i,j} = 1,$$

   where $W_{i,j}^*$ is the weight given to control unit j in the synthetic control unit corresponding to treated unit i. A synthetic control estimate of the effect of the treatment on treated unit i is

$$\widehat{\tau}_i = Y_i - \sum_{j=n_1+1}^{n} W_{i,j}^* Y_j.$$

2. Averaging the treatment effects on the treated produces a synthetic control estimate of τ,

$$\widehat{\tau} = \frac{1}{n_1} \sum_{i=1}^{n_1} \left[Y_i - \sum_{j=n_1+1}^{n} W_{i,j}^* Y_j \right]. \tag{4}$$

Notice that $\widehat{\tau}$ is the estimator in equation (2) reweighting each nontreated unit, $j = n_1 + 1, \ldots, n$, by $V_j = (n_0/n_1) \sum_{i=1}^{n_1} W_{i,j}^*$, with $W_{i,j}^* = 0$ for $i \geq n_1 + 1$.

To simplify the exposition, so far we have described a simple cross-sectional setting. The extension to the more common panel data setting for synthetic controls, where the same units are observed for a number of periods—before and after the intervention happens for the treated—is immediate and we will use it in later sections. Panel data settings with multiple treated units also raise the possibility that different treated units adopt the treatment at different points in time. Staggered adoption of a treatment (Athey and Imbens, 2018; Ben-Michael et al., 2019) can easily be accommodated in the synthetic control framework, although it creates some implementation challenges related to the choice of a meaningful average of the individual treatment effects as a target parameter, and the fact that the donor pool changes in time. Moreover, even in cross-sectional settings or when all treated units adopt the treatment at the same time, τ in equation (1) is by no means the only possible target parameter of interest. For instance, if the data consist of a number of cities or states, one may wish to calculate a population weighted treatment effect. This is, again, easy to implement in a synthetic control framework like the one in this article, where the effect of the treatment is estimated separately for each treated unit. Abadie (2020) provides an introduction to synthetic control estimation and discusses feasibility and data requirements.

2.2. Penalized Synthetic Control

The main contribution of this article is to propose a penalized version of the synthetic control estimator in equation (3). For treated unit i and given a positive penalization constant λ, the penalized synthetic control weights, $W^*_{i,j}(\lambda)$, solve

$$\begin{aligned} \min_{W_i \in \mathbb{R}^{n_0}} \quad & \left\| X_i - \sum_{j=n_1+1}^{n} W_{i,j} X_j \right\|^2 + \lambda \sum_{j=n_1+1}^{n} W_{i,j} \|X_i - X_j\|^2 \\ \text{s.t.} \quad & W_{i,n_1+1} \geq 0, \ldots, W_{i,n} \geq 0, \\ & \sum_{j=n_1+1}^{n} W_{i,j} = 1. \end{aligned} \tag{5}$$

The penalized synthetic control estimates are

$$\widehat{\tau}_i = Y_i - \sum_{j=n_1+1}^{n} W^*_{i,j}(\lambda) Y_j.$$

for the unit-level treatment effects, τ_i, and

$$\widehat{\tau}(\lambda) = \frac{1}{n_1}\sum_{i=1}^{n_1}\left[Y_i - \sum_{j=n_1+1}^{n} W^*_{i,j}(\lambda)Y_j\right] \tag{6}$$

for the average effect on the treated, τ.

The tuning parameter λ sets the trade-off between componentwise fit and aggregate fit. The choice of the value of λ is important and will be discussed in Section 4. The penalized synthetic control estimator encompasses both the synthetic control estimator and the nearest-neighbor matching as special polar cases. At one end of the spectrum, as $\lambda \to 0$, the penalized estimator becomes the synthetic control that minimizes the sum of pairwise matching discrepancies among the set of synthetic controls that best reproduce the characteristics of the treated units. Our motivation to choose among synthetic controls that fit the treated unit equally well by minimizing the sum of pairwise matching discrepancies is to reduce worst-case interpolation biases. At the other end of the spectrum, as $\lambda \to \infty$, the penalized estimator becomes the one-match nearest-neighbor matching with replacement estimator in Abadie and Imbens (2006).

Let X_0 be the $(p\times n_0)$ matrix with column j equal to X_{n_1+j}, and let Δ_i be the $(n_0\times 1)$ vector with j-th element equal to $\|X_i - X_{n_1+j}\|^2$. Moreover, let $\Delta_i^{NN} = \min_{j=1,\dots,n_0}\|X_i - X_{n_1+j}\|^2$ be the smallest discrepancy between unit i and the units in the donor pool. Finally, let $W_i^*(\lambda)$ be a solution to (5), and $\Delta_i^*(\lambda) = \|X_i - X_0 W_i^*(\lambda)\|^2$ be the square of the discrepancy between unit i and the (penalized) synthetic control. The next lemma establishes bounds on $\Delta_i^*(\lambda)$ and $\Delta_i' W_i^*(\lambda)$.

Lemma 1 (Discrepancy Bounds) *For any* $\lambda \geq 0$

$$0 \leq \Delta_i^*(\lambda) \leq \Delta_i^{NN},$$

and for $\lambda > 0$

$$\Delta_i^{NN} \leq \Delta_i' W_i^*(\lambda) \leq \frac{1+\lambda}{\lambda}\Delta_i^{NN}.$$

All proofs are in the appendix.

The first result in Lemma 1 states that the synthetic unit is contained in a closed ball of center X_i and radius equal to the distance to the nearest-neighbor, $\sqrt{\Delta_i^{NN}}$. The second result implies that the tuning parameter λ controls the compound discrepancy between the treated unit and the units that contribute to the synthetic control, $\Delta_i' W_i^*(\lambda)$.

The specific penalty term in equation (5) is one of many possible alternatives. For instance, in the spirit of elastic nets, one could add an L_2 penalty term, $\gamma(W_{i,n_1+1}^2 + \cdots + W_{i,n}^2)$, to the objective function in equation (5). The L_1 penalty term in equation (5) has the advantage of producing easy-to-interpret sparse solutions, which is also a feature of the standard synthetic control estimator. In the absence of the penalty term (that is, when $\lambda = 0$), the problem in (5) can be solved by projecting X_i on the convex hull of X_0. Existence of sparse solutions follows from Carathéodory's theorem. However, if $\lambda = 0$ the solution to the problem in (5) may not be unique if X_i belongs to the convex hull of the columns of X_0. Adopting $\lambda > 0$ penalizes solutions with potentially large interpolation biases created by large matching discrepancies and produces uniqueness and sparsity as stated in the following result.

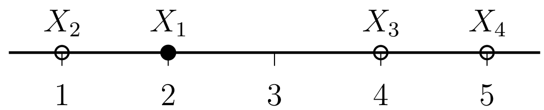
Theorem 1 (Uniqueness and Sparsity) *Suppose that any submatrix composed by rows of $[X_0' \; 1_{n_0} \; \Delta_i]$ has full rank, where 1_{n_0} is the $(n_0 \times 1)$ vector of ones. Then, if $\lambda > 0$ the optimization problem in equation (5) admits a unique solution $W_i^*(\lambda)$ with at most $p + 1$ non-zero components.*

Notice that the condition that any submatrix composed by rows of $[X_0' \; 1_{n_0} \; \Delta_i]$ has full rank implies that there are no two control units with the same values of the predictors. It also implies that there is no set of control units of cardinality $p + 2$ or larger such that the values of the predictors belong to a sphere with center at X_i.

Example: Consider a simple numerical example with only one covariate. Suppose, there is one treated unit with $X_1 = 2$ and three control units with $X_2 = 1$, $X_3 = 4$ and $X_4 = 5$. This simple setting is depicted in Figure 1.

Notice that X_1 belongs to $[1, 5]$, the convex hull of the columns of X_0, and $\Delta_1 = (1,\, 4,\, 9)'$. Consider first the case with $\lambda = 0$. Then, $W^*(0) = (2/3,\, 1/3,\, 0)'$ and $W^{**}(0) = (3/4,\, 0,\, 1/4)'$

Figure 1: A simple example



are the only two sparse solutions (with number of non-zero weights not greater than $p+1 = 2$) to (5). The first sparse solution, $W^*(0)$, interpolates $X_1 = 2$ using $X_2 = 1$ and $X_3 = 4$. The second sparse solution, $W^{**}(0)$ is of lower quality relative to $W^*(0)$ in terms of compound discrepancy, as it uses an interpolation scheme that replaces X_3 with X_4, an observation farther away from X_1. As a result, $W^*(0)$ is preferred over $W^{**}(0)$ in terms of worst case interpolation bias. However, the better compound fit of $W^*(0)$ is not reflected in a better value in the objective function in (3). Moreover, because any convex combination of $W^*(0)$ and $W^{**}(0)$ is also a solution, the problem in (3) has an infinite number of solutions, $\mathcal{W}_0^* = \{aW^*(0) + (1-a)W^{**}(0) : a \in [0,1]\}$. Let $\bar{V}(a) = aW^*(0) + (1-a)W^{**}(0)$. The compound discrepancy of $\bar{V}(a)$ is

$$\Delta_i'\bar{V}(a) = 3 - a.$$

$W^*(0)$, which is obtained making $a = 1$, produces the lowest compound discrepancy among all the solutions to equation (3).

When $\lambda > 0$, however, the program (5) has a unique solution, which is sparse:

$$W^*(\lambda) = \begin{cases} (2+\lambda/2,\ 1-\lambda/2,\ 0)'/3 & \text{if } 0 < \lambda \leq 2, \\ (1,\ 0,\ 0)' & \text{if } \lambda > 2. \end{cases}$$

Notice that $W^*(\lambda)$ never puts any weight on X_4. As $\lambda \to \infty$, $W^*(\lambda)$ selects the nearest-neighbor match, and as $\lambda \to 0$, $W^*(\lambda)$ converges to $W^*(0)$, the (non-penalized) synthetic control in $\mathcal{W}_0^*$ with the smallest compound discrepancy. □

2.3. Geometric Properties of Penalized Synthetic Controls

In this section, we use Delaunay tessellations to characterize the geometric properties of penalized synthetic control estimators. The results in this section provide a geometric interpretation of penalized synthetic controls, and imply that there exists an estimator for the case $\lambda \to 0$ that does not depend on approximating the limit estimate with the one obtained for an arbitrary small value of λ.

A Delaunay triangulation tessellates the convex hull of a set of points, $\{x_1, \ldots, x_n\}$, in $\mathbb{R}^2$ into triangles with vertices in $\{x_1, \ldots, x_n\}$. Each triangle of the Delaunay triangulation of $\{x_1, \ldots, x_n\}$ is such that its circumscribing circle does not contain any point in $\{x_1, \ldots, x_n\}$ in its interior. Delaunay triangulations generalize to higher dimensions, in which case they are often referred to as Delaunay tessellations. A Delaunay tesselation in $\mathbb{R}^3$ is a collection of tetrahedrons with vertices in $\{x_1, \ldots, x_n\}$ such that their circumscribing spheres do not contain points in $\{x_1, \ldots, x_n\}$ in their interiors. More generally, a Delaunay tesselation in $\mathbb{R}^p$ is a collection of p-simplices with vertices in $\{x_1, \ldots, x_n\}$ such that their circumscribing hyperspheres do not contain points of $\{x_1, \ldots, x_n\}$ in their interiors (see, e.g., Boissonnat and Yvinec, 1998; Okabe et al., 2000). We will refer to the simplices of a Delaunay tessellation as Delaunay simplices. The set $\{x_1, \ldots, x_n\}$, along with the collection of segments connecting the vertices of each p-simplex of a Delaunay tessellation, constitutes the Delaunay graph induced by the tessellation. For the remainder of this section, we will assume that every Delaunay tessellation is done on the convex hull of a set of points in general quadratic position. We say that n points in $\mathbb{R}^p$ are in general quadratic position when *(i)* for $k = 2, \ldots, p$, no $k+1$ points lie in a $(k-1)$-dimensional hyperplane of $\mathbb{R}^p$ *(non-collinearity)*, and *(ii)* no $p+2$ points lie on the boundary of an hypersphere in $\mathbb{R}^p$ *(non-cosphericity)* (see, e.g., Okabe et al., 2000). If all the points in the set $\{x_1, \ldots, x_n\}$ are in general quadratic position, then the Delaunay tessellation of the convex hull of $\{x_1, \ldots, x_n\}$ exists and is unique. The assumption of general quadratic position is fairly innocuous. Realizations of random vectors drawn form a distribution that is continuous with respect to the Lebesgue measure are in general quadratic position with probability one.

The next theorem provides a characterization of the units contributing to a particular synthetic control, $X_0 W_i^*(\lambda)$ with $\lambda > 0$, as vertices of the Delaunay simplex containing $X_0 W_i^*(\lambda)$ in the Delaunay tessellation of $X_{n_1+1}, \ldots, X_n$.

Theorem 2 (Delaunay Property I) *Let $W_i^*(\lambda)$ be a solution to the penalized synthetic* *control problem in* (5) *with* $\lambda > 0$*. Consider the Delaunay tessellation induced by the columns of* X_0*. Then, for any control unit* $j = n_1 + 1, \ldots, n$*, such that* X_j *is not a vertex of the Delaunay simplex containing* $X_0 W_i^*(\lambda)$*, it holds that* $W_{i,j}^*(\lambda) = 0$*.*

This result provides a notion of proximity between a synthetic control and each of the units that contribute to it. Theorem 2 also provides a simple way to compute the solution for the "pure synthetic control" case ($\lambda \to 0$) that does not entail the choice of an arbitrarily small value of λ to use in (5). Recall that when $\lambda = 0$, the problem of minimizing $\|X_i - X_0 W\|$ subject to the weight constraints has typically infinite solutions if X_1 belongs to the convex hull of the columns of X_0, in which case $X_i = X_0 W$ for all solutions. In the presence of multiple solutions, the pure synthetic control case selects the solution that produces the lowest compound discrepancy, $W'\Delta_i$, among all W such that $X_i = X_0 W$. Directly solving (5) for $\lambda \to 0$ requires, in practice, a choice of a small value for λ. It also creates computational difficulties, as the minimization problem is close to one with multiple solutions. Theorem 2 provides a solution to these problems, because it implies that the solution of (5) for $\lambda \to 0$ can assign positive weights only to the vertices of the simplex in the Delaunay tessellation of $X_{n_1+1}, \ldots, X_n$ that contains the projection of X_i on the convex hull of the columns of X_0. As a result, it is enough to solve (5) allowing positive weights only on the observations that represent the vertices of the Delaunay face that contains the projection of X_i on the convex hull of the columns of X_0. In high-dimensional settings, however, the large computation costs of Delaunay triangulations may make this approach unfeasible.

Consider the Delaunay graph induced by the Delaunay tessellation of the convex hull of a set of points representing the predictor values for a treated unit, i, and for all the units in the donor pool. The next theorem shows that, in such a graph, the treated unit is connected to all the untreated units that contribute to the synthetic control of unit i.

Theorem 3 (Delaunay Property II) *Let* $W_i^*(\lambda)$ *be a solution to the penalized synthetic control problem in* (5) *with* $\lambda > 0$*. Consider the Delaunay tessellation induced by the columns of* X_0 *and the treated* X_i*, and denote* $\mathcal{I}_i$ *as the indices of the points in* $\{X_{n_1+1}, \ldots, X_n\}$ *that are connected to* X_i *in the corresponding Delaunay graph. For any* $j \notin \mathcal{I}_i$*, it holds that*

$W^*_{i,j}(\lambda) = 0$.

Theorem 3 provides a notion of proximity between a treated unit and the units contributing to its penalized synthetic control. It therefore restricts the donor pool to these units connected to the treated and as such provides a way to simplify the computation of the synthetic control for $\lambda > 0$.

Figure 2 illustrates Lemma 1 and Theorems 2 and 3 in two dimensions. The top-left panel (a) displays the treated unit (black cross) and the Delaunay triangulation of untreated units. The top-right panel (b) draws the trajectory of the synthetic unit as λ changes (as λ increases, the solution drifts toward the nearest neighbor and away from the treated – solid black line) and the circle centered on the treated of radius equal to the distance between the treated and its nearest neighbor. Notice that the synthetic unit is never located outside of this circle, as per Lemma 1. The bottom left panel (c) shows the four untreated units that have a non-zero weight across some solutions of the penalized synthetic control as λ changes (black dots). They are the vertices of the two triangles where the synthetic unit is located, as per Theorem 2. The bottom-right panel (d) shows that these units are also connected to the treated in the augmented Delaunay triangulation (that includes the treated unit), as per Theorem 3.

Notice that being connected to the treated unit in the augmented Delaunay triangulation is a necessary but not sufficient condition for a unit to contribute to the synthetic control. This can easily be seen in Figure 2 (d). For example, the nearest neighbor to the treated unit is connected to the treated unit in the augmented Delaunay graph. This is, in fact, a general property: the (undirected) nearest neighbor graph is always a subgraph of the Delaunay graph. However, there are positive values of λ for which the penalized synthetic control estimator puts zero weight on the nearest neighbor of the treated unit. This is implied by the fact that the penalized synthetic control does not lie on a Delaunay simplex with the nearest neighbor as one of the vertices and Theorem 2. Although the objective function in equation (6) is a combination of the the objective functions minimized by the unpenalized synthetic control and the nearest neighbor matching estimator, the penalized synthetic controls estimator is not in general a combination of the unpenalized synthetic control estimator and the nearest neighbor matching estimator.

2.4. Bias-Corrected Synthetic Control

We will also consider bias-corrected versions of synthetic control estimators. We adopt a bias correction analogous to that implemented in Rubin (1973) and Abadie and Imbens (2011) for matching estimators. Let $\widehat{\mu}_0(x)$ be a regression predictor of the outcome, Y_i, of an untreated unit with covariate values $X_i = x$. A bias-corrected version of the synthetic control estimator in equation (6) is

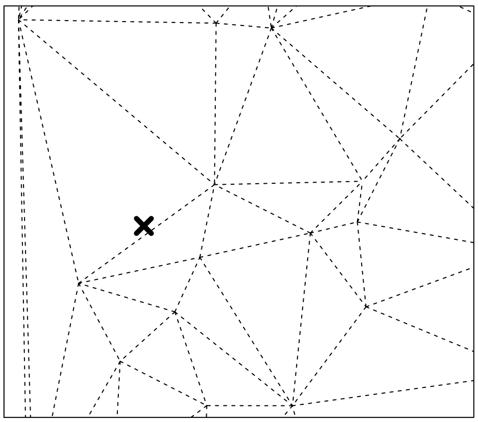
$$\widehat{\tau}_{BC}(\lambda) = \frac{1}{n_1} \sum_{i=1}^{n_1} \left[\big(Y_i - \widehat{\mu}_0(X_i)\big) - \sum_{j=n_1+1}^{n} W^*_{i,j}(\lambda) \big(Y_j - \widehat{\mu}_0(X_j)\big) \right]. \tag{7}$$

Like in Abadie and Imbens (2011), the bias correction in equation (7) adjusts for mismatches between the characteristics of the treated units and the characteristics of each of the units that contribute to the synthetic controls. Depending on the setting and the nature and quantity of data, $\widehat{\mu}_0$ can be a parametric or a non-parametric regression. A bias correction of this type has been independently studied in Ben-Michael et al. (2019), who propose using ridge regression to estimate $\widehat{\mu}_0(x)$.

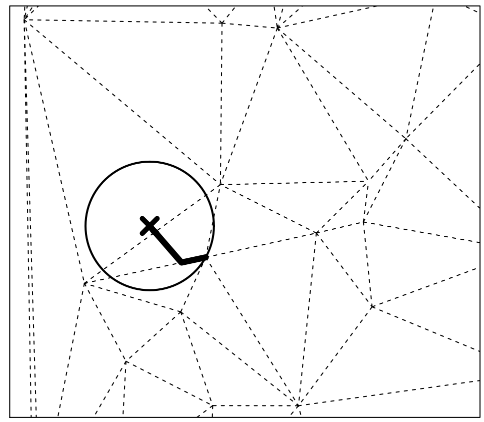
3. Permutation Inference

In this section, we adapt the inferential framework in Abadie et al. (2010) to the penalized synthetic control estimators of Section 2. Like in Abadie et al. (2010), our inferential exercises compare the value of a test statistic to its permutation distribution induced by random reassignment of the treatment variable in the data set. Aside from simulation errors, this inferential exercise is exact by construction, regardless of the number of units in the data. We next describe two possible implementations that employ different test statistics and permutation schemes. Alternative test statistics and permutation schemes are possible and, in practice, the choice among them should take into account the nature of the parameter(s) of interest (e.g., individual vs. aggregate effects), the characteristics of the intervention that

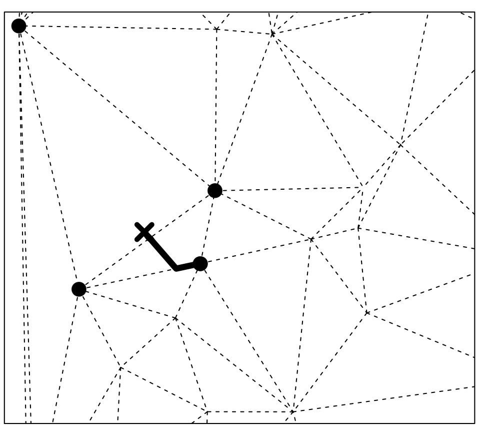
Figure 2: Geometric properties of penalized synthetic control estimator



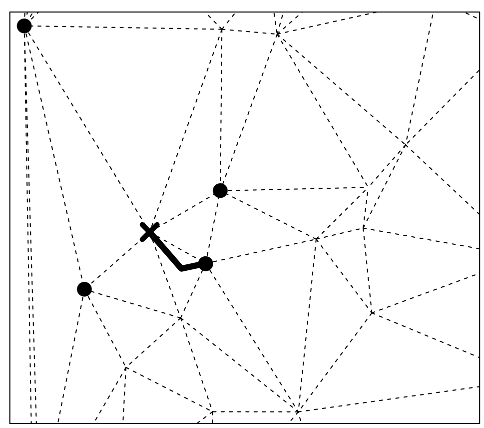
(a) Treated unit (black cross) on the Delaunay triangulation of untreated units (dashed lines).



(b) The synthetic unit as λ changes (solid black line) and the circle centered on the treated of radius equal to the distance between the treated and its nearest neighbor.



(c) The four untreated (black dots) that have a non-zero weight in some solutions of the penalized synthetic control as λ changes.



(d) Treated unit and synthetic units on the Delaunay triangulation augmented with the treated unit.

is the object of the analysis, and the structure of the data set. Randomized reassignment of the treatment in the data is taken here as a benchmark against which we evaluate the rareness of the value of a test statistic, and it may not reflect the actual and typically unknown treatment assignment process (see Abadie et al., 2010, 2015). Firpo and Possebom (2018) propose a procedure to assess the sensitivity of permutation inference to deviations from the reassignment benchmark. The permutation procedure outlined in this section is conditional on the data and its validity does not depend on the nature of the mechanism used to generate the data set. Alternative inferential procedures for synthetic controls have been proposed by Chernozhukov et al. (2019) and Cattaneo et al. (2019), among others, and they are summarized in Abadie (2020). While this section focuses exclusively on p-values, permutation distributions are easy to visualize and report, and they contain important additional information, like the signs and magnitudes of the test statistics (Abadie, 2020). In addition, as in Firpo and Possebom (2018), confidence intervals around synthetic control estimates can be obtained by inverting the results on statistical tests based on the p-values in this section.

3.1. Inference on Aggregate Effects

Here we outline a simple permutation procedure that employs a test statistic, $\widehat{T}$, that measures aggregate effects for the treated. Examples of aggregate statistics of this type are the synthetic controls estimators in equations (6) and (7). Similar to Abadie et al. (2010), in a panel data setting $\widehat{T}$ can be based on the ratio between the aggregate mean square prediction error in a post-intervention period and a pre-intervention period. Let Y_{it} be the observed outcome for unit i at time t, and let $\widehat{\tau}_{it}$ be as in equations (6) and (7) but with Y_i and Y_j replaced by Y_{it} and Y_{jt}, respectively. Then, the ratio between the aggregate mean square prediction error in a post-intervention period $\mathcal{T}_1 \subseteq \{T_0+1,\ldots,T\}$ and a pre-intervention period $\mathcal{T}_0 \subseteq \{1,\ldots,T_0\}$ is

$$\sum_{t\in\mathcal{T}_1}\left(\sum_{i=1}^{n_1}\widehat{\tau}_{it}(\lambda)\right)^2 \Big/ \sum_{t\in\mathcal{T}_0}\left(\sum_{i=1}^{n_1}\widehat{\tau}_{it}(\lambda)\right)^2. \tag{8}$$

Let $\boldsymbol{D}^{obs} = (D_1, ..., D_n)$ be the observed treatment assignments in the data. We will write $\widehat{T}(\boldsymbol{D}^{obs})$ to indicate the value of the test statistic in the data, and $\widehat{T}(\boldsymbol{D})$ to indicate the value of the test statistic when the treatment values are reassigned in the data as indicated in $\boldsymbol{D}$. The procedure is as follows:

1. Compute the test statistic in the original data $\widehat{T}(\boldsymbol{D}^{obs})$.

2. At each iteration, $b = 1, \ldots, B$, permute at random the components of $\boldsymbol{D}^{obs}$ to obtain $\widehat{T}(\boldsymbol{D}^{(b)})$.

3. Calculate p-values as the frequencies across iterations of the values of $\widehat{T}(\boldsymbol{D}^{(b)})$ more extreme than $\widehat{T}(\boldsymbol{D}^{obs})$. Typically, for two-sided tests:

$$\widehat{p} = \frac{1}{B+1}\left(1 + \sum_{b=1}^{B} \mathbf{1}\left\{|\widehat{T}(\boldsymbol{D}^{(b)})| \geq |\widehat{T}(\boldsymbol{D}^{obs})|\right\}\right).$$

   For one sided tests:

$$\widehat{p} = \frac{1}{B+1}\left(1 + \sum_{b=1}^{B} \mathbf{1}\left\{\widehat{T}(\boldsymbol{D}^{(b)}) \geq \widehat{T}(\boldsymbol{D}^{obs})\right\}\right),$$

   or

$$\widehat{p} = \frac{1}{B+1}\left(1 + \sum_{b=1}^{B} \mathbf{1}\left\{\widehat{T}(\boldsymbol{D}^{(b)}) \leq \widehat{T}(\boldsymbol{D}^{obs})\right\}\right).$$

3.2. Inference Based on the Sum of Rank Statistics of Unit-Level Treatment Effects Estimates

Similar to Dube and Zipperer (2015), we propose a test based on the rank statistics of the unit-level treatment effects. Unlike the test in Dube and Zipperer (2015), we calculate the permutation distribution directly from the data. The test we employ is based on the sum of the ranks of $n_1 \times (B+1)$ unit-level test statistics for the treated in the data and in B random permutation of observed treatments. Individual treatment effects, $\widehat{T}_i$, may be based on differences in outcomes between treated and synthetic controls,

$$Y_i - \sum_{j=n_1+1}^{n} W^*_{i,j}(\lambda) Y_j,$$

bias corrected versions of the unit-level treatment effects,

$$(Y_i - \widehat{\mu}_0(X_i)) - \left(\sum_{j=n_1+1}^{n} W^*_{i,j}(\lambda)Y_j - \widehat{\mu}_0(X_j)\right).$$

The test statistic may be based on unit-level versions of the mean squared prediction error ratio in equation (8). The procedure is implemented as follows:

1. Compute unit-level test statistic for the treated, $\widehat{T}_i$, for $i = 1, \ldots, n_1$, under the actual treatment assignment, $\boldsymbol{D}^{obs}$.

2. At each iteration $b = 1, \ldots, B$, permute at random the components of $\boldsymbol{D}^{obs}$ to obtain $\widehat{T}_i(\boldsymbol{D}^{(b)})$ for the treated. Denote these estimates $\widehat{T}_1^{(b)}, \ldots, \widehat{T}_{n_1}^{(b)}$ (in arbitrary order).

3. Calculate the ranks $R_1, \ldots, R_{n_1}$, $R_1^{(1)}, \ldots, R_{n_1}^{(1)}$, …, $R_1^{(B)}, \ldots, R_{n_1}^{(B)}$ associated to the $n_1 \times (B+1)$ individual treatment effect estimates $\widehat{T}_1, \ldots, \widehat{T}_{n_1}, \widehat{T}_1^{(1)}, \ldots, \widehat{T}_{n_1}^{(1)}$, …, $\widehat{T}_1^{(B)}, \ldots,$ $\widehat{T}_{n_1}^{(B)}$ (or of their absolute values or negative values) and the sums of ranks for each permutation, $SR = \sum_{i=1}^{n_1} R_i$, $SR^{(b)} = \sum_{i=1}^{n_1} R_i^{(b)}$, $b = 1, \ldots, B$.

4. Calculate p-values as:

$$\widehat{p} = \frac{1}{B+1}\left(1 + \sum_{b=1}^{B} \mathbf{1}\left\{SR^{(b)} \geq SR\right\}\right).$$

4. Penalty Choice

We present two data-driven selectors for the penalty term, λ. In the context of treatment effects estimation, cross-validation is complicated by the absence of data on a ground truth (that is, on the values of Y_0 for the treated units in the post-intervention periods, see Athey and Imbens, 2016). Since synthetic controls are often applied to panel data, we consider a balanced panel data setting with T periods and $T_0 < T$ pre-intervention periods. As before, we define Y_{it} as the outcome for unit i at time t. Adaptation of (5) and (6) to the panel data setting is straightforward by allowing X_i to potentially include multiple pre-intervention values of the outcome variable and of other predictors of post-intervention outcomes.

The first selector proposed in this section is based on cross-validation on the outcomes on the untreated units in the post-intervention period. The second selector uses a strategy similar to the model selection procedure in Abadie et al. (2015), minimizing mean squared prediction error (MSPE) in a hold-out pre-intervention period.

4.1. Leave-One-Out Cross-Validation of Post-Intervention Outcomes for the Untreated

This section discusses a leave-one-out cross-validation procedure to select λ by minimizing mean squared prediction error for the untreated units in the post-intervention period. The procedure is as follows:

1. For each control unit $i = n_1 + 1, ..., n$, and each post-intervention period, $t = T_0 + 1, \ldots, T$, calculate

$$\widehat{\tau}_{it}(\lambda) = Y_{it} - \sum_{\substack{j=n_1+1 \\ j\neq i}}^{n} W^*_{i,j}(\lambda) Y_{jt},$$

   where $W^*_{i,j}(\lambda)$ is a synthetic control for unit i that is produced by the donor pool $\{n_1 + 1, \ldots, n\}\backslash\{i\}$.

2. Choose λ to minimize some measure of loss, such as the sum of the squared prediction errors for the individual outcomes,

$$\sum_{i=n_1+1}^{n} \sum_{t=T_0+1}^{T} \left(\widehat{\tau}_{it}(\lambda)\right)^2,$$

   or for the average outcomes

$$\sum_{t=T_0+1}^{T} \left(\sum_{i=n_1+1}^{n} \widehat{\tau}_{it}(\lambda)\right)^2.$$

4.2. Pre-Intervention Holdout Validation on the Outcomes of the Treated

An alternative selector of λ is based on validation over the outcomes for the treated on a hold out pre-intervention period. This is similar in spirit to the model selection procedure in Abadie et al. (2015). To simplify the exposition, and because it may be the most natural choice, we will only describe the case when the validation period comes immediately before the intervention, although other choices are possible. Let h and k be the lengths of the training and validation periods, respectively. The validation period comprises the k dates immediately before the intervention, and the training period comprises the h dates immediately before the validation period. The procedure is as follows:

1. For each treated individual, i, and validation period, $t \in \{T_0 - k + 1, \ldots, T_0\}$, compute

$$\widehat{\tau}_{it}(\lambda) = Y_{it} - \sum_{j=n_1+1}^{n} W^*_{i,j}(\lambda) Y_{jt},$$

   where $W^*_{i,j}$ solve (5) with $X_1, \ldots, X_n$ measured in the training period.

2. Choose λ to minimize a measure of error, such as the sum of the squared prediction for the individual outcomes,

$$\sum_{i=1}^{n_1} \sum_{t=T_0-k+1}^{T_0} \left(\widehat{\tau}_{it}(\lambda)\right)^2,$$

   or the squared prediction error of the aggregate outcomes,

$$\sum_{t=T_0-k+1}^{T_0} \left(\sum_{i=1}^{n_1} \widehat{\tau}_{it}(\lambda)\right)^2,$$

   in the validation period.

5. Simulations

This section reports the results of a Monte Carlo experiment that investigates the properties of the penalized synthetic control estimator relative to its unpenalized version ($\lambda = 0$) and to the nearest-neighbor matching estimator in a panel data framework.

The data generating process is as follows. Let X_{mi} be the m-th component of X_i. The simulation design includes two periods: a pre-intervention period ($t = 1$), and a post-intervention period ($t = 2$). Irrespective of the treatment status, the outcome at date $t \in \{1, 2\}$ is generated by $Y_{it} = \left(\sum_{m=1}^{p} X^r_{mi}\right)/\beta + \varepsilon_{it}$ with r a positive constant governing the degree of linearity of the outcome function. Hence, the treatment effects, τ_i, are equal to zero. The error terms ε_{it} are generated as independent values from the standard normal distribution. For the n_1 treated units, X_i, is a vector of dimension p with independent entries uniformly distributed on $[.1, .9]$. For the n_0 control units, X_i is a vector of the same dimension with independent entries distributed as $\sqrt{U}$, where U is uniform on $[0, 1]$. We set $\beta = \sqrt{\text{var}\left(\sum_{m=1}^{p} X_{mi}^r | D_i = 1\right)}$, so that $\text{var}(Y_{i,t}|D_i = 1) = 2$ and the signal-to-noise ratio for the treated is equal to one.

We compare the performances of synthetic control and matching estimators with M matches per unit (see, e.g. Abadie and Imbens, 2006). We will consider these two estimators with a fixed choice and a data-driven choice of λ and M. Under the fixed procedure, we impose $\lambda \to 0$ for the synthetic control and $M = 1$ in the matching estimator, encompassing both polar cases of the penalized synthetic control estimator highlighted in this paper. The case $\lambda \to 0$ is referred to as the "pure synthetic control". Among all the solutions to the unpenalized synthetic control optimization problem in equation (3), it selects the one with the smallest componentwise matching discrepancy, $\sum_{j=n_1+1}^{n} W_{i,j} \|X_i - X_j\|^2$. The computation of the pure synthetic control estimator is based on the result in Theorem 2 and discussion thereafter. The pure synthetic control estimator is not to be confused with the non-penalized synthetic control ($\lambda = 0$), for which we also report results, and which does not take into account the compound discrepancy. The data-driven choice of λ and M uses the first period outcome to minimize the mean square error (MSE) over that period. In other words, we follow the second procedure in Section 4. At each simulation step, λ and M are chosen so as to minimize mean square error,

$$\frac{1}{n_1} \sum_{i=1}^{n_1} \left(Y_{i1} - \sum_{j=n_1+1}^{n} W_{i,j}^*(\lambda) Y_{j1} \right)^2,$$

and

$$\frac{1}{n_1} \sum_{i=1}^{n_1} \left(Y_{i1} - \frac{1}{M} \sum_{j \in \mathcal{J}_M(i)} Y_{j1} \right)^2,$$

with respect to λ and M, respectively, where $\mathcal{J}_M(i)$ is the set of indices of the M control units that are the nearest to treated unit i as measured by the Euclidean norm. The parameter

λ is selected over a grid of positive values. This implies that the penalized synthetic control estimator, which is sparse, does not nest the unpenalized one, which is not necessarily sparse if the treated unit falls inside the convex hull defined by the values of the predictors in the donor pool. The number of matches is selected over the set of positive integers not greater than 20. We also report a bias-corrected version of the estimators as in Section 2.4, based on a linear specification for the regression function, $\widehat{\mu}_0$.

For each configuration and each estimator, we report four statistics computed on the estimates of the treatment effects on the treated units in the second period. The first statistic is the individual root mean square error (RMSE indiv.), computed as the square root of average individual-level MSE across simulations,

$$\left(\frac{1}{B}\sum_{b=1}^{B}\frac{1}{n_1}\sum_{i=1}^{n_1}\left(\widehat{\tau}_{i2}^{(b)}\right)^2\right)^{1/2}.$$

The second is the aggregate-level RMSE (RMSE aggreg.) across simulations,

$$\left(\frac{1}{B}\sum_{b=1}^{B}\left(\frac{1}{n_1}\sum_{i=1}^{n_1}\widehat{\tau}_{i2}^{(b)}\right)^2\right)^{1/2}.$$

The third is the absolute value of the bias across simulations (|Bias|),

$$\left|\frac{1}{B}\sum_{b=1}^{B}\frac{1}{n_1}\sum_{i=1}^{n_1}\widehat{\tau}_{i2}^{(b)}\right|.$$

The last is the average density defined as the average number of untreated units used as controls for each treated unit, i.e., number of non-zero entries of $W_i^*(\lambda)$ or number of matches in the optimized matching procedure.

The results are reported in Tables 1, 2 and 3 for $n_0 \in \{20, 40, 100\}$ respectively, each time with $n_1 = 10$. Table 4 reports results for $n_1 = 100$, $n_0 = 400$. Each table is divided into sixteen blocks corresponding to a particular value of (p, r). Each block is divided into two parts: the upper half reports the results without bias-correction and the lower half reports results with a bias-correction based on a linear specification of the regression function. Results are color-coded column-by-column within each half-block on a continuous color scale. For the upper half-block, the scale varies from dark blue (minimum column value) to light yellow (maximum column value). For the lower half-block, the scale varies from bright red (minimum column value) to light yellow (maximum column value).

Some clear patterns emerge from the results in Tables 1-3. First, for most parameter values the penalized and the pure synthetic control estimators outperform the matching procedures across all three measures of performance. This advantage appears to be increasing with p, the dimension of the covariates. Second, in terms of aggregate RMSE and bias, the unpenalized synthetic control estimator shows mixed results, especially when p is small and r is large, but catches up with the pure synthetic control estimator as p increases, which is expected. Indeed, the pure and unpenalized synthetic control estimators coincide for treated units outside of the convex hull of the untreated. As the dimensionality of the matching variables increases, the probability that a treated unit falls outside the convex hull of the untreated increases. In terms of individual RMSE, the unpenalized synthetic control estimator behaves very well at the cost of large reductions in sparsity and, therefore, at the cost of interpretability of the individual estimates. Third, the advantage of the penalized and pure synthetic control estimators with respect to the bias slightly decreases as the degree of the outcome function r increases. When r is relatively large, the matching procedure displays a low bias as expected, albeit at the expense of a very large individual RMSE. These three observations are magnified in Table 4 where the penalized synthetic control performs consistently well in each of the sixteen blocks. The biases of the estimators go down substantially when we adopt the bias-correction procedure of Section 2.4. Here, it is more difficult to rank estimators based on the simulation, as the amount of bias corrected by the procedure is different for each estimator in a way that may be directly linked to the simulation design. That said, the overall patterns of relative performance of the bias-corrected estimators is similar to that of the the non-corrected estimators, albeit with more muted differences in performance. Overall, the penalized synthetic control estimator strikes a favorable bias-variance trade-off in Tables 1-4 by combining the strength of matching and (unpenalized) synthetic control.

Table 1: Monte-Carlo Simulations, $n_1 = 10$, $n_0 = 20$

	$r = 1$				$r = 1.2$				$r = 1.4$				$r = 2$			
	RMSE indiv.	RMSE aggreg.	|Bias|	Density	RMSE indiv.	RMSE aggreg.	|Bias|	Density	RMSE indiv.	RMSE aggreg.	|Bias|	Density	RMSE indiv.	RMSE aggreg.	|Bias|	Density
$p = 2$, average number of treated outside convex hull: 4.35																
Pen. Synth.	1.3681	0.6014	0.2096	2.2814	1.3618	0.5987	0.2096	2.2823	1.3567	0.5964	0.2108	2.2853	1.3493	0.5919	0.2135	2.2796
Unpen. Synth.	1.2953	0.6297	0.2084	11.7803	1.2921	0.6353	0.2367	11.7803	1.2914	0.6430	0.2646	11.7803	1.3023	0.6767	0.3466	11.7803
Pure Synth.	1.3437	0.6008	0.2009	2.5124	1.3395	0.6000	0.2062	2.5124	1.3364	0.5998	0.2112	2.5124	1.3319	0.6017	0.2260	2.5124
Matching	1.5280	0.6368	0.2357		1.5233	0.6330	0.2293		1.5198	0.6295	0.2229		1.5157	0.6219	0.2060	
Opt. Matching	1.3603	0.6749	0.4174	4.5260	1.3541	0.6709	0.4095	4.4750	1.3493	0.6673	0.4033	4.4100	1.3381	0.6591	0.3935	4.4120
Pen. Synth. (BC)	1.3261	0.5714	0.0045		1.3263	0.5713	0.0061		1.3281	0.5715	0.0154		1.3441	0.5773	0.0423	
Unpen. Synth. (BC)	1.2535	0.5968	0.0140		1.2554	0.5970	0.0320		1.2608	0.5995	0.0497		1.2935	0.6183	0.1030	
Pure Synth. (BC)	1.3034	0.5708	0.0065		1.3041	0.5700	0.0015		1.3068	0.5704	0.0036		1.3234	0.5766	0.0175	
Matching (BC)	1.4492	0.5948	0.0031		1.4504	0.5948	0.0148		1.4539	0.5962	0.0324		1.4744	0.6071	0.0814	
Opt. Matching (BC)	1.2271	0.5285	0.0063		1.2293	0.5292	0.0303		1.2359	0.5319	0.0531		1.2676	0.5505	0.1184	
$p = 4$, average number of treated outside convex hull: 8.84																
Pen. Synth.	1.4731	0.8102	0.5713	2.9922	1.4749	0.8215	0.5868	2.9868	1.4768	0.8344	0.6024	2.9741	1.4902	0.8659	0.6398	2.9361
Unpen. Synth.	1.4453	0.8016	0.5590	4.9819	1.4482	0.8199	0.5860	4.9819	1.4529	0.8387	0.6120	4.9819	1.4771	0.8983	0.6873	4.9819
Pure Synth.	1.4480	0.8012	0.5600	3.3532	1.4509	0.8178	0.5839	3.3532	1.4553	0.8347	0.6068	3.3532	1.4768	0.8877	0.6728	3.3532
Matching	1.7010	0.8992	0.6264		1.6989	0.8994	0.6240		1.6980	0.8996	0.6212		1.7033	0.9022	0.6133	
Opt. Matching	1.5795	0.9682	0.7764	3.6150	1.5787	0.9698	0.7750	3.5250	1.5787	0.9709	0.7712	3.4870	1.5763	0.9789	0.7711	3.4050
Pen. Synth. (BC)	1.3051	0.6096	0.0078		1.3096	0.6132	0.0197		1.3180	0.6204	0.0317		1.3601	0.6493	0.0715	
Unpen. Synth. (BC)	1.2897	0.6168	0.0151		1.2938	0.6211	0.0157		1.3024	0.6276	0.0167		1.3491	0.6574	0.0181	
Pure Synth. (BC)	1.2927	0.6156	0.0141		1.2969	0.6200	0.0178		1.3050	0.6266	0.0219		1.3488	0.6558	0.0326	
Matching (BC)	1.4455	0.6431	0.0045		1.4485	0.6468	0.0248		1.4556	0.6533	0.0540		1.4967	0.6859	0.1372	
Opt. Matching (BC)	1.2815	0.6003	0.0132		1.2878	0.6038	0.0445		1.2990	0.6121	0.0767		1.3469	0.6527	0.1671	
$p = 8$, average number of treated outside convex hull: 10.00																
Pen. Synth.	1.8514	1.3112	1.1722	3.7048	1.8707	1.3385	1.2007	3.6512	1.8874	1.3615	1.2250	3.5926	1.9428	1.4326	1.2958	3.4703
Unpen Synth.	1.8275	1.3144	1.1826	4.3612	1.8494	1.3498	1.2216	4.3612	1.8717	1.3843	1.2589	4.3612	1.9426	1.4846	1.3650	4.3612
Pure Synth.	1.8275	1.3144	1.1826	4.3601	1.8494	1.3498	1.2216	4.3601	1.8717	1.3843	1.2589	4.3601	1.9426	1.4846	1.3650	4.3601
Matching	2.0989	1.3945	1.2228		2.1094	1.4080	1.2361		2.1204	1.4206	1.2480		2.1580	1.4569	1.2802	
Opt. Matching	2.0103	1.4916	1.3637	3.2010	2.0191	1.5073	1.3789	3.0990	2.0299	1.5179	1.3874	2.9480	2.0575	1.5491	1.4146	2.7920
Pen. Synth. (BC)	1.5372	0.8046	0.0167		1.5440	0.8090	0.0405		1.5556	0.8176	0.0670		1.6139	0.8557	0.1408	
Unpen. Synth. (BC)	1.5172	0.7980	0.0117		1.5213	0.8007	0.0283		1.5303	0.8062	0.0451		1.5824	0.8363	0.0926	
Pure Synth. (BC)	1.5172	0.7980	0.0116		1.5213	0.8008	0.0282		1.5304	0.8062	0.0451		1.5824	0.8363	0.0926	
Matching (BC)	1.6450	0.8274	0.0310		1.6501	0.8320	0.0665		1.6604	0.8402	0.1021		1.7179	0.8838	0.2046	
Opt. Matching (BC)	1.5497	0.8108	0.0220		1.5548	0.8142	0.0564		1.5676	0.8247	0.0927		1.6244	0.8671	0.1958	
$p = 10$, average number of treated outside convex hull: 10.00																
Pen. Synth.	2.0497	1.5653	1.4361	3.9785	2.0759	1.6012	1.4736	3.9163	2.1005	1.6322	1.5060	3.8594	2.1721	1.7179	1.5894	3.6503
Unpen Synth.	2.0307	1.5705	1.4510	4.7351	2.0618	1.6144	1.4975	4.7351	2.0928	1.6568	1.5421	4.7351	2.1871	1.7789	1.6683	4.7351
Pure Synth.	2.0307	1.5705	1.4510	4.7351	2.0618	1.6144	1.4975	4.7351	2.0928	1.6568	1.5421	4.7351	2.1871	1.7789	1.6683	4.7351
Matching	2.2908	1.6316	1.4763		2.3075	1.6520	1.4956		2.3241	1.6710	1.5132		2.3760	1.7249	1.5610	
Opt. Matching	2.2098	1.7203	1.5950	3.0870	2.2270	1.7413	1.6142	2.9670	2.2438	1.7601	1.6319	2.8680	2.2917	1.8147	1.6816	2.7000
Pen. Synth. (BC)	1.7845	1.0655	0.0166		1.7881	1.0679	0.0427		1.7961	1.0721	0.0714		1.8489	1.1045	0.1600	
Unpen. Synth. (BC)	1.7705	1.0598	0.0172		1.7725	1.0604	0.0383		1.7794	1.0638	0.0596		1.8264	1.0888	0.1204	
Pure Synth. (BC)	1.7705	1.0598	0.0172		1.7725	1.0604	0.0383		1.7794	1.0638	0.0596		1.8264	1.0888	0.1204	
Matching (BC)	1.8667	1.0867	0.0245		1.8691	1.0881	0.0626		1.8769	1.0931	0.1008		1.9285	1.1272	0.2112	
Opt. Matching (BC)	1.7970	1.0698	0.0192		1.8008	1.0724	0.0568		1.8103	1.0779	0.0940		1.8629	1.1130	0.2027	

Table 2: Monte-Carlo Simulations, $n_1 = 10$, $n_0 = 40$

	$r=1$				$r=1.2$				$r=1.4$				$r=2$			
	RMSE indiv.	RMSE aggreg.	|Bias|	Density	RMSE indiv.	RMSE aggreg.	|Bias|	Density	RMSE indiv.	RMSE aggreg.	|Bias|	Density	RMSE indiv.	RMSE aggreg.	|Bias|	Density
$p=2$, average number of treated outside convex hull: 2.88																
Pen. Synth.	1.2955	0.4948	0.1258	2.4458	1.2932	0.4934	0.1275	2.4432	1.2920	0.4945	0.1296	2.4466	1.2876	0.4940	0.1332	2.4393
Unpen. Synth.	1.1985	0.5144	0.0973	27.6213	1.1970	0.5211	0.1420	27.6213	1.1989	0.5322	0.1850	27.6213	1.2202	0.5835	0.3059	27.6213
Pure Synth.	1.2695	0.4923	0.1132	2.6839	1.2673	0.4927	0.1192	2.6839	1.2658	0.4933	0.1245	2.6839	1.2641	0.4963	0.1385	2.6839
Matching	1.4757	0.5442	0.1626		1.4715	0.5410	0.1574		1.4684	0.5382	0.1523		1.4636	0.5322	0.1389	
Opt. Matching	1.2582	0.5516	0.3226	5.8850	1.2498	0.5483	0.3216	5.9820	1.2440	0.5433	0.3167	6.0020	1.2333	0.5350	0.3034	6.1120
Pen. Synth. (BC)	1.2707	0.4749	0.0103		1.2726	0.4757	0.0067		1.2754	0.4785	0.0036		1.2852	0.4853	0.0081	
Unpen. Synth. (BC)	1.1772	0.5024	0.0083		1.1790	0.5035	0.0313		1.1847	0.5091	0.0693		1.2183	0.5449	0.1762	
Pure Synth. (BC)	1.2494	0.4782	0.0076		1.2504	0.4791	0.0085		1.2524	0.4805	0.0089		1.2623	0.4865	0.0088	
Matching (BC)	1.4237	0.5078	0.0224		1.4245	0.5077	0.0108		1.4266	0.5086	0.0006		1.4385	0.5148	0.0318	
Opt. Matching (BC)	1.1637	0.4380	0.0071		1.1612	0.4376	0.0126		1.1648	0.4388	0.0322		1.1893	0.4570	0.0915	
p = 4, average number of treated outside convex hull: 7.66																
Pen. Synth.	1.3772	0.6611	0.4131	3.3715	1.3769	0.6693	0.4291	3.3603	1.3786	0.6797	0.4459	3.3563	1.3897	0.7059	0.4848	3.3040
Unpen. Synth.	1.3373	0.6500	0.4034	10.9692	1.3397	0.6704	0.4395	10.9692	1.3446	0.6920	0.4744	10.9692	1.3725	0.7628	0.5745	10.9692
Pure Synth.	1.3470	0.6496	0.4024	3.7334	1.3489	0.6646	0.4293	3.7334	1.3524	0.6801	0.4550	3.7334	1.3712	0.7297	0.5282	3.7334
Matching	1.6240	0.7609	0.4792		1.6202	0.7572	0.4749		1.6177	0.7537	0.4704		1.6183	0.7462	0.4584	
Opt. Matching	1.4690	0.8333	0.6701	4.6600	1.4634	0.8299	0.6667	4.5690	1.4598	0.8283	0.6653	4.5340	1.4529	0.8283	0.6662	4.5050
Pen. Synth. (BC)	1.2462	0.5056	0.0031		1.2487	0.5058	0.0056		1.2527	0.5061	0.0066		1.2815	0.5177	0.0201	
Unpen. Synth. (BC)	1.2107	0.5047	0.0018		1.2128	0.5056	0.0201		1.2189	0.5088	0.0377		1.2565	0.5296	0.0887	
Pure Synth. (BC)	1.2214	0.5030	0.0008		1.2229	0.5032	0.0099		1.2274	0.5049	0.0183		1.2551	0.5166	0.0423	
Matching (BC)	1.4267	0.5576	0.0096		1.4283	0.5573	0.0338		1.4330	0.5590	0.0579		1.4615	0.5739	0.1254	
Opt. Matching (BC)	1.2074	0.4860	0.0061		1.2085	0.4868	0.0210		1.2132	0.4887	0.0480		1.2428	0.5106	0.1209	
p = 8, average number of treated outside convex hull: 9.97																
Pen. Synth.	1.6953	1.1269	0.9992	4.3286	1.7109	1.1534	1.0274	4.2726	1.7261	1.1782	1.0546	4.2336	1.7735	1.2441	1.1228	4.0561
Unpen Synth.	1.6654	1.1153	0.9917	5.0143	1.6846	1.1511	1.0323	5.0143	1.7046	1.1860	1.0713	5.0143	1.7706	1.2884	1.1827	5.0143
Pure Synth.	1.6655	1.1154	0.9918	4.9575	1.6846	1.1512	1.0323	4.9575	1.7047	1.1861	1.0713	4.9575	1.7706	1.2883	1.1825	4.9575
Matching	1.9804	1.2263	1.0740		1.9864	1.2345	1.0819		1.9932	1.2421	1.0887		2.0196	1.2650	1.1075	
Opt. Matching	1.8812	1.3390	1.2200	3.6280	1.8845	1.3451	1.2254	3.4810	1.8894	1.3506	1.2284	3.3950	1.9083	1.3736	1.2457	3.2840
Pen. Synth. (BC)	1.2816	0.5653	0.0127		1.2875	0.5685	0.0259		1.2965	0.5735	0.0400		1.3449	0.6019	0.0909	
Unpen. Synth. (BC)	1.2607	0.5649	0.0152		1.2633	0.5665	0.0176		1.2701	0.5697	0.0207		1.3108	0.5873	0.0288	
Pure Synth. (BC)	1.2607	0.5649	0.0151		1.2634	0.5665	0.0176		1.2702	0.5697	0.0207		1.3108	0.5874	0.0290	
Matching (BC)	1.4708	0.6122	0.0214		1.4741	0.6155	0.0556		1.4817	0.6221	0.0899		1.5267	0.6580	0.1880	
Opt. Matching (BC)	1.3084	0.5818	0.0072		1.3128	0.5851	0.0416		1.3220	0.5910	0.0773		1.3733	0.6336	0.1805	
p = 10, average number of treated outside convex hull: 10.00																
Pen. Synth.	1.8726	1.3592	1.2501	4.6745	1.8981	1.3961	1.2893	4.5992	1.9215	1.4282	1.3231	4.5124	1.9877	1.5135	1.4104	4.2991
Unpen Synth.	1.8494	1.3567	1.2503	5.4268	1.8770	1.3995	1.2971	5.4268	1.9050	1.4410	1.3419	5.4268	1.9926	1.5616	1.4697	5.4268
Pure Synth.	1.8494	1.3567	1.2503	5.4268	1.8770	1.3995	1.2971	5.4268	1.9050	1.4410	1.3419	5.4268	1.9926	1.5616	1.4697	5.4268
Matching	2.1648	1.4732	1.3334		2.1759	1.4867	1.3470		2.1875	1.4993	1.3593		2.2263	1.5360	1.3934	
Opt. Matching	2.0708	1.5651	1.4577	3.6240	2.0802	1.5762	1.4687	3.4790	2.0901	1.5882	1.4806	3.3780	2.1193	1.6236	1.5126	3.2200
Pen. Synth. (BC)	1.3179	0.6141	0.0064		1.3227	0.6167	0.0207		1.3316	0.6211	0.0387		1.3788	0.6485	0.0978	
Unpen. Synth. (BC)	1.2966	0.6077	0.0085		1.2988	0.6077	0.0153		1.3054	0.6098	0.0226		1.3462	0.6260	0.0432	
Pure Synth. (BC)	1.2966	0.6077	0.0085		1.2988	0.6077	0.0153		1.3054	0.6098	0.0226		1.3462	0.6260	0.0432	
Matching (BC)	1.4854	0.6578	0.0057		1.4874	0.6582	0.0425		1.4942	0.6625	0.0795		1.5396	0.6953	0.1854	
Opt. Matching (BC)	1.3534	0.6279	0.0009		1.3573	0.6291	0.0366		1.3659	0.6356	0.0730		1.4120	0.6696	0.1823	

Table 3: Monte-Carlo Simulations, $n_1 = 10$, $n_0 = 100$

	$r = 1$				$r = 1.2$				$r = 1.4$				$r = 2$			
	RMSE indiv.	RMSE aggreg.	|Bias|	Density	RMSE indiv.	RMSE aggreg.	|Bias|	Density	RMSE indiv.	RMSE aggreg.	|Bias|	Density	RMSE indiv.	RMSE aggreg.	|Bias|	Density
$p = 2$, average number of treated outside convex hull: 1.43																
Pen. Synth.	1.2724	0.4375	0.0688	2.5804	1.2711	0.4367	0.0708	2.5783	1.2700	0.4373	0.0715	2.5806	1.2680	0.4373	0.0745	2.5795
Unpen. Synth.	1.1475	0.4517	0.0553	75.6452	1.1504	0.4633	0.1205	75.6452	1.1578	0.4826	0.1815	75.6452	1.1985	0.5668	0.3450	75.6452
Pure Synth.	1.2475	0.4281	0.0647	2.8435	1.2470	0.4285	0.0699	2.8435	1.2466	0.4290	0.0741	2.8435	1.2464	0.4303	0.0838	2.8435
Matching	1.4454	0.4905	0.0816		1.4430	0.4895	0.0786		1.4414	0.4887	0.0758		1.4390	0.4876	0.0687	
Opt. Matching	1.1725	0.4575	0.2407	8.3450	1.1673	0.4538	0.2356	8.4790	1.1618	0.4517	0.2307	8.5130	1.1498	0.4460	0.2271	8.8670
Pen. Synth. (BC)	1.2648	0.4317	0.0236		1.2650	0.4313	0.0235		1.2654	0.4326	0.0226		1.2684	0.4341	0.0205	
Unpen. Synth. (BC)	1.1399	0.4470	0.0169		1.1441	0.4547	0.0806		1.1528	0.4706	0.1400		1.1976	0.5463	0.2990	
Pure Synth. (BC)	1.2405	0.4228	0.0264		1.2411	0.4233	0.0299		1.2420	0.4239	0.0326		1.2455	0.4259	0.0378	
Matching (BC)	1.4264	0.4794	0.0164		1.4267	0.4798	0.0108		1.4277	0.4805	0.0054		1.4328	0.4835	0.0090	
Opt. Matching (BC)	1.1176	0.3791	0.0249		1.1186	0.3782	0.0090		1.1203	0.3801	0.0054		1.1300	0.3863	0.0425	
p = 4, average number of treated outside convex hull: 5.82																
Pen. Synth.	1.2868	0.5276	0.2525	3.6931	1.2821	0.5322	0.2684	3.6955	1.2810	0.5377	0.2818	3.6891	1.2856	0.5562	0.3205	3.6450
Unpen. Synth.	1.2250	0.5073	0.2250	33.7853	1.2269	0.5291	0.2744	33.7853	1.2323	0.5541	0.3219	33.7853	1.2664	0.6410	0.4558	33.7853
Pure Synth.	1.2469	0.5113	0.2326	4.1437	1.2477	0.5232	0.2597	4.1437	1.2499	0.5358	0.2854	4.1437	1.2634	0.5771	0.3572	4.1437
Matching	1.5503	0.6203	0.3304		1.5460	0.6166	0.3248		1.5431	0.6134	0.3194		1.5425	0.6073	0.3067	
Opt. Matching	1.3286	0.6638	0.4936	5.7240	1.3232	0.6602	0.4892	5.6550	1.3171	0.6570	0.4831	5.6230	1.3114	0.6563	0.4770	5.7180
Pen. Synth. (BC)	1.2196	0.4541	0.0202		1.2181	0.4531	0.0152		1.2212	0.4542	0.0125		1.2410	0.4621	0.0072	
Unpen. Synth. (BC)	1.1604	0.4474	0.0238		1.1630	0.4487	0.0156		1.1700	0.4545	0.0533		1.2116	0.4919	0.1596	
Pure Synth. (BC)	1.1835	0.4479	0.0162		1.1849	0.4485	0.0009		1.1885	0.4505	0.0169		1.2085	0.4620	0.0610	
Matching (BC)	1.4308	0.5160	0.0220		1.4332	0.5181	0.0412		1.4378	0.5214	0.0598		1.4618	0.5365	0.1103	
Opt. Matching (BC)	1.1562	0.4231	0.0299		1.1587	0.4282	0.0527		1.1641	0.4345	0.0773		1.1957	0.4608	0.1436	
p = 8, average number of treated outside convex hull: 9.79																
Pen. Synth.	1.5478	0.9490	0.8178	5.0248	1.5631	0.9779	0.8521	4.9698	1.5787	1.0064	0.8834	4.8960	1.6196	1.0743	0.9562	4.6851
Unpen Synth.	1.5221	0.9349	0.8024	6.6565	1.5394	0.9741	0.8489	6.6565	1.5581	1.0126	0.8935	6.6565	1.6231	1.1262	1.0207	6.6565
Pure Synth.	1.5222	0.9352	0.8027	5.6929	1.5395	0.9738	0.8485	5.6929	1.5581	1.0118	0.8924	5.6929	1.6223	1.1236	1.0176	5.6929
Matching	1.8664	1.0824	0.9207		1.8668	1.0849	0.9225		1.8684	1.0873	0.9238		1.8810	1.0964	0.9285	
Opt. Matching	1.7385	1.1782	1.0613	4.3520	1.7385	1.1810	1.0642	4.2450	1.7376	1.1838	1.0656	4.2330	1.7405	1.1895	1.0683	4.1050
Pen. Synth. (BC)	1.1927	0.4670	0.0120		1.1979	0.4686	0.0120		1.2047	0.4731	0.0098		1.2414	0.4932	0.0122	
Unpen. Synth. (BC)	1.1703	0.4655	0.0145		1.1722	0.4671	0.0293		1.1780	0.4705	0.0431		1.2146	0.4888	0.0826	
Pure Synth. (BC)	1.1705	0.4655	0.0147		1.1723	0.4670	0.0289		1.1780	0.4702	0.0420		1.2135	0.4877	0.0795	
Matching (BC)	1.4286	0.5245	0.0046		1.4297	0.5267	0.0282		1.4346	0.5319	0.0608		1.4686	0.5622	0.1524	
Opt. Matching (BC)	1.2186	0.4753	0.0069		1.2228	0.4767	0.0272		1.2276	0.4819	0.0620		1.2664	0.5178	0.1645	
p = 10, average number of treated outside convex hull: 9.97																
Pen. Synth.	1.6937	1.1442	1.0379	5.5262	1.7173	1.1818	1.0793	5.4363	1.7390	1.2146	1.1155	5.3607	1.7991	1.2992	1.2051	5.0745
Unpen Synth.	1.6639	1.1232	1.0205	6.3829	1.6884	1.1678	1.0704	6.3829	1.7138	1.2113	1.1184	6.3829	1.7960	1.3377	1.2551	6.3829
Pure Synth.	1.6640	1.1232	1.0204	6.2551	1.6884	1.1677	1.0703	6.2551	1.7138	1.2111	1.1181	6.2551	1.7959	1.3373	1.2547	6.2551
Matching	2.0378	1.3018	1.1663		2.0431	1.3094	1.1741		2.0490	1.3165	1.1811		2.0731	1.3383	1.2013	
Opt. Matching	1.9132	1.3796	1.2830	3.7810	1.9166	1.3863	1.2898	3.6950	1.9208	1.3930	1.2961	3.6600	1.9408	1.4155	1.3174	3.5530
Pen. Synth. (BC)	1.2011	0.4775	0.0042		1.2059	0.4784	0.0050		1.2129	0.4811	0.0104		1.2540	0.5012	0.0401	
Unpen. Synth. (BC)	1.1796	0.4655	0.0088		1.1812	0.4655	0.0006		1.1865	0.4671	0.0090		1.2215	0.4794	0.0332	
Pure Synth. (BC)	1.1796	0.4656	0.0089		1.1812	0.4656	0.0004		1.1865	0.4672	0.0087		1.2213	0.4795	0.0328	
Matching (BC)	1.4430	0.5533	0.0046		1.4447	0.5550	0.0398		1.4507	0.5600	0.0750		1.4896	0.5917	0.1747	
Opt. Matching (BC)	1.2477	0.4965	0.0026		1.2497	0.5001	0.0345		1.2560	0.5062	0.0706		1.2981	0.5381	0.1748	

Table 4: Monte-Carlo Simulations, $n_1 = 100$, $n_0 = 400$

	$r = 1$				$r = 1.2$				$r = 1.4$				$r = 2$			
	RMSE indiv.	RMSE aggreg.	|Bias|	Density	RMSE indiv.	RMSE aggreg.	|Bias|	Density	RMSE indiv.	RMSE aggreg.	|Bias|	Density	RMSE indiv.	RMSE aggreg.	|Bias|	Density
$p = 2$, average number of treated outside convex hull: 1.43																
Pen. Synth.	1.2359	0.1521	0.0004	2.9157	1.2359	0.1519	0.0016	2.9155	1.2358	0.1519	0.0029	2.9156	1.2356	0.1520	0.0051	2.9156
Unpen. Synth.	1.0819	0.1965	0.0030	189.1356	1.0838	0.2073	0.0656	189.1356	1.0898	0.2350	0.1276	189.1356	1.1226	0.3489	0.2859	189.1356
Pure Synth.	1.2322	0.1510	0.0017	2.9586	1.2321	0.1509	0.0006	2.9586	1.2320	0.1509	0.0024	2.9586	1.2319	0.1509	0.0055	2.9586
Matching	1.4241	0.1625	0.0089		1.4235	0.1623	0.0076		1.4231	0.1622	0.0065		1.4227	0.1620	0.0036	
Opt. Matching	1.0623	0.1651	0.1056	13.7820	1.0588	0.1630	0.1025	14.0870	1.0557	0.1607	0.0989	14.3150	1.0494	0.1547	0.0903	14.9920
Pen. Synth. (BC)	1.2352	0.1519	0.0076		1.2353	0.1516	0.0058		1.2354	0.1517	0.0048		1.2359	0.1519	0.0034	
Unpen. Synth. (BC)	1.0812	0.1966	0.0093		1.0833	0.2056	0.0590		1.0895	0.2318	0.1207		1.1229	0.3435	0.2784	
Pure Synth. (BC)	1.2316	0.1509	0.0080		1.2316	0.1508	0.0059		1.2317	0.1508	0.0044		1.2322	0.1508	0.0020	
Matching (BC)	1.4199	0.1625	0.0104		1.4201	0.1627	0.0124		1.4205	0.1629	0.0143		1.4222	0.1637	0.0192	
Opt. Matching (BC)	1.0442	0.1271	0.0070		1.0438	0.1285	0.0164		1.0441	0.1302	0.0260		1.0484	0.1387	0.0522	
p = 4, average nb. treated outside convex hull: 31.64																
Pen. Synth.	1.1939	0.2045	0.1116	4.4127	1.1941	0.2140	0.1296	4.4056	1.1953	0.2243	0.1470	4.3925	1.2022	0.2536	0.1899	4.3451
Unpen. Synth.	1.1396	0.2255	0.1021	82.2579	1.1431	0.2628	0.1706	82.2579	1.1515	0.3083	0.2351	82.2579	1.1983	0.4568	0.4106	82.2579
Pure Synth.	1.1862	0.2020	0.1012	4.5760	1.1872	0.2148	0.1252	4.5760	1.1891	0.2283	0.1472	4.5760	1.1980	0.2696	0.2053	4.5760
Matching	1.4817	0.2698	0.2012		1.4788	0.2662	0.1969		1.4770	0.2630	0.1929		1.4771	0.2564	0.1842	
Opt. Matching	1.1872	0.4041	0.3773	8.2250	1.1805	0.4018	0.3750	8.3720	1.1744	0.3995	0.3726	8.5480	1.1620	0.3962	0.3695	9.0720
Pen. Synth. (BC)	1.1701	0.1668	0.0088		1.1708	0.1681	0.0226		1.1726	0.1712	0.0352		1.1825	0.1829	0.0633	
Unpen. Synth. (BC)	1.1158	0.1977	0.0104		1.1194	0.2117	0.0754		1.1281	0.2411	0.1364		1.1768	0.3647	0.3022	
Pure Synth. (BC)	1.1634	0.1714	0.0095		1.1644	0.1741	0.0300		1.1664	0.1788	0.0485		1.1765	0.1996	0.0969	
Matching (BC)	1.4182	0.1745	0.0057		1.4189	0.1745	0.0060		1.4210	0.1754	0.0172		1.4324	0.1817	0.0463	
Opt. Matching (BC)	1.0731	0.1397	0.0078		1.0727	0.1401	0.0107		1.0736	0.1428	0.0293		1.0851	0.1632	0.0812	
p = 8, average nb. treated outside convex hull: 89.06																
Pen. Synth.	1.3329	0.5593	0.5220	6.3437	1.3467	0.6031	0.5698	6.2703	1.3624	0.6433	0.6130	6.1805	1.4094	0.7375	0.7112	5.8088
Unpen Synth.	1.3220	0.5492	0.5093	12.4059	1.3370	0.6053	0.5699	12.4059	1.3551	0.6600	0.6281	12.4059	1.4250	0.8174	0.7921	12.4059
Pure Synth.	1.3240	0.5486	0.5092	6.6374	1.3387	0.6000	0.5647	6.6374	1.3557	0.6499	0.6178	6.6374	1.4190	0.7932	0.7674	6.6374
Matching	1.7224	0.7350	0.7026		1.7202	0.7335	0.7009		1.7196	0.7320	0.6991		1.7268	0.7305	0.6967	
Opt. Matching	1.5189	0.8892	0.8681	4.8290	1.5147	0.8907	0.8698	4.8240	1.5118	0.8918	0.8708	4.8260	1.5077	0.8961	0.8746	4.8420
Pen. Synth. (BC)	1.1361	0.1927	0.0028		1.1390	0.1938	0.0272		1.1458	0.1982	0.0459		1.1776	0.2157	0.0652	
Unpen. Synth. (BC)	1.1274	0.1986	0.0049		1.1298	0.2040	0.0464		1.1370	0.2173	0.0859		1.1811	0.2844	0.1967	
Pure Synth. (BC)	1.1298	0.1969	0.0048		1.1318	0.2012	0.0412		1.1377	0.2117	0.0756		1.1739	0.2660	0.1720	
Matching (BC)	1.4188	0.2025	0.0010		1.4201	0.2040	0.0268		1.4245	0.2096	0.0541		1.4525	0.2423	0.1293	
Opt. Matching (BC)	1.1259	0.1770	0.0008		1.1267	0.1791	0.0294		1.1311	0.1869	0.0599		1.1589	0.2319	0.1465	
p = 10, average nb. treated outside convex hull: 97.20																
Pen. Synth.	1.4496	0.7692	0.7434	7.0704	1.4722	0.8209	0.7972	6.9771	1.4959	0.8685	0.8466	6.8578	1.5611	0.9738	0.9544	6.3061
Unpen Synth.	1.4396	0.7574	0.7308	8.4734	1.4631	0.8157	0.7914	8.4734	1.4886	0.8720	0.8495	8.4734	1.5762	1.0330	1.0141	8.4734
Pure Synth.	1.4401	0.7572	0.7306	7.3541	1.4635	0.8144	0.7900	7.3541	1.4888	0.8696	0.8470	7.3541	1.5747	1.0274	1.0084	7.3541
Matching	1.8549	0.9581	0.9342		1.8562	0.9606	0.9365		1.8587	0.9627	0.9384		1.8744	0.9708	0.9454	
Opt. Matching	1.6860	1.1074	1.0909	4.2200	1.6866	1.1129	1.0964	4.1840	1.6871	1.1154	1.0987	4.1180	1.6924	1.1281	1.1108	4.0750
Pen. Synth. (BC)	1.1277	0.1949	0.0017		1.1315	0.1964	0.0203		1.1391	0.2010	0.0362		1.1785	0.2177	0.0358	
Unpen. Synth. (BC)	1.1213	0.1956	0.0026		1.1238	0.1983	0.0304		1.1307	0.2064	0.0615		1.1720	0.2521	0.1494	
Pure Synth. (BC)	1.1219	0.1954	0.0027		1.1244	0.1979	0.0290		1.1309	0.2055	0.0590		1.1701	0.2484	0.1437	
Matching (BC)	1.4190	0.2073	0.0057		1.4214	0.2113	0.0381		1.4275	0.2202	0.0702		1.4629	0.2666	0.1596	
Opt. Matching (BC)	1.1430	0.1809	0.0040		1.1455	0.1855	0.0382		1.1528	0.1963	0.0730		1.1874	0.2529	0.1708	

6. Empirical Application

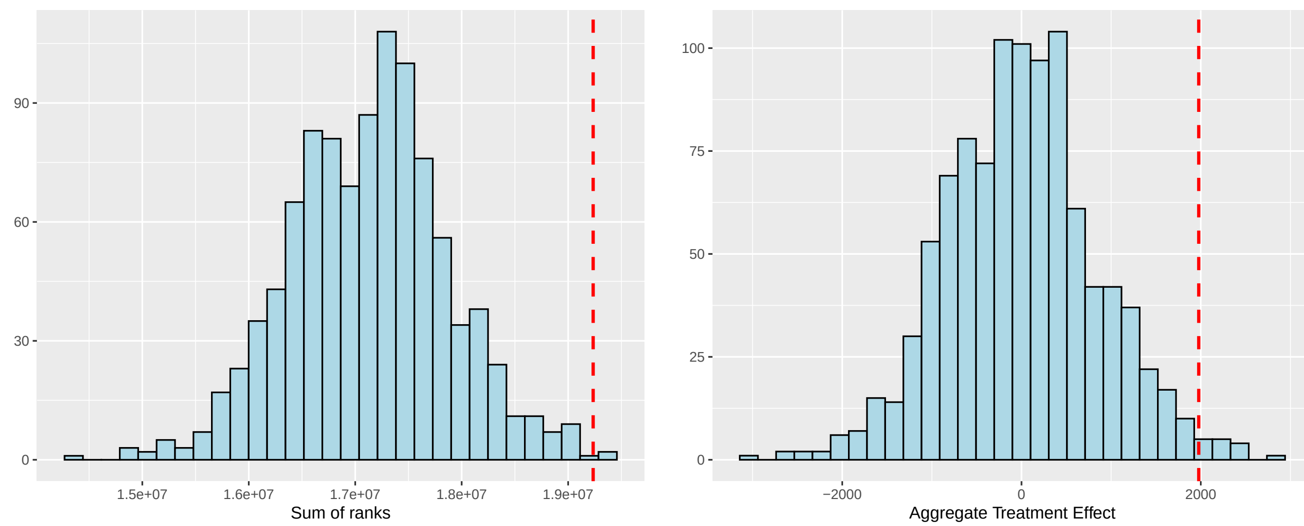
Starting with LaLonde (1986), many studies have applied data from the National Supported Work Demonstration (NSW) to demonstrate the applicability and performance of alternative estimators of treatment effects (see, e.g., Dehejia and Wahba, 2002; Smith and Todd, 2005; Abadie and Imbens, 2011). This section provides an empirical application of penalized synthetic control estimators using NSW data. The NSW program was aimed at improving employment opportunities for individuals at the margins of the labor market by providing them with temporary subsidized jobs. It targeted individuals with low levels of education or criminal records, former drug addicts, and mothers who received welfare benefits for several years. In this application, the quantity of interest is the impact of the participation in the program on 1978 yearly earnings in dollars for this specific population. In the original experiment, individuals from the targeted population were randomly split between a treatment arm ($n_1 = 185$) and a control arm ($n_0 = 260$). On that sample, the ATET estimate is \$1,794, which provides an experimental benchmark. A second control group, extracted from the Panel Study of Income Dynamics (PSID, $n_0 = 2,490$), has been made available to study estimators based on observational data (LaLonde, 1986; Dehejia and Wahba, 1999). We use this second sample to illustrate the penalized synthetic control estimator. We include in X_i the 10 covariates of the dataset (age, education, black, hispanic, married, no degree, income in 1974, income in 1975, no earnings in 1974, no earnings in 1975). Before computing estimates, we divide each of the predictors in X_i by their standard deviations in the treated sample. For the 1974 and 1975 income variables, which feature long right tails in their distributions, we first discard values above the .9 quantiles before computing the standard deviation that we use to rescale these variables.

We compare three penalized synthetic control estimators: one with a fixed (and small) value of λ ($\lambda = .1$) and two with data-driven penalties – one that minimizes individual RMSE, and one that minimizes the bias. We also report the results from nearest-neighbor matching estimators: a one-match-with-replacement procedure, and one where the optimal number of neighbors minimizes the individual RMSE. For the latter estimator, the number of matches is selected over the set of positive integers not greater than 30. In order to select the penalty level λ (and the number of neighbors in the matching procedure), we adapt the strategy proposed in Section 4.1. In particular, to reduce the computational burden, we select 170 untreated units that are close to the treated units, and for each one of them construct a synthetic unit using all the other (2,489) untreated units. The goal is to select a penalty level in a setting as close as possible to the one where the estimator is going to be applied. The set of these 170 "placebo-treated" untreated units is constructed as the union of all four nearest-neighbors of each treated unit. The selected penalty level $\widehat{\lambda}$ is then used to compute the synthetic units for each of the 185 treated, using the 2,490 untreated. In the PSID dataset, several untreated individuals are identical in terms of covariates, which can create ties in matches and non-unique solutions in penalized synthetic control weights. To make the solution unique, we consolidate identical untreated individuals into a single individual whose outcome is an average of the outcomes of its components before computing the synthetic control weights.

Table 5 reports the results. All penalized synthetic control estimators, as well as the one-match nearest-neighbor matching, in columns (1) to (4), come fairly close to the treated sample in terms of average values of the predictors. These four estimators yield treatment effects ranging from \$1,881 to \$2,171, in the ballpark of the experimental benchmark, \$1,794. The matching procedure in column (5) selects 23 neighbors, and yields a matched sample that is substantially different than the sample of treated units in the average value of the predictors. The estimated treatment effect for this matching estimator is \$983, markedly smaller than the experimental benchmark. The synthetic control estimators in columns (1) and (3), which employ positive values of λ, are very sparse. In contrast, the synthetic control estimator in column (2) employs $\lambda = 0$ and produces some highly dense estimates, as evidenced by the value of its maximal density.

Figure 3 displays the results of permutation tests described in Section 3 for an estimator with a fixed level $\lambda = .1$ – column (1) in Table 5. We consider two tests statistics: the sum of ranks and the aggregate treatment effect. The p-values are computed using one-sided tests (no effect vs. positive effect) and 1,000 permutations. In both cases, the effect lies at the right tail of the distribution and is significant at the 1% level for the sum-of-rank statistics and at 5% level for the aggregate treatment effect statistics.

Figure 3: Permutation tests for the NSW data



Note: Results are obtained for 1,000 permutations, using a fixed value of $\lambda = .1$. The left panel displays the histogram of the sum-of-ranks statistics. The right panel displays the histogram of the aggregate treatment effect. The red doted line is the value of the statistics for the observed assignment. p-values for the one sided tests are 0.003 and 0.013 respectively.

7. Conclusions

In this article, we propose a penalized synthetic control estimator that trades-off pairwise matching discrepancies with respect to the characteristics of each unit in the synthetic control against matching discrepancies with respect to the characteristics of the synthetic control unit as a whole. We study the properties of this estimator and propose data-driven choices of the penalization parameter. We show that the penalized synthetic control estimator is unique and sparse, which makes it particularly convenient for empirical applications with multiple treated units, where the focus of the analysis may be on average treatment effects. We propose a bias-correction for the penalized synthetic control estimator, and extend the inferential methods for synthetic controls in Abadie et al. (2010) to settings with multiple treated units. We show that the penalized synthetic control estimator performs well in

Table 5: LaLonde (1986) dataset, Results

	Treated	Untreated (Experimental)	Untreated (PSID)	Pen. Synthetic λ^{fixed} (1)	Pen. Synthetic $\widehat{\lambda}^{RMSE}$ (2)	Pen. Synthetic $\widehat{\lambda}^{bias}$ (3)	Matching $M = 1$ (4)	Matching $\widehat{M}^{RMSE}$ (5)
Married	0.19	0.15	0.87	0.20	0.20	0.21	0.21	0.44
Black	0.84	0.83	0.25	0.82	0.81	0.83	0.84	0.68
Hispanic	0.06	0.11	0.03	0.05	0.05	0.05	0.05	0.02
No Degree	0.71	0.83	0.31	0.67	0.66	0.67	0.68	0.61
No Earnings 1974	0.71	0.75	0.09	0.67	0.68	0.67	0.69	0.54
No Earnings 1975	0.60	0.68	0.10	0.63	0.63	0.63	0.63	0.62
Age	25.8	25.1	34.9	27.2	27.4	26.6	26.3	29.5
Education	10.4	10.1	12.1	10.4	10.3	10.4	10.3	10.6
Earnings 1974	2,095.6	2,107.0	19,428.8	2,225.1	2,256.6	2,235.6	2,275.4	2,421.4
Earnings 1975	1,532.1	1,266.9	19,063.3	1,602.4	1,619.6	1,582.4	1,598.7	1,733.2
Treatment Effect		1,794.3		1,977.3	2,171.4	1,881.4	2,138.8	982.9
λ				0.1	0	0.95		
Min. Density				1	2	1	1	23
Median Density				4	6	2	1	23
Max. Density				8	1,021	6	2	26
Active Untreated Units		260	2490	193	1,664	124	67	511

Note: The upper part of the table displays the average characteristics of the individuals in the sample. For the columns "Pen. Synthetic" (resp. "Matching"), it is an average weighted by the synthetic control (resp. matching) weights. The lower part of the table displays the resulting treatment effect, the corresponding value of λ and statistics regarding the weights. Here, density counts the number of non-zero elements in a vector of synthetic weights, $W_i^*(\lambda)$. The median, min. and max. density rows report the median, minimal and maximal densities observed for a synthetic unit. Active untreated units refers to untreated units who receive a non-zero weight in at least one synthetic unit. Notice that for matching estimators the density can differ from the chosen number of neighbors when there are ties.

simulations. The inferential methods proposed in this article are conditional on the sample and do not rely on the sampling mechanism. Sampling-based inference, which requires an approximation to sampling distribution of the penalized synthetic control estimator, is an interesting avenue for future research.

Appendix

Notation

For any real matrix X, let $\mathcal{CH}(X)$ and $\mathcal{DT}(X)$ be the convex hull and the Delaunay tesselation of the columns of X, respectively. We recall that $\mathcal{DT}(X)$ is a partition of $\mathcal{CH}(X)$. μ^{Leb} denotes the Lebesgue measure.

Proof of Lemma 1

Notice that if the first result in Lemma 1 does not hold, then $W_i^*(\lambda)$ cannot be a solution to the problem in equation (5). We start by proving the upper bound in the second inequality. Since $W_i^*(\lambda)$ minimizes (5), it follows that

$$(X_i - X_0 W_i^*(\lambda))' (X_i - X_0 W_i^*(\lambda)) + \lambda \Delta_i' W_i^*(\lambda) \leq (X_i - X_{NN_i})' (X_i - X_{NN_i}) + \lambda \Delta_i^{NN}.$$

Therefore,

$$\lambda \Delta_i' W_i^*(\lambda) \leq (1+\lambda) \Delta_i^{NN},$$

and the result follows from $\lambda > 0$. The lower bound follows from the definition of Δ_i^{NN}. □

Proof of Theorem 1

Without loss of generality, consider the case with only one treated, $n_1 = 1$. The penalized synthetic control estimator is calculated as the vector of weights that solves

$$\begin{aligned} \min_{W} \quad & f_\lambda(W) = (X_1 - X_0 W)'(X_1 - X_0 W) + \lambda W' \Delta_1, \\ \text{s.t.} \quad & W \in \mathcal{W}, \end{aligned} \tag{A.1}$$

where $\mathcal{W} = \{W \in [0,1]^{n_0} \,|\, W' 1_{n_0} = 1\}$. It is easy to check that the feasible set, $\mathcal{W}$, is convex and compact. Because f_λ is continuous and $\mathcal{W}$ is compact, it follows that the function attains a minimum on $\mathcal{W}$. Moreover, $X_0' X_0$ is positive semi-definite, so f_λ is convex.

Suppose that more than one solution exists. In particular, assume that W_1 and W_2 are solutions, with $f_\lambda(W_1) = f_\lambda(W_2) = f_\lambda^*$. Then, for any $a \in (0,1)$ we have that $aW_1 + (1-a)W_2 \in \mathcal{W}$. Because f_λ is convex, we obtain

$$f_\lambda(aW_1 + (1-a)W_2) \leq a f_\lambda(W_1) + (1-a) f_\lambda(W_2) = f_\lambda^*.$$

This implies that the problem has either a unique solution or infinitely many. In addition, if there are multiple solutions, they all produce the same fitted values $X_0 W$. To prove this, suppose there are two solutions W_1 and W_2 such that $X_0 W_1 \neq X_0 W_2$. Then, because $\|x - c\|^2$ is strictly convex in c, for $a \in (0,1)$ we obtain

$$\begin{aligned} f_\lambda(aW_1 + (1-a)W_2) &= \|X_1 - X_0(aW_1 + (1-a)W_2)\|^2 + \lambda(aW_1 + (1-a)W_2)'\Delta_1 \\ &< a\|X_1 - X_0 W_1\|^2 + (1-a)\|X_1 - X_0 W_2\|^2 + \lambda(aW_1 + (1-a)W_2)'\Delta_1 \\ &= a f_\lambda^* + (1-a) f_\lambda^* \\ &= f_\lambda^*, \end{aligned}$$

which contradicts that W_1 and W_2 are solutions. As a result, if W_1 and W_2 are solutions, then $X_0 W_1 = X_0 W_2$. Moreover, $\lambda > 0$ implies $W_1' \Delta_1 = W_2' \Delta_1$. Let $A = [X_0' \; 1_{n_0} \; \Delta_1]$. It follows that, if W_1 and W_2 are solutions, then $A'(W_1 - W_2) = 0_{p+2}$ (where 0_{p+2} is a $(p+2) \times 1$ vector of zeros).

Karush-Kuhn-Tucker conditions imply:

$$X_j'(X_1 - X_0 W) - \frac{\lambda}{2}\Delta_{1,j} = \pi - \gamma_j$$
$$W_j \geq 0, \ W' 1_{n_0} = 1, \ \gamma_j \geq 0, \ \gamma_j W_j = 0.$$

Stacking the first n_0 conditions above and pre-multiplying by W', we obtain

$$W' X_0'(X_1 - X_0 W) - \frac{\lambda}{2} W' \Delta_1 = \pi.$$

From this equation, it follows that the value of π is unique across solutions, because $X_0' W$ and $W' \Delta_1$ are unique across solutions. Given that π is unique, the equations

$$X_j'(X_1 - X_0 W) - \frac{\lambda}{2}\Delta_{1,j} = \pi - \gamma_j.$$

imply that the γ_j's are unique across solutions. Let $\widetilde{X}_0$ be the submatrix of X_0 formed by the columns associated with zero γ_j's, and define $\widetilde{W}$, $\widetilde{\Delta}_1$, and $1_{\widetilde{n}_0}$ analogously, where $\widetilde{n}_0$ is the number of columns of $\widetilde{X}_0$. Then,

$$\widetilde{X}_0'(X_1 - \widetilde{X}_0 \widetilde{W}) = \frac{\lambda}{2}\widetilde{\Delta}_1 + \pi 1_{\widetilde{n}_0}. \tag{A.2}$$

Notice that if $\lambda > 0$, then $\|X_1 - X_0 W\| = 0$ implies that $\widetilde{\Delta}_1$ is a constant vector. We therefore obtain that if $\lambda > 0$ and $\widetilde{\Delta}_1$ is not constant, then it must be the case that $\|X_1 - X_0 W\| > 0$.

Let $\widetilde{A} = [\widetilde{X}_0' \ 1_{\widetilde{n}_0} \ \widetilde{\Delta}_1]$. Consider the case $\widetilde{n}_0 \geq p+2$. In this case $\widetilde{A}$ has full column rank, which implies that equation (A.2) cannot hold if $\lambda > 0$. As a result, when $\lambda > 0$, the solution to (A.1) has $p+1$ non-zero components at most.

Consider now the case $\widetilde{n}_0 \leq p+1$. For this case $\widetilde{A}$ has full row rank. Moreover, if $\widetilde{W}_1$ and $\widetilde{W}_2$ are solutions, it must be the case that $\widetilde{A}'(\widetilde{W}_1 - \widetilde{W}_2) = 0_{p+2}$. However, because $\widetilde{A}$ has full row rank the system $\widetilde{A}' z = 0_{p+2}$ admits only the trivial solution, $z = 0_{\widetilde{n}_0}$, which implies that the solution to (A.1) is unique. □

Lemma A.1 (Optimality of Delaunay for the Compound Discrepancy, Rajan, 1994) *Let $Z \in \mathcal{CH}(X_0)$. Consider a solution $\widetilde{W} = (\widetilde{W}_{n_1+1}, \ldots, \widetilde{W}_n)'$ of the problem*

$$\min_{W \in [0,1]^{n_0}} \sum_{j=n_1+1}^{n} W_j \|X_j - Z\|^2, \tag{A.3}$$

$$\text{s.t.} \quad X_0 W = Z, \ \sum_{j=n_1+1}^{n} W_j = 1. \tag{A.4}$$

Then, non-zero values of $\widetilde{W}_j$ occur only among the vertices of the Delaunay simplex containing Z.

We restate the proof of Lemma 10 in Rajan (1994) for clarity and note that it does not rely on general quadratic position of the set of points.

Proof of Lemma A.1

For a point $X \in \mathbb{R}^p$, consider the transformation $\phi : X \to (X, \|X\|^2)$. The images under ϕ of points in $\mathbb{R}^p$ belong to the paraboloid of revolution $\mathcal{P}$ with vertical axis and equation $x_{p+1} = \sum_{i=1}^{p} x_i^2$. By Theorem 17.3.1 in Boissonnat and Yvinec (1998), the Delaunay tessellation of the convex hull of the n_0 points $X_{n_1+1}, \dots, X_n$ in $\mathbb{R}^p$ are obtained by projecting onto $\mathbb{R}^p$ the faces of the lower envelope of the convex hull of the n_0 points $\phi(X_{n_1+1}), \dots, \phi(X_n)$ obtained by lifting the X_j's onto the paraboloid $\mathcal{P}$.

Now consider points $\left(\sum_{j=n_1+1}^{n} W_j X_j, \sum_{j=n_1+1}^{n} W_j \|X_j\|^2\right)$ subject to the constraints in (A.4). These points are equal to $\left(Z, \sum_{j=n_1+1}^{n} W_j \|X_j - Z\|^2 + \|Z\|^2\right)$ and belong to the convex hull of $\phi(X_{n_1+1}), \dots, \phi(X_n)$. Hence, a solution of (A.3) for a fixed Z is given by such a point with the lowest $(p+1)$-th coordinate. It is a point on the lower envelope of the convex hull of $\phi(X_{n_1+1}), \dots, \phi(X_n)$, so Z belongs to a p-simplex of the Delaunay tessellation. As a consequence, non-zero entries of $\widetilde{W}$ occur only among the vertices of the face of the Delaunay tessellation of the columns of X_0 containing Z. □

Proof of Theorem 2

It is enough to prove that the result holds for one treated unit, so we consider the case $n_1 = 1$ and drop the treated units subscripts from the notation. We proceed by contradiction. Suppose that the synthetic control weights are given by the vector $W^*(\lambda) = (W_2^*(\lambda), \dots, W_n^*(\lambda))'$, and that $W_j^*(\lambda) > 0$ for j which is not a vertex of the Delaunay simplex in $\mathcal{DT}(X_0)$ containing $X_0 W^*(\lambda)$. Because $X_0 W^*(\lambda) \in \mathcal{CH}(X_0)$, it follows from Lemma A.1 that we can always choose a n_0-vector of weights $\widetilde{W} \in [0,1]^{n_0}$, such that *(i)* $X_0 \widetilde{W} = X_0 W^*(\lambda)$, *(ii)* $\sum_{j=2}^{n} \widetilde{W}_j = 1$, *(iii)* $\widetilde{W}_j = 0$ for any j that is not a vertex of the Delaunay simplex containing $X_0 W^*(\lambda)$, and *(iv)* $\widetilde{W}$ induces a lower compound discrepancy than $W^*(\lambda)$ relative to $X_0 \widetilde{W} = X_0 W^*(\lambda)$,

$$\sum_{j=2}^{n} \widetilde{W}_j \|X_j - X_0 \widetilde{W}(\lambda)\|^2 < \sum_{j=2}^{n} W_j^*(\lambda) \|X_j - X_0 W^*(\lambda)\|^2. \tag{A.5}$$

For any $W \in [0,1]^{n_0}$ it can be easily seen that

$$\sum_{j=2}^{n} W_j \|X_j - X_1\|^2 = \sum_{j=2}^{n} W_j \|X_j - X_0 W\|^2 + \|X_1 - X_0 W\|^2. \tag{A.6}$$

Combining equations (A.5) and (A.6) with the fact that $\|X_1 - X_0 \widetilde{W}\|^2 = \|X_1 - X_0 W^*(\lambda)\|^2$, we obtain

$$\sum_{j=2}^{n} \widetilde{W}_j \|X_j - X_1\|^2 < \sum_{j=2}^{n} W_j^*(\lambda) \|X_j - X_1\|^2.$$

As a result

$$\|X_1 - X_0 \widetilde{W}\|^2 + \lambda \sum_{j=2}^{n} \widetilde{W}_j \|X_j - X_1\|^2 < \|X_1 - X_0 W^*(\lambda)\|^2 + \lambda \sum_{j=2}^{n} W_j^*(\lambda) \|X_j - X_1\|^2,$$

which contradicts the premise that $W^*(\lambda)$ is a solution to (5). □

Proof of Theorem 3

Since the columns of $[X_1\ X_0]$ are in general quadratic position, the augmented Delaunay triangulation, $\mathcal{DT}([X_1 : X_0])$, exists and is unique. Without loss of generality, consider the case of a single treated unit, and normalize X_1 to be at the origin. Let X_0^* be the submatrix of X_0 formed by the columns of X_0 that are connected to X_1 in the augmented triangulation, $\mathcal{DT}([X_1 : X_0])$, and $\mathcal{UT}(X_1, X_0)$ be the union of the Delaunay simplices that have X_1 as a vertex in $\mathcal{DT}([X_1 : X_0])$. Consider a point $z \in \mathcal{CH}([X_1 : X_0])\backslash\mathcal{UT}(X_1, X_0)$. We will first show that z cannot be equal to $X_0W^*(\lambda)$. Because z does not belong to $\mathcal{UT}(X_1, X_0)$ and because the set $\mathcal{CH}([X_1 : X_0])$ is convex, it is always possible to find a point $v \in \mathcal{CH}([X_1 : X_0])\backslash\mathcal{UT}(X_1, X_0)$ on the line segment that connects z and X_1 such that $\|X_1 - v\| < \|X_1 - z\|$ (or, equivalently, $\|v\| < \|z\|$). For any point in $x \in \mathcal{CH}([X_1 : X_0])$ consider the set of non-negative weights, $w_1(x), \ldots, w_n(x)$, such that: *(i)* $\sum_{i=1}^n w_i(x) = 1$, *(ii)* $\sum_{i=1}^n w_i(x)X_i = x$, and *(iii)* if X_i is not a vertex of the Delaunay simplex containing x, then $w_i(x) = 0$. If $x \in \mathcal{CH}([X_1 : X_0])\backslash\mathcal{UT}(X_1, X_0)$, then the Delaunay simplex containing x in $\mathcal{DT}([X_1 : X_0])$ is the same as the the Delaunay simplex contain x in $\mathcal{DT}(X_0)$ (Devillers and Teillaud, 2003; Boissonnat et al., 2009). Therefore, by Theorem 2, if $x \in \mathcal{CH}([X_1 : X_0])\backslash\mathcal{UT}(X_1, X_0)$ and $X_0W^* = x$, then $W^*(\lambda) = (w_2(x), \ldots, w_n(x))'$. Now, let $f(x) = \sum_{i=1}^n w_i(x)\|X_i\|^2 = \sum_{i=1}^n w_i(x)\|X_1 - X_i\|^2$. This function is convex because it is the lower boundary of the convex hull of $\{(X_1, \|X_1\|^2), \ldots, (X_n, \|X_n\|^2)\}$ (Rajan, 1994), and is minimized at $x = X_1$. As we move from z to v, we travel in the direction of the minimun of $f(x)$. Because $f(x)$ is a convex function, it follows that $f(v) < f(z)$. Because $\|v\| < \|z\|$ and $f(v) < f(z)$, it follows that $X_0W^*(\lambda) \neq z$, regardless of the value of λ. This implies that $X_0W^*(\lambda)$ must belong to $\mathcal{UT}(X_1, X_0)$ and the result follows from Theorem 2. □

References

Abadie, A. (2020). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*. Forthcoming. Available online at `https://www.aeaweb.org/articles/pdf/doi/10.1257/jel.20191450`.

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(490):493–505.

Abadie, A., Diamond, A., and Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science*, 59(2):495–510.

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

Abadie, A. and Imbens, G. W. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica*, 74(1):235–267.

Abadie, A. and Imbens, G. W. (2011). Bias-corrected matching estimators for average treatment effects. *Journal of Business & Economic Statistics*, 29(1):1–11.

Acemoglu, D., Johnson, S., Kermani, A., Kwak, J., and Mitton, T. (2016). The value of connections in turbulent times: Evidence from the united states. *Journal of Financial Economics*, 121:368–391.

Amjad, M., Shah, D., and Shen, D. (2018). Robust synthetic control. *Journal of Machine Learning Research*, 19(22):1–51.

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, Princeton, NJ.

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., and Wager, S. (2018). Synthetic Difference in Differences. arXiv:1812.09970.

Athey, S., Bayati, M., Doudchenko, N., Imbens, G., and Khosravi, K. (2017). Matrix Completion Methods for Causal Panel Data Models. arXiv:1710.10251.

Athey, S. and Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(7):7353–7360.

Athey, S. and Imbens, G. (2018). Design-based analysis in difference-in-differences settings with staggered adoption.

Ben-Michael, E., Feller, A., and Rothstein, J. (2019). Synthetic controls and weighted event studies with staggered adoption.

Ben-Michael, E., Feller, A., and Rothstein, J. (2019). The Augmented Synthetic Control Method. arXiv:1811.04170.

Bohn, S., Lofstrom, M., and Raphael, S. (2014). Did the 2007 Legal Arizona Workers Act reduce the state's unauthorized immigrant population? *Review of Economics and Statistics*, 96(2):258–269.

Boissonnat, J.-D., Devillers, O., and Hornus, S. (2009). Incremental construction of the Delaunay graph in medium dimension. In *Annual Symposium on Computational Geometry*, pages 208–216, Aarhus, Denmark.

Boissonnat, J.-D. and Yvinec, M. (1998). *Algorithmic Geometry.* Cambridge University Press, New York, NY, USA.

Cattaneo, M. D., Feng, Y., and Titiunik, R. (2019). Prediction intervals for synthetic control methods. arXiv:1912.07120.

Chernozhukov, V., Wüthrich, K., and Zhu, Y. (2019). An Exact and Robust Conformal Inference Method for Counterfactual and Synthetic Controls. arXiv:1712.09089v6.

Cunningham, S. and Shah, M. (2018). Decriminalizing indoor prostitution: Implications for sexual violence and public health. *The Review of Economic Studies*, 85(3):1683–1715.

Dehejia, R. and Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *The Review of Economics and Statistics*, 84(1):151–161.

Dehejia, R. H. and Wahba, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association*, 94(448):1053–1062.

Devillers, O. and Teillaud, M. (2003). Perturbations and vertex removal in a 3D Delaunay triangulation. In *14th ACM-Siam Symposium on Discrete Algorithms (SODA)*, pages 313–319, Baltimore, MA, United States.

Doudchenko, N. and Imbens, G. W. (2016). Balancing, regression, difference-in-differences and synthetic control methods: A synthesis. *NBER Working Papers*, 22791.

Dube, A. and Zipperer, B. (2015). Pooling Multiple Case Studies Using Synthetic Controls: An Application to Minimum Wage Policies. IZA Discussion Papers 8944, Institute for the Study of Labor (IZA).

Firpo, S. and Possebom, V. (2018). Synthetic Control Method: Inference, Sensitivity Analysis and Confidence Sets. *Journal of Causal Inference*, 6(2):1–26.

Gobillon, L. and Magnac, T. (2016). Regional policy evaluation: Interactive fixed effects and synthetic controls. *Review of Economics and Statistics*, 98(3):535–551.

Hackmann, M. B., Kolstad, J. T., and Kowalski, A. E. (2015). Adverse selection and an individual mandate: When theory meets practice. *American Economic Review*, 105(3):1030–1066.

Hirano, K., Imbens, G. W., and Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*, 71(4):1161–1189.

Imbens, G. W. and Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction.* Cambridge University Press, New York, NY, USA.

Kleven, H. J., Landais, C., and Saez, E. (2013). Taxation and international migration of superstars: Evidence from the European football market. *American Economic Review*, 103(5):1892–1924.

Kreif, N., Grieve, R., Hangartner, D., Turner, A. J., Nikolova, S., and Sutton, M. (2016). Examination of the synthetic control method for evaluating health policies with multiple treated units. *Health Economics*, 25(12):1514–1528.

LaLonde, R. J. (1986). Evaluating the econometric evaluations of training programs with experimental data. *The American Economic Review*, 76(4):604–620.

Okabe, A., Boots, B., Sugihara, K., and Chiu, S. N. (2000). *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams.* Series in Probability and Statistics. John Wiley and Sons, Inc.

Rajan, V. T. (1994). Optimality of the Delaunay triangulation in $\mathbb{R}^d$. *Discrete & Computational Geometry*, 12(2):189–202.

Rosenbaum, P. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

Rubin, D. B. (1973). The use of matched sampling and regression adjustment to remove bias in observational studies. *Biometrics*, 29(1):185–203.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688.

Smith, J. and Todd, P. (2005). Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics*, 125(1-2):305–353.