Estimation and Inference for Synthetic Control Methods with Spillover Effects*

Jianfei Cao† Connor Dowd‡

March 8, 2023

Abstract

Current estimation and inference procedures for synthetic control methods do not allow for the existence of spillover effects, which are plausible in many applications. In this paper, we consider estimation and inference for synthetic control methods, allowing for spillover effects. We propose estimators for both direct treatment effects and spillover effects and show that they are asymptotically unbiased. In addition, we propose an inferential procedure and show that it is asymptotically unbiased. Our estimation and inference procedure applies to cases with multiple treated units and/or multiple post-treatment periods, and to ones where the underlying factor model is either stationary or cointegrated. In simulations, we confirm that the presence of spillovers renders current methods biased with distorted sizes, whereas our method yields properly-sized tests and retains reasonable power. We apply our method to a classic empirical example that investigates the effect of California's tobacco control program as in Abadie et al. (2010) and find evidence of spillovers.

1 Introduction

The synthetic control method is 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

*We thank Max Farrell and Christian Hansen for their invaluable guidance. We are grateful to Tetsuya Kaji, Ruey Tsay, Alexander Torgovitsky, Yinchu Zhu, and other seminar participants at the Booth Econometrics Seminar, the 2018 Midwest Econometrics Group Conference, the UChicago Econometrics Advising Group, the 2019 China Meeting of the Econometric Society, and the 2019 North America Summer Meeting of the Econometric Society. We thank Ziyao Wang for his excellent research assistantship.

†Department of Economics, Northeastern University, j.cao@northeastern.edu.

‡Joint Committee on Taxation, connor.dowd@jct.gov.

methods do not allow for the existence of spillover effects, which are plausible in many applications. Spillover effects are a particular concern in the realm of synthetic controls because observations which are highly predictive, and thus recieve substantial synthetic control weight, may be particularly prone to experiencing spillovers from treated observations. Those spillovers can bias treatment effect estimates and inference routines. This paper alleviates these concerns by showing that given some knowledge about the spillover effects, it is possible to provide asymptotically unbiased estimators and inference in the presence of spillovers. Our results extend to scenarios with multiple treated units and periods and cases with stationary or cointegrated factor models.

The synthetic control method (SCM) has gained popularity in empirical studies since its introduction in Abadie and Gardeazabal (2003). The SCM can estimate treatment effects when we observe panel data with only a few treated units and post-treatment periods. This setting is common in program evaluation, where we often consider state-level policies and have state-level aggregate data. The SCM models the relationship between the treated and untreated units using pre-treatment data. Then it uses the post-treatment data of the untreated units to predict the counterfactual values of the treated unit, i.e., to form the "synthetic control". The treatment effect estimate is given by the difference between the outcome and the predicted counterfactual outcome. The SCM exploits the pre-treatment data to form better counterfactual values, and so in comparative case studies, it is often favored over other program evaluation methods such as difference-in-differences. See Abadie and Cattaneo (2018) for a review and comparison of econometric methods used in program evaluation.

However, the SCM and its variants assume that untreated units are not affected by the treatment. That is, they rely on the Stable Unit Treatment Value Assumption (SUTVA). This dependence is natural and often implicit since the SCM uses post-treatment control units to predict the counterfactual values of the treated units. SCM is not alone in this, this assumption is shared with many other similar procedures like difference-in-differences. The structure of SCM leaves it more vulnerable to problems arising from this, often flawed, assumption. For example, SUTVA implies (among other things) that when California imposes a tax increase on cigarettes, nobody shifts their cigarette purchases to adjacent states like Nevada. We illustrate in Section 5 that SUTVA is violated in this well-known synthetic controls application. In other similar applications with aggregated geographic data and treatment at the level of some geographic unit, similar concerns about neighboring geographies may be substantial.

In the presence of a spillover effect, SCM can be severely biased. Intuitively, the reason is that post-treatment controls are contaminated by the spillover effect, resulting in a biased estimator of the counter-factual value of the treated unit, which in turn implies a biased treatment effect estimate. Contamination inducing bias is a

standard problem in program evaluation, even observed within difference-in-differences and RCTs. For a fixed, small contamination, the potential scale of this bias is worse for the SCM. If by chance or design the spillover is concentrated in control units that the synthetic control method puts significant weight on, the bias will be substantially worse than in difference-in-differences. Because synthetic controls deliberately weights some observations very heavily, the worst case version of this problem can be much larger than in difference-in-differences or RCTs. Moreover, it is possible the spillovers propagate along the same channels as the underlying factor model, which would mean that the SCM may actively select for units that will induce bias to give weight to. In this situation, even generous assumptions, like the spillover having a fixed number of affected units, can create a large asymptotic bias. We will illustrate the possible effects of this bias at length in our simulation section.

The problem of spillover effects can be partly solved by eliminating contaminated units in estimation or applying methods developed for multiple treated units (see Cavallo et al., 2013; Firpo et al., 2018; Kreif et al., 2016; Robbins et al., 2017; Xu, 2017). However, those methods may be concerning for two reasons. First, the contaminated units are often the most important control units that can be useful in forming the synthetic control. Simply eliminating them in estimation can potentially cause efficiency loss. Second, there are cases where most or even all control units are affected by the spillover, which cannot be solved by throwing away affected control units. This is also true for methods for multiple treated units since they use only the units that are not affected by the treatment in order to form the synthetic control. Third, often the measured objective is only a relative of the planner's true objective function. If, for example, California wanted to reduce the prevalance of cigarette smoke in California, cigarette purchases that shift across the border to Nevada are not merely a statistical problem, but a fundamental concern about the efficacy of the policy – which the planner would want to know about. Dropping the observations undermines that goal, while our procedure enables planners to learn about these possible side effects.

The goal of this paper is to relax the SUTVA condition and to perform estimation and testing. Particularly, we look at the cases where there are spillover effects, which are defined by a Rubin model as the difference between the actual outcomes and the counterfactual ones. To facilitate estimation, we assume some knowledge about the spillover effects is known. Specifically, the treatment effect and the spillover effects are linear in some unknown parameters.[1] We give examples where this assumption is plausible. Thanks to the known spillover structure, we can propose an asymptotically unbiased estimator for the treatment and spillover effects. We also characterize the

[1]When multiple post-treatment periods are available and the researcher is willing to restrict the variability of spillover structures, the structures are estimable in principle. See Manresa (2013) for example. We do not explore this direction in the paper and will leave it for future research.

asymptotic distribution of the estimator. Compared with existing methods, our method uses information from all known units in estimation. Our method can also often deal with cases where all units are contaminated by spillovers, at the cost of assuming more structure on the spillovers.

We follow the setup in Ferman and Pinto (2021), where we focus on cases with an impect pre-treatment fit and make use of the demeaned version of SCM. Ferman and Pinto (2021) show that the demeaned SCM ensures asymptotically unbiased estimation of treatment effects, which cannot be consistently estimated by the standard SCM estimator. We also rely on the imperfect pre-treatment fit to identify the null distribution of the proposed statistic. In terms of the asymptotic framework, we consider cases with many pre-treatment periods and a fixed number of control units. As suggested by Ferman and Pinto (2021), even in cases where large-T asymptotics is not justifed, our results can be interpreted as the SCM weights not converging to weights that reconstruct the factor loadings of the treated unit even when the number of pre-treatment periods is large. Besides, Monte-Carlo simulation shows that our methods produce reasonable estimation and testing results as long as there is a moderate number of pre-treatment periods.

Additionally, we propose an inferential procedure based on Andrews (2003)'s end-of-sample instability test, or P-test. We generalize the P-test to the synthetic control method with and without spillover effects. The theoretical result is of interest in its own right. Similar to the P-test, our testing procedures use the idea of approximating the null distribution of the statistic using pre-treatment data. We show the validity of the proposed method and compare it with the standard placebo test through a simulation study. Tangent to the main idea of this paper, our method alleviates the problem of selection into treatment, which is a major threat to the placebo test.

We give high-level conditions under which our methods are valid. Specifically, our conditions adapt to factor models with either stationary or cointegrated common factors, which are often used to justify the usage of synthetic control methods. Furthermore, we consider extensions where treatment applies to multiple units or periods, and where there are extra covariates.

We examine an empirical example from Abadie et al. (2010). In 1989 California implemented a cigarette tax. Abadie et al. (2010) gather data from 50 states starting in 1970 for comparison. They dismiss 12 states for potentially being affected by spillovers or later treatment, leaving 38 states to be used in estimation. Despite this precaution, we find evidence of spillover effects (i.e., SUTVA violations) in most years after the tax increase within that subset of 38 states. We also find evidence of spillover effects in the 12 excluded states. Moreover, our estimates are smaller in magnitude than those in Abadie et al. (2010) in the first four post-treatment years.

This paper mainly contributes to three developing literatures. First, it complements the fast-developing literature on statistical inference in SCM by providing formal statistical results without assuming SUTVA. Due to SCM's 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-differences, and use control units to form the null distribution of the statistic. In this particular setting with only a few treated units, difference-in-difference estimator can be treated as a special case of the SCM with equal weights. In Ferman and Pinto (2017) and Hahn and Shi (2017), similar ideas are used to conduct placebo tests which permute across units. Among all, Chernozhukov et al. (2021) is the most related to our work, since they also use outcomes across *periods* rather than *units* like the above citations. Li (2020) proposes a testing procedure that is based on the idea of projection onto convex sets and results in Fang and Santos (2019). Despite the popularity, very few works consider the existence of spillovers. As far as we are aware, the only inferential methods that allow for spillover effects are the ones proposed by Stefano and Mellace (2020) and Grossi et al. (2020). Stefano and Mellace (2020) propose the "inclusive synthetic control estimator" that requires a perfect pre-treatment fit (Assumption 2 of their paper), and an inferential procedure that is a modification to the placebo test. Compared with their method, our method allows for imperfect pre-treatment fit. Grossi et al. (2020) consider a penalized version of SCM that is similar to Abadie and LHour (2021) and assume that the units can be clustered into exchangeable groups. Our method does not require such an exchangeability assumption. Furthermore, our estimation and testing procedure applies to factor models with cointegrated common factors, which is of special interest even in cases without spillover effects.

We also contribute to the literature on spillover effects. This fast-growing literature looks into both estimation of treatment effects in the presence of spillover effects, as well as estimation of spillover effects themselves. For example, Vazquez-Bare (2022) considers a framework where observations are grouped into clusters, and spillover effects are allowed within a cluster, but not across clusters. It discusses estimation of heterogeneous treatment effects as a function of the number of treated units within the same cluster, and spillover effects as a function of whether the unit is treated, and number of treated units within the same cluster. Basse et al. (2019) and Rosenbaum (2007) use a randomization test for inference in the presence of spillover effects. Also, see Basse et al. (2019) and Vazquez-Bare (2022) for a literature review on spillover effects. However, this literature rarely looks at the panel data setting with only a few treated units and short post-treatment periods. This limitation is in part because we

usually do not have enough information about the spillover effects in this particular setting. We avoid this problem by requiring the assumption that the spillover structures be pre-specified and follow a pattern that is linear in some underlying parameters. With that specification, 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 applies this idea to OLS, IV, and GMM. Chernozhukov et al. (2021) propose a permutation test 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 (2006) extend Andrews (2003)'s results to the cointegrated cases, we also show that our method is still valid for a cointegrated factor model.

The remainder of this paper is organized as follows. Section 2 introduces a potential outcome framework with spillover effects. Section 3 proposes an estimator and derives its asymptotic distribution. Section 4 considers the P-test introduced by Andrews (2003) and Andrews and Kim (2006), and explains how it can be applied in our settings. We present an empirical example of our method in Section 5 and in Section 6 we present Monte Carlo simulation results. Section 7 discusses some extensions of our methods, including cases with multiple treated units and/or multiple post-treatment periods, and cases with extra covariates. Section 8 concludes. All proofs are in the appendix.

2 Model Specification

2.1 A Rubin model with spillover effects

We start our discussion with a Rubin potential outcome framework. We consider a standard synthetic control setting where only one unit is treated and only one period is available after the treatment is implemented. We consider cases with multiple treated units and multiple post-treatment periods in Section 7.

In Rubin's model with violation of SUTVA, the potential outcomes are functions of treatment assignments on all units. Assume the outcome of unit i at time t is

$$y_{i,t} = y_{i,t}(d_t),$$

where $d_t = (d_{1,t}, \ldots, d_{N,t})'$ and $d_{i,t} = 1$ if unit i has been treated at time t. Assume unit 1 is treated between time T and $T+1$, and there are another $N-1$ units that are not directly treated by the policy. Thus, we observe an $N \times (T+1)$ panel as shown in Figure 1.

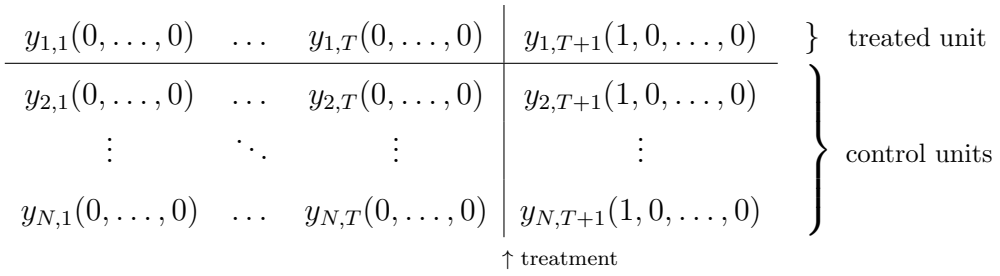


Figure 1

Note that we only observe outcomes with $d_{T+1} = (0,\ldots,0)'$ or $d_{T+1} = (1,0,\ldots,0)'$. This is the fundamental limitation of the dataset we are currently studying. Unless other homogeneity conditions are assumed, we cannot say anything about $y_{i,T+1}(d_{T+1})$ for $d_{T+1} \notin \{(0,\ldots,0)', (1,0,\ldots,0)\}$ because only a few units are treated and only a few post-treatment periods are available. For notation simplicity, let

$$\begin{cases} y_{i,t}(0) = y_{i,t}(0,\ldots,0) \\ y_{i,t}(1) = y_{i,t}(1,0,\ldots,0) \end{cases}$$

for each (i,t). Let $\alpha_i = y_{i,T+1}(1) - y_{i,T+1}(0)$ be the potential deviation from unit i's counterfactual outcome $y_{i,T+1}(0)$ where no unit is treated at time $T+1$. That is, α_1 is the direct treatment effect on unit 1, while α_i with $i \neq 1$ is the indirect effect or spillover effect. The whole effect vector $\alpha = (\alpha_1,\ldots,\alpha_N)'$ can be of interest in our setting. Throughout, we consider the case where N is fixed and T goes to infinity.

2.2 Known spillover structures

Throughout the paper, we assume that some knowledge about the spillover effects is known. Namely, assume that the full effect vector α is a linear transformation of some unknown parameter $\gamma \in \mathbb{R}^k$, i.e., $\alpha = A\gamma$. Typically, γ has fewer dimensions than α does. Note that the linearity is not particularly restrictive of actual spillover structures – but rather mostly imposes knowledge requirements on the estimator. Consider the extreme example where we impose no restrictions on the possible spillovers. This is a special case of $\alpha = A\gamma$ with A being the identity matrix and $\gamma = \alpha$.

Here are some examples that fit this framework.

Example 1. *Assume that a subset of control units, but not all of them, are equally*

affected by the spillover effects, i.e.

$$A = \begin{bmatrix} 1 & 0 \\ 0_{k\times 1} & 1_{k\times 1} \\ 0_{l\times 1} & 0_{l\times 1} \end{bmatrix}, \quad \gamma = \begin{bmatrix} \alpha_1 \\ b \end{bmatrix}, \quad A\gamma = \begin{bmatrix} \alpha_1 \\ b \\ \vdots \\ b \\ 0_{l\times 1} \end{bmatrix},$$

where α_1 is the treatment effect and b is the homogeneous spillover effect.

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

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

Example 3. *Assume that the spillover effect is likely to take place at some known locations, but not at other locations, while the sizes of spillover effects are allowed to vary across those units. For example, assume there are potential 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 $A\gamma$. WLOG order the units by increasing distance from unit 1, and let p be the number of units experiencing spillovers. Then*

$$A = \begin{bmatrix} 1 & 0_{1\times p} \\ 0_{p\times 1} & I_p \\ 0_{(N-p-1)\times 1} & 0_{(N-p-1)\times p} \end{bmatrix}, \ \gamma = \begin{bmatrix} \alpha_1 \\ \alpha_{k_1} \\ \vdots \\ \alpha_{k_p} \end{bmatrix}, \quad A\gamma = \begin{bmatrix} \alpha_1 \\ \alpha_{k_1} \\ \vdots \\ \alpha_{k_p} \\ 0_{(N-p-1)\times 1} \end{bmatrix}.$$

Thus, the units indexed $2, ..., (p+1)$ each experience their own size spillover effect.

The assumptions in Example 3 are often plausible. We give an empirical example in Section 5. If misspecification of the spillover structure is a concern, one can choose an A matrix that incorporates more potential spillovers, i.e., a bigger p. The consequences of misspecification are discussed in Section 3.4.2.

3 An Asymptotically Unbiased Estimator

3.1 SCM without spillover effects

We apply a version of SCM proposed in Ferman and Pinto (2021), which starts with obtaining the synthetic control weights by solving the optimization problem

$$\begin{bmatrix} \widehat{a}_1 \\ \widehat{b}_1 \end{bmatrix} = \underset{\tilde{a}\in\mathbb{R},\tilde{b}\in W^{(1)}}{\arg\min} \sum_{t=1}^{T} (y_{1,t} - \tilde{a} - Y_t'\tilde{b})^2, \tag{1}$$

where $Y_t = (y_{1,t}, \ldots, y_{N,t})'$ and $W^{(1)} = \{(w_1, \ldots, w_N)' \in \mathbb{R}_+^N : w_1 = 0, \sum_{j=2}^N w_j = 1\}$. This restricts the estimation weights such that they sum to 1, are all non-negative, and the same-unit weight is 0, as well as including an intercept term. An estimator of the treatment effect α_1 is given by

$$\widehat{\alpha}_1 = y_{1,T+1} - (\widehat{a} + Y_{T+1}'\widehat{b}),$$

i.e., the counter-factual value $y_{1,T+1}(0)$ is approximated by $\widehat{a} + Y_{T+1}'\widehat{b}$. Here, we do not restrict the intercept but require the other coefficients to be positive and sum up to one.[2] The intercept term a is important in our setting because it takes out the bias by recentering the estimator. Ferman and Pinto (2021) show that when the pre-treatment is imperfect, this estimator is asymptotically unbiased, while the original SCM as in Abadie et al. (2010) has bias.

3.2 The proposed estimator under spillovers

In order to back out the spillover effects, we proceed as follows. We first define the individual synthetic control weights and their limits. Namely, for each i, let the individual-specific synthetic control weights (and the intercept) be

$$\begin{bmatrix} \widehat{a}_i \\ \widehat{b}_i \end{bmatrix} = \underset{\tilde{a}\in\mathbb{R},\tilde{b}\in W^{(i)}}{\arg\min} \sum_{t=1}^{T} (y_{i,t} - \tilde{a} - Y_t\tilde{b}')^2, \tag{2}$$

where $W^{(i)} = \{(w_1, \ldots, w_N)' \in \mathbb{R}_+^N : w_i = 0, \sum_{j=1}^N w_j = 1\}$. These are the same restrictions as above (sum-to-one, non-negativity, same-weight is 0). Then, let the

[2] Other choices of constraint set for $(\widehat{a}_1, \widehat{b}_1')'$ include $\{0\} \times \{0\} \times \Delta_{N-1}$ as in the original synthetic control method of Abadie and Gardeazabal (2003) and Abadie et al. (2010), and $\mathbb{R} \times \{0\} \times \mathbb{R}_+^{N-1}$ as in the modified synthetic control of Li (2020), where $\Delta_{N-1} = \{w \in \mathbb{R}^{N-1} : w_i \geq 0 \text{ for each } i, \sum_{i=1}^{N-1} w_i = 1\}$ is a $(N-1)$-dimensional simplex. See Doudchenko and Imbens (2017) for a discussion of choosing restriction sets.

probability limit of the intercept and weights be

$$a_i = \text{plim}_{T\to\infty}\ \widehat{a}_i,\ \ b_i = \text{plim}_{T\to\infty}\ \widehat{b}_i,$$

and we only consider cases where they are well-defined. We show later by Lemma 1 in Section 3.3 that a_i and b_i exist for each i in factor models with stationary or cointegrated common factors. In general, a_i and b_i do not coincide with the weights that reconstruct the factor loadings (Ferman and Pinto, 2021).

For each (i,t), define the specification error by

$$u_{i,t} = y_{i,t}(0) - (a_i + Y_t(0)'b_i). \tag{3}$$

Define $a = (a_1, \ldots, a_N)'$ and $B = (b_1, \ldots, b_N)'$. Stacking Equation (3) for all i's gives $u_t = Y_t(0) - (a + BY_t(0))$, where $u_t = (u_{1,t}, \ldots, u_{N,t})'$. Since $\alpha = Y_{T+1}(1) - Y_{T+1}(0)$, we have at period $T+1$

$$u_{T+1} = (I - B)(Y_{T+1} - \alpha) - a, \tag{4}$$

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

We form estimators for (a, B) using synthetic control methods as in (2). We do that for each $i = 1, \ldots, N$, as if each i is the treated unit and other units are controls. Then, the estimators for a and B are $\widehat{a} = (\widehat{a}_1, \ldots, \widehat{a}_N)'$ and $\widehat{B} = (\widehat{b}_1, \ldots, \widehat{b}_N)'$ respectively. Define $M = (I - B)'(I - B)$ and let $\widehat{M} = (I - \widehat{B})'(I - \widehat{B})$ be an estimator for M.

Recall that the effect vector is $\alpha = A\gamma$. Let an estimator of γ be such that

$$\begin{aligned}\widehat{\gamma} &= \arg\min_{g\in\mathbb{R}^k} \|(I - \widehat{B})(Y_{T+1} - Ag) - \widehat{a}\| \\ &= (A'\widehat{M}A)^{-1}A'(I - \widehat{B})'((I - \widehat{B})Y_{T+1} - \widehat{a}).\end{aligned} \tag{5}$$

Note that the FOC implies $A'(I - \widehat{B})'\widehat{u}_{T+1} = 0$, where $\widehat{u}_{T+1} = (I - \widehat{B})(Y_{T+1} - \widehat{\alpha}) - \widehat{a}$, i.e., it requires that some weighted sum of the residuals be zero. Under that condition, the treatment and spillover effect vector α can be estimated by $\widehat{\alpha} = A\widehat{\gamma}$.

Assumption 1. (a) $\{u_t\}_{t\geq 1}$ *is stationary, and has mean zero.*
(b) $\|\widehat{a} - a\| = o_p(1)$, $\|\widehat{B} - B\| = o_p(1)$.
(c) $\|(\widehat{B} - B)Y_{T+1}(0)\| = o_p(1)$.
(d) $A'MA$ *is non-singular.*

Part (a) generally requires that there is no regime shift or structural break. Part (b) and (c) requires that there are at least a moderate number of pre-treatment periods so

that the synthetic control weights are well-estimated. We show later that Assumption 1(a)-(c) are satisfied in factor models with either stationary or cointegrated common factors. We will discuss Part (d) later in Section 3.4.1.

Theorem 1. *Suppose Assumption 1 holds. Then,* $\widehat{\alpha} - (\alpha + Gu_{T+1}) \to_p 0$ *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 in Ferman and Pinto (2021), as it is inconsistent but asymptotically unbiased. Note that consistent estimators are impossible because only one post-treatment period with one treated unit is available, so the error term for that one unit and period does not shrink in any limit we consider.

3.3 The factor model as an example

Factor models are often used to justify the usage of synthetic control methods. Here we show that our assumptions are satisfied by factor models with stationary and cointegrated common factors. We follow Ferman and Pinto (2021) and consider a factor model such that for $i = 1, \ldots, N$ and $t = 1, \ldots, T+1$,

$$y_{i,t}(0) = \eta_t + \lambda_t' \mu_i + \varepsilon_{i,t}, \tag{6}$$

where λ_t is F-dimensional common factors with a fixed F, and $\varepsilon_{i,t}$ is the noise that is uncorrelated with λ_t. For notation simplicity, we write $Y_t(0) = (y_{1,t}(0), \ldots, y_{N,t}(0))'$, $Y_t = (y_{1,t}, \ldots, y_{N,t})'$, and $\varepsilon_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{n,t})'$.

We focus on two sets of conditions in our discussion.

Condition ST (model with stationary common factors)**.** *Assume* $\{(\eta_t, \lambda_t, \varepsilon_t)\}_{t \geq 1}$ *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

$$a_i = E[y_{i,1}(0) - Y_1(0)'b_i], \; b_i = \underset{w \in W^{(i)}}{\arg\min} \; (w - e_i)'\Omega_y(w - e_i),$$

where e_i is a unit vector with one at the i-th entry and zeros everywhere else, and $W^{(i)} = \{(w_1, \ldots, w_N) \in \mathbb{R}_+^N : w_i = 0, \sum_{j \neq i} w_j = 1\}$. Note that in general b_i does not recover the factor structure, because $\mu_i \neq (\mu_1, \ldots, \mu_N)b_i$ 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 asymptotically unbiased and the test proposed below to be valid is to include an intercept in the optimization problem (2).

Condition CO (model with cointegrated $\mathcal{I}(1)$ common factors)**.** *Rewrite Equation* (6) *as*

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

and η_t *can be either in* λ_t^1 *or* λ_t^0*. Assume* $\{(\lambda_t^0, \varepsilon_t)\}_{t\geq 1}$ *is stationary, ergodic for the first and second moments, and has a finite* 4*-th moment. Without loss of generality,* $E[\varepsilon_{i,t}] = 0$*. 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* $T^{-1/2}y_{i,[rT]}(0) \Rightarrow \nu_i(r)$*, 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* $W^{(i)} = \{(w_1, \ldots, w_N) \in \mathbb{R}_+^N : w_i = 0, \sum_{j\neq i} w_j = 1\}$*. 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.*

Note that Condition CO puts restrictions on the factor loadings. The restrictions are similar to those in Ferman and Pinto (2021).

The relevance of the factor model is given by the following lemma:

Lemma 1. *Suppose* $A'MA$ *is non-singular. Then, either Condition ST or Condition CO implies Assumption 1.*

Thus, results derived in Theorem 1 apply to factors models with Condition ST or Condition CO.

3.4 Discussion

3.4.1 Invertibility of $A'MA$

In Assumption 1(d), we require that $A'MA$ is invertible. First, note the invertibility of $A'MA$ is testable in principle. Recall that $M = (I-B)'(I-B)$, so that $A'MA = A'(I-B)'(I-B)A$. A is identified by the econometrician ahead of time. We can consistently estimate B so the data informs us of the validity of this assumption.

To understand this assumption better, we replace α by $A\gamma$ in Equation (4) and have

$$(I-B)A\gamma = (I-B)Y_{T+1} - a - u_{T+1}. \tag{7}$$

Equation (7) is the key to learning α. Under mild regularity conditions, a and B are identified from the model and learned by the synthetic control method. We do not observe u_{T+1}, but the distribution of u_{T+1} can be learned using pre-treatment data under stationarity of $\{u_t\}_{t\geq 1}$. Therefore, if $A'MA$ is non-singular, or equivalently, $(I-B)A$ has full rank, we can form an estimator of γ whose limiting distribution is identified by multiplying both sides of Equation (7) by $(A'MA)^{-1}A'(I-B)'$. Note

that we do not point-identify γ or α. This is because we have only one observation of the outcome in post-treatment periods.

We illustrate the invertibility of $A'MA$ in the following toy example.

Example 4. *Assume there are 3 units in total, where unit 1 is treated. WLOG, let the synthetic control weight matrix B be*

$$B = \begin{bmatrix} 0 & w_1 & 1-w_1 \\ w_2 & 0 & 1-w_2 \\ w_3 & 1-w_3 & 0 \end{bmatrix}.$$

Suppose the researcher first assumes unit 2 and 3 are equally exposed to the spillover effects. That is, they assume

$$A_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}, \gamma = \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix}, \text{ and } \alpha = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_2 \end{bmatrix}.$$

Then, $A'MA$ is non-invertible, because

$$(I-B)A_1 = \begin{bmatrix} 1 & -1 \\ -w_2 & w_2 \\ -w_3 & w_3 \end{bmatrix}.$$

Intuitively, the problem here is there are two control observations we want to take a difference from, to determine the treatment effect.

If they instead assume only one of the controls is exposed to the spillover effects, $A'MA$ is non-singular in general. In this case,

$$A_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \gamma = \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix}, \text{ and } \alpha = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ 0 \end{bmatrix},$$

and

$$(I-B)A_2 = \begin{bmatrix} 1 & -w_1 \\ -w_2 & 1 \\ -w_3 & w_3 - 1 \end{bmatrix}.$$

It can be shown that $(I-B)A_2$ always has full rank for $(w_1, w_2, w_3) \in [0,1]^3$.

This applies to more general settings. That is, if all controls are equally hit by the spillover effects, then $(I-B)A$ does not have full rank and $A'MA$ is non-invertible. Allowing a few units to be exempt from the spillover effects makes $(I-B)A$ have full

rank in general.

A more interesting case is Example 3, where we only restrict the range of spillover effects and allow the levels to vary. In this case, $(I-B)A$ can be obtained by eliminating columns that correspond to units that are neither treated nor exposed to spillover effects. Again, as long as at least one control is not exposed to the spillover effects, $(I-B)A$ has full rank in general. This assumption is more convincing if a moderate number of columns are eliminated from $(I-B)$, i.e., only a few units are exposed to the spillover effects.

3.4.2 Structure misspecification

Misspecification of the spillover structure can lead to asymptotic bias in treatment effect estimation. Therefore, we suggest that the researcher be conservative about choosing the structure. That is, if a certain unit is suspected to be affected by spillover effects, it should be included in the spillover structure.

On other hand, we can show that the misspecification bias of the proposed method is only a function of the "missed" spillover effects, while that of the usual SCM is a function of all spillovers. For concreteness, we follow Example 3 and start with the case where unit 2 and 3 are affected by spillover effects. The usual synthetic control estimator is

$$\widehat{\alpha}_{1,SCM} = Y_{1,T+1} - \left(\widehat{a} + \sum_{i=2}^{N} \widehat{b}_{1,i} Y_{i,T+1}(0)\right) - \widehat{b}_{1,2}\alpha_2 - \widehat{b}_{1,3}\alpha_3,$$

so the (asymptotic) bias is

$$\delta_{1,SCM} = -b_{1,2}\alpha_2 - b_{1,3}\alpha_3.$$

Suppose we only include unit 2 as the spillover unit, then the bias of our treatment effect estimator can be shown to be

$$\delta_{1,SP} = -\frac{\det\left([\tilde{b}_1, \tilde{b}_2]'[\tilde{b}_2, \tilde{b}_3]\right)}{\det(A'MA)}\alpha_3,$$

where $\tilde{b}_i = e_i - (b_{1,i}, b_{2,i}, \ldots, b_{N,i})'$ and e_i is the unit vector with one on the i-th entry and zeros everywhere else. Generally, there is no guarantee that either $\delta_{1,SCM}$ or $\delta_{1,SP}$ is smaller.

For the general case, suppose $\alpha = (\alpha_1, \ldots, \alpha_{k_1}, \ldots, \alpha_{k_2}, 0, \ldots, 0)'$. The bias of the usual synthetical control is

$$\delta_{1,SCM} = -\sum_{i=2}^{k_2} b_{1,i}\alpha_i.$$

Suppose we only include Unit 2, 3, ..., k_1 in the spillover structure, then the asymptotic bias of our estimator for the whole vector α is

$$\delta_{SP} = (A(A'MA)^{-1}A'(I-B)'(I-B) - I)(0,\ldots,0,\alpha_{k_1+1},\ldots,\alpha_{k_2},0,\ldots,0)'. \quad (8)$$

Therefore, the bias for the treatment effect estimator is a linear combination of omitted spillovers in the form of

$$\delta_{1,SP} = \sum_{i=k_1+1}^{k_2} c_i\alpha_i,$$

where c_i is determined in the first entry of Equation 8. Again, the asymptotic bias $\delta_{1,SP}$ is only a linear combination of the "missed" spillovers, but it is generally difficult to compare the misspecification bias of our estimator and the usual estimator that does not account for spillover effects at all.

4 Statistical Inference

In this section, we discuss formal results on inference. At a high level, our test uses pre-treatment data to form the null distribution of a pre-specified post-treatment quantity. We only consider cases with imperfect pre-treatment fit to facilitate the identification of the null distribution. In Section 4.1, we consider the case without spillover effects, and state the assumptions under which Andrews' P test (Andrews, 2003) is valid. This result might be interesting by its own interest. In Section 4.2, we generalize P test to cases where spillover effects cannot be ignored, and allow for a more general set of hypotheses.

4.1 Cases without spillover effects

Suppose for now that there are no spillover effects ($\alpha_2 = \cdots = \alpha_N = 0$). We want to test for the existence of treatment effect on unit 1. The null and alternative hypotheses of interest are

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

The test procedure we consider here is the end-of-sample instability test (P-test) in Andrews (2003). The usage of Andrews' test in the context of synthetic control methods is mentioned in Ferman and Pinto (2019), where they focus on the difference-in-differences estimator. We formalize this idea and derive conditions under which Andrews' test delivers valid inference results.

We assume that α_1 is not a function of T under H_1. That is, we consider fixed, not

local, alternatives, as in Andrews (2003) and Andrews and Kim (2006). Specifically, 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

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

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 hypotheses become $H_0 : a_1^* = a_1$ and $H_1 : a_1^* \neq a_1$, respectively. Let the synthetic control regression residuals be $\widehat{u}_{1,t} = y_{1,t} - \widehat{a}_1 - Y_t'\widehat{b}_1$. If there is no treatment effect, the distribution of $\widehat{u}_{1,T+1}$ should be asymptotically equivalent to that of $\widehat{u}_{1,t}$ for $t \leq T$. Using this idea, define the test statistic by

$$P = \widehat{u}_{1,T+1}^2.$$

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

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

Then, $P = (y_{1,T+1} - x_{T+1}'\widehat{\beta}_1)^2 = P_{T+1}(\widehat{\beta}_1)$. The pre-treatment counterparts are defined by $P_t = P_t(\widehat{\beta}_1^{(t)})$, where $\widehat{\beta}_1^{(t)} = \widehat{\beta}_1$ for each t.[3] For a significance level of τ, we reject H_0 if P is larger than the $(1-\tau)$-quantile of $\{P_t\}_{t=1}^T$.

To establish the validity of the proposed test, let P_∞ be a random variable with the same distribution as $P_{T+1}(\beta_1)$ with $\beta_1 = (a_1, b_1')'$. Define the empirical CDF of $\{P_t\}_{t=1}^T$ by

$$\widehat{F}_{P,T}(x) = \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}\{P_t \leq x\},$$

and let $F_P(x)$ be the distribution function of $P_1(\beta_1)$. We reject H_0 if $P > \widehat{q}_{P,1-\tau}$, where $\widehat{q}_{P,1-\tau} = \inf\{x \in \mathbb{R} : \widehat{F}_{P,T}(x) \geq 1-\tau\}$. Finally, let $q_{P,1-\tau}$ be the $(1-\tau)$-quantile of $P_1(\beta_1)$.

Assumption 2. (a) $\{u_t\}_{t\geq 1}$ *are stationary, ergodic, and have mean zero.*
(b) $E[|u_t|] < \infty$.
(c) $\exists$ *a non-random sequence of positive definite matrices* $\{C_T\}_{T\geq 1}$ *such that* $\max_{t\leq T+1} \|C_T^{-1} x_t\| = O_p(1)$

[3]Readers can also use leave-one-estimator to construct P_t as in Andrews (2003) and Andrews and Kim (2006). For $t = 1, \ldots, T$, the leave-one-out estimator $\widehat{\beta}_1^{(t)}$ is defined by the synthetic control weight estimator using only observations indexed by $s = 1, \ldots, t-1, t+1, \ldots, T$.

(d) $\|C_T(\widehat{\beta}_1 - \beta_1)\| = o_p(1)$, *and* $\max_{t=1,\ldots,T} \|C_T(\widehat{\beta}_1^{(t)} - \beta_1)\| = o_p(1)$.
(e) *The distribution function of* $P_1(\beta_1)$ *is continuous and increasing at its* $(1-\tau)$*-quantile.*

Assumption 2 is similar to those in Andrews (2003). Part (a) does not allow for a structural break. Part (b) and (c) are moment conditions. Part (d) requires at least a moderate number of pre-treatment periods so that the synthetic control weights are well-estimated. Part (e) generally requires that u_t follows a continuous distribution.

Theorem 2. *Suppose Assumption 2 holds. Then, as* $T \to \infty$,
(a) $P \to_d P_\infty$ *under* H_0 *and* H_1,
(b) $\widehat{F}_{P,T}(x) \to_p F_P(x)$ *for all* x *in a neighborhood of* $q_{P,1-\tau}$ *under* H_0 *and* H_1,
(c) $\widehat{q}_{P,1-\tau} \to_p q_{P,1-\tau}$ *under* H_0 *and* H_1,
(d) $\Pr(P > \widehat{q}_{P,1-\tau}) \to \tau$ *under* H_0.

Theorem 2 states that the distribution of our test statistic P can be approximated by the empirical distribution of $\{P_t\}_{t=1}^T$. Specifically, Part (d) shows that the proposed test is asymptotically valid in the sense that the rejection probability under the null is convergent to the nominal level.

We also show the relevance of the factor model in this context by the following lemma:

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

4.2 Cases with spillover effects

In this section, we generalize Section 4.1 to cases allowing 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 hypotheses we consider are $H_0 : C\alpha = d$ and $H_1 : C\alpha \neq d$, with known C and d. 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 effectively makes Section 4.1 a special case of our test, although Theorem 2 has slightly stronger results than Theorem 3 does. If we want to test that there is a spillover, then we can let $C = [0_{(N-1)\times 1}\ I_{N-1}] \in \mathbb{R}^{(N-1)\times N}$ and $d = (0, \ldots, 0)' \in \mathbb{R}^{(N-1)\times 1}$.

The test statistic we consider here is $P = (C\widehat{\alpha} - d)' W_T (C\widehat{\alpha} - d)$ for some weighting matrix $W_T \to_p W$. Recall $G = A(A'MA)^{-1}A'(I-B)$ and can be consistently estimated by $\widehat{G} = A(A'\widehat{M}A)^{-1}A'(I-\widehat{B})$ if $\widehat{B} \to_p B$. By Theorem 1, P is asymptotically equivalent

to $u'_{T+1}G'C'WCGu_{T+1}$. To construct critical values, define

$$P_t(\theta) = (Y_t - \theta x_t)'G'C'WCG(Y_t - \theta x_t),$$

and

$$\widehat{P}_t(\theta) = (Y_t - \theta x_t)'\widehat{G}'C'W_TC\widehat{G}(Y_t - \theta x_t),$$

for some $\theta \in \mathbb{R}^{N\times(N+1)}$, $x_t = (1, Y_t')'$, and $\widehat{G} = A(A'\widehat{M}A)^{-1}A'(I - \widehat{B})'$. Let $\widehat{P}_t = \widehat{P}_t(\widehat{\theta}^{(t)})$, where $\widehat{\theta}^{(t)} = \widehat{\theta}$ for each t.[4] For a significance level of τ, we reject H_0 if P is larger than the $(1-\tau)$-quantile of $\{\widehat{P}_t\}_{t=1}^T$.

To establish the validity of the proposed test, let $P_\infty = P_1(\theta_0)$ for $\theta_0 = [a\ B]$. Define

$$\widehat{F}_{P,T}(x) = \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}\{\widehat{P}_t \le x\},$$

and let $F_P(x)$ be the distribution function of P_∞. Finally, let $\widehat{q}_{P,1-\tau} = \inf\{x \in \mathbb{R} : \widehat{F}_{P,T}(x) \ge 1-\tau\}$, and $q_{P,1-\tau}$ be the $(1-\tau)$-quantile of P_∞. The assumptions and validity of the proposed testing procedure are given as follows.

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

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

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

Assumption 3(b)-(f) are similar to Assumption 2 as well as those in Andrews (2003).

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

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

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

(c) $\widehat{q}_{P,1-\tau} \to_p q_{P,1-\tau}$,

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

Just like Theorem 2, Theorem 3 shows that we can approximate the null distribution of P using its pre-treatment counterparts. Part (d) shows the asymptotic validity of the test proposed in this section.

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

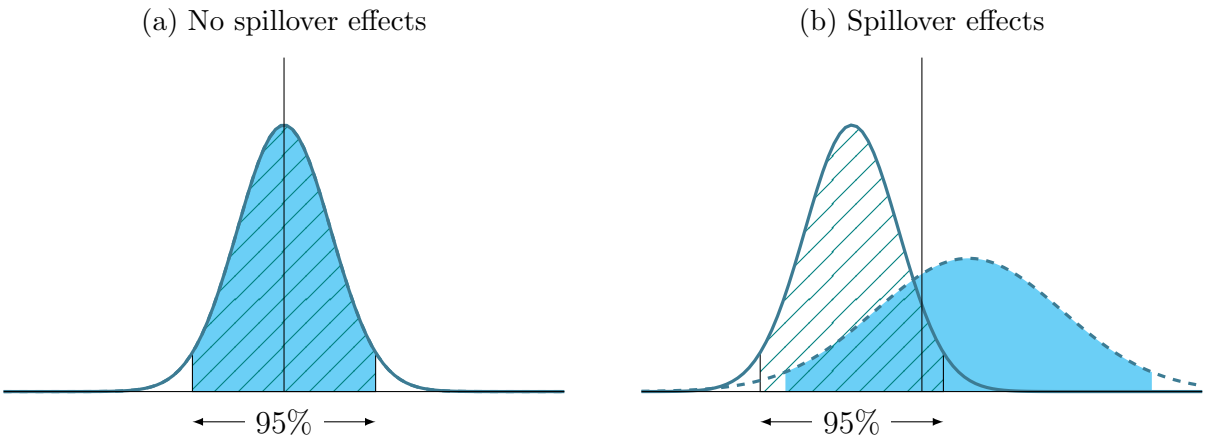


Figure 2: Placebo test. Area with lines is 95% probability region of the error of the treated unit. Filled area is 95% probability region of null distribution formed in placebo test. A test is rejected when the error of the treated units falls outside of the filled area.

Again, we show the relevance of the factor model in this context by the following lemma:

Lemma 3. *Suppose that $A'MA$ is non-singular and the distribution function of $P_1(\theta_0)$ is continuous and increasing at its $(1-\tau)$-quantile. Then, Assumption 3 is satisfied if either of these holds:*

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

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

4.3 Other testing procedures

When we allow for the existence of non-zero spillover effects, the existing testing procedures will have poor performance. Here we intuitively explain what happens to the placebo test as in Abadie et al. (2010) and Andrews' test as in Andrews (2003) in the presence of spillover effects.

Suppose we want to test for the treatment effect being zero and are not aware of the spillover effects. Placebo test and Andrews' test are similar in the sense that they use data to form the null distribution of $u_{1,T+1}$ in order to perform hypothesis testing. The difference is that the placebo test exploits variations of $\{\widehat{u}_{i,T+1}\}_{i=1}^{N}$, while Andrews' test uses variations of $\{\widehat{u}_{1,t}\}_{t=1}^{T+1}$.

We look at the placebo test first. When there is no spillover effect, the distribution of $\widehat{u}_{1,T+1}$ and that of any element in $\{\widehat{u}_{i,T+1}\}_{i=2}^{N}$ coincide asymptotically. As shown in Figure 2(b), when there are positive spillover effects, we will underestimate the treatment effect and the density function of $\widehat{u}_{1,T+1}$ moves to the left. At the same time, some of the control units shift to the right because of the positive spillovers, so

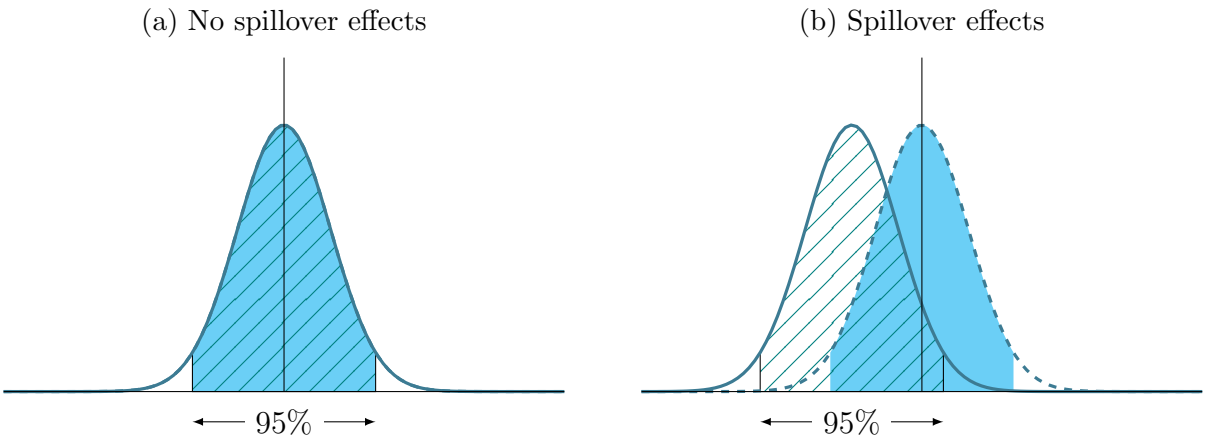
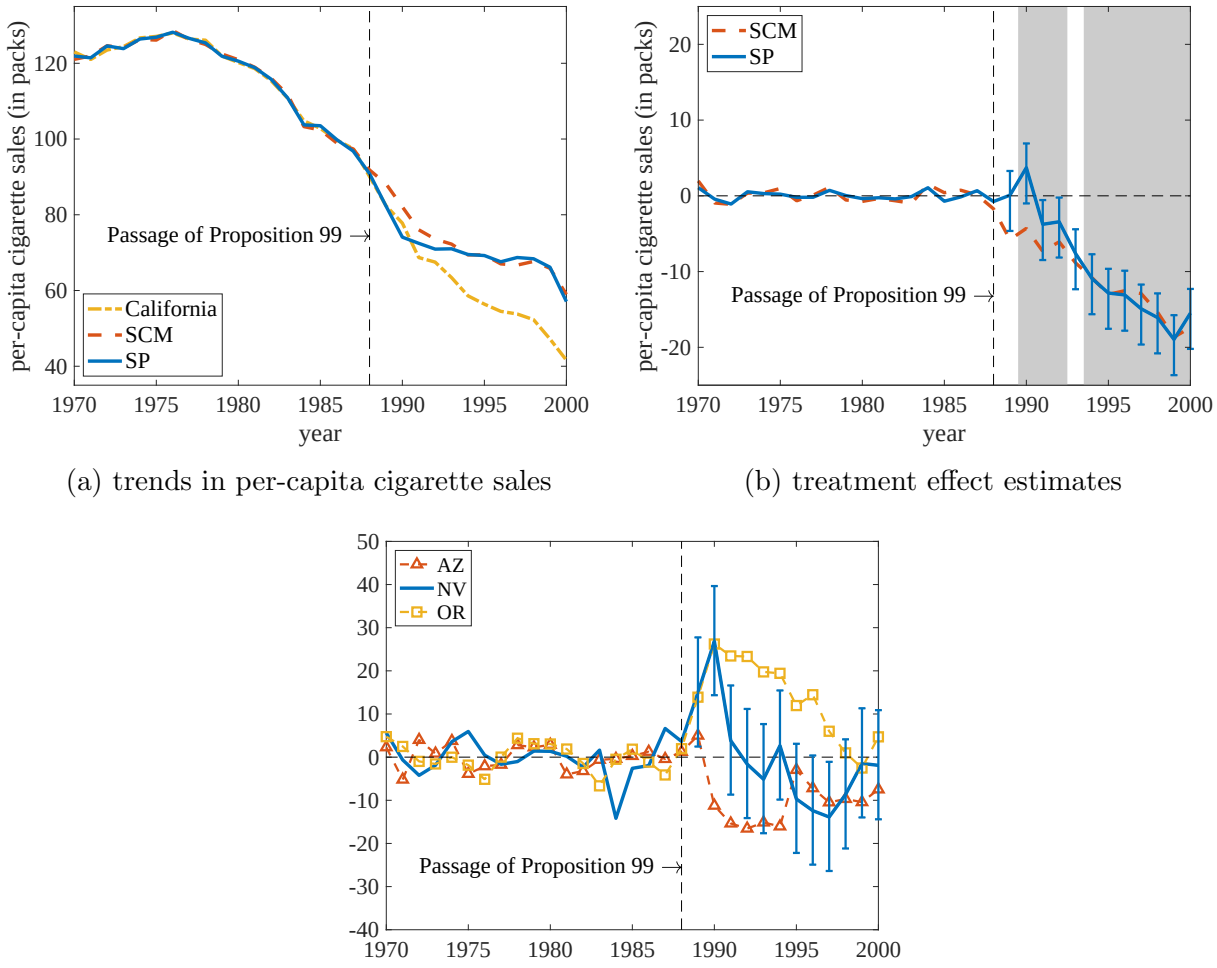


Figure 3: Andrews' test. Area with lines is 95% probability region of the error of the treated unit. Filled area is 95% probability region of null distribution formed in Andrews' test. A test is rejected when the error of the treated units falls outside of the filled area.

density of $\{\widehat{u}_{i,T+1}\}_{i=2}^{N}$ moves to the right and gets wider. In terms of test performance, the shift of $\widehat{u}_{1,T+1}$ is offset by the wider density of $\{\widehat{u}_{i,T+1}\}_{i=2}^{N}$ (harder to reject H_0), which explains why the empirical sizes of placebo test often do not deviate too much from the nominal size, even in the presence of spillovers (see, for example, Table 3 for $T = 50$ and 200 cases in Section 6). In essence, the placebo test becomes much more conservative and has low power as shown in Table 4.

Now we consider Andrews' test. When there is no spillover effect, the distribution of $\widehat{u}_{1,T+1}$ and that of any element in $\{\widehat{u}_{1,t}\}_{t=1}^{T}$ coincide asymptotically. As shown in Figure 3(b), when there is a positive spillover effect, we underestimate the treatment effect and the density function of $\widehat{u}_{1,T+1}$ shifts to the left, while the density of $\{\widehat{u}_{1,t}\}_{t=1}^{T}$ doesn't, since they are pre-treatment and the spillover only happens after the treatment. This results in an invalid test.

Although not the main focus of this paper, selection into treatment can be a threat to the placebo test in practice. For example, if a unit is more likely to be treated when its own outcome y is higher than those of other units, then the placebo test tends to over-reject the zero treatment effect hypothesis, even without spillover effects. This form of selection is not a problem for Andrews' test and the test proposed in this paper, since they use the variation across time periods instead of different units.



(a) trends in per-capita cigarette sales

(b) treatment effect estimates

(c) spillover effect estimates

Figure 4: Panel (a): Per-capita cigarette sales in California, synthetic California (SCM), and spillover-adjusted synthetic California (SP). SCM is obtained by using standard synthetic control method. SP is using our estimation procedure, which accounts for spillover effects. The vertical line indicates the start of treatment. Panel (b): Per-capita cigarette sales gap between California and (spillover-adjusted) synthetic California (with 95% confidence intervals for SP). The lines to the right of passage of Proposition 99 are treatment effect estimates. Shaded area denotes our test rejects that there is no spillover effect in those years. Panel (c): Spillover effect estimates for Arizona, Nevada, and Oregon (95% confidence intervals are shown only for Nevada for clarity of exposition).

5 Estimating the Effects of California's Proposition 99

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

Proposition 99 intended to disincentivize smoking, which was 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 cigarette consumption. 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 (and other vice taxes like gambling and alcohol) is that measured local consumption might fall as people move their purchasing behavior across legal boundaries, particularly in early years. In order to accommodate this, we allow for a spillover affecting states neighboring California, i.e., Arizona, Nevada, and Oregon. One might also think that there could be policy contamination whereby culturally close states also enact policies with similar targets. Our method can allow for this kind of spillover in our estimation. ADH took that type of problem into account, and 12 states which experienced legislative changes in the ensuing years were removed in that paper.

The data used is per capita cigarette consumption in the 50 states plus the District of Columbia running from 1970 to 2000. In 1989 California enacted Proposition 99, so all periods from 1989 onwards are considered post-treatment periods. We replicate this program evaluation 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 spillover effects in states that are geographically close or have experience policy contamination, but not the others.[5] Those spillover effects are allowed to be different for different states and different time periods. We also perform hypothesis testing on both treatment effects and spillover effects. Since we have multiple post-treatment periods, we treat each post-treatment period as if it is the year right after the policy implementation. The details are outlined in Section 7.3.

The results are shown in Figure 4. The standard synthetic control method that is similar to Abadie et al. (2010) is indexed by SCM and our method is SP. Figure 4a shows the "synthetic California" and 4b elaborates on this by specifically looking at

[5]The states that are considered exposed to spillovers include AK, AZ, DC, FL, HI, MA, MD, MI, NJ, NV, NY, OR, and WA.

the estimated treatment effects. Figure 4c plots the estimated spillover effects for the three neighboring states of California. The error bars denote 95% confidence intervals that are built by inverting the test proposed in Section 4.2. As you can see, the scale of the error bars is visually larger than the amount of variation in the pre-period – this is a consequence of estimating and then adjusting for spillovers – which adds uncertainty to our main effect estimates.

As Figure 4a shows, our estimated consumption in the "synthetic California" does not differ substantially from what a standard SCM would predict, especially for later periods. However, our estimates of the first two post-treatment periods (1989 and 1990) are not significantly different from zero at a 95% level, in contrast with SCM. This difference may result from the over-estimation of the scale of the treatment effects by SCM in the presence of spillover effects. From the tests of spillover effects (shaded area of Figure 4b), we see that likely there were substantial spillover effects. One potential cause of spillovers may be that consumers in California shifted their purchasing to the nearby states, Arizona, Nevada, and Oregon. Since similar laws with a tax increase on cigarettes were passed in Arizona and Oregon in 1994 and 1996, separately, it is difficult to distinguish the spillover effects of Proposition 99 from anticipation effects as well as direct effects of their own laws. Nevada however did not pass any such laws in this period, so the treatment effect estimates for Nevada are more reliable. From Figure 4c, we observe that Nevada has experienced significant spillover effects in the first two periods of the passage of Proposition 99 and mostly insignificant effects afterward, with the exception of 1997. This is consistent with our conjecture and may provide more evidence on how the effects of the policy have propagated.

The non-significant effects of policy right after the implementation could be explained by the addictive behavior of cigarette consumption. The persistence of cigarette consumption has been extensively studied and well-understood by both the rational addiction and the medical literature (Baltagi and Griffin, 2001; Baumeister, 2017; Becker et al., 1994; Benowitz, 1992; Christelis and de Galdeano, 2011; Labeaga, 1999; Miura, 2019; Vleeming et al., 2002). Compared with SCM, our results are more consistent with an addiction story, that tobacco consumption is addictive and unlikely to drop immediately after the policy, but rather slowly transition to a lower equilibrium.

6 Monte Carlo Simulations

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

6.1 Estimation with spillover effects

In this subsection, we examine the finite sample performance of our estimation procedure proposed in Section 2.2. The model considered here is similar to Li (2020), where $y_{i,t}(0)$ follows a factor model structure. We show 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									
SCM	-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)
SP	-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									
SCM	-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)
SP	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									
SCM	-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)
SP	-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)

Notes: The numbers without parentheses are empirical bias in simulation. The ones with parentheses are empirical variance. SCM is the standard synthetic control method assuming no spillover effects. SP is the estimation procedure proposed in this paper that takes spillover effects into account. *No spillover effects* stands for the cases where the true DGP has no spillover effects. *Concentrated spillover effects* is the case where 1/3 of the control units receive a spillover effect (of the same level). *Spreadout spillover effects* is the case where 2/3 of the control units receive a spillover effect.

6.1.1 Stationary case

The underlying factor model is

$$y_{i,t}(0) = \eta_t + \lambda_t' \mu_i + \varepsilon_{i,t},$$

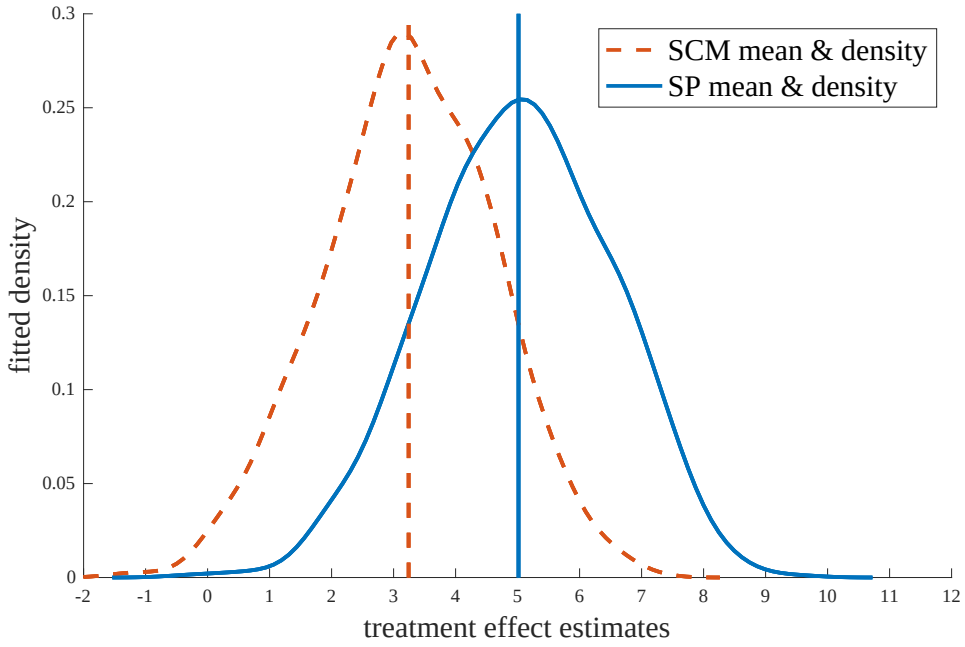


Figure 5: Distribution of treatment effect estimates. The true treatment effect is 5. SCM is using the standard synthetic control method assuming no spillover effects. SP is the estimation procedure proposed in this paper that takes spillover effects into account. Estimates are fitted using kernel density.

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

$$\begin{aligned}
\eta_t &= 1 + 0.5\eta_{t-1} + \nu_{0,t}, \\
\lambda_{1,t} &= 0.5\lambda_{1,t-1} + \nu_{1,t}, \\
\lambda_{2,t} &= 1 + \nu_{2,t} + 0.5\nu_{2,t-1}, \\
\lambda_{3,t} &= 0.5\lambda_{3,t-1} + \nu_{3,t} + 0.5\nu_{3,t-1},
\end{aligned}$$

and $\varepsilon_{i,t}$ and $\nu_{j,s}$ is i.i.d. $N(0,1)$ for each (i,j,s,t). Each entry of μ_i is drawn from an independent uniform distribution on $[0,1]$ and fixed for all repetitions. 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.

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. SCM is the original synthetic control method, and SP is the

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									
SCM	-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)
SP	-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									
SCM	-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)
SP	-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									
SCM	-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)
SP	-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)

Notes: The numbers without parentheses are empirical bias in simulation. The ones with parentheses are empirical variance. SCM is the standard synthetic control method assuming no spillover effects. SP is the estimation procedure proposed in this paper that takes spillover effects into account. *No spillover effects* stands for the cases where the true DGP has no spillover effects. *Concentrated spillover effects* is the case where 1/3 of the control units receive a spillover effect (of the same level). *Spreadout spillover effects* is the case where 2/3 of the control units receive a spillover effect.

corrected synthetic control method proposed in Section 3. Throughout the simulations, assume that we know the coverage of spillover effects but no other information, so A is constructed as in Example 3. For *No spillover effects*, we are being conservative in our use of the SP estimator and run it as if 1/3 of the control units are exposed to spillover effects. To better compare results, we also fit the simulation results using kernel density for the $(N,T) = (10,50)$ case with concentrated spillover effects and plot it in Figure 5.

The empirical bias and variance (in parentheses) of the treatment effect estimator using two methods are shown in Table 1. Throughout, SP produces virtually unbiased estimates, while the usual SCM has a bias that increases as spillovers propagate. For all cases in *Concentrated spillover effects* and five out of nine cases in *Spreadout spillover effects*, SP has a smaller empirical variance than SCM does.

6.1.2 $\mathcal{I}(1)$ case

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

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

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

$$\begin{aligned}
\lambda_{1,t} &= \lambda_{1,t-1} + 0.5\nu_{1,t}, \\
\lambda_{2,t} &= \lambda_{2,t-1} + 0.5\nu_{2,t}, \\
\lambda_{3,t} &= 0.5\lambda_{3,t-1} + \nu_{3,t},
\end{aligned}$$

and $\varepsilon_{i,t}$ 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 while other settings are the same as the stationary case.

The results are shown in Table 2. Similarly as in the stationary case, SP produces virtually unbiased results, while SCM is biased. One thing different here is that SP often has a larger variance than SCM does, except for four out of nine cases in *Concentrated spillover effects*. SP has an especially large variance when $N = 10$ and $T = 200$.

6.2 Test for treatment effects

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 6.1.1 (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 et al. (2010) and Hahn and Shi (2017). Andrews' test is as in Andrews (2003). SP is the spillover-adjust test proposed in Section 4.2.

Among the three testing procedures, SP test has mostly correct sizes and outperforms the other two methods in power. The placebo test has correct sizes in some cases but has lower power, and Andrews' test over-rejects under the null. The reasons are discussed in Section 4.3.

It is worth mentioning that Andrews' test and SP test may experience over-rejection in cases with small T. For example, in the case with $(N,T) = (50,15)$ in Table 3, Andrews' test rejects the null 14.1% of the time in *No spillover effects*, and SP test

Table 3: Empirical rejection rate of testing for treatment effects under $H_0 : \alpha_1 = 0$.

	$N = 10$			$N = 30$			$N = 50$		
	$T = 15$	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
Andrews	0.076	0.061	0.060	0.108	0.082	0.065	0.141	0.078	0.072
SP	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
Andrews	0.411	0.207	0.224	0.417	0.279	0.346	0.519	0.346	0.184
SP	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
Andrews	0.576	0.478	0.399	0.685	0.563	0.616	0.741	0.621	0.544
SP	0.036	0.035	0.042	0.034	0.042	0.046	0.030	0.042	0.044

Notes: SP is the estimation procedure proposed in this paper that takes spillover effects into account. *No spillover effects* stands for the cases where the true DGP has no spillover effects. *Concentrated spillover effects* is the case where 1/3 of the control units receive a spillover effect. *Spreadout spillover effects* is the case where 2/3 of the control units receive a spillover effect of the same level.

rejects the null 10.9% of the time in *Concentrated spillover effects*. This is because Andrews-type tests rely on variation across time periods to deliver valid inference, and may experience over-rejection when it observes insufficient variation.

6.3 Test for existence of spillover effects

In this section, we examine the power of the proposed test against the null hypothesis that there are no spillover effects. We also look into its behavior when the range of the spillover effect is not correctly specified. In this set of experiments, the level of spillover effects varies from 0 to 2, corresponding to the strength of alternative hypotheses. We set $(N, T) = (20, 50)$ and $\alpha_1 = 5$. There are 9 units that are affected by spillover effects. Other settings follow exactly as in Section 6.1.1 (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 4.2 are plotted in Figure 6. Here *Include too few* misses half of the units that are actually affected by the treatment (assuming that unit 1 as well as four 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, five of which are actually not affected by spillover effects.

Among the three cases, *Include too many* is still a correct specification but is

Table 4: Empirical rejection rate of testing for treatment effects under $H_1 : \alpha_1 \neq 0$.

	$N = 10$			$N = 30$			$N = 50$		
	$T = 15$	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
Andrews	0.797	0.948	0.926	0.785	0.901	0.983	0.797	0.972	0.827
SP	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
Andrews	0.651	0.765	0.329	0.704	0.754	0.542	0.680	0.746	0.737
SP	0.860	0.932	0.991	0.957	0.918	0.967	0.834	0.816	0.853
Spread-out spillover effects									
Placebo	0.000	0.000	0.000	0.348	0.378	0.331	0.305	0.255	0.294
Andrews	0.337	0.403	0.277	0.563	0.414	0.278	0.406	0.309	0.343
SP	0.866	0.978	0.981	0.969	0.950	0.991	0.909	0.985	0.974

Notes: SP is the estimation procedure proposed in this paper that takes spillover effects into account. *No spillover effects* stands for the cases where the true DGP has no spillover effects. *Concentrated spillover effects* is the case where 1/3 of the control units receive a spillover effect. *Spread-out spillover effects* is the case where 2/3 of the control units receive a spillover effect of the same level.

supposed to be more conservative, so it has less power than *Correct specification*. Note that the range of spillover effects for *Include too few* is correctly specified only when the level of spillover effects is zero. Its power curve is similar to that of *Include too many*.

7 Extensions

7.1 An estimator with a smaller variance

In this section, we show that it is possible to form an estimator of α with possibly lower variance other than $\widehat{\alpha}$ proposed in Section 3.2. The idea is to minimize $\|W^{1/2}\widehat{u}_{T+1}\|$ instead of $\|\widehat{u}_{T+1}\|$, where $W \in \mathbb{R}^N$ is some positive definite matrix and $\widehat{u}_{T+1} = (Y_{T+1} - \widehat{\alpha}) - \widehat{a} - \widehat{B}(Y_{T+1} - \widehat{\alpha})$. The resulting estimator, as a function of W, is

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

where $\widehat{M}_W = (I - \widehat{B})'W(I - \widehat{B})$. The corresponding estimator for α is $\widehat{\alpha}_W = A\widehat{\gamma}_W$. In the spirit of GMM with an efficient weighting matrix, let $\Omega = Cov[u_1]$ and W_T^e

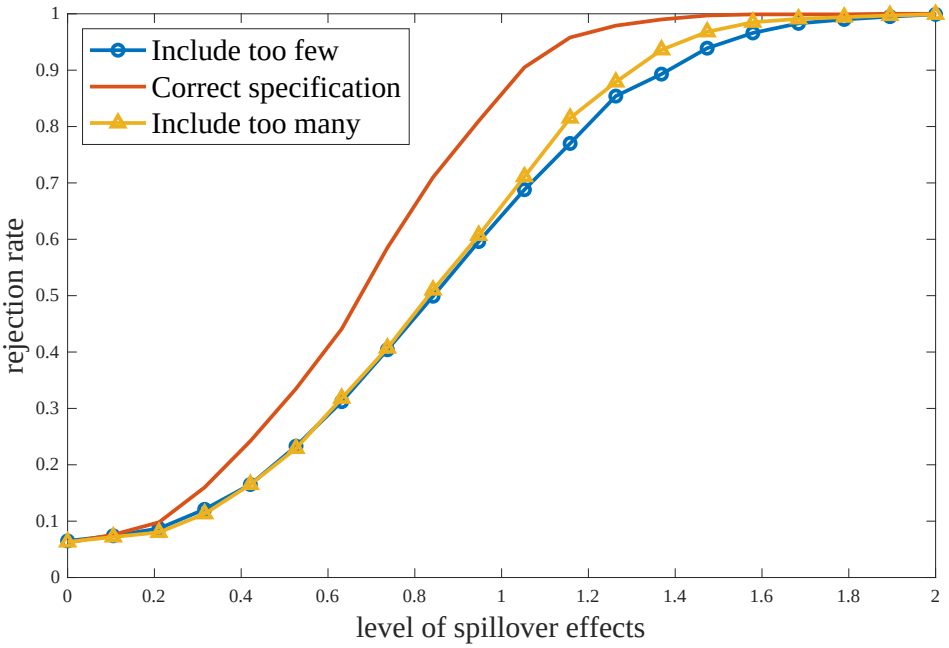


Figure 6: Empirical rejection rate of testing for the existence of spillover effects. There are 20 units in total, half of which are affected by the treatment. *Include too few* is assuming only 5 of them are affected by the treatment. *Correct specification* assumes the researcher knows exactly which set of units are affected. *Include too many* assumes 15 units are affected, 5 of which are, in fact, not affected.

be a consistent estimator of Ω^{-1}. Then an estimator of α with lower variance can be achieved by $\widehat{\alpha}^e = \widehat{\alpha}_{W_T^e}$.

Let $M_W = (I-B)'W(I-B)$, $G_W = A(A'M_WA)^{-1}A'(I-B)'W$ 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, $\widehat{\alpha}_{W_T} - (\alpha + G_W u_{T+1}) \to_p 0$, and specifically, $\widehat{\alpha}^e - (\alpha + G^e u_{T+1}) \to_p 0$, as $T \to \infty$. Moreover, $(Cov[G_W u_{T+1}] - Cov[G^e u_{T+1}])$ is positive semi-definite.*

Proposition 1 states that $\widehat{\alpha}^e$ always has a smaller asymptotic variance than $\widehat{\alpha}$. In practice, we need to estimate Ω, and for that we would need a relatively large sample size (large T) to have a good approximation.

7.2 Multiple treated units

Our method readily extends to cases where multiple units are treated. In our setting, the treatment and spillover effects can be estimated in the same way, since the spillover effects can be interpreted as the indirect treatment effect. With a corrected specified structure matrix A, we can perform estimation and inference 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

$$A = \begin{bmatrix} I_3 \\ 0_{1\times 3} \end{bmatrix},$$

and the resulting estimator $\widehat{\gamma} = (\widehat{\gamma}_1, \widehat{\gamma}_2, \widehat{\gamma}_3)'$ by (5) is such that $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$ are the treatment effect estimator for unit 1 and unit 2, respectively, and $\widehat{\gamma}_3$ is the spillover effect estimator for unit 3. Tests can be performed accordingly. If the researcher wants to test for the hypothesis that there are no spillover effects, the null is then $H_0 : C\alpha = d$, where $C = (0, 0, 1, 0)$ and $d = 0$.

7.3 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

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

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

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

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 introduced in previous sections.

To answer questions such as whether there is spillover effect at all, we can extend Andrews' instability test discussed above. Consider the null hypothesis $H_0 : C_t\alpha_t = d_t$ for $t = T + 1, \ldots, T + m$. Let $\widehat{P}_t$ be constructed as in Section 4.2 for $t = 1, \ldots, T$. For $t = T + 1, \ldots, T + m$, let $\widehat{P}_t = (C_t\widehat{\alpha}_t - d_t)'W_T(C_t\widehat{\alpha}_t - d_t)$. For $t = 1, \ldots, T + 1$, let

$$P^{(t)} = \sum_{s=0}^{m-1} \widehat{P}_{t+s},$$

each of which contains information from period t through $t+m-1$. The test statistic is then $P^{(T+1)}$, and we use its pre-treatment counterparts $\{P^{(t)}\}_{t=1}^{T}$ to form the null distribution.

7.4 Including covariates

Many empirical researchers are interested in including extra covariates when using synthetic control methods. Our framework can be combined with existing methods such as Abadie et al. (2010) and Li (2020), and be readily adapted to settings with covariates. For example, suppose we have a vector of observable variables $z_{i,t}$ and want to estimate the treatment effects, while being worried about spillover effects. Following Li (2020), we estimate the least square coefficients for the model

$$y_{i,t}(0) = a_i + \sum_{j \neq i} b_{i,j} y_{j,t}(0) + z_{i,t}'\pi + u_{i,t},$$

with the simplex constraints on $b_{i,j}$ and obtain coefficient estimates $(\widehat{a}_i, \widehat{b}_i, \widehat{\pi}_i)$. This is done for each i. Let $\widehat{g}_t = (z_{1,t}'\widehat{\pi}_i, \ldots, z_{N,t}'\widehat{\pi}_N)'$. Under appropriate regularity conditions, the results of the paper apply when the intercept estimator $\widehat{a}$ is replaced by $\widehat{a} + \widehat{g}_t$ at time t. The treatment effects estimator now becomes

$$\widehat{\gamma} = (A'\widehat{M}A)^{-1}A'(I - \widehat{B})'((I - \widehat{B})Y_{T+1} - \widehat{a} - \widehat{g}_{T+1}).$$

8 Conclusion

The synthetic control method is a powerful tool in treatment effect estimation in the panel data settings, but it may have a large bias in the presence of spillover effects. In this paper, we relax the SUTVA assumption and propose an estimation and testing procedure that is robust to the presence of spillover effects. Our method requires a specification of the spillover structure. We derive a set of conditions under which 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 under a set of conditions. We show that our conditions are satisfied by the commonly used factor models, with either stationary or cointegrated common factors. Our methods can be extended to cases with multiple treated units and multiple post-treatment periods and with extra covariates. Simulation results certify the validity of our estimation and testing procedure in the presence of spillover effects. The proposed procedure is illustrated through an application to Abadie et al. (2010)'s California

tobacco control program data, where we find evidence of spillover effects as well as the main treatment effect.

Appendix

Proof of Lemma 1. (i) First assume that $A'MA$ is non-singular and Condition ST holds. The proof follows Ferman and Pinto (2021), 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 (b). It suffices to show $|\widehat{a}_i - a_i| = o_p(1)$ and $\|\widehat{b}_i - b_i\| = o_p(1)$ 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

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

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 $\widehat{b}_1$ is numerically equivalent to $(0, \widehat{v}')'$, otherwise the minimization problem in forming $\widehat{a}_1$ and $\widehat{b}_1$ may have a lower objective evaluated at $(\bar{y}_1 - \sum_{j=2}^{N} \bar{y}_j \widehat{v}_j, 0, \widehat{v}')'$. Now we let $\widehat{Q}(v)$ denote the objective function such that

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

and its population analog be $Q(v) = (-1, v')\Omega_y(-1, v')'$. 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 an $(N-1)$-dimensional simplex. (iii) $Q(v)$ is continuous since it has a quadratic form.

(iv) To see uniform convergence, note

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

where $\|\cdot\|_F$ is the Frobenius norm. The second inequality is by ergodicity for the second moments. Therefore, $\widehat{v}\to_p v_0$. This implies $\|\widehat{b}_1-b_1\|=o_p(1)$. By ergodicity,

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

This shows part (b) 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.

Part (c) follows from part (b) and the stationarity of $\{Y_{T+1}(0)\}_{T\geq 1}$. Part (d) is assumed. Thus, Assumption 1 holds under the invertibility of $A'MA$ and Condition ST.

(ii) Now, we assume that $A'MA$ is non-singular and Condition CO holds.

We first show part (c). We will show $\|Y_{T+1}(0)'(\widehat{b}_1-b_1)\|=o_p(1)$ and other i's follows the same strategy. Since the synthetic control estimator can be written as a projection of the OLS estimator onto a closed convex set, we will first derive the asymptotic properties of the OLS estimator, and then use the properties of projections to obtain the desired results. For examples of this strategy, see Li (2020) and Yu et al. (2019). 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 $\theta_1,\theta_2\in\mathbb{R}^N$, $\langle\theta_1,\theta_2\rangle_D=\theta_1'D\theta_2$. 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 $\theta,\theta'\in\mathbb{R}^N$,

$$\|\Pi_D\theta-\Pi_D\theta'\|_D\leq\|\theta-\theta'\|_D. \tag{9}$$

With some abuse of notation, let $x_t=Y_t-T^{-1}\sum_{s=1}^{T}Y_s$. Then, $\widehat{b}_1$ 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 $\Sigma_T=T^{-1}\sum_{t=1}^{T}x_tx_t'$.

Appendix A.2 in Li (2020) establishes that $\widehat{b}_1 = \Pi_{\Sigma_T}\tilde{b}_1$. Thus, we have

$$\begin{aligned}
\|\widehat{b}_1 - b_1\| &= \|\Sigma_T^{-1/2}\Sigma_T^{1/2}(\widehat{b}_1 - b_1)\| \\
&\leq \|\Sigma_T^{-1/2}\|_F \cdot \|\Sigma_T^{1/2}(\widehat{b}_1 - b_1)\| \\
&= \|\Sigma_T^{-1/2}\|_F \cdot \|\widehat{b}_1 - b_1\|_{\Sigma_T} \\
&= \|\Sigma_T^{-1/2}\|_F \cdot \|\Pi_{\Sigma_T}\tilde{b}_1 - \Pi_{\Sigma_T}b_1\|_{\Sigma_T} \\
&\leq \|\Sigma_T^{-1/2}\|_F \cdot \|\tilde{b}_1 - b_1\|_{\Sigma_T} \\
&= \|\Sigma_T^{-1/2}\|_F \cdot \|\Sigma_T^{1/2}\|_F \cdot \|\tilde{b}_1 - b_1\| \\
&= O_p(1)o_p(T^{-1/2}) \\
&= o_p(T^{-1/2}), \qquad (10)
\end{aligned}$$

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 (9). To see the fifth equality, note

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

so

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

where the second equality is standard results for $\mathcal{I}_1$ process (see Hamilton, 1994, part (g) and (i) of Proposition 18.1). Also, $\|\tilde{b}_1 - b_1\| = o_p(T^{-1/2})$ is by Proposition 19.2 in Hamilton (1994). This shows (10). Apply part (a) of Proposition 18.1 in Hamilton (1994), we have

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

Now we show part (b). Again, it suffices to show $|\widehat{a}_i - a_i| = o_p(1)$ and $\|\widehat{b}_i - b_i\| = o_p(1)$. We consider the $i = 1$ case and other cases follow the same strategy. We have showed $\|\widehat{b}_i - b_i\| = o_p(1)$ in part (c) of the proof. Section A.6.1 in Ferman and Pinto (2021) establishes that

$$[\mu_1^1 \ \mu_2^1 \ \ldots \ \mu_N^1](b_1 - e_1) = 0, \qquad (11)$$

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

$$\begin{aligned}
\widehat{a}_1 &= [\bar{y}_1\ \bar{y}_2\ \dots\ \bar{y}_N](e_1 - \widehat{b}_1) \\
&= [\bar{y}_1\ \bar{y}_2\ \dots\ \bar{y}_N](e_1 - b_1) + [\bar{y}_1\ \bar{y}_2\ \dots\ \bar{y}_N](b_1 - \widehat{b}_1) \\
&= \left\{ \frac{1}{T}\sum_{t=1}^{T} \left((\lambda_t^0)'[\mu_1^0\ \dots\ \mu_N^0] + [\varepsilon_{1,t}\ \dots\ \varepsilon_{N,t}] \right) \right\} (e_1 - b_1) + \\
&\quad \left(\frac{1}{\sqrt{T}}[\bar{y}_1\ \bar{y}_2\ \dots\ \bar{y}_N] \right) \sqrt{T}(b_1 - \widehat{b}_1) \\
&= E[\lambda_t^0]'[\mu_1^0\ \dots\ \mu_N^0](e_1 - b_1) + o_p(1) + O_p(1)o_p(1) \\
&\to_p E[\lambda_t^0]'[\mu_1^0\ \dots\ \mu_N^0](e_1 - b_1). \\
&= a_1. \qquad (12)
\end{aligned}$$

The third equality is by (11). The fourth equality is by stationarity of $\{(\lambda_t^0, \varepsilon_t)\}_{t\geq 1}$ and results in part (d) of the proof. This shows part (b) of the Assumption 1 .

Combining (11) and (12), we have part (a) in Assumption 1. Part (d) is assumed.

□

Proof of Theorem 1. Using formula of $\widehat{\gamma}$ in Equation (5), we have

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

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 (b) and (c) in Assumption 1. Therefore,

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

The third equality is by (b) 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

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

and

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

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 Hayashi, 2000, Proposition 3.5). 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 (2006). Let

$$L_{1,T}(\epsilon) = \left\{ \|C_T(\widehat{\beta}_1 - \beta_1)\| \leq \epsilon, \max_{t=1,\ldots,T} \|C_T(\widehat{\beta}_1^{(t)} - \beta_1)\| \leq \epsilon \right\},$$
$$L_{2,T}(c) = \left\{ \max_{t \leq T+1} \|C_T^{-1} x_t\| \leq c \right\}.$$

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 $\beta = \widehat{\beta}_1$ or $\beta = \widehat{\beta}_1^{(t)}$ for some $t = 1, \ldots, T$, we have

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

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,

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

To see the last equality, pick $\epsilon > 0$. By continuity, $\exists \delta > 0$ such that for each $y \in$

$(x-\delta, x+\delta)$, $|\Pr(P_{T+1}(\beta_1) \le y) - \Pr(P_{T+1}(\beta_1) \le x)| < \epsilon$. Therefore,

$$\begin{aligned}
& \Pr(P_{T+1}(\widehat{\beta}_1) \le x + g_t(\epsilon_T, c_T)) \\
= \; & \Pr(\{P_{T+1}(\widehat{\beta}_1) \le x + g_t(\epsilon_T, c_T)\} \cap \{|g_t(\epsilon_T, c_T)| \ge \delta\}) \\
& + \Pr(\{P_{T+1}(\widehat{\beta}_1) \le x + g_t(\epsilon_T, c_T)\} \cap \{|g_t(\epsilon_T, c_T)| < \delta\}) \\
\le \; & \Pr(|g_t(\epsilon_T, c_T)| \ge \delta) + \Pr(P_{T+1}(\widehat{\beta}_1) \le y) \\
< \; & \Pr(P_{T+1}(\beta_1) \le x) + o(1).
\end{aligned}$$

Similarly, $\Pr(P_{T+1}(\widehat{\beta}_1) \le x) \ge \Pr(P_{T+1}(\beta_1) \le x) + o(1)$. This shows part (a).

To 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 $x \le 0$, $k(x) \in [0,1]$ for $x \in (0,1)$, and $k(x) = 0$ for $x \ge 1$. 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

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

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 $c_T = 1/\sqrt{\epsilon_T}$. Also, define,

$$\widehat{F}_T(x, \{\beta_1\}) = \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}\{P_t(\beta_1) \le x\},$$

i.e., $\widehat{F}_T(x, \{\beta_1\})$ is the empirical cdf of P_t as if the true parameter β_1 is known.

We write

$$|\widehat{F}_{P,T}(x) - F_P(x)| \le \sum_{i=1}^{4} D_{i,T},$$

for

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

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

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

Thus, for any $\delta > 0$,

$$\begin{aligned}\Pr(D_{1,T} > \delta) &\leq \Pr(\{D_{1,T} > \delta\} \cap L_T) + \Pr(L_T^c) \\ &\leq \Pr\left(\frac{1}{T}\sum_{t=1}^{T} \mathbb{1}\left\{P_t(\widehat{\beta}_1^{(t)}) - x \in (-g_t(\epsilon_T, c_T), h_T + g_t(\epsilon_T, c_T)\right\} > \delta\right) + o(1) \\ &\leq \frac{E\mathbb{1}\left\{P_t(\widehat{\beta}_1^{(t)}) - x \in (-g_t(\epsilon_T, c_T), h_T + g_t(\epsilon_T, c_T)\right\}}{\delta} + o(1), \qquad (13)\end{aligned}$$

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, (13) implies $\Pr(D_{1,T} > \delta) \leq o(1)$ and thus $D_{1,T} = o_p(1)$.

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

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

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

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

Therefore,

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

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 $\widehat{F}_{P,T}(x) \to_p F_P(x)$ for $x \in$

$(q_{P,1-\tau} - \epsilon, q_{P,1-\tau} + \epsilon)$. Note

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

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

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

Again, the inequality is by definition of $\widehat{q}_{P,1-\tau}$, 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 $q_{P,1-\tau}$ is $(1-\tau)$-quantile of P_∞. Therefore,

$$\Pr(P > \widehat{q}_{P,1-\tau}) = 1 - \Pr(P \leq \widehat{q}_{P,1-\tau}) = 1 - \Pr(P + (q_{P,1-\tau} - \widehat{q}_{P,1-\tau}) \leq q_{P,1-\tau}) \to \tau,$$

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 $\eta_t, \lambda_t, \varepsilon_t$.

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

$$D_T = \begin{bmatrix} 1 & 0 \\ 0 & T^\tau I_N \end{bmatrix}. \tag{14}$$

Then, we have

$$\max_{t \leq T+1} \|D_T^{-1} x_t\| = \max_{t \leq T+1} \left\| \begin{bmatrix} 1 \\ T^{-\tau} Y_t \end{bmatrix} \right\| = \sqrt{1 + \left(\max_{t \leq T+1} \|T^{-\tau} Y_t\| \right)^2}. \tag{15}$$

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

$$\begin{aligned}
\Pr\left(\max_{t\leq T+1} \|T^{-\tau}Y_t\| > \epsilon\right) &= \Pr\left(\bigcup_{t\leq T+1} \|Y_t\| > T^{\tau}\epsilon\right) \\
&\leq \left(\sum_{t=1}^{T} \Pr(\|Y_t\| > T^{\tau}\epsilon)\right) + \Pr(\|Y_{T+1}(0) + \alpha\| > T^{\tau}\epsilon) \\
&= \frac{TE[\|Y_t\|^{2+\delta}]}{T^{\tau(2+\delta)}\epsilon^{2+\delta}} + o(1) \\
&= o(1). \qquad (16)
\end{aligned}$$

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 (15) and (16), we obtain part (c).

For part (d), we use D_T defined in (14). Following the same reasoning as in (10), for each $i = 1, \ldots, N$, we have

$$\begin{aligned}
\|\widehat{b}_i - b_i\| &\leq \|\Sigma_T^{-1/2}\|_F \cdot \|\Sigma_T^{1/2}\|_F \cdot \|\tilde{b}_i - b_i\| \\
&= O_p(1)O_p(T^{-1/2}) \\
&= O_p(T^{-1/2}). \qquad (17)
\end{aligned}$$

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

$$\begin{aligned}
\|D_T(\widehat{\beta}_i - \beta_i)\| &= \left\| \begin{bmatrix} 1 & 0 \\ 0 & T^{\tau-1/2}I_N \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & T^{1/2}I_N \end{bmatrix} (\widehat{\beta}_i - \beta_i) \right\| \\
&\leq \left\| \begin{bmatrix} 1 & 0 \\ 0 & T^{\tau-1/2}I_N \end{bmatrix} \right\|_F \left\| \begin{bmatrix} \widehat{a}_i - a_i \\ \sqrt{T}(\widehat{b}_i - b_i) \end{bmatrix} \right\| \\
&= \sqrt{1 + NT^{2\tau-1}} \|O_p(1)\| \\
&= o_p(1).
\end{aligned}$$

The second equality is due to (17). The last equality is because $\tau < 1/2$. Therefore, $\|(\widehat{\theta} - \theta_0)D_T\|_F = \sqrt{\sum_{i=1}^{N} \|D_T(\widehat{\beta}_i - \beta_i)\|^2} = o_p(1)$. Also, since $\widehat{\theta}^{(t)} = \widehat{\theta}$ for each t, $\max_{t=1,\ldots,T} \|(\widehat{\theta}^{(t)} - \theta_0)D_T\|_F = \|(\widehat{\theta} - \theta_0)D_T\|_F = o_p(1)$. This shows part (d).

Part (e) is assumed.

Part (f) is trivial if $W_T = I$. Assume now $W_T = (C\widehat{G}(T^{-1}\sum_{t=1}^{T} \widehat{u}_t\widehat{u}_t')\widehat{G}'C')^{-1}$.

Then,

$$\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T}\widehat{u}_t\widehat{u}_t' \\
= \;& (I-\widehat{B})\left(\frac{1}{T}\sum_{t=1}^{T}Y_tY_t'\right)(I-\widehat{B})' - (I-\widehat{B})\left(\frac{1}{T}\sum_{t=1}^{T}Y_t\right)\widehat{a}' \\
& -\widehat{a}\left(\frac{1}{T}\sum_{t=1}^{T}Y_t'\right)(I-\widehat{B})' + \widehat{a}\widehat{a}' \\
\to \;& E[u_tu_t'],
\end{aligned}$$

by ergodicity and Assumption 1(b). Therefore, $\widehat{W}_T \to_p W = (CGE[u_tu_t']G'C')^{-1}$.

This concludes part (i) of Lemma 3.

(ii) Assume Condition CO holds.

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

By (11), 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

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

Then, we have

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

The second equality is because

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

by the continuous mapping theorem.

To show Part (d), we combine (10) and (12), and have

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

Therefore, $\|(\widehat{\theta} - \theta_0)D_T\|_F = \sqrt{\sum_{i=1}^N \|D_T(\widehat{\beta}_i - \beta_i)\|^2} = o_p(1)$. The second half of Part (d) is also satisfied since $\widehat{\theta}^{(t)} = \widehat{\theta}$ 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

$$\begin{aligned}
L_{1,T}(\epsilon) &= \left\{ \|(\widehat{\theta} - \theta_0)D_T\|_F \le \epsilon, \max_{t=1,\dots,T} \|(\widehat{\theta}^{(t)} - \theta_0)D_T\|_F \le \epsilon \right\}, \\
L_{2,T}(c) &= \left\{ \max_{t \le T+1} \|D_T^{-1} x_t\| \le c \right\}, \\
L_{3,T}(\eta) &= \left\{ \|\widehat{G}' C' W_T C \widehat{G} - G' C' W C G\|_F < \eta \right\}.
\end{aligned}$$

By Assumption 3(d), there exists a positive sequence $\{\epsilon_T\}_{T \ge 1}$ 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 $\Pr(L_{2,T}(c_T)) \to 1$. By Assumption 1(c) and Assumption 2(f), there exists a positive sequence $\{\eta_T\}_{T \ge 1}$ such that $\eta_T \to 0$ and $\Pr(L_{3,T}(\eta_T)) \to 1$. Let $L_T = L_{1,T}(\epsilon_T) \cap L_{2,T}(c_T) \cap L_{3,T}(\eta_T)$, then we have $\Pr(L_T) \to 1$ and $\Pr(L_T^c) \to 0$.

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

$$|\widehat{P}_t(\theta) - P_t(\theta_0)| \le |\widehat{P}_t(\theta) - P_t(\theta)| + |P_t(\theta) - P_t(\theta_0)|. \tag{18}$$

Note that

$$\begin{aligned}
|\widehat{P}_t(\theta) - P_t(\theta)| &= \left| (Y_t - \theta x_t)' (\widehat{G}' C' W_T C \widehat{G}) - G' C' W C G)(Y_t - \theta x_t) \right| \\
&\le \|Y_t - \theta x_t\|^2 \|(\widehat{G}' C' W_T C \widehat{G} - G' C' W C G\|_F \\
&\le \|u_t + (\theta_0 - \theta) x_t\|^2 \cdot \eta_T \\
&\le (\|u_t\| + \|(\theta_0 - \theta) D_T D_T^{-1} x_t\|)^2 \eta_T \\
&\le (\|u_t\| + \|(\theta_0 - \theta) D_T\|_F \|D_T^{-1} x_t\|)^2 \eta_T \\
&\le (\|u_t\| + \epsilon_T c_T)^2 \eta_T
\end{aligned} \tag{19}$$

and

$$
\begin{aligned}
|P_t(\theta) - P_t(\theta_0)| &= |(Y_t - \theta x_t)'G'C'WCG(Y_t - \theta x_t) - (Y_t - \theta_0 x_t)'G'C'WCG(Y_t - \theta_0 x_t)| \\
&\leq |(Y_t - \theta x_t)'G'C'WCG(Y_t - \theta x_t) - (Y_t - \theta x_t)'G'C'WCG(Y_t - \theta_0 x_t)| \\
&\quad + |(Y_t - \theta x_t)'G'C'WCG(Y_t - \theta_0 x_t) - (Y_t - \theta_0 x_t)'G'C'WCG(Y_t - \theta_0 x_t)| \\
&= |(u_t + (\theta_0 - \theta)x_t)'G'C'WCG(\theta_0 - \theta)x_t| + |((\theta_0 - \theta)x_t)'G'C'WCGu_t| \\
&\leq \|u_t + (\theta_0 - \theta)D_T D_T^{-1} x_t\| \|G'C'WCG\|_F \|(\theta_0 - \theta)D_T D_T^{-1} x_t\| \\
&\quad + \|(\theta_0 - \theta)D_T D_T^{-1} x_t\| \|G'C'WCG\|_F \|u_t\| \\
&\leq (\|u_t\| + \epsilon_T c_T)\|G'C'WCG\|_F \epsilon_T c_t + \epsilon_T c_T \|G'C'WCG\|_F \|u_t\| \\
&= (2\|u_t\| + \epsilon_T c_T)\|G'C'WCG\|_F \epsilon_T c_t. \qquad (20)
\end{aligned}
$$

Combining (18), (19), and (20), we have $|\widehat{P}_t(\theta) - P_t(\theta_0)| \leq g(\epsilon_T, c_T, \eta_T)$, where $g_t(\epsilon_T, c_T, \eta_T) = (\|u_t\| + \epsilon_T c_T)^2 \eta_T + (2\|u_t\| + \epsilon_T c_T)\|G'C'WCG\|_F \epsilon_T c_t$. 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,

$$
\begin{aligned}
P &= (C\widehat{\alpha} - d)'W_T(C\widehat{\alpha} - d) \\
&= (C(\alpha + Gu_{T+1} + o_p(1)) - d)'(W + o_p(1))(C(\alpha + Gu_{T+1} + o_p(1)) - d) \\
&= (CGu_{T+1} + o_p(1))'(W + o_p(1))(CGu_{T+1} + o_p(1)) \\
&= u'_{T+1}G'C'WCGu_{T+1} + o_p(1).
\end{aligned}
$$

The second equality is by Theorem 1. Since $P_\infty = u'_1 G'C'WCGu_1$, 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

Abadie, A. and Cattaneo, M. D. (2018). Econometric methods for program evaluation. *Annual Review of Economics*, 10:465–503.

Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of californias tobacco control program. *Journal of the American Statistical Association*, 105:493–505.

Abadie, A. and Gardeazabal, J. (2003). The economic costs of conflict: A case study of the basque country. *American Economic Review*, 93:113–132.

Abadie, A. and LHour, J. (2021). A penalized synthetic control estimator for disaggregated data. *Journal of the American Statistical Association*, 116:1817–1834.

Andrews, D. W. K. (2003). End-of-sample instability tests. *Econometrica*, 71:1661–1694.

Andrews, D. W. K. and Kim, J. (2006). Tests for cointegration breakdown over a short time period. *Journal of Business & Economic Statistics*, 24:379–394.

Baltagi, B. H. and Griffin, J. M. (2001). The econometrics of rational addiction. *Journal of Business & Economic Statistics*, 19:449–454.

Basse, G. W., Feller, A., and Toulis, P. (2019). Randomization tests of causal effects under interference. *Biometrika*, 106:487–494.

Baumeister, R. F. (2017). Addiction, cigarette smoking, and voluntary control of action: Do cigarette smokers lose their free will? *Addictive Behaviors Reports*, 5:67.

Becker, G., Crossman, M., and Murphy, K. M. (1994). An empirical analysis of cigarette addiction. *American Economic Review*, 84:396–418.

Benowitz, N. L. (1992). Cigarette smoking and nicotine addiction. *Medical Clinics of North America*, 76:415–437.

Cavallo, E., Galiani, S., Noy, I., and Pantano, J. (2013). Catastrophic natural disasters and economic growth. *Review of Economics and Statistics*, 95:1549–1561.

Chernozhukov, V., Wthrich, K., and Zhu, Y. (2021). An exact and robust conformal inference method for counterfactual and synthetic controls. *Journal of the American Statistical Association*, 116:1849–1864.

Christelis, D. and de Galdeano, A. S. (2011). Smoking persistence across countries: A panel data analysis. *Journal of Health Economics*, 30:1077–1093.

Conley, T. G. and Taber, C. R. (2011). Inference with difference in differences with a small number of policy changes. *Review of Economics and Statistics*, 93:113–125.

Doudchenko, N. and Imbens, G. W. (2017). Balancing, regression, difference-in-differences and synthetic control methods: A synthesis. *arXiv preprint arXiv:1610.07748*.

Fang, Z. and Santos, A. (2019). Inference on directionally differentiable functions. *The Review of Economic Studies*, 86:377–412.

Ferman, B. and Pinto, C. (2017). Placebo tests for synthetic controls. Working paper.

Ferman, B. and Pinto, C. (2019). Inference in differences-in-differences with few treated groups and heteroskedasticity. *The Review of Economics and Statistics*, 101:452–467.

Ferman, B. and Pinto, C. (2021). Synthetic controls with imperfect pretreatment fit. *Quantitative Economics*, 12:1197–1221.

Firpo, S., Possebom, V., Sergio, F., and Vitor, P. (2018). Synthetic control method: Inference, sensitivity analysis and confidence sets. *Journal of Causal Inference*, 6:1–26.

Grossi, G., Mariani, M., Mattei, A., Lattarulo, P., and Öner, Ö. (2020). Synthetic control group methods in the presence of interference: The direct and spillover effects of light rail on neighborhood retail activity. Working paper.

Hahn, J. and Shi, R. (2017). Synthetic control and inference. *Econometrics*, 5:52.

Hamilton, J. D. (1994). *Time series analysis*. Princeton University Press.

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

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

Labeaga, J. M. (1999). A double-hurdle rational addiction model with heterogeneity: Estimating the demand for tobacco. *Journal of Econometrics*, 93:49–72.

Li, K. T. (2020). Statistical inference for average treatment effects estimated by synthetic control methods. *Journal of the American Statistical Association*, 115:2068–2083.

Manresa, E. (2013). Estimating the structure of social interactions using panel data.

Miura, T. (2019). Does time preference affect smoking behavior? A dynamic panel analysis. *Journal of Behavioral and Experimental Economics*, 78:170–180.

Newey, W. K. and McFadden, D. (1994). Chapter 36 large sample estimation and hypothesis testing. *Handbook of Econometrics*, 4:2111–2245.

Robbins, M. W., Saunders, J., and Kilmer, B. (2017). A framework for synthetic control methods with high-dimensional, micro-level data: Evaluating a neighborhood-specific crime intervention. *Journal of the American Statistical Association*, 112:109–126.

Rosenbaum, P. R. (2007). Interference between units in randomized experiments. *Journal of the American Statistical Association*, 102:191–200.

Stefano, R. D. and Mellace, G. (2020). The inclusive synthetic control method. *SSRN Electronic Journal*.

Vazquez-Bare, G. (2022). Identification and estimation of spillover effects in randomized experiments. *Journal of Econometrics*.

Vleeming, W., Rambali, B., and Opperhuizen, A. (2002). The role of nitric oxide in cigarette smoking and nicotine addiction. *Nicotine and Tobacco Research*, 4:341–348.

Xu, Y. (2017). Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis*, 25:57–76.

Yu, M., Gupta, V., and Kolar, M. (2019). Constrained high dimensional statistical inference. *arXiv preprint arXiv:1911.07319*.

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