WORKING PAPER

Synthetic Control Estimation Beyond Case Studies:

Does the Minimum Wage Reduce Employment?

David Powell

RAND Labor & Population

WR-1142
July, 2017
This paper series made possible by the NIA funded RAND Center for the Study of Aging (P30AG012815) and the RAND Labor and Population Unit.

RAND CORPORATION

For more information on this publication, visit www.rand.org/pubs/working_papers/WR1142.html

Published by the RAND Corporation, Santa Monica, Calif.

Synthetic Control Estimation Beyond Case Studies: Does the Minimum Wage Reduce Employment?*

David Powell
RAND †

First Version: October 2015
This Version: June 2017

Abstract

Panel data are often used in empirical work to account for fixed additive time and unit effects. The synthetic control estimator relaxes the assumption of additive effects for case studies in which a treated unit adopts a single policy. This paper generalizes the case study synthetic control estimator to estimate treatment effects for multiple discrete or continuous variables, jointly estimating the treatment effects and synthetic controls for each unit. Applying the estimator to study the disemployment effects of the minimum wage for teenagers, I estimate an elasticity of -0.44, substantially larger in magnitude than estimates generated using additive fixed effects.

Keywords: synthetic control estimation, finite inference, minimum wage, teen employment, interactive fixed effects, correlated clusters
JEL classification: C33, J21, J31, J38

*I am grateful for helpful feedback from seminar participants at several seminars. I am especially appreciate of suggestions from Abby Alpert, Tom Chang, John Donohue, Michael Dworsky, Beth Ann Griffin, Kathleen Mullen, David Neumark, Michael Robbins, Stefan Staubli, and Jeffrey Wenger.
†Contact information: dpowell@rand.org

1 Introduction

Empirical research often relies on panel data to account for fixed differences across units and common time effects, using differential changes in the policy variables to identify the relationship between policies and the outcome of interest. These additive fixed effect models include separate additive unit ("state") fixed effects and additive time fixed effects. The assumption of these models is that an unweighted average of all states acts as a valid counterfactual for each state. In this paper, I introduce a synthetic control estimator which generates an empirically-optimal control for each unit using a *weighted* average of the other units, permitting consistent estimation of treatment effects even in the presence of non-parallel trends and systematic shocks. This estimator builds on and generalizes previous work by Abadie and Gardeazabal (2003), Abadie et al. (2010), and Abadie et al. (2015).

This previous work introduced a technique to estimate causal effects for case studies and I refer to this method as the "case study synthetic control" (CSSC) estimator. The innovation of this approach was to use the period before the treatment (e.g., terrorism outbreak in Abadie and Gardeazabal (2003); the adoption of a tobacco control program in California in Abadie et al. (2010)) to construct a comparison group which is most appropriate for the group exposed to the treatment. The CSSC estimator uses a data-driven method to create a weighted average of the other units to act as a counterfactual and has been called "[a]rguably the most important innovation in the evaluation literature in the last 15 years" (Athey and Imbens (2017)). While it has become increasingly common in applied work to test for pre-existing trends through event study analysis, it is not generally clear how to proceed if this test fails. A typical motivation for the use of the synthetic control estimator is that the "parallel trends" assumption required by additive fixed effect models often does not hold in practice. CSSC permits estimation of specifications with interactive fixed effects which generate non-parallel trends, permitting causal inference even when additive fixed effect models are inappropriate. These properties should be beneficial more generally.

The estimator introduced in this paper generalizes the CSSC estimator beyond case studies, allowing for the inclusion of one or more discrete or continuous treatment variables in the model. The CSSC estimator was designed for case studies but has been applied to settings in which multiple units adopt the same policy. For such applications, CSSC requires units to be divided into ever-treated and never-treated units so that the never-treated units can potentially be used as controls for the treated units. There are many applications where this separation is difficult such as the minimum wage application of this paper. Every state has a minimum wage and these state-level minimum wages are regularly changing over time. In fixed effects models, it is unnecessary to make the distinction between ever-treated and never-treated units. This paper introduces a synthetic control equivalent. This framework nests the additive fixed effects model which is a workhorse in empirical microeconomics, permitting additional flexibility to estimate causal effects in the presence of differential state-level trends and shocks. The estimator jointly estimates the parameters associated with the treatment variables and the weights necessary to create synthetic controls for each unit. The estimator performs well in simulations, and I apply this estimator to study the employment effects of the minimum wage.

Understanding the employment ramifications of minimum wage increases is an essential part of the debate over the efficacy of federal, state, and local policies to change the minimum wage. The minimum wage rose in 19 states at the beginning of 2017 (12 due to legislative action and 7 due to automatic cost-of-living increases).[1] In 2016, California and Washington D.C. passed legislation to gradually raise the minimum wage to \$15 per hour. Arizona, Colorado, and Maine voted to increase the minimum wage to \$12 by 2020; Washington voted to increase its minimum wage to \$13.50 by that time. Minimum wage increases are common at the state-level and political momentum to increase the federal minimum wage grew during the 2016 presidential campaign, including the "Fight for \$15."

The economic literature on the relationship between the minimum wage and employment has often exploited state-level minimum wage increases to test whether states which increase the minimum wage experience different employment rate changes relative to states with stagnant minimum wages over the same time period. This additive fixed effects approach assumes that, on average, adopting states and non-adopting states would have experienced the same employment trends in the absence of the policy adoption.

Recent work has questioned the usefulness of this approach with evidence that estimated disemployment effects in the traditional fixed effects approach are not robust to the inclusion of state-specific trends (Allegretto et al. (2011)) or limiting comparisons to contiguous counties which straddle state borders (Dube et al. (2010)). In response, Neumark et al. (2014b) argues that these new approaches do not produce more appropriate comparison groups. In more recent work, Neumark et al. (2014a) states, "A central issue in estimating

[1] http://www.ncsl.org/research/labor-and-employment/state-minimum-wage-chart.aspx, last accessed February 22, 2017

the employment effects of minimum wages is the appropriate comparison group for states (or other regions) that adopt or increase the minimum wage." This problem likely resonates with many empirical applications. Synthetic control estimation offers a solution by empirically selecting appropriate control groups, and the approach introduced in this paper should have broad applicability.

The application of CSSC has proven quite useful in applied work.[2] While initially motivated by case study applications, the CSSC estimator is often used in circumstances where multiple units are treated (e.g., Billmeier and Nannicini (2013); Donohue et al. (2017); Powell et al. (2015)) by applying the CSSC estimator to each treated unit and then aggregating the estimates. In this paper, I extend the synthetic control approach beyond case studies and applications where the treatment variable is represented by a single indicator variable. I will refer to this estimator as the generalized synthetic control (GSC) estimator. GSC parallels additive fixed effects estimators which allow for multiple discrete or continuous treatment variables. However, instead of assuming the outcome variable is a function of additive state fixed effects and additive time fixed effects, the synthetic control estimator permits estimation of a specification with interactive fixed effects while nesting more traditional additive fixed effect models. Models with additive time fixed effects are equivalent to subtracting off the mean value of the outcome and the explanatory variables for each time period, giving equal weight to each state for this de-meaning. The synthetic control estimator generates a data-driven control for each state which does not constrain the weights on the control units to be equal, assigning different weights to generate an optimal control. GSC is equivalent to a *weighted* de-meaning of the outcome and treatment variables.

Moreover, GSC has advantages even in applications with a single treatment variable represented by a binary variable, including case studies. For case studies, it may be desirable or even necessary to control for time-varying covariates. The GSC estimator, by permitting multiple explanatory variables, uses control variables in a way that parallels linear panel data models – by estimating the independent relationship between each control variable and the outcome. Similarly, there are advantages in the case where multiple units implement the

[2]See Pinotti (2015); Nonnemaker et al. (2011); Hinrichs (2012); Cavallo et al. (2013); Saunders et al. (2014); Ando (2015); Bauhoff (2014); Fletcher et al. (2014); Cunningham and Shah (2014); Mideksa (2013); Nannicini and Billmeier (2011); Torsvik and Vaage (2014); Munasib and Rickman (2015); Munasib and Guettabi (2013); Eren and Ozbeklik (2016); Kreif et al. (2015); Montalvo (2011); Hackmann et al. (2015); Adhikari and Alm (2016); Adhikari et al. (2016); Zhang et al. (2016); Quast and Gonzalez (2016); Grier and Maynard (2016); Dillender et al. (2016) for a small subset of examples.

same policy. The CSSC approach presumes the existence of (several) non-adopting units. In cases where every state eventually adopts the policy of interest, CSSC will not have any states to generate a synthetic control as it necessary to divide units into "ever-treated" and "never-treated" states. Additive fixed effect estimators do not require this distinction; GSC provides the synthetic control equivalent.

This paper also introduces an inference procedure for synthetic control estimation, building on complementary work which develops an inference method for panel data in the presence of a finite number of dependent clusters (Powell (2017b)). This method has two important qualities for the GSC estimator. First, it is appropriate for fixed N, which is desirable since many panel data applications involve a relatively small number of units. Second, it permits dependence across units, which is essential in this context. The estimator uses other states as controls for each unit, mechanically inducing correlations across units. The inference procedure will adjust for this dependence to calculate appropriate p-values.

In the next section, I further discuss the CSSC estimator to motivate the need for a more general synthetic control technique and include background on the minimum wage debate in the economics literature. In Section 3, I introduce the generalized synthetic control estimator. I discuss an appropriate inference procedure in Section 4. Section 5 includes simulation results. In Section 6, I apply GSC to estimate the relationship between state minimum wages and the employment rate of teenagers. Section 7 concludes.

2 Background

2.1 Case Study Synthetic Control Estimation

This paper builds on the synthetic control estimation technique discussed in Abadie and Gardeazabal (2003), Abadie et al. (2010), and Abadie et al. (2015). In the CSSC framework, there are $J+1$ units and T time periods. Unit 1 is exposed to the treatment in periods T_0+1 to T and unexposed in periods 1 to T_0. All other units are unexposed in all time periods. Outcomes are defined by a factor model:

$$Y_{it} = \alpha_{it} D_{it} + \delta_t + \boldsymbol{\lambda}_t \boldsymbol{\mu}_i + \epsilon_{it}, \tag{1}$$

where D_{it} represents the treatment variable and is equal to 1 for state $i = 1$ and time periods $t > T_0$, 0 otherwise. $\boldsymbol{\lambda}_t$ is a $1 \times F$ vector of common unobserved factors and $\boldsymbol{\mu}_i$ is a $F \times 1$ vector of factor loadings. Gobillon and Magnac (2016) compares the synthetic control approach to panel data models with interactive fixed effects (e.g., Bai (2009)). Both approaches are useful for relaxing the restriction that the fixed effects are additive.

Unit 1 is the treated unit while all other units are part of the "donor pool," the units that may potentially compose part of the synthetic control. The purpose of the approach is to find a weighted combination of units (which exists by assumption) in the donor pool such that

$$\sum_{j=2}^{J+1} w_j^* Y_j = Y_1, \tag{2}$$

where w_j^* represents the weight on unit j. If this condition holds, then $\sum_{j=2}^{J+1} w_j^* Y_{jt}$ provides an estimate of Y_{1t} when $D_{1t} = 0$.

The weights are constrained to be non-negative and to sum to one. They are generated to minimize the difference between the pre-treated outcomes of the treated unit and the unit's synthetic control. I assume that there are no other control variables, which are included additively in equation (1) in Abadie and Gardeazabal (2003) and Abadie et al. (2010), and that all pre-treatment outcomes (and only pre-treatment outcomes) are used to generate the synthetic control weights. This is common in applications using CSSC (see Cavallo et al. (2013) for one example). Kaul et al. (2015) notes that the CSSC estimator will not use the covariates when each pre-treatment outcome is used to generate the synthetic control weights. I assume that no additional covariates are used with the CSSC estimator since GSC will permit a more standard treatment of the covariates. Due to concerns of overfitting, it is often recommended that covariates and a limited set of pre-treatment outcomes be used to create the synthetic control weights. In this paper, I assume large T such that overfitting is less likely to be a concern.[3] The GSC estimator below will use control variables in a manner that parallels traditional regression analysis by letting them have a direct effect on the outcome variable and estimating that effect jointly with the other parameters.

[3] If overfitting is a concern, the same approaches used in the CSSC literature can be applied to GSC. Namely, it is common to use only a select few pre-intervention outcomes (or the mean in the pre-period) as well as some covariates to generate the weights. For GSC, it is also possible to generate synthetic controls using only covariates and a subset of $Y_{it} - D_{it}'\mathbf{b}$ values.

Let $\mathbf{W} = (w_2, \ldots, w_{J+1})'$ such that $w_j \geq 0$ for all $2 \leq j \leq J+1$ and $\sum_{j=2}^{J+1} w_j = 1$. Let $\mathbf{X}_1 = (Y_{11}, \ldots, Y_{1T_0})$, a $(T_0 \times 1)$ vector of pre-intervention outcomes for the treated unit. $\mathbf{X}_0$ is a $(T_0 \times J)$ of the same variables for the units in the donor pool. Let $||\cdot||$ represent the Euclidian norm. The weights are estimated using

$$\mathbf{W}^* = \underset{\mathbf{W}}{\operatorname{argmin}} \, ||\mathbf{X}_1 - \mathbf{X}_0\mathbf{W}||$$

$$\text{s.t.} \quad w_j \geq 0 \quad \text{for all} \quad 2 \leq j \leq J+1 \quad \text{and} \quad \sum_{j=2}^{J+1} w_j = 1.$$

In words, the CSSC estimator involves a constrained optimization to find the weighted average of the donor states which is "closest" to the treated unit in terms of pre-treated outcomes. Because I assume that all pre-treatment outcomes are used to create the synthetic control weights, I do not consider weighting matrices when minimizing the norm in the equation above. The CSSC approach uses a two-step process in which some variables included in $\mathbf{X}$ are weighted more than others if they are better predictors of the pre-intervention outcomes. Since I use the pre-intervention outcomes themselves, this weighting matrix is unnecessary. See Kaul et al. (2015) for details. The treatment effect estimate for period $t > T_0$ is

$$\hat{\alpha}_{it} = Y_{it} - \sum_{j=2}^{J+1} w_j^* Y_{jt}. \tag{3}$$

CSSC permits estimation of a treatment effect for each post-treatment time period, though it is common to report an aggregate post-treatment effect. For inference, Abadie and Gardeazabal (2003) and Abadie et al. (2010) recommend a permutation test in which a treatment effect is estimated for each unit in the donor pool as if it were treated (and Unit 1 is included in the donor pool). Under the null hypothesis that there is no treatment effect, these placebo tests generate a distribution of estimates which would be observed randomly. The placement of $\hat{\alpha}_{it}$ in this distribution generates an appropriate p-value. Ando and Sävje (2013) discusses possible limitations of the placebo test approach for synthetic control estimation. Alternate approaches and test statistics have been mentioned in the literature (see Firpo and Possebom (2016) for an extensive discussion).

CSSC has proven valuable in applied work as a means of relaxing the parallel trends assumption required by additive fixed effect models. The framework of equation (1) nests the more typical additive fixed effect model which assumes

$$\boldsymbol{\lambda}_t = \begin{bmatrix} 1 & \phi_t \end{bmatrix} \quad \text{and} \quad \boldsymbol{\mu}_i = \begin{bmatrix} \gamma_i \\ 1 \end{bmatrix}, \tag{4}$$

such that $\boldsymbol{\lambda}_t \boldsymbol{\mu}_i = \gamma_i + \phi_t$. A primary contribution of this paper is that the motivation behind CSSC should apply more generally to panel data models, not just case studies.

2.2 Finite Inference

Panel data models are often applied in cases where the number of units is relatively small. The minimum wage analysis of this paper uses 50 states plus Washington D.C. The donor pool in Abadie et al. (2010) includes 38 states; Abadie and Gardeazabal (2003) uses 16 Spanish regions; Abadie et al. (2015) uses a sample of 16 OECD countries. Two notable exceptions in the CSSC context are Robbins et al. (2016), which studies a crime intervention policy implemented by city block and includes a total of 3,601 blocks in the data, and Ando (2015) with over 800 units in the donor pool.

In this paper, I apply an inference procedure for fixed N, which is developed more generally for panel data with a fixed number of units. This method is introduced in Powell (2017b) and applied here for the synthetic control estimator as a special case. Additive fixed effects create problems when N is small because they induce mechanical correlations across clusters. Inference procedures developed for a small number of clusters are frequently not valid for models with state and year fixed effects (see discussions in Hansen (2007); Bester et al. (2011); Ibragimov and Müller (2010)). In linear models, the inclusion of time fixed effects is equivalent to de-meaning the variables for each time period, creating a correlation across all units of order $O(\frac{1}{N})$.

The synthetic control estimator introduced in this paper also generates dependence across clusters as the synthetic control for each state is a weighted average of other states. Consequently, it is difficult to consider each unit as independent of the other units. Consider a case where unit 2 receives a large weight as part of the synthetic control for unit 1 and unit 1 receives a large weight as part of the synthetic control for unit 2. The estimator subtracts off the synthetic control value of the outcome and the treatment variables for each unit, and it is thus unlikely that de-meaned units 1 and 2 should be considered independent when testing a statistical hypothesis.

I apply an inference procedure which accounts for empirical correlations across units. These correlations can also arise for reasons other than the estimation procedure, such as similarities across states due to geography, political institutions, industry composition, or even similarities on unobserved dimensions. The inference procedure accounts for this dependence regardless of the source and there are no restrictions on the magnitudes of the correlations across units.

2.3 Minimum Wage and Employment

A vast literature has debated whether minimum wage increases affect employment rates (see Neumark et al. (2014b) for a review). The empirical literature has recently focused on determining appropriate controls for states with increasing minimum wages. State minimum wages increase due to state legislative changes, binding federal minimum wage changes, or cost-of-living adjustments. The literature typically uses all sources of variation while controlling for fixed differences across states and common time effects.[4] The potential of the CSSC estimator to generate appropriate control units has been recognized recently in the literature, though it it difficult to appropriately apply the CSSC method to this application. I briefly discuss prior uses of the CSSC estimator found in the literature to study the possible disemployment effects of the minimum wage.

First, Sabia et al. (2012) studies the employment effects of the 2004-2006 New York minimum wage increase, finding large employment reductions. They use CSSC estimation for part of their analysis, though they do not use any pre-intervention outcomes to construct the synthetic control. Instead, the synthetic control is based only on other covariates.

More recently, Dube and Zipperer (2015) (also discussed in Allegretto et al. (2017)) uses CSSC estimation to study minimum wage employment effects. To implement CSSC, the authors select on states with no minimum wage changes for two years prior to a minimum wage increase and with at least one year of post-treatment data. Further selection criteria are also necessary and limit the sample to the study of 29 of the 215 state-level minimum wage increases over their time period (their calculation). The donor pools for each of the treated states are limited to states with no minimum wage changes over the same time period and occasionally consist of relatively few states. Furthermore, the synthetic controls

[4] Recent work (Clemens and Wither (2014); Clemens (2015)) has exploited the differential effects of federal minimum wage increases on states in which the increases were binding relative to states in which they were not.

are generated for a relatively short time period around the time of adoption of the treated state. These selection criteria are unnecessary using the GSC estimator introduced below as the estimator will include all states and time periods and use all available variation, paralleling additive fixed effects estimators.

Moreover, Dube and Zipperer (2015) creates the synthetic controls for each adopting state based on pre-intervention outcomes. These outcomes, however, are themselves treated given that the minimum wage (potentially) causally affects employment. Consequently, the synthetic control for each treated state using the CSSC estimator is, in fact, only valid if the minimum wage does not affect employment and, otherwise, does not estimate the model represented in equation (1). The GSC estimator introduced below accounts for the causal impact of the treatment variables when creating the synthetic controls for each state.

Overall, it is difficult to use the traditional CSSC estimator to study the effects of the minimum wage. The minimum wage is constantly changing across states, making it difficult to isolate "treated" states and "control" states. More generally, this problem arises whenever there are multiple treatments, multiple treated units, treatment variables that are continuous, etc. In a fixed effects specification, the designation of units as "treated" and "control" is unnecessary. This paper introduces a synthetic control equivalent.

3 Generalized Synthetic Control Estimator

This section develops the GSC estimator. It permits the inclusion of multiple treatment variables, which can be discrete or continuous, in the specification. I follow the notation developed for the CSSC method and used in Section 2.1 as closely as possible. I assume that there are (small) N units and (large) T time periods. I model outcomes as

$$Y_{it} = \mathbf{D}'_{it}\boldsymbol{\alpha}_{\mathbf{0}} + \delta_t + \boldsymbol{\lambda}_t\boldsymbol{\mu}_i + \epsilon_{it}, \tag{5}$$

where $\mathbf{D}_{it}$ is a $K \times 1$ vector of treatment variables for state i at time period t. As before, $\boldsymbol{\lambda}_t$ is a $1 \times F$ vector of common unobserved factors and $\boldsymbol{\mu}_i$ is an $F \times 1$ vector of factor loadings. The value of including the $\boldsymbol{\lambda}_t\boldsymbol{\mu}_i$ term corresponds to the benefits discussed for the CSSC method. This term permits flexible, non-linear state-level trends and shocks that may be correlated with the policy variables. Similar to the CSSC estimator, the additive fixed effects equivalent to equation (5) is nested in this framework. Equation (5) also nests specifications which include parameterized unit-specific trends (e.g., state-specific linear trends). While a state-time interaction term would be collinear with $\mathbf{D}_{it}$, interactive fixed effects permit the inclusion of a term which varies flexibly at the state and time level. The primary difference between equation (5) and the specification represented in equation (1) is that $\mathbf{D}_{it}$ is a vector and not limited to one indicator variable. Related, the parameter $\boldsymbol{\alpha_0}$ in equation (5) does not vary by state or time (unlike the parameter in equation (1)). However, it is possible to allow heterogeneity by defining $\mathbf{D}_{it}$ as the interaction of the treatment variables with state or time indicators. In fact, equation (5) nests the specification estimated by CSSC (equation (1)).

The estimator will require simultaneous estimation of a synthetic control for each state. Let

$$\boldsymbol{w}_i = (w_i^1, \ldots, w_i^{i-1}, w_i^{i+1}, \ldots, w_i^N),$$

where $\boldsymbol{w}_i$ is the vector of weights on all other states to generate the synthetic control for state i; w_i^j represents the weight given unit j for the creation of the synthetic control for unit i. The weights are constrained, as before, to be non-negative and to sum to one. I define the set of possible weighting vectors to create the unit i synthetic control by

$$\mathcal{W}_i = \left\{ \boldsymbol{w}_i \;\middle|\; \sum_{j \neq i} w_i^j = 1, w_i^j \geq 0 \text{ for all } j \right\}.$$

3.1 Main Assumptions

In this section, I discuss the assumptions necessary for identification of $\boldsymbol{\alpha}_0$. The weights themselves are not necessarily identified as multiple sets of weights may be appropriate. This will not affect identification of the parameters of interest. Define $\mathbf{D}_i \equiv (\mathbf{D}_{i1}, \ldots, \mathbf{D}_{iT})$.

A1 ***Outcomes*****:** $Y_{it} = \mathbf{D}_{it}'\boldsymbol{\alpha}_0 + \delta_t + \boldsymbol{\lambda}_t\boldsymbol{\mu}_i + \epsilon_{it}$

A2 ***Independence*****:** There exists $\boldsymbol{w}_i \in \mathcal{W}_i$ such that

$$E\left[\boldsymbol{\lambda}_t\boldsymbol{\mu}_i + \epsilon_{it} \;\middle|\; \mathbf{D}_i\right] = \sum_{j \neq i} w_i^j \left[\boldsymbol{\lambda}_t\boldsymbol{\mu}_j + \epsilon_{jt}\right].$$

A3 ***Rank Condition*:**

$$E\begin{bmatrix} \mathbf{D}'_{i1}-\sum\limits_{j\neq i}\phi_i^j\mathbf{D}'_{j1} & \sum\limits_{j\neq i}\phi_i^j Y_{i1} \\ \vdots & \vdots \\ \mathbf{D}'_{iT}-\sum\limits_{j\neq i}\phi_i^j\mathbf{D}'_{jT} & \sum\limits_{j\neq i}\phi_i^j Y_{iT} \end{bmatrix} \text{ is rank } K+1 \text{ for all } \boldsymbol{\phi}_i \in \mathcal{W}_i.$$

A1 defines the outcome as a function of the treatment variables, a time fixed effect, (an arbitrary number of) interactive fixed effects, and an observation-specific disturbance term. The advantage of the synthetic control estimator is that it accounts for $\boldsymbol{\lambda}_t\boldsymbol{\mu}_i$, which may be arbitrarily correlated with the treatment variables. The dimensions of $\boldsymbol{\lambda}_t$ and $\boldsymbol{\mu}_i$ are both unknown and neither vector is estimated.

A2 assumes the existence of a synthetic control for each unit, equivalent to the CSSC estimator assumption (see equation (2)). **A2** assumes that there exists a weighted combination of other units which equals the conditional expected value of the interactive fixed effects plus the error term. The CSSC assumption shown in equation (2) as well as the corresponding GSC assumption **A2** may seem restrictive. Complementary work (Powell (2017a)) relaxes the CSSC assumption and the approach discussed in that paper can also easily be applied here.[5] The motivation for this paper is to extend the synthetic control method beyond case studies and the approach of this paper most directly generalizes the CSSC approach. However, as modifications are made in terms of estimating synthetic control weights or the conditions necessary to estimate synthetic control weights for CSSC, those approaches should also be applicable to GSC. Note that **A2** only assumes existence of weights and does not assume uniqueness of the synthetic control for each unit. To the extent that **A2** does not hold for specific units, these units can be excluded[6] or downweighted. The weighting procedure discussed in Section 3.2.2 suggests weighting based on synthetic control fit.

[5]Using the approach in Powell (2017a), **A2** can be replaced by $E\left[\boldsymbol{\lambda}_t\boldsymbol{\mu}_i+\epsilon_{it}\;\middle|\;\mathbf{D}_i\right]=\sum\limits_{j\neq i}w_i^j E\left[\boldsymbol{\lambda}_t\boldsymbol{\mu}_j\right]$.

[6]The units can be used to generate synthetic controls for other units but excluded as units to use in the objective function.

A3 is a rank condition which assumes $T \geq K+1$, though this requirement can be relaxed.[7] **A3** assumes independent variation in each element of $\mathbf{D}'_{it} - \sum_{j \neq i} \phi_i^j \mathbf{D}'_{jt}$ for all possible $\boldsymbol{\phi}_i \in W_i$. In addition, the treatment variables must also vary independently of the outcomes. Since the synthetic control of each unit will be estimated by using the "untreated" outcomes of the other units, it is important that variation in the treatment variables cannot be replicated by a weighted average of the outcomes. This permits independent variation in the unobserved interactive fixed effects.

3.2 Estimation

The estimator simultaneously estimates the parameters associated with each treatment variable while constructing a synthetic control for each unit. The synthetic control is an estimate of $Y_{it} - \boldsymbol{D}'_{it}\hat{\boldsymbol{\alpha}}$ using a constrained weighted average of $(Y_{jt} - \boldsymbol{D}'_{jt}\hat{\boldsymbol{\alpha}})$ for all $j \neq i$. $\sum_{j \neq i} \left(\hat{w}_i^j \left(Y_{jt} - \boldsymbol{D}'_{jt}\hat{\boldsymbol{\alpha}} \right) \right)$ acts as the control, constructed empirically to minimize the sum of the square of the residuals. I refer to $Y_{it} - \boldsymbol{D}'_{it}\hat{\boldsymbol{\alpha}}$ as the estimated "untreated outcome" for unit i, the outcome that would be observed if all treatment variables were equal to zero.

3.2.1 Implementation

The estimator chooses an estimate for $\boldsymbol{\alpha}_0$ and jointly constructs a synthetic control for each unit. The synthetic controls are constructed by selecting weights in $\mathcal{W}_i$. The estimates solve

$(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{w}}_1, \ldots, \hat{\boldsymbol{w}}_N) =$

$$\underset{\boldsymbol{b}, \boldsymbol{\phi}_1 \in \mathcal{W}_1, \ldots, \boldsymbol{\phi}_N \in \mathcal{W}_N}{\operatorname{argmin}} \left\{ \frac{1}{2NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[Y_{it} - \boldsymbol{D}'_{it}\boldsymbol{b} - \sum_{j \neq i} \left(\phi_i^j \left(Y_{jt} - \boldsymbol{D}'_{jt}\boldsymbol{b} \right) \right) \right]^2 \right\}. \tag{6}$$

$\sum_{j \neq i} \left(\hat{w}_i^j \left(Y_{jt} - \boldsymbol{D}'_{jt}\hat{\boldsymbol{\alpha}} \right) \right)$ is used as a control for unit i. It would be inappropriate to construct a synthetic control based solely on Y_{it} since Y_{it} is causally impacted by $\boldsymbol{D}_{it}$. Instead, the

[7] It is possible to include the treatment variables (and outcomes) for all units which would change this requirement to $NT \geq K+1$.

GSC estimator uses the "untreated outcomes" as possible controls by subtracting off the causal effects of the policy variables from the outcomes.

With only one or two treatment variables, the above minimization can be implemented with the following straightforward approach. First, create a grid of possible values for $\boldsymbol{\alpha}_0$ and search over that grid. For each "guess" $\boldsymbol{b}$, calculate $Y_{it} - \boldsymbol{D}'_{it}\boldsymbol{b}$ and then estimate the synthetic control for each unit i. The estimation of each synthetic control involves a constrained minimization. Then, calculate the argument in equation (6). The $\boldsymbol{b}$ that minimizes this objection function is $\hat{\boldsymbol{\alpha}}$.

This procedure assumes that grid-searching is possible. For each $\boldsymbol{b}$, one can simply use the Stata `synth` command (Abadie et al. (2014)) or R package (Abadie et al. (2011))[8] developed for the CSSC estimator to obtain a synthetic control for each state with $Y_{it} - \boldsymbol{D}'_{it}\boldsymbol{b}$ as the outcome. Other estimation methods of the synthetic control weights and the parameters of interest are also possible.

3.2.2 Weighting

With CSSC estimation, it is customary to visually check the fit of the synthetic control with the treated unit in the pre-period. It may not be possible to generate an appropriate synthetic control for the treated unit, casting doubts in such circumstances on the estimate generated by the CSSC method. For the estimator of this paper, the parallel concern is that Assumption **A2** may not hold. I employ a two-step procedure which places more weight on units with better synthetic control fits. The steps are as follows:

1. For each state i,

$$\widehat{\Omega}_i \equiv \min_{\boldsymbol{b}_i, \boldsymbol{\phi}_i \in \mathcal{W}_i} \left\{ \frac{1}{2T} \sum_{t=1}^{T} \left[Y_{it} - \boldsymbol{D}'_{it}\boldsymbol{b}_i - \sum_{j \neq i} \left(\phi_i^j \left(Y_{jt} - \boldsymbol{D}'_{jt}\boldsymbol{b}_i \right) \right) \right]^2 \right\}. \tag{7}$$

In the first step, I allow the parameters ($\boldsymbol{b}_i$) to vary when estimating the variance for each unit, though this flexibility is unnecessary.[9] The purpose of the first step is simply

[8]More information about these packages can be found at http://web.stanford.edu/jhain/synthpage.html

[9]The benefit of permitting such flexibility is that a poor synthetic control fit for one unit may affect the aggregate (preliminary) estimates for the treatment effects and, consequently, the fits of the other synthetic controls. By estimating the parameters by unit initially, one simply gets a measure of the best fit for each unit and this measure is unaffected by poor synthetic control fits for other units.

to estimate a measure of fit by state. It is possible that $\boldsymbol{\alpha}_0$ is not identified for a unit by itself. However, this will not cause problems in estimating the variance for that unit.

2. Minimize the weighted sum of squared residuals:

$(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{w}}_1, \dots, \hat{\boldsymbol{w}}_N) =$

$$\underset{\boldsymbol{b},\boldsymbol{\phi}_1\in\mathcal{W}_1,\dots,\boldsymbol{w}_N\in\mathcal{W}_N}{\operatorname{argmin}} \left\{ \frac{1}{2NT} \sum_{i=1}^{N} \widehat{\Omega}_i^{-1} \sum_{t=1}^{T} \left[Y_{it} - \boldsymbol{D}_{it}'\boldsymbol{b} - \sum_{j\neq i} \left(\phi_i^j \left(Y_{jt} - \boldsymbol{D}_{jt}'\boldsymbol{b} \right) \right) \right]^2 \right\}. \quad (8)$$

This method places more weight on units with better synthetic control fits. I employ this two-step method in the simulations.[10] Alternate weighting schemes, such as excluding[11] states with synthetic control fits below a certain threshold, are also possible.

3.2.3 Comparison to Additive Time Fixed Effects

Given estimated synthetic control weights, each de-meaned observation in the objective function can be represented by

$$\left(Y_{it} - \sum_{j\neq i} \hat{w}_i^j Y_{jt} \right) - \left(\boldsymbol{D}_{it} - \sum_{j\neq i} \hat{w}_i^j \boldsymbol{D}_{jt} \right)' \boldsymbol{b},$$

for treatment parameter values $\boldsymbol{b}$. Thus, given the estimated synthetic control weights, the optimization is equivalent to a regression of $Y_{it} - \sum_{j\neq i} \hat{w}_i^j Y_{jt}$ on $\boldsymbol{D}_{it} - \sum_{j\neq i} \hat{w}_i^j \boldsymbol{D}_{jt}$. In a model with additive time fixed effects instead of interactive fixed effects, the outcome and treatment variables are de-meaned, giving equal weight to all units such that $\hat{w}_i^j = \frac{1}{N}$ for all (i, j). The additional flexibility of GSC by not constraining the weights in this manner illustrates the potential of GSC to accommodate systematic trends and shocks that are correlated with the treatment variables.

[10] The weighting procedure makes little difference in the final estimates of the minimum wage application (see Table 3). Though not shown, the simulation results below (Tables 1 and 2) were also similar whether the two-step approach was used or not.

[11] As noted before, the excluded states can (and should) still be used to create synthetic controls for other states. They are excluded from the first summation in equation (6) (or equation (8))

3.2.4 Benefits Even With a Binary Treatment Variable

The primary motivation of the GSC estimator is to extend synthetic control estimation to applications with multiple treatment variables, multiple adopters, non-binary treatments, etc. However, it is instructive to discuss the usefulness of the GSC method in more basic cases, including those in which the CSSC estimator is typically used, such as case studies or applications with multiple adopters of the same policy. I briefly discuss four benefits of the GSC approach in cases in which the CSSC estimator is often used.

First, in the case study approach, it is often desirable to condition on additional covariates. CSSC assumes that the synthetic control weights also difference out the covariates, which may be difficult in some cases. For example, one of the control variables may not be in the convex hull of the other units, violating the CSSC assumptions. Similarly, it may be necessary to condition on a separate policy which is itself adopted by only the treated unit and a few control units. Again, it would be difficult for this control variable to be in the convex hull of the other units unless the researcher limits the donor pool only to the few states adopting this other policy. The GSC framework permits joint estimation of the effect of the treatment variable and the covariates in a manner that parallels traditional regression techniques.

Second, it is common to apply the CSSC estimator to applications where multiple units adopt the same treatment. However, there is no accepted method for aggregating the multiple estimates. The GSC estimator aggregates these estimates naturally by estimating a single parameter across all treated units. If unit-specific estimates are desired, they are still available in the GSC framework by interacting the treatment variables with unit-specific indicators.

Third, even in a traditional case study application with no covariates, CSSC and GSC have a slight but important difference. CSSC creates a synthetic control for the treated unit and estimates the treatment effect given the post-adoption difference in the outcome of the treated unit and its synthetic control. GSC estimates synthetic controls for all units. These other units may provide additional information if the treated unit is itself part of the synthetic control of one or more of the other units. In fact, consider a case where the treated unit is not in the convex hull of the control units such that CSSC cannot create an appropriate synthetic control. However, the treated unit may be part of an appropriate synthetic control for some of the control units and the treatment effect can be estimated using equation (8).[12] While I have suggested a two-step estimation procedure in this paper to downweight units with poor synthetic control fits, I have previously noted that it is also possible to exclude units with especially poor fits. In a case where the treated unit is clearly outside the convex hull, this unit can be excluded from the first summation of equation (8) and GSC will rely on the units in which the treated unit composes part of the synthetic control. An estimate of the treatment effect is then identified (assuming $w_i^1 > 0$ for some i). Alternatively, weighting by the inverse of the variance will upweight units with good synthetic control fits. The units receiving the largest weights may not be the actual treated unit, but they contain especially useful information in such cases if the treated unit is part of the synthetic control. This insight is discussed in-depth for case study applications in Powell (2017a) but applies here as well.

Fourth, when most units are treated, it is unclear how to construct the donor pools for each treated unit using the CSSC approach. For example, imagine a policy which all states adopt over the sample period but at different times. The CSSC donor pool is empty under the requirement that the donor pool consists only of untreated states. With a fixed effects model, it is unnecessary to separate the states into ever-treated states and never-treated states. With CSSC estimation, however, this distinction is required. For example, Donohue et al. (2017) studies right-to-carry (RTC) laws. Only 9 states had never passed RTC legislation by the end of their time period and the authors hypothesize that is too few to construct proper synthetic controls for each adopting state and to conduct appropriate inference. To expand the donor pool, Donohue et al. (2017) includes some adopting states in the donor pool based on the time of adoption. Similarly, Billmeier and Nannicini (2013) expands the donor pool for their application by restricting the analysis to countries with a sufficient number of countries which were not treated within 10 years of the treated country (meeting other restrictions as well).

The synthetic control estimator introduced above parallels models with additive fixed effects. Treated units are jointly used as controls, and dividing units into treatment units and control units is unnecessary. The GSC estimator combines the benefits of the CSSC model (interactive fixed effects, non-parallel trends) with the benefits of traditional panel data models (using all states as possible treated and control units).

[12] Imagine a case in which the value of the outcome variable for the treated unit is always above the values of the other units. It is impossible to create a weighted average of the other units to match the treated unit. However, the treated unit is potentially part of a synthetic control for at least one of the other units.

3.3 Identification

In the next sections, I discuss properties of the GSC estimates. Identification of $\boldsymbol{\alpha}_0$ holds under **A1-A3**:

Theorem 3.1 (Identification)**.** *If* ***A1-A3*** *hold, then* $E\left\{\left[Y_{it} - \boldsymbol{D}_{it}'\boldsymbol{b} - \sum_{j\neq i}\left(\phi_i^j\left(Y_{jt} - \boldsymbol{D}_{jt}'\boldsymbol{b}\right)\right)\right]^2\right\}$ *has a unique minimum and* $\boldsymbol{b} = \boldsymbol{\alpha_0}$ *at this minimum.*

Theorem 3.1 states that $\boldsymbol{\alpha}_0$ is unique. The weights are not necessarily unique but unique weights are not required for identification of the treatment effects. Intuitively, it is straightforward to show that $E\left[Y_{it}\middle|\mathbf{D}_i\right] = \mathbf{D}_{it}'\boldsymbol{\alpha}_0 + \sum_{j\neq i} w_i^j\left(Y_{jt} - \mathbf{D}_{jt}'\boldsymbol{\alpha}_0\right)$. This conditional expectation is unique by **A3**. I include a discussion in Appendix Section A.1.

Throughout this paper, I consider the weights which generate the synthetic controls as nuisance parameters which are necessary for consistent estimation of $\boldsymbol{\alpha}_0$. To discuss properties of the GSC estimates, however, it will be helpful to make assumptions about the synthetic untreated outcomes. For a given $\boldsymbol{b}$, define the corresponding synthetic control weights for each unit as

$$\widehat{\boldsymbol{w}}_i(\boldsymbol{b}) \equiv \underset{\boldsymbol{\phi}_i\in\mathcal{W}_i}{\operatorname{argmin}} \frac{1}{2T}\sum_{t=1}^{T}\left[Y_{it} - \boldsymbol{D}_{it}'\boldsymbol{b} - \sum_{j\neq i}\phi_i^j\left(Y_{jt} - \boldsymbol{D}_{jt}'\boldsymbol{b}\right)\right]^2. \tag{9}$$

These are the synthetic control weights given treatment effects equal to $\boldsymbol{b}$. Also, define

$$\Gamma_{it}(\boldsymbol{b}) \equiv \sum_{j\neq i}\hat{w}_i^j(\boldsymbol{b})\left[Y_{jt} - \boldsymbol{D}_{it}'\boldsymbol{b}\right],$$

The benefit of defining $\Gamma_{it}(\boldsymbol{b})$ is that it links the value of the parameter estimates to the synthetic control for each unit. The regularity conditions are more intuitive when discussing restrictions on the synthetic controls instead of the weights themselves.

3.4 Consistency

I discuss consistency for $T \to \infty$ and assume that N is fixed. Consistency of $\hat{\boldsymbol{\alpha}}$ requires additional regularity conditions.

A4 ***Within-Unit Dependence*****:** For each i, $(Y_{it}, \boldsymbol{D}_{it})$ is a strongly mixing sequence in t with α of size $-\frac{r}{r-2}$, $r > 2$.

A5 ***Compactness*****:** Assume $\boldsymbol{\alpha}_0 \in \text{interior}(\Theta)$, where Θ is compact.

A6 ***Dominance Condition*****:** $\sup_{(i,t)} E\left[(Y_{it} - \boldsymbol{D}_{it}'\boldsymbol{b})^2\right]^{r+\delta} < \Delta < \infty$ for all $\boldsymbol{b} \in \Theta$ and for any $\delta > 0$.

A7 ***Continuity*****:** $\Gamma_{it}(\boldsymbol{b})$ continuous in $\boldsymbol{b}$ with probability one for all $\boldsymbol{b} \in \Theta$.

A4 permits dependence within each unit. Other dependence structures would also generate similar theoretical results. I impose no assumptions about independence or dependence *across* units. The inference procedure will allow for strong dependence across units. **A5** assumes that the parameter is in the interior of a compact parameter space. **A6** is necessary for uniform convergence. **A7** disallows large jumps in the synthetic controls for small changes in the treatment parameters. This assumption does not rule out large changes in the weights for any synthetic control as $\boldsymbol{b}$ changes. Given that the weights are not necessarily unique, we might expect that small changes $\boldsymbol{b}$ may generate large changes in the synthetic control weights. However, this does not imply large changes in $\Gamma_{it}(\boldsymbol{b})$. Since $Y_{it} - \boldsymbol{D}_{it}'\boldsymbol{b}$ is continuous in $\boldsymbol{b}$ and the synthetic control fits this term, assuming continuity of $\Gamma_{it}(\boldsymbol{b})$ is likely unrestrictive.

Under these assumptions, uniform convergence of the sample objective function holds and, consequently, the estimates are consistent by Theorem 2.1 of Newey and McFadden (1994):

Theorem 3.2 (Consistency)**.** *If* ***A1*-*A7*** *hold, then* $\widehat{\boldsymbol{\alpha}} \xrightarrow{p} \boldsymbol{\alpha}_0$.

See Appendix A.1 for a more detailed discussion. For that discussion and for the next section, it is helpful to define

$$\widehat{Q}(\boldsymbol{b}) \equiv -\frac{1}{2NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[Y_{it} - \boldsymbol{D}_{it}'\boldsymbol{b} - \Gamma_{it}(\boldsymbol{b})\right]^2,$$

where $\widehat{Q}(\boldsymbol{b})$ represents the GSC objective function that is maximized.

3.5 Asymptotic Normality

The proposed inference procedure uses the gradient of the objective function and relies on the gradient converging to a normally-distributed random variable. This requires asymptotic normality of the estimates of $\boldsymbol{\alpha}_0$. This section discusses conditions under which the estimates converge at a $\sqrt{T}$ rate. Define the sample average of the gradient as

$$\nabla_{\boldsymbol{\alpha}}\widehat{Q}(\boldsymbol{\alpha}_0) \equiv \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\left[\boldsymbol{D}_{it} + \frac{\partial\Gamma_{it}(\boldsymbol{\alpha}_0)}{\partial\boldsymbol{\alpha}}\right]\left[Y_{it} - \boldsymbol{D}_{it}'\boldsymbol{\alpha}_0 - \Gamma_{it}(\boldsymbol{\alpha}_0)\right]\right\}.$$

Here, the synthetic controls are considered a function of the treatment parameters. The gradient reflects the movement in the objective function with respect to the treatment parameters, which includes changes in the synthetic untreated outcomes. When discussing inference below, it is more straightforward not to nest the synthetic controls in this manner, but it is useful notation for this section. I impose additional regularity conditions (let k index each condition):

A8: $H \equiv E\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathbf{D}_{it} + \frac{\partial\Gamma_{it}(\boldsymbol{\alpha}_0)}{\partial\boldsymbol{\alpha}}\right)\left(\mathbf{D}_{it} + \frac{\partial\Gamma_{it}(\boldsymbol{\alpha}_0)}{\partial\boldsymbol{\alpha}}\right)'\right]$ nonsingular.

A9: (a) $E|\nabla_{\boldsymbol{\alpha}}Q_{itk}(\boldsymbol{\alpha}_0)|^r < \Delta < \infty$ for all (i,t,k).

(b) $\Sigma \equiv \lim_{T\to\infty} \text{Var}\left(\sqrt{NT}\nabla_{\boldsymbol{\alpha}}Q(\boldsymbol{\alpha}_0)\right)$ exists and is finite.

A10: $\Gamma_{it}(\boldsymbol{b})$ twice continuously differentiable for all $\boldsymbol{b} \in \Theta$.

A11: $\sup_{(i,t)} E\left|\left|\frac{\partial^2\Gamma_{it}(\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}'}\left[Y_{it} - \boldsymbol{D}_{it}'\boldsymbol{b} - \Gamma_{it}(\boldsymbol{b})\right] - \left(\mathbf{D}_{it} + \frac{\partial\Gamma_{it}(\boldsymbol{b})}{\partial\boldsymbol{\alpha}}\right)\left(\mathbf{D}_{it} + \frac{\partial\Gamma_{it}(\boldsymbol{b})}{\partial\boldsymbol{\alpha}}\right)'\right|\right|^{r+\delta} < \Delta < \infty$ for some $\delta > 0$ and for all $\boldsymbol{b} \in \Theta$.

A8 states that the Hessian is full rank. **A9** (with **A4**) implies that a central limit theorem holds for the gradient. **A11** is necessary for uniform convergence of the Hessian matrix.

Theorem 3.3 (Asymptotic Normality)**.** *If **A1-A11** hold, then* $\sqrt{T}\left(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\right) \xrightarrow{d} N(0, \frac{1}{N}H^{-1}\Sigma H^{-1})$.

I include a discussion of this theorem in Appendix A.1. As before, I do not impose any assumptions about independence across units. Estimation of Σ is likely complicated when N is fixed and correlations across units exist. Estimation of H is also likely complicated since it requires estimation of $\frac{\partial \Gamma_{it}(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}}$. These concerns motivate the inference procedure discussed below. The inference procedure will not require estimation of H or Σ and will account for dependence across units.

4 Inference

I discuss an inference procedure that is valid for fixed N as $T \to \infty$. The procedure permits proper inference even if the units are dependent. It is common in applied research using panel data to account for within-unit dependence (i.e., "clustering by state"). Because of this standard, I will use "units" and "clusters" interchangeably in this section, though it is possible to construct clusters including multiple states using the proposed inference method.[13]

Permitting cross-unit dependence is important in this context and, more generally, whenever a fixed number of units are used to construct measures of variables to explicitly "difference out" from another unit. This approach generates mechanical correlations across clusters. For the GSC method, the estimator constructs a synthetic control for each unit i through a weighted (constrained) average of $Y_{jt} - \boldsymbol{D}'_{jt}\widehat{\boldsymbol{\alpha}}$. Consequently, each unit potentially composes part of the synthetic controls for other units, generating correlations across units. This section follows the inference procedure developed in Powell (2017b) for panel data estimators more generally.

The inference method relies on a Wald statistic given the mean and variance of the cluster-specific scores evaluated at the restricted (imposing the null hypothesis) estimates. It then simulates the distribution of the Wald statistic by perturbing the score functions using weights generated by the Rademacher distribution, equal to 1 with probability $\frac{1}{2}$ and -1 with probability $\frac{1}{2}$. The approach uses the panel nature of the data to estimate the relationship between the scores for each cluster and isolate the independent component for each cluster. Once the independent functions are isolated, appropriate inference is possible through simulation. This method is straightforward to implement. This paper considers a

[13]Given that the proposed method adjusts for dependence across clusters, there is likely little gain in aggregating units into larger clusters. However, if the researcher believes that condition **I2** below does not appropriately model dependence across clusters, then there are benefits in creating larger clusters. The inference approach of this paper is well-suited for this aggregating since it is valid for a small number of clusters.

null hypothesis of the form

$$H_0 : a(\boldsymbol{\alpha}_0) = 0, \tag{10}$$

which is rank L. This framework allows for nonlinear hypotheses. I assume the null hypothesis only involves the treatment parameters and not the synthetic control weights. Formally, I impose the following conditions:

I1 *(Null Hypothesis): (i) $a(\boldsymbol{\alpha}_0) = 0$ with $a(\cdot)$ continuously differentiable; (ii) $A \equiv \nabla_{\boldsymbol{\alpha}} a(\boldsymbol{\alpha}_0)$ has rank L.*

The null hypothesis will be tested by imposing the restriction when minimizing the objective function:

$(\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{w}}_1, \ldots, \tilde{\boldsymbol{w}}_N) =$

$$\underset{\boldsymbol{b}, \boldsymbol{\phi}_1 \in \mathcal{W}_1, \ldots, \boldsymbol{\phi}_N \in \mathcal{W}_N}{\operatorname{argmin}} \left\{ \frac{1}{2NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[Y_{it} - \boldsymbol{D}_{it}' \boldsymbol{b} - \sum_{j \neq i} \left(\phi_i^j \left(Y_{jt} - \boldsymbol{D}_{jt}' \boldsymbol{b} \right) \right) \right]^2 \right\} \quad \text{subject to } a(\boldsymbol{b}) = 0. \tag{11}$$

In words, the null hypothesis is imposed and the other parameters (including the synthetic control weights) are estimated.[14] The inference procedure will use the gradient with respect to $\boldsymbol{\alpha}$ for each unit i evaluated at the restricted estimates, which I denote

$$h_i(\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{w}}_i) \equiv \frac{1}{T} \sum_{t=1}^{T} \left[\mathbf{D}_{it} - \sum_{j \neq i} \tilde{w}_j^i \boldsymbol{D}_{jt} \right] \left[Y_{it} - \boldsymbol{D}_{it}' \tilde{\boldsymbol{\alpha}} - \sum_{j \neq i} \tilde{w}_j^i \left[Y_{jt} - \boldsymbol{D}_{jt}' \tilde{\boldsymbol{\alpha}} \right] \right]. \tag{12}$$

$h_i(\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{w}}_i)$ is a $K \times 1$ vector. An advantage of the inference procedure introduced in Powell (2017b) is that it is unnecessary to use all elements of the gradient. It only uses a subset of these elements ($L \leq \tilde{K} \leq K$), defined by the conditions below. This property permits use of the gradient with respect to the treatment parameters only and avoids requiring estimation of the gradient with respect to the synthetic control weights, which may be complicated.

The test statistic below does not (necessarily) use all elements of $\widehat{h}(\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{w}}) \equiv \frac{1}{N} \sum_{i=1}^{N} h_i(\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{w}}_i)$

[14] It is recommend that the weights adjusting for synthetic control fit, constructed in equation (7), are used here. I do not explicitly include them in the equations in the section but this is done without loss of generality. Note that the equation (7) weights do not need to be re-estimated for inclusion in equation (11).

where $\tilde{\boldsymbol{w}} = (\tilde{\boldsymbol{w}}_1, \cdots, \tilde{\boldsymbol{w}}_N)$ but, instead, relies on a subset. I refer to this vector as $\widehat{g}(\tilde{\boldsymbol{\alpha}})$, which is $\tilde{K} \times 1$, while suppressing the dependence of this function on the synthetic control weights. This vector has the identification condition that for all k,

$$E\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} g_{itk}(\boldsymbol{\alpha}_0)\right] = 0 \quad \text{if and only if} \quad a(\boldsymbol{\alpha}_0) = 0, \tag{13}$$

where k indexes the element and $g_{itk}(\tilde{\boldsymbol{\alpha}}) = h_{itk'}(\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{w}}_i)$ for some k'. This condition eliminates elements which are equal to zero regardless of whether the null hypothesis is true or not. The test statistic below is valid if elements which are always equal to zero are used, but their inclusion may reduce power.

The motivation for the proposed inference procedure is that the GSC estimator induces correlations across units. These correlations occur because of dependence within the same time period since the synthetic control for $Y_{it} - \boldsymbol{D}_{it}'\tilde{\boldsymbol{\alpha}}$ is a weighted average of $Y_{jt} - \boldsymbol{D}_{jt}'\tilde{\boldsymbol{\alpha}}$ for $j \neq i$. Imagine if the control for unit i involves a substantial weight on unit j, and vice versa. Examining equation (12), it would be difficult to consider the scores for units i and j independent. The inference method introduced in this section uses co-movements in the scores across units to estimate the dependence between units, and then isolates the independent components from each unit for proper inference. The next assumption permits this dependence across units:

I2 *(Dependence Across Clusters):* $g_{itk}(\boldsymbol{\alpha}_0)$ *can be expressed as*

$$g_{itk}(\boldsymbol{\alpha}_0) = \sum_{j=i+1}^{N} c_{jk}^{i} \times g_{jtk}(\boldsymbol{\alpha}_0) + \mu_{itk}, \tag{14}$$

with $E\left[g_{jtk}(\boldsymbol{\alpha}_0)\mu_{itk}\right] = 0$ *and* $E\left[\mu_{itk}\mu_{jsk}\right] = 0$ *for all* s, t, k, *and* $i \neq j$.

I model each element of the gradient (denoted by k) for unit i as correlated with the gradients of other units. The above setup is a triangular framework such that g_{itk} is a function of $g_{i+1,tk}, \ldots, g_{Ntk}$. There are no constraints on the c_{jk}^{i} terms, permitting arbitrarily-strong correlations. The triangular framework is itself not a restriction. If there is a common component in two or more units, it is important to keep this information in one unit and eliminate it in all other units, implying a triangular framework. Moreover, there are no restrictions in **I2** on within-unit dependence. The above setup separates the gradient of each unit into a dependent component ($\sum_{j=i+1}^{N} c^i_{jk} \times g_{jtk}(\boldsymbol{\alpha}_0)$) and an independent component (μ_{itk}).

Condition **I2** models dependence as operating through correlations in the same time period. The focus on these contemporaneous co-movements is because the dependence induced by the estimation procedure operates through correlations in the same time period. The estimator uses the untreated outcomes of other units in period t to control for unit i's untreated outcome in period t. Given within-unit dependence as well, this approach may induce correlations across all observations in unit i with all observations in the units composing the synthetic control for unit i. However, by eliminating the contemporaneous co-movements, the inference method eliminates dependence across all observations in different units and then constructs asymptotically independent functions. While the primary motivation for the inference method is to account for the mechanical correlations caused by the GSC estimator, condition **I2** permits other (observable and unobservable) economic sources of dependence across units that are unrelated to the estimator.

To highlight, condition **I2** does *not* assume that there is only cross-cluster dependence in the same time period. Equation (14) assumes that lags and leads for other clusters are not *independently* related to $g_{itk}(\boldsymbol{\alpha}_0)$. This setup still permits strong correlations between $g_{it}(\boldsymbol{\alpha}_0)$ and $g_{js}(\boldsymbol{\alpha}_0)$ as long as those correlations operate through $g_{jt}(\boldsymbol{\alpha}_0)$ or $g_{is}(\boldsymbol{\alpha}_0)$. Given dependence within-cluster i and within-cluster j as well as correlations across clusters in each time period, this framework permits a rich level of dependence across observations in clusters i and j. The idea behind the inference method is that by eliminating the correlation across clusters i and j within the same period, the approach eliminates dependence across all observations in clusters i and j. Note further that the inference procedure permits finite dependence across $g_{it}(\boldsymbol{\alpha}_0)$ and $g_{js}(\boldsymbol{\alpha}_0)$ that does not operate through $g_{jt}(\boldsymbol{\alpha}_0)$ or $g_{is}(\boldsymbol{\alpha}_0)$. Powell (2017b) discusses extensions of this model to include leads and lags in equation (14), which further relaxes the assumption that correlations must operate through the period-t scores. That modification can be applied here.

The mechanical dependence generated by the GSC estimator operates through period-t dependence and is modeled appropriately by equation (14). Equation (14) also permits additional sources of dependence. Units may be dependent for a variety of reasons such as geographical proximity, similar industry compositions, comparable political institutions, and unobservable factors. The inference method estimates these co-movements, regardless of the underlying source.

The goal is to create functions which are asymptotically independent across clusters. It is straightforward to estimate equation (14), plugging in $\tilde{\boldsymbol{\alpha}}$ as a consistent estimate for $\boldsymbol{\alpha}_0$, for each cluster i and condition k using OLS. Estimation of equation (14) estimates the co-movements across clusters and allows one to subtract off the dependent component of the score to isolate the independent component. Given estimates $\hat{c}^i_{jk}$, it is possible to construct $\widehat{\mu}_{itk} = g_{itk}(\tilde{\boldsymbol{\alpha}}) - \sum_{j=i+1}^{N} \hat{c}^i_{jk} \times g_{jtk}(\tilde{\boldsymbol{\alpha}})$ and, subsequently,

$$s_{ik}(\tilde{\boldsymbol{\alpha}}) \equiv \frac{1}{T}\sum_{i=1}^{T} \widehat{\mu}_{itk}, \tag{15}$$

where $s_{ik}(\tilde{\boldsymbol{\alpha}})$ is the asymptotically-independent score function for condition k. In practice, equation (14) can be estimated with a constant. The estimated constant is $s_{ik}(\tilde{\boldsymbol{\alpha}})$. Define $g^i_{tk}(\boldsymbol{\alpha}_0) \equiv \big(g_{i+1,tk}(\boldsymbol{\alpha}_0), \ldots, g_{Ntk}(\boldsymbol{\alpha}_0)\big)'$ and $\boldsymbol{c}^i_k \equiv \left(c^i_{i+1,k}, \ldots, c^i_{N,k}\right)'$ so that equation (14) can be rewritten as $g_{itk}(\boldsymbol{\alpha}_0) = g^i_{tk}(\boldsymbol{\alpha}_0)'\boldsymbol{c}^i_k + \mu_{itk}$, which will be helpful notation in the subsequent analysis. Define $\hat{\boldsymbol{c}}^i_k \equiv \left[\frac{1}{T}\sum_{t=1}^{T} g^i_{tk}(\tilde{\boldsymbol{\alpha}}) g^i_{tk}(\tilde{\boldsymbol{\alpha}})'\right]^{-1} \left[\frac{1}{T}\sum_{t=1}^{T} g^i_{tk}(\tilde{\boldsymbol{\alpha}}) g_{itk}(\tilde{\boldsymbol{\alpha}})\right]$, which represents the estimates from a regression of $g_{itk}(\tilde{\boldsymbol{\alpha}})$ on $\big(g_{i+1,tk}(\tilde{\boldsymbol{\alpha}}), \ldots, g_{Ntk}(\tilde{\boldsymbol{\alpha}})\big)$.

For inference, the method uses a Wald statistic defined by:

$$S = \left(\frac{1}{N}\sum_{i=1}^{N} s_i(\tilde{\boldsymbol{\alpha}})\right)' \widehat{\Sigma}(\tilde{\boldsymbol{\alpha}})^{-1} \left(\frac{1}{N}\sum_{i=1}^{N} s_i(\tilde{\boldsymbol{\alpha}})\right), \tag{16}$$

where $\widehat{\Sigma}(\tilde{\boldsymbol{\alpha}}) = \frac{1}{N-1}\sum_{i=1}^{N}\left(s_i(\tilde{\boldsymbol{\alpha}}) - \left(\frac{1}{N}\sum_{j=1}^{N} s_j(\tilde{\boldsymbol{\alpha}})\right)\right)\left(s_i(\tilde{\boldsymbol{\alpha}}) - \left(\frac{1}{N}\sum_{j=1}^{N} s_j(\tilde{\boldsymbol{\alpha}})\right)\right)'$. This approach avoids estimation of H and Σ, defined in Section 3.5, which would likely be difficult to estimate. Instead, it uses the empirical mean and variance of $s_i(\tilde{\boldsymbol{\alpha}})$ across clusters.

To derive p-values, the inference method simulates the distribution of this test statistic by perturbing the $s_i(\tilde{\boldsymbol{\alpha}})$ functions. The weights used to perturb the functions meet the following conditions:

I3 (Weights)*: Assume i.i.d weights $\{W_i\}_{i=1}^N$ independent of $s_i(\boldsymbol{\alpha}_0)$ drawn from the Rademacher distribution.*

To simulate the distribution of the test statistic, consider weights $W_i^{(d)}$ which satisfy condition **I3** and construct a simulated test statistic, indexed by (d):

$$S^{(d)} = \left(\frac{1}{N} \sum_{i=1}^{N} W_i^{(d)} s_i(\tilde{\boldsymbol{\alpha}}) \right)' \widehat{\Sigma}^{(d)}(\tilde{\boldsymbol{\alpha}})^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} W_i^{(d)} s_i(\tilde{\boldsymbol{\alpha}}) \right), \tag{17}$$

$$\widehat{\Sigma}^{(d)}(\tilde{\boldsymbol{\alpha}}) = \frac{1}{N-1} \sum_{i=1}^{N} \left(W_i^{(d)} s_i(\tilde{\boldsymbol{\alpha}}) - \left(\frac{1}{N} \sum_{j=1}^{N} W_j^{(d)} s_j(\tilde{\boldsymbol{\alpha}}) \right) \right) \left(W_i^{(d)} s_i(\tilde{\boldsymbol{\alpha}}) - \left(\frac{1}{N} \sum_{j=1}^{N} W_j^{(d)} s_j(\tilde{\boldsymbol{\alpha}}) \right) \right)'.$$

Given weights $W_i^{(d)}$, it is possible to simulate the distribution of $s_i(\tilde{\boldsymbol{\alpha}})$ for each i. The Rademacher distribution provides weights that are equal to 1 with probability $\frac{1}{2}$ and equal to -1 with probability $\frac{1}{2}$. Consequently, $E[W_i] = 0$ and $E[W_i^2] = 1$ such that the gradient functions and the perturbed gradient functions will match asymptotically given that they converge to a normal distribution. In addition, $E[W_i^3] = 0$ and $E[W_i^4] = 1$ which provides asymptotic refinement for symmetric distributions. The limitation of the Rademacher distribution is that when N is small, there are only a limited number of unique simulated test statistics that it can generate. If the number of units is less than 10, Powell (2017b) recommends using the weights introduced in Webb (2014).

The weights are independent across units which is why it is necessary to have independent functions since any cross-unit dependence will not be preserved by the weights. The test statistic and simulated test statistic converge to the same distribution. The following theorem is discussed extensively in Appendix Section A.2:

Theorem 4.1. *Assume **A1-A11, I1-I3**, (i) $E|g_{itk}(\boldsymbol{\alpha})^2|^{(r+\delta)} < \Delta < \infty$ for all i, t, k and for some $\delta > 0$, for $\boldsymbol{\alpha}$ in some neighborhood of $\boldsymbol{\alpha}_0$; (ii) $E\left[\frac{1}{T}\sum_{t=1}^{T} g_{tk}^i(\boldsymbol{\alpha}_0) g_{tk}^i(\boldsymbol{\alpha}_0)'\right]$ is uniformly positive definite. Then, S and $S^{(d)}$ converge to the same distribution.*

Given the result in Theorem 4.1, it is straightforward to create the test statistic S and then simulate its distribution under the null hypothesis. This inference approach is simple to implement. The inference steps are as follows:

1. Estimate the constrained parameters using equation (11).

2. Create the gradient for each unit using equation (12).
3. Estimate equation (14) for each unit and condition. The estimated constants are the $s_i(\tilde{\boldsymbol{\alpha}})$ functions.[15]
4. Create the test statistic defined by equation (16).
5. Simulate the test statistic using Rademacher weights $D = 999$ times (equation (17)).
6. The p-value is the fraction of simulations in which the simulated value of the test statistic is greater than S: $\hat{p} = \frac{1}{D}\sum_{d=1}^{D} \mathbf{1}(S < S^{(d)})$.

Large values of S imply that we should rarely observe the distribution of gradients created by the null hypothesis and that we should reject it. By Theorem 4.1, we can simulate how often we should see a value as large as S if the null hypothesis were true.

The proposed inference method produces valid inference given a finite number of dependent clusters. There are also no assumptions about homogeneity across clusters. This inference procedure is simple to implement and computationally inexpensive. It only requires one estimation of the model under the null hypothesis. Once estimated, the perturbations use the scores evaluated at the restricted estimates. In the next section, I show rejection rates of a true null hypothesis to test the usefulness of the proposed approach.

5 Simulations

In this section, I report the results of simulations using the synthetic control estimator of this paper and the more traditional additive fixed effects estimator. In Section 5.1, I generate data with a continuous treatment effect which is a function of interactive fixed effects. In Section 5.2, I simulate a differences-in-differences design in which all units adopt the policy but at different times.

[15] Note that $s_N(\tilde{\boldsymbol{\alpha}}) = g_N(\tilde{\boldsymbol{\alpha}})$.

5.1 Continuous Treatment Variables

The data are generated by

$$\begin{aligned} \text{Treatment Variable:} \quad d_{it} &= \phi_{it} + \mu_i^{(1)}\lambda_t^{(1)} + 2\mu_i^{(2)}\lambda_t^{(2)}, \\ \text{Outcome:} \quad y_{it} &= 2\mu_i^{(1)}\lambda_t^{(1)} + \mu_i^{(2)}\lambda_t^{(2)} + \epsilon_{it}, \end{aligned}$$

where $\mu_i^{(1)}, \mu_i^{(2)} \sim U(0,1)$; $\epsilon_{it} \sim N(0,1)$; $\lambda_t^{(1)} \sim N(0,400)$; $\lambda_t^{(2)} \sim N(0,100)$; $\phi_{it} \sim U(0,50)$. Both the outcome and the treatment variable are functions of the interactive fixed effects. Not accounting for these interactive fixed effects should lead to biased estimates. I generate the above data for $N = 30$ and $T = 30$. The treatment effect is zero, and the treatment variable is continuous. In Panel A of Table 1, I report Mean Bias, Median Absolute Deviation (MAD) and Root-Mean-Square Error (RMSE). The first estimator is an additive fixed effects estimator which includes both state and time fixed effects. I also show results in which state-specific trends are included. The second estimator is the synthetic control estimator of this paper. The fixed effects estimator, as expected, performs poorly. Even when flexible state-specific trends are included, the treatment effect estimate is biased. In fact, the inclusion of state-specific trends does not reduce the bias.

GSC performs well. The mean bias is 0.0010. In Panel B of Table 1, I test the inference procedure discussed in Section 4. The inference procedure works well at multiple significance levels, even though the procedure benefits from T much greater than N and the simulations use $N = T$. The inference method rejects the null hypothesis at a rate of 0.055 for significance level 0.05. Though not shown, the fixed effects models with and without trends (adjusting for clustering at the state level and using a t-distribution with 29 degrees of freedom) reject in 100% of simulations.

5.2 Difference-in-Differences

I generate similar data as above but simulate a difference-in-differences setup with a discrete treatment variable. This treatment variable is equal to 0 for all units initially. Once the treatment is adopted, the unit never repeals the treatment. Finally, all units adopt treatment by period $t = 28$. It would be difficult to use CSSC to estimate the relevant parameter with these data since it is not clear which states to consider "treated" and which to include as

Table 1: Simulation Results: Continuous Treatment Variable

	Panel A: Estimates		
	Mean Bias	Median Absolute Deviation	Root Mean Squared Error
Fixed Effects	0.2964	0.29	0.3034
+ Linear Trends	0.2955	0.29	0.3025
+ Quadratic Trends	0.2955	0.29	0.3028
+ Cubic Trends	0.2952	0.29	0.3027
Synthetic Control	**0.0010**	**0.00**	**0.0064**
	Panel B: Rejection Rates		
Significance Level:	0.10	0.05	0.01
Synthetic Control	**0.109**	**0.055**	**0.010**

The "fixed effects" estimator refers to a model with additive state and time fixed effects. The trends refer to state-specific trends of 1, 2, or 3 degrees. The "synthetic control" estimator is introduced in this paper. The inference procedure is discussed in Section 4.

part of the donor pool for those treated units. The data are generated in the following manner:

$$\text{For} \quad a_{it} = \mathbf{1}\left(\phi_{it} + \mu_i^{(1)}\lambda_t^{(1)} + 2\mu_i^{(2)}\lambda_t^{(2)} > 45\right),$$

$$\text{Treatment Variable:} \quad d_{it} = \begin{cases} 0 & \text{if} \quad t = 1 \\ 1 & \text{if} \quad d_{i,t-1} = 1 \quad \text{or} \quad a_{it} = 1 \quad \text{or} \quad t \geq 28 \\ 0 & \text{otherwise} \end{cases}$$

$$\text{Outcome:} \quad y_{it} = 2\mu_i^{(1)}\lambda_t^{(1)} + \mu_i^{(2)}\lambda_t^{(2)} + \epsilon_{it},$$

where $\mu_i^{(1)}, \mu_i^{(2)} \sim U(0,1)$; $\epsilon_{it} \sim N(0,1)$; $\lambda_t^{(1)} \sim N(0,400)$; $\lambda_t^{(2)} \sim N(0,100)$; $\phi_{it} \sim U(0,50)$. This data generating process produces a serially-correlated treatment dummy variable equal to 1 about 50% of the time. As before, I set $N = 30$ and $T = 30$. The simulation results are presented in Table 2. The additive fixed effects estimator performs especially poorly as one would expect given the correlation between the (omitted) interactive fixed effects and the treatment variable. Adding flexible state-specific trends does not reduce this bias and, in fact, appears to increase the bias. The synthetic control estimator performs well; the mean bias is 0.0010. The proposed inference method also rejects at approximately the correct rates.

Table 2: Simulation Results: Difference-in-Differences

	Mean Bias	Median Absolute Deviation	Root Mean Squared Error
	Panel A: Estimates		
Fixed Effects	1.9680	2.51	4.0256
+ Linear Trends	1.8180	2.79	4.3696
+ Quadratic Trends	3.1529	3.22	5.3867
+ Cubic Trends	3.9879	4.18	6.5266
Synthetic Control	**0.0010**	**0.02**	**0.0276**
	Panel B: Rejection Rates		
Significance Level:	0.10	0.05	0.01
Synthetic Control	**0.108**	**0.057**	**0.013**

The "fixed effects" estimator refers to a model with additive state and time fixed effects. The trends refer to state-specific trends of 1, 2, or 3 degrees. The "synthetic control" estimator is introduced in this paper. The inference procedure is discussed in Section 4.

6 Employment Effects of Minimum Wage

Recent work has suggested that the traditional fixed effects strategy in the minimum wage literature is inappropriate. I first replicate findings using the additive fixed effect specification typically estimated in the literature:

$$\ln E_{st} = \alpha_s + \gamma_t + \beta \ln(MW_{st}) + \epsilon_{st}, \tag{18}$$

where E_{st} is the employment rate in state s, quarter t for 16-19 year olds. The log of the employment rate is modeled as a function of state fixed effects, time fixed effects, and the log of the minimum wage (MW). I do not control for the overall state unemployment rate or the relative size of the youth population, though these controls are frequently included in the literature. These controls are usually included because of concerns that minimum wage policy reacts to state economic conditions, but they are potentially problematic if they are also affected by changes in the minimum wage. An advantage of the synthetic control estimator is that it should make inclusion of these (potentially endogenous) control variables unnecessary.

I use the data set constructed for Dube and Zipperer (2015) and Allegretto et al.

(2017).[16] This allows for straightforward comparisons of estimates generated by the synthetic control approach and fixed effects estimation. The data are state-level minimum wage information merged with employment rates aggregated from the Current Population Survey (CPS) for 1979-2014. The data are quarterly and aggregated by state. The outcome of interest is the employment rate of the 16-19 population. There are 51 units and 144 time periods.

Table 3 replicates results found in Table 2 of Allegretto et al. (2017) without conditioning on the control variables mentioned above. Columns (1)-(6) show that the estimate varies between -0.243 and 0.083 depending on the inclusion and flexibility of state-specific trends. Columns (7)-(12) include division-time fixed effects such that identification originates from minimum wage changes within a Census division. The estimates vary between -0.228 and 0.125. In general, the estimates appear sensitive to the inclusion of state-specific trends and choice of control group. GSC should account for these considerations.

Table 3: Effect of Minimum Wage on Teen Employment

Outcome:	ln(Employment / Population)					
	Additive State and Time Fixed Effects					
	(1)	(2)	(3)	(4)	(5)	(6)
ln(Minimum Wage)	-0.243***	0.044	0.053	0.083	0.017	0.026
	[-0.406, -0.080]	[-0.120, 0.208]	[-0.139, 0.245]	[-0.097, 0.262]	[-0.174, 0.209]	[-0.166, 0.218]
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
State Trends	No	Linear	Quadratic	Cubic	Quartic	Quintic
Division-specific Time Effects	No	No	No	No	No	No
	Additive State and Division-Time Fixed Effects					
	(7)	(8)	(9)	(10)	(11)	(12)
ln(Minimum Wage)	-0.228*	0.106	0.125	0.105	0.077	0.084
	[-0.469, 0.013]	[-0.056, 0.267]	[-0.105, 0.356]	[-0.070, 0.281]	[-0.099, 0.252]	[-0.098, 0.266]
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
State Trends	No	Linear	Quadratic	Cubic	Quartic	Quintic
Division-specific Time Effects	Yes	Yes	Yes	Yes	Yes	Yes
	Synthetic Control Estimates					
	(13)	(14)				
ln(Minimum Wage)	-0.45**	-0.44***				
	[-1.19, -0.02]	[-0.77, -0.02]				
Two-Step Weighting?	No	Yes				

Notes: $N = 7344$. Significance Levels: *10%, **5%, ***1%. 95% Confidence Intervals in brackets are adjusted for clustering at state-level. Hypothesis testing for synthetic control estimator use simulation technique discussed in text. Inference for the synthetic control estimates accounts for within-state clustering and cross-sectional correlations. Two-step weighting uses the estimation procedure discussed in Section 3.2.2.

I report the main empirical results of this paper in Columns (13) and (14) of Table 3.

[16] Code and data can be found at Arindrajit Dube's site: http://arindube.com/working-papers/, last accessed January 15, 2016.

In column (13), I report the GSC estimate which does not adjust for synthetic control fit and weights all states equally (i.e., using equation (5)). I estimate an elasticity of -0.45, larger in magnitude than estimates typically found in the literature. In Section 3.2.2, I discussed the merits of a two-step procedure which places more weight on states with better synthetic control fits. Such a two-step procedure could also be employed for fixed effect estimators but this is not typically done in applied work. Column (14) reports the estimates using the two-step procedure (i.e., using equation (8)). I estimate an elasticity of -0.44, implying that a 10% increase in the minimum wage decreases the teen employment rate by 4.4%. This estimate is statistically significant from zero at the 1% level. These estimates are larger in magnitude than those typically generated by additive fixed effect models, suggesting that those models produce estimates that are biased against finding a relationship between the minimum wage and employment.

7 Discussion and Conclusion

Many empirical applications rely on panel data to exploit differential changes in the policy variables to study their impacts on the outcome. It is common to condition on state and time fixed effects, implicitly assuming that the average unit acts as an appropriate control for each unit. This paper introduces a synthetic control estimator which generalizes the case study synthetic control estimator of Abadie and Gardeazabal (2003) and Abadie et al. (2010). It also nests traditional additive fixed effect models. The estimator jointly estimates the parameters associated with the treatment variables while creating synthetic controls for each unit. These synthetic controls permit the treatment variables to co-vary with flexible state-specific trends and shocks. Instead of assuming that the national average is an appropriate control for each treated unit, the generalized synthetic control estimator permits additional flexibility in constructing the controls for each unit.

Additive fixed effects models are not limited to applications in which the treatment is represented by a single binary variable. It is also unnecessary in such models to divide units into “ever-treated” units and “never-treated” units. This paper introduces a synthetic control equivalent, permitting researchers to account for one or more explanatory variables, which can be discrete or continuous. The introduced estimation technique combines the benefits of more traditional panel data estimators (valid for continuous treatment variables, multiple explanatory variables, policies adopted by all units over time) with the benefits of synthetic control estimation (conditioning on flexible interactive fixed effects and permitting non-parallel trends). This paper also discusses the benefits of the proposed approach even in traditional "case study" settings.

Inference is potentially complicated in this context given that the estimator generates dependence across units and inference methods often rely on (asymptotic) independence across units. I introduced a simple procedure which is valid even when the number of units is fixed, there is heterogeneity across units, and there is cross-unit dependence. The estimator and inference procedure work well in simulations. I provide evidence of the usefulness of the estimator by estimating the effect of the minimum wage on teenage employment rates. The recent literature on this topic has noted that choice of appropriate control groups for states increasing their minimum wages is essential for consistent estimation and potentially difficult. The estimator of this paper creates optimal data-driven control groups. I estimate larger employment reductions than often found in the minimum wage literature, suggesting that additive fixed effect models are producing estimates biased towards zero. The estimator in this paper should be useful more broadly for applications using panel data.

A Appendix

A.1 Properties of Estimator

Theorem 3.1 (Identification)**.** *If **A1**-**A3** hold, then* $E\left[\left(Y_{it} - \boldsymbol{D}'_{it}\boldsymbol{b} - \sum_{j\neq i}\left(\phi_i^j\left(Y_{jt} - \boldsymbol{D}'_{jt}\boldsymbol{b}\right)\right)\right)^2\right]$ *has a unique minimum and* $\boldsymbol{b} = \boldsymbol{\alpha}_0$ *at this minimum.*

Proof.

$$E\left[\left(Y_{it} - \boldsymbol{D}'_{it}\boldsymbol{b} - \sum_{j\neq i}\left(\phi_i^j\left(Y_{jt} - \boldsymbol{D}'_{jt}\boldsymbol{b}\right)\right)\right)^2\right]$$

is minimized for $\boldsymbol{b}, \boldsymbol{\phi}_i \in \mathcal{W}_i$ such that $\boldsymbol{D}'_{it}\boldsymbol{b} + \sum_{j\neq i}\left(\phi_i^j\left(Y_{jt} - \boldsymbol{D}'_{jt}\boldsymbol{b}\right)\right)$ is equal to the conditional expectation of Y_{it}. I first show that $\boldsymbol{D}'_{it}\boldsymbol{\alpha}_0 + \sum_{j\neq i}\left(w_i^j\left(Y_{jt} - \boldsymbol{D}'_{jt}\boldsymbol{\alpha}_0\right)\right)$ is the expected value of Y_{it}.

$$\begin{aligned} E\left[Y_{it} - \mathbf{D}'_{it}\boldsymbol{\alpha}_0 - \delta_t \middle| \mathbf{D}_i\right] &= E\left[\boldsymbol{\lambda}_t\boldsymbol{\mu}_i + \epsilon_{it}\middle|\mathbf{D}_i\right] \quad \text{by } \mathbf{A1}, \\ &= \sum_{j\neq i} w_i^j\left[\boldsymbol{\lambda}_t\boldsymbol{\mu}_j + \epsilon_{jt}\right] \quad \text{by } \mathbf{A2}. \end{aligned}$$

The above implies that

$$E\left[Y_{it}\middle|\mathbf{D}_i\right] = \mathbf{D}'_{it}\boldsymbol{\alpha}_0 + \sum_{j\neq i} w_i^j\left(Y_{jt} - \mathbf{D}'_{jt}\boldsymbol{\alpha}_0\right).$$

Thus,

$$E\left[Y_{it}\middle|\mathbf{D}_i\right] = \left(\mathbf{D}'_{it} - \sum_{j\neq i} w_i^j\mathbf{D}'_{jt}\right)\boldsymbol{\alpha}_0 + \sum_{j\neq i} w_i^j Y_{jt}$$

By **A3**, we know that this is unique at $\boldsymbol{b} = \boldsymbol{\alpha}_0$. □

Theorem 3.2 (Consistency)**.** *If **A1**-**A7** hold, then* $\widehat{\boldsymbol{\alpha}} \xrightarrow{p} \boldsymbol{\alpha}_0$.

Proof. By the triangle inequality, the definition of $\mathcal{W}_i$, and **A6**,

$$E\left[\left(Y_{it} - \boldsymbol{D}_{it}'\boldsymbol{b} - \sum_{j\neq i}\left[w_i^j(\boldsymbol{b})\left(Y_{jt} - \boldsymbol{D}_{jt}'\boldsymbol{b}\right)\right]\right)^2\right]^{r+\delta} \leq 2\sup_{(i,t)} E\left[\left(Y_{it} - \boldsymbol{D}_{it}'\boldsymbol{b}\right)^2\right]^{r+\delta} < \Delta' < \infty,$$

where $\Delta' \equiv 2\Delta$. Note also by **A7** that the objective function is continuous at each $\boldsymbol{b}$ with probability one. Given these properties and conditions **A4** and **A5**, a uniform law of large numbers (see Theorem 2.3 in White and Domowitz (1984)) implies that uniform convergence of the sample objective function to its expectation holds. It also implies continuity of the expectation.

Consistency of $\widehat{\boldsymbol{\alpha}}$ then follows from Theorem 2.1 of Newey and McFadden (1994).

□

Theorem 3.3 (Asymptotic Normality)**.** *If* ***A1-A11*** *hold, then* $\sqrt{T}\,(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) \xrightarrow{d} N(0, \frac{1}{N}H^{-1}\Sigma H^{-1})$.

Proof. $\widehat{\boldsymbol{\alpha}}$ solves the first order condition:

$$\nabla_{\boldsymbol{\alpha}}\widehat{Q}(\widehat{\boldsymbol{\alpha}}) = 0.$$

Expanding around $\boldsymbol{\alpha}_0$,

$$\nabla_{\boldsymbol{\alpha}}\widehat{Q}(\boldsymbol{\alpha}_0) + \widehat{H}(\boldsymbol{\alpha}^*)\,(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) = 0,$$

where $\boldsymbol{\alpha}^*$ represents a different intermediate point between $\widehat{\boldsymbol{\alpha}}$ and $\boldsymbol{\alpha}_0$ for each row and

$$\widehat{H}_i(\boldsymbol{b}) \equiv \frac{1}{T}\sum_{t=1}^{T}\left\{\left[\frac{\partial^2\Gamma_{it}(\boldsymbol{b})}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}'}\right]\left[Y_{it} - \boldsymbol{D}_{it}'\boldsymbol{b} - \Gamma_{it}(\boldsymbol{b})\right] - \left[\boldsymbol{D}_{it} + \frac{\partial\Gamma_{it}(\boldsymbol{b})}{\partial\boldsymbol{\alpha}}\right]\left[\boldsymbol{D}_{it} + \frac{\partial\Gamma_{it}(\boldsymbol{b})}{\partial\boldsymbol{\alpha}}\right]'\right\},$$

$$\widehat{H}(\boldsymbol{b}) \equiv \frac{1}{N}\sum_{i=1}^{N}\widehat{H}_i(\boldsymbol{b}).$$

$\widehat{H}(\boldsymbol{b})$ is continuous at each $\boldsymbol{b}$ (by **A10**). By **A11**, the uniform law of large numbers (Theorem

2.3 in White and Domowitz (1984)) implies

$$\sup_{\boldsymbol{b}\in\Theta} \|\widehat{H}(\boldsymbol{b}) - H(\boldsymbol{b})\| \xrightarrow{p} 0,$$

where $H(\boldsymbol{b}) \equiv E\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\partial^2\Gamma_{it}(\boldsymbol{b})}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}'}\left[Y_{it} - \boldsymbol{D}_{it}'\boldsymbol{b} - \Gamma_{it}(\boldsymbol{b})\right] - \left(\mathbf{D}_{it} + \frac{\partial\Gamma_{it}(\boldsymbol{b})}{\partial\boldsymbol{\alpha}}\right)\left(\mathbf{D}_{it} + \frac{\partial\Gamma_{it}(\boldsymbol{b})}{\partial\boldsymbol{\alpha}}\right)'\right]$.

By the continuous mapping theorem, consistency, and **A8**,

$$\sqrt{T}\left(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\right) = -H^{-1}\sqrt{T}\nabla_{\boldsymbol{\alpha}}\widehat{Q}(\boldsymbol{\alpha}_0) + o_p.$$

By the central limit theorem under conditions **A4** and **A9**,

$$\sqrt{T}\nabla_{\boldsymbol{\alpha}}\widehat{Q}(\boldsymbol{\alpha}_0) \xrightarrow{d} N(0, \frac{1}{N}\Sigma). \tag{19}$$

By the continuous mapping theorem,

$$\sqrt{T}\left(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\right) \xrightarrow{d} N\left(0, \frac{1}{N}H^{-1}\Sigma H^{-1}\right).$$

□

A.2 Inference

Theorem 4.1. *Assume* ***A1-A11***, ***I1-I3****, (i)* $E|g_{itk}(\boldsymbol{\alpha})^2|^{(r+\delta)} < \Delta < \infty$ *for all* i, t, k *and for some* $\delta > 0$*, for* $\boldsymbol{\alpha}$ *in some neighborhood of* $\boldsymbol{\alpha}_0$*; (ii)* $E\left[\frac{1}{T}\sum_{t=1}^{T} g_{tk}^i(\boldsymbol{\alpha}_0)g_{tk}^i(\boldsymbol{\alpha}_0)'\right]$ *is uniformly positive definite. Then,* S *and* $S^{(d)}$ *converge to the same distribution*

Proof. I suppress the dependence of the objective function on $\boldsymbol{w}$ and model the synthetic controls as a function of the treatment effect parameters. Let

$$\widehat{\mathcal{L}}(\boldsymbol{\alpha}) = \widehat{Q}(\boldsymbol{\alpha}) + a(\boldsymbol{\alpha})'\boldsymbol{\gamma}, \tag{20}$$

where $\boldsymbol{\gamma}$ is an $L \times 1$ vector of Lagrange multipliers and $\widehat{Q}(\boldsymbol{\alpha}) \equiv -\frac{1}{2NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Big(Y_{it} - \boldsymbol{D}_{it}'\boldsymbol{\alpha} -$

$\Gamma(\boldsymbol{\alpha})\Big)^2$. The constrained estimator maximizes equation (20), solving the following set of equations:

$$\begin{pmatrix} \sqrt{NT}\nabla_{\boldsymbol{\alpha}}\widehat{Q}(\tilde{\boldsymbol{\alpha}}) + \nabla_{\boldsymbol{\alpha}} a(\tilde{\boldsymbol{\alpha}})'\sqrt{NT}\tilde{\boldsymbol{\gamma}} \\ \sqrt{NT}a(\tilde{\boldsymbol{\alpha}}) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

Note first that under the given conditions $\tilde{\boldsymbol{\alpha}} \xrightarrow{p} \boldsymbol{\alpha}_0$. The same conditions discussed for $\widehat{Q}(\boldsymbol{\alpha})$ for Theorem 3.2 above also apply to $\widehat{\mathcal{L}}(\boldsymbol{\alpha})$ under **I1**. Consistency again follows from Theorem 2.1 of Newey and McFadden (1994).

Next, I show that $\sqrt{NT}\,(\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0)$ converges to a normally distributed random variable. A mean value expansion around $\boldsymbol{\alpha}_0$ gives

$$\sqrt{NT}\nabla_{\boldsymbol{\alpha}}\widehat{Q}(\tilde{\boldsymbol{\alpha}}) = \sqrt{NT}\nabla_{\boldsymbol{\alpha}}\widehat{Q}(\boldsymbol{\alpha}_0) + \nabla_{\boldsymbol{\alpha}\boldsymbol{\alpha}}\widehat{Q}(\boldsymbol{\alpha}^*)\sqrt{NT}(\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0), \tag{21}$$

where $\boldsymbol{\alpha}^*$ represents a vector of points between $\tilde{\boldsymbol{\alpha}}$ and $\boldsymbol{\alpha}_0$. Then, by consistency of $\tilde{\boldsymbol{\alpha}}$,

$$\begin{bmatrix} \sqrt{NT}\nabla_{\boldsymbol{\alpha}}\widehat{Q}(\boldsymbol{\alpha}_0) \\ 0 \end{bmatrix} + \begin{bmatrix} H & A' \\ A & 0 \end{bmatrix} \begin{bmatrix} \sqrt{NT}\,(\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) \\ \sqrt{NT}\tilde{\boldsymbol{\gamma}} \end{bmatrix} + o_p = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

$$\begin{bmatrix} \sqrt{NT}\,(\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) \\ \sqrt{NT}\tilde{\boldsymbol{\gamma}} \end{bmatrix} = -\begin{bmatrix} H^{-1} - H^{-1}A'(AH^{-1}A')^{-1}AH^{-1} & H^{-1}A'(AH^{-1}A')^{-1} \\ (AH^{-1}A')^{-1}AH^{-1} & -(AH^{-1}A')^{-1} \end{bmatrix} \begin{bmatrix} \sqrt{NT}\nabla_{\boldsymbol{\alpha}}\widehat{Q}(\boldsymbol{\alpha}_0) \\ 0 \end{bmatrix} + o_p.$$

Thus, $\sqrt{NT}(\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0)$ converges to a normally-distributed random variable given the result in equation (19). Since N is fixed, $\sqrt{T}(\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0)$ also converges to a normally-distributed random variable.

A mean value expansion for unit i specifically yields

$$\sqrt{T}\nabla_{\boldsymbol{\alpha}}\widehat{Q}_i(\tilde{\boldsymbol{\alpha}}) = \sqrt{T}\nabla_{\boldsymbol{\alpha}}\widehat{Q}_i(\boldsymbol{\alpha}_0) + \nabla_{\boldsymbol{\alpha}\boldsymbol{\alpha}}\widehat{Q}_i(\boldsymbol{\alpha}^*)\sqrt{T}(\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0),$$

$\sqrt{T}\nabla_{\boldsymbol{\alpha}}\widehat{Q}_i(\boldsymbol{\alpha}_0)$ converges to a normally-distributed variable (by **A4** and **A9**). By a uniform law of large numbers, $\sup_{\boldsymbol{b}\in\Theta}\|\nabla_{\boldsymbol{\alpha}\boldsymbol{\alpha}}\widehat{Q}_i(\boldsymbol{b}) - H_i(\boldsymbol{b})\| \xrightarrow{p} 0$. Thus, $\nabla_{\boldsymbol{\alpha}\boldsymbol{\alpha}}\widehat{Q}_i(\boldsymbol{\alpha}^*)$ converges to H_i by consistency of $\tilde{\boldsymbol{\alpha}}$. Convergence of $\sqrt{T}(\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0)$ to a normally-distributed variable was shown above. Thus, by definition of $s_i(\tilde{\boldsymbol{\alpha}})$,

$$\sqrt{T} s_i(\tilde{\boldsymbol{\alpha}}) \xrightarrow{d} N(0, \tilde{V}_i),$$

for (unspecified) asymptotic variance $\tilde{V}_i$. By condition **I3**,

$$\sqrt{T} W_i s_i(\tilde{\boldsymbol{\alpha}}) \xrightarrow{d} N(0, \tilde{V}_i),$$

By (i) and **A4**, the law of large numbers implies

$$\frac{1}{T}\sum_{t=1}^{T} g_{tk}^i(\boldsymbol{\alpha}_0) g_{tk}^i(\boldsymbol{\alpha}_0)' \xrightarrow{p} E\left[\frac{1}{T}\sum_{t=1}^{T} g_{tk}^i(\boldsymbol{\alpha}_0) g_{tk}^i(\boldsymbol{\alpha}_0)'\right]. \tag{22}$$

By condition **A10**, $g_{itk}(\boldsymbol{\alpha})$ is continuous at $\boldsymbol{\alpha}_0$ with probability one for all i, t, k. Then, by consistency of $\tilde{\boldsymbol{\alpha}}$ and equation (22),

$$\frac{1}{T}\sum_{t=1}^{T} g_{tk}^i(\tilde{\boldsymbol{\alpha}}) g_{tk}^i(\tilde{\boldsymbol{\alpha}})' \xrightarrow{p} E\left[\frac{1}{T}\sum_{t=1}^{T} g_{tk}^i(\boldsymbol{\alpha}_0) g_{tk}^i(\boldsymbol{\alpha}_0)'\right].$$

This result holds by Lemma 4.3 in Newey and McFadden (1994). Similarly,

$$\frac{1}{T}\sum_{t=1}^{T} g_{tk}^i(\tilde{\boldsymbol{\alpha}}) g_{itk}(\tilde{\boldsymbol{\alpha}}) \xrightarrow{p} E\left[\frac{1}{T}\sum_{t=1}^{T} g_{tk}^i(\boldsymbol{\alpha}_0) g_{itk}(\boldsymbol{\alpha}_0)\right]$$

By assumption (ii) and the definition of g_{itk} in **I2**,

$$\left[\frac{1}{T}\sum_{t=1}^{T} g_{tk}^i(\tilde{\boldsymbol{\alpha}}) g_{tk}^i(\tilde{\boldsymbol{\alpha}})'\right]^{-1} \left[\frac{1}{T}\sum_{t=1}^{T} g_{tk}^i(\tilde{\boldsymbol{\alpha}}) g_{itk}(\tilde{\boldsymbol{\alpha}})\right]$$

$$\xrightarrow{p} \quad \boldsymbol{c}_k^i + E\left[\frac{1}{T}\sum_{t=1}^{T} g_{tk}^i(\boldsymbol{\alpha}_0) g_{tk}^i(\boldsymbol{\alpha}_0)'\right]^{-1} E\left[\frac{1}{T}\sum_{t=1}^{T} g_{tk}^i(\boldsymbol{\alpha}_0)\mu_{itk}\right] = \boldsymbol{c}_k^i$$

Thus, $\hat{\boldsymbol{c}}_k^i \xrightarrow{p} \boldsymbol{c}_k^i$. Then, it is straightforward to show that $\widehat{\mu}_{itk} \xrightarrow{p} \mu_{itk}$ such that

$$\text{Cov}\left(\frac{1}{T}\sum_{s=1}^{T}\widehat{\mu}_{isk}, \frac{1}{T}\sum_{s=1}^{T}\widehat{\mu}_{jsk'}\right) \xrightarrow{p} E\left[\left(\frac{1}{T}\sum_{s=1}^{T}\mu_{isk}\right)\left(\frac{1}{T}\sum_{s=1}^{T}\mu_{jsk'}\right)\right] = 0 \text{ for all } k, k'.$$

Thus, $s_i(\tilde{\boldsymbol{\alpha}})$ and $s_j(\tilde{\boldsymbol{\alpha}})$ are asymptotically uncorrelated and the weights in **I3** preserve the asymptotic variance such that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sqrt{T}s_i(\tilde{\boldsymbol{\alpha}}) \xrightarrow{d} N(0, \frac{1}{N}\sum_{i=1}^{N}\tilde{V}_i),$$
$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sqrt{T}W_i s_i(\tilde{\boldsymbol{\alpha}}) \xrightarrow{d} N(0, \frac{1}{N}\sum_{i=1}^{N}\tilde{V}_i).$$

By the continuous mapping theorem, S and $S^{(d)}$ converge to the same distribution.

□

References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller, "Synthetic control methods for comparative case studies: Estimating the effect of Californias tobacco control program," *Journal of the American Statistical Association*, 2010, *105* (490).

_ , _ , and _ , "Synth: Stata module to implement synthetic control methods for comparative case studies," 2014.

_ , _ , and _ , "Comparative politics and the synthetic control method," *American Journal of Political Science*, 2015, *59* (2), 495–510.

_ , _ , _ et al., "Synth: An R Package for Synthetic Control Methods in Comparative Case Studies," *Journal of Statistical Software*, 2011, *42* (i13).

_ and Javier Gardeazabal, "The economic costs of conflict: A case study of the Basque Country," *American Economic Review*, 2003, pp. 113–132.

Adhikari, Bibek and James Alm, "Evaluating the Economic Effects of Flat Tax Reforms Using Synthetic Control Methods," *Southern Economic Journal*, 2016.

_ , Romain A Duval, Bingjie Hu, Prakash Loungani et al., "Can Reform Waves Turn the Tide? Some Case Studies Using the Synthetic Control Method," Technical Report, International Monetary Fund 2016.

Allegretto, Sylvia A, Arindrajit Dube, and Michael Reich, "Do minimum wages really reduce teen employment? Accounting for heterogeneity and selectivity in state panel data," *Industrial Relations: A Journal of Economy and Society*, 2011, *50* (2), 205–240.

Allegretto, Sylvia, Arindrajit Dube, Michael Reich, and Ben Zipperer, "Credible research designs for minimum wage studies: A response to Neumark, Salas, and Wascher," *ILR Review*, 2017, *70* (3), 559–592.

Ando, Michihito, "Dreams of urbanization: Quantitative case studies on the local impacts of nuclear power facilities using the synthetic control method," *Journal of Urban Economics*, 2015, *85*, 68–85.

_ and Fredrik Sävje, "Hypothesis Testing with the Synthetic Control Method," 2013.

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

Bai, Jushan, "Panel data models with interactive fixed effects," *Econometrica*, 2009, *77* (4), 1229–1279.

Bauhoff, Sebastian, "The effect of school district nutrition policies on dietary intake and overweight: A synthetic control approach," *Economics & Human Biology*, 2014, *12*, 45–55.

Bester, C Alan, Timothy G Conley, and Christian B Hansen, "Inference with dependent data using cluster covariance estimators," *Journal of Econometrics*, 2011, *165* (2), 137–151.

Billmeier, Andreas and Tommaso Nannicini, "Assessing economic liberalization episodes: A synthetic control approach," *Review of Economics and Statistics*, 2013, *95* (3), 983–1001.

Cavallo, Eduardo, Sebastian Galiani, Ilan Noy, and Juan Pantano, "Catastrophic natural disasters and economic growth," *Review of Economics and Statistics*, 2013, *95* (5), 1549–1561.

Clemens, Jeffrey, "The minimum wage and the Great Recession: Evidence from the Current Population Survey," Technical Report, National Bureau of Economic Research 2015.

— **and Michael Wither**, "The minimum wage and the Great Recession: Evidence of effects on the employment and income trajectories of low-skilled workers," Technical Report, National Bureau of Economic Research 2014.

Cunningham, Scott and Manisha Shah, "Decriminalizing indoor prostitution: Implications for sexual violence and public health," Technical Report, National Bureau of Economic Research 2014.

Dillender, Marcus O, Carolyn J Heinrich, and Susan N Houseman, "Health insurance reform and part-time work: Evidence from Massachusetts," *Labour Economics*, 2016.

Donohue, John, Abhay Aneja, and Kyle D. Weber, "Right-to-Carry Laws and Violent Crime: A Comprehensive Assessment using Panel Data and a State-Level Synthetic Control Analysis," Technical Report, National Bureau of Economic Research 2017.

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

— **, T William Lester, and Michael Reich**, "Minimum wage effects across state borders: Estimates using contiguous counties," *Review of Economics and Statistics*, 2010, *92* (4), 945–964.

Eren, Ozkan and Serkan Ozbeklik, "What Do Right-to-Work Laws Do? Evidence from a Synthetic Control Method Analysis," *Journal of Policy Analysis and Management*, 2016, *35* (1), 173–194.

Firpo, Sergio and Vítor Possebom, "Synthetic Control Estimator: A Generalized Inference Procedure and Confidence Sets," *Unpublished manuscript*, 2016.

Fletcher, Jason M, David E Frisvold, and Nathan Tefft, "Non-linear effects of soda taxes on consumption and weight outcomes," *Health Economics*, 2014.

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

Grier, Kevin and Norman Maynard, "The economic consequences of Hugo Chavez: A synthetic control analysis," *Journal of Economic Behavior & Organization*, 2016, *125*, 1–21.

Hackmann, Martin B, Jonathan T Kolstad, and Amanda E Kowalski, "Adverse selection and an individual mandate: When theory meets practice," *The American Economic Review*, 2015, *105* (3), 1030–1066.

Hansen, Christian B, "Asymptotic properties of a robust variance matrix estimator for panel data when T is large," *Journal of Econometrics*, 2007, *141* (2), 597–620.

Hinrichs, Peter, "The effects of affirmative action bans on college enrollment, educational attainment, and the demographic composition of universities," *Review of Economics and Statistics*, 2012, *94* (3), 712–722.

Ibragimov, Rustam and Ulrich K Müller, "t-Statistic based correlation and heterogeneity robust inference," *Journal of Business & Economic Statistics*, 2010, *28* (4), 453–468.

Kaul, Ashok, Stefan Klößner, Gregor Pfeifer, and Manuel Schieler, "Synthetic Control Methods: Never Use All Pre-Intervention Outcomes as Economic Predictors," 2015.

Kreif, Noémi, Richard Grieve, Dominik Hangartner, Alex James Turner, Silviya Nikolova, and Matt Sutton, "Examination of the synthetic control method for evaluating health policies with multiple treated units," *Health Economics*, 2015.

Mideksa, Torben K, "The economic impact of natural resources," *Journal of Environmental Economics and Management*, 2013, *65* (2), 277–289.

Montalvo, Jose G, "Voting after the bombings: A natural experiment on the effect of terrorist attacks on democratic elections," *Review of Economics and Statistics*, 2011, *93* (4), 1146–1154.

Munasib, Abdul and Dan S Rickman, "Regional economic impacts of the shale gas and tight oil boom: A synthetic control analysis," *Regional Science and Urban Economics*, 2015, *50*, 1–17.

_ **and Mouhcine Guettabi**, "Florida Stand Your Ground Law and Crime: Did It Make Floridians More Trigger Happy?," *Available at SSRN 2315295*, 2013.

Nannicini, Tommaso and Andreas Billmeier, "Economies in transition: How important is trade openness for growth?," *Oxford Bulletin of Economics and Statistics*, 2011, *73* (3), 287–314.

Neumark, David, JM Ian Salas, and William Wascher, "More on recent evidence on the effects of minimum wages in the United States," *IZA Journal of Labor Policy*, 2014, *3* (1), 24.

—, —, and —, "Revisiting the Minimum Wage–Employment Debate: Throwing Out the Baby with the Bathwater?," *Industrial & Labor Relations Review*, 2014, *67* (3 suppl), 608–648.

Newey, Whitney K. and Daniel McFadden, "Large sample estimation and hypothesis testing," *Handbook of Econometrics*, 1994, *4*, 2111–2245.

Nonnemaker, James, Mark Engelen, and Daniel Shive, "Are methamphetamine precursor control laws effective tools to fight the methamphetamine epidemic?," *Health Economics*, 2011, *20* (5), 519–531.

Pinotti, Paolo, "The economic costs of organised crime: Evidence from Southern Italy," *The Economic Journal*, 2015, *125* (586), F203–F232.

Powell, David, "Imperfect Synthetic Controls: Did the Massachusetts Health Care Reform Save Lives?," 2017. `https://sites.google.com/site/davidmatthewpowell/imperfect-synthetic-controls`.

—, "Panel Data Inference with Dependent Clusters," 2017.

—, Rosalie Liccardo Pacula, and Mireille Jacobson, "Do Medical Marijuana Laws Reduce Addictions and Deaths Related to Pain Killers?," Working Paper 21345, National Bureau of Economic Research July 2015.

Quast, Troy and Fidel Gonzalez, "Sex Work Regulation and Sexually Transmitted Infections in Tijuana, Mexico," *Health Economics*, 2016.

Robbins, Michael W, Jessica Saunders, and Beau Kilmer, "A framework for synthetic control methods with high-dimensional, micro-level data: Evaluating a neighborhood-specific crime intervention," *Journal of the American Statistical Association*, 2016, (just-accepted).

Sabia, Joseph J, Richard V Burkhauser, and Benjamin Hansen, "Are the effects of minimum wage increases always small? New evidence from a case study of New York state," *Industrial & Labor Relations Review*, 2012, *65* (2), 350–376.

Saunders, Jessica, Russell Lundberg, Anthony A Braga, Greg Ridgeway, and Jeremy Miles, "A Synthetic Control Approach to Evaluating Place-Based Crime Interventions," *Journal of Quantitative Criminology*, 2014, pp. 1–22.

Torsvik, Gaute and Kjell Vaage, "Gatekeeping versus monitoring: Evidence from a case with extended self-reporting of sickness absence," 2014.

Webb, Matthew D, "Reworking wild bootstrap based inference for clustered errors," Technical Report, Queen's Economics Department Working Paper 2014.

White, Halbert and Ian Domowitz, "Nonlinear regression with dependent observations," *Econometrica: Journal of the Econometric Society*, 1984, pp. 143–161.

Zhang, Jun, Chunping Zhong, and Ming Yi, "Did Olympic Games improve air quality in Beijing? Based on the synthetic control method," *Environmental Economics and Policy Studies*, 2016, *18* (1), 21–39.