Journal of Business & Economic Statistics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ubes20

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

David Powell

JOURNAL OF BUSINESS & ECONOMIC STATISTICS
2021, VOL. 00, NO. 0, 1–13
https://doi.org/10.1080/07350015.2021.1927743

Taylor & Francis
Taylor & Francis Group

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

David Powell

RAND, Santa Monica, CA

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 fixed effects for comparative case studies in which a treated unit adopts a single policy. This article generalizes the synthetic control estimator to estimate parameters associated with multiple discrete or continuous explanatory variables, jointly estimating the parameters and synthetic controls for each unit. I apply the estimator to study the disemployment effects of the minimum wage, estimating that increases in the minimum wage reduce employment.

ARTICLE HISTORY
Received August 2019
Accepted April 2021

KEYWORDS
Finite inference; Interactive fixed effects; Minimum wage; Synthetic control estimation; Teen employment

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 explanatory variables to identify relationships with the outcome of interest. These additive (or two-way) fixed-effects models include separate unit ("state") fixed effects and time fixed effects with the assumption that an unweighted average of all units acts as a valid counterfactual for each unit. In this article, 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 the parameters associated with the explanatory variables even in the presence of nonparallel trends and systematic shocks. This estimator builds on work by Abadie and Gardeazabal (2003), Abadie, Diamond, and Hainmueller (2010), and Abadie, Diamond, and Hainmueller (2015).

This previous work introduced a technique to estimate causal effects for comparative case studies and I refer to this method as the "comparative case study synthetic control" (CSSC) estimator. The innovation of this approach is to use the period before treatment to construct a comparison group which is most appropriate for the group exposed to the treatment. The CSSC estimator has been called "[a]rguably the most important innovation in the evaluation literature in the last 15 years" (Athey and Imbens 2017). A typical motivation for the synthetic control estimator is that the "parallel trends" assumption required by additive fixed-effects models often does not hold in practice. CSSC potentially permits causal inference even in these cases. This feature should be beneficial beyond comparative case studies.

The estimator introduced in this article generalizes the CSSC estimator, allowing for the inclusion of one or more discrete or continuous explanatory variables in the model. The CSSC estimator has been used in circumstances where multiple units are treated (e.g., Billmeier and Nannicini 2013; Acemoglu et al. 2016; Donohue, Aneja, and Weber 2019) by applying the CSSC estimator to each treated unit and then aggregating the estimates. However, it is difficult to apply CSSC to settings in which there are not clear *treated* and *never-treated* groups, such as the minimum wage application of this article. Every state has a minimum wage and state-level minimum wages are regularly changing over time. In this article, I extend the synthetic control approach beyond applications in which the treatment variable is represented by a single indicator variable. I refer to this estimator as the generalized synthetic control (GSC) estimator. GSC parallels additive fixed-effects estimators which permit joint estimation of the relationships between the outcome and multiple explanatory variables. However, instead of assuming the outcome variable is a function of additive state and time fixed effects, GSC allows for more flexible and systematic outcome trajectories.

CSSC estimates a treatment effect for one unit in each time period, placing no restrictions on treatment effect heterogeneity. GSC, in many contexts, must restrict parameter heterogeneity to some extent in a manner similar to many two-way fixed-effects models (in which parameter homogeneity is typically assumed). I develop GSC with a focus on permitting unit-specific parameter heterogeneity. Gibbons, Serrato, and Urbancic (2018) considered unit-specific treatment heterogeneity and show that typical two-way fixed-effects models do not provide consistent estimates of the average treatment effect (ATE). Słoczyński (2020) makes a related point for OLS estimation more generally. An emerging literature discusses potential problems with two-way fixed-effects models given heterogeneous treatment effects and staggered adoption (Borusyak and Jaravel 2017; Athey and

CONTACT David Powell dpowell@rand.org RAND, 1776 Main St., Santa Monica, CA 90402.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/UBES.

Imbens 2018; Goodman-Bacon 2018; Callaway and Sant'Anna 2020; De Chaisemartin and D'Haultfœuille 2020a,b; Sun and Abraham 2020). There is less work analyzing how two-way fixed-effects models operate given multiple and/or nonbinary explanatory variables with slope heterogeneity.[1] Given that the traditional synthetic control framework does not require restrictions on treatment effect heterogeneity, I dedicate most of the discussion and development of GSC to permitting some parameter heterogeneity while acknowledging that some tradeoffs are necessary.

Since the estimator uses other units as controls for each unit, it mechanically induces correlations across units. This article suggests an inference procedure for synthetic control estimation given this dependence. The inference procedure will adjust for cross-unit dependence to calculate appropriate p-values.

In the next section, I further discuss the CSSC estimator to motivate the gains of a more general synthetic control estimation technique and include background on the minimum wage debate in the economics literature. In Section 3, I introduce the generalized synthetic control estimator. I briefly 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. CSSC Estimation

This article builds on the synthetic control estimation technique discussed in Abadie and Gardeazabal (2003), Abadie, Diamond, and Hainmueller (2010), and Abadie, Diamond, and Hainmueller (2015). In the CSSC framework, there are n 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

$$Y_{it}^N = \boldsymbol{\lambda}_t \boldsymbol{\mu}_i + \epsilon_{it}, \tag{1}$$

$$Y_{it} = Y_{it}^N + \alpha_{it} D_{it},$$

where D_{it} represents the treatment variable and is equal to 1 for unit $i = 1$ and time periods $t > T_0$, 0 otherwise. Y_{it}^N represents the untreated outcome. $\boldsymbol{\lambda}_t$ is a $1 \times F$ vector of common unobserved factors and $\boldsymbol{\mu}_i$ is a $F \times 1$ vector of factor loadings.

Unit 1 is the treated unit while all other units are part of the "donor pool." The purpose of the approach is to find a weighted combination of units (which exists by assumption) in the donor pool such that for all $t \leq T_0$:

$$\sum_{j=2}^{n} w_j Y_{jt}^N = Y_{1t}^N, \tag{2}$$

where w_j represents the weight on unit j.

The weights are constrained to be nonnegative and to sum to one to avoid extrapolation. They are generated to minimize the difference between the pretreated outcomes of the treated unit and the unit's synthetic control. To facilitate comparisons with GSC, I assume that there are no other control variables, which are included additively (with no unit-specific parameter heterogeneity) in Abadie and Gardeazabal (2003) and Abadie, Diamond, and Hainmueller (2010), and that all pretreatment outcomes (and only pretreatment outcomes) are used to generate the synthetic control weights. This is common in applications using CSSC (see, e.g., Cavallo et al. 2013).[2] The GSC estimator will use control variables in a manner that parallels traditional regression analysis by letting them have direct effects on the outcome variable and estimating those effects jointly with the other parameters. I discuss this issue further in Section 3.3.4.

Let $\mathbf{W} = (w_2, \ldots, w_n)'$ such that $w_j \geq 0$ for all $2 \leq j \leq n$ and $\sum_{j=2}^{n} w_j = 1$. The weights are estimated using

$$\widehat{\mathbf{W}} = \operatorname{argmin}_{\boldsymbol{\phi}_j} \sum_{t=1}^{T_0} \left(Y_{1t} - \sum_{j=2}^{n} \phi_j Y_{jt}\right)^2$$

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

The CSSC estimator involves a constrained optimization to find the weighted average of the donor units which is "closest" to the treated unit in terms of pretreated outcomes. The treatment effect estimate for period $t > T_0$ is $\hat{\alpha}_{1t} = Y_{1t} - \sum_{j=2}^{n} \hat{w}_j Y_{jt}$. CSSC permits estimation of a treatment effect for each posttreatment time period, though it is common to report an aggregate posttreatment effect.

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

$$\boldsymbol{\lambda}_t = \begin{bmatrix} 1 & \eta_t \end{bmatrix} \quad \text{and} \quad \boldsymbol{\mu}_i = \begin{bmatrix} \gamma_i \\ 1 \end{bmatrix}, \tag{3}$$

such that $\boldsymbol{\lambda}_t \boldsymbol{\mu}_i = \gamma_i + \eta_t$. A primary contribution of this article is that the motivation behind CSSC should apply more generally to panel data models.

However, there are costs to the Equation (1) framework. For example, two-way fixed-effects models do not require equation (2). An emerging literature seeks to relax or address the CSSC assumptions imposed in Equation (2) including Ben-Michael, Feller, and Rothstein (2018), Arkhangelsky et al. (2020), Xu (2017), Amjad, Shah, and Shen (2018), Ferman and Pinto (2018), and Powell (2020).[3] This article maintains the original Abadie, Diamond, and Hainmueller (2010) "convex hull" assumption (but applied to multiple units). As this new literature finds further ways to relax the CSSC assumption, it should be straightforward to relax the corresponding GSC assumption.

[1] With some exceptions, most of the recent staggered adoption literature assumes no covariates. See Callaway and Sant'Anna (2020) for a discussion of this issue.

[2] Kaul et al. (2021) noted that the CSSC estimator will not use the covariates when each pretreatment outcome is used to generate the synthetic control weights.

[3] See Hollingsworth and Wing (2020) for implementation of a synthetic control-type estimator given high-dimensional data.

Gobillon and Magnac (2016) compared the synthetic control approach to panel data models with interactive fixed effects (e.g., Bai 2009; Ahn, Lee, and Schmidt 2013). Both approaches are useful for relaxing the restriction that the fixed effects are additive. The synthetic control method requires the treated unit to be (approximately) in the convex hull of the other outcomes and avoids extrapolating out of sample. Interactive effect models often require that the number of factors is known or estimated (Bai and Ng 2002),[4] a requirement not imposed by synthetic control estimation.

2.2. Minimum Wage and Employment

A vast literature has debated whether minimum wage increases affect employment rates.[5] The empirical literature has recently focused on determining appropriate controls for states with increasing minimum wages.[6] The potential of the CSSC estimator to generate appropriate control units has been recognized recently in the literature, though it is difficult to apply CSSC estimation to this application.

Dube and Zipperer (2015) (also discussed in Allegretto et al. 2017) used CSSC estimation to study minimum wage employment effects.[7] The authors selected on states with no minimum wage changes for two years prior to a minimum wage increase and with at least one year of posttreatment 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. 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. These selection criteria are unnecessary using the GSC estimator introduced below.

Moreover, Dube and Zipperer (2015) created 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 state using the CSSC estimator in this manner is only valid if the minimum wage does not affect employment.

3. GSC Estimator

This section develops the GSC estimator. I follow the notation developed for the CSSC method and used in Section 2.1 as closely as possible. I assume that there are (fixed) n units and (large) T time periods. I model outcomes as

$$Y_{it}^N = \boldsymbol{\lambda}_t \boldsymbol{\mu}_i + \epsilon_{it}, \qquad (4)$$
$$Y_{it} = Y_{it}^N + \mathbf{D}_{it}' \boldsymbol{\alpha}_i,$$

where $\mathbf{D}_{it}$ is a $K \times 1$ vector of (continuous and/or discrete) explanatory variables for unit i at time period t such that $K \times n$ is small. While CSSC differentiates between the primary policy variable of interest and covariates, I will not make such a distinction, similar to many additive fixed-effect models. However, Sections 3.3.3 and 3.3.4 consider possible benefits of categorizing the explanatory variables.

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. Additive fixed-effects models are nested in this framework. However, it is worth noting that I assume large T throughout this article and a convex hull assumption, assumptions not required in two-way fixed-effects models.[8]

A primary difference between Equation (4) and the specification represented in Equation (1) is that $\mathbf{D}_{it}$ is a vector and not limited to one indicator variable. The tradeoff is that the parameters cannot have unlimited heterogeneity in many cases covered by this model. In the setup above, I permit the parameters to vary by unit, similar to the additive fixed-effects model with slope heterogeneity discussed in Gibbons, Serrato, and Urbancic (2018), Wooldridge (2005), Juhl and Lugovskyy (2014), Blomquist and Westerlund (2013), Pesaran and Yamagata (2008), and Campello, Galvao, and Juhl (2019). Simplifying assumptions, such as $\boldsymbol{\alpha}_i \equiv \boldsymbol{\alpha}$, are possible and may be useful, especially when K is large (see Section 3.3.3). I will also discuss models with additional heterogeneity (see Section 3.3.2).

Most empirical applications implementing two-way fixed-effects models do not consider slope heterogeneity. My baseline model includes unit-specific slope heterogeneity as a way to highlight the tradeoff within synthetic control models between parameter heterogeneity and the specifications which can be estimated. This parameter heterogeneity may also be independently important in these models (and in panel data models more broadly).

The estimator involves simultaneous estimation of a synthetic control for each unit. Let $\boldsymbol{w}_i = (w_1^i, \ldots, w_{i-1}^i, w_{i+1}^i, \ldots, w_n^i)$, where $\boldsymbol{w}_i$ is the vector of weights on all other units to generate the synthetic control for unit i; w_j^i represents the weight given unit j for the creation of the synthetic control for unit i. The weights are constrained, as before, to be nonnegative and to sum to one. I define the set of possible weighting vectors to create the unit i synthetic control by $\mathcal{W}_i = \left\{ \boldsymbol{w}_i \;\middle|\; \sum_{j \neq i} w_j^i = 1, w_j^i \geq 0 \text{ for all } j \right\}$.

3.1. Estimation

I begin by discussing the estimation procedure. GSC simultaneously estimates the parameters associated with each explanatory

[4]Moon and Weidner (2015) suggested that, under certain assumptions, including "too many" interactive fixed effects is not problematic.

[5]See Neumark, Salas, and Wascher (2014), Cengiz et al. (2019), Neumark (2019), Manning (2020), and Neumark and Shirley (2021) for discussions and reviews of the literature.

[6]Recent work (Clemens and Wither 2019; Clemens and Strain 2018; 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. Previous work has used contiguous counties as controls (Dube, Lester, and Reich 2010). Finally, Cengiz et al. (2019) used all sources of variation but study changes in employment changes by wage bins to identify the impact on low-wage employment.

[7]Sabia, Burkhauser, and Hansen (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.

[8]When T is small, it may be difficult to use the longitudinal nature of the data to learn about underlying unit-specific outcome trajectories so assumptions, such as the "parallel trends" assumption in many panel data designs, must be imposed.

variable while constructing a synthetic control for each unit. The synthetic control is an estimate of $Y_{it}-\boldsymbol{D}'_{it}\hat{\boldsymbol{\alpha}}_i$ using a constrained weighted average of $(Y_{jt}-\boldsymbol{D}'_{jt}\hat{\boldsymbol{\alpha}}_j)$ for all $j \neq i$. I refer to $Y_{it}-\boldsymbol{D}'_{it}\hat{\boldsymbol{\alpha}}_i$ as the estimated "untreated outcome" for unit i, the outcome that would be observed if all explanatory variables were equal to zero.[9]

3.1.1. Implementation

The estimator chooses estimates for $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n)$ and jointly constructs a synthetic control for each unit: $(\hat{\boldsymbol{\alpha}}_1, \ldots, \hat{\boldsymbol{\alpha}}_n, \hat{\boldsymbol{w}}_1, \ldots, \hat{\boldsymbol{w}}_n) =$

$$\operatorname{argmin}_{\boldsymbol{b}_1,\ldots,\boldsymbol{b}_n,\boldsymbol{\phi}_1\in\mathcal{W}_1,\ldots,\boldsymbol{\phi}_n\in\mathcal{W}_n} \left\{ \frac{1}{2nT} \sum_{i=1}^{n} \hat{\Omega}_i^{-1} \sum_{t=1}^{T} \left[Y_{it} - \boldsymbol{D}'_{it}\boldsymbol{b}_i - \sum_{j\neq i} \left(\phi_j^i \left(Y_{jt} - \boldsymbol{D}'_{jt}\boldsymbol{b}_j \right) \right) \right]^2 \right\}. \tag{5}$$

$\sum_{j\neq i} \left(\hat{w}_j^i \left(Y_{jt} - \boldsymbol{D}'_{jt}\hat{\boldsymbol{\alpha}}_j \right) \right)$ is used as a control for unit i. It would be inappropriate to construct a synthetic control based solely on Y_{it}. Instead, the GSC estimator uses the "untreated outcomes" as possible controls. $\hat{\Omega}$ is a weighting matrix discussed in the next section.

Given the large number of parameters that need to be estimated, I use Nelder-Mead optimization to minimize the objective function when generating all simulation and application results.

3.1.2. Weighting Matrix

With CSSC estimation, it is customary to visually check the fit of the synthetic control with the treated unit in the pretreatment 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 GSC, the parallel concern is that the convex hull assumption may not hold for a specific unit or units. I employ a two-step procedure which places more weight on units with better synthetic control fits. The gains of this approach are discussed later. The estimation steps are as follows:

1. For each unit i,
$$\hat{\Omega}_i \equiv \min_{\boldsymbol{b},\boldsymbol{\phi}_i\in\mathcal{W}_i} \left\{ \frac{1}{2T} \sum_{t=1}^{T} \left[Y_{it} - \boldsymbol{D}'_{it}\boldsymbol{b}_i - \sum_{j\neq i} \left(\phi_j^i \left(Y_{jt} - \boldsymbol{D}'_{jt}\boldsymbol{b}_j \right) \right) \right]^2 \right\}. \tag{6}$$
2. Minimize the weighted sum of squared residuals using Equation (5).

In the first step, I estimate the parameters using only one unit at a time[10] to estimate a measure of fit by unit.[11] The method proposed in this section places more weight on units with better synthetic control fits. Alternate weighting schemes, such as excluding[12] units with synthetic control fits below a certain threshold, are also possible.

3.1.3. Average of Unit-Specific Estimates

The proposed estimation procedure estimates unit-specific parameters. To summarize the estimates, I define the average estimates (AE) as

$$\boldsymbol{\alpha}^{\text{AE}} \equiv \sum_i \pi_i \boldsymbol{\alpha}_i, \tag{7}$$

where π_i represents population frequency. In the case of a binary treatment, $\boldsymbol{\alpha}^{\text{AE}}$ is often interpreted as an average treatment effect. For continuous variables, the parameters represent marginal effects, the average (across units) relationships between the outcome and an increase of 1 in each explanatory variable.

3.2. Main Assumptions

In this section, I discuss assumptions implying identification of $\boldsymbol{\alpha}$. 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.

A1. Outcomes: $Y_{it}^N = \boldsymbol{\lambda}_t\boldsymbol{\mu}_i + \epsilon_{it}; \quad Y_{it} = Y_{it}^N + \mathbf{D}'_{it}\boldsymbol{\alpha}_i.$

A2. Convex hull: For each i, there exists either (a) $\boldsymbol{w}_i \in \mathcal{W}_i$ such that $\sum_{j\neq i} w_j^i Y_{jt}^N = Y_{it}^N$ or (b) $\boldsymbol{w}_k \in \mathcal{W}_k$ with $w_i^k > 0$ for some k such that $\sum_{j\neq k} w_j^k Y_{jt}^N = Y_{kt}^N$.

A3. Rank condition:
$$E\begin{bmatrix} \mathbf{D}'_{11} \cdots \mathbf{D}'_{n1} & \sum_{j\neq i} \phi_j^i Y_{i1} \\ \vdots & \vdots \\ \mathbf{D}'_{1T} \cdots \mathbf{D}'_{nT} & \sum_{j\neq i} \phi_j^i Y_{iT} \end{bmatrix}$$
is rank $K \times n + 1$ for all $\boldsymbol{\phi}_i \in \mathcal{W}_i$.

A4. Weighting matrix: For all i and j, $\frac{\Omega_i}{\Omega_j} < \infty$.

A1 defines the outcome and "untreated" outcome. *A2* is the "convex hull" assumption and is similar to the CSSC assumption expressed in Equation (2). In the CSSC context, part (b) of *A2* relaxes the convex hull assumption (i.e., Equation (2)). If unit 1 outcomes are outside of the convex hull of the outcomes of the other units, then it is not possible to estimate causal effects using CSSC without extrapolation.[13] However, GSC constructs synthetic controls for all units. Unit 1 might be part of the synthetic control for unit k (i.e., $w_1^k > 0$). If unit k's outcomes are within the convex hull of the outcomes of the other units, then it may be possible to estimate $\boldsymbol{\alpha}_1$ using only unit k and its synthetic control since the synthetic control includes unit 1.[14]

[9]The important characteristic is to think about outcomes across units when evaluated at the same $\boldsymbol{D} = \boldsymbol{d}$ since there are cases in which we may not consider $\boldsymbol{d} = \boldsymbol{0}$ as the untreated case. This normalization is not critical as the estimation would be unaffected by evaluating all observations at the same $\boldsymbol{d}$.

[10]The benefit of permitting such flexibility is that a poor synthetic control fit for one unit may affect the (preliminary) parameter estimates 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.

[11]It is possible that $\boldsymbol{\alpha}$ is not identified for a unit by itself. However, this will not cause problems in estimating the variance for that unit.

[12]The excluded units can (and should) still be used to create synthetic controls for other units. They are excluded from the first summation in Equation (5).

[13]Modifications of CSSC have considered ways to extrapolate outside of the convex hull to permit estimation (e.g., Doudchenko and Imbens (2016)).

[14]This insight is introduced in Powell (2020). In the CSSC context, the idea is that unit 1 is part of the synthetic control for unit k. At the time of policy

For GSC, *A2* is required to hold for all i while CSSC only imposes it for unit 1 so *A2* is more restrictive in this respect given that GSC must construct counterfactuals for all units. However, it is worth highlighting that GSC does not require that every unit have an appropriate synthetic control. Instead, the $\boldsymbol{\alpha}_i$ parameters can still be consistently estimated using unit k (and, in principle, *only* unit k) if condition *A2(b)* holds.

As discussed in Section 2.1, the CSSC assumption shown in equation (2) as well as the corresponding GSC assumption *A2* may seem restrictive.[15] An emerging literature has suggested approaches to relax the CSSC assumption for comparative case studies and those modifications can potentially be applied here.

A3 is a rank condition which assumes $T \geq K \times n + 1$, though this requirement can be relaxed.[16] *A3* assumes independent variation in each element of $\mathbf{D}_{it}$ across i. 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 explanatory variables cannot be replicated by a weighted average of the outcomes. As one test of *A3*, it is recommended to estimate the empirical analog of this matrix, plugging in $\hat{\boldsymbol{w}}$ for $\boldsymbol{\phi}$ and testing whether it is full rank.

A4 is a regularity condition governing the "fit" weighting matrix, where

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

3.3. Discussion of Model and Modifications

3.3.1. Relationship with Comparative Case Studies

The primary motivation of the GSC estimator is to extend synthetic control estimation, but it is instructive to discuss the usefulness of the GSC method in cases in which the CSSC estimator is typically used. First, it is often desirable to condition on additional covariates. CSSC assumes that the synthetic control weights also (approximately) satisfy $\boldsymbol{X}_1 = \sum_{j\neq 1}\hat{w}_j\boldsymbol{X}_j$ (see Abadie, Diamond, and Hainmueller 2010, eq. 2), where $\boldsymbol{X}$ represents other covariates. The unit's synthetic control then also accounts for any independent effects of the covariates by canceling out those covariates under an assumption that the relationship between the outcome and covariates does not vary across units. It may be difficult for the synthetic control weights to satisfy both Equation (2) and these additional conditions. For example, it may be critical to condition on another policy adopted by the treated unit. The convex hull assumption would require selecting only on units which also adopted this 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.

Arkhangelsky et al. (2020) introduced synthetic difference-in-differences (SDID), which includes regression adjustment. SDID involves estimating synthetic control weights both by unit and time, followed by estimating a two-way fixed-effects model using the product of those weights as regression weights and including any time-varying covariates in this last step. However, the synthetic counterfactuals do not take into account that the outcomes in the pretreatment period are impacted by the covariates. This means that the synthetic controls may not be appropriate since they were generated using "treated" outcomes.

Second, even in a traditional case study application with no covariates, CSSC and GSC have an important difference, which was discussed above in the context of *A2*. Note that it is straightforward in the GSC framework, for a comparative case study analysis, to allow for the treatment effect to vary by time, making the GSC and CSSC estimators equivalent with the exception of condition *A2*. GSC estimates synthetic controls for all units. These other units and their synthetic controls may provide additional information if the treated unit is itself part of the synthetic control of one or more of the other units. Consider a case in which pre-intervention outcomes of the treated unit are outside of the convex hull of the pre-intervention outcomes of the control units such that CSSC cannot create an appropriate synthetic control. 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 (5).

3.3.2. Heterogeneous Effects

The baseline model permits unit-specific parameter heterogeneity. Equation (4) allows for additional heterogeneity by interacting the explanatory variables accordingly. There are, however, limits in many contexts to how much heterogeneity can be explicitly modeled and estimation becomes increasingly difficult. In comparison, CSSC places no restrictions on treatment heterogeneity. Thus, the cost of extending synthetic control estimation to more contexts is that some sacrifices need to be made in terms of how much heterogeneity is permitted in the parameters. This tradeoff is not unique to synthetic control estimation and often imposed for additive fixed-effects models (and many other estimators).

To permit additional parameter heterogeneity, condition *A1* can be modified in the following way (let $\boldsymbol{D} \equiv (\boldsymbol{D}_1, \ldots, \boldsymbol{D}_n)$):

A1'. Outcomes: (a) $Y_{it}^N = \boldsymbol{\lambda}_t\boldsymbol{\mu}_i + \epsilon_{it}$; (b) $Y_{it} = Y_{it}^N + \mathbf{D}_{it}'\boldsymbol{\alpha}_{it}$, such that $E[\boldsymbol{\alpha}_{it}|\boldsymbol{D}] = \boldsymbol{\alpha}_i$.

This modification permits additional heterogeneity in the parameters ($\boldsymbol{\alpha}_{it}$). The restriction is that this heterogeneity cannot be systematically related to the values of the explanatory variables. Alternative assumptions are also possible. The identification proof below is unaffected by replacing *A1* with *A1'*.

3.3.3. Homogeneous Effects

On the other hand, there may be advantages to assuming and estimating homogeneous effects, similar to what is commonly done with many panel data models. This restriction reduces the computational complexity of the optimization problem and overfitting concerns as $K \times n$ approaches T. In these cases, it is straightforward to assume $\boldsymbol{\alpha}_i = \boldsymbol{\alpha}$ for all i and estimate a single parameter for each explanatory variable or a subset of the

adoption ($t = T_0 + 1$), $Y_{kt} - \sum_{j\neq k} w_j^k Y_{jt}$ provides evidence of the policy effect in unit 1 since the difference in the policy variable for unit k (relative to its synthetic control) is equal to $-w_1^k D_{1,T_0+1} = -w_1^k$. Since this is nonzero, the policy effect is identified.

[15] See discussions in Ben-Michael, Feller, and Rothstein (2018), Arkhangelsky et al. (2020), Amjad, Shah, and Shen (2018), Ferman and Pinto (2018), Xu (2017), and Powell (2020).

[16] Section 3.3.3 discusses assuming homogenous effects for some parameters which would relax this requirement.

explanatory variables. It is straightforward to modify condition *A1* to include a set of variables, $\tilde{\mathbf{D}}$, with homogenous parameters such that $Y_{it} = Y_{it}^N + \mathbf{D}_{it}'\boldsymbol{\alpha}_i + \tilde{\mathbf{D}}_{it}'\boldsymbol{\beta}$.

As one example of how to categorize variables, CSSC (and most panel data applications) assumes no unit-specific heterogeneity in the parameters associated with the covariates. In that spirit, policy variables could be denoted by $\boldsymbol{D}$ while the additional covariates could be denoted by $\tilde{\boldsymbol{D}}$. However, this distinction is not necessary.

3.3.4. Estimating Synthetic Controls Using Covariates

As a simplification, I discussed CSSC without covariates in Section 2.1. However, CSSC uses preintervention characteristics and selects preintervention outcomes (or averages) in a two-step process, partially due to concerns of overfitting when T is not large. GSC can construct synthetic control weights in the same manner. Define $\boldsymbol{Z}_i(\boldsymbol{b}_i)$ as an $L \times 1$ vector consisting of covariates (excluding $\boldsymbol{D}_i$), averages of those covariates, values of $Y_{it} - \boldsymbol{D}_{it}'\boldsymbol{b}_i$, and averages over multiple time periods of $Y_{it} - \boldsymbol{D}_{it}'\boldsymbol{b}_i$. Each element of $\boldsymbol{Z}_i(\boldsymbol{b}_i)$ is indexed by ℓ and the elements themselves can be defined by time period (e.g., the unemployment rate for time period s).

Optimization proceeds as before, though I write it here as occurring in nested steps

$$\hat{\boldsymbol{\alpha}} = \operatorname{argmin}_{\boldsymbol{b}_1,\ldots,\boldsymbol{b}_N} \left\{ \frac{1}{2nT} \sum_{i=1}^{n} \hat{\Omega}^{-1} \sum_{t=1}^{T} \left[Y_{it} - \boldsymbol{D}_{it}'\boldsymbol{b}_i - \sum_{j \neq i} \left(\hat{w}_j^i(\boldsymbol{b}) \left(Y_{jt} - \boldsymbol{D}_{jt}'\boldsymbol{b}_j \right) \right) \right]^2 \right\},$$

where $\hat{\boldsymbol{w}}(\boldsymbol{b}) = \operatorname{argmin}_{\boldsymbol{\phi}_1 \in \mathcal{W}_1,\ldots,\boldsymbol{\phi}_n \in \mathcal{W}_n}$

$$\left\{ \frac{1}{2nL} \sum_{i=1}^{n} \sum_{\ell=1}^{L} V_{i\ell} \left[Z_{i\ell}(\boldsymbol{b}_i) - \sum_{j \neq i} \phi_j^i Z_{j\ell}(\boldsymbol{b}_j) \right]^2 \right\}.$$

where $\boldsymbol{V}$ represents a weighting matrix to place more weight on certain elements of $\boldsymbol{Z}$ (which may vary by i) and can be chosen in the same manner as for CSSC (see Abadie, Diamond, and Hainmueller 2010, p. 496). Thus, GSC permits covariates to be used in the same manner as CSSC while also providing an opportunity to estimate the relationship between the covariates and the outcome. As in the previous section, the researcher can separate the variables into categories. One set of variables ($\boldsymbol{Z}$) can include those in which (a) the researcher is not interested in estimating the parameters associated with those variables and (b) the convex hull assumption holds for those variables. This distinction is not required and the development of GSC in this article has simply assumed that all variables are included in $\boldsymbol{D}$.

3.4. Properties

3.4.1. Identification

In the next sections, I discuss properties of the GSC estimates. Identification of $\boldsymbol{\alpha}$ holds under *A1–A4*:

Theorem 3.1 (Identification). If *A1–A4* hold, then

$$E\left\{ \frac{1}{2nT} \sum_{i=1}^{n} \Omega_i^{-1} \sum_{t=1}^{T} \left[Y_{it} - \boldsymbol{D}_{it}'\boldsymbol{b}_i - \sum_{j \neq i} \left(\phi_j^i \left(Y_{jt} - \boldsymbol{D}_{jt}'\boldsymbol{b}_j \right) \right) \right]^2 \right\}$$

has a unique minimum and $\boldsymbol{b} = \boldsymbol{\alpha}$ at this minimum.

Theorem 3.1 states that $\boldsymbol{\alpha}$ is unique. The weights are not necessarily unique but unique weights are not required for identification of $\boldsymbol{\alpha}$. A similar result holds if *A1* is replaced by *A1'*. I include a discussion in Appendix A (supplementary material).

3.4.2. Consistency

To discuss properties of the GSC estimates, it is helpful to make assumptions about the synthetic untreated outcomes. For a given $\boldsymbol{b}$, define the corresponding synthetic control weights for each unit as

$$\widehat{\boldsymbol{w}}_i(\boldsymbol{b}) \equiv \operatorname{argmin}_{\boldsymbol{\phi}_i \in \mathcal{W}_i} \frac{1}{2T} \sum_{t=1}^{T} \left[Y_{it} - \boldsymbol{D}_{it}'\boldsymbol{b}_i - \sum_{j \neq i} \phi_j^i \left(Y_{jt} - \boldsymbol{D}_{jt}'\boldsymbol{b}_j \right) \right]^2. \tag{8}$$

Also, define $\Gamma_{it}(\boldsymbol{b}) \equiv \sum_{j \neq i} \hat{w}_j^i(\boldsymbol{b}) \left[Y_{jt} - \boldsymbol{D}_{jt}'\boldsymbol{b}_j \right]$. The benefit of defining $\Gamma_{it}(\boldsymbol{b})$ is that it links the parameter estimates to the synthetic control for each unit. Consistency (as $T \to \infty$) of $\hat{\boldsymbol{\alpha}}$ requires additional regularity conditions:

- *A5. Within-unit dependence*: For each i, $(Y_{it}, \boldsymbol{D}_{it})$ is a strongly mixing sequence in t of size $-\frac{r}{r-2}$, $r > 2$.
- *A6. Compactness*. Assume $\boldsymbol{\alpha} \in \text{interior}(\Theta)$, where Θ is compact.
- *A7. Dominance condition*: $\sup_{(i,t)} E\left[\left(Y_{it} - \boldsymbol{D}_{it}'\boldsymbol{b}_i \right)^2 \right]^{r+\delta} < \Delta < \infty$ for all $\boldsymbol{b} \in \Theta$ and for any $\delta > 0$.
- *A8. Continuity*: $\Gamma_{it}(\boldsymbol{b})$ continuous in $\boldsymbol{b}$ with probability one for all $\boldsymbol{b} \in \Theta$.

A5 permits dependence within each unit. Other dependence structures would generate similar theoretical results. I impose no assumptions about independence *across* units and the inference procedure will allow for strong dependence across units. *A6* assumes that the parameter is in the interior of a compact parameter space. *A7* is necessary for uniform convergence. *A8* disallows large jumps in the synthetic controls for small changes in the treatment parameters. Since $Y_{it} - \boldsymbol{D}_{it}'\boldsymbol{b}_i$ is continuous in $\boldsymbol{b}_i$ and the synthetic control fits this term, assuming continuity of $\Gamma_{it}(\boldsymbol{b})$ is likely unrestrictive. Under these assumptions, the estimates are consistent:

Theorem 3.2 (Consistency). If *A1–A8* hold, then $\widehat{\boldsymbol{\alpha}} \xrightarrow{p} \boldsymbol{\alpha}$ as $T \to \infty$.

This result directly implies consistency of $\hat{\boldsymbol{\alpha}}^{\text{AE}}$. See Appendix A (supplementary material) for a more detailed discussion of Theorem 3.2. For that discussion and for the next section, it is helpful to define $\widehat{Q}(\boldsymbol{b}) \equiv -\frac{1}{2nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left[Y_{it} - \boldsymbol{D}_{it}'\boldsymbol{b}_i - \Gamma_{it}(\boldsymbol{b}) \right]^2$.

3.4.3. Asymptotic Normality

Define the sample average of the gradient as

$$\nabla_{\boldsymbol{\alpha}} \widehat{Q}(\boldsymbol{\alpha}) = \left(\nabla_{\boldsymbol{\alpha}_1} \widehat{Q}(\boldsymbol{\alpha}), \ldots, \nabla_{\boldsymbol{\alpha}_n} \widehat{Q}(\boldsymbol{\alpha})\right),$$

$$\text{where} \quad \nabla_{\boldsymbol{\alpha}_k} \widehat{Q}(\boldsymbol{\alpha}) \equiv \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left\{ \left[\boldsymbol{D}_{it} \times \mathbf{1}(i=k) + \frac{\partial \Gamma_{it}(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}_k} \right] \left[Y_{it} - \boldsymbol{D}_{it}' \boldsymbol{\alpha}_i - \Gamma_{it}(\boldsymbol{\alpha}) \right] \right\}.$$

The gradient reflects the movement in the objective function with respect to the parameters, which includes changes in the synthetic untreated outcomes, governed by Equation (8). Let k index each condition corresponding to the K explanatory variables:

A9. $E[\nabla^2_{\boldsymbol{\alpha}\boldsymbol{\alpha}'} Q(\boldsymbol{\alpha})]$ nonsingular.
A10. (a) $E|\nabla_{\boldsymbol{\alpha}} Q_{itk}(\boldsymbol{\alpha})|^r < \Delta < \infty$ for all (i,t,k). (b) $\Sigma \equiv \lim_{T\to\infty} \operatorname{var}\left(\sqrt{nT}\nabla_{\boldsymbol{\alpha}} \widehat{Q}(\boldsymbol{\alpha})\right)$ exists and is finite.
A11. $\Gamma_{it}(\boldsymbol{b})$ twice continuously differentiable for all $\boldsymbol{b} \in \Theta$.
A12. $\sup_{(i,j,k,t)} E\left[\left(\frac{\partial^2 \Gamma_{it}(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}_j \partial \boldsymbol{\alpha}_k'} \left[Y_{it} - \boldsymbol{D}_{it}'\boldsymbol{b}_i - \Gamma_{it}(\boldsymbol{b})\right] - \left(\mathbf{D}_{it} \times \mathbf{1}(i=j) + \frac{\partial \Gamma_{it}(\boldsymbol{b})}{\partial \boldsymbol{\alpha}_j}\right)\left(\mathbf{D}_{it} \times \mathbf{1}(i=k) + \frac{\partial \Gamma_{it}(\boldsymbol{b})}{\partial \boldsymbol{\alpha}_k}\right)'\right)^2\right]^{r+\delta} < \Delta < \infty$ for some $\delta > 0$ and for all $\boldsymbol{b} \in \Theta$.

Theorem 3.3 (Asymptotic normality). If *A1–A12* hold, then $\sqrt{nT}\,(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \xrightarrow{d} N(0, H^{-1}\Sigma H^{-1})$ as $T \to \infty$.

I include a discussion in Appendix A (supplementary material). 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 $\frac{\partial \Gamma_{it}(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}$. These concerns motivate the inference procedure discussed below.

4. Inference

4.1. Hypothesis Testing

I discuss an inference procedure that is valid for fixed n as $T \to \infty$. The procedure permits arbitrary within-unit dependence as well as cross-unit dependence. The traditional randomization approach typically used with CSSC is difficult to apply more generally to cases with multiple and/or nonbinary variables.[17] Permitting cross-unit dependence is important since GSC generates mechanical correlations across units. Each unit potentially composes part of the synthetic controls for other units.[18]

The inference method relies on a Wald statistic given the mean and variance of orthogonal score functions generated from 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 co-movements between the scores for each cluster and construct independent functions. Since the synthetic control weights are fixed over time by unit, the induced correlations occur due to movements in contemporaneous time periods across units.

This section considers a null hypothesis of the form H_0 : $\boldsymbol{\alpha}^{\text{AE},(k)} = \overline{a}$, where k denotes the parameter. The null hypothesis will be tested by imposing the restriction when minimizing the objective function: $(\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{w}}_1, \ldots, \tilde{\boldsymbol{w}}_n) =$

$$\operatorname{argmin}_{\boldsymbol{b}, \boldsymbol{\phi}_1 \in \mathcal{W}_1, \ldots, \boldsymbol{\phi}_n \in \mathcal{W}_n} \left\{ \frac{1}{2nT} \sum_{i=1}^{n} \hat{\Omega}_i^{-1} \sum_{t=1}^{T} \left[Y_{it} - \boldsymbol{D}_{it}'\boldsymbol{b}_i - \sum_{j \neq i} \left(\phi_j^i \left(Y_{jt} - \boldsymbol{D}_{jt}'\boldsymbol{b}_j \right) \right) \right]^2 \right\}$$

$$\text{subject to } \boldsymbol{\alpha}^{\text{AE},(k)} = \overline{a}. \tag{9}$$

The null hypothesis is imposed and the other parameters, including the synthetic control weights, are estimated.[19] Null hypotheses constraining $\boldsymbol{\alpha}^{(\boldsymbol{k})}$ are also possible in this framework. The inference procedure uses the function:

$$h_{it}^{(k)}(\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{w}}_i) \equiv \left(D_{it}^{(k)} - \sum_{j \neq i} \tilde{w}_j^i D_{jt}^{(k)} \right) \left[Y_{it} - \boldsymbol{D}_{it}'\tilde{\boldsymbol{\alpha}}_i - \sum_{j \neq i} \tilde{w}_j^i \left[Y_{jt} - \boldsymbol{D}_{jt}'\tilde{\boldsymbol{\alpha}}_j \right] \right]. \tag{10}$$

$h_i^{(k)}(\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{w}}_i)$ is a $T \times 1$ vector. Under the null hypothesis, $E[h_i^{(k)}(\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{w}}_i)] = 0$. There are n moments related to $D^{(k)}$, but they are aggregated together in Equation (10) to make the orthogonalization process below more straightforward.[20]

The idea is to construct orthogonal functions, which I model as follows:

$$s_{mt}(\tilde{\boldsymbol{\alpha}}) = \sum_{i=1}^{n} c_{im} h_{it}^{(k)}(\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{w}}_i), \tag{11}$$

where $\operatorname{cov}(s_{mt}(\tilde{\boldsymbol{\alpha}}), s_{\ell t}(\tilde{\boldsymbol{\alpha}}))=0$ for $m \neq \ell$. The c_{im} parameters above are weights $(m = 1, \ldots, n)$ estimated using principal components analysis to construct orthogonal "scores."

This approach models dependence as operating through correlations in the same time period. This restriction does *not* assume that observations in different units and time periods are independent (i.e., h_{it} is independent of h_{js}). The focus on contemporaneous co-movements is because the dependence induced by GSC 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 arbitrary within-unit dependence, the estimator potentially induces 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

[17] Firpo and Possebom (2018) discussed some concerns and modifications even in the traditional context.

[18] This section parallels the inference procedure developed in Powell (2019) for panel data estimators more generally.

[19] Equation (6) weights do not need to be re-estimated for inclusion in Equation (9).

[20] An advantage of the inference procedure is that it is unnecessary to use all elements of the gradient. This property permits use of the gradient with respect to the parameter of interest only and avoids requiring estimation of the gradient with respect to the synthetic control weights, which may be complicated. The test statistic is valid if other elements are also used, but their inclusion could reduce power since they are equal to zero regardless of whether the null hypothesis is true or not.

co-movements, the inference method eliminates dependence across all observations in different units and then constructs asymptotically independent functions. Define

$$s_m(\tilde{\boldsymbol{\alpha}}) \equiv \frac{1}{T}\sum_{t=1}^{T} s_{mt}(\tilde{\boldsymbol{\alpha}}), \tag{12}$$

where $s_m(\tilde{\boldsymbol{\alpha}})$ is an asymptotically independent score function. For inference, the method uses a Wald statistic defined by

$$S = \left(\frac{1}{n}\sum_{k=1}^{n} s_k(\tilde{\boldsymbol{\alpha}})\right)' \widehat{\Sigma}(\tilde{\boldsymbol{\alpha}})^{-1}\left(\frac{1}{n}\sum_{k=1}^{n} s_k(\tilde{\boldsymbol{\alpha}})\right), \tag{13}$$

where

$$\widehat{\Sigma}(\tilde{\boldsymbol{\alpha}}) = \frac{1}{n-1}\sum_{k=1}^{n}\left(s_k(\tilde{\boldsymbol{\alpha}}) - \left(\frac{1}{n}\sum_{j=1}^{n} s_j(\tilde{\boldsymbol{\alpha}})\right)\right)\left(s_k(\tilde{\boldsymbol{\alpha}}) - \left(\frac{1}{n}\sum_{j=1}^{n} s_j(\tilde{\boldsymbol{\alpha}})\right)\right)'.$$

To derive p-values, the inference method simulates the distribution of this test statistic by perturbing the $s_m(\tilde{\boldsymbol{\alpha}})$ functions. Given asymptotically independent and symmetric[21] functions with the same mean under the null hypothesis, the inference approach follows directly from Canay, Romano, and Shaikh (2017). The weights $\{W_k\}_{k=1}^{n}$ used to perturb the functions are iid, independent of $s_k(\boldsymbol{\alpha})$, and drawn from the Rademacher distribution. To simulate the distribution of the test statistic, consider weights $W_k^{(d)}$ which satisfy these conditions and construct a simulated test statistic, indexed by (d)

$$S^{(d)} = \left(\frac{1}{N}\sum_{k=1}^{N} W_k^{(d)} s_k(\tilde{\boldsymbol{\alpha}})\right)' \widehat{\Sigma}^{(d)}(\tilde{\boldsymbol{\alpha}})^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} W_k^{(d)} s_k(\tilde{\boldsymbol{\alpha}})\right), \tag{14}$$

$$\widehat{\Sigma}^{(d)}(\tilde{\boldsymbol{\alpha}}) = \frac{1}{n-1}\sum_{k=1}^{n}\left(W_k^{(d)} s_k(\tilde{\boldsymbol{\alpha}}) - \left(\frac{1}{n}\sum_{j=1}^{n} W_j^{(d)} s_j(\tilde{\boldsymbol{\alpha}})\right)\right) \times \left(W_k^{(d)} s_k(\tilde{\boldsymbol{\alpha}}) - \left(\frac{1}{n}\sum_{j=1}^{n} W_j^{(d)} s_j(\tilde{\boldsymbol{\alpha}})\right)\right)'.$$

The p-value is the fraction of simulations in which the simulated value of the test statistic is greater than S: $\hat{p} = \frac{1}{D}\sum_{d=1}^{D} \mathbf{1}(S < S^{(d)})$, given D simulations. Large values of S imply that we should rarely observe the distribution of gradients generated under the null hypothesis and that we should reject it. The proposed inference method produces valid inference given a finite number of heterogenous, dependent clusters.

Multiple hypothesis testing (i.e., $H_0 : a(\boldsymbol{\alpha}^{\text{AE}}) = \mathbf{0}$) is also possible following this approach. In this case, $s_{mt}(\tilde{\boldsymbol{\alpha}}) = \sum_{i=1}^{n}\sum_{k\in\mathcal{K}} c_{im} h_{it}^{(k)}(\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{w}}_i)$, where $\mathcal{K}$ is the set of conditions not equal to zero when the null is not true. The inference procedure then follows the steps outlined above.[22]

[21] This property holds due to the asymptotic normality of the estimates, discussed in the previous section.

[22] In this case, there will be more than n independent score functions so the test statistic must be modified accordingly.

4.2. Diagnostics

In the CSSC context, it is typical to visually evaluate the "fit" of the synthetic control for the treated unit. This approach is similar to event study estimation often paired with difference-in-differences designs in which leads of a policy variable are included in the specification to test whether there are differential effects prior to policy changes.[23] In the absence of anticipation effects (Alpert 2016; Malani and Reif 2015), meaningful and statistically significant relationships between policy leads and the outcome suggest that the empirical approach is not using proper counterfactuals. The same method can be applied with GSC by including leads of the explanatory variable(s) of interest.

However, the proposed GSC estimator estimates unit-specific parameters for all variables (or a subset, as discussed in Section 3.3.3) so including even a few leads may drastically increase the number of parameters. As a result, I suggest post-estimation testing. The estimation of the parameters implicitly imposed the null hypothesis $H_0 : \boldsymbol{\alpha}^s = 0$, where I denote the parameters associated with the s-lead policy variables as $\boldsymbol{\alpha}^s$. The relevant moment condition related to lead s for variable k is

$$h_{it}^{s,k}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{w}}_i) \equiv \left(D_{i,t+s}^{(k)} - \sum_{j\neq i}\hat{w}_j^i D_{j,t+s}^{(k)}\right) \left[Y_{it} - \boldsymbol{D}_{it}'\hat{\boldsymbol{\alpha}}_i - \sum_{j\neq i}\hat{w}_j^i\left[Y_{jt} - \boldsymbol{D}_{jt}'\hat{\boldsymbol{\alpha}}_j\right]\right]. \tag{15}$$

It is important to note that the *unrestricted* moment conditions are used here (i.e., $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{w}}$). The inference procedure proceeds as before.

If the leads are directly included as variables in the estimation, then it is possible to use the inference procedure to test $H_0 : \boldsymbol{\alpha}^{\text{AE},(k),s} = 0$. The null hypothesis is that, on average, the leads have no effect on the outcome. The proposed approach, instead, tests $H_0 : \boldsymbol{\alpha}^{(\boldsymbol{k}),\boldsymbol{s}} = 0$, whether the leads in each unit are jointly equal to 0. These are different tests and both can be employed with GSC. Given that we would often not expect policies—in the absence of anticipation effects—to impact any unit prior to adoption, this latter and simpler test should be adequate in many applications.

5. Simulations

In this section, I report the results of simulations using GSC. The data are generated by $Y_{it} = \beta_i d_{it} + 5 \times \sum_{k=1}^{2}\lambda_t^{(k)}\mu_i^{(k)} + \epsilon_{it}$,

$$\lambda_t^{(1)} = \begin{cases} 0.1t & \text{if} \quad t \leq 20 \\ 2 & \text{if} \quad 20 < t < 30 \\ 2 + \frac{t-30}{30} & \text{if} \quad t \geq 30 \end{cases},$$

$$\lambda_t^{(2)} = \begin{cases} 2 & \text{if} \quad t \leq 5 \\ -4 & \text{if} \quad 5 < t \leq 20 \\ 6 & \text{if} \quad 20 < t \leq 40 \\ 1 & \text{if} \quad t > 40 \end{cases},$$

[23] Critiques and modifications of this approach are discussed in Rambachan and Roth (2019) and Bilinski and Hatfield (2018).

Table 1. Main simulation results.

Panel A: Estimates

	Model 1			Model 2			Model 3			Model 4		
	Mean Bias	MAD	RMSE	Mean Bias	MAD	RMSE	Mean Bias	MAD	RMSE	Mean Bias	MAD	RMSE
GSC	0.011	0.020	0.031	0.005	0.006	0.010	−0.001	0.012	0.020	−0.002	0.012	0.018
Additive fixed effects	0.850	0.854	0.885	0.846	0.844	0.856	0.501	0.496	0.525	0.263	0.254	0.281
+ Linear trend	0.696	0.699	0.727	0.692	0.690	0.701	0.423	0.421	0.443	0.173	0.169	0.191
+ Quadratic trend	0.642	0.645	0.672	0.638	0.637	0.647	0.396	0.395	0.415	0.135	0.131	0.151
+ Cubic trend	0.355	0.353	0.376	0.351	0.349	0.357	0.258	0.254	0.273	0.136	0.131	0.152
With slope heterogeneity	0.513	0.518	0.539	0.520	0.516	0.528	0.211	0.213	0.231	0.113	0.108	0.139
Interactive fixed effects	−0.027	0.032	0.047	−0.011	0.012	0.017	0.065	0.059	0.081	0.062	0.056	0.084
With slope heterogeneity	2.711	0.008	7.894	0.021	0.003	0.654	2.230	0.005	6.374	−0.001	0.007	0.010

NOTES: 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. "With Slope Heterogeneity" means the unit-specific parameters were estimated and averaged. I implemented the interactive fixed effects estimator assuming 2 interactive fixed effects. MAD = median absolute deviation; RMSE = root mean squared error.

Panel B: GSC rejection rate at size 0.05

	Model 1	Model 2	Model 3	Model 4
Rejection rate	0.044	0.041	0.033	0.056

NOTE: I use the Section 4 inference procedure.

and $\mu_i^{(1)}, \mu_i^{(2)} \sim U(0,1)$; $\epsilon_{it} \sim N(0, \frac{1}{4})$. I let the relationship between the policy variable and the outcome vary by unit and the variation is dependent on the factor loadings: $\beta_i = \sum_{k=1}^{2} \mu_i^{(k)} - \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{2} \mu_i^{(k)}$. I estimate several models, varying construction of the treatment variables as well as the size of n and T:

$$\text{Model 1:} \quad d_{it} = \sum_{k=1}^{2} \lambda_t^{(k)} \mu_i^{(k)} + U_{it}, T = 50, n = 10,$$

$$\text{Model 2:} \quad d_{it} = \sum_{k=1}^{2} \lambda_t^{(k)} \mu_i^{(k)} + U_{it}, T = 50, n = 30,$$

$$\text{Model 3:} \quad d_{it} = \sum_{k=1}^{2} \lambda_t^{(k)} \mu_i^{(k)} + \tilde{U}_{it} \times (\beta_i + 3), T = 50, n = 10,$$

$$\text{Model 4:} \quad d_{it} = \sum_{k=1}^{2} \lambda_t^{(k)} \mu_i^{(k)} + U_{it}, T = 20, n = 10,$$

where $U_{it} \sim U(0, 10)$, $\tilde{U}_{it} \sim U(0, 5)$. Both the outcome and the treatment variable are functions of the same interactive fixed effects. I intentionally generate outcomes which would be difficult to fit with unit-specific polynomials. The treatment variable is continuous such that CSSC is not applicable. Model 2 increases the number of units from 10 to 30. Model 3 allows within-unit variation in the explanatory variable to be a function of β_i. Model 4 repeats Model 1 but for a smaller number of time periods.

In Panel A of Table 1, I report mean bias, median absolute deviation (MAD) and root-mean-square error (RMSE) from 1000 simulations per model. The first row provides the results for GSC. The next 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. These results are provided simply to illustrate that more traditional two-way fixed-effects estimators are not valid given the above data-generating process, even when flexible unit-specific trends are included. I also provide two-way fixed-effects estimator results in which each unit is permitted to have its own slope.[24] Finally, I estimate the model using an interactive fixed-effects estimator,[25] assuming (correctly) that there are two interactive fixed effects. Since I am providing the estimator with the correct information, the results for this estimator may represent a best case scenario. I provide interactive fixed-effects results with and without slope heterogeneity.

The additive fixed-effects estimators, as expected, perform poorly, even when state-specific trends are included. The interactive fixed-effects models generally perform well. Notably, when slope heterogeneity is modeled, most of the interactive fixed-effects estimates in the simulations are close to the true value, as can be seen in the MAD statistics. However, in rare cases, the estimator provides estimates that are considerably biased, producing large mean biases and RMSE values.

GSC performs well across all models. In Panel B, I provide rejection rates for the GSC estimator for size 0.05 using the

[24] I do not include unit-specific trends when also allowing for slope heterogeneity.

[25] I implement this procedure in Stata using `regife`.

Table 2. Additional simulation results.

	Model 5			Model 6			Model 7		
	Mean bias	MAD	RMSE	Mean bias	MAD	RMSE	Mean bias	MAD	RMSE
GSC	0.000	0.005	0.008	0.002	0.008	0.012	1.011	1.012	1.012
Fixed effects	0.100	0.094	0.111	0.000	0.006	0.008	1.698	1.701	1.710
+ Linear trend	0.100	0.095	0.110	0.000	0.006	0.008	1.570	1.573	1.579
+ Quadratic trend	0.100	0.095	0.110	0.000	0.006	0.009	1.525	1.527	1.534
+ Cubic trend	0.100	0.096	0.110	0.000	0.006	0.009	1.291	1.290	1.295
With slope heterogeneity	0.000	0.004	0.006	0.000	0.006	0.008	1.411	1.415	1.417
Interactive fixed effects	0.047	0.063	0.086	0.000	0.006	0.009	0.973	0.973	0.974
With slope heterogeneity	0.019	0.006	0.117	0.023	0.009	0.135	3.012	1.002	6.241

NOTES: 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. "With Slope Heterogeneity" means the unit-specific parameters were estimated and averaged. I implemented the interactive fixed-effects estimator assuming two interactive fixed effects. MAD, median absolute deviation; RMSE, root-mean-squared error.

proposed inference procedure. The inference procedure rejects at appropriate rates across models with some evidence that it underrejects.

The motivations for the next simulations are (a) to test how GSC fares when the additive fixed-effects model is appropriate and (b) to consider a case in which the explanatory variable is endogenous even conditional on unobserved interactive fixed effects. The models are the same as before with limited changes (variables and distributions defined above):

$$\text{Model 5: } Y_{it} = \beta_i d_{it} + 5 \times \sum_{k=1}^{2} \left(\lambda_t^{(k)} + \mu_i^{(k)} \right) + \epsilon_{it},$$

$$d_{it} = \sum_{k=1}^{2} \left(\lambda_t^{(k)} + \mu_i^{(k)} \right) + \tilde{U}_{it} \times (\beta_i + 3).$$

Model 6: Same as Model 5 with $\beta_i = 0$ for all i,

$$d_{it} = \sum_{k=1}^{2} \left(\lambda_t^{(k)} + \mu_i^{(k)} \right) + U_{it}.$$

$$\text{Model 7: } Y_{it} = \beta_i d_{it} + 5 \times \sum_{k=1}^{2} \lambda_t^{(k)} \mu_i^{(k)} + \tilde{U}_{it} + \epsilon_{it},$$

$$d_{it} = \sum_{k=1}^{2} \lambda_t^{(k)} \mu_i^{(k)} + \tilde{U}_{it}.$$

Model 5 is an additive fixed-effects model with slope heterogeneity. Model 6 uses a constant parameter of interest. In Model 7, both the outcome and explanatory variable are functions of $\tilde{U}_{it}$, a confounder. The results from 1000 simulations are presented in Table 2.

GSC performs well even when the true model is an additive fixed-effects specification. This is not necessarily surprising given that interactive fixed-effects nest additive fixed effects. In terms of MAD and RMSE, the fixed-effects estimators (with slope heterogeneity in Model 5; with and without slope heterogeneity in Model 6) outperform GSC. This result also is perhaps not surprising since these estimators impose a correct assumption, which should improve performance.

Finally, Model 7 provides a case in which there is an omitted variable, correlated with both the explanatory variable and the outcome, which is difficult to approximate using other units. All of the estimators, including GSC, perform poorly for this model.

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-effects specification typically estimated in the literature in which the outcome is the log of the employment rate for 16–19 years old and the explanatory variable is the log of the minimum wage. I use the dataset constructed for Dube and Zipperer (2015) and Allegretto et al. (2017).[26] The data are state-level minimum wage information merged with employment rates aggregated from the Current Population Survey (CPS) for 1979–2014. There are 51 units and 144 time periods (quarter-years). It is not possible to apply the CSSC estimator to this application since there are no "untreated" observations. Instead, it is necessary to use GSC given the inclusion of a continuous policy variable. I also present results for an interactive fixed-effects model using Bai (2009), similar to the approach applied in Totty (2017). I present results in Table 3.

The additive fixed-effects estimator produces an elasticity estimate of −0.268, which is consistent with estimates found in the literature using this approach. However, including state-specific trends in Column 2 produces a positive elasticity estimate of 0.110.[27] In Columns 3 and 4, I permit state-specific slope heterogeneity and estimate positive elasticities. Interestingly, comparing the estimates without state-specific trends (Columns 1 and 3), state-specific slope heterogeneity changes the elasticity estimate considerably. In Column 5, I use an interactive fixed-effects estimator and estimate an elasticity of −0.077. This estimate increases in magnitude to −0.165 (and is statistically different from zero) when the elasticity parameters are permitted to vary by state.

The GSC estimate is provided in Column 6. I estimate an elasticity of −0.178 and can statistically reject that the average effect is zero.[28] Thus, while the equivalent two-way fixed-effects estimator with slope heterogeneity produces a positive elasticity estimate (Column 3), GSC more flexibly accounts for differences in underlying outcome trends and estimates that minimum wage increases reduce employment rates. Notably, the GSC

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

[27]See Meer and West (2016) for concerns about state-specific trends in this context.

[28]If I do not allow for state-specific parameter heterogeneity, I estimate an even larger elasticity.

Table 3. Minimum wage elasticity estimates.

				Panel A: Estimates				
	ln(Employment rate)							
Outcome	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln(MW)	−0.268***	0.110	0.105	0.135	−0.077	−0.165**	−0.178**	−0.004
p-value	[0.000]	[0.470]	[0.274]	[0.454]	[0.144]	[0.026]	[0.045]	[0.772]
Fixed Effects / Estimator	Additive	Additive	Additive	Additive	Interactive	Interactive	GSC	GSC
Trends	No	Yes	No	Yes	No	No	No	No
Parameters	Constant	Constant	By State	By State	Constant	By State	By State	By State
Ages	16–19	16–19	16–19	16–19	16–19	16–19	16–19	25–55

NOTES: $N = 7344$. Significance Levels: *10%, **5%, and ***1%. *P*-values in brackets reflect test that average elasticity estimate is equal to 0. For Columns 1–6, I use a cluster covariance matric estimator to estimate standard errors to implement this test. For Columns 7–8, I use the inference procedure discussed in Section 4. I assume three common factors for the interactive fixed-effects estimators. The "Parameters" row refers to whether state-specific parameters were estimated and aggregated (weighted by population) or whether a constant effect was assumed across all states. "Trends" refers to whether state-specific cubic trends are included. The final column is a placebo test, studying a population which should be less affected by minimum wage policy. All analyses are population-weighted.

	Panel B: P-Values from tests of significance of leads	
	GSC	Additive fixed effects
ln(MW)	0.045	0.133
Lead 1 Period	0.777	0.083
Lead 2 Period	0.529	0.165
Lead 3 Period	0.802	0.227

NOTES: This panel uses the inference procedure discussed in Section 4.2 and present the resulting *p*-values. The additive fixed effects model refers to the model estimated in Panel A, Column 1 (given that it is the specification often used in the literature). The *p*-value for the contemporaneous value of ln(MW) is different than the one shown in Panel A since I use Section 4.2 procedure, which accounts for cross-state dependence (also a potential issue for fixed-effects estimation), here.

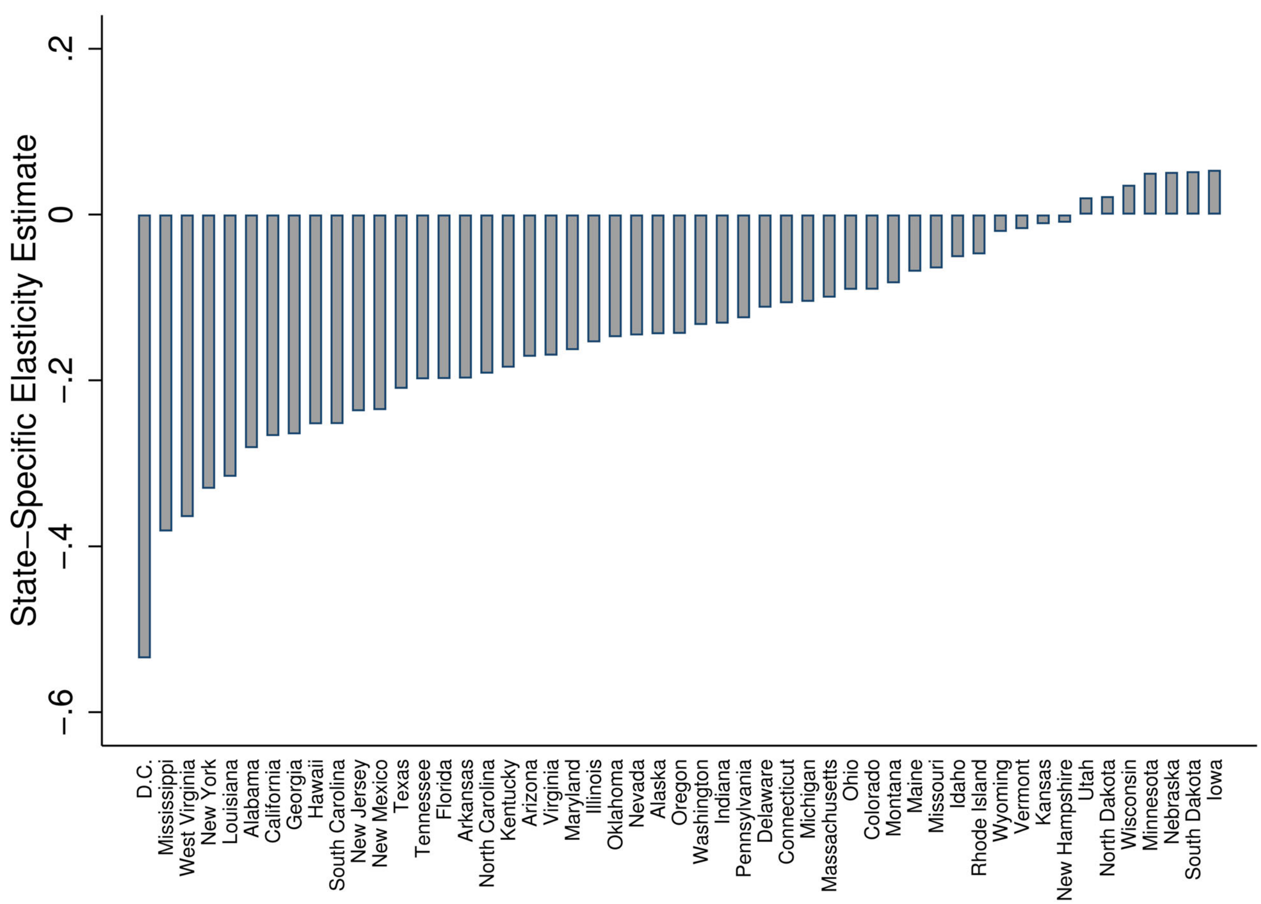

Figure 1. State-specific elasticity estimates.
NOTES: GSC state-specific estimates. These estimates are the inputs for the average elasticity estimate shown in Table 3, Panel A, Column 7.

estimate is similar to the interactive fixed-effects estimate with slope heterogeneity. I present the state-specific estimates which are the inputs for the Column 6 average estimate graphically in Figure 1. Fourty-four of the 51 estimates are negative.

Finally, in Column 8, I use GSC to estimate the elasticity for ages 25–55. The minimum wage is less binding for this age group so we would not expect to see much of an employment impact. I estimate a small elasticity of −0.004, which suggests that GSC is able to appropriately account for any confounding trends correlated with minimum wage increases and that the Column 7 estimate reflects a causal response.

In Panel B of Table 3, I implement the test discussed in Section 4.2. The assumption is that future minimum wages, conditional on the current minimum wage, should not predict current employment rates. The first column shows that the null hypothesis that these leads are equal to zero is never rejected (with p-values ranging from 0.529 to 0.802) for the GSC estimator. The p-values are much smaller for the traditional additive fixed-effects model (Column 1 of Panel A). The p-value for the contemporaneous minimum wage variable is different than the one shown in Panel A because I apply the Section 4 procedure here. Additive time fixed effects can also induce dependence across units, which is consistent with an increase in the p-value when I account for this dependence. I can reject that the next period's minimum wage is not predictive of employment rates at the 10% significance level.

7. Discussion and Conclusion

Many empirical applications rely on panel data to exploit differential changes in the explanatory variables to study their impacts on the outcome. This article introduces a synthetic control estimator which generalizes the comparative case study synthetic control estimator of Abadie and Gardeazabal (2003) and Abadie, Diamond, and Hainmueller (2010). The estimator jointly estimates the parameters associated with the explanatory variables while creating synthetic controls for each unit. Instead of assuming that the national average is an appropriate control for each treated unit, the GSC estimator permits additional flexibility.

Additive fixed-effects models are not limited to applications in which the treatment is represented by a single binary variable. This article 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 with the benefits of synthetic control estimation.

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 briefly 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 estimator in this article should be useful more broadly for applications using panel data.

Supplementary Materials

Proofs and replication files are available in supplementary appendix.

Funding

I am grateful for funding from the CDC (R01CE02999), the Center for Causal Inference, and the Methods Lab. I received helpful feedback from seminar participants at Uber, the University of Calgary Empirical Microeconomics Workshop, and the Center for Causal Inference as well as from Abby Alpert, Tom Chang, John Donohue, Michael Dworsky, Beth Ann Griffin, Kathleen Mullen, David Neumark, Michael Robbins, Stefan Staubli, and Jeffrey Wenger.

References

Abadie, A. (2010), "Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects," *Journal of Economic Literature*, 2019.

Abadie, A., Diamond, A., and Hainmueller, J. (2010), "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program," *Journal of the American Statistical Association*, 105, 493–505. [1,2,5,6,12]

——— (2015), "Comparative Politics and the Synthetic Control Method," *American Journal of Political Science*, 59, 495–510. [1,2]

Abadie, A., and Gardeazabal, J. (2003), "The Economic Costs of Conflict: A Case Study of the Basque Country," *American Economic Review*, 113–132. [1,2,12]

Acemoglu, D., Johnson, S., Kermani, A., Kwak, J., and Mitton, T. (2016), "The Value of Connections in Turbulent Times: Evidence From the United States," *Journal of Financial Economics*, 121, 368–391. [1]

Ahn, S. C., Lee, Y. H., and Schmidt, P. (2013), "Panel Data Models With Multiple Time-Varying Individual Effects," *Journal of Econometrics*, 174, 1–14. [3]

Allegretto, S., Dube, A., Reich, M., and Zipperer, B. (2017), "Credible Research Designs for Minimum Wage Studies: A Response to Neumark, Salas, and Wascher," *ILR Review*, 70, 559–592. [3,10]

Alpert, A. (2016), "The Anticipatory Effects of Medicare Part D on Drug Utilization," *Journal of Health Economics*, 49, 28–45. [8]

Amjad, M., Shah, D., and Shen, D. (2018), "Robust synthetic control," *The Journal of Machine Learning Research*, 19, 802–852. [2,5]

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., and Wager, S. (2020), "Synthetic Difference in Differences," Technical Report, National Bureau of Economic Research. [2,5]

Athey, S., and Imbens, G. W. (2017), "The State of Applied Econometrics: Causality and Policy Evaluation," *Journal of Economic Perspectives*, 31, 3–32. [1]

——— (2018), "Design-based Analysis in Difference-in-Differences Settings with Staggered Adoption," Technical Report, National Bureau of Economic Research 2018. [2]

Bai, J. (2009), "Panel Data Models With Interactive Fixed Effects," *Econometrica*, 77, 1229–1279. [3,10]

Bai, J., and Ng, S. (2002), "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191–221. [3]

Ben-Michael, E., Feller, A., and Rothstein, J. (2018), "The Augmented Synthetic Control Method," arXiv:1811.04170. [2,5]

Bilinski, A., and Hatfield, L. A. (2018), "Nothing to See here? Non-Inferiority Approaches to Parallel Trends and Other Model Assumptions," arXiv:1805.03273. [8]

Billmeier, A., and Nannicini, T. (2013), "Assessing Economic Liberalization Episodes: A Synthetic Control Approach," *Review of Economics and Statistics*, 95, 983–1001. [1]

Blomquist, J., and Westerlund, J. (2013), "Testing Slope Homogeneity in Large Panels With Serial Correlation," *Economics Letters*, 121, 374–378. [3]

Borusyak, K., and Jaravel, X. (2017), "Revisiting Event Study Designs," Available at SSRN 2826228. [1]

Callaway, B., and Sant'Anna, P. H. C. (2020), "Difference-in-Differences With Multiple Time Periods," *Journal of Econometrics*. [2]

Campello, M., Galvao, A. F., and Juhl, T. (2019), "Testing for Slope Heterogeneity Bias in Panel Data Models," *Journal of Business & Economic Statistics*, 37, 749–760. [3]

Canay, I. A, Romano, J. P. and Shaikh, A. M. (2017), "Randomization Tests Under an Approximate Symmetry Assumption," *Econometrica*, 85(3), 1013–1030. [8]

Cavallo, E., Galiani, S., Noy, I., and Pantano, J. (2013), "Catastrophic Natural Disasters and Economic Growth," *Review of Economics and Statistics*, 95, 1549–1561. [2]

Cengiz, D., Dube, A., Lindner, A., and Zipperer, B. (2019), "The Effect of Minimum Wages on Low-Wage Jobs," *The Quarterly Journal of Economics*, *134* (3), 1405–1454. [3]

Clemens, J. (2015), "The Minimum Wage and the Great Recession: Evidence From the Current Population Survey," Technical Report, National Bureau of Economic Research. [3]

Clemens, J., and Strain, M. R. (2018), "The Short-Run Employment Effects of Recent Minimum Wage Changes: Evidence From the American Community Survey," *Contemporary Economic Policy*, 36, 711–722. [3]

Clemens, J., and Wither, M. (2019), "The Minimum Wage and the Great Recession: Evidence of Effects on the Employment and Income Trajectories of Low-Skilled Workers," *Journal of Public Economics*, *170*, 53–67. [3]

De Chaisemartin, C., and D'Haultfœuille, X. (2020), "Difference-in-Differences Estimators of Intertemporal Treatment Effects," arXiv:2007.04267. [2]

——— (2020), "Two-Way Fixed Effects Estimators With Heterogeneous Treatment Effects," *American Economic Review*, 110, 2964–96. [2]

Donohue, J. J., Aneja, A., and Weber, K. D. (2019), "Right-to-Carry Laws and Violent Crime: A Comprehensive Assessment Using Panel Data and a State-Level Synthetic Control Analysis," *Journal of Empirical Legal Studies*, 16, 198–247. [1]

Doudchenko, N., and Imbens, G. W. (2016), "Balancing, Regression, Difference-in-Differences and Synthetic Control Methods: A Synthesis," Technical Report, National Bureau of Economic Research. [4]

Dube, A., and Zipperer, B. (2015), "Pooling Multiple Case Studies Using Synthetic Controls: An Application to Minimum Wage Policies," IZA Discussion Paper No. 8944. [3,10]

Dube, A., Lester, T. W., and Reich, M. (2010), "Minimum Wage Effects Across State Borders: Estimates Using Contiguous Counties," *Review of Economics and Statistics*, 92, 945–964. [3]

Ferman, B., and Pinto, C. (2018), "Revisiting the Synthetic Control Estimator," Working Paper. [2,5]

Firpo, S., and Possebom, V. (2018), "Synthetic Control Method: Inference, Sensitivity Analysis and Confidence Sets," *Journal of Causal Inference*, 6. [7]

Gibbons, C. E., Suárez Serrato, J. C., and Urbancic, M. B. (2018), "Broken or Fixed Effects?," *Journal of Econometric Methods*, 8. [1,3]

Gobillon, L., and Magnac, T. (2016), "Regional Policy Evaluation: Interactive Fixed Effects and Synthetic Controls," *Review of Economics and Statistics*, 98, 535–551. [3]

Goodman-Bacon, A. (2018), "Difference-in-Differences With Variation in Treatment Timing," Technical Report, National Bureau of Economic Research. [2]

Hollingsworth, A., and Wing, C. (2020), "Tactics for Design and Inference in Synthetic Control Studies: An Applied Example Using High-Dimensional Data," Available at SSRN 3592088. [2]

Juhl, T., and Lugovskyy, O. (2014), "A Test for Slope Heterogeneity in Fixed Effects Models," *Econometric Reviews*, 33, 906–935. [3]

Kaul, A., Klößner, S., Pfeifer, G., and Schieler, M. (2021), "Standard Synthetic Control Methods: The Case of Using All Pre-Intervention Outcomes Together With Covariates," *Journal of Business & Economic Statistics*, 1–34. [2]

Malani, A., and Reif, J. (2015), "Interpreting Pre-Trends as Anticipation: Impact on Estimated Treatment Effects From Tort Reform," *Journal of Public Economics*, 124, 1–17. [8]

Manning, A. (2020), "The Elusive Employment Effect of the Minimum Wage," *Journal of Economic Perspectives*, 35, 3–26 [3]

Meer, J., West, J. (2016), "Effects of the Minimum Wage on Employment Dynamics," *Journal of Human Resources*, 51, 500–522. [10]

Moon, H. R., and Weidner, M. (2015), "Linear Regression for Panel with Unknown Number of Factors as Interactive Fixed Effects," *Econometrica*, 83, 1543–1579. [3]

Neumark, D. (2019), "The Econometrics and Economics of the Employment Effects of Minimum Wages: Getting from Known Unknowns to Known Knowns," *German Economic Review*, 20, 293–329. [3]

Neumark, D., and Shirley, P. (2021), "Myth or Measurement: What Does the New Minimum Wage Research Say about Minimum Wages and Job Loss in the United States?," Working Paper No. 28388, National Bureau of Economic Research. [3]

Neumark, D., Ian Salas, J. M., and Wascher, W. (2014), "Revisiting the Minimum Wage–Employment Debate: Throwing Out the Baby with the Bathwater?," *Industrial & Labor Relations Review*, 67, 608–648. [3]

Pesaran, M. H., and Yamagata, T. (2008), "Testing Slope Homogeneity in Large Panels," *Journal of Econometrics*, 142, 50–93. [3]

Powell, D. (2019), "Panel Data Inference with Dependent Clusters," Available at *https://sites.google.com/site/davidmatthewpowell/inference-with-correlated-clusters*. [7]

Powell, D. (2020), "Imperfect Synthetic Controls: Did the Massachusetts Health Care Reform Save Lives?" Available at *https://sites.google.com/site/davidmatthewpowell/imperfect-synthetic-controls*. [2,4,5]

Rambachan, A., and Roth, J. (2019), "An Honest Approach to Parallel Trends," Unpublished manuscript, Harvard University. [8]

Sabia, J. J., Burkhauser, R. V., and Hansen, B. (2012), "Are the Effects of Minimum Wage Increases Always Small? New Evidence From a Case Study of New York State," *Industrial & Labor Relations Review*, 65, 350–376. [3]

Słoczyński, T. (2020), "Interpreting OLS Estimands When Treatment Effects Are Heterogeneous: Smaller Groups Get Larger Weights," *Review of Economics and Statistics*, 1–27. [1]

Sun, L., and Abraham, S. (2020), "Estimating Dynamic Treatment Effects in Event Studies With Heterogeneous Treatment Effects," *Journal of Econometrics*. [2]

Totty, E. (2017), "The Effect of Minimum Wages on Employment: A Factor Model Approach," *Economic Inquiry*, 55, 1712–1737. [10]

Wooldridge, J. M. (2005), "Fixed-Effects and Related Estimators for Correlated Random-coefficient and Treatment-Effect Panel Data Models," *Review of Economics and Statistics*, 87, 385–390. [3]

Xu, Y. (2017), "Generalized Synthetic Control Method: Causal Inference With Interactive Fixed Effects Models," *Political Analysis*, 25, 57–76. [2,5]