A Penalized Synthetic Control Estimator for Disaggregated Data

(preliminary and incomplete)

Alberto Abadie — MIT

Jérémy L'Hour — CREST

October 3, 2018

Abstract

Synthetic control methods are commonly applied in empirical research to estimate the effects of treatments or interventions of interest on aggregate outcomes. A synthetic control estimator compares the outcome of a treated unit – that is, a unit exposed to the intervention of interest – 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 in settings with disaggregated data, that is, when the sample 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, abadie@mit.edu. Jérémy L'Hour, CREST, ENSAE, 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 and Clément de Chaisemartin.

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 a 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 traditional 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 sparsity, non-negativity of the weights, and weights that sum to one. Like for nearest neighbor matching estimators, most of the synthetic control weights are equal to zero and a small number of untreated units contribute positive weights to reproduce the counterfactual of each treated observation without the treatment. 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 incorporating 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 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 most applications of the synthetic control framework have focused on cases where only one or a few aggregate units are exposed to the intervention of interest, the method has found recent applications in contexts with disaggregated data, where samples contain large numbers of treated and untreated units, and the interest lies 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 each particular treated unit is unlikely to fall in the convex hull of the untreated units, especially if the number of untreated units is small. As a result, lack of uniqueness is not often a problem 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), a large number of units in the donor pool creates a setting where single untreated units may provide close matches to the treated units in the sample. Therefore, in such a setting the researcher faces a trade-off between minimizing the covariate discrepancy between each treated unit and its synthetic control as a whole (pure synthetic control case) and minimizing the covariate discrepancy between each treated unit and each unit that contributes to its synthetic control (pure matching case).

This paper provides a generalized synthetic control framework for estimation and inference. The framework builds on synthetic controls and introduces 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 matching variables. Moreover, it can be shown 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 the formal properties of the penalized synthetic control estimator and propose data driven choices of the penalization parameter.

Our approach belongs to the recent literature on "machine learning" estimators for program evaluation problems. Following Doudchenko and Imbens (2016) which represents synthetic controls as a solution to complete an outcome matrix with missing entries, Athey et al. (2017) assumes an underlying sparse factor structure for the outcome under no treatment and adapts matrix completion techniques to estimate a counterfactual. Their estimator penalizes the complexity of the factor structure, while our approach penalizes the discrepancy between the treated unit and each control unit that enters the synthetic unit.

2. Penalized Synthetic Control

2.1. Synthetic Control for Disaggregated Data

We code treatment using a binary variable, D, so $D = 1$ for treated individuals and $D = 0$ otherwise. To define the object of interest we adopt the potential outcome notation in Rubin (1974). Let Y_1 and Y_0 be random variables representing potential outcomes under treatment and under no treatment, respectively. The treatment effect is $Y_1 - Y_0$. Realized outcomes are defined as

$$Y = \begin{cases} Y_1 & \text{if } D = 1, \\ Y_0 & \text{if } D = 0. \end{cases}$$

Let X be a $(p \times 1)$-vector of pre-treatment predictors of Y_0. Consider the distributions of the triple (Y_1, Y_0, X) under treatment and no treatment, with $E[\cdot|D = 1]$ and $E[\cdot|D = 0]$ denoting the corresponding expectation operators, and $E[\cdot|X, D = 1]$ and $E[\cdot|X, D = 0]$ denoting expectations conditional on X. Let P_1 and P_0 be the probability measures that describe the distribution of X for treated and nontreated, respectively.

Assumption 1 (Sampling) *$\{(Y_{1i}, X_i)\}_{i=1,\ldots,n_1}$ are n_1 independent draws from the distribution of (Y_1, X) and $\{(Y_{0i}, X_i)\}_{i=n_1+1,\ldots,n}$ are n_0 independent draws from the distribution of (Y_0, X).*

Combining data for treated and nontreated we obtain the pooled sample, $\{(Y_i, D_i, X_i)\}_{i=1}^{n}$, $n = n_0 + n_1$. To simplify notation, we reorder the observations in the sample so that the n_1 treated observations are first and the n_0 untreated observations are last. The quantity of interest is the average treatment effect on the treated (ATET):

$$\tau = E[Y_1 - Y_0|D = 1]. \tag{1}$$

Assumption 2 (Nested support) *$P_1 \ll P_0$, that is, P_1 is absolutely continuous with respect to P_0.*

Assumption 3 (Unconfoundedness I) $E[Y_0|X, D = 1] = E[Y_0|X, D = 0]$.

Versions of assumptions 2 and 3 are ubiquitous in the program evaluation literature (see, e.g., Imbens, 2004). Assumption 2 states that there is no value of X for which individuals are always treated. In other words, for any treated, it should be possible to find a non-treated with the same value of the covariates in the population. Assumption 3 states that conditionally on a set of observed covariates or confounding factors, X, the expected potential outcome without the treatment is the same for treated and control individuals. Graphical causal structures that support Assumption 3 are studied in Pearl (2000) and the subsequent literature.

Notice that, under these two assumptions, the counterfactual $E[Y_0|D = 1]$ can be expressed as a weighted average of the outcome among the untreated,

$$\tau = E[Y|D = 1] - E[VY|D = 0], \tag{2}$$

where $V = dP_1/dP_0$. Many econometric estimators of τ based on Assumptions 2 and 3, whether explicitly or implicitly, employ a sample analog of equation (2),

$$\frac{1}{n_1}\sum_{i=1}^{n} Y_i D_i - \frac{1}{n_0}\sum_{i=1}^{n} Y_i(1 - D_i)V_i. \tag{3}$$

Popular estimators of this type in micro-econometrics include most notably regression (Angrist and Pischke, 2008; Abadie et al., 2015), propensity score weighting (Rosenbaum and Rubin, 1983; Hirano et al., 2003) and matching (Smith and Todd, 2005). For example, in the case of the pair-matching estimator, the weight V_i given to control unit i is equal to an integer counting 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.

While these assumptions are enough to recover the average treatment effect in equation (1), identification of a wide variety of parameters can be attained by strengthening the identifying conditions as in Assumptions 2′ and 3′ below.

Assumption 2′ (Common support) $P_1 \ll P_0$ *and* $P_0 \ll P_1$.

Assumption 3′ (Unconfoundedness II) $Y_1, Y_0 \perp\!\!\!\perp D|X$.

Parameters identified by the addition of Assumptions 2′ and 3′ include quantile treatment effects, that is, differences in the quantiles of the distributions of potential outcomes (Firpo, 2007), bounds on the distribution of the treatment effect (Firpo and Ridder, 2008), or counterfactual distributions (Chernozhukov et al., 2013), among others. They also include parameters describing conditional features of the distribution of potential outcomes (see, e.g., Crump et al., 2008) and regression parameters obtained after imposing the same distribution of X for treated and non-treated (Ho et al., 2007; Abadie and Spiess, 2016). While, for the sake of clarity, this article focuses on the estimation of average treatment effects, the generalized synthetic control method outlined here can be applied to estimate any of the parameters above. Moreover, Assumptions 1-3, which are adopted here for simplicity, are not the only possible identification conditions in a synthetic control setting, nor necessarily the least restrictive ones. In particular, Abadie et al. (2010) show that under a factor-structure condition on the regression residual of the outcome on the covariates for the untreated, using synthetic controls that match pre-treatment outcomes for the treated help control for unobserved confounding that arises from heterogeneity in the factor loadings.

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, $i = 1, \dots, n_1$, compute the n_0-vector of weights $W_i^* = (W_{i,n_1+1}^*, \dots, W_{i,n}^*)$ that solves

$$\begin{aligned} \min_{W_i \in \mathbb{R}^{n_0}} \quad & \left\| X_i - \sum_{j=n_1+1}^{n} W_{i,j} X_j \right\|^2 \\ \text{s.t.} \quad & W_{i,n_1+1} \geq 0, \dots, W_{i,n} \geq 0, \\ & \sum_{j=n_1+1}^{n} W_{i,j} = 1, \end{aligned} \tag{4}$$

   where $W_{i,j}^*$ is the weight given to control unit j in the synthetic control unit corresponding to treated unit i.

2. Estimate τ using the mean difference between the realized outcome and the synthetic outcome for the treated

$$\widehat{\tau} = \frac{1}{n_1} \sum_{i=1}^{n_1} \left[Y_i - \sum_{j=n_1+1}^{n} W_{i,j}^* Y_j \right], \tag{5}$$

Notice that $\widehat{\tau}$ is the estimator in equation (3) reweighting each nontreated unit, $j = n_1 + 1, \dots, n$, by $V_j = (n_0/n_1) S_j$, where $S_j = \sum_{i=1}^{n_1} W_{i,j}^*$ is the sum of the synthetic control weights assigned to unit j. Lemma A.2 in the appendix derives some properties of S_j.

While, to simplify notation, we described here a cross-sectional setting only, the extension to the more common panel data setting for synthetic controls is immediate and we will use it later.

Notice that, when X_i belongs to the convex hull of $\{X_{n_1+1}, \dots, X_n\}$, the value of the objective function (4) at the minimum is equal to zero and multiple solutions may exist. By Carathéodory's theorem, a solution with at most $p + 1$ non-zero weights exists in that case. On the other hand, it is easy to show that if X_i does not belong to the convex hull of $\{X_{n_1+1}, \dots, X_n\}$ and under weak regularity conditions (that is, if observations are in "general position") the solution is unique and involves at most $p + 1$ non-zero weights, see Theorem 1 below.

2.2. Penalized Synthetic Control

The main contribution of this article is to propose an alternative, penalized version of the synthetic control estimator in equation (4). For treated unit i and given a positive constant λ, the penalized synthetic control weights, $W^*_{i,j}(\lambda)$, solve

$$\begin{aligned} \min_{W_i \in \mathbb{R}^{n_0}} \quad & \left\| X_i - \sum_{j=n_1+1}^{n} W_{i,j} X_j \right\|^2 + \lambda \sum_{j=n_1+1}^{n} W_{i,j} \|X_i - X_j\|^2 \\ \text{s.t.} \quad & W_{i,n_1+1} \geq 0, \ldots, W_{i,n} \geq 0, \\ & \sum_{j=n_1+1}^{n} W_{i,j} = 1. \end{aligned} \tag{6}$$

The penalized synthetic control estimator is then given by

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

The tuning parameter λ sets the trade-off between componentwise and aggregate fit. The choice of the value of λ is important and will be discussed in Section 3. 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 interpolations 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 $\Delta_i^{NN} = \min_{j=1,\ldots,n_0} \|X_i - X_{n_1+j}\|^2$, be the smallest discrepancy between unit i and the units in the donor pool. Finally, let $W_i^*(\lambda)$ be a solution to (6), 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.

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

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

and for $\lambda > 0$

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

The first inequality 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 inequality shows that the tuning parameter λ controls the compound discrepancy between the treated unit and the units that contribute to the synthetic control, $\Delta_i' W_i^*(\lambda)$. All proofs are in the appendix.

Some remarks are in order to justify the choice of the penalization term in equation (6). First, notice that the penalty term is linear rather than quadratic in the weights. This has the advantage of producing easy-to-interpret sparse solutions, similarly to a matching procedure.

Notice also that the optimization problem in (6) can be solved via quadratic programming, like the standard synthetic control in (5). To see why notice that, in matrix notation, program (6) is

$$\begin{aligned} &\min_{W \in \mathbb{R}^{n_0}} \; (X_i - X_0 W)' (X_i - X_0 W) + \lambda \Delta_i' W \\ &\text{s.t.} \quad 1_{n_0}' W = 1, W \geq 0, \end{aligned} \tag{8}$$

where 1_{n_0} is the $(n_0 \times 1)$ vector of ones and the inequality restriction applies to each component of W.

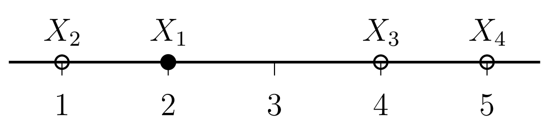
A third remark has to do with uniqueness of the solution. In the absence of the penalty term (that is, when $\lambda = 0$), the problem in (6) and (8) can be solved by projecting X_i on the convex hull of X_0. Existence of sparse solutions follows from Carathéoroy's theorem. However, if $\lambda = 0$ the solution to the problem in (6) and (8) may not be unique, especially 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 (i) $\lambda > 0$; (ii) any submatrix of X_0' consisting of no more than p rows has full row rank; (iii) 1_{p+1} does not belong to the column space of any submatrix of X_0' consisting of $p+1$ rows; (iv) any submatrix composed by $p+2$ or more rows of $(X_0'\ 1_{n_0}\ \Delta_i)$ has full column rank. Then, the optimization problem in equation (6) admits a unique solution $W_i^*(\lambda)$ with at most $p+1$ non-zero components.*

Condition *(i)* imposes a non-zero penalization on the compound discrepancy. Condition *(ii)* implies that any subset of p or fewer control observations are not linearly dependent in the values of the predictors. Conditions *(iii)* and *(iv)* require that there is no subset of $p+1$ or more control observations with values of the predictors that fall in a lower-dimensional affine subspace. In addition, condition *(iv)* requires 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.

Figure 1: A simple example



Notice that X_1 belongs to $[1, 5]$, the convex hull of the columns of X_0. Consider first the case with $\lambda = 0$. Then, $W^*(0) = (2/3\ \ 1/3\ \ 0)'$ and $W^{**}(0) = (3/4\ \ 0\ \ 1/4)'$ are the only two sparse solutions (with number of non-zero weights not greater than $p+1 = 2$) to (6). 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 (e.g., under a Lipschitz bound on $E[Y|X, W = 0]$). However, the better compound fit of $W^*(0)$ is not reflected in a better value in the objective function in (4). Moreover, because any convex combination of $W^*(0)$ and $W^{**}(0)$ is also a solution, the program in (4) 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

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

From Figure 1, it is apparent that $W^*(0)$, which is obtained making $a = 1$, produces the lowest compound discrepancy among all the solutions to equation (4).

When $\lambda > 0$, however, the program (6) 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. □

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 face of the Delaunay complex containing $X_0 W_i^*(\lambda)$ in the Delaunay tessellation of $X_{n_1+1}, \ldots, X_n$.

Theorem 2 (Delaunay Property) *Let $W_i^*(\lambda)$ be a solution to the penalized synthetic control problem in (6) 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 face of the Delaunay complex containing $X_0 W_i^*(\lambda)$ it holds that $W_{i,j}^*(\lambda) = 0$.*

This result along with the first part of Lemma 1, which bounds $\|X_i - X_0 W_i^*(\lambda)\|$, provides a notion of proximity between each treated unit X_i and the untreated units that contribute to its synthetic control. Theorem 2 provides also 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 (6). Recall that when $\lambda = 0$, the problem of minimizing $\|X_i - X_0 W\|$ subject to the weight constraints may have multiple (infinite) number of solutions, 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 (6) for an arbitrarily small value of λ requires, in practice, a choice for λ. It also creates computational difficulties, as the minimization problem is close to one with multiple solutions and the dimension of W may be large. Theorem 2 implies that the solution of (6) for $\lambda \to 0$ assigns 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.

2.3. Bias-Corrected Synthetic Control

We will also consider bias-corrected versions of synthetic control estimator. We adopt a bias correction analogous to that implemented in Abadie and Imbens (2011) for matching estimators. Let $\mu_0(x) = E[Y|X = x, D = 0]$, and let $\widehat{\mu}_0(x)$ be an estimator of $\mu_0(x)$. A bias-corrected version of the synthetic control estimator in equation (7) is

$$\widehat{\tau}_{BC}(\lambda) = \frac{1}{n_1} \sum_{i=1}^{n_1} \left[\big(Y_i - \widehat{\mu}_0(X_i)\big) - \sum_{j=n_1+1}^{n} W^*_{i,j}(\lambda) \big(Y_j - \widehat{\mu}_0(X_j)\big) \right]. \tag{9}$$

3. Penalty Choice

In this section we present two data-driven selectors for the penalty term, λ. In the context of treatment effects estimation, cross-validation (CV) 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). 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-intervetion period.

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

This section discusses a leave-one-out cross-validation procedure to find an optimal value λ by minimizing mean squared prediction error for the untreated units in the post-intervention period. Consider a balanced panel data setting with T periods and $T_0 < T$ pre-intervention periods. Let Y_{it} be the outcome for unit i at time t. 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 mean squared prediction error for the individual outcomes,

$$\frac{1}{n_0(T - T_0)} \sum_{i=1}^{n_0} \sum_{t=T_0+1}^{T} \left(\widehat{\tau}_{it}(\lambda) \right)^2.$$

3.2. 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 is the most natural choice, we will assume that the validation period is at the end of the pre-intervention period, although other choices are possible. The procedure is as follows:

1. Split the pre-intervention period that contains T_0 dates into $T_0 - k$ initial training dates and k subsequent validation dates.

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

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

   where $W^*_{i,j}$ solve (6) with X measured in the training period.

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

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

or the squared prediction error of the aggregate outcomes,

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

Notice that the cross-validation procedures delineated can also be applied here to guide model selection (i.e., choice of V) as in Abadie et al. (2015).

4. 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. We next describe three 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 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 sample value of a test statistic, and 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.

4.1. Inference on Aggregate Effects

Here we outline a simple permutation procedure that employs 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 (7) and (9). 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-intervetion period $\mathcal{T}_1 \subseteq \{T_0+1,\ldots,T\}$ and a pre-intervention period $\mathcal{T}_0 \subseteq \{1,\ldots,T_0\}$,

$$\sum_{t\in\mathcal{T}_0}\left(\sum_{i=1}^{n_1}\widehat{\tau}_{it}(\lambda)\right)^2 \Big/ \sum_{t\in\mathcal{T}_1}\left(\sum_{i=1}^{n_1}\widehat{\tau}_{it}(\lambda)\right)^2. \tag{10}$$

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

1. Compute the treatment effect estimate in the original sample $\widehat{T}(\boldsymbol{D}^{obs})$.

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 frequency across iterations of 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

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

4.2. 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 ranks of individual treatment effects in the ordered sample combining the $n_1 \times (B+1)$ unit-level treatment effects for the actual assignments and B random permutations. 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,

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

or unit-level versions of the mean squared prediction error ratio in equation (10). The test is implemented as follows:

1. Compute unit-level measures treatment effects 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 treatment effects $\widehat{T}_i(\boldsymbol{D}^{(b)})$ for the treated. Denote these estimates $\widehat{T}_1^{(b)}, \ldots, \widehat{T}_{n_1}^{(b)}$ (in arbitrary order).

3. Calculate the ranks $R_1, \ldots, R_{n_1},\ R_1^{(1)}, \ldots, R_{n_1}^{(1)},\ \ldots,\ R_1^{(B)}, \ldots, R_{n_1}^{(B)}$ associated to the $n_1 \times (B+1)$ individual treatment effect estimates $\widehat{T}_1, \ldots, \widehat{T}_{n_1}, \widehat{T}_1^{(1)}, \ldots, \widehat{T}_{n_1}^{(1)}, \ldots, \widehat{T}_1^{(B)}, \ldots, \widehat{T}_{n_1}^{(B)}$ (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:

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

5. Monte Carlo Experiment

We report the results of a Monte Carlo experiment that investigates the finite sample 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.

Irrespective of the treatment status, the outcome at time $t \in \{1, 2\}$ is generated by $Y_{i,t} = \left(\sum_{j=1}^{p} X_{i,j}^{r}\right)/\beta + \varepsilon_{i,t}$ with r a positive real governing the degree of linearity of the outcome function. For any t, $\varepsilon_{i,t} \perp\!\!\!\perp X_i$ and $\varepsilon_{i,t} \sim \mathcal{N}(0, 1)$. For the n_1 treated units, X_i, is a vector of dimension p with iid entries distributed as $\mathcal{U}[a, b]$, with $a, b > 0$. For the n_0 control units, X_i is a vector of the same dimension with iid entries distributed according to density $\frac{2x}{b-a+2h}\mathbf{1}\{a - h < x < b + h\}$ with $h > 0$, *i.e.* each entry has the same distribution as $\sqrt{U}$ where $U \sim \mathcal{U}[a - h, b + h]$. The larger the parameter h, the smaller the support of the treated relative to that of the controls. β is a constant set so that $\beta^2 = \mathrm{V}\left(\sum_{j=1}^{p} X_{i,j}^{r} | D_i = 1\right)$

$$\beta = \sqrt{p}\sqrt{\frac{b^{2r+1} - a^{2r+1}}{(b-a)(2r+1)} - \left(\frac{b^{r+1} - a^{r+1}}{(b-a)(r+1)}\right)^2}.$$

As a consequence $\mathrm{V}(Y_{i,t}|D_i = 1) = 2$ and the signal-to-noise ratio for the treated is equal to one.

We compare the performances of synthetic control and matching estimators. Both procedures have a tuning parameter, λ for the synthetic control and the number of neighbors M for matching. We will consider these two estimators with a fixed and a data-driven choices of tuning parameter. Under the fixed procedure, $\lambda \to 0$ in 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" case that reproduces the treated as closely as possible by taking a convex combination of the vertices of the Delaunay triangle that contains the treated, when possible. Its computation is based on the remark following Theorem 2. It is not to be confused with the non-penalized synthetic control which does not take into account the compound discrepancy, for which we also report results. The data-driven choice of tuning parameter uses the first period outcome to minimize the MSE over that period. At each simulation step, both λ and M are chosen so as to minimize their respective criteria

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

and

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

where $\mathcal{J}_M(i)$ is the set of indices of the M control units that are the nearest to treated unit i. We also report a bias-corrected version of the estimators as in Section 2.3, based on a XXX specification.

The degree of the outcome function, r, is the key parameter governing the relative performances of the candidate estimators. When $r = 1$, the outcome function is linear, which suggests emphasizing the "synthetic control" part in (6) at the cost of a larger component-wise discrepancy, so we expect the pure synthetic control to perform well, while the 1-to-1 matching should do relatively worse, exhibiting a notably larger variance. As a consequence, the data-driven λ is likely to be small and the penalized synthetic control is likely to perform similarly as the pure synthetic control. As r increases, the synthetic control with $\lambda \to 0$ should suffer from a larger interpolation bias, while the performance of the 1-to-1 matching should improve. As a consequence, the data-driven λ is expected to increase so that the penalized synthetic control strikes a favorable compromise in the bias-variance trade-off.

Results are reported in Tables 1, 2 and 3 for $n_0 = 20, 40, 100$ respectively. For each configuration and each estimator $\hat{\tau}$, we report four statistics computed on the treated sample in the second period. The first is the individual-level Root Mean Squared Error (RMSE) defined as

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

the second is the aggregate-level RMSE

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

the third is the aggregate absolute 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 sparsity defined as the average number of control units used in the match to a given treated unit, *i.e.* number of non-zero entries in $W_i^*(\lambda)$ or number of matches in the optimized matching procedure.

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

	$r=1$				$r=1.2$				$r=1.4$				$r=1.8$			
	RMSE indiv	RMSE	|Bias|	Sparsity	RMSE indiv	RMSE	|Bias|	Sparsity	RMSE indiv	RMSE	|Bias|	Sparsity	RMSE indiv	RMSE	|Bias|	Sparsity
								$p=2$								
Pen. Synth.	1.3659	0.6008	0.2090	2.6892	1.3601	0.5980	0.2092	2.6864	1.3545	0.5955	0.2094	2.6882	1.3495	0.5930	0.2128	2.6716
Unpen. Synth.	1.2953	0.6297	0.2084	11.7805	1.2921	0.6353	0.2367	11.7805	1.2914	0.6430	0.2646	11.7805	1.2967	0.6639	0.3194	11.7805
Pure Synth.	1.3437	0.6008	0.2009	2.5126	1.3395	0.6001	0.2062	2.5126	1.3364	0.5998	0.2113	2.5126	1.3327	0.6007	0.2210	2.5126
Matching	1.5280	0.6368	0.2357		1.5233	0.6330	0.2293		1.5198	0.6295	0.2229		1.5161	0.6240	0.2112	
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	1.3408	0.6615	0.3962	4.3960
Pen. Synth. (BC)	1.4650	0.6700	0.0139		1.4642	0.6701	0.0106		1.4626	0.6698	0.0085		1.4615	0.6680	0.0103	
Unpen. Synth. (BC)	1.4017	0.6669	0.0178		1.4015	0.6668	0.0159		1.4015	0.6667	0.0149		1.4015	0.6668	0.0158	
Pure Synth. (BC)	1.4435	0.6637	0.0127		1.4433	0.6635	0.0104		1.4432	0.6634	0.0093		1.4433	0.6635	0.0104	
Matching (BC)	1.5748	0.6932	0.0163		1.5746	0.6930	0.0133		1.5745	0.6930	0.0118		1.5746	0.6930	0.0133	
Opt. Matching (BC)	1.3972	0.6436	0.0086		1.3977	0.6434	0.0050		1.3982	0.6432	0.0043		1.3987	0.6412	0.0074	
								$p=4$								
Pen. Synth.	1.4728	0.8098	0.5717	3.0254	1.4747	0.8214	0.5870	3.0210	1.4767	0.8343	0.6025	3.0070	1.4841	0.8557	0.6290	2.9891
Unpen. Synth.	1.4453	0.8016	0.5590	4.9822	1.4482	0.8199	0.5860	4.9822	1.4529	0.8387	0.6120	4.9822	1.4674	0.8779	0.6624	4.9822
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	1.4683	0.8696	0.6510	3.3532
Matching	1.7010	0.8992	0.6264		1.6989	0.8994	0.6240		1.6980	0.8996	0.6212		1.7003	0.9009	0.6156	
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	1.5746	0.9756	0.7701	3.4250
Pen. Synth. (BC)	1.7718	0.8883	0.0351		1.7725	0.8883	0.0390		1.7726	0.8874	0.0410		1.7715	0.8855	0.0387	
Unpen. Synth. (BC)	1.7607	0.8909	0.0394		1.7612	0.8915	0.0431		1.7615	0.8919	0.0451		1.7612	0.8916	0.0433	
Pure Synth. (BC)	1.7636	0.8906	0.0394		1.7641	0.8912	0.0431		1.7644	0.8915	0.0451		1.7641	0.8912	0.0433	
Matching (BC)	1.8526	0.9013	0.0348		1.8531	0.9020	0.0391		1.8534	0.9024	0.0414		1.8531	0.9020	0.0392	
Opt. Matching (BC)	1.7614	0.8816	0.0483		1.7626	0.8815	0.0523		1.7647	0.8827	0.0563		1.7639	0.8813	0.0555	
								$p=6$								
Pen. Synth.	1.6739	1.0840	0.9056	3.4314	1.6838	1.1053	0.9296	3.4012	1.6931	1.1244	0.9516	3.3716	1.7130	1.1572	0.9867	3.3012
Unpen Synth.	1.6440	1.0731	0.9006	4.0064	1.6551	1.0998	0.9327	4.0064	1.6672	1.1259	0.9635	4.0064	1.6950	1.1778	1.0222	4.0064
Pure Synth.	1.6439	1.0733	0.9007	3.9082	1.6550	1.0998	0.9327	3.9082	1.6671	1.1259	0.9633	3.9082	1.6948	1.1775	1.0218	3.9082
Matching	1.9247	1.1643	0.9535		1.9266	1.1699	0.9584		1.9294	1.1752	0.9623		1.9382	1.1855	0.9688	
Opt. Matching	1.8273	1.2486	1.0939	3.3640	1.8314	1.2577	1.1027	3.2470	1.8336	1.2641	1.1068	3.1490	1.8371	1.2726	1.1118	3.0340
Pen. Synth. (BC)	2.5247	1.3032	0.0135		2.5246	1.3033	0.0093		2.5244	1.3028	0.0071		2.5267	1.3033	0.0072	
Unpen. Synth. (BC)	2.5198	1.3017	0.0087		2.5195	1.3017	0.0045		2.5195	1.3018	0.0021		2.5196	1.3018	0.0042	
Pure Synth. (BC)	2.5199	1.3019	0.0088		2.5196	1.3019	0.0045		2.5195	1.3020	0.0022		2.5197	1.3020	0.0042	
Matching (BC)	2.5672	1.3081	0.0169		2.5669	1.3081	0.0123		2.5668	1.3081	0.0099		2.5670	1.3082	0.0121	
Opt. Matching (BC)	2.5276	1.2936	0.0065		2.5278	1.2935	0.0027		2.5268	1.2926	0.0003		2.5264	1.2917	0.0031	

Note: Results based on 1,000 replications. Parameters are: $n_0 = 20$,$n_1 = 10$,$a = .1$,$b = .9$,$h = .1$. We allow the matching procedure to choose up to 20 matches. The bias-correction is based on a quadratic specification of the regression function.

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

	$r=1$				$r=1.2$				$r=1.4$				$r=1.8$			
	RMSE indiv	RMSE	$\lvert$Bias$\rvert$	Sparsity	RMSE indiv	RMSE	$\lvert$Bias$\rvert$	Sparsity	RMSE indiv	RMSE	$\lvert$Bias$\rvert$	Sparsity	RMSE indiv	RMSE	$\lvert$Bias$\rvert$	Sparsity
									$p=2$							
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	1.2884	0.4950	0.1322	2.4441
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	1.2108	0.5640	0.2667	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	1.2643	0.4952	0.1340	2.6839
Matching	1.4757	0.5442	0.1626		1.4715	0.5410	0.1574		1.4684	0.5382	0.1523		1.4645	0.5339	0.1430	
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	1.2351	0.5367	0.3064	6.0650
Pen. Synth. (BC)	1.2940	0.4877	0.0097		1.2951	0.4879	0.0091		1.2958	0.4894	0.0091		1.2955	0.4904	0.0102	
Unpen. Synth. (BC)	1.2015	0.4943	0.0029		1.2017	0.4945	0.0045		1.2019	0.4946	0.0050		1.2017	0.4945	0.0040	
Pure Synth. (BC)	1.2732	0.4909	0.0088		1.2733	0.4911	0.0087		1.2734	0.4912	0.0085		1.2733	0.4911	0.0085	
Matching (BC)	1.4396	0.5159	0.0182		1.4396	0.5159	0.0167		1.4397	0.5159	0.0159		1.4396	0.5159	0.0166	
Opt. Matching (BC)	1.1978	0.4521	0.0007		1.1937	0.4506	0.0015		1.1930	0.4491	0.0028		1.1923	0.4475	0.0052	
									$p=4$							
Pen. Synth.	1.3770	0.6609	0.4131	3.5888	1.3767	0.6694	0.4297	3.5829	1.3786	0.6800	0.4465	3.5764	1.3845	0.6967	0.4739	3.5418
Unpen. Synth.	1.3373	0.6500	0.4034	10.9695	1.3397	0.6704	0.4395	10.9695	1.3446	0.6920	0.4744	10.9695	1.3612	0.7383	0.5417	10.9695
Pure Synth.	1.3470	0.6496	0.4024	3.7335	1.3489	0.6646	0.4293	3.7335	1.3524	0.6801	0.4550	3.7335	1.3637	0.7127	0.5042	3.7335
Matching	1.6240	0.7609	0.4792		1.6202	0.7572	0.4749		1.6177	0.7537	0.4704		1.6168	0.7482	0.4619	
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	1.4522	0.8282	0.6655	4.4880
Pen. Synth. (BC)	1.3283	0.5526	0.0151		1.3295	0.5539	0.0164		1.3299	0.5549	0.0174		1.3312	0.5540	0.0180	
Unpen. Synth. (BC)	1.2977	0.5446	0.0102		1.2979	0.5448	0.0109		1.2980	0.5449	0.0114		1.2979	0.5448	0.0112	
Pure Synth. (BC)	1.3055	0.5475	0.0112		1.3056	0.5477	0.0119		1.3058	0.5478	0.0125		1.3056	0.5476	0.0123	
Matching (BC)	1.4811	0.5918	0.0176		1.4813	0.5920	0.0198		1.4814	0.5922	0.0211		1.4813	0.5920	0.0201	
Opt. Matching (BC)	1.2979	0.5353	0.0028		1.2976	0.5360	0.0054		1.2977	0.5361	0.0067		1.2953	0.5375	0.0050	
									$p=6$							
Pen. Synth.	1.5532	0.9201	0.7561	3.9598	1.5614	0.9407	0.7799	3.9287	1.5703	0.9596	0.8026	3.8941	1.5884	0.9941	0.8401	3.8165
Unpen Synth.	1.5258	0.9096	0.7506	5.5584	1.5366	0.9396	0.7873	5.5584	1.5489	0.9694	0.8227	5.5584	1.5779	1.0284	0.8906	5.5584
Pure Synth.	1.5260	0.9088	0.7496	4.4455	1.5366	0.9379	0.7852	4.4455	1.5487	0.9667	0.8195	4.4455	1.5768	1.0238	0.8852	4.4455
Matching	1.8304	1.0225	0.8268		1.8303	1.0245	0.8281		1.8313	1.0263	0.8288		1.8372	1.0305	0.8298	
Opt. Matching	1.6899	1.1015	0.9652	4.0070	1.6892	1.1046	0.9677	3.9500	1.6862	1.1048	0.9655	3.8870	1.6874	1.1095	0.9685	3.7430
Pen. Synth. (BC)	1.4268	0.6225	0.0308		1.4273	0.6234	0.0285		1.4293	0.6249	0.0270		1.4331	0.6260	0.0268	
Unpen. Synth. (BC)	1.4087	0.6166	0.0313		1.4087	0.6167	0.0302		1.4087	0.6168	0.0293		1.4087	0.6167	0.0298	
Pure Synth. (BC)	1.4093	0.6166	0.0307		1.4093	0.6167	0.0295		1.4093	0.6168	0.0287		1.4093	0.6167	0.0291	
Matching (BC)	1.5824	0.6621	0.0284		1.5823	0.6621	0.0262		1.5823	0.6620	0.0249		1.5823	0.6620	0.0258	
Opt. Matching (BC)	1.4287	0.6186	0.0155		1.4291	0.6169	0.0129		1.4293	0.6184	0.0098		1.4294	0.6196	0.0121	

Note: Results based on 1,000 replications. Parameters are: $n_0 = 20, n_1 = 10, a = .1, b = .9, h = .1$. We allow the matching procedure to choose up to 20 matches. The bias-correction is based on a quadratic specification of the regression function.

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

	$r=1$				$r=1.2$				$r=1.4$				$r=1.8$			
	RMSE indiv	RMSE	|Bias|	Sparsity	RMSE indiv	RMSE	|Bias|	Sparsity	RMSE indiv	RMSE	|Bias|	Sparsity	RMSE indiv	RMSE	|Bias|	Sparsity
								$p=2$								
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	1.2684	0.4372	0.0730	2.5786
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	1.1825	0.5359	0.2932	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	0.4298	0.0809	2.8435
Matching	1.4454	0.4905	0.0816		1.4430	0.4895	0.0786		1.4414	0.4887	0.0758		1.4394	0.4878	0.0708	
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	1.1529	0.4481	0.2291	8.7890
Pen. Synth. (BC)	1.2689	0.4325	0.0239		1.2689	0.4321	0.0240		1.2686	0.4329	0.0233		1.2684	0.4329	0.0221	
Unpen. Synth. (BC)	1.1391	0.4268	0.0240		1.1395	0.4274	0.0302		1.1397	0.4277	0.0326		1.1394	0.4273	0.0289	
Pure Synth. (BC)	1.2440	0.4236	0.0273		1.2442	0.4238	0.0283		1.2442	0.4239	0.0286		1.2441	0.4238	0.0280	
Matching (BC)	1.4298	0.4816	0.0163		1.4299	0.4816	0.0157		1.4299	0.4817	0.0153		1.4299	0.4816	0.0156	
Opt. Matching (BC)	1.1285	0.3815	0.0249		1.1282	0.3804	0.0221		1.1275	0.3815	0.0209		1.1239	0.3805	0.0242	
								$p=4$								
Pen. Synth.	1.2868	0.5277	0.2525	3.7105	1.2821	0.5321	0.2683	3.7108	1.2809	0.5377	0.2818	3.7048	1.2828	0.5488	0.3072	3.6746
Unpen. Synth.	1.2207	0.5049	0.2251	42.3364	1.2223	0.5258	0.2723	42.3365	1.2273	0.5495	0.3177	42.3365	1.2456	0.6030	0.4042	42.3365
Pure Synth.	1.2469	0.5113	0.2326	4.1567	1.2477	0.5232	0.2597	4.1567	1.2499	0.5358	0.2854	4.1567	1.2579	0.5628	0.3339	4.1567
Matching	1.5503	0.6203	0.3304		1.5460	0.6166	0.3248		1.5431	0.6134	0.3194		1.5415	0.6087	0.3102	
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	1.3126	0.6553	0.4785	5.7070
Pen. Synth. (BC)	1.2362	0.4595	0.0220		1.2334	0.4587	0.0217		1.2331	0.4590	0.0230		1.2343	0.4585	0.0234	
Unpen. Synth. (BC)	1.1723	0.4408	0.0303		1.1722	0.4408	0.0287		1.1722	0.4408	0.0282		1.1723	0.4408	0.0293	
Pure Synth. (BC)	1.1997	0.4499	0.0208		1.1996	0.4499	0.0197		1.1996	0.4499	0.0195		1.1996	0.4499	0.0203	
Matching (BC)	1.4397	0.5175	0.0189		1.4397	0.5176	0.0205		1.4398	0.5176	0.0215		1.4397	0.5176	0.0208	
Opt. Matching (BC)	1.1788	0.4303	0.0274		1.1784	0.4320	0.0276		1.1782	0.4335	0.0299		1.1792	0.4342	0.0291	
								$p=6$								
Pen. Synth.	1.3851	0.7141	0.5295	4.5109	1.3902	0.7331	0.5582	4.4799	1.3968	0.7516	0.5836	4.4481	1.4150	0.7856	0.6276	4.3590
Unpen Synth.	1.3590	0.6955	0.5093	15.4066	1.3674	0.7274	0.5539	15.4066	1.3782	0.7597	0.5968	15.4066	1.4064	0.8255	0.6793	15.4066
Pure Synth.	1.3623	0.6961	0.5080	5.0435	1.3703	0.7248	0.5480	5.0435	1.3802	0.7535	0.5864	5.0435	1.4053	0.8117	0.6600	5.0435
Matching	1.6925	0.8442	0.6350		1.6903	0.8424	0.6327		1.6895	0.8408	0.6303		1.6923	0.8392	0.6262	
Opt. Matching	1.5349	0.9168	0.7874	4.7030	1.5319	0.9158	0.7877	4.6830	1.5257	0.9134	0.7854	4.6810	1.5204	0.9135	0.7859	4.6750
Pen. Synth. (BC)	1.2275	0.4716	0.0088		1.2280	0.4699	0.0091		1.2291	0.4701	0.0102		1.2333	0.4687	0.0109	
Unpen. Synth. (BC)	1.2059	0.4644	0.0071		1.2058	0.4646	0.0064		1.2058	0.4646	0.0064		1.2058	0.4645	0.0070	
Pure Synth. (BC)	1.2100	0.4671	0.0084		1.2100	0.4672	0.0077		1.2100	0.4672	0.0078		1.2100	0.4671	0.0083	
Matching (BC)	1.4352	0.5163	0.0113		1.4352	0.5162	0.0128		1.4352	0.5162	0.0138		1.4352	0.5162	0.0132	
Opt. Matching (BC)	1.2278	0.4536	0.0130		1.2290	0.4534	0.0137		1.2262	0.4526	0.0174		1.2236	0.4523	0.0165	

Note: Results based on 1,000 replications. Parameters are: $n_0 = 100$,$n_1 = 10$,$a = .1$,$b = .9$,$h = .1$. We allow the matching procedure to choose up to 20 matches. The bias-correction is based on a quadratic specification of the regression function.

Table 4: Extended Monte-Carlo Simulations I, $n_0 = 20$

	$r=1$ RMSE indiv	RMSE	|Bias|	Sparsity	$r=1.2$ RMSE indiv	RMSE	|Bias|	Sparsity	$r=1.4$ RMSE indiv	RMSE	|Bias|	Sparsity	$r=1.8$ RMSE indiv	RMSE	|Bias|	Sparsity	$r=2$ RMSE indiv	RMSE
$p=2$, average nb. 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	1.3509	0.5931	0.2136	2.2793	1.3493	0.5919
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	1.2967	0.6639	0.3194	11.7803	1.3023	0.6767
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	1.3327	0.6007	0.2210	2.5124	1.3319	0.6017
Matching	1.5280	0.6368	0.2357		1.5233	0.6330	0.2293		1.5198	0.6295	0.2229		1.5161	0.6240	0.2112		1.5157	0.6219
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	1.3408	0.6615	0.3962	4.3960	1.3381	0.6591
Pen. Synth. (BC)	1.3261	0.5714	0.0045		1.3263	0.5713	0.0061		1.3281	0.5715	0.0154		1.3371	0.5743	0.0326		1.3441	0.5773
Unpen. Synth. (BC)	1.2535	0.5968	0.0140		1.2554	0.5970	0.0320		1.2608	0.5995	0.0497		1.2802	0.6105	0.0852		1.2935	0.6183
Pure Synth. (BC)	1.3034	0.5708	0.0065		1.3041	0.5700	0.0015		1.3068	0.5704	0.0036		1.3167	0.5739	0.0132		1.3234	0.5766
Matching (BC)	1.4492	0.5948	0.0031		1.4504	0.5948	0.0148		1.4539	0.5962	0.0324		1.4662	0.6025	0.0658		1.4744	0.6071
Opt. Matching (BC)	1.2271	0.5285	0.0063		1.2293	0.5292	0.0303		1.2359	0.5319	0.0531		1.2546	0.5432	0.0976		1.2676	0.5505
$p=4$, average nb. 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	1.4842	0.8558	0.6290	2.9558	1.4902	0.8659
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	1.4674	0.8779	0.6624	4.9819	1.4771	0.8983
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	1.4683	0.8696	0.6510	3.3532	1.4768	0.8877
Matching	1.7010	0.8992	0.6264		1.6989	0.8994	0.6240		1.6980	0.8996	0.6212		1.7003	0.9009	0.6156		1.7033	0.9022
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	1.5746	0.9756	0.7701	3.4250	1.5763	0.9789
Pen. Synth. (BC)	1.3051	0.6096	0.0078		1.3096	0.6132	0.0197		1.3180	0.6204	0.0317		1.3424	0.6372	0.0564		1.3601	0.6493
Unpen. Synth. (BC)	1.2897	0.6168	0.0151		1.2938	0.6211	0.0157		1.3024	0.6276	0.0167		1.3305	0.6461	0.0181		1.3491	0.6574
Pure Synth. (BC)	1.2927	0.6156	0.0141		1.2969	0.6200	0.0178		1.3050	0.6266	0.0219		1.3315	0.6448	0.0295		1.3488	0.6558
Matching (BC)	1.4455	0.6431	0.0045		1.4485	0.6468	0.0248		1.4556	0.6533	0.0540		1.4801	0.6731	0.1105		1.4967	0.6859
Opt. Matching (BC)	1.2815	0.6003	0.0132		1.2878	0.6038	0.0445		1.2990	0.6121	0.0767		1.3282	0.6372	0.1385		1.3469	0.6527
$p=6$, average nb. treated outside convex hull: 9.96																		
Pen. Synth.	1.6740	1.0841	0.9057	3.4302	1.6838	1.1053	0.9296	3.4010	1.6931	1.1244	0.9515	3.3709	1.7130	1.1572	0.9867	3.3007	1.7252	1.1732
Unpen Synth.	1.6440	1.0731	0.9006	4.0065	1.6551	1.0998	0.9327	4.0065	1.6672	1.1259	0.9635	4.0065	1.6950	1.1778	1.0222	4.0065	1.7108	1.2037
Pure Synth.	1.6439	1.0733	0.9007	3.9083	1.6550	1.0998	0.9327	3.9083	1.6671	1.1259	0.9633	3.9083	1.6948	1.1775	1.0218	3.9083	1.7104	1.2033
Matching	1.9247	1.1643	0.9535		1.9266	1.1699	0.9584		1.9294	1.1752	0.9623		1.9382	1.1855	0.9688		1.9444	1.1910
Opt. Matching	1.8273	1.2486	1.0939	3.3640	1.8314	1.2577	1.1027	3.2470	1.8336	1.2641	1.1068	3.1490	1.8371	1.2726	1.1118	3.0340	1.8389	1.2793
Pen. Synth. (BC)	1.4342	0.7237	0.0182		1.4344	0.7233	0.0014		1.4395	0.7254	0.0168		1.4632	0.7365	0.0594		1.4811	0.7472
Unpen. Synth. (BC)	1.4126	0.7166	0.0108		1.4119	0.7157	0.0002		1.4160	0.7174	0.0108		1.4367	0.7273	0.0324		1.4525	0.7351
Pure Synth. (BC)	1.4125	0.7166	0.0109		1.4118	0.7157	0.0002		1.4158	0.7174	0.0110		1.4363	0.7272	0.0329		1.4521	0.7350
Matching (BC)	1.5638	0.7498	0.0181		1.5639	0.7496	0.0162		1.5689	0.7530	0.0507		1.5917	0.7693	0.1183		1.6087	0.7815
Opt. Matching (BC)	1.4368	0.7190	0.0099		1.4379	0.7198	0.0224		1.4412	0.7226	0.0577		1.4641	0.7397	0.1260		1.4813	0.7529
$p=8$, average nb. 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	1.9256	1.4105	1.2737	3.4969	1.9428	1.4326
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	1.9182	1.4514	1.3303	4.3612	1.9426	1.4846
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	1.9182	1.4514	1.3303	4.3601	1.9426	1.4846
Matching	2.0989	1.3945	1.2228		2.1094	1.4080	1.2361		2.1204	1.4206	1.2480		2.1445	1.4448	1.2697		2.1580	1.4569
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	2.0476	1.5395	1.4063	2.8230	2.0575	1.5491
Pen. Synth. (BC)	1.5372	0.8046	0.0167		1.5440	0.8090	0.0405		1.5556	0.8176	0.0670		1.5920	0.8414	0.1161		1.6139	0.8557
Unpen. Synth. (BC)	1.5172	0.7980	0.0117		1.5213	0.8007	0.0283		1.5303	0.8062	0.0451		1.5613	0.8242	0.0776		1.5824	0.8363
Pure Synth. (BC)	1.5172	0.7980	0.0116		1.5213	0.8008	0.0282		1.5304	0.8062	0.0451		1.5614	0.8243	0.0776		1.5824	0.8363
Matching (BC)	1.6450	0.8274	0.0310		1.6501	0.8320	0.0665		1.6604	0.8402	0.1021		1.6948	0.8665	0.1715		1.7179	0.8838
Opt. Matching (BC)	1.5497	0.8108	0.0220		1.5548	0.8142	0.0564		1.5676	0.8247	0.0927		1.6018	0.8486	0.1619		1.6244	0.8671
$p=10$, average nb. 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	2.1484	1.6903	1.5637	3.7224	2.1721	1.7179
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	2.1553	1.7387	1.6271	4.7351	2.1871	1.7789
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	2.1553	1.7387	1.6271	4.7351	2.1871	1.7789
Matching	2.2908	1.6316	1.4763		2.3075	1.6520	1.4956		2.3241	1.6710	1.5132		2.3582	1.7072	1.5455		2.3760	1.7249
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	2.2752	1.7968	1.6661	2.7540	2.2917	1.8147
Pen. Synth. (BC)	1.7845	1.0655	0.0166		1.7881	1.0679	0.0427		1.7961	1.0721	0.0714		1.8266	1.0904	0.1299		1.8489	1.1045
Unpen. Synth. (BC)	1.7705	1.0598	0.0172		1.7725	1.0604	0.0383		1.7794	1.0638	0.0596		1.8068	1.0782	0.1010		1.8264	1.0888
Pure Synth. (BC)	1.7705	1.0598	0.0172		1.7725	1.0604	0.0383		1.7794	1.0638	0.0596		1.8068	1.0782	0.1010		1.8264	1.0888
Matching (BC)	1.8667	1.0867	0.0245		1.8691	1.0881	0.0626		1.8769	1.0931	0.1008		1.9070	1.1129	0.1755		1.9285	1.1272
Opt. Matching (BC)	1.7970	1.0698	0.0192		1.8008	1.0724	0.0568		1.8103	1.0779	0.0940		1.8406	1.0982	0.1670		1.8629	1.1130
$p=12$, average nb. treated outside convex hull: 10.00																		
Pen. Synth.	2.1987	1.7540	1.6492	4.2172	2.2318	1.7964	1.6931	4.1505	2.2648	1.8359	1.7339	4.0764	2.3256	1.9050	1.8022	3.8954	2.3548	1.9381
Unpen Synth.	2.1875	1.7653	1.6651	5.0670	2.2262	1.8162	1.7183	5.0670	2.2644	1.8654	1.7692	5.0670	2.3403	1.9600	1.8663	5.0670	2.3785	2.0062
Pure Synth.	2.1875	1.7653	1.6651	5.0670	2.2262	1.8162	1.7183	5.0670	2.2644	1.8654	1.7692	5.0670	2.3403	1.9600	1.8663	5.0670	2.3785	2.0062
Matching	2.4266	1.8308	1.7058		2.4479	1.8565	1.7310		2.4688	1.8806	1.7542		2.5107	1.9260	1.7971		2.5321	1.9482
Opt. Matching	2.3532	1.8969	1.7942	3.0080	2.3753	1.9235	1.8211	2.9080	2.3960	1.9493	1.8460	2.8600	2.4331	1.9894	1.8824	2.6610	2.4555	2.0133
Pen. Synth. (BC)	2.0291	1.2078	0.0458		2.0366	1.2151	0.0729		2.0489	1.2247	0.1015		2.0895	1.2535	0.1613		2.1165	1.2730
Unpen. Synth. (BC)	2.0210	1.2071	0.0466		2.0271	1.2137	0.0700		2.0386	1.2233	0.0941		2.0765	1.2511	0.1413		2.1021	1.2689
Pure Synth. (BC)	2.0210	1.2071	0.0466		2.0271	1.2137	0.0700		2.0386	1.2233	0.0941		2.0765	1.2511	0.1413		2.1021	1.2689
Matching (BC)	2.0854	1.2109	0.0372		2.0917	1.2179	0.0741		2.1038	1.2286	0.1115		2.1436	1.2602	0.1850		2.1705	1.2805
Opt. Matching (BC)	2.0368	1.2026	0.0415		2.0454	1.2107	0.0774		2.0586	1.2228	0.1140		2.1008	1.2551	0.1874		2.1285	1.2767

6. Empirical Applications

6.1. The Value of Connections in Turbulent Times

This section revisits Acemoglu et al. (2016) on the effect of the announcement of the appointment of Tim Geithner as Treasury Secretary on November 21, 2008 on stock returns of firms that were connected to him. To choose λ we employ the pre-intervention holdout procedure of Section 3.2. The training sample uses stock returns over a 250-day window that ends 30 days prior Geithner announcement. The validation sample to select the tuning parameter λ uses returns on the following 30-day window. Abnormal returns are defined as the difference between a connected firm returns and its synthetic match returns. The measure of the announcement effect is the Cumulative Abnormal Returns (CAR) defined as the sum of abnormal returns since the announcement day.

Our methodology differs in a few ways from the original study. To mitigate complications caused by lack of uniqueness of the synthetic control estimator, Acemoglu et al. (2016) construct synthetic controls on the basis of pretreatment stock returns and restrict the units entering each synthetic control to the 20 untreated units with the highest correlation in returns with the treated unit during the training window. This is a clever ad-hoc solution to the non-uniqueness problem described in Section 1, but it does not easily generalize to contexts where synthetic controls are constructed on the basis of multiple characteristics, and leaves unaddressed the issue of how to decide on the maximum number of units that contribute to the synthetic controls. Instead, we use the full sample of control units and apply our penalized version of the synthetic control without this pre-selection step. Moreover, the original study re-weights the CAR of each treated by goodness-of-fit instead of using a simple average across the treated (see equation (7) in their paper). The authors argue that treated firms for which their corresponding synthetic unit fits better its returns over the pre-treatment period should be emphasized because they contain more information. While this assertion makes intuitive sense, especially for cases when a lack of common support prevents a particular treated unit from being well reproduced by a convex combination of control units, the properties of such an estimator are unknown and not covered in the theoretical part of our work. As a consequence, we only use a simple average. Lastly, the 10% firms most correlated with Citigroup are discarded, as in the original study, to exclude the effect of the announcement of Citigroup's bailout that could confound the effect of Geithner announcement.

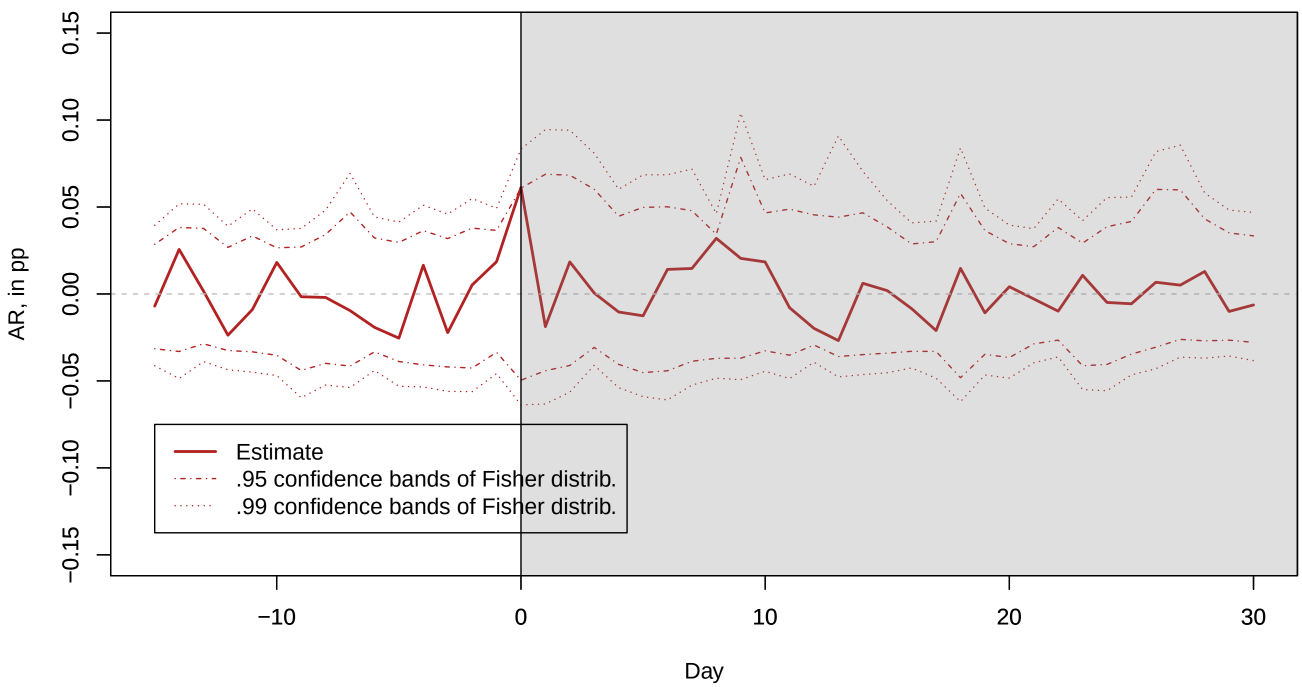
Table 5 displays our results which are qualitatively similar to the original study, albeit more muted: significance is only obtained at the 5% level in the corrected inference procedure. Figure 2 displays the Geithner announcement effect on stock returns versus the Fisher distribution under the no-treatment effect assumption. With the selected penalty level of .1, we find that the median number of active controls – defined as having a positive weight in the synthetic unit – for each treated unit is 26.7 (min: 20, max: 40) which is substantially more than in the original analysis where active controls are limited to be 20 or less. Another key difference in our inference procedure is that we recompute the cross-validated optimal λ^{opt} and corresponding synthetic control weights for every treated under the given permutation for each random permutation, as explained in Section 4. These two observations help explaining the difference between our results and the original study.

Table 5: Connections to Geithner and Reactions to Treasury Secretary Announcement, Synthetic Control Inference. Replication of Acemoglu et al. (2016)

	Estimate	Q 0.5%	Q 2.5%	Q 5 %	Q 95%	Q 97.5%	Q 99.5%
	Penalized Synthetic Control						
Day 1, CAR[0,1]	0.061**	- 0.064	- 0.050	- 0.042	0.049	0.061	0.083
	[-0.0019; 0.1250]						
Day 10, CAR[0,10]	0.138*	- 0.128	- 0.093	- 0.075	0.126	0.150	0.202
	[-0.0084; 0.2850]						
	Corrected Inference						
Day 1, CAR[0,1]	0.061**	- 0.065	- 0.049	- 0.042	0.045	0.058	0.087
Day 10, CAR[0,10]	0.138*	- 0.123	- 0.091	- 0.073	0.116	0.142	0.202
	Bias-corrected Estimator						
Day 1, CAR[0,1]	0.058	- 0.108	- 0.080	- 0.067	0.064	0.077	0.105
Day 10, CAR[0,10]	0.125	- 0.229	- 0.171	- 0.142	0.156	0.186	0.247
Cross-val. (RMSE) λ	0.08						
Mean nb. of active controls	26.7						
	Non-Penalized Synthetic Control						
Day 1, CAR[0,1]	0.060**	- 0.070	- 0.054	- 0.046	0.047	0.060	0.082
Day 10, CAR[0,10]	0.114*	- 0.155	- 0.124	- 0.108	0.094	0.119	0.171
	Corrected Inference						
Day 1, CAR[0,1]	0.060**	- 0.068	- 0.053	- 0.045	0.044	0.057	0.087
Day 10, CAR[0,10]	0.114*	- 0.165	- 0.126	- 0.111	0.084	0.114	0.171
	Bias-corrected Estimator						
Day 1, CAR[0,1]	0.058	- 0.110	- 0.082	- 0.068	0.063	0.076	0.104
Day 10, CAR[0,10]	0.119	- 0.238	- 0.178	- 0.150	0.149	0.180	0.243
Mean nb. of active controls	40.8						
Sample size (n)	525						
Nb. in treatment group (n_1)	12						

Note: This table displays Cumulative Abnormal Returns (CAR) on day 1 and 10 corresponding to panels B and C, columns 2 and 3, of Table 5 in Acemoglu et al. (2016). Results are obtained on their base sample which excludes the 10% firms whose returns are most correlated with Citigroup. We define being treated as at least one meeting between the firm and Geithner in 2007-08. The estimate column corresponds to the difference between the treated returns and synthetic control returns accumulated for the said number of days since announcement. The number between brackets are Fisher confidence intervals at 95% levels, based on 5,000 permutations. The quantiles displayed in the other columns are computed as quantiles of the Fisher distribution under the no-effect assumption. 20,000 random permutations have been used. Corrected inference discards permuted treated units for which the in-sample MSPE was three times larger than the mean MSPE for original treated units. Bias-corrected inference relies on a linear specification for the regression function. Asterisks denote significance levels (** = 5%, * = 10%).

Figure 2: Abnormal Returns after Geithner Announcement, non-corrected inference



Note: The confidence bands are computed as quantiles of the Fisher distribution under the no-effect assumption. They do not define a confidence interval for the treatment effect. When the solid red line goes out of these bands, it means the effect is significant. 20,000 permutations are used. The shadowed grey area is the post-announcement period.

Appendix

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 (6). We start by proving the upper bound in the second inequality. Since $W_i^*(\lambda)$ minimizes (6), 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, $n_1 = 1$. Program (8) is

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

where $\mathcal{W} = \{W \in [0,1]^{n_0} \,|\, W' \iota_{n_0} = 1\}$. 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, $X)'X_0$ is positive semi-definite, so f_λ is convex.

Suppose that more than one solution exist. 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, now $\lambda > 0$ implies $W_1' \Delta_1 = W_2' \Delta_1$.

Karush-Kunh-Tucker conditions imply:

$$\begin{aligned} & X_j'(X_1 - X_0 W) - \frac{\lambda}{2} \Delta_{1,j} = \pi - \gamma_j \\ & W_j \geq 0, \; W' \iota_{n_0} = 1, \; \gamma_j \geq 0, \; \gamma_j W_j = 0. \end{aligned}$$

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

$$W'X_0'(X_1 - X_0W) - \frac{\lambda}{2}W'\Delta_1 = \pi.$$

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

$$X_j'(X_1 - X_0W) - \frac{\lambda}{2}\Delta_{1,j} = \pi - \gamma_j$$

imply that the γ_j's are unique across solutions. As a result, for any solution the subvector of W formed by the rows associated with non-zero γ_j's is equal to zero. 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,

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

As an aside, notice that if $\lambda > 0$, then $\|X_1 - X_0W\| = 0$ 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 $\|X_1 - X_0W\| > 0$.

Let $Az = c$, where

$$A = \begin{pmatrix} \widetilde{X}_0'\widetilde{X}_0 & 1_{\widetilde{n}_0} \\ 1_{\widetilde{n}_0}' & 0 \end{pmatrix}, \quad z = \begin{pmatrix} \widetilde{W} \\ \pi \end{pmatrix}, \quad c = \begin{pmatrix} \widetilde{X}_0'X_1 - (\lambda/2)\widetilde{\Delta}_1 \\ 1 \end{pmatrix}.$$

Any solution to the program has to satisfy this set of linear equations, and the non-negativity constraint on $\widetilde{W}$.

Assume that any submatrix of X_0' consisting of no more than p rows has full row rank. Suppose $\widetilde{n}_0 \leq p$. Then, $\widetilde{X}_0'\widetilde{X}_0$ has full rank equal to $\widetilde{n}_0$. Augmenting $\widetilde{X}_0'\widetilde{X}_0$ by adding a column $1_{\widetilde{n}_0}$ does not change the rank, because $(\widetilde{X}_0'\widetilde{X}_0\ \ 1_{\widetilde{n}_0})$ only has $\widetilde{n}_0$ rows. Now, adding the row $(1_{\widetilde{n}_0}'\ \ 0)$ increases by one the number of linearly independent rows (because if $u \neq 0$, $u'1_{\widetilde{n}_0} = 0$, and $u'\widetilde{X}_0'\widetilde{X}_0 = 1_{\widetilde{n}_0}'$, then $u'\widetilde{X}_0'\widetilde{X}_0u = 0$, which cannot be true because $\widetilde{X}_0'\widetilde{X}_0$ is full rank). Therefore, A is full rank and there is a unique solution for z.

Now, assume $\widetilde{n}_0 = p+1$. Then, $\widetilde{X}_0'\widetilde{X}_0$ has rank p. Assume that 1_{p+1} does not belong to the column space of $\widetilde{X}_0'$. Then, 1_{p+1} does not belong to the column space of $\widetilde{X}_0'\widetilde{X}_0$ (because the columns of $\widetilde{X}_0'$ and $\widetilde{X}_0'\widetilde{X}_0$ span the same space.) As a result, the rank of $(\widetilde{X}_0'\widetilde{X}_0\ \ 1_{p+1})$ is $p+1$. Moreover, because $1_{p+1}'$ does not belong to the row space of $\widetilde{X}_0'\widetilde{X}_0$, adding the row $(1_{p+1}'\ \ 0)$ increases the rank of A by one. In this case, again, A is full rank and there is a unique solution for z.

Now, assume $\widetilde{n}_0 > p+1$. In this case, A is rank deficient and $\widetilde{X}_0'\widetilde{X}_0$ is rank deficient. By the same arguments as before $\text{rank}(A) = p+2$. Consider the rank of the augmented matrix $(A\ \ c)$

$$\text{rank}(A\ \ c) = \text{rank}\left(\begin{bmatrix} \widetilde{X}_0'\widetilde{X}_0 & 1_{\widetilde{n}_0} & \widetilde{X}_0'X_1 - (\lambda/2)\widetilde{\Delta}_1 \\ 1_{\widetilde{n}_0}' & 0 & 1 \end{bmatrix}\right).$$

Assume that any matrix composed by $p+2$ or more rows of $(\widetilde{X}_0'\ \ 1_{\widetilde{n}_0}\ \ \Delta_1)$ has full column rank. Assume also that $\lambda > 0$. Then, there is no linear combination of the columns of $(\widetilde{X}_0'\widetilde{X}_0\ \ 1_{\widetilde{n}_0})$ that is equal to $\widetilde{X}_0'X_1 - (\lambda/2)\widetilde{\Delta}_1$. As a result,

$$\text{rank}(A\ \ c) = \text{rank}(A) + 1$$

and the system has no solution. We conclude that, under the stated assumptions, if $\lambda > 0$ then there is a unique solution W with at most $p+1$ non-zero components.

In contrast, for $\lambda = 0$ it is easy to obtain examples with multiple solutions for the case $\widetilde{n}_0 > p+1$, even if any matrix composed by $p+2$ or more rows of $(\widetilde{X}_0'\ \ 1_{n_0}\ \ \Delta_1)$ has full column rank. □

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

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

$$\text{s.t.} \quad X_0 W = Z, \ \sum_{j=n_1+1}^{n} W_j = 1 \tag{A.2}$$

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

Proof of Lemma A.1

The proof of this lemma closely follows the proof of Lemma 10 in Rajan (1994) but does not rely on general position of the set of points. 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 faces of the Delaunay complex of the n_0 points $X_{n_1+1}, \dots, X_n$ in $\mathbb{R}^p$ are obtained by projecting 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 a point $\left(\sum_{j=n_1+1}^{n} W_j X_j, \sum_{j=n_1+1}^{n} W_j \|X_j\|^2\right)$ subject to the constraints in A.2. This point is equal to $\left(Z, \sum_{j=n_1+1}^{n} W_j \|X_j - Z\|^2 + \|Z\|^2\right)$ and belongs to the convex hull of $\phi(X_{n_1+1}), \dots, \phi(X_n)$. Hence, a solution of A.1 for a fixed Z is given by 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-face of the Delaunay complex. As a consequence, the only non-zero entries of $\widetilde{W}$ occur only among the vertices of the face of the Delaunay complex 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. 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 face of the Delaunay complex $DT(X_0)$ 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 an n_0-vector of weights $\widetilde{W} \in [0,1]^{n_0}$, such that *(i)* $X_0\widetilde{W} = X_0 W^*(\lambda)$, *(ii)* $\sum_{j=n_1+1}^{n} \widetilde{W}_j = 1$, *(iii)* $\widetilde{W}_j = 0$ for any j that is not a vertex of the face of the Delaunay complex containing $X_0W^*(\lambda)$, and *(iv)* $\widetilde{W}$ induces a lower compound discrepancy than $W^*(\lambda)$ relative to $X_0\widetilde{W} = X_0W^*(\lambda)$,

$$\sum_{j=2}^{n} \widetilde{W}_j \|X_j - X_0\widetilde{W}(\lambda)\|^2 < \sum_{j=2}^{n} W_j^*(\lambda)\|X_j - X_0W^*(\lambda)\|^2. \tag{A.3}$$

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

$$\sum_{j=2}^{n} W_j \|X_j - X_1\|^2 = \sum_{j=2}^{n} W_j \|X_j - X_0W\|^2 + \|X_1 - X_0W\|^2. \tag{A.4}$$

Combining equations (A.3) and (A.4) with the fact that $\|X_1 - X_0\widetilde{W}\|^2 = \|X_1 - X_0W^*(\lambda)\|^2$, 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

$$\|X_1 - X_0\widetilde{W}\|^2 + \lambda \sum_{j=2}^{n} \widetilde{W}_j \|X_j - X_1\|^2 < \|X_1 - X_0W^*(\lambda)\|^2 + \lambda \sum_{j=2}^{n} W_j^*(\lambda)\|X_j - X_1\|^2,$$

which contradicts the premise that $W^*(\lambda)$ is a solution to (6). $\square$

Lemma A.2 (Sum of Weights) *For $j = n_1+1, ..., n$, denote $S_j(\lambda) = \sum_{i=1}^{n_1} W_{i,j}^*(\lambda)$, the sum of weights given to a particular control unit across all the synthetic units. Under Assumption 1, for any $\lambda > 0$:*

1. *$\sum_{j=n_1+1}^{n} S_j(\lambda) = n_1$ almost surely,*
2. *$E[S_j(\lambda)] = n_1/n_0$ for every $j = n_1+1, ..., n$,*
3. *$\rho[S_j(\lambda), S_k(\lambda)] = -1/(n_0-1)$ for any $j \neq k$, where $\rho[S_j(\lambda), S_k(\lambda)] = Cov[S_j(\lambda), S_k(\lambda)]/V[S_j(\lambda)]$.*

Proof of Lemma A.2

The first assertion holds because each of the n_1 synthetic units is created as a convex combination of control units. The second assertion is a consequence of the previous one, the linearity of the expectation operator and exchangeability. For the third assertion, notice that the first statement of the lemma implies $\mathrm{V}[\sum_{j=n_1+1}^{n} S_j(\lambda)] = 0$ which in combination with exchangeability leads to:

$$n_0 \mathrm{V}[S_{n_1+1}(\lambda)] + n_0(n_0-1)\mathrm{Cov}[S_{n_1+1}(\lambda), S_{n_1+2}(\lambda)] = 0. \tag{A.5}$$

A consequence of equation A.5 is that $\rho[S_{n_1+1}(\lambda), S_{n_1+2}(\lambda)] = \mathrm{Cov}[S_{n_1+1}(\lambda), S_{n_1+2}(\lambda)]/\mathrm{V}[S_{n_1+1}(\lambda)] = -1/(n_0-1)$. $\square$

References

Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American Statistical Association*, 105(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.

Abadie, A. and Spiess, J. (2016). Robust post-matching inference. *Working Paper*.

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.

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

Athey, S., Bayati, M., Doudchenko, N., Imbens, G., and Khosravi, K. (2017). Matrix Completion Methods for Causal Panel Data Models. *ArXiv e-prints*.

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

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

Chernozhukov, V., Fernández-Val, I., and Melly, B. (2013). Inference on counterfactual distributions. *Econometrica*, 81(6):2205–2268.

Crump, R. K., Hotz, V. J., Imbens, G. W., and Mitnik, O. A. (2008). Nonparametric tests for treament effect heterogeneity. *Review of Economics and Statistics*, 90(3):389–405.

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.

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. (2007). Efficient semiparametric estimation of quantile treatment effects. *Econometrica*, 75(1):259–276.

Firpo, S. and Possebom, V. (2018). Synthetic control method: Inference, sensitivity analysis and confidence sets. *Working Paper*.

Firpo, S. and Ridder, G. (2008). Bounds on functionals of the distribution of treatment effects. *Working Paper*.

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.

Ho, D. E., Imai, K., King, G., and Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, 15(3):199–236.

Imbens, G. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and Statistics*, 86(1):4–29.

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.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, NY, USA.

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. (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.