Estimation and Statistical Inference for Synthetic Control Methods with Spillover Effects

Jianfei Cao Connor Dowd*

October 31, 2018

Abstract

Synthetic control methods are often used in treatment effect estimation with panel data where only a few units are treated and a small number of post-treatment periods are available. Current estimation and inference procedures for synthetic control methods do not allow for the existence of spillover effects. However, this assumption doesn't hold in many empirical applications. In this paper, we consider estimation and statistical inference for synthetic control methods, allowing for spillover effects. We propose estimators for both direct treatment effects and spillover effects and show they are asymptotically unbiased. Moreover, we propose an inferential procedure that is based on Andrews (2003)'s end-of-sample instability tests (P-test). Similar to Andrews' results, we show this procedure is generally inconsistent but asymptotically unbiased. In simulations, we confirm that the presence of spillovers renders current methods biased, whereas our methods yields corrects size and has good power properties. We apply our method to an empirical example that investigates the effect of California's tobacco control program as in Abadie et al. (2010).

1 Introduction

The synthetic control method (SCM) has gained popularity in empirical studies since its introduction in Abadie and Gardeazabal (2003). SCM is often used in treatment effect estimation where we observe a panel of data going back in time, but only a few units are treated and a small number of post-treatment periods. This happens frequently in the US when we consider state polices and have state-level aggregate data. One common approach to this setting is differences-in-differences, which treats the control group as parallel to the treatment group, and looks at the average difference between the two. The basic insight of SCM is to model the relationship between the treated and untreated units using pre-treatment data,

*The authors thank Max Farrell and Christian Hansen for invaluable guidance.

essentially trying to create a data-driven weighted average. Then SCM uses the post-treatment data to predict the counter-factual values of the treated unit, which gives us the "synthetic control".

However, SCM and all of its variants assume explicitly or implicitly that untreated units are not affected by the treatment. That is, they rely on the Stable Unit Treatment Value Assumption (SUTVA). This is natural, since SCM uses post-treatment control units to predict the counter-factual values of the treated units. If SUTVA does not hold, the resulting treatment effect estimator could be heavily biased.

In this paper, we relax SUTVA and look into the case where spillover effects are allowed. We consider a model with spillover effects and provide a method that estimates treatment effect and spillover effects simultaneously. We also propose an inferential procedure that allows people to test for hypotheses such as no treatment effect or no spillover effects.

To fix ideas, we consider the Rubin's potential outcome model, with only one unit being treated. That is, let

$$y_{i,t} = \begin{cases} y_{i,t}(1), \text{ if } d_t = 1, \\ y_{i,t}(0), \text{ otherwise,} \end{cases} \tag{1}$$

and

$$d_t = \begin{cases} 1, \text{ if } i = 1 \text{ and } t = T+1, \\ 0, \text{ otherwise,} \end{cases} \tag{2}$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T+1$. The treatment indicator $d_t = 1$ means unit 1 is treated at time t, and $d_t = 0$ means otherwise. Note that only unit 1 is treated (at $t = T+1$). We consider the case where N is fixed and T goes to infinity. For simplicity, we only consider the case with one post-treatment period. Everything discussed here can be naturally extended to cases with multiple post-treatment periods.

Let $\alpha_i = y_{i,T+1}(1) - y_{i,T+1}(0)$, which can be the treatment effect or spillover effect depending on i. If we are interested in estimating the treatment effect α_1, a popular choice is the synthetic control estimator. Let $x_t = (1, y_{1,t}, y_{2,t}, \ldots, y_{N,t})'$, then a synthetic control weight estimator for a given closed convex set $\Lambda \subset \mathbb{R}^{N+1}$ is

$$\hat{\beta}_\Lambda = \underset{\beta \in \Lambda}{\arg\min} \sum_{t=1}^{T} (y_{1,t} - x_t'\beta)^2. \tag{3}$$

An estimator of α_1 is given by

$$\hat{\alpha}_1 = y_{1,T+1} - x_{T+1}'\hat{\beta}_\Lambda, \tag{4}$$

i.e. the counter-factual value $y_{1,T+1}(0)$ is approximated by $x_{T+1}'\hat{\beta}_\Lambda$. Popular choices of Λ include the original synthetic control method in Abadie and Gardeazabal (2003) and Abadie et al. (2010) $\{0\} \times \{0\} \times \Delta_{N-1}$, The demeaned synthetic control in Ferman and Pinto (2016) $\mathbb{R} \times \{0\} \times \Delta_{N-1}$, and the modified synthetic control in Li (2017) $\mathbb{R} \times \{0\} \times \mathbb{R}_+^{N-1}$.[1] We will mainly use the demeaned synthetic control restriction in this paper. That is, we do not restrict the intercept but require other coefficients

[1]Where $\Delta_{N-1} = \{\theta \in \mathbb{R}^{N-1} : \theta_i \geq 0 \text{ for each } i, \sum_{i=1}^{N-1} \theta_i = 1\}$ is a $(N-1)$-dimensional simplex.

to be positive and sum up to one. Intuitively, this can be seen as relaxing the common notion that SCM requires your treated unit to be on the 'interior' of your data. The demeaned SCM can also be seen as imposing a large degree of regularization on the data – as for $n-1$ coefficients, the LASSO style restriction of summing to no more than 1 is imposed. Note that the choice of restrictions depends on what a researcher believes about the DGP and that our analysis extends to most choices. See Doudchenko and Imbens (2016) for a discussion.

Under SUTVA and regularity conditions within a factor model, this estimator is shown to be inconsistent but asymptotically unbiased by Ferman and Pinto (2016). However, in the presence of spillover effect, this estimator can be severely biased. The reasons are (i) x_{T+1} is contaminated by the spillover effect, which results in a biased estimator of $y_{1,T+1}(0)$. (ii) The spillover often happens in the control units that the synthetic control method puts a lot of weight on, since they are more related to the treated unit.

The goal of this paper is to relax the SUTVA condition and to perform estimation and tests. We consider a factor model as the data generating process. To facilitate the estimation, we assume that the treatment effect and the spillover effects are linear in some underlying parameters. For each unit, we estimate a linear model between it and all the other units, using SCM with pre-treatment data. Thanks to the linear spillover structure, we are able to obtain asymptotically unbiased estimators for the treatment and spillover effects. We also characterize the asymptotic distribution of the estimator.

Furthermore, we propose an inferential procedure based on Andrews (2003)'s end-of-sample instability test (P-test). We first generalize P-test to the synthetic control method without spillover effects, and then generalize it further to incorporate cases with spillover effects. Similar to P-test, our testing procedures use the idea of approximating the null distribution of the statistic using pre-treatment data.

This paper mainly contributes to three streams of literature. First, it complements the fast-developing literature on synthetic control inference by relaxing SUTVA. Due to its popularity among empirical researchers, many formal results have been developed for statistical inference in similar settings. For example, Conley and Taber (2011) consider hypothesis testing in a similar data structure where only a few units are treated and both pre- and post-treatment periods are short. They consider *difference-in-difference*, which can be treated as a special case of SCM, and use control units to form the null distribution of the statistic. In Ferman and Pinto (2017) and Hahn and Shi (2016), similar ideas are used to conduct placebo tests that permute across observed units. Among all, Chernozhukov et al. (2017) is the most related to our work, since they also use outcomes across time periods rather than across units like the above citations. Unlike other works, Li (2017) proposes an innovative testing procedure that is based on the idea of projection onto convex sets and results in Fang and Santos (2014). However, none of the papers mentioned above allows for existence of spillover effects. Our methods provides formal statistical results in this setting, without assuming SUTVA.

Second, we contribute to the spillover effects estimation literature. In this literature, Manski (1993) introduces the reflection problem and since then the linear-in-means model has become a prominent tool in studies of spillover effects. Duflo and Saez (2003) use a randomized experiment approach to deal with the spillover effects. For the panel data settings, Manresa (2016) assumes a stable network structure across time and uses Lasso to estimate the spillover effects. However, the literature seldom looks at the panel data setting with only a few treated units and short post-treatment periods. This is partly because we don't usually have enough information about the spillover effects in this particular setting. In our setting we are able to delimit the amount of information needed ahead of time. Specifically, we overcome this problem by requiring that the spillover structures be pre-specified and follow a pattern that is linear in some underlying parameters. With that pre-specification, under our identification condition we can estimate the spillover effects and perform statistical tests on the spillovers.

Third, our results extend the literature on Andrews (2003)'s end-of-sample instability tests. Andrews (2003) uses data across time periods to approximate the null distribution of the test statistic, and apply this idea to OLS, IV, and GMM. Chernozhukov et al. (2017) propose a permutation method that is more general, but similar in cases where serial correlation matters. We extend this idea to the SCM case, and further to more complicated cases with spillover effects. As Andrews and Kim (2012) extends Andrews (2003)'s results to the co-integrated cases, we also show that our method is still valid for a co-integrated factor model.

The remainder of this paper is organized as follows. Section 2 introduces a factor model with spillover effects, proposes an estimator of the spillover effects and derives its asymptotic distribution. Section 3 considers P-test introduced by Andrews (2003) and Andrews and Kim (2012), and explains how it can be applied in our settings. Section 4 extends our methods to cases with multiple treated units and/or multiple post-treatment periods. Monte Carlo simulation results are presented in Section 5. In Section 6, we present an empirical example of our method. Section 7 concludes.

2 Estimation

2.1 A factor model with linear spillover effects

We follow Ferman and Pinto (2016) and consider a factor model such that for $i = 1, \dots, N$ and $t = 1, \dots, T+1$,

$$y_{i,t}(0) = \delta_t + \lambda_t' \mu_i + \epsilon_{i,t}, \tag{5}$$

where λ_t is F-dimensional common factors, and $\epsilon_{i,t}$ is noise that is uncorrelated with λ_t. For notation simplicity, write $Y_t(0) = (y_{1,t}(0), \dots, y_{n,t}(0))'$, $Y_t = (y_{1,t}, \dots, y_{n,t})'$, and $\epsilon_t = (\epsilon_{1,t}, \dots, \epsilon_{n,t})'$.

As a reminder we are looking at additive treatment effects of the form:

$$Y_t(1) = Y_t(0) + \alpha$$

Assume that the spillover effect is a linear transformation of some unknown parameter $\gamma \in \mathbb{R}^k$, i.e. $\alpha = A\gamma$. Typically, γ has much less dimensions than α does. Here are some examples that fit in this framework.

Example 1. *Assume the spillover effect shrink as the geometric distance goes up, where for $i = 2, \ldots, N$, $\alpha_i = b\exp(-d_i)$ for some b and d_i is the distance between unit 1 and unit i. Then, we have*

$$A = \begin{bmatrix} 1 & 0 \\ 0 & \exp(-d_2) \\ \vdots & \vdots \\ 0 & \exp(-d_N) \end{bmatrix}, \ \gamma = \begin{bmatrix} \alpha_1 \\ b \end{bmatrix}.$$

Example 2. *Assume every control units are equally affected by the spillover effects, i.e.*

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{bmatrix}, \ \gamma = \begin{bmatrix} \alpha_1 \\ b \end{bmatrix}.$$

Example 3. *Assume the spillover effect takes place at some known locations but not the others, while the levels of spillover effects are allowed to vary across those units. For example, there are spillovers at locations whose distance to unit 1 is less than $\bar{d}$. Then, the treatment and spillover effect vector can also be represented by $A\gamma$. Write the index set of spillover-exposed units as $\{k_j\}_{j=1}^p = \{i \in \{2, \ldots, N\} : d_i \leq \bar{d}\}$, then*

$$A = \begin{bmatrix} 1 & 0 \\ 0 & A^* \end{bmatrix}, \ \gamma = \begin{bmatrix} \alpha_1 \\ \alpha_{k_1} \\ \vdots \\ \alpha_{k_p} \end{bmatrix},$$

where $A^* \in \mathbb{R}^{(N-1)\times p}$ is such that the k_j-th row of A^* is a row unit vector with one at the j-th entry, and it has zeros everywhere else.

In order to back out the spillover effects, we re-formulate the model in the following way. For each i and t, we write $y_{i,t}(0)$ as a constant plus a weighted average of other outcomes at time t. That is, let

$$y_{i,t}(0) = a_i + Y_t(0)'b_i + u_{i,t}, \tag{6}$$

where the i-th entry of b_i is 0, where a_i and b_i are probability limit of the synthetic control weight estimator and $u_{i,t}$ is defined by this equation. Namely, let

$$\begin{pmatrix} \hat{a}_i \\ \hat{b}_i \end{pmatrix} = \underset{(a,b)\in W_i}{\arg\min} \sum_{t=1}^{T} (y_{i,t} - a - Y_t b')^2, \tag{7}$$

where $W_i = \{\beta = (\beta_0, \beta_1, \ldots, \beta_N)' \in \mathbb{R} \times \mathbb{R}_+^N : \beta_i = 0, \sum_{j=1}^N \beta_j = 1\}$. Then, let

$$a_i = \text{plim } \hat{a}_i,\ \ b_i = \text{plim } \hat{b}_i, \tag{8}$$

and we only consider cases where they are well-defined (see Lemma 1 for examples where $\hat{a}_i$ and $\hat{b}_i$ converge). Stacking equation (6) for all i's gives

$$Y_t(0) = a + BY_t(0) + u_t, \tag{9}$$

where $a = (a_1, \ldots, a_N)'$, i-th row of B is b_i' and $u_t = (u_{1,t}, \ldots, u_{N,t})'$. For $t = T+1$, this becomes

$$(I - B)(Y_{T+1} - \alpha) - a = u_{T+1}, \tag{10}$$

where $Y_{T+1} = (y_{1,T+1}, \ldots, y_{N,T+1})'$. We want to use this equation to estimate the spillover effect.

Note that (6) is not obtained by solving for λ_t using (5). In general, a_i and b_i will the probability limit of regressing $y_{i,t}$ on other outcomes using synthetic control methods, which typically does not coincide with the weights that reconstruct the factor loadings (Ferman and Pinto (2016)).

Define $M = (I - B)'(I - B)$. Our identification condition is:

Condition ID. *Identification Condition $A'MA$ is non-singular*

Remarks: 1. This implies $(I - B)A$ has full rank, this means that the submatrix of $(I - B)$ obtained by eliminating columns that correspond to irrelevant units has full rank.

2. Clearly then A must perform some degree of dimension reduction for us to identify point estimates, the extent of which depends on the variation in B.

3. Variation in B is what will identify our parameters.

We focus on two sets of technical conditions in our discussion.

Condition ST (model with stationary common factors)**.** *Assume $\{(\delta_t, \lambda_t, \epsilon_t)\}_{t\geq 1}$ is stationary, ergodic for the first and second moments, and has finite $(2+\delta)$-moment for some $\delta > 0$. Assume $cov[Y_t(0)] = \Omega_y$ is positive definite.*

Remarks: 1. We show in the proof of Lemma 1 that in this case (under the identification condition):

$$b_i = \underset{w \in W^{(i)}}{\arg\min}\ (w - e_i)'\Omega_y(w - e_i), \tag{11}$$

$$a_i = E[y_{i,1}(0) - Y_1(0)'b_i], \tag{12}$$

where e_i is a unit vector with one at the i-th entry and zeros everywhere else, and $W^{(i)} = \{(w_1, \dots, w_N) \in \mathbb{R}^N_+ : w_i = 0, \sum_{j \neq i} w_j = 1\}$. Note that b_i does not in general recover factor structure, in the sense that $\mu_i \neq (\mu_1, \dots, \mu_N)b_i$ in general.

2. We do not impose any restriction on the factor loadings $\{\mu_i\}_{i=1}^N$ except for Ω_y being positive definite. In the stationary case, the key for the treatment estimator to be unbiased and the test proposed below to be valid is to include an intercept in the optimization problem (7).

Condition CO (model with cointegrated $\mathcal{I}(1)$ common factors)**.** *Rewrite Equation* (5) *as*

$$y_{i,t}(0) = (\lambda_t^1)'\mu_i^1 + (\lambda_t^0)'\mu_i^0 + \epsilon_{i,t}, \tag{13}$$

and δ_t can be either in λ_t^1 or λ_t^0. Assume $\{(\lambda_t^0, \epsilon_t)\}_{t \geq 1}$ is stationary, ergodic for the first and second moments, and has finite 4*-th moment. Without loss of generality, $E[\epsilon_{i,t}] = 0$. Assume $\{\lambda_t^1\}_{t \geq 1}$ is $\mathcal{I}(1)$. Further assume for each i, $y_{i,t}(0)$ is such that weak convergence holds for $T^{-1/2}\sum_{t=1}^{[rT]} y_{i,t}(0) \Rightarrow \nu_i(r)$, where $\Rightarrow$ is weak convergence and process $\nu_i(r)$ is defined on* $[0,1]$ *and has bounded continuous sample path almost surely. For each i, let $W^{(i)} = \{(w_1, \dots, w_N) \in \mathbb{R}^N_+ : w_i = 0, \sum_{j \neq i} w_j = 1\}$. Assume for each i, there exists $w^{(i)} \in W^{(i)}$ such that $\mu_i^1 = \sum_{j=1}^N w_j^{(i)}\mu_j^1$. That is, $(w^{(i)} - e_i)$ is a cointegrating vector for $Y_t(0)$, where e_i is a unit vector with i-th entry being one and zeros everywhere else.*

2.2 Estimation

We form estimators for (a, B) using synthetic control methods as in (7). We do that for each $i = 1, \dots, N$, pretending i is the treated unit and other units are controls. Then, the estimators for a and B are $\hat{a} = (\hat{a}_1, \dots, \hat{a}_N)'$ and $\hat{B} = (\hat{b}_1, \dots, \hat{b}_N)'$, respectively. Let $\hat{M} = (I - \hat{B})'(I - \hat{B})$ be an estimator for M. Let an estimator of γ be such that

$$\begin{aligned}\hat{\gamma} &= \underset{g \in \mathbb{R}^k}{\arg\min} \|(I - \hat{B})(Y_{T+1} - Ag) - \hat{a}\| \\ &= (A'\hat{M}A)^{-1}A'(I - \hat{B})'((I - \hat{B})Y_{T+1} - \hat{a}).\end{aligned} \tag{14}$$

Note that the FOC implies

$$A'(I - B)'u_{T+1} = 0, \tag{15}$$

i.e. it requires some weighted sum of the residuals to be zero. Then the treatment and spillover effect vector α can be estimated by $\hat{\alpha} = A\hat{\gamma}$.

Assumption 1. (a) $\{u_t\}_{t\geq 1}$ *is stationary, and has mean zero.*

(b) $A'MA$ *is non-singular.*

(c) $\|\hat{a} - a\| = o_p(1)$, $\|\hat{B} - B\| = o_p(1)$

(d) $\|(\hat{B} - B)Y_{T+1}(0)\| = o_p(1)$.

Remark: 1. Part (c) is fairly strong in the sense that it excludes polynomial time trends.

Lemma 1. *Either Condition ST + Condition ID or Condition CO + Condition ID implies Assumption 1.*

Theorem 1. *Suppose Assumption 1 holds. Then,* $\hat{\alpha} - (\alpha + Gu_{T+1}) \to_p 0$ *as* $T \to \infty$*, where* $G = A(A'MA)^{-1}A'(I - B)'$*. Moreover,* $E[Gu_{T+1}] = 0$.

The structure of the limiting distribution is similar to the case as in Ferman and Pinto (2016), as it's inconsistent but asymptotically unbiased, in the sense that the difference between the estimator and the true value has mean zero. Note that linearity of α in γ is crucial here in order for us to obtain the asymptotic unbiasedness.

2.3 Efficiency

Now we form an estimator of α with possibly lower variance. For a positive define matrix $W \in \mathbb{R}^N$, we minimize $\|W^{1/2}\epsilon_{T+1}\|$ instead of $\|\epsilon_{T+1}\|$. The resulting estimator is

$$\begin{aligned}\hat{\gamma}_W &= \underset{g\in\mathbb{R}^k}{\arg\min}\|W^{1/2}((I - \hat{B})(Y_{T+1} - Ag) - \hat{a})\| \\ &= (A'\hat{M}_W A)^{-1}A'(I - \hat{B})'W((I - \hat{B})Y_{T+1} - \hat{a}),\end{aligned} \tag{16}$$

where $\hat{M}_W = (I - \hat{B})'W(I - \hat{B})$. The corresponding estimator for α is $\hat{\alpha}_W = A\hat{\gamma}_W$. Let $\Omega = Cov[u_1]$ and W_T^e be a consistent estimator of Ω^{-1}. Then the efficient estimator of α is defined by

$$\hat{\alpha}^e = \hat{\alpha}_{W_T^e} \tag{17}$$

Let $M_W = (I - B)'W(I - B)$, $G_W = A(A'M_W A)^{-1}A'(I - B)'W$ for some weighting matrix W, $W^e = \Omega^{-1}$, $M^e = M_{W^e}$, and $G^e = G_{W^e}$. Then,we have the following results.

Proposition 1. *Suppose Assumption 1 holds,* W_T *is a consistent estimator for* W*, and* W_T^e *is a consistent estimator for* W^e*. Then,* $\hat{\alpha}_{W_T} - (\alpha + G_W u_{T+1}) \to_p 0$*, and specifically,* $\hat{\alpha}^e - (\alpha + G^e u_{T+1}) \to_p 0$*, as* $T \to \infty$*. Moreover,* $(Cov[G_W u_{T+1}] - Cov[G^e u_{T+1}])$ *is positive semi-definite.*

In practice, we need to estimate Ω, and the drawback is that we need relatively large sample size (large T) to have a good approximation.

3 Andrews' P-Test

In this section, we discuss formal results on inference. In Section 3.1, we consider the case without spillover effects, and state the assumptions under which Andrews' P test is valid. In Section 3.2, we generalize P test to cases where spillover effects cannot be ignored.

At a high level, the test operates by using a leave-one-out procedure across time periods. That for each time period $j = 1, ..., T$, we hold out j and estimate the weights and intercepts using $t = 1, ..., j-1, j+1, ..., T$. Then we can combine those weights with our observed outcomes in time period j to create a 'treatment effect estimate' α_{-j}. When we perform this procedure for each pre-treatment time period, we create a null distribution for α_{T+1}. Then we can compare that null distribution to our observed effects to perform inference on a variety of hypotheses.

3.1 Cases without spillover effects

Suppose for now there is no spillover effects, i.e. $\alpha_2 = \cdots = \alpha_N = 0$. We want to test for the existence of treatment effect on unit 1. The null and alternative hypothesis of interest are

$$\begin{cases} H_0 : \alpha_1 = 0, \\ H_1 : \alpha_1 \neq 0. \end{cases} \tag{18}$$

The test procedure we consider here is the end-of-sample instability test (P-test) in Andrews (2003). A relevant work is Chernozhukov et al. (2017). Since we allow for serial correlation and only look at one post-treatment period, we will use *moving blocks* permutation with each block having only one unit, when applying the test procedure in Chernozhukov et al. (2017). This effectively makes their test exactly P-test in Andrews (2003).

We assume the α_1 is independent of T under H_1. That is, we consider fixed, not local, alternatives, as in Andrews (2003) and Andrews and Kim (2012). That is, α_1 does not change as T grows, which facilitates our analysis of the test statistic under H_1.

Now we translate our hypothesis into the linear formulation considered in Andrews (2003). Namely, we have

$$y_t = \begin{cases} a_1 + Y_t' b_1 + u_{1,t}, & \text{for } t = 1, \ldots, T, \\ a_1^* + Y_t' b_1 + u_{1,t}, & \text{for } t = T+1. \end{cases} \tag{19}$$

A non-zero treatment effect is equivalent to a shift in the intercept a_1 (or equivalently, change of the distribution of $u_{1,t}$, at $t = T+1$). The null and alternative hypothesis (18) become

$$\begin{cases} H_0 : a_1^* = a_1, \\ H_1 : a_1^* \neq a_1. \end{cases} \tag{20}$$

Let the synthetic control regression residuals be $\hat{u}_{1,t} = y_{1,t} - \hat{a}_1 - Y_t'\hat{b}_1$. The test statistic is defined by

$$P = \hat{u}_{T+1}^2. \tag{21}$$

For notational simplicity, let $\hat{\beta}_1 = (\hat{a}_1, \hat{b}_1')'$ and $x_t = (1, Y_t')'$. For $\beta \in \mathbb{R}^{N+1}$, define

$$P_t(\beta) = (y_{1,t} - x_t'\beta)^2. \tag{22}$$

Then, $P = (y_{1,T+1} - x_{T+1}'\hat{\beta}_1)^2 = P_{T+1}(\hat{\beta}_1)$. Let P_∞ be a random variable with the same distribution as $P_{T+1}(\beta_1)$ with $\beta_1 = (a_1, b_1')'$.

Let $P_t = P_t(\hat{\beta}_1^{(t)})$, where $\hat{\beta}_1^{(t)} = \hat{\beta}_1$ for each t.[2] Define

$$\hat{F}_{P,T}(x) = \frac{1}{T}\sum_{j=1}^{T} \mathbb{1}\{P_t \leq x\}, \tag{23}$$

and let $F_P(x)$ be the distribution function of $P_1(\beta_1)$. Finally, let $\hat{q}_{P,1-\alpha} = \inf\{x \in \mathbb{R} : \hat{F}_{P,T}(x) \geq 1-\alpha\}$, and $q_{P,1-\alpha}$ be the $(1-\alpha)$-quantile of $P_1(\beta_1)$.

Assumption 2. (a) $\{u_t\}_{t\geq 1}$ *are stationary, ergodic, and has mean zero.*

(b) $E[|u_t|] < \infty$.

(c) $\exists$ *a non-random sequence of positive definite matrices* $\{C_T\}_{T\geq 1}$ *such that* $\max_{t\leq T+1} \|C_T^{-1}x_t\| = O_p(1)$

(d) $\|C_T(\hat{\beta}_1 - \beta_1)\| = o_p(1)$, *and* $\max_{t=1,\ldots,T} \|C_T(\hat{\beta}_1^{(t)} - \beta_1)\| = o_p(1)$.

(e) *The distribution function of* $P_1(\beta_1)$ *is continuous and increasing at its* $(1-\alpha)$*-quantile.*

Lemma 2. *Suppose the distribution function of* $P_1(\beta_1)$ *is continuous and increasing at its* $(1-\alpha)$*-quantile. Then, Condition ID and either Condition ST or Condition CO implies Assumption 2.*

Theorem 2. *Suppose Assumption 2 holds. Then, as* $T \to \infty$,

(a) $P \to_d P_\infty$ *under* H_0 *and* H_1,

(b) $\hat{F}_{P,T}(x) \to_p F_P(x)$ *for all* x *in a neighborhood of* $q_{P,1-\alpha}$ *under* H_0 *and* H_1,

(c) $\hat{q}_{P,1-\alpha} \to_p q_{P,1-\alpha}$ *under* H_0 *and* H_1,

(d) $\Pr(P > \hat{q}_{P,1-\alpha}) \to \alpha$ *under* H_0.

3.2 Cases with spillover effects

Now we allow for non-zero spillover effects. We propose a testing procedure that is based on Andrews' P-test and accounts for the spillover effect. The null and alternative hypothesis we consider are $H_0 : C\alpha = d$ and $H_1 : C\alpha \neq d$, with C and d known. For example, we want to test for the hypothesis that there is no treatment effect at the treated unit (unit 1), then we let $C = (1, 0, 0, \ldots, 0) \in \mathbb{R}^{1\times N}$ and $d = 0$. This

[2] You can also use leave-one-estimator to construct P_t as in Andrews (2003) and Andrews and Kim (2012). For $t = 1, \ldots, T$, the leave-one-out estimator $\hat{\beta}_1^{(t)}$ is defined by the synthetic control weight estimator using only observations indexed by $s = 1, \ldots, t-1, t+1, \ldots, T$.

effectively makes Section 3.1 a special case of our test (although Theorem 2 has slightly stronger results than Theorem 3 does). Another example is that we want to test for whether there is spillover effect, then we can let $C = [\mathbf{0}\ I_{N-1}] \in \mathbb{R}^{(N-1)\times N}$ and $d = (0, \ldots, 0)' \in \mathbb{R}^{(N-1)\times 1}$.

The test statistic we consider here is $P = (C\hat{\alpha} - d)' W_T (C\hat{\alpha} - d)$ for some weighting matrix $W_T \to_p W$. Recall $G = A(A'MA)^{-1}A'(I - B)$ and can be consistently estimated by $\hat{G} = A(A'\hat{M}A)^{-1}A'(I - \hat{B})$ if $\hat{B} \to_p B$. By Theorem 1, P is asymptotically equivalent to $u_{T+1}' G'C'WCGu_{T+1}$. To construct critical values, define

$$P_t(\theta) = (Y_t - \theta x_t)' G'C'WCG(Y_t - \theta x_t), \tag{24}$$

and

$$\hat{P}_t(\theta) = (Y_t - \theta x_t)' \hat{G}'C'W_T C\hat{G}(Y_t - \theta x_t), \tag{25}$$

for some $\theta \in \mathbb{R}^{N\times(N+1)}$, $x_t = (1, Y_t')'$, and $\hat{G} = A(A'\hat{M}A)^{-1}A'(I - \hat{B})'$. Let $\hat{P}_t = \hat{P}_t(\hat{\theta}^{(t)})$, where $\hat{\theta}^{(t)} = \hat{\theta}$ for each t.[3] Let $P_\infty = P_1(\theta_0)$ for $\theta_0 = [a\ B]$. Define

$$\hat{F}_{P,T}(x) = \frac{1}{T}\sum_{j=1}^{T} \mathbb{1}\{\hat{P}_t \le x\}, \tag{26}$$

and let $F_P(x)$ be the distribution function of P_∞. Finally, let $\hat{q}_{P,1-\alpha} = \inf\{x \in \mathbb{R} : \hat{F}_{P,T}(x) \ge 1 - \alpha\}$, and $q_{P,1-\alpha}$ be the $(1-\alpha)$-quantile of P_∞.

Assumption 3. (a) *Assumption 1 holds.*

(b) *$\{u_t\}_{t\ge1}$ is ergodic and $E[\|u_t\|] < \infty$.*

(c) *There exists a non-random sequence of positive definite matrices $\{D_T\}_{T\ge1}$ such that $\max_{t\le T+1} \|D_T^{-1} x_t\| = O_p(1)$.*

(d) *$\|(\hat{\theta} - \theta_0)D_T\|_F = o_p(1)$, and $\max_{t=1,\ldots,T} \|(\hat{\theta}^{(t)} - \theta_0)D_T\|_F = o_p(1)$, where $\|\cdot\|_F$ is the Frobenius norm.*

(e) *The distribution function of $P_1(\theta_0)$ is continuous and increasing at its $(1-\alpha)$-quantile.*

(f) *$W_T \to_p W$ as $T \to \infty$.*

Lemma 3. *Suppose the distribution function of $P_1(\theta_0)$ is continuous and increasing at its $(1-\alpha)$-quantile. Also assume Condition ID holds. Then, Assumption 3 is satisfied if either of these holds:*

(i) *Condition ST with $W_T = I$ or $W_T = (CG(T^{-1}\sum_{t=1}^{T} \hat{u}_t\hat{u}_t')G'C')^{-1}$;*

(ii) *Condition CO with $W_T = I$.*

Theorem 3. *Suppose Assumption 3 holds. Then, under H_0, as $T \to \infty$,*

(a) *$P \to_d P_\infty$,*

(b) *$\hat{F}_{P,T}(x) \to_p F_P(x)$ for all x in a neighborhood of $q_{P,1-\alpha}$,*

[3]Similar to the case without spillover effects, the leave-one-out estimator $\hat{\theta}^{(t)} = [\hat{a}^{(t)}\ \hat{B}^{(t)}]$ is defined by the synthetic control weight estimator using only observations indexed by $s = 1, \ldots, t-1, t+1, \ldots, T$.

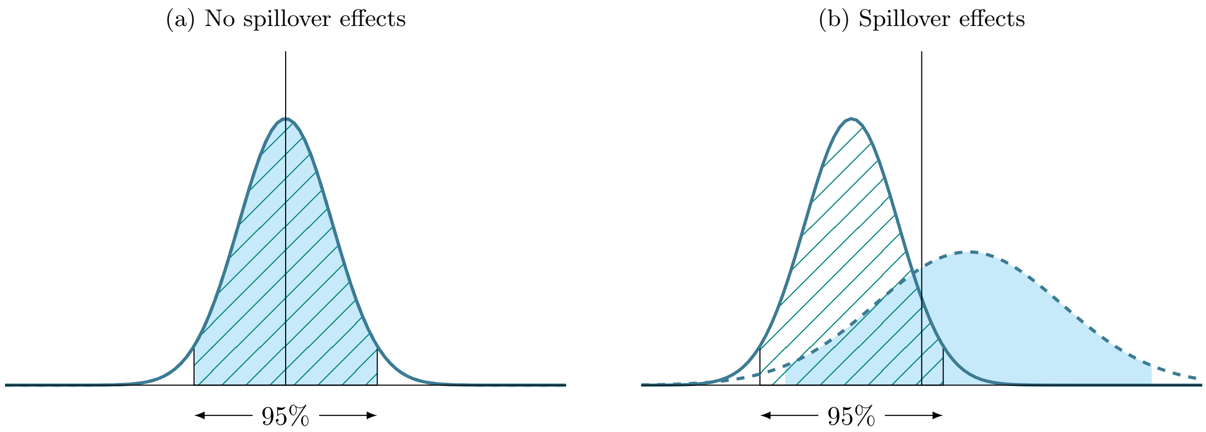

Figure 1: Placebo test

(c) $\hat{q}_{P,1-\alpha} \to_p q_{P,1-\alpha}$,

(d) $\Pr(P > \hat{q}_{P,1-\alpha}) \to \alpha$.

3.3 Other testing procedures

When we allow for existence of non-zero spillover effects, the existing testing procedures will have poor performance. Here we intuitively explain what happens to placebo test as in Abadie and Gardeazabal (2003) and CWZ test as in Chernozhukov et al. (2017) in the presence of spillover effects.

Now suppose we want to test for treatment effect being zero and are not aware of the spillover effects. Placebo test and CWZ test are similar in the sense that placebo test exploits variations of $\{\hat{u}_{i,T+1}\}_{i=1}^{N}$ while CWZ test uses variations of $\{\hat{u}_{1,t}\}_{t=1}^{T+1}$.

We look at the placebo test first. When there is no spillover effect, the distribution of $\hat{u}_{1,T+1}$ and distribution of $\{\hat{u}_{i,T+1}\}_{i=2}^{N}$ overlap asymptotically. As shown in Figure 1, when there is positive spillover effects, we will underestimate the treatment effect and the density function of $\hat{u}_{1,T+1}$ moves to the left; some of the control units shift to the right because of the positive spillovers, so density of $\{\hat{u}_{i,T+1}\}_{i=2}^{N}$ moves to the right but gets wider. In terms of test, the shift of $\hat{u}_{1,T+1}$ is offset by the wider density of $\{\hat{u}_{i,T+1}\}_{i=2}^{N}$ (harder to reject H_0), which explains why in Table 3 the empirical sizes of placebo test for $T = 50$ and 200 cases are not too far away from 0.05. It effectively becomes much more conservative and has low power as shown in Table 4.

Now we consider CWZ test. When there is no spillover effect, the distribution of $\hat{u}_{1,T+1}$ and distribution of $\{\hat{u}_{1,t}\}_{t=1}^{T}$ overlap asymptotically. As shown in Figure 2, when there is positive spillover effect, we underestimate the treatment effect and the density function of $\hat{u}_{1,T+1}$ shifts to the left; density of $\{\hat{u}_{1,t}\}_{t=1}^{T}$ since they are pre-treatment but spillover only happens after the treatment. This results in an invalid test.

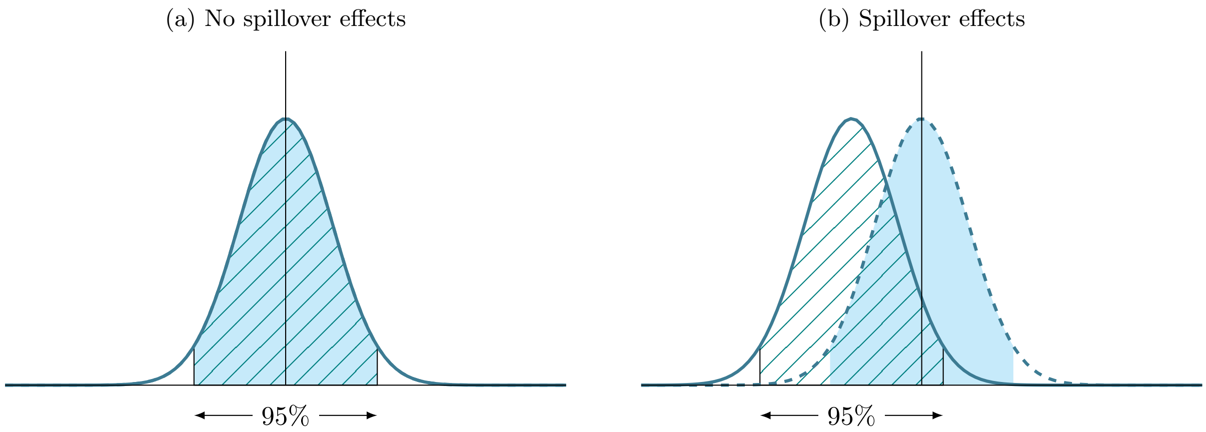

Figure 2: CWZ test

4 Extensions

4.1 Multiple treated units

Our method readily extends to cases where multiple units are treated. In our setting, spillover effects are not distinguished from treatment effects, since one can think of spillover as the treatment on the units that are not directly treated. With a correctly specified structure matrix A, we can perform estimation and testing just as previous sections. For example, suppose $N = 4$, unit 1 and unit 2 are treated, unit 3 is affected by spillover effect, and unit 4 is neither treated nor exposed to spillover effect. Then we can specify

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \tag{27}$$

and the resulting estimator $\hat{\gamma} = (\hat{\gamma}_1, \hat{\gamma}_2, \hat{\gamma}_3)'$ by (14) is such that $\hat{\gamma}_1$ and $\hat{\gamma}_2$ are the treatment effects estimator for unit 1 and unit 2, respectively, and $\hat{\gamma}_3$ is the spillover effect estimator for unit 3. Test can be performed accordingly. If one wants to test for the hypothesis that there is no spillover effects, the null is then $H_0 : C\alpha = d$, where $C = (0, 0, 1, 0)$ and $d = 0$.

4.2 Multiple post-treatment time periods

Suppose now we have observations of $\{y_{i,t}\}$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T + m$. Treatment is received at $t = T + 1$. The model becomes (with some abuse of notation)

$$Y_t = \begin{cases} Y_t(0), \text{ if } t \leq T \\ Y_t(0) + \alpha_t, \text{ otherwise.} \end{cases} \tag{28}$$

Note that we do not allow for spillovers in time. That is, the treatment effect or spillover effects cannot affect future selves. For each $t = T+1, \ldots, T+m$, we need to specify the spillover structure matrix A_t. Then, an estimator of α_t is

$$\hat{\alpha}_t = A_t(A_t'\hat{M}A_t)^{-1}A_t'(I-\hat{B})'((I-\hat{B})Y_t - \hat{a}). \tag{29}$$

That is, we treat $T+s$ period as $T+1$ and do the same procedure as before. For each $t = T+1, \ldots, T+m$, we can perform separate tests as introduces in previous sections.

To answer simultaneous questions such as whether there is spillover effect at all, we can extend P-test discussed above. Consider the null hypothesis $H_0 : C_t\alpha_t = d_t$ for $t = T+1, \ldots, T+m$. Let $\hat{P}_t$ be constructed as in Section 3.2 for $t = 1, \ldots, T$. For $t = T+1, \ldots, T+m$, let $\hat{P}_t = (C_t\hat{\alpha}_t - d_t)'W_T(C_t\hat{\alpha}_t - d_t)$. We the now form

$$P^{(t)} = \sum_{s=0}^{m-1} \hat{P}_{t+s} \tag{30}$$

for $t = 1, \ldots, T+1$. The test statistic is then $P^{(T+1)}$, and we use $\{P^{(t)}\}_{t=1}^{T}$ to form its null distribution.

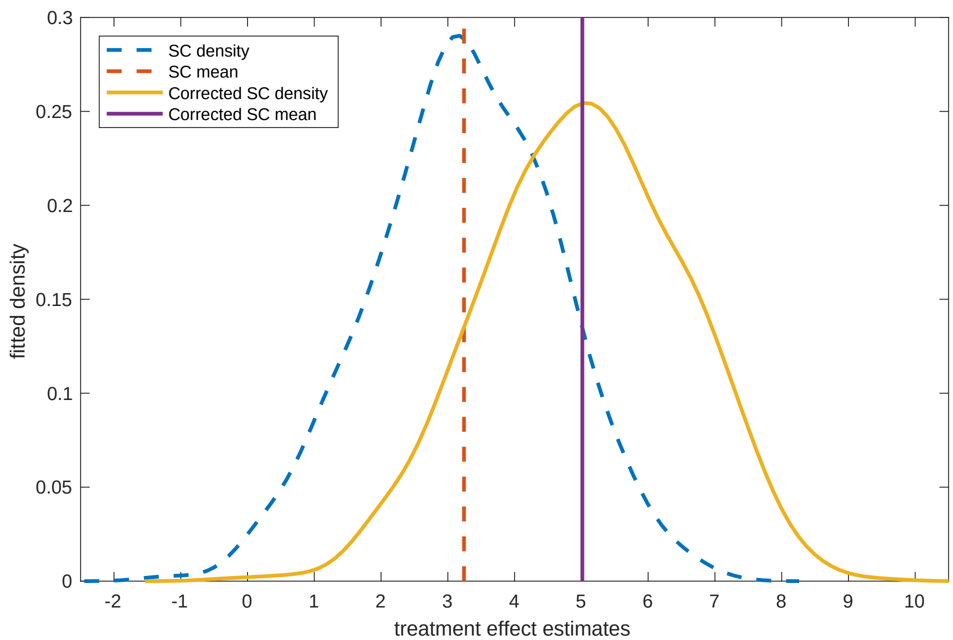

Figure 3: Distribution of treatment effect estimates.

5 Simulation

We present Monte Carlo simulation results in this section. For each case considered, we use 1000 simulation repetitions.

5.1 Estimation with spillover effects

In this subsection we examine the finite sample performance of our estimation procedure proposed in Section 2.2. The model consider here is similar to Li (2017), where $y_{i,t}(0)$ follows a factor model structure. We consider both stationary and $\mathcal{I}(1)$ case.

Table 1: Treatment effect estimation with stationary common factors.

	$N = 10$			$N = 30$			$N = 50$		
	$T = 15$	50	200	15	50	200	15	50	200
No spillover effects									
SC	-0.062	0.011	-0.003	0.114	-0.005	0.016	0.037	-0.041	-0.033
	(2.113)	(1.249)	(1.586)	(1.642)	(1.244)	(1.273)	(1.408)	(1.290)	(1.182)
CSC	-0.077	0.013	0.018	0.091	-0.012	0.010	0.042	-0.031	-0.040
	(2.618)	(1.417)	(1.710)	(1.974)	(1.362)	(1.486)	(1.741)	(1.516)	(1.270)
Concentrated spillover effects									
SC	-1.326	-0.986	-1.333	-0.756	-0.880	-1.543	-1.492	-1.070	-0.796
	(2.714)	(1.451)	(2.065)	(1.958)	(1.654)	(1.392)	(1.912)	(1.638)	(1.461)
CSC	0.267	0.025	0.140	0.248	0.038	0.025	-0.133	-0.055	0.110
	(2.554)	(1.425)	(1.756)	(1.897)	(1.435)	(1.250)	(1.700)	(1.581)	(1.408)
Spreadout spillover effects									
SC	-2.378	-1.910	-2.114	-2.245	-1.859	-2.398	-2.147	-2.112	-2.154
	(2.493)	(1.470)	(1.696)	(2.029)	(1.472)	(1.369)	(1.791)	(1.538)	(1.313)
CSC	-0.048	0.007	0.029	0.090	-0.025	0.018	0.037	-0.048	-0.028
	(2.740)	(1.438)	(2.061)	(2.231)	(1.296)	(1.602)	(1.643)	(1.450)	(1.290)

5.1.1 stationary case

The underlying factor model is

$$y_{i,t}(0) = \delta_t + \lambda_t' \mu_i + \epsilon_{i,t}, \tag{31}$$

Table 2: Treatment effect estimation with $\mathcal{I}(1)$ common factors.

	$N=10$			$N=30$			$N=50$		
	$T=15$	50	200	15	50	200	15	50	200
No spillover effects									
SC	-0.023	-0.018	-0.043	0.036	-0.088	-0.031	0.041	0.038	-0.038
	(1.873)	(1.642)	(1.772)	(1.708)	(1.539)	(1.900)	(1.915)	(1.810)	(1.866)
CSC	-0.021	-0.057	-0.017	0.037	-0.053	-0.044	0.007	0.013	-0.017
	(2.460)	(2.249)	(4.523)	(2.116)	(2.121)	(2.184)	(2.308)	(1.849)	(1.952)
Concentrated spillover effects									
SC	-1.185	-1.400	-2.234	-1.206	-2.026	-1.954	-1.316	-1.408	-2.325
	(2.421)	(1.854)	(1.856)	(2.269)	(1.921)	(2.079)	(2.449)	(2.043)	(1.976)
CSC	-0.021	-0.057	-0.017	0.037	-0.053	-0.044	0.007	0.013	-0.017
	(2.460)	(2.249)	(4.523)	(2.116)	(2.121)	(2.184)	(2.308)	(1.849)	(1.952)
Spreadout spillover effects									
SC	-2.088	-2.599	-2.885	-2.233	-2.536	-2.465	-2.219	-2.402	-2.889
	(2.390)	(1.779)	(1.795)	(2.101)	(1.759)	(2.037)	(2.249)	(1.921)	(1.900)
CSC	-0.029	0.027	-0.022	0.047	-0.008	0.010	0.022	0.006	-0.045
	(2.452)	(3.447)	(7.367)	(2.357)	(2.412)	(2.740)	(2.418)	(2.279)	(2.712)

where $\lambda_t = (\lambda_{1,t}, \lambda_{2,t}, \lambda_{3,t})'$,

$$\delta_t = 1 + 0.5\delta_{t-1} + \nu_{0,t}, \tag{32}$$

$$\lambda_{1,t} = 0.5\lambda_{1,t-1} + \nu_{1,t}, \tag{33}$$

$$\lambda_{2,t} = 1 + \nu_{2,t} + 0.5\nu_{2,t-1}, \tag{34}$$

$$\lambda_{3,t} = 0.5\lambda_{3,t-1} + \nu_{3,t} + 0.5\nu_{3,t-1}, \tag{35}$$

and $\epsilon_{i,t}$ and $\nu_{j,s}$ is i.i.d. $N(0,1)$ for each (i,j,s,t). Each entry of μ_i is drew from independent uniform distribution on $[0,1]$ and fixed for each repetition. At $t = T+1$, the observed outcome is $y_{i,T+1} = y_{i,T+1}(0) + \alpha_i$, where α_i is either treatment effect or spillover effect and is specified below. The treatment effect is set to 5 and the spillover effect is 3.

The empirical bias and variance (in parenthesis) of the treatment effect estimator using two methods are shown in Table 1. We consider three spillover patterns. *No spillover effects* is the case where unit 1 receives a treatment effect of 5 at $t = T+1$ and other units are not affected. *Concentrated spillover effects* is the case where 1/3 of the control units receive a spillover effect of 3. *Spreadout spillover effects* is the case where 2/3 of the control units receive a spillover effect of 3. SC is the original synthetic control method, and CSC is the corrected synthetic control method proposed in Section 2.2. Throughout we assume the coverage of spillover effect is known, but not other information, so A is constructed as in Example 3. For *No spillover effects*, we are being conservative and pretend 1/3 of the control units are exposed to spillover effects.

To better compare results, we fit the simulation results using kernel density for the $(N, T) = (10, 50)$ case with concentrated spillover effects and plot it in Figure 3.

5.1.2 $\mathcal{I}(1)$ case

For the $\mathcal{I}(1)$ case, the underlying factor model follows

$$y_{i,t}(0) = \lambda_t' \mu_i + \epsilon_{i,t}, \tag{36}$$

where $\lambda_t = (\lambda_{1,t}, \lambda_{2,t}, \lambda_{3,t})'$,

$$\lambda_{1,t} = \lambda_{1,t-1} + 0.5\nu_{1,t}, \tag{37}$$

$$\lambda_{2,t} = \lambda_{2,t-1} + 0.5\nu_{2,t}, \tag{38}$$

$$\lambda_{3,t} = 0.5\lambda_{3,t-1} + \nu_{3,t}, \tag{39}$$

and $\epsilon_{i,t}$ and $\nu_{j,s}$ follows i.i.d. $N(0, 1)$ for each (i, j, s, t). The factor loadings are constructed such that Condition CO is satisfied. Namely, we let $\mu_1 = (1, 0, 0)'$, $\mu_2 = (0, 1, 0)'$, $\mu_3 = (1, 0, 0)'$, $\mu_4 = (0, 1, 0)'$, and for μ_j with $j = 5, \ldots, N$, we draw independent uniform distribution on $[0, 1]$ for each entry and then normalize each loading vector such that three entries of each μ_j sum up to one. The constructed factor loadings are fixed for each repetition. Other settings are same with the stationary case. The results are shown in Table 2.

5.2 Test for treatment effect

In this section we compare test procedures against the null hypothesis $H_0 : \alpha_1 = 0$, i.e. the treatment effect is zero. The results are shown in Table 3 and Table 4. The DGP is exactly the same as in Section 5.1.1 (the stationary case), except that $\alpha_1 = 0$ (the null) for Table 3 and $\alpha_1 = 5$ (the alternative) for Table 4. Placebo test is as in Abadie and Gardeazabal (2003) and Hahn and Shi (2016). CWZ test is proposed by Chernozhukov et al. (2017). In this set-up, where data are serially correlated and we only look at one period after the treatment, CWZ is exactly our test (or Andrew's P-test) when assuming there is no spillover effect. P-test is the spillover-adjust test proposed in Section 3.2.

Among the three testing procedures, P-test has correct sizes and outperforms the other two methods in power. Placebo test has correct sizes in some cases but has lower power, and CWZ test over-rejects under null. The reasons are discussed in Section 3.3.

Table 3: Empirical rejection rate of testing for treatment effect under null.

	$N = 10$			$N = 30$			$N = 50$		
	$T = 15$	50	200	15	50	200	15	50	200
No spillover effects									
Placebo	0.000	0.000	0.000	0.072	0.053	0.062	0.034	0.031	0.040
CWZ	0.076	0.061	0.060	0.108	0.082	0.065	0.141	0.078	0.072
P-test	0.048	0.049	0.058	0.055	0.064	0.052	0.066	0.046	0.059
Concentrated spillover effects									
Placebo	0.000	0.000	0.000	0.066	0.046	0.116	0.035	0.029	0.026
CWZ	0.411	0.207	0.224	0.417	0.279	0.346	0.519	0.346	0.184
P-test	0.065	0.050	0.043	0.111	0.069	0.061	0.109	0.092	0.054
Spreadout spillover effects									
Placebo	0.000	0.000	0.000	0.129	0.063	0.147	0.060	0.059	0.072
CWZ	0.576	0.478	0.399	0.685	0.563	0.616	0.741	0.621	0.544
P-test	0.036	0.035	0.042	0.034	0.042	0.046	0.030	0.042	0.044

5.3 Test for existence of spillover effects

In this section we examine the power of the proposed test against the null hypothesis that there is no spillover effects. We also look into its behavior when the range of spillover effect is not exactly correctly specified. In this set of experiments, the level of spillover effects (the alternatives) vary from 0 to 2. We set $(N, T) = (20, 50)$ and $\alpha_1 = 5$. There are 9 units that is affected by spillover effects. Other settings follows exactly as in Section 5.1.1 (the stationary case). The model for the range of spillover is as in Example 3.

The empirical rejection rates against various levels of spillover effects using our method proposed in Section 3.2 are plotted in Figure 4. Here *not include enough* misses half of the units that are actually affected by the treatment (assuming that unit 1 and 4 other units are affected), *correct specification* assumes we know exactly which units are affected, and *include too many* assumes 15 units are affected in estimation (including unit 1, 9 units that are affected by spillover effects, and 5 units that are actually not affected).

The simulation results show that the proposed test is quite robust to model misspecification. Among the three cases, *include too many* is still correct specification but supposed to be more conservative, so it has less power than *correct specification* does. The range of spillover effects is misspecified in *not include enough*, but the test is still able to have correct size under null[4] and reasonable power under alternatives.

[4]The model is always correctly specified under null.

Table 4: Empirical rejection rate of testing for treatment effect under alternative.

	$N = 10$			$N = 30$			$N = 50$		
	$T = 15$	50	200	15	50	200	15	50	200
No spillover effects									
Placebo	0.000	0.000	0.000	0.908	0.939	0.966	0.922	0.936	0.931
CWZ	0.797	0.948	0.926	0.785	0.901	0.983	0.797	0.972	0.827
P-test	0.835	0.956	0.923	0.823	0.937	0.965	0.839	0.964	0.993
Concentrated spillover effects									
Placebo	0.000	0.000	0.000	0.461	0.502	0.448	0.465	0.434	0.464
CWZ	0.651	0.765	0.329	0.704	0.754	0.542	0.680	0.746	0.737
P-test	0.860	0.932	0.991	0.957	0.918	0.967	0.834	0.816	0.853
Spreadout spillover effects									
Placebo	0.000	0.000	0.000	0.348	0.378	0.331	0.305	0.255	0.294
CWZ	0.337	0.403	0.277	0.563	0.414	0.278	0.406	0.309	0.343
P-test	0.866	0.978	0.981	0.969	0.950	0.991	0.909	0.985	0.974

6 Empirical Example

To demonstrate our method, we've used it on the classic SCM example from Abadie et al. (2010) (ADH), which look at the effect of Prop 99 on California cigarette consumption. In this section, we will walk through the results from our method, with interruptions to point out quirks and key features.

Prop 99 intended to disincentivize smoking, primarily achieved by introducing a 0.25 tax on each pack of cigarettes. By measuring sales in California, ADH and others have attempted to determine the effect of the policy on smoking rates. However, traditional SCM is not guaranteed to produce an unbiased treatment effect estimator in the presence of spillover effects. In this tobacco control program example, we are concerned about two kinds of spillover effects. The first spillover is based on concerns about "leakage". A common problem with cigarette taxes is that measured local consumption might fall as people move their purchasing behavior across legal boundaries. In order to accommodate this, we allowed for a spillover affecting states neighboring California or one of its neighbors. The second spillover type we considered was a cultural change. If tobacco is discouraged in California, it might reduce the cultural appeal of smoking. Reasoning that the northeast is culturally close to the west coast, we allowed for the northeastern states to experience this cultural spillover.

One might also think that there could be a policy contamination whereby culturally close states also enact policies with similar targets. Our method can allow for this kind of spillover in our estimation. However, the initial paper was worried about that type of problem, and so 12 states which experienced legislative changes in the ensuing years were removed in that paper (and thus in our data).

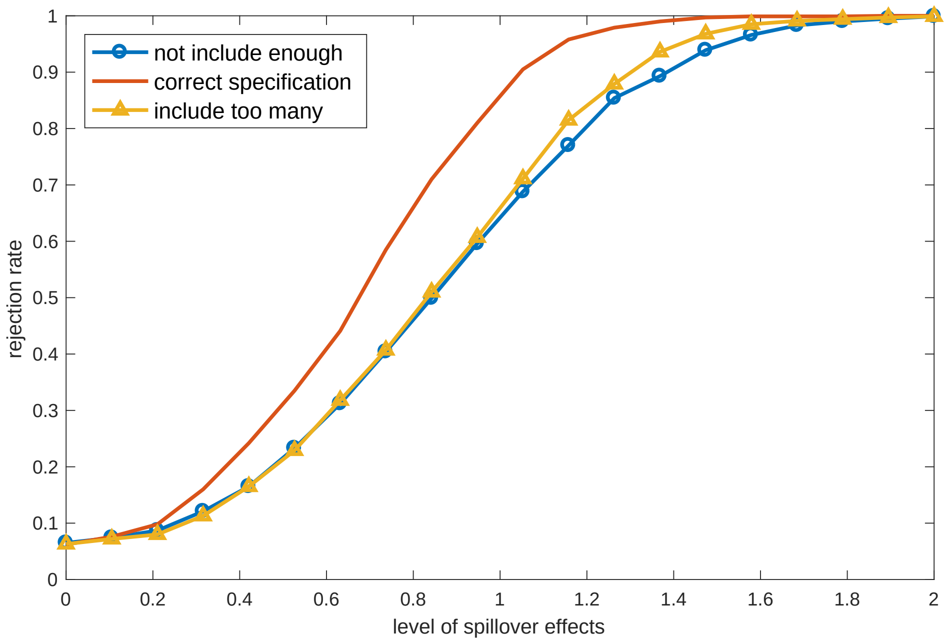

Figure 4: Empirical rejection rate of testing for existence of spillover effects.

The data is per capita cigarrette consumption in 38 of the 50 states running from 1970 to 2000. In 1989 California enacted prop 99 and so all periods from 1989 onwards will be considered our treatment periods. With 38 coefficients to estimate (37 other states and an intercept) and 19 pre-periods, we only get point identified weight matrices because of the regularization implicit in assuming simplex weights.

We replicate the program evaluation in ADH using the method introduced in previous sections, allowing for possible spillover effects. We use the spillover structure as in Example 3. That is, we allow for arbitrary spillover effects in those geographically close and culturally similar states as described in the first paragraph, but not the others. We also perform hypothesis testing on both treatment effect and spillover effects.

The results are shown in Figure 5 and Figure 6. Abadie et al. (2010)'s method is indexed by *synthetic control* and our method is *corrected synthetic control*. Figure 5 shows the "synthetic California" and Figure 6 elaborates on this by specifically looking at the estimated treatment effects. The error bars are built using the methods described in this paper, at a significance level of 90%. We do not use a 95% significance level because there are only 19 pre-treatment periods. With 19 observations, a 1-in-20 event has unknown magnitude. The dashed vertical line indicates the time of treatment. The shaded region represents time periods in which our test rejects the null of no spillover at the 90% level.

As Figure 5 shows, our estimated consumption in the "synthetic California" does not differ qualitatively from what a standard SCM would predict. Quantitatively, Figure 6 shows that our results are more

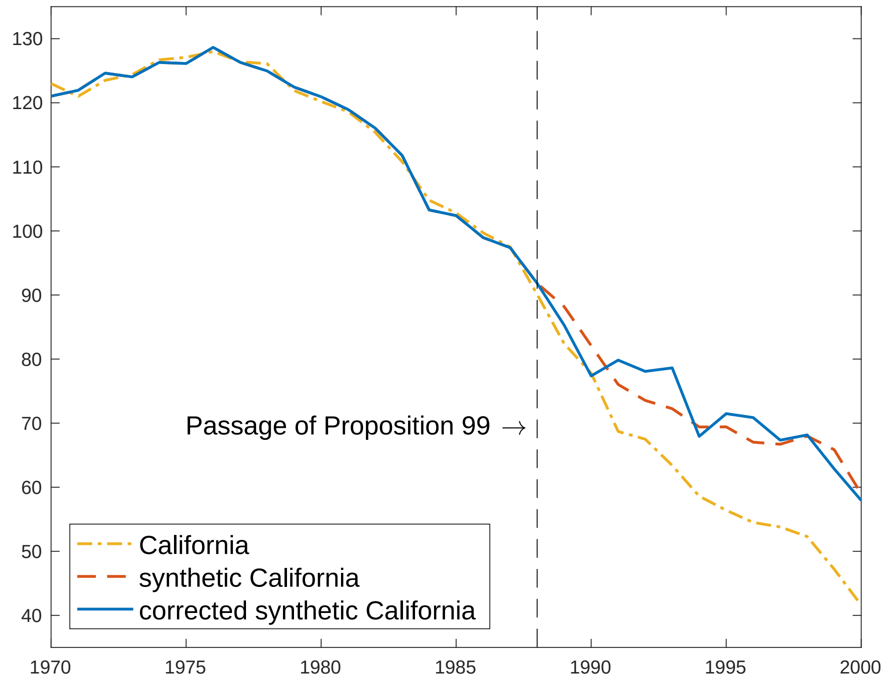

Figure 5: Trends in per-capita cigarette sales: California, synthetic California, and corrected synthetic California.

consistent with an addiction story, that tobacco consumption is addictive and should not fall immediately after the policy. From the tests of spillover effects (shaded area of Figure 6), we see that likely there were substantial spillover effects, which in some periods lead to statistically significant changes in the treatment effect estimates. This is particularly impressive given that the original paper considered some spillover effect types and removed those states from the data set to avoid these contamination problems.

7 Conclusion

Synthetic control method is a powerful tool in treatment effect estimation in the panel data settings, but it relies heavily on SUTVA. In this paper, we relax this assumption and propose an estimation and testing procedure that is robust to violation of SUTVA. Our method requires specification of the spillover structure, which can be fairly weak (Example 3). We consider both stationary and co-integrated cases, and show that our estimators are asymptotically unbiased. We develop a testing procedure based on Andrews (2003)'s end-of-sample instability tests, and show that it is asymptotically unbiased. Simulation results indicate our estimator beats current methods in estimation precision in the presence of spillover effects. In terms of testing, our method outperforms Chernozhukov et al. (2017)'s test in size and placebo tests in power. Also, in simulation our testing procedure is fairly robust to minor mis-specification of spillover structure. Finally, we illustrate our method by applying it to Abadie et al. (2010)'s California

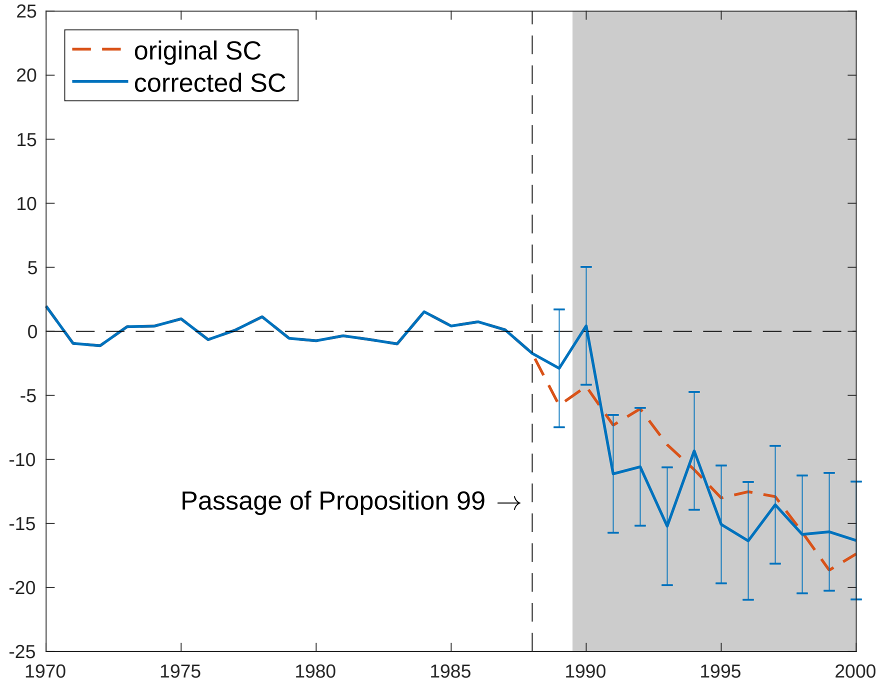

Figure 6: Per-capita cigarette sales gap between California and (corrected) synthetic California (with 90% confidence interval).

tobacco control program data.

Appendix

Proof of Lemma 1. First assume Condition ST holds. The proof follows Ferman and Pinto (2016), except that we do not assume that there is a set of weights that reconstruct the factor loadings and belong to the simplex.

We first show part (c). It suffices to show $|\hat{a}_i - a_i| = o_p(1)$ and $\|\hat{b}_i - b_i\| = o_p(1)$ for each i, i.e. a_i and b_i are well-defined. We show it for the $i = 1$ case and other cases follow the same strategy. Let $\bar{y}_j = T^{-1}\sum_{t=1}^{T} y_{j,t}$. Write down an (equivalent) optimization problem

$$\hat{v} = \arg\min_{v \in V} \left((y_{1,t} - \bar{y}_1) - \sum_{j=2}^{N} (y_{j,t} - \bar{y}_j) v_j \right)^2, \tag{40}$$

where $V = \{v = (v_2, \ldots, v_N) \in \mathbb{R}_+^{N-1} : \sum_{j=2}^{N} v_j = 1\}$. The objective is strictly convex (with probability approaching one), so the solution is unique. Note that it implies $\hat{b}_1$ is numerically equivalent to $(0, \hat{v}')'$, otherwise the minimization problem in forming $\hat{a}_1$ and $\hat{b}_1$ may have a lower objective evaluated at $(\bar{y}_1 - \sum_{j=2}^{N} \bar{y}_j \hat{v}_j, 0, \hat{v}')'$. Now we let $\hat{Q}(v)$ denote the objective function such that

$$\hat{Q}(v) = \frac{1}{T} \sum_{t=1}^{T} \left((y_{1,t} - \bar{y}_1) - \sum_{j=2}^{N} (y_{j,t} - \bar{y}_j) v_j \right)^2, \tag{41}$$

and its population analog be

$$Q(v) = \begin{pmatrix} -1 \\ v \end{pmatrix}' \Omega_y \begin{pmatrix} -1 \\ v \end{pmatrix}. \tag{42}$$

Let v_0 be a minimizer of $Q(v)$ in V. We verify the conditions for consistency (see Newey and McFadden (1994), Theorem 2.1): (i) Since Ω_y is positive definite, $Q(v)$ is strictly convex. Also, V is convex. Therefore, $Q(v)$ is uniquely minimized at v_0. (ii) V is compact, since it is a $(N-1)$-dimensional simplex. (iii) $Q(v)$ is continuous, since it has a quadratic form. (iv) To see uniform convergence, note

$$\begin{aligned}
\sup_{v \in V} |\hat{Q}(v) - Q(v)| &= \sup_{v \in V} \left| \begin{pmatrix} -1 \\ v \end{pmatrix}' \left(\frac{1}{T} \sum_{t=1}^{T} (Y_t - \bar{Y})(Y_t - \bar{Y})' - \Omega_y \right) \begin{pmatrix} -1 \\ v \end{pmatrix} \right| \\
&\leq \sup_{v \in V} \left\| \begin{pmatrix} -1 \\ v \end{pmatrix} \right\|^2 \left\| \frac{1}{T} \sum_{t=1}^{T} (Y_t - \bar{Y})(Y_t - \bar{Y})' - \Omega_y \right\|_F \\
&\leq N \cdot o_p(1) \\
&= o_p(1),
\end{aligned} \tag{43}$$

where $\|\cdot\|_F$ is the Frobenius norm. The second inequality is by ergodicity for the second moments.

Therefore, $\hat{v} \to_p v_0$. This implies $\|\hat{b}_1 - b_1\| = o_p(1)$. By ergodicity,

$$\hat{a}_1 = \bar{y}_1 - [\bar{y}_2 \ \bar{y}_3 \ \dots \ \bar{y}_N]\hat{v} \to_p E[y_{1,t}(0) - Y_t(0)'b_1] = a_1. \tag{44}$$

This shows part (c) and $E[u_{1,t}] = 0$ by definition of $u_{i,t}$. We also have that $\{u_t\}_{t\geq 1}$ is stationary since it is a linear combination of stationary and ergodic processes. This shows part (a) in Assumption 1.

Condition ST assumes part (b). Part (d) follows from (c) and the stationarity of $\{Y_{T+1}(0)\}_{T\geq 1}$. Thus, Assumption 1 holds under Condition ST.

Now we instead assume Condition CO holds.

We first show part (d). We will show $\|Y_{T+1}(0)'(\hat{b}_1 - b_1)\| = o_p(1)$ and other i's follows the same strategy. We follows Li (2017)'s strategy of treating the synthetic control weight estimator as a projection of the OLS estimator onto a closed convex set. Namely, for some positive definite matrix $D \in \mathbb{R}^N$, let $\mathbb{R}^N$ be a Hilbert space with the inner product $\langle \cdot, \cdot \rangle_D$ such that for $\theta_1, \theta_2 \in \mathbb{R}^N$,

$$\langle \theta_1, \theta_2 \rangle_D = \theta_1' D \theta_2. \tag{45}$$

The norm $\| \cdot \|_D$ is defined accordingly, i.e. $\|\theta\|_D = \sqrt{\theta' D \theta}$, for $\theta \in \mathbb{R}^N$. For a closed convex set $\Lambda \subset \mathbb{R}^N$, define a projection Π_D such that for each $\theta \in \mathbb{R}^N$, $\Pi_D \theta = \arg\min_{\theta' \in \Lambda} \|\theta - \theta'\|_D$. Zarantonello (1971) shows that for each $\theta, \theta' \in \mathbb{R}^N$,

$$\|\Pi_D \theta - \Pi_D \theta'\|_D \leq \|\theta - \theta'\|_D. \tag{46}$$

With some abuse of notation, let $x_t = Y_t - T^{-1}\sum_{s=1}^T Y_s$. Then, $\hat{b}_1$ is the synthetic control weight estimators of regressing $(y_{1,t} - T^{-1}\sum_{s=1}^T y_{1,s})$ on x_t, subject to $\{0\} \times \Delta_{N-1}$ with Δ_{N-1} being an $(N-1)$-dimensional simplex. Let $\tilde{b}_1$ be the OLS estimator of regressing $(y_{1,t} - T^{-1}\sum_{s=1}^T y_{1,s})$ on x_t. Let $\Sigma_T = T^{-1}\sum_{t=1}^T x_t x_t'$.

Appendix A.2 in Li (2017) establishes that $\hat{b}_1 = \Pi_{\Sigma_T}\tilde{b}_1$. Thus, we have

$$\begin{aligned}
\|\hat{b}_1 - b_1\| &= \|\Sigma_T^{-1/2}\Sigma_T^{1/2}(\hat{b}_1 - b_1)\| \\
&\leq \|\Sigma_T^{-1/2}\|_F \cdot \|\Sigma_T^{1/2}(\hat{b}_1 - b_1)\| \\
&= \|\Sigma_T^{-1/2}\|_F \cdot \|\hat{b}_1 - b_1\|_{\Sigma_T} \\
&= \|\Sigma_T^{-1/2}\|_F \cdot \|\Pi_{\Sigma_T}\tilde{b}_1 - \Pi_{\Sigma_T} b_1\|_{\Sigma_T} \\
&\leq \|\Sigma_T^{-1/2}\|_F \cdot \|\tilde{b}_1 - b_1\|_{\Sigma_T} \\
&= \|\Sigma_T^{-1/2}\|_F \cdot \|\Sigma_T^{1/2}\|_F \cdot \|\tilde{b}_1 - b_1\| \\
&= O_p(1) o_p(T^{-1/2}) \\
&= o_p(T^{-1/2}),
\end{aligned} \tag{47}$$

where $\|\cdot\|_F$ is the Frobenius norm of a matrix. The third equality is because $b_1 \in \{0\} \times \Delta_{N-1}$. The second inequality is by (46). To see the fifth equality, note

$$\Sigma_T = T\left(\frac{1}{T^2}\sum_{t=1}^{T} Y_t Y_t' - \left(\frac{1}{T^{3/2}}\sum_{t=1} Y_t\right)\left(\frac{1}{T^{3/2}}\sum_{t=1} Y_t\right)'\right), \tag{48}$$

so

$$\|\Sigma_T^{-1/2}\|_F \cdot \|\Sigma_T^{1/2}\|_F = \mathrm{tr}(\Sigma_T^{-1})\mathrm{tr}(\Sigma_T) = O_p(1) \cdot \frac{1}{T} \cdot T \cdot O_p(1) = O_p(1), \tag{49}$$

where the second equality is standard results for $\mathcal{I}_1$ process (see part (g) and (i) of Proposition 18.1 in Hamilton (1994) for example). Also, $\|\tilde{b}_1 - b_1\| = o_p(T^{-1/2})$ is by Proposition 19.2 in Hamilton (1994). This shows (47). Apply part (a) of Proposition 18.1 in Hamilton (1994), we have

$$\|Y_{T+1}(0)'(\hat{b}_1 - b)\| = \|(T^{-1/2}Y_{T+1}(0))'(T^{-1/2}(\hat{b}_1 - b))\| = O_p(1)o_p(1) = o_p(1). \tag{50}$$

Now we show part (c). Again, it suffices to show $|\hat{a}_i - a_i| = o_p(1)$ and $\|\hat{b}_i - b_i\| = o_p(1)$. We consider the $i = 1$ case and other cases follow the same strategy. We have showed $\|\hat{b}_i - b_i\| = o_p(1)$ in part (d) of the proof. Section A.6.1 in Ferman and Pinto (2016) establishes that

$$[\mu_1^1 \ \mu_2^1 \ \ldots \ \mu_N^1](b_1 - e_1) = 0, \tag{51}$$

where e_i is the unit vector with one at the i-th entry. Thus,

$$\begin{aligned}
\hat{a}_1 &= [\bar{y}_1 \ \bar{y}_2 \ \ldots \ \bar{y}_N](e_1 - \hat{b}_1) \\
&= [\bar{y}_1 \ \bar{y}_2 \ \ldots \ \bar{y}_N](e_1 - b_1) + [\bar{y}_1 \ \bar{y}_2 \ \ldots \ \bar{y}_N](b_1 - \hat{b}_1) \\
&= \left\{\frac{1}{T}\sum_{t=1}^{T}\left((\lambda_t^0)'[\mu_1^0 \ \ldots \ \mu_N^0] + [\epsilon_{1,t} \ \ldots \ \epsilon_{N,t}]\right)\right\}(e_1 - b_1) + \left(\frac{1}{\sqrt{T}}[\bar{y}_1 \ \bar{y}_2 \ \ldots \ \bar{y}_N]\right)\sqrt{T}(b_1 - \hat{b}_1) \\
&= E[\lambda_t^0]'[\mu_1^0 \ \ldots \ \mu_N^0](e_1 - b_1) + o_p(1) + O_p(1)o_p(1) \\
&\to_p E[\lambda_t^0]'[\mu_1^0 \ \ldots \ \mu_N^0](e_1 - b_1). \\
&= a_1
\end{aligned} \tag{52}$$

The third equality is by (51). The fourth equality is by stationarity of $\{(\lambda_t^0, \epsilon_t)\}_{t\geq 1}$ and results in part (d) of the proof. This shows part (c) of the proof.

Combining (51) and (52), we have part (a) in Assumption 1. Part (b) is assumed in Condition CO. □

Proof of Theorem 1. Using formula of $\hat{\gamma}$ in Equation (14), we have

$$\begin{aligned}\hat{\gamma} &= (A'\hat{M}A)^{-1}A'(I-\hat{B})'((I-\hat{B})Y_{T+1}(0) + (I-\hat{B})\alpha - \hat{a}) \\ &= (A'\hat{M}A)^{-1}A'(I-\hat{B})'(u_{T+1} + (B-\hat{B})Y_{T+1}(0) + (a-\hat{a}) + (I-\hat{B})A\gamma) \\ &= (A'\hat{M}A)^{-1}A'(I-\hat{B})'u_{T+1} + o_p(1) + o_p(1) + \gamma. \end{aligned} \tag{53}$$

The first equality is by $Y_{T+1} = Y_{T+1}(0) + \alpha$. The second equation is because $Y_{T+1}(0) = a + BY_{T+1}(0) + u_{T+1}$. The third equation is by (c) and (d) in Assumption 1. Therefore,

$$\begin{aligned}\hat{\alpha} - (\alpha + Gu_{T+1}) &= A(A'\hat{M}A)^{-1}A'(I-\hat{B})'u_{T+1} + A\gamma + o_p(1) - \alpha - Gu_{T+1} \\ &= (A(A'\hat{M}A)^{-1}A'(I-\hat{B}) - G)'u_{T+1} + o_p(1) \\ &= o_p(1)O_p(1) + o_p(1) \\ &= o_p(1). \end{aligned} \tag{54}$$

The third equality is by (c) in Assumption 1 and stationarity of $\{u_t\}_{t\geq 1}$. □

Proof of Proposition 1. The proof for the first half of the proposition is similar to the proof for Theorem 1, and thus is omitted. Too see the second half, note

$$Cov[G_W u_{T+1}] = A(Q'WQ)^{-1}Q'W\Omega WQ(Q'WQ)^{-1}A' \tag{55}$$

and

$$Cov[G^e u_{T+1}] = A(Q'\Omega Q)^{-1}A', \tag{56}$$

where $Q = (I-B)A$. It suffices to show $((Q'WQ)^{-1}Q'W\Omega WQ(Q'WQ)^{-1} - (Q'\Omega Q)^{-1})$ is positive semi-definite. Note that the first term is asymptotic variance of using W as the weighting matrix in GMM exercise and the second term is the one using the efficient weighting matrix (see Proposition 3.5 in Hayashi (2000)). Thus, $(Cov[G_W u_{T+1}] - Cov[G^e u_{T+1}])$ is positive semi-definite. □

Proof of Lemma 2. Since Assumption 3 implies Assumption 2, we only need to show Lemma 3. □

Proof of Theorem 2. We follow the proof of Theorem 2 in Andrews and Kim (2012). Let

$$\begin{aligned} L_{1,T}(\epsilon) &= \left\{ \|C_T(\hat{\beta}_1 - \beta_1)\| \leq \epsilon, \max_{t=1,\ldots,T} \|C_T(\hat{\beta}_1^{(t)} - \beta_1)\| \leq \epsilon \right\}, \\ L_{2,T}(c) &= \left\{ \max_{t\leq T+1} \|C_T^{-1}x_t\| \leq c \right\}. \end{aligned} \tag{57}$$

By Assumption 2(d), there exists a positive sequence $\{\epsilon_T\}_{T\geq 1}$ such that $\epsilon_T \to 0$ and $\Pr(L_{1,T}(\epsilon_T)) \to 1$. Let $c_T = 1/\sqrt{\epsilon_T}$. So we have $c_T \to \infty$ and $c_T\epsilon_T \to 0$. By Assumption 2(c), we must have $\Pr(L_{2,T}(c_T)) \to 1$. Let $L_T = L_{1,T}(\epsilon_T) \cap L_{2,T}(c_T)$, then we have $\Pr(L_T) \to 1$ and $\Pr(L_T^c) \to 0$.

Suppose L_T holds. Then, for $\beta = \hat{\beta}_1$ or $\beta = \hat{\beta}_1^{(t)}$ for some $t = 1, \dots, T$, we have

$$
\begin{aligned}
|P_t(\beta) - P_t(\beta_1)| &= \left|(\beta - \beta_1)' x_t x_t' (\beta - \beta_1) - 2x_t'(\beta - \beta_1)u_{1,t}\right| \\
&= \left|(\beta - \beta_1)' C_T' (C_T')^{-1} x_t x_t' C_T^{-1} C_T (\beta - \beta_1) - 2x_t' C_T^{-1} C_T (\beta - \beta_1) u_{1,t}\right| \\
&\leq \|C_T(\beta - \beta_1)\|^2 \|C_T^{-1} x_t\|^2 + 2\|C_T^{-1} x_t\| \|C_T(\beta - \beta_1)\| |u_{1,t}| \\
&\leq \epsilon_T^2 c_T^2 + 2\epsilon_T c_T |u_{1,t}|. \qquad (58)
\end{aligned}
$$

Define $g_t(\epsilon_T, c_T) = \epsilon_T^2 c_T^2 + 2\epsilon_T c_T |u_{1,t}|$. Note that $g_t(\epsilon_T, c_T)$ is identically distributed across t for a fixed T, by Assumption 2(a).

We first prove part (a). Let x be some continuous point of distribution function of $P_{T+1}(\beta_1)$. Then,

$$
\begin{aligned}
\Pr(P_{T+1}(\hat{\beta}_1) \leq x) &= \Pr(\{P_{T+1}(\hat{\beta}_1) \leq x\} \cap L_T) + \Pr(\{P_{T+1}(\hat{\beta}_1) \leq x\} \cap L_T^c) \\
&\leq \Pr(P_{T+1}(\hat{\beta}_1) \leq x + g_t(\epsilon_T, c_T)) + \Pr(L_T^c) \\
&\leq \Pr(P_{T+1}(\beta_1) \leq x) + o(1). \qquad (59)
\end{aligned}
$$

To see the last equality, pick $\epsilon > 0$. By continuity, $\exists \delta > 0$ such that for each $y \in (x - \delta, x + \delta)$, $|\Pr(P_{T+1}(\beta_1) \leq y) - \Pr(P_{T+1}(\beta_1) \leq x)| < \epsilon$. Therefore,

$$
\begin{aligned}
\Pr(P_{T+1}(\hat{\beta}_1) \leq x + g_t(\epsilon_T, c_T)) &= \Pr(\{P_{T+1}(\hat{\beta}_1) \leq x + g_t(\epsilon_T, c_T)\} \cap \{|g_t(\epsilon_T, c_T)| \geq \delta\}) \\
&\quad + \Pr(\{P_{T+1}(\hat{\beta}_1) \leq x + g_t(\epsilon_T, c_T)\} \cap \{|g_t(\epsilon_T, c_T)| < \delta\}) \\
&\leq \Pr(|g_t(\epsilon_T, c_T)| \geq \delta) + \Pr(P_{T+1}(\hat{\beta}_1) \leq y) \\
&< \Pr(P_{T+1}(\beta_1) \leq x) + o(1). \qquad (60)
\end{aligned}
$$

Similarly,

$$
\Pr(P_{T+1}(\hat{\beta}_1) \leq x) \geq \Pr(P_{T+1}(\beta_1) \leq x) + o(1). \qquad (61)
$$

This shows part (a).

Too see part (b), let $k : \mathbb{R} \to \mathbb{R}$ be a monotonically decreasing and everywhere differentiable function that has bounded derivative and satisfies $k(x) = 1$ for $x \leq 0$, $k(x) \in [0, 1]$ for $x \in (0, 1)$, and $k(x) = 0$ for $x \geq 1$. For example, let $k(x) = \cos(\pi x)/2 + 1/2$ for $x \in (0, 1)$. Given some $\{\beta^{(t)}\}_{t=1}^T$, a smoothed df is defined by

$$
\hat{F}_T(x, \{\beta^t\}, h_T) = \frac{1}{T} \sum_{t=1}^{T} k\left(\frac{P_t(\beta^{(t)}) - x}{h_T}\right), \qquad (62)
$$

for some sequence of positive constants $\{h_T\}$ such that $h_T \to 0$ and $c_T \epsilon_T / h_T \to 0$. For example, we let $h_T = \epsilon_T^{1/4}$ when $c_T = 1/\sqrt{\epsilon_T}$.

We write

$$|\hat{F}_{P,T}(x) - F_P(x)| \leq \sum_{i=1}^{4} D_{i,T}, \tag{63}$$

for

$$\begin{aligned} D_{1,T} &= |\hat{F}_{P,T}(x) - \hat{F}_T(x, \{\hat{\beta}_j\}, h_T)|, \\ D_{2,T} &= |\hat{F}_T(x, \{\hat{\beta}_j\}, h_T) - \hat{F}_T(x, \{\beta_1\}, h_T)|, \\ D_{3,T} &= |\hat{F}_T(x, \{\beta_1\}, h_T) - \hat{F}_T(x, \{\beta_1\})|, \text{ and} \\ D_{4,T} &= |\hat{F}_T(x, \{\beta_1\}, h_T) - F_P(x\})|. \end{aligned} \tag{64}$$

We want to show that all four terms vanish. First note that

$$D_{1,T} \leq \frac{1}{T} \sum_{t=1}^{T} \mathbb{1} \left\{ \frac{P_t(\hat{\beta}_1^{(t)}) - x}{h_T} \in (0, 1) \right\}. \tag{65}$$

Thus, for any $\delta > 0$,

$$\begin{aligned} \Pr(D_{1,T} > \delta) &\leq \Pr(\{D_{1,T} > \delta\} \cap L_T) + \Pr(L_T^c) \\ &\leq \Pr\left(\frac{1}{T} \sum_{t=1}^{T} \mathbb{1} \left\{ P_t(\hat{\beta}_1^{(t)}) - x \in (-g_t(\epsilon_T, c_T), h_T + g_t(\epsilon_T, c_T) \right\} > \delta \right) + o(1) \\ &\leq \frac{E\mathbb{1} \left\{ P_t(\hat{\beta}_1^{(t)}) - x \in (-g_t(\epsilon_T, c_T), h_T + g_t(\epsilon_T, c_T) \right\}}{\delta} + o(1), \end{aligned} \tag{66}$$

where the last inequality is by Markov's inequality. Recall $\Pr(P_1(\beta_1) \neq x) = 1$ and $g_t(\epsilon_T, c_T) \to 0$ a.s., so $\mathbb{1}\{P_t(\beta_1) - x \in \{-g_t(\epsilon_T, c_T), h_T + g_t(\epsilon_T, c_T)\} \to 0$ a.s.. By the dominated convergence theorem, (66) implies $\Pr(D_{1,T} > \delta) \leq o(1)$ and thus $D_{1,T} = o_p(1)$.

For $D_{2,T}$, we have

$$\begin{aligned} D_{2,T} &= \left| \frac{1}{T} \sum_{t=1}^{T} k' \left(\frac{\tilde{P}_t - x}{h_T} \right) \frac{P_t(\hat{\beta}_1^{(t)}) - P_t(\beta_1)}{h_T} \right| \\ &\leq \frac{\bar{k}}{T} \sum_{t=1}^{T} \frac{g_t(\epsilon_T, c_T)}{h_T}. \end{aligned} \tag{67}$$

The equality is by the mean value theorem and we have $\tilde{P}_t$ lies between $P_t(\hat{\beta}_1^{(t)})$ and $P_t(\beta_1)$. In the inequality, $\bar{k}$ is a bound for the derivative of k. Also, note

$$E\left[\frac{g_t(\epsilon_T, c_T)}{h_T} \right] = \frac{\epsilon_T^2 c_T^2}{h_T} + 2 \frac{\epsilon_T c_T}{h_T} E|u_{1,t}| = o(1). \tag{68}$$

Therefore,

$$\begin{aligned}
\Pr(D_{2,T} > \delta) &\leq \Pr(\{D_{2,T} > \delta\} \cap L_T) + \Pr(L_T^c) \\
&\leq \Pr\left(\frac{\bar{k}}{T}\sum_{t=1}^{T}\frac{g_t(\epsilon_T, c_T)}{h_T} > \delta\right) + o(1) \\
&\leq \bar{k}\frac{Eg_t(\epsilon_T, c_T)}{\delta h_T} \\
&\to 0. \qquad (69)
\end{aligned}$$

The third inequality is by Markov's inequality. This shows $D_{2,T} = o_p(1)$.

$D_{3,T}$ is similar to the $D_{1,T}$ case. Finally, by stationary and ergodicity of $u_{1,t}$, we have $D_{4,T} = o_p(1)$. This shows part (b).

Now we show part (c). Pick any small ϵ such that $\hat{F}_{P,T}(x) \to_p F_P(x)$ for $x \in (q_{P,1-\alpha} - \epsilon, q_{P,1-\alpha} + \epsilon)$. Note

$$\begin{aligned}
\Pr(\hat{q}_{P,1-\alpha} > q_{P,1-\alpha} + \epsilon) &\leq \Pr(\hat{F}_{P,T}(q_{P,1-\alpha} + \epsilon) < 1 - \alpha) \\
&= \Pr(\hat{F}_{P,T}(q_{P,1-\alpha} + \epsilon) - F_P(q_{P,1-\alpha} + \epsilon) < (1-\alpha) - F_P(q_{P,1-\alpha} + \epsilon)) \\
&\to 0. \qquad (70)
\end{aligned}$$

The inequality is by definition of $\hat{q}_{P,1-\alpha}$. The convergence is because of part (e) of Assumption 2 and part (b) of Theorem 2. Similarly,

$$\begin{aligned}
\Pr(\hat{q}_{P,1-\alpha} < q_{P,1-\alpha} - \epsilon) &\leq \Pr(\hat{F}_{P,T}(q_{P,1-\alpha} - \epsilon) \geq 1 - \alpha) \\
&= \Pr(\hat{F}_{P,T}(q_{P,1-\alpha} - \epsilon) - F_P(q_{P,1-\alpha} - \epsilon) \geq (1-\alpha) - F_P(q_{P,1-\alpha} - \epsilon)) \\
&\to 0. \qquad (71)
\end{aligned}$$

Again, the inequality is by definition of $\hat{q}_{P,1-\alpha}$, and the convergence is because of part (e) of Assumption 2 and part (b) of Theorem 2.

Finally, we show part (d). Under null, P_∞ and $P_1(\beta_1)$ have the same distribution, so $q_{P,1-\alpha}$ is $(1-\alpha)$-quantile of P_∞. Therefore,

$$\begin{aligned}
\Pr(P > \hat{q}_{P,1-\alpha}) &= 1 - \Pr(P \leq \hat{q}_{P,1-\alpha}) \\
&= 1 - \Pr(P + (q_{P,1-\alpha} - \hat{q}_{P,1-\alpha}) \leq q_{P,1-\alpha}) \\
&\to \alpha, \qquad (72)
\end{aligned}$$

where the convergence is by combining part (a) and (c). This concludes our proof. □

Proof of Lemma 3. (i) Assume Condition ST holds.

By Lemma 1, part (a) of Assumption 3 holds.

Part (b) is because u_t is a linear combination of $\delta_t, \lambda_t, \epsilon_t$.

For part (c), pick some τ such that $1/(2+\delta) < \tau < 1/2$, where δ is defined in Condition ST. Let

$$D_T = \begin{bmatrix} 1 & 0 \\ 0 & T^{\tau} I_N \end{bmatrix}. \tag{73}$$

Then, we have

$$\max_{t \le T+1} \|D_T^{-1} x_t\| = \max_{t \le T+1} \left\| \begin{pmatrix} 1 \\ T^{-\tau} Y_t \end{pmatrix} \right\| = \sqrt{1 + \left(\max_{t \le T+1} \|T^{-\tau} Y_t\| \right)^2}. \tag{74}$$

Also, for any $\epsilon > 0$, note

$$\begin{aligned} \Pr\left(\max_{t \le T+1} \|T^{-\tau} Y_t\| > \epsilon \right) &= \Pr\left(\bigcup_{t \le T+1} \|Y_t\| > T^{\tau} \epsilon \right) \\ &\le \left(\sum_{t=1}^{T} \Pr(\|Y_t\| > T^{\tau}\epsilon) \right) + \Pr(\|Y_{T+1}(0) + \alpha\| > T^{\tau}\epsilon) \\ &= \frac{T E[\|Y_t\|^{2+\delta}]}{T^{\tau(2+\delta)} \epsilon^{2+\delta}} + o(1) \\ &= o(1). \end{aligned} \tag{75}$$

The second equality is due to Markov inequality and stationarity of $\{Y_{T+1}(0)\}_{t+1}$. The last equality is because $\tau > 1/(2+\delta)$. Combining (74) and (75), we obtain part (c).

For part (d), we use D_T defined in (73). Following the same reasoning as in (47), for each $i = 1, \ldots, N$, we have

$$\begin{aligned} \|\hat{b}_i - b_i\| &\le \|\Sigma_T^{-1/2}\|_F \cdot \|\Sigma_T^{1/2}\|_F \cdot \|\tilde{b}_i - b_i\| \\ &= O_p(1) O_p(T^{-1/2}) \\ &= O_p(T^{-1/2}). \end{aligned} \tag{76}$$

The first equality is because $\{Y_t(0)\}_{t \ge 1}$ is ergodic for the second moment, and $\tilde{b}_i$ is the OLS estimator for b_i. Thus,

$$
\begin{aligned}
\|D_T(\hat\beta_i-\beta_i)\| &= \left\| \begin{pmatrix} 1 & 0 \\ 0 & T^{\tau-1/2}I_N \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & T^{1/2}I_N \end{pmatrix} (\hat\beta_i-\beta_i) \right\| \\
&\leq \left\| \begin{pmatrix} 1 & 0 \\ 0 & T^{\tau-1/2}I_N \end{pmatrix} \right\|_F \left\| \begin{pmatrix} \hat a_i - a_i \\ \sqrt{T}(\hat b_i - b_i) \end{pmatrix} \right\| \\
&= \sqrt{1+NT^{2\tau-1}}\|O_p(1)\| \\
&= o_p(1). \qquad (77)
\end{aligned}
$$

The second equality is due to (76). The last equality is because $\tau < 1/2$. Therefore,

$$
\|(\hat\theta-\theta_0)D_T\|_F = \sqrt{\sum_{i=1}^N \|D_T(\hat\beta_i-\beta_i)\|^2} = o_p(1). \qquad (78)
$$

Also, since $\hat\theta^{(t)} = \hat\theta$ for each t,

$$
\max_{t=1,\dots,T} \|(\hat\theta^{(t)}-\theta_0)D_T\|_F = \|(\hat\theta-\theta_0)D_T\|_F = o_p(1). \qquad (79)
$$

This shows part (d).

Part (e) is assumed.

Part (f) is trivial is $W_T = I$. Assume now $W_T = (C\hat G(T^{-1}\sum_{t=1}^T \hat u_t \hat u_t')\hat G' C')^{-1}$. Then,

$$
\begin{aligned}
\frac{1}{T}\sum_{t=1}^T \hat u_t \hat u_t' &= (I-\hat B)\left(\frac{1}{T}\sum_{t=1}^T Y_t Y_t'\right)(I-\hat B)' - (I-\hat B)\left(\frac{1}{T}\sum_{t=1}^T Y_t\right)\hat a' - \hat a\left(\frac{1}{T}\sum_{t=1}^T Y_t'\right)(I-\hat B)' + \hat a \hat a' \\
&\to E[u_t u_t'], \qquad (80)
\end{aligned}
$$

by ergodicity and Assumption 1(c). Therefore, $\hat W_T \to_p W = (CGE[u_t u_t']G'C')^{-1}$.

This concludes part (i) of Lemma 3.

(ii) Assume Condition CO holds.

By Lemma 1, Assumption 1 holds. This shows Part (a).

By (51), u_t is a linear combination of λ_t^o and ϵ_t, so $\{u_t\}_{t\geq 1}$ is ergodic and has finite first moment. This shows Part (b).

Now we show Part (c). Let

$$
D_T = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{T}\cdot I_N \end{bmatrix}. \qquad (81)
$$

Then, we have

$$
\begin{aligned}
\max_{t\leq T+1}\|D_T^{-1}x_t\| &= \sqrt{1+\left(\max_{t\leq T+1}\|T^{-1/2}Y_t\|\right)^2} \\
&\leq \sqrt{1+\sum_{i=1}^{N}\left(\max_{t\leq T+1}|T^{-1/2}y_{i,t}|\right)^2} \\
&\leq \sqrt{1+\sum_{i=1}^{N}\left(T^{-1}|\alpha_i|+\max_{t\leq T+1}|T^{-1/2}y_{i,t}(0)|\right)^2} \\
&= \sqrt{1+\sum_{i=1}^{N}(o(1)+O_p(1))^2} \\
&= O_p(1)
\end{aligned}
\tag{82}
$$

The second equality is because

$$
\max_{t\leq T+1}|T^{-1/2}y_{i,t}(0)| \Rightarrow \max_{r\in[0,1]}\nu_i(r) \tag{83}
$$

by the continuous mapping theorem.

To show Part (d), we combine (47) and (52), and have

$$
\|D_T(\hat{\beta}_i-\beta_i)\| = \left\|\begin{pmatrix}\hat{a}_i-a_i \\ \sqrt{T}(\hat{b}_i-b_i)\end{pmatrix}\right\| = o_p(1). \tag{84}
$$

Therefore,

$$
\|(\hat{\theta}-\theta_0)D_T\|_F = \sqrt{\sum_{i=1}^{N}\|D_T(\hat{\beta}_i-\beta_i)\|^2} = o_p(1). \tag{85}
$$

The second half of Part (d) is also satisfied since $\hat{\theta}^{(t)}=\hat{\theta}$ for each t.

Part (e) is assumed and Part (f) is trivial for $W_T=I$. □

Proof of Theorem 3. We use similar strategy as we do in the proof of Theorem 2. Let

$$
\begin{aligned}
L_{1,T}(\epsilon) &= \left\{\|(\hat{\theta}-\theta_0)\|_F\leq\epsilon,\ \max_{t=1,\dots,T}\|(\hat{\theta}^{(t)}-\theta_0)\|_F\leq\epsilon\right\}, \\
L_{2,T}(c) &= \left\{\max_{t\leq T+1}\|D_T^{-1}x_t\|\leq c\right\}, \\
L_{3,T}(\eta) &= \left\{\|\hat{G}'C'W_TC\hat{G}-G'C'WCG\|_F<\eta\right\}.
\end{aligned}
\tag{86}
$$

By Assumption 3(d), there exists a positive sequence $\{\epsilon_T\}_{T\geq1}$ such that $\epsilon_T\to0$ and $\Pr(L_{1,T}(\epsilon_T))\to1$. Let $c_T=1/\sqrt{\epsilon_T}$. So we have $c_T\to\infty$ and $c_T\epsilon_T\to0$. By Assumption 2(c), we must have $\Pr(L_{2,T}(c_T))\to1$. By Assumption 1(c) and Assumption 2(f), there exists a positive sequence $\{\eta_T\}_{T\geq1}$ such that $\eta_T\to0$ and $\Pr(L_{3,T}(\eta_T))\to1$. Let $L_T=L_{1,T}(\epsilon_T)\cap L_{2,T}(c_T)\cap L_{3,T}(\eta_T)$, then we have $\Pr(L_T)\to1$ and $\Pr(L_T^c)\to0$.

Suppose L_T holds. Then, for some $\theta = \hat{\theta}$ or $\theta = \hat{\theta}^{(t)}$ and for some $t = 1, \ldots, T$, we have

$$|\hat{P}_t(\theta) - P_t(\theta_0)| \leq |\hat{P}_t(\theta) - P_t(\theta)| + |P_t(\theta) - P_t(\theta_0)|. \tag{87}$$

Note that

$$\begin{aligned}
|\hat{P}_t(\theta) - P_t(\theta)| &= \left|(Y_t - \theta x_t)'(\hat{G}'C'W_T C\hat{G}) - G'C'WCG)(Y_t - \theta x_t)\right| \\
&\leq \|Y_t - \theta x_t\|^2 \|(\hat{G}'C'W_T C\hat{G} - G'C'WCG\|_F \\
&\leq \|u_t + (\theta_0 - \theta)x_t\|^2 \cdot \eta_T \\
&\leq (\|u_t\| + \|(\theta_0 - \theta)D_T D_T^{-1} x_t\|)^2 \eta_T \\
&\leq (\|u_t\| + \|(\theta_0 - \theta)D_T\|_F \|D_T^{-1} x_t\|)^2 \eta_T \\
&\leq (\|u_t\| + \epsilon_T c_T)^2 \eta_T
\end{aligned} \tag{88}$$

and

$$\begin{aligned}
|P_t(\theta) - P_t(\theta_0)| &= |(Y_t - \theta x_t)'G'C'WCG(Y_t - \theta x_t) - (Y_t - \theta_0 x_t)'G'C'WCG(Y_t - \theta_0 x_t)| \\
&\leq |(Y_t - \theta x_t)'G'C'WCG(Y_t - \theta x_t) - (Y_t - \theta x_t)'G'C'WCG(Y_t - \theta_0 x_t)| \\
&\quad + |(Y_t - \theta x_t)'G'C'WCG(Y_t - \theta_0 x_t) - (Y_t - \theta_0 x_t)'G'C'WCG(Y_t - \theta_0 x_t)| \\
&= |(u_t + (\theta_0 - \theta)x_t)'G'C'WCG(\theta_0 - \theta)x_t| + |((\theta_0 - \theta)x_t)'G'C'WCGu_t| \\
&\leq \|u_t + (\theta_0 - \theta)D_T D_T^{-1} x_t\| \|G'C'WCG\|_F \|(\theta_0 - \theta)D_T D_T^{-1} x_t\| \\
&\quad + \|(\theta_0 - \theta)D_T D_T^{-1} x_t\| \|G'C'WCG\|_F \|u_t\| \\
&\leq (\|u_t\| + \epsilon_T c_T)\|G'C'WCG\|_F \epsilon_T c_t + \epsilon_T c_T \|G'C'WCG\|_F \|u_t\| \\
&= (2\|u_t\| + \epsilon_T c_T)\|G'C'WCG\|_F \epsilon_T c_t.
\end{aligned} \tag{89}$$

Combining (87), (88), and (89), we have

$$|\hat{P}_t(\theta) - P_t(\theta_0)| \leq g(\epsilon_T, c_T, \eta_T), \tag{90}$$

where

$$g_t(\epsilon_T, c_T, \eta_T) = (\|u_t\| + \epsilon_T c_T)^2 \eta_T + (2\|u_t\| + \epsilon_T c_T)\|G'C'WCG\|_F \epsilon_T c_t. \tag{91}$$

By Assumption 1(a), $g_t(\epsilon_T, c_T, \eta_T)$ is identically distributed across t for a fixed T.

To show part (a), note that under null,

$$
\begin{aligned}
P &= (C\hat{\alpha} - d)' W_T (C\hat{\alpha} - d) \\
&= (C(\alpha + G u_{T+1} + o_p(1)) - d)'(W + o_p(1))(C(\alpha + G u_{T+1} + o_p(1)) - d) \\
&= (CG u_{T+1} + o_p(1))'(W + o_p(1))(CG u_{T+1} + o_p(1)) \\
&= u_{T+1}' G' C' W C G u_{T+1} + o_p(1). \qquad (92)
\end{aligned}
$$

The second equality is by Theorem 1. Since $P_\infty = u_1' G' C' W C G u_1$, we have $P \to_d P_\infty$ by stationary of $\{u_t\}_{t \geq 1}$.

Part (b)-(d) can be shown using the same strategy as in the proof of Theorem 2, with $g_t(\epsilon_T, c_T, \eta_T)$ in place of $g_t(\epsilon_T, c_T)$, and θ in place of β, so is omitted here. □

References

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

A. Abadie, A. Diamond, and J. Hainmueller. Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association*, 105(490):493–505, June 2010.

D. W. K. Andrews. End-of-Sample Instability Tests. *Econometrica*, 71(6):1661–1694, 2003.

D. W. K. Andrews and J.-Y. Kim. Tests for Cointegration Breakdown Over a Short Time Period. *Journal of Business & Economic Statistics*, 24(4):379–394, Jan. 2012.

V. Chernozhukov, K. Wuthrich, and Y. Zhu. An Exact and Robust Conformal Inference Method for Counterfactual and Synthetic Controls. *arXiv.org*, Dec. 2017.

T. G. Conley and C. R. Taber. Inference with "Difference in Differences" with a Small Number of Policy Changes. *The Review of Economics and Statistics*, 93(1):113–125, Jan. 2011.

N. Doudchenko and G. W. Imbens. Balancing, Regression, Difference-In-Differences and Synthetic Control Methods: A Synthesis. Oct. 2016.

E. Duflo and E. Saez. The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment. *The Quarterly Journal of Economics*, 118(3):815–842, 2003.

Z. Fang and A. Santos. Inference on Directionally Differentiable Functions, Apr. 2014. URL `https://arxiv.org/abs/1404.3763`.

B. Ferman and C. Pinto. Placebo Tests for Synthetic Controls. Apr. 2017.

B. Ferman and C. C. d. X. Pinto. Revisiting the synthetic control estimator. *Textos para discussão*, June 2016.

J. Hahn and R. Shi. Synthetic Control and Inference. 2016.

J. Hamilton. *Time Series Analysis*. Princeton University Press, Jan. 1994.

F. Hayashi. *Econometrics*. Princeton University Press, 2000.

K. Li. Statistical Inference for Average Treatment Effects Estimated by Synthetic Control Methods. Sept. 2017.

E. Manresa. Estimating the Structure of Social Interactions Using Panel Data. 2016.

C. F. Manski. Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies*, 60(3):531–542, July 1993.

W. K. Newey and D. McFadden. Chapter 36 Large sample estimation and hypothesis testing. In *Handbook of Econometrics Volumn 4*, pages 2111–2245. Elsevier, 1994.

E. H. Zarantonello. Projections on Convex Sets in Hilbert Space and Spectral Theory. In *Contributions to Nonlinear Functional Analysis*, pages 237–424. Elsevier, 1971.