Causal Inference under Temporal and Spatial Interference

Ye Wang*

Abstract

Many social events and policies generate spillover effects in both time and space. Their occurrence influences not only the outcomes of interest in the future, but also these outcomes in nearby areas. In this paper, we propose a design-based approach to estimate the direct and indirect/spillover treatment effects of any event or policy under the assumption of sequential ignorability, when both temporal and spatial interference are present. The proposed estimators are shown to be consistent and normally distributed if the degree of interference dependence does not grow too fast relative to the sample size. The conventional difference-in-differences (DID) or two-way fixed effects model, nevertheless, leads to biased estimates in this scenario. We apply the method to examine the impact of Hong Kong's Umbrella Movement on the result of the ensuing election and how an institutional reform affects real estate assessment in New York State.

Keywords: *Causal inference, Design-based, Interference, Spillover effect, TSCS data, IPTW estimator, Stein's method, Difference-in-differences*

*Predoctoral researcher at the 21st Century China Center, UCSD and PhD Candidate at the Wilf Family Department of Politics, New York University. ye.wang@nyu.edu

1 Introduction

To understand the impact generated by social events or policies is the goal of many empirical studies in social sciences. In practice, researchers often assume that the impact is contemporary and resides within the unit of interest. In time-series cross-sectional (TSCS) data analysis, for example, the two-way fixed effects model is one of the most popular approaches. But its validity relies on assumptions such as the effect disappears once the treatment is lifted, and while some units (the treatment group) are affected by the treatment, others (the control group) are not (Imai and Kim, 2018). The credibility of these assumptions, nevertheless, is not always guaranteed. If one state cuts its corporation tax, the productivity of companies in neighboring states will also be influenced even several years later. Once a protest breaks out, plenty of people living nearby may get inspired and change their choice in following elections. Such a "spillover effect" can travel further as time goes by. Statistically speaking, the treatment imposed on unit i at period t can affect the outcome of unit j at period s. In other words, we have interference both over time and across units. As the units of interest are often located on a geographic space, we call the former temporal interference and the latter spatial interference.

In this paper, we investigate the problem of causal identification in TSCS data when both types of interference are present. There have been attempts to deal with either temporal or spatial interference in the literature of causal inference (Hudgens and Halloran, 2008; Aronow and Samii, 2017; Blackwell and Glynn, 2018). But most existent methods impose stringent restrictions on the structure of interference— for example, spillover only occurs among adjacent units— that are neither practical nor testable. Progress has been made in recent years to account for general interference under which arbitrary spillover is allowed to exist. Yet these works concentrate on experimental designs rather than the analysis in observational studies (Savje, Aronow and Hudgens, 2018; Aronow, Samii and Wang, 2019; Chin, 2019), and none of them considers scenarios where temporal and spatial interference coexist.

Drawing connections between the literature on TSCS data analysis and experimental design under interference, this paper presents a comprehensive discussion on what causal conclusions could be reached under general interference and common assumptions on TSCS datasets. It takes a design-based perspective and requires little structural restrictions on either the heterogeneity of treatment effects or the functional form of spillover. We first introduce a series of causal quantities based on the idea of marginalization. Generalizing from "circle means" defined in Aronow, Samii and Wang (2019), we propose that researchers can construct theory-driven "spillover mapping" to aggregate each unit's influence on the others into meaningful estimands. Essentially, these estimands describe the expected difference caused by the change of treatment assignment history at a representative unit on itself or its neighbors within any particular time period.

We then show that the defined estimands can be non-parametrically identified and consistently estimated via semi-parametric inverse probability of treatment weighting (IPTW) estimators under the assumption of sequential ignorability. The assumption states that the treatment assignment at period t is independent to the potential outcomes in the same period after conditioning on the values of treatments and outcomes in history. Estimation proceeds by first predicting the propensity score for each observation (i, t)[1], and then plugging the predicted values into a Horvitz-Thompson or Hajek estimator with the spillover mapping's value as the outcome. The estimators can be augmented by employing diffusion models to predict the outcome values with a higher accuracy.

We further deploy Stein's method to prove that the proposed estimators are asymptotically normal when the degree of interference dependence does not grow too fast as the sample size increases (Stein, 1972; Ross, 2011; Chatterjee, 2014; Ogburn et al., 2020). Valid confidence intervals could be obtained either analytically or via Fisher's randomization test (FRT). We demonstrate that the spatial/two-way HAC variance estimator (Conley, 1999; Cameron, Gelbach and Miller, 2012) offers asymptotically valid estimates for the

[1]It can be done via statistical models (Hirano, Imbens and Ridder, 2003; Ertefaie, Hejazi and van der Laan, 2020) or covariate balancing techniques (Imai and Ratkovic, 2015; Kallus and Santacatterina, 2018)

analytic variances. If unobservable attributes ("fixed effects") affect both the treatment and the outcome, however, the introduced estimands are identifiable only under strong (and often implausible) assumptions on the structure of spillover. Consequently, estimates from conventional DID or the two-way fixed effects model converge to a weighted sum of individualistic effects that has no substantive interpretation. The bias of estimation does not diminish even as the sample size grows to infinity.

The framework could be extended to incorporate various possibilities. First, researchers can account for the contagion of treatment by revising the propensity score model accordingly. The estimators preserve their properties as long as the model is correctly specified. Second, after adding more restrictions on the outcomes, we can also estimate the lag effect of treatments in previous periods, as in Blackwell and Glynn (2018). Finally, we can replace the geographic space with a social network, and measure the distance between units with the length of the path that connects them. We demonstrate that such a setting requires stronger assumptions as the "small world" phenomenon tends to make the degree of interference dependence grow too rapidly.

The proposed method works well in both simulation and real-world examples. We apply it to re-analyze two empirical studies. The first one is Wang and Wong (2019), in which the authors investigate the impact of Hong Kong's 2014 Umbrella Movement on the opposition's vote share in the ensuing election. Under weaker assumptions, our result confirms the authors' original conclusion that the protest reduced the opposition's vote share in constituencies that were close to the protest sites. The second study is Sances (2016), an examination of how the change of selection mechanism reshapes the incentives of real estate assessors in New York State. Our estimates of effects caused by temporal interference are consistent with the original findings, but we are able to show that the estimates are unlikely to be contaminated by spatial interference.

The rest of the paper is organized as follows: Section 2 summarizes previous studies that are related to the current work. Section 3 describes the basic framework, including causal quantities of interest and identification assumptions. Section 4 discusses the identification and estimation of the estimands under the assumption of sequential ignorability. Section 5 discusses the same issue under the alternative assumption of strict exogeneity. Section 6 analyzes statistical properties of the proposed estimators. Section 7 explores potential extensions. Section 8 presents simulation results. Section 9 illustrates the application of the method with empirical studies and section 10 concludes.

2 Related studies

Interference, also known as "spillover effect," "peer effect," or "diffusion" in social sciences, refers to the phenomenon that one observation's outcome can be influenced by the treatment of other observations. In the cross-sectional setting, it means that unit i's outcome is dependent on another unit j's treatment. When the time dimension exists, it could also be that unit i's outcome at period t relies on its own treatment status at period s. In either case, the classic stable unit treatment value assumption (SUTVA) in causal inference is violated. For simplicity, we use spatial interference or between-unit interference to call the former, and temporal interference or within-unit interference to call the latter. We distinguish interference from contagion, which means that one observation's outcome is affected by the outcome of other observations (Ogburn and VanderWeele, 2017).

This paper is built upon the recently burgeoning literature on experimental design under interference. Most early works in this field explicitly specify the structure of interference. For example, Hudgens and Halloran (2008) proposed the split-plot design under "partial interference" in which they allow for arbitrary interference within pre-specified strata but none between the strata. They also distinguished the treatment's direct effect from the indirect effect caused by interference and developed a variance estimator under homogeneity. This assumption has been adopted by a series of following studies (Tchetgen-Tchetgen and VanderWeele, 2010; Sinclair, McConnell and Green, 2012; Liu and Hudgens, 2014; Baird et al., 2016; Basse and Feller, 2018). In the scenario with social networks, Aronow and Samii (2017) suggested an IPTW estimator under the assumption that the "exposure mapping" from assignment to actual exposure is known to the researcher. Such an approach has also been extended by later works (Paluck, Shepherd and Aronow, 2016; Ogburn et al., 2020; An and VanderWeele, 2019; Leung, 2020).

Realizing that these structural assumptions are unrealistic, more scholars are now working under the assumption of general interference, in which one's outcome may be affected by the treatment assignment of all the units in arbitrary ways. Eckles, Karrer and Ugander (2017) argued that clustering design could be used to reduce the bias with unknown interference. Savje, Aronow and Hudgens (2018) and Li and Wager (2020) claimed that the difference-in-means estimator converges to the expected average treatment effect (EATE) when the degree of interference dependence satisfies certain conditions under any experimental design. Chin (2019) showed that the same estimator is normally distributed in large sample. Closest to the current work, Aronow, Samii and Wang (2019) investigated causal inference under general interference in field experiments. They constructed a spatial estimator that has a similar form as the difference-in-means estimator, with the outcome value of each subject replaced by its "circle mean," the average outcome value of subjects that are d distance units away from it. They showed that such an estimator is unbiased and consistent for the corresponding EATE. Nevertheless, they focused only on experimental designs with a bipartite structure where treatment assignment and outcome measure occur at two different levels. The current paper generalizes their idea to one-partite TSCS datasets in observational studies, which are more common in social sciences.

The model-based approach is popular in observational studies when researchers worry about between-unit interference. Common choices include linear and generalized linear models (Manski, 1995; Bramoullé, Djebbari and Fortin, 2009; Manski, 2012; Bowers, Fredrickson and Panagopolous, 2013; Goldsmith-Pinkham and Imbens, 2013; Blume et al., 2015), spatial econometric models (Beck, Gleditsch and Beardsley, 2006; LeSage and Pace, 2009), and network models (Graham, 2008; Acemoglu, García-Jimeno and Robinson, 2015; Leung, 2020). The fundamental problem of the model-based approach is that social interactions can be too complicated to be captured by any single model, as argued by Angrist (2014). Liu, Hudgens and Becker-Dreps (2016) are one of the pioneers to adopt the design-based perspective to analyze general interference in observational studies. The authors suggested a Hajek estimator and proved its consistency. Zigler and Papadogeorgou (2018) discussed the estimation of indirect effect in bipartite designs under similar assumptions. Recently, this approach has been further advanced by van der Laan (2014), Ogburn et al. (2020), and Papadogeorgou et al. (2020). The first two papers considered treatment assignment that satisfies sequential ignorability in dynamic social networks. They showed that causal identification is possible if the true network formation process is known—a prerequisite that is often too strong for practitioners. The third paper analyzed the direct effect of a "stochastic intervention strategy" in time series. The proposed estimator is robust to general interference. But none of the papers emphasized the estimation of indirect effects. Neither did they discuss identification under different assumptions.

Techniques to handle temporal interference are developed almost independently in the literature. Motivated by dynamic experiments[2] in medicine, biostatisticians have been studying cases where one's outcome is jointly decided by treatment assignments in the history since Robins (1986). Under the assumption of sequential ignorability, Robins, Hernan and Brumback (2000) proposed the marginal structural models (MSM), in which researchers can exploit IPTW estimators to balance the probability for a particular history to appear and obtain unbiased estimates of the history's cumulative effect. Bang and Robins (2005) further developed a doubly robust estimator for MSM. These methods have been introduced into social sciences by Blackwell (2013) and Blackwell and Glynn (2018). However, model-based approaches, especially the two-way fixed effects model (DID included), are still the first choice for social scientists to analyze TSCS data. In

[2] In a dynamic experiment, treatment in the next stage will be determined by the treatment and the outcome up to now

contrast to sequential ignorability, these approaches typically assume that all error terms are "strictly exogenous" to the treatment assignment at any period after conditioning on some unobservable factors (e.g., fixed effects). They then generate causal estimates by adjusting the response surface (outcome model) to remove the influence of these factors.

But as pointed out by Imai and Kim (2018), the two-way fixed effects model works only when the outcome is not affected by the past history ("no carryover" and "no feedback"), thus does not allow for within-unit interference. Otherwise, the estimates will be biased. Strezhnev (2018) and De Chaisemartin and d'Haultfoeuille (2020) further demonstrated that the two-way fixed effects estimator could be decomposed into a weighted sum of a series of DID estimators. Under temporal interference, the parallel trends assumption is violated for DID estimators in post-treatment periods, which causes the bias. The only exception is data that have a generalized DID or staggered adoption structure (Athey and Imbens, 2018). Then, the cumulative treatment effect on the treated (ATT) can be estimated through the idea of counterfactual estimation (Xu, 2017; Athey et al., 2018; Liu, Wang and Xu, 2019), where the model is only used to impute the missing counterfactual outcomes. But none of these methods allows for or ever discusses between-unit interference.

This paper aims at building a bridge between these separate branches of literature and focuses on TSCS data analysis in observational studies when both temporal and spatial interference are likely to occur. Such a situation is common when we are interested in the effect of policies or events, yet has not been paid much attention to by either methodologists or statisticians. In the following section, we will introduce estimands that describe causal effects of interest and clarify assumptions that are needed for their identification.

3 The framework

3.1 Set up

We work within the potential outcome framework (Neyman, 1923; Rubin, 1974). Suppose there are N contiguous units located on a surface[3], $\mathcal{X}$. Denote the geographic location of unit $i \in \{1, 2, \ldots, N\}$ as $\mathbf{x}_i = (x_{i1}, x_{i2}) \in \mathcal{X}$ and the time-invariant distance matrix among the N units as $\mathbf{D} = \{d_{ij}\}_{N \times N}$. Each of the units is observed for T consecutive periods. In each period t, we know the outcome Y_{it} and the treatment status $Z_{it} \in \{0, 1\}$ for each unit i. We use uppercase letters to represent random variables (e.g. Y_{it}), lowercase letters to represent their specific value (e.g. y_{it}), and boldface ones for corresponding vectors (e.g. $\mathbf{y}_t = (y_{1t}, y_{2t}, \ldots, y_{Nt})$).

For simplicity, we denote a random variable or its value in a particular period with subscript and their past history with superscript. For example, $\mathbf{Y}_i^{1:t} = (Y_{i1}, Y_{i2}, \ldots, Y_{it})$ and $\mathbf{Z}_i^{1:t} = (Z_{i1}, Z_{i2}, \ldots, Z_{it})$ are i's outcome values and treatment status up to period t, respectively; $\mathbf{Z}^{t:s} = (Z_{1t}, \ldots, Z_{Nt}, \ldots, Z_{1s}, \ldots, Z_{Ns})$ is the history of treatment assignment for all the units between period t and period s. Moreover, we use $\mathbf{Z}^{t:s} \setminus \mathbf{Z}_i^{t':s'}$ to denote the same history without unit i's treatment status between period t' and period s'.

General interference is allowed in this setting. The outcome for each unit i at period t, Y_{it}, is jointly decided by the entire history of treatment assignment of all the N units: $Y_{it} = Y_{it}(\mathbf{Z}^{1:T}) = Y_{it}(\mathbf{Z}_1^{1:T}, \mathbf{Z}_2^{1:T}, \ldots, \mathbf{Z}_N^{1:T})$. Thus, there are $2^{N \times T}$ different possible values for each Y_{it}. Notice that when there is no interference, $Y_{it} = Y_{it}(Z_{it})$ and SUTVA holds. With only temporal interference, $Y_{it} = Y_{it}(\mathbf{Z}_i) = Y_{it}(Z_{i1}, Z_{i2}, \ldots, Z_{iT})$; while with only spatial interference, $Y_{it} = Y_{it}(\mathbf{Z}_t) = Y_{it}(Z_{1t}, Z_{2t}, \ldots, Z_{Nt})$.

[3] In general, the surface can be a finite-dimensional metric space. But we only consider the case with two dimensions for simplicity.

3.2 Define causal quantities

We are interested in the causal effect generated by the change of a treatment assignment history. For each unit i in period t, we define the individualistic treatment effect of history $\mathbf{z}^{1:T}$ relative to history $\tilde{\mathbf{z}}^{1:T}$ as:

$$\tau_{it}(\mathbf{z}^{1:T}, \tilde{\mathbf{z}}^{1:T}) = Y_{it}(\mathbf{z}^{1:T}) - Y_{it}(\tilde{\mathbf{z}}^{1:T}).$$

Clearly, there are plenty of possible effects for each observation, not all of which are of substantive interest for researchers. Therefore, we focus on several quantities with particular forms in this paper. The first one is the individualistic contemporary direct effect: $\tau_{it}(1, 0; \mathbf{Z}^{1:T} \setminus Z_{it}) = Y_{it}(1, \mathbf{Z}^{1:T} \setminus Z_{it}) - Y_{it}(0, \mathbf{Z}^{1:T} \setminus Z_{it})$. It captures the effect of observation (i, t)'s treatment, Z_{it}, on its own outcome, Y_{it}. Next one is the individualistic cumulative direct effect: $\tau_{it}(\mathbf{z}^{s:t}, \tilde{\mathbf{z}}^{s:t}; \mathbf{Z}^{1:T} \setminus \mathbf{Z}_i^{s:t}) = Y_{it}(\mathbf{z}^{s:t}, \mathbf{Z}^{1:T} \setminus \mathbf{Z}_i^{s:t}) - Y_{it}(\tilde{\mathbf{z}}^{s:t}, \mathbf{Z}^{1:T} \setminus \mathbf{Z}_i^{s:t})$, where $s \leq t$. This is the cumulative effect of unit i's history between period s and t on its own outcome in period t and driven by temporal interference.

Then, we define the individualistic contemporary indirect effect as $\tau_{it}(1, 0; \mathbf{Z}^{1:T} \setminus Z_{jt}) = Y_{it}(1, \mathbf{Z}^{1:T} \setminus Z_{jt}) - Y_{it}(0, \mathbf{Z}^{1:T} \setminus Z_{jt})$. This effect reflects how observation (j, t)'s treatment influences the outcome of unit i in the same period and it is non-zero only under spatial interference. Finally, we use the individualistic cumulative indirect effect to describe the cumulative impact of unit j's history on unit i's outcome: $\tau_{it}(\mathbf{z}^{s:t}, \tilde{\mathbf{z}}^{s:t}; \mathbf{Z}^{1:T} \setminus \mathbf{Z}_j^{s:t}) = Y_{it}(\mathbf{z}^{s:t}, \mathbf{Z}^{1:T} \setminus \mathbf{Z}_j^{s:t}) - Y_{it}(\tilde{\mathbf{z}}^{s:t}, \mathbf{Z}^{1:T} \setminus \mathbf{Z}_j^{s:t})$. This quantity captures the influence of both types of interference.

It is easy to notice that all the quantities defined above are random variables since their values hinge on the treatment status of other observations. To fix their values and generate well-defined estimands, we follow the literature and marginalize these quantities over the distribution of treatment assignment of other observations. The marginalized contemporary direct effect, cumulative direct effect, contemporary indirect effect, and cumulative indirect effect are defined as:

$$\begin{aligned}
\tau_{it} &= \mathrm{E}_{\mathbf{Z}^{1:T} \setminus Z_{it}} \left[\tau_{it}(1, 0; \mathbf{Z}^{1:T} \setminus Z_{it}) \right] \\
\tau_{it}(\mathbf{z}^{s:t}, \tilde{\mathbf{z}}^{s:t}) &= \mathrm{E}_{\mathbf{Z}^{1:T} \setminus \mathbf{Z}_i^{s:t}} \left[\tau_{it}(\mathbf{z}^{s:t}, \tilde{\mathbf{z}}^{s:t}; \mathbf{Z}^{1:T} \setminus \mathbf{Z}_i^{s:t}) \right] \\
\tau_{it;j} &= \mathrm{E}_{\mathbf{Z}^{1:T} \setminus Z_{jt}} \left[\tau_{it}(1, 0; \mathbf{Z}^{1:T} \setminus Z_{jt}) \right] \\
\tau_{it;j}(\mathbf{z}^{s:t}, \tilde{\mathbf{z}}^{s:t}) &= \mathrm{E}_{\mathbf{Z}^{1:T} \setminus \mathbf{Z}_j^{s:t}} \left[\tau_{it}(\mathbf{z}^{s:t}, \tilde{\mathbf{z}}^{s:t}; \mathbf{Z}^{1:T} \setminus \mathbf{Z}_j^{s:t}) \right]
\end{aligned},$$

respectively[4].

Even after marginalization, none of these individualistic quantities is identifiable from data. We must further aggregate them in substantively meaningful ways. For the two direct effects, we simply take the average over the N units, and define the average contemporary direct effect and average cumulative direct effect in period t as:

$$\begin{aligned}
\tau_t &= \frac{1}{N} \sum_{i=1}^{N} \tau_{it} \\
\tau_t(\mathbf{z}^{s:t}, \tilde{\mathbf{z}}^{s:t}) &= \frac{1}{N} \sum_{i=1}^{N} \tau_{it}(\mathbf{z}^{s:t}, \tilde{\mathbf{z}}^{s:t})
\end{aligned}.$$

To construct average indirect effects, we introduce the concept of "spillover mapping," a linear mapping μ from $\mathcal{R}^N$ to $\mathcal{R}$ that captures the influence of a unit on the others. The value of μ for any unit i at period t is $\mu_i(\mathbf{x}) = \sum_{j=1}^{N} w_{ij} x_j$, where the time-invariant weights satisfy $w_{ij} \geq 0$ and $\sum_{j=1}^{N} w_{ij} = 1$. One example is the "circle mean" defined in Aronow, Samii and Wang (2019): $\mu_i(\mathbf{x}; d) = \frac{\sum_{j=1}^{N} \mathbf{1}\{d_{ji}=d\} x_j}{\sum_{j=1}^{N} \mathbf{1}\{d_{ji}=d\}}$ for any i. Here d is an arbitrary value of distance. It is a natural choice when the focus is how the spillover effects vary with distance. But there are other options, such as the average value of all the units within a range, $\mu_i(\mathbf{x}; d) = \frac{\sum_{j=1}^{N} \mathbf{1}\{d_{ji} \leq d\} x_j}{\sum_{j=1}^{N} \mathbf{1}\{d_{ji} \leq d\}}$, or the average value of all the first-degree neighbors of unit i. Researchers can construct different mappings by choosing the weights $\{w_{ij}\}_{j=1}^{N}$ according to theory or the study's purpose. We say Y_{jt} composes μ_i if $w_{ij} > 0$. We rely on the circle means $\mu(.; d)$ for illustration in what follows. But all the analytic results apply

[4]We write τ_{it} and $\tau_{it;j}$ rather than $\tau_{it}(1, 0)$ and $\tau_{it;j}(1, 0)$ for simplicity.

to general spillover mappings. $\mu(.;d)$ can be understood as a general spillover mapping that is indexed by d. Now, the average contemporary/cumulative indirect effect (for circle means) is defined as:

$$\tau_t(d) = \frac{1}{N}\sum_{i=1}^{N}\mu_i(\tau_{(1:N,t);i};d) = \frac{1}{N}\sum_{i=1}^{N}\frac{\sum_{j=1}^{N}\mathbf{1}\{d_{ji}=d\}\tau_{jt;i}}{\sum_{j=1}^{N}\mathbf{1}\{d_{ji}=d\}}$$
$$\tau_t(\mathbf{z}^{s:t},\tilde{\mathbf{z}}^{s:t};d) = \frac{1}{N}\sum_{i=1}^{N}\mu_i(\tau_{(1:N,t);i}(\mathbf{z}^{s:t},\tilde{\mathbf{z}}^{s:t});d) = \frac{1}{N}\sum_{i=1}^{N}\frac{\sum_{j=1}^{N}\mathbf{1}\{d_{ji}=d\}\tau_{jt;i}(\mathbf{z}^{s:t},\tilde{\mathbf{z}}^{s:t})}{\sum_{j=1}^{N}\mathbf{1}\{d_{ji}=d\}}$$

where $\tau_{(1:N,t);i} = (\tau_{1t,i},\tau_{2t,i},\ldots,\tau_{Nt,i})$ is the marginalized contemporary indirect effect generated by unit i on each of the N units.

These two estimands describe the expected difference generated by the change in a unit's treatment status/history on units that are d away from it in distance[5]. Notice that the average cumulative indirect effect includes all the other three effects as its special cases. For example, it equals to the average cumulative direct effect when $d = 0$ and equals to the average contemporary indirect effect when $s = t$. Hence, our formal analysis focuses on the average cumulative indirect effect. But the results generalize naturally to the other three estimands.

All these estimands are known as belonging to the family of expected average treatment effect (EATE) in the literature. They reflect the effect generated by the change of the key variable when the values of other relevant variables are fixed at their expectations. In what follows, we use EATE and the most general estimand, the average cumulative indirect effect, interchangeably. In practice, we often set the reference history $\tilde{\mathbf{z}}^{s:t}$ to be $0^{s:t}$. Then, the estimand measures the effect generated by $\mathbf{z}^{s:t}$ relative to the status quo on unit i between periods s and t[6].

[5]Clearly, the interpretation hinges on the choice of spillover mapping.
[6]We discuss alternative estimands of interest in the section on extensions.

3.3 Assumptions

Following the literature on TSCS data analysis, we make the following assumptions on both the treatment assignment process and the potential outcomes:

Assumption 1 (No reverse causality)**.** *If* $\mathbf{Z}^{1:t} = \tilde{\mathbf{Z}}^{1:t}$, *then*

$$Y_{it}(\mathbf{Z}^{1:T}) = Y_{it}(\tilde{\mathbf{Z}}^{1:T}).$$

for any i and t.

This assumption requires that the potential outcome of any unit i at period t is not affected by treatments assigned in the future. It will be violated if units anticipate the occurrence of the treatment in advance and adjust their behavior accordingly. When units have little agency in controlling the outcome, the assumption is more plausible. Under assumption 1, we can write $Y_{it}(\mathbf{Z}^{1:T})$ as $Y_{it}(\mathbf{Z}^{1:t})$.

Assumption 2 (Sequential ignorability)**.**

$$\mathbf{Z}_t \perp Y_{it}(\mathbf{Z}_t, \mathbf{Z}^{1:T} \setminus \mathbf{Z}_t) | \mathbf{Z}^{1:(t-1)}, \mathbf{Y}^{1:(t-1)}, \mathbf{X}^{1:t}$$

and

$$\mathbf{Z}_1 \perp Y_{i1}(\mathbf{Z}_1, \mathbf{Z}^{1:T} \setminus \mathbf{Z}_1) | \mathbf{X}^1$$

for any i and t.

This is the most crucial assumption for our formal analysis. It suggests that treatment assignment in period t, $\mathbf{Z}_t$, is "as-if" at random conditional on past treatment assignments, past outcomes, and covariates that are not affected by $\mathbf{Z}_t$. It is a variant of the "select on observables" assumption in cross-sectional studies and has been widely adopted in TSCS data analysis (Robins, Hernan and Brumback, 2000; Blackwell, 2013; Blackwell and Glynn, 2018). Such an assumption indicates that the dataset is generated by a hypothetical dynamic experiment and we can utilize the information contained in the history to estimate the propensity scores at period t, $P(\mathbf{Z}_t = \mathbf{z}_t|\mathbf{Z}^{1:(t-1)}, \mathbf{Y}^{1:(t-1)}, \mathbf{X}^{1:t})$, which play a key role in the identification of the estimands.

But unlike the assumption of strict exogeneity which the DID analysis or fixed effects models build upon, no unobservable confounder is allowed to exist under sequential ignorability. Therefore, if both the outcome and the assignment process are affected by some unobservable variables, the assumption will no longer hold. However, such a concern is partly alleviated in the spatial setting as we possess the geo-locations of the units. Suppose the unit fixed effects vary continuously on the surface $\mathcal{X}$, then their influence could be approximated by a smooth function of these coordinates. We will further discuss the tradeoff between this assumption and strict exogeneity in following sections. Actually, we will show that approaches that rely on strict exogeneity, such as DID or fixed effects models, no longer return meaningful or consistent estimates when spatial interference exists.

Theoretically speaking, Assumptions 1 and 2 are sufficient for the identification of our estimands. Yet in practice, more structural assumptions are often necessary to facilitate the estimation. For example, estimating propensity scores may be impossible when the dataset is high-dimensional with long time-series or a rich set of covariates. We thus add the following two assumptions on the treatment assignment process.

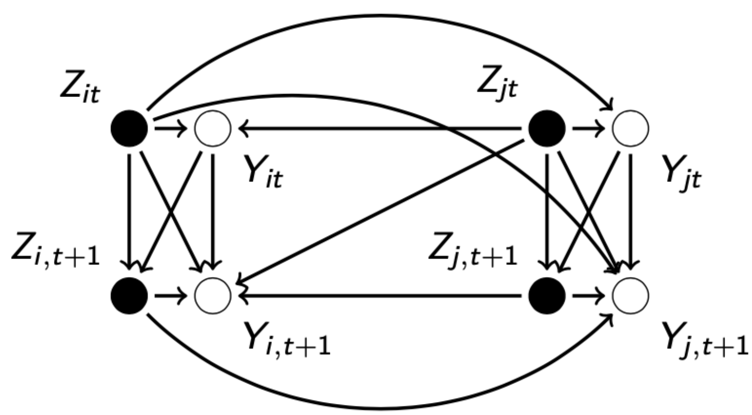
Assumption 3 (Bernoulli design)**.** *In any period t, Z_{it} is independent to each other for any i.* $0 < P(\mathbf{Z}_i^{1:t} = \mathbf{z}^{1:t}) < 1$.

The first part of the assumption is usually implicit in observational studies. We write it down to emphasize the fact that Z_{it} can be dependent on the history, $\mathbf{Z}^{1:(t-1)}$, but not on the treatment status of other units in the same period, $\mathbf{Z}_t \setminus Z_{it}$. The second part is the common requirement of positivity or overlapping: each possible history for unit i should have a positive probability to occur on its support.

Assumption 4 (No contagion). *For any i and t, the probability $P(Z_{it} = z)$ is decided only by unit i's own history.*

By assuming no contagion, we block the direct causal path from $\mathbf{Z}_j^{1:(t-1)}$ to Z_{it} for $j \neq i$. Now $\mathbf{Z}_j^{1:(t-1)}$ can only affect Z_{it} via $\mathbf{Y}_i^{1:(t-1)}$. In other words, the set of confounders excludes $\mathbf{Z}_j^{1:(t-1)}$, but may include $\mathbf{Z}_i^{1:(t-1)}$ and $\mathbf{Y}_i^{1:(t-1)}$. This assumption reduces the dimension of the space of propensity scores. We will explore its relaxation in the section on extensions. Figure 1 uses a directed acyclic graph (DAG) to illustrate the relationship between the treatment variables and outcomes under Assumptions 1-4[7].

Figure 1: A DAG illustration



Under these four assumptions, we can write the propensity score for each observation (i, t) as $P(Z_{it}|\mathbf{Z}_i^{1:(t-1)}, \mathbf{Y}_i^{1:(t-1)}, \mathbf{X}_i^{1:t})$. The propensity score for the entire history of treatment assignment for unit i up to period t, $\mathbf{Z}_i^{1:t}$, can be expressed as $W_{it} = \prod_{s=1}^{t} P(Z_{is}|\mathbf{Z}_i^{1:(s-1)}, \mathbf{Y}_i^{1:(s-1)}, \mathbf{X}_i^{1:s})$. The expression can be further simplified with more structural restrictions imposed. For example, we may assume that sequential ignorability holds after conditioning on the treatment status and outcome value in the previous period. Then, we have $W_{it} = \prod_{s=1}^{t} P(Z_{is}|Z_{i,s-1}, Y_{i,s-1}, \mathbf{X}_{is})$.

[7]We consider the simplest case with two units i, j and two periods $t, t+1$. Covariates are omitted to save space. Note that the graph's structure will be more complicated without Assumptions 1 and 4. For example, there will exist an arrow from $Z_{i,t+1}$ to Y_{it} as well as an arrow from Z_{jt} to $Z_{i,t+1}$.

4 Identification and estimation under the assumption of sequential ignorability

4.1 Identification

We now discuss the identification of our estimands when Assumptions 1-4 are satisfied. We leave out covariates in what follows to save space. To gain some intuition, let's first look at the average contemporary direct effect. We have the following result:

$$\begin{aligned}
\tau_t &= \frac{1}{N}\sum_{i=1}^{N}\tau_{it} \\
&= \frac{1}{N}\sum_{i=1}^{N}\mathrm{E}_{\mathbf{Z}^{1:t}}\left[\frac{Z_{it}Y_{it}(\mathbf{Z}^{1:t})}{P(Z_{it}=1)}\right] - \frac{1}{N}\sum_{i=1}^{N}\mathrm{E}_{\mathbf{Z}^{1:t}}\left[\frac{(1-Z_{it})Y_{it}(\mathbf{Z}^{1:t})}{1-P(Z_{it}=1)}\right].
\end{aligned}$$

Notice that the estimand is decomposed into two moments (expectations), each of which can be obtained via repeating the assignment process. Intuitively, the decomposition holds because after reweighting with the propensity score, each outcome value is an unbiased estimate of the expected outcome that marginalizes over the treatment status of other observations. Now, for the EATE (the average cumulative indirect effect) at distance d, we can similarly show that:

$$\begin{aligned}
&\tau_t(\mathbf{z}^{s:t},\tilde{\mathbf{z}}^{s:t};d) = \frac{1}{N}\sum_{i=1}^{N}\mu_i(\tau_{(1:N),t}(\mathbf{z}^{s:t},\tilde{\mathbf{z}}^{s:t});d) \\
=&\frac{1}{N}\sum_{i=1}^{N}\mathrm{E}_{\mathbf{Z}^{1:t}}\left[\frac{\mathbf{1}\{\mathbf{Z}_i^{s:t}=\mathbf{z}^{s:t}\}\mu_i(\mathbf{Y}_t(\mathbf{Z}^{1:t});d)}{P(\mathbf{Z}_i^{s:t}=\mathbf{z}^{s:t})}\right] - \frac{1}{N}\sum_{i=1}^{N}\mathrm{E}_{\mathbf{Z}^{1:t}}\left[\frac{\mathbf{1}\{\mathbf{Z}_i^{s:t}=\tilde{\mathbf{z}}^{s:t}\}\mu_i(\mathbf{Y}_t(\mathbf{Z}^{1:t});d)}{P(\mathbf{Z}_i^{s:t}=\tilde{\mathbf{z}}^{s:t})}\right],
\end{aligned}$$

where $\mathbf{Y}_t(\mathbf{Z}^{1:t}) = (Y_{1t}(\mathbf{Z}^{1:t}), Y_{2t}(\mathbf{Z}^{1:t}), \ldots, Y_{Nt}(\mathbf{Z}^{1:t}))$. Again, the estimand equals to the difference between two expectations. Treating $\mu_i(\mathbf{Y}_t(\mathbf{Z}^{1:t});d)$ as the outcome and $\mathbf{1}\{\mathbf{Z}_i^{s:t} = \mathbf{z}^{s:t}\}$ as the treatment, then the same intuition applies for why we have the equality. We formally summarize the results in the following theorem:

Theorem 1. *Under Assumptions 1-4, the average contemporary direct effect, the average cumulative direct effect, the average contemporary indirect effect, and the average cumulative indirect effect are all non-parametrically identifiable.*

4.2 The IPTW estimators

Theorem 1 leads to a natural estimator for the EATE— the quantity under the expectation sign in the results on identification[8]:

$$\hat{\tau}_t(\mathbf{z}^{s:t}, \tilde{\mathbf{z}}^{s:t}; d) = \frac{1}{N}\sum_{i=1}^{N} \frac{\mathbf{1}\{\mathbf{Z}_i^{s:t} = \mathbf{z}^{s:t}\}\mu_i(\mathbf{Y}_t(\mathbf{Z}^{1:t}); d)}{\hat{P}(\mathbf{Z}_i^{s:t} = \mathbf{z}^{s:t})} - \frac{1}{N}\sum_{i=1}^{N} \frac{\mathbf{1}\{\mathbf{Z}_i^{s:t} = \tilde{\mathbf{z}}^{s:t}\}\mu_i(\mathbf{Y}_t(\mathbf{Z}^{1:t}); d)}{\hat{P}(\mathbf{Z}_i^{s:t} = \tilde{\mathbf{z}}^{s:t})}.$$

It takes the form of the Horvitz-Thompson estimator, one of the IPTW estimators[9]. Note that the propensity scores in the expression are unknown and have to be estimated. Treating them as nuisance parameters, the estimator also belongs to the class of semi-parametric estimators whose statistical properties have been well studied (Van der Vaart, 2000). To estimate the propensity scores, researchers may fit a simple logistic regression model or employ more advanced techniques such as highly adaptive LASSO (Ertefaie, Hejazi and van der Laan, 2020), covariates balancing propensity score (Imai and Ratkovic, 2015), or kernel optimal weighting(Kallus and Santacatterina, 2018).

There is a well-known shortcoming of the Horvitz-Thompson estimator: when some of the estimated propensity scores have extreme values that are close to 0 or 1, the estimates may deviate greatly from their true values (Glynn and Quinn, 2010). Therefore, researchers could "stabilize" the estimation by using the Hajek estimator instead, which replaces the denominator N with the weighted sum of the treatment indicator:

[8] For the average contemporary direct effect, the estimator simplifies to be:

$$\hat{\tau}_t = \frac{1}{N}\sum_{i=1}^{N} \frac{Z_{it}Y_{it}}{\hat{P}(Z_{it} = 1)} - \frac{1}{N}\sum_{i=1}^{N} \frac{(1 - Z_{it})Y_{it}}{1 - \hat{P}(Z_{it} = 1)}$$

[9] See van der Laan (2014), Ogburn et al. (2020), Aronow, Samii and Wang (2019), and Papadogeorgou et al. (2020) for similar estimation strategies.

$$\hat{\tau}_{t,HA}(\mathbf{z}^{s:t}, \tilde{\mathbf{z}}^{s:t}; d) = \frac{\sum_{i=1}^{N} \mathbf{1}\{\mathbf{Z}_i^{s:t} = \mathbf{z}^{s:t}\} \mu_i(\mathbf{Y}_t(\mathbf{Z}^{1:t}); d) / \hat{P}(\mathbf{Z}_i^{s:t} = \mathbf{z}^{s:t})}{\sum_{i=1}^{N} \mathbf{1}\{\mathbf{Z}_i^{s:t} = \mathbf{z}^{s:t}\} / \hat{P}(\mathbf{Z}_i^{s:t} = \mathbf{z}^{s:t})} - \frac{\sum_{i=1}^{N} \mathbf{1}\{\mathbf{Z}_i^{s:t} = \tilde{\mathbf{z}}^{s:t}\} \mu_i(\mathbf{Y}_t(\mathbf{Z}^{1:t}); d) / \hat{P}(\mathbf{Z}_i^{s:t} = \tilde{\mathbf{z}}^{s:t})}{\sum_{i=1}^{N} \mathbf{1}\{\mathbf{Z}_i^{s:t} = \tilde{\mathbf{z}}^{s:t}\} / \hat{P}(\mathbf{Z}_i^{s:t} = \tilde{\mathbf{z}}^{s:t})}$$

From Theorem 1, it is straightforward to see that the Horvitz-Thompson estimator is unbiased if we know the true values of the nuisance parameters while the Hajek estimator is not. Establishing large sample properties for the two estimators requires stronger assumptions. We will return to this topic in section 6.

4.3 The augmented estimators

The estimators proposed above can be "augmented" by approximating the response surfaces, Y_{it} and $\mu_i(\mathbf{Y}_t; d)$, more precisely. The idea is to estimate a diffusion model for $\mathrm{E}\left[Y_{it} | \mathbf{Z}^{1:t}, \mathbf{Y}^{1:(t-1)}, \mathbf{X}^{1:t}\right]$ and plug the fitted value $\hat{Y}_{it} = \hat{Y}_{it}(\mathbf{z}^{1:t}, \mathbf{y}^{1:(t-1)}, \mathbf{x}^{1:t})$ into the IPTW estimators[10]. For example, we can assume that the true effect function is homogeneous, additive and proportional to the distance to a treated unit: $g_{it}(\mathbf{Z}_t) = \sum_{d=d_1}^{d_{\tilde{d}}} \beta_d \sum_{j=1}^{N} Z_{jt} \mathbf{1}\{d_{ij} = d\}$[11]. Furthermore, the effect is assumed to be non-cumulative in time. Then, we can fit the following model to predict Y_{it}:

$$Y_{it} = g_{it}(\mathbf{Z}_t) + h(\mathbf{Z}^{1:(t-1)}, \mathbf{Y}^{1:(t-1)}, \mathbf{X}^{1:t}) + \varepsilon_{it}.$$

where h is a function with a known form. After fitting the model, we can obtain the fitted values $\hat{Y}_{it}$ for each observation.

If the above model captures the true data generating process, then β_d equals to the average contemporary indirect effect. In this sense, the augmented estimator is doubly robust: it is unbiased when either the propensity score or the response surface model is correctly specified. However, it is highly unlikely that our diffusion model is close to

[10] For the spillover mapping, we also have the fitted value $\hat{\mu}_i = \mu_i(\hat{\mathbf{Y}}_t; d)$.

[11] $\tilde{d}$ is the maximal range that the effect from a unit could spread to.

the true one (just see how many restrictions we have imposed on $g_{it}(\mathbf{Z}_t)$!). Even if the selected model is a precise approximation of the true DGP, we still need correct propensity score estimates to marginalize the effects when the effect function is non-additive. But a "wrong model" is still helpful in enhancing the efficiency of the estimation. We may understand the augmented estimators from the perspective of residual balancing (Liu et al., 2019; Athey, Imbens and Wager, 2018): we use the diffusion model to reduce noises and then weighting to balance the remaining influences of the confounders. The augmented estimator for the average cumulative indirect effect has the following form:

$$
\begin{aligned}
&\hat{\tau}_{t,Aug}(\mathbf{z}^{s:t}, \tilde{\mathbf{z}}^{s:t}; d) \\
=&\frac{1}{N}\sum_{i=1}^{N}\left[\frac{\mathbf{1}\{\mathbf{Z}_i^{s:t} = \mathbf{z}^{s:t}\}\left[\mu_i(\mathbf{Y}_t(\mathbf{Z}^{1:t}); d) - \mu_i(\hat{\mathbf{Y}}_t(\mathbf{Z}^{1:t}); d)\right]}{\hat{P}(\mathbf{Z}_i^{s:t} = \mathbf{z}^{s:t})} + \mu_i(\hat{\mathbf{Y}}_{t;i}(\mathbf{z}^{s:t}); d)\right] \\
&- \frac{1}{N}\sum_{i=1}^{N}\left[\frac{\mathbf{1}\{\mathbf{Z}_i^{s:t} = \tilde{\mathbf{z}}^{s:t}\}\left[\mu_i(Y_t(\mathbf{Z}^{1:t}); d) - \mu_i(\hat{\mathbf{Y}}_t(\mathbf{Z}^{1:t}); d)\right]}{\hat{P}(\mathbf{Z}_i^{s:t} = \tilde{\mathbf{z}}^{s:t})} + \mu_i(\hat{\mathbf{Y}}_{t;i}(\tilde{\mathbf{z}}^{s:t}); d)\right]
\end{aligned}
$$

where $\hat{\mathbf{Y}}_t(\mathbf{Z}^{1:t})$ is the predicted outcome values for $\mathbf{Y}_t(\mathbf{Z}^{1:t})$ and $\hat{\mathbf{Y}}_{t;i}(\mathbf{z}^{s:t}); d)$ is the estimate of the marginalized outcome $\mathrm{E}\left[Y_{it}|\mathbf{z}^{s:t}, \mathbf{Z}^{s:t} \setminus \mathbf{Z}_k^{s:t}\right]$[12]. We demonstrate more details of the augmented estimator in the appendix and show that the weighted DID estimator proposed by Strezhnev (2018) can be seen as a variant of it. The estimator's statistical properties are similar to those of the IPTW estimators and will also be discussed in section 6.

[12] When the effect is indeed additive, $\frac{1}{N}\sum_{i=1}^{N}(\mu_i(\hat{\mathbf{Y}}_{t;i}(\mathbf{z}^{s:t}); d) - \mu_i(\hat{\mathbf{Y}}_{t;i}(\tilde{\mathbf{z}}^{s:t}); d)) = \beta_d$. Otherwise, it has to be weighted by the estimates of propensity scores.

5 Identification and estimation under the assumption of strict exogeneity

5.1 The basic idea

As discussed, another common approach in TSCS data analysis is the fixed effects model (including DID as a special case). Their identification relies on the assumption of strict exogeneity. It typically indicates the mean independence of error terms to the treatment conditional on unobservable variables that have a low-dimensional approximation, such as the unit and period fixed effects. Since these confounders are unobservable, we are no longer able to estimate the propensity score for each observation. Nevertheless, if we assume SUTVA holds and the fixed effects influence the outcome in a certain manner (e.g. linearly), it is possible to remove their impact via adjusting the response surface[13]. Then, the difference between the treated and control observations can be purely attributed to the treatment, which results in an unbiased and consistent estimate of the average treatment effect on the treated (ATT).

This is also the idea behind most model-based approaches in TSCS data analysis. Instead of considering the assignment process, researchers choose to predict the counterfactual outcome for the treated observations based upon assumptions on the response surface, or the outcome's trajectory. These assumptions, such as parallel trends or mean independence, enable researchers to obtain valid estimates for the ATT without knowing the propensity scores. When the dataset has a generalized DID or staggered adoption structure, such an approach is robust to temporal interference, since the prediction of counterfactual outcomes only relies on untreated observations where no interference exists. Yet when spatial interference is also present, the model-based approaches can no longer return any result of substantive interest. The reason is simple: without the propensity scores, we

[13] For example, we can eliminate the unit fixed effects' influence by subtracting the unit-specific mean from each observation.

do not know to what extent the response surfaces of different groups have contaminated each other, or to which unit's treatment the observed effect should be attributed.

5.2 An example

To see the point more clearly, let's consider a simple case where there are N units in two groups and two periods. The treatment group receives the treatment $Z_{i2} = 1$ only in period 2 and the control group remains untreated in both periods. The unobservable propensity score for each unit is denoted as p_i. We employ the following data generating process:

$$Y_{i1} = \mu + \alpha_i + \xi_1 + \varepsilon_{i1}$$
$$Y_{i2} = \mu + g_i(\mathbf{Z}_2) + \alpha_i + \xi_2 + \varepsilon_{i2}$$

where μ is the grand intercept; α_i and ξ_t represent unit and period fixed effects, respectively; ε_{it} is the idiosyncratic error term and $g_i()$ stands for the treatment's impact on unit i in period 2. We assume that the N units are drawn from a larger population and $\mathbb{E}\left[\varepsilon_{it}|\alpha_i, \xi_t, \mathbf{Z}^{1:2}\right] = 0$, where the expectation $\mathbb{E}$ is taken over samplings[14]. In other words, ε_{it} and $\mathbf{Z}$ are conditionally mean-independent. This is the assumption of "strict exogeneity" and it implies the parallel trends assumption in the DID analysis:

$$\mathbb{E}\left[Y_{i2}(0^{1:2}) - Y_{i1}(0^{1:2})|Z_{i2} = 1\right] = \mathbb{E}\left[Y_{i2}(0^{1:2}) - Y_{i1}(0^{1:2})|Z_{i2} = 0\right].$$

Suppose we want to estimate the direct effect generated by the treatment. The conventional DID estimator for the effect in this example is equivalent to the within estimator of two-way fixed effects model, and it takes the following form:

[14] We can assume strict exogeneity within the sample, but we use its population version to be consistent with the convention in the literature.

$$
\begin{aligned}
\tau_{DID} =& \frac{1}{N_1}\sum_{i=1}^{N} Z_{i2}(Y_{i2}-Y_{i1}) - \frac{1}{N_0}\sum_{i=1}^{N}(1-Z_{i2})(Y_{i2}-Y_{i1}) \\
\rightarrow& \frac{\mathbb{E}\left[p_i \mathrm{E}_{\mathbf{Z}^{1:2}\setminus Z_{i2}} g_i(1,\mathbf{Z}^{1:2}\setminus Z_{i2})\right]}{\mathbb{E}p_i} - \frac{\mathbb{E}\left[(1-p_i)\mathrm{E}_{\mathbf{Z}^{1:2}\setminus Z_{i2}} g_i(0,\mathbf{Z}^{1:2}\setminus Z_{i2})\right]}{\mathbb{E}(1-p_i)} \\
\neq& \mathbb{E}\left[\mathrm{E}_{\mathbf{Z}^{1:2}\setminus Z_{i2}} g_i(1,\mathbf{Z}^{1:2}\setminus Z_{i2})\right] - \mathbb{E}\left[\mathrm{E}_{\mathbf{Z}^{1:2}\setminus Z_{i2}} g_i(0,\mathbf{Z}^{1:2}\setminus Z_{i2})\right]
\end{aligned}
$$

The quantity in the last line is the estimand of interest, the average contemporary direct effect τ_2[15], for the whole population. Obviously, the estimate's limit equals to the estimand of interest only if p_i and $\mathrm{E}_{\mathbf{Z}^{1:2}\setminus Z_{i2}} g_i(z_{i2},\mathbf{Z}^{1:2}\setminus Z_{i2})$ are uncorrelated. This occurs, for example, when $\mathrm{E}_{\mathbf{Z}^{1:2}\setminus Z_{i2}} g_i(z_{i2},\mathbf{Z}^{1:2}\setminus Z_{i2})$ is homogeneous. In practice, however, the value of both p_i and $\mathrm{E}_{\mathbf{Z}^{1:2}\setminus Z_{i2}} g_i(z_{i2},\mathbf{Z}^{1:2}\setminus Z_{i2})$ could be decided by the unit fixed effect α_i and the correlation between them always exists. Then, the DID estimate is asymptotically biased and inconsistent. The result holds for the general two-way fixed effects model as its estimate equals to a weighted average of a series of DID estimates, as shown in Strezhnev (2018). When there is no interference, $\mathrm{E}_{\mathbf{Z}^{1:2}\setminus Z_{i2}} g_i(0,\mathbf{Z}^{1:2}\setminus Z_{i2}) = 0$ and $\mathrm{E}_{\mathbf{Z}^{1:2}\setminus Z_{i2}} g_i(1,\mathbf{Z}^{1:2}\setminus Z_{i2}) = \tau_i$. The above expression boils down to $\frac{\mathbb{E}[p_i\tau_i]}{\mathbb{E}p_i}$, which is the exact definition of the ATT. From the expression, we know the bias of the DID estimate from the estimand τ_2 is:

[15] As we only have two periods, there is no temporal interference and the average contemporary direct effect is the same as the average cumulative direct effect.

$$
\begin{aligned}
Bias_{DID} =& \tau_{DID} - \tau_2 \\
=& \frac{\mathbb{E}\left[p_i \mathrm{E}_{\mathbf{Z}^{1:2} \setminus Z_{i2}} g_i(1, \mathbf{Z}^{1:2} \setminus Z_{i2})\right]}{\mathbb{E} p_i} - \frac{\mathbb{E}\left[(1-p_i) \mathrm{E}_{\mathbf{Z}^{1:2} \setminus Z_{i2}} g_i(0, \mathbf{Z}^{1:2} \setminus Z_{i2})\right]}{\mathbb{E}(1-p_i)} \\
& - \left(\mathbb{E}\left[\mathrm{E}_{\mathbf{Z}^{1:2} \setminus Z_{i2}} g_i(1, \mathbf{Z}^{1:2} \setminus Z_{i2})\right] - \mathbb{E}\left[\mathrm{E}_{\mathbf{Z}^{1:2} \setminus Z_{i2}} g_i(0, \mathbf{Z}^{1:2} \setminus Z_{i2})\right]\right) \\
=& \frac{\mathrm{Cov}\left[p_i, \mathrm{E}_{\mathbf{Z}^{1:2} \setminus Z_{i2}} g_i(1, \mathbf{Z}^{1:2} \setminus Z_{i2})\right]}{\mathbb{E} p_i} - \frac{\mathrm{Cov}\left[1-p_i, \mathrm{E}_{\mathbf{Z}^{1:2} \setminus Z_{i2}} g_i(0, \mathbf{Z}^{1:2} \setminus Z_{i2})\right]}{\mathbb{E}(1-p_i)} \\
=& \frac{\mathrm{Cov}\left[p_i, \mathrm{E}_{\mathbf{Z}^{1:2} \setminus Z_{i2}} g_i(1, \mathbf{Z}^{1:2} \setminus Z_{i2})\right]}{\mathbb{E} p_i} + \frac{\mathrm{Cov}\left[p_i, \mathrm{E}_{\mathbf{Z}^{1:2} \setminus Z_{i2}} g_i(0, \mathbf{Z}^{1:2} \setminus Z_{i2})\right]}{\mathbb{E}(1-p_i)}
\end{aligned}
$$

The direction of the bias depends on the correlation between the unobservable propensity score p_i and the magnitude of the marginalized effect. When both of them are positively correlated with α_i, for instance, DID will overestimate the average contemporary direct effect. Otherwise it may underestimate the effect. Similar conclusion holds for other estimands such as the average cumulative indirect effect[16].

5.3 Discussion

In TSCS data analysis, sequential ignorability and strict exogeneity are the two most common assumptions. They lead to the design-based (e.g. IPTW) and model-based (e.g. DID and FE models) approaches, respectively. The latter assumption is often considered as weaker as it permits unobservable attributes to affect both the outcome and the treatment. But as we have seen in this section, when between-unit interference becomes a serious concern, the model-based approaches are no longer feasible for estimating the causal effects. Instead, the design-based approaches still work with some modest adjustments. This articulates a tradeoff that empirical researchers have to make based on their substantive

[16]Notice that the bias does not diminish to zero even if SUTVA holds. In that case, it is the difference between the ATT and the ATE and equals to zero only when the effect is homogeneous across the two treatment groups.

knowledge: which is more likely to bias the study's estimates, unobservable confounders or interference?

That said, the cost of ignoring fixed effects might be smaller than what researchers would have imagined. When the units of interest are small in size and contiguous[17], we can assume that their unobservable attributes vary continuously on the geography. Hence, the influence of unit fixed effects can be well approximated by a smooth function of the geographic coordinates of the units. In section 8, we show through simulations that the approximation of unit fixed effects that are continuously distributed via a quadratic polynomial of the coordinates is fairly accurate. When units in the dataset are large in size and distinct from each other (e.g. countries), we do not recommend researchers to make this assumption or apply the method under discussion. In these cases, between-unit interference is also less likely to appear.

The problem of the model-based approaches has been well documented by Aronow and Samii (2016) in the cross-sectional setting. But unlike in their paper, p_i here is unknown under the asumption of strict exogeneity. Hence, we cannot adjust the weights to obtain consistent estimates. The result above suggests that under spatial interference, the conventional DID estimator converges to a weighted sum of marginalized outcomes, which usually has no substantive meaning. One may modify the estimator by explicitly modeling the effect function $E[g_i(\mathbf{Z})|\mathbf{Z}^t, \mathbf{Y}^{t-1}, \mathbf{X}^t]$. But such a model can hardly be correct in reality[18]. Essentially, the tradeoff is about how we want to remain agnostic. If we possess little knowledge on the structure of interference but a clear picture on the assignment process, then the design-based approaches should be preferable. Instead, when we are aware of how the effects diffuse among the units but uncertain about the set of confounders, trying to model the effect function while allowing for the existence of fixed effects may be a more reasonable choice.

[17]Examples include towns, polling stations, or even pixels on maps.

[18]Nevertheless, as we have demonstrated in the previous section, these models can be used to augment the IPTW estimators and enhance their efficiency.

6 Statistical properties of the IPTW estimators

6.1 The dependency graph

To establish the large sample properties of the proposed estimators, we need their variance to gradually decrease to zero as $N \to \infty$[19]. Yet this may not happen when the interdependence among the observations caused by interference (or more formally, interference dependence) is too strong. Suppose that each observation interferes with all the others, we actually have only one valid observation and the variance does not change with the sample size. Therefore, we must restrict the maximal degree of interference dependence among the observations.

We measure the degree of interference dependence using the dependency graph (Savje, Aronow and Hudgens, 2018). For the outcome Y, it is represented by a $NT \times NT$ adjacency matrix $\mathcal{G}_Y = \{g_{(it,js);Y}\}_{NT\times NT}$. Each observation (i,t) in the sample is treated as a node in the graph. We say the observation (i,t) interferes with the observation (j,s) in Y if the former's outcome is influenced by the latter's treatment status. By definition, an observation always interferes in Y with itself. A link between observations (i,t) and (j,s) in $\mathcal{G}_Y$ exists when one interferes in Y with the other, or both of them interfere in Y with a third observation. Two observations are dependent in Y if and only if they are linked in $\mathcal{G}_Y$. We emphasize "in Y" since the dependency graph's structure holds only for the particular variable. More formally, we define a dummy variable $I_{(it,js);Y}$ to indicate whether there is interference in Y between (i,t) and (j,s), then,

$$I_{(it,js);Y} = \begin{cases} 1, \text{ if } i = j \text{ and } t = s; \\ 1, \text{ if } Y_{it}(\mathbf{Z}^{1:T}) \neq Y_{it}(\tilde{\mathbf{Z}}^{1:T}) \text{ and } \mathbf{Z}^{1:T} \setminus \mathbf{Z}_{js} = \tilde{\mathbf{Z}}^{1:T} \setminus \mathbf{Z}_{js}; \\ 0, \text{ otherwise.} \end{cases}$$

[19]To be more accurate, here N is the number of units with the assignment history $\mathbf{z}^{s:t}$ or $\tilde{\mathbf{z}}^{s:t}$.

Then, $g_{(it,js);Y} = 1$ if and only if $I_{(it,js);Y} = 1$ or $I_{(it,lq);Y} I_{(js,lq);Y} = 1$ for some observation (l, q). Suppose we reorder the sample such that the T observations of each unit are next to each other, then $\{g_{(it,js);Y}\}_{NT \times NT}$ is a block matrix with $N \times N$ blocks. Values on the diagonal of each block reflect the degree of interference dependence driven by temporal interference, while values of blocks off the diagonal are non-zero only when there is spatial interference.[20] Two units i and j are defined as dependent in Y if $g_{(it,js);Y} = 1$ for some periods t and s and we denote this as $g_{ij;Y} = 1$. We further denote the set of units that depend on unit i as $\mathcal{B}(i;Y)$ and its number as $b_{i;Y}$. Clearly, $b_{i;Y} = \sum_{j=1}^{N} g_{ij;Y} = \sum_{j \in \mathcal{B}(i;Y)} g_{ij;Y}$. The specific structure of the graph $\mathcal{G}_Y$ is often unknown to researchers, hence cannot be used to construct exposure mappings as in Aronow and Samii (2017). But it is still a helpful device for defining theoretical concepts that will appear in the derivation of analytic results.

As the exposure mapping μ's values are used as the outcome when estimating the indirect effects, we also need to define the interference and interference dependence in μ among the observations. Two observations (i, t) and (j, s) are dependent in $\mu(.; d)$ if one observation (i', t) that composes $\mu_i(\mathbf{Y}_t; d)$ and one observation (j', s) that composes $\mu_j(\mathbf{Y}_s; d)$ are dependent in Y. Two units i and j are dependent in $\mu(.; d)$ if (i, t) and (j, s) are dependent in $\mu(.; d)$ for some t and s. We similarly use $\mathcal{B}(i; d)$[21] to denote the set of units that are dependent on unit i in $\mu(.; d)$ and $b_{i;d}$ to denote the cardinality of $\mathcal{B}(i; d)$. Obviously, the degree of interference dependence in $\mu(.; d)$ may differ from that in Y. In the case of circle means, the two are equal when $d = 0$, and the former will become higher when d is larger. We make the following assumption on the maximal degree of interference dependence:

Assumption 5 (Limited degree of interference dependence)**.** $\max_{i \in \{1,2,...,N\}} b_{i;d} = o_P(N^{1/2})$.

[20] If SUTVA holds, then $\{g_{(it,js);Y}\}_{NT \times NT}$ becomes a diagonal matrix where only elements on the diagonal equal to 1 and others equal to zero.

[21] $\mathcal{B}(i; d)$ is short for $\mathcal{B}(i; \mu(.; d))$ as $\mu(.; d)$ is indexed by d.

By definition, $\max_{i\in\{1,2,\dots,N\}} b_{i;d}$ is the maximal number of units that a unit could be dependent on in $\mu(.;d)$ within the sample. The assumption thus indicates that the maximum grows to infinity at the relatively slower rate ($o_P(N^{1/2})$) than the number of units does. When the sample includes more units, the dependency graph becomes "larger" more rapidly than it becomes "denser." Under finite sample, it means that the magnitude of $\max_{i\in\{1,2,\dots,N\}} b_{i;d}$ is negligible compared to N. The assumption's credibility is determined by the dependence in Y among the units as well as the form of μ. For circle means, the assumption is less likely to be satisfied as d increases. The last assumption we need for the asymptotic results is the boundedness of the potential outcomes. This assumption, combined with the linearity of μ, ensure the existence of all the moments of $\mu_i(\mathbf{Y}_t)$.

Assumption 6 (Bounded potential outcomes)**.** *There exist a constant $\tilde{y}$ such that $|Y_{it}(\mathbf{z}^{1:t})| \leq \tilde{y}$ for all i, t, and $\mathbf{z}^{1:t}$.*

6.2 Asymptotic distribution and inferece

With Assumptions 1-6, we are able to show that the variances of both the IPTW and the augmented estimators converge to zero in the limit, as long as the estimates for propensity scores are sufficiently accurate. The variance $V_t(d)$ can be decomposed into three parts: the classic Neyman variance without interference $V_{1t}(d)$, the variance induced by interference $V_{2t}(d)$, and the variance from estimating nuisance parameters $V_{3t}(d)$. The sum of $V_{1t}(d)$ and $V_{2t}(d)$ is the variance of the estimate under the true values of the nuisance parameters. Their expressions remain unchanged no matter what techniques we choose to estimate the nuisance parameters. We present the variance expressions only for the Hajek estimator:

$$
\begin{aligned}
& V_{1t}(d) \\
=& \frac{1}{N^2}\sum_{i=1}^{N}\frac{\mathrm{E}\left[\mu_i(\mathbf{Y}_t(\mathbf{z}^{s:t},\mathbf{Z}^{1:T}\setminus\mathbf{Z}_i^{s:t});d)-\bar{\mu}_t^1(d)\right]^2}{W_{it}} \\
&+\frac{1}{N^2}\sum_{i=1}^{N}\frac{\mathrm{E}\left[\mu_i(\mathbf{Y}_t(\tilde{\mathbf{z}}^{s:t},\mathbf{Z}^{1:T}\setminus\mathbf{Z}_i^{s:t});d)-\bar{\mu}_t^0(d)\right]^2}{\tilde{W}_{it}} \\
&-\frac{1}{N^2}\sum_{i=1}^{N}\mathrm{E}^2\left[\left(\mu_i(\mathbf{Y}_t(\mathbf{z}^{s:t},\mathbf{Z}^{1:T}\setminus\mathbf{Z}_i^{s:t});d)-\bar{\mu}_t^1(d)\right)-\left(\mu_i(\mathbf{Y}_t(\tilde{\mathbf{z}}^{s:t},\mathbf{Z}^{1:T}\setminus\mathbf{Z}_i^{s:t});d)-\bar{\mu}_t^0(d)\right)\right]
\end{aligned}
$$

$$
\begin{aligned}
& V_{2t}(d)=\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\in\mathcal{B}(i;d)}\sum_{\mathbf{a},\mathbf{b}=\mathbf{z}^{s:t}}^{\tilde{\mathbf{z}}^{s:t}}(-1)^{\mathbf{1}\{\mathbf{a}=\mathbf{b}\}}* \\
& \mathrm{Cov}\left[\mu_i(\mathbf{Y}_t(\mathbf{a},\mathbf{Z}^{1:T}\setminus\mathbf{Z}_i^{s:t});d)-\bar{\mu}_t^{\mathbf{1}\{\mathbf{a}=\mathbf{z}^{s:t}\}}(d),\mu_j(\mathbf{Y}_t(\mathbf{b},\mathbf{Z}^{1:T}\setminus\mathbf{Z}_j^{s:t});d)-\bar{\mu}_t^{\mathbf{1}\{\mathbf{b}=\tilde{\mathbf{z}}^{s:t}\}}(d)\right].
\end{aligned}
$$

where $\bar{\mu}_t^1(d)$ is the averages of $\mathrm{E}\left[\mu_i(\mathbf{Y}_t(\mathbf{z}^{s:t},\mathbf{Z}^{1:T}\setminus\mathbf{Z}_i^{s:t});d)\right]$ and $\bar{\mu}_t^0(d)$ is the average of $\mathrm{E}\left[\mu_i(\mathbf{Y}_t(\tilde{\mathbf{z}}^{s:t},\mathbf{Z}^{1:T}\setminus\mathbf{Z}_i^{s:t});d)\right]$. $V_{1t}(d)$ converges to zero as suggested by statistical canons. $V_{3t}(d)$ converges to zero under the correct model specification. Assumptions 5-6 ensure that $V_{2t}(d)$ becomes sufficiently small when N is large. Asymptotic unbiasedness and consistency then follows from Markov's inequality:

Theorem 2. *Under Assumptions 1-6, when the estimates of propensity scores are consistent for their true values, both the IPTW estimators and the augmented estimators are asymptotically unbiased and consistent.*

The proof of asymptotic normality is based on the combination of the theory on semi-parametric models and Stein's method. The method was originally proposed by Stein (1972) to show that a random variable obeys the normal distribution if it satisfies a particular differential equation. It was further developed into a series of central limit theorems for dependent observations (Chatterjee et al., 2008; Chin, 2019). Essentially, asymptotic normality holds as long as the functional of interest is robust to small perturbation of the treatment assignment, which translates into limited interference dependence among the observations. To save space, we only report the result for the Hajek estimator[22].

Theorem 3. *Under Assumptions 1-6, when the estimates of the propensity scores converge to their true values at a rate of* $O_P\left(\sqrt{\frac{\tilde{b}_d}{N}}\right)$ *or higher, estimates from the proposed Hajek estimator of the average cumulative indirect effect and its average over periods converge to normal distributions:*

$$\sqrt{\frac{N}{\tilde{b}_d}}(\hat{\tau}_{t,HA}(\mathbf{z}^{s:t}, \tilde{\mathbf{z}}^{s:t}; d) - \tau_t(\mathbf{z}^{s:t}, \tilde{\mathbf{z}}^{s:t}; d)) \rightarrow N(0, V_t(d)).$$

where $\tilde{b}_d$ *is a constant that satisfies* $1 \leq \tilde{b}_d \leq \max_{i \in \{1,2,\ldots,N\}} b_{i;d}$.

It is worth noting that we may no longer have $\sqrt{N}$-convergence due to interference. When each unit interferes with the same number of other units, $\tilde{b}_d = b(i;d)$ for any i, the convergence rate is discounted by a scale of $\frac{1}{\sqrt{b(i;d)}}$. If we assume that there is a hard bound for all the $b(i;d)$s that does not change with the sample size as in Aronow, Samii and Wang (2019), then $\tilde{b}_d = 1$ and $\sqrt{N}$-convergence remains. This assumption is convincing if there is a minimum distance between each pair of units and the effect declines as distance rises. Otherwise, the scale $\tilde{b}_d$ is in-between 1 and $\max_{i \in \{1,2,\ldots,N\}} b_{i;d}$. In practice, it can be difficult to pin down the value of $\tilde{b}_d$. Researchers may try different ones and examine whether the findings are robust to the variation. We require the propensity score model to have at least the same rate of convergence. The requirement is satisfied by many approaches such the classic logistic model (assuming the specification is correct), CBPS (Fan et al., 2016), sieve estimator (Hirano, Imbens and Ridder, 2003), and highly adaptive LASSO (Ertefaie, Hejazi and van der Laan, 2020).

To estimate $V_t(d)$, a convenient approach is to rely on the regression representation of the estimators. For example, the Hajek estimator has a representation of the following

[22] The results are "high-level" in the sense that they omit regularity conditions for specific models that are chosen to estimate the nuisance parameters. See the appendix for results for the Horvitz-Thompson estimator and the augmented estimators

form:

$$
\begin{aligned}
&\begin{pmatrix} \hat{\alpha}(d) \\ \hat{\tau}_{t,HA}(\mathbf{z}^{s:t}, \tilde{\mathbf{z}}^{s:t}; d) \end{pmatrix} \\
=&\arg\min_{\alpha_d,\tau_d} \sum_{i=1}^{N} \left(\frac{\mathbf{1}\{\mathbf{Z}_i^{s:t} = \mathbf{z}^{s:t}\}\hat{P}(\mathbf{Z}_i^{s:t} = \tilde{\mathbf{z}}^{s:t}) - \mathbf{1}\{\mathbf{Z}_i^{s:t} = \tilde{\mathbf{z}}^{s:t}\}\hat{P}(\mathbf{Z}_i^{s:t} = \mathbf{z}^{s:t})}{\hat{P}(\mathbf{Z}_i^{s:t} = \tilde{\mathbf{z}}^{s:t})\hat{P}(\mathbf{Z}_i^{s:t} = \mathbf{z}^{s:t})} \right) * \\
&(\mu_i(\mathbf{Y}_t; d) - \alpha_d - \tau_d \mathbf{1}\{\mathbf{Z}_i^{s:t} = \mathbf{z}^{s:t}\})^2
\end{aligned}
$$

We show in the appendix that OLS estimates from the above regression are the same as those from the Hajek estimator. Moreover, the regression's spatial HAC variance (Conley, 1999) estimate is asymptotically valid for the first two parts of the Hajek estimator's asymptotic variance, $V_{1t}(d) + V_{2t}(d)$. Finally, the estimate of $V_{3t}(d)$ can be acquired using the selected model of nuisance parameters.

7 Extensions

7.1 Contagious treatment assignment

In reality, policy or event that appeared in one place may not only affect the outcome of interest, but also the probability for the same type of policy or event to emerge in other places. States often follow each other's policy innovation, and a small-scale protest sometimes develops into a storm of revolution. In this case, our Assumption 4 is no longer satisfied. Z_{it} is a function of the unit's own history, as well as the past treatment assignments or even past outcomes of other units. Now, we have both interference from Z_{js} to Y_{it} and contagion from Z_{js} to Z_{it}. The basic framework does not have to change much to accommodate this extension. The only difference is that we need to modify the model for the propensity scores, which must account for the contagion of the treatment.

Various options have been proposed in both biostatistics and social sciences to describe the phenomenon of contagion. For example, biostatisticians have been relying on the

SIR (Susceptible-Infected-Recovered) model to investigate the contagion of diseases. In political science, Egami (2018) developed an approach using static causal graphs to identify the causal effect of contagion. Researchers may use their substantive knowledge to decide which model is the best for the context. As long as the model satisfies our requirement in Theorem 3, the main results remain valid. Of course, the existence of contagion may further consume the degree of freedom and reduce the convergence rate of the propensity score model. Researchers should keep this in mind when extending the method along this direction.

7.2 Alternative estimands

The main estimand we have discussed so far, the average cumulative indirect effect, is defined for each period t. We can further aggregate all the $\tau_t(\mathbf{z}^{s:t}, \tilde{\mathbf{z}}^{s:t}; d)$s over periods:

$$\tau(\mathbf{z}^{s:t}, \tilde{\mathbf{z}}^{s:t}; d) = \frac{1}{t-s+1} \sum_{t'=s}^{t} \tau_{t'}(\mathbf{z}^{s:t}, \tilde{\mathbf{z}}^{s:t}; d)$$

This quantity could be estimated via averaging over $\hat{\tau}_t(\mathbf{z}^{s:t}, \tilde{\mathbf{z}}^{s:t}; d)$s and it has the same statistical properties. The only thing that differs is its asymptotic variance, which now includes a part driven by temporal interference. Therefore, when estimating it, we should use the two-way HAC variance estimator (Cameron, Gelbach and Miller, 2012) rather than the spatial HAC one.

Moreover, as pointed out by Blackwell and Glynn (2018), different histories of treatment assignment in the sample can be combined into one group if we believe that their effects are comparable. For example, if the total number of treated periods, $g(\mathbf{Z}_i^{s:t}) = \sum_{t'=s}^{t} Z_{it'}$, is what matters, we can estimate the effect generated by $g(\mathbf{Z}_i^{s:t}) = g$ relative to $g(\mathbf{Z}_i^{s:t}) = g'$: $\tau(g, g'; d)$. Estimation and inference of this estimand proceed in similar ways.

Besides the cumulative effect of the treatment assignment history, we may also be interested in the "lag effects" it generates. In theory, we can define the historic direct/indirect effect by replacing the time index t in the cumulative direct/indirect effect with t', where $s \leq t' < t$. When $s = t' = t - 1$, for instance, it describes the effect of treatment assignment in the previous period. However, if no restriction is imposed on the heterogeneity of treatment effects, it is impossible to distinguish the lag effects from the heterogeneous effects in any period, as discussed in Blackwell and Glynn (2018). In the staggered adoption scenario, for instance, we may compare the estimated effect in period t for cohort $t - 1$ and the estimated effect in the same period for cohort $t - 2$ at any distance. Now cohort $t - 2$ has received treatment for two periods and cohort $t - 1$ just one period. Assuming that the effect is the same for both cohorts in the first period, then the difference between the two estimates reflects the lag effect of treatment in period $t - 1$. Yet this assumption may be not be valid in practice, especially when cohorts differ greatly from each other. Researchers thus must be cautious if they try to make claims on the lag effects.

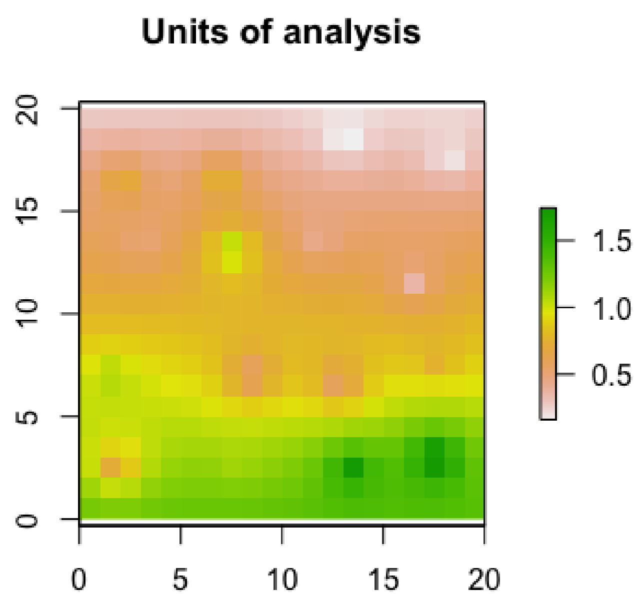
7.3 Apply the method to network setting

Our framework can also be applied to settings where units are located in a social network. Now, the Euclidean or Geodesic distance on the geographic surface is replaced by the length of the shortest path between any pair of nodes. Again, all the analyses will remain the same. Nevertheless, the assumption on local interference will be harder to satisfy in networks, due to the well-known "small world" phenomenon. If there exists a short path between any pair of nodes, and the true effect declines slowly along the path, then each node may interfere with most of other nodes and the assumption fails. Therefore, one must have a clear understanding of the network formation process and the true effect function before applying the method. Suppose the network consists of several separated clusters. As more nodes enter the network, the number of new clusters rather than the maximal size of each cluster rises, then Assumption 5 interference will still be guaranteed. But it may no longer be the case after some weak ties, or "bridges," are added into the network. For the direct effects, though, Li and Wager (2020) showed that the IPTW estimators are always consistent and asymptotically normal under slightly stronger assumptions.

8 Simulation

In this section, we test the performance of the proposed estimators via simulation. We use a simulated dataset with 400 units and 5 periods. The units are generated as a raster object. They lie on a 20×20 square and are contiguous to each other (see Figure 2).

Figure 2: Simulated units and the geography



Notes: The figure shows the location of units in the simulated dataset where the color of each cell indicates the value of α_i of the corresponding unit.

The potential outcome when there is no treatment, $Y_{it}(\mathbf{0}^T)$, is generated as follows:

$$Y_{it}(\mathbf{0}^T) = \mu + 0.3 * X_{1,it} + 0.5 * X_{2,it} + \alpha_i + \xi_t + \varepsilon_{it}.$$

where μ is the grand intercept; α_i and ξ_t represent unit and period fixed effects, respectively;

$X_{1,it}$ and $X_{2,it}$ are two covariates; ε_{it} is the idiosyncratic shock. We set $\mu = 5$ and draw each of ξ_t, $X_{2,it}$, and ε_{it} independently from the standard normal distribution. α_i is interpolated using a fitted kriging model with 16 random points on the surface. The distribution of α_i is displayed in Figure 2. $X_{2,it}$ obeys a normal distribution centered around α_i with unit standard deviation.

The treatment is adopted in a staggered way. It gradually kicks off from the third period. We thus have two pre-treatment periods and four types of histories[23]. The propensity score of unit i at period t is decided by the outcome and treatment status in the previous period as well as the covariates, but not the fixed effects. To be specific, we have:

$$P(Z_{it} = 1) = Logit(-2 + 0.2 * X_{1,it} + 0.2 * X_{2,it} + 0.05 * Y_{i,t-1} + 0.4 * Z_{i,t-1} + \nu_{it}).$$

where $\nu_{it} \sim N(0, 1)$. On average, the number of treated units rises from 85 in period 3 to 210 in period 5.

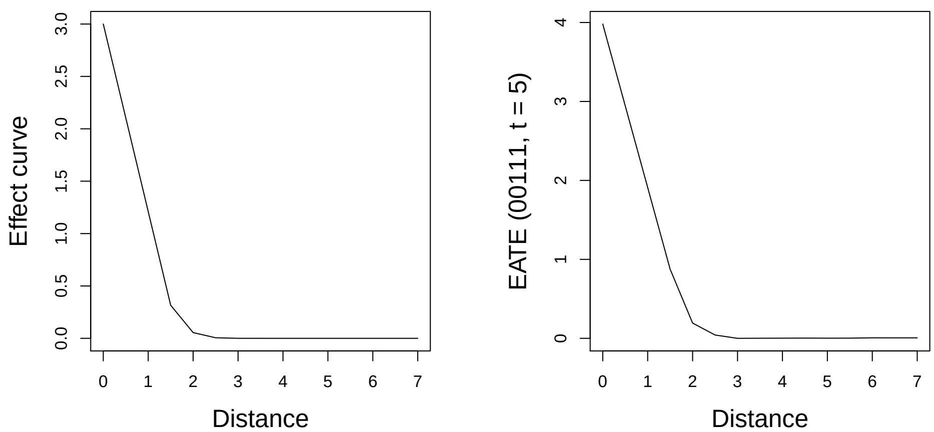
We assume that the exists an effect function emanating from each unit whose value gradually declines with the distance to that unit[24]. Effects from different units are additive. The total effect each unit receives is proportional to the value of its fixed effect α_i. Effects in the previous period have a lag effect with a "discount rate" of 0.6. The spillover mapping is set to be the circle mean and the propensity scores are estimated via a pooled logistic regression. Figure 3 shows the effect function and the corresponding EATE for the treatment assignment history $(0^{1:2}, 1^{3:5})$ in period 5 at each distance d. The EATE is larger than the effect function as it is amplified by the value of fixed effects as well as the lagged effect.

Let's first check the bias of the proposed IPTW estimators for the EATEs. We repeat the treatment assignment process for 1000 times and plot the estimates against the EATE curve. To save space, only results for period 5 from the Hajek estimator are presented. On

[23] They are $(0^{1:5})$, $(0^{1:4}, 1^5)$, $(0^{1:3}, 1^{4:5})$, and $(0^{1:2}, 1^{3:5})$.

[24] We show results for other effect functions in the appendix. This function is chosen as the results are easier to interpret.

Figure 3: The effect function and the expected average treatment effects

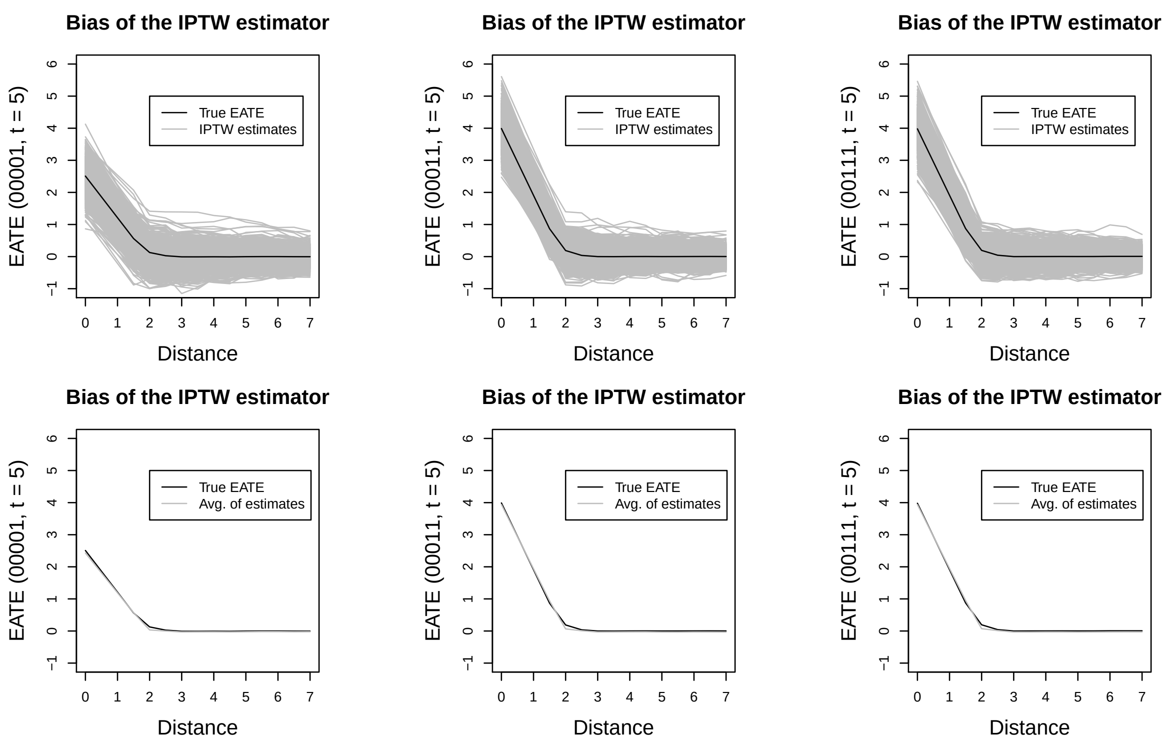


Notes: The left figure shows how the effect function varies with the distance to a unit and the right one shows the same for the expected average treatment effect (EATE) for the treatment assignment history $(\mathbf{0}^{1:2}, \mathbf{1}^{3:5})$ in period 5.

the top of Figure 4 we present the estimate for each assignment history. On the bottom, the average of these estimates is compared against the corresponding EATE. Even though the sample size is only moderate (400 units), the bias of the Hajek estimator is negligible, confirming our conclusion that the bias of the IPTW estimators gradually diminishes.

Next, we allow the unit fixed effect α_i to positively affect the propensity score. The sequential ignorability assumption no longer holds. Since we also make the effect for each unit to be positively related to α_i, our analysis in section 5 suggests that the DID estimates of the EATEs should have an upward bias. The results are shown in Figure 5. As predicted, the average of the DID estimates is much larger than the true EATE at all distances. Under this setting, the IPTW estimators would also be biased if we simply ignore the unit fixed effects when estimating the propensity scores. Nevertheless, these fixed effects vary continuously in the simulated dataset. Hence, we try control their influence by adding a polynomial of geographic coordinates into the estimating equation of propensity scores. Results using this approach are presented in Figure 6. We can see that the bias becomes much smaller. Therefore, the proposed estimators dominate the DID approach

Figure 4: The bias of the IPTW estimator under sequential ignorability

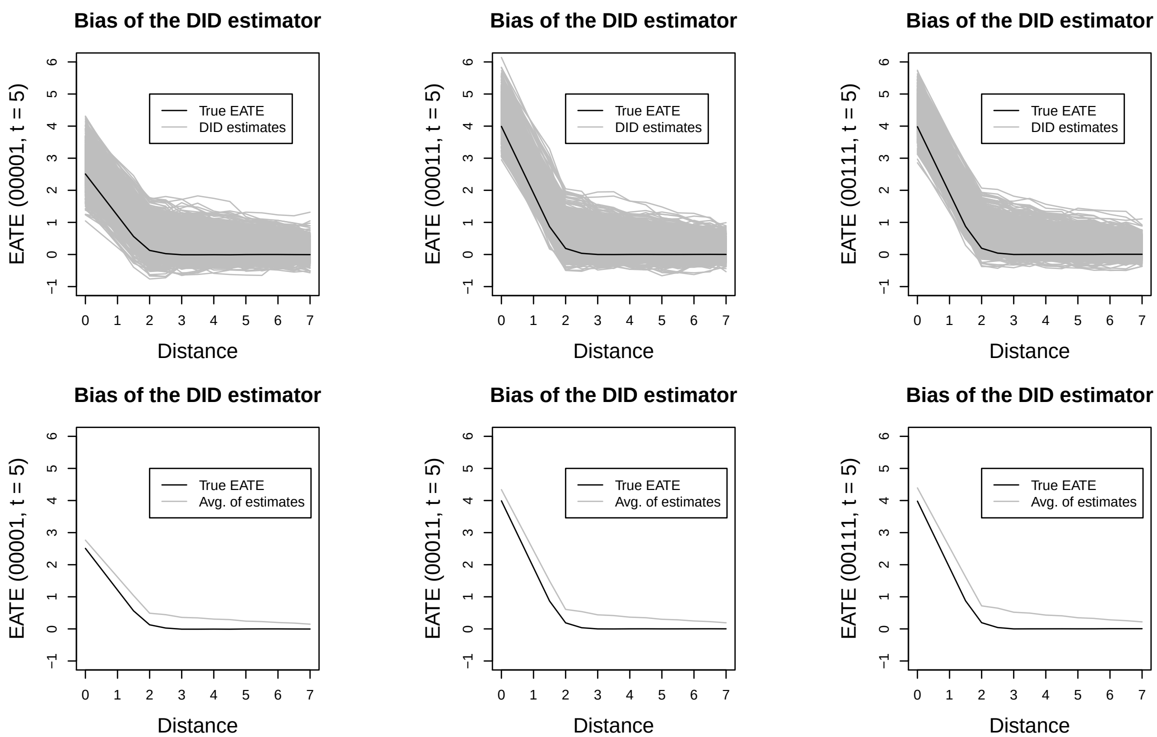


Notes: Figures on the top show the estimates from the Hajek estimator for all the 1000 assignments and figures on the bottom compare the averages of estimates against the EATEs. The difference between the gray and black curves on the bottom is the bias of the estimator. The effect function is monotonic and sequential ignorability holds.

under the assumption of continuous unit fixed effects.

We have more simulation results included in the appendix. They suggest that 1. the IPTW estimators work well under various effect functions, 2. the bias of the DID estimator diminishes when the effect is homogeneous, and 3. the augmented estimators indeed improve the efficiency of estimation. All these results are consistent with our theoretical derivations.

Figure 5: The bias of the DID estimator when unit fixed effects are confounders

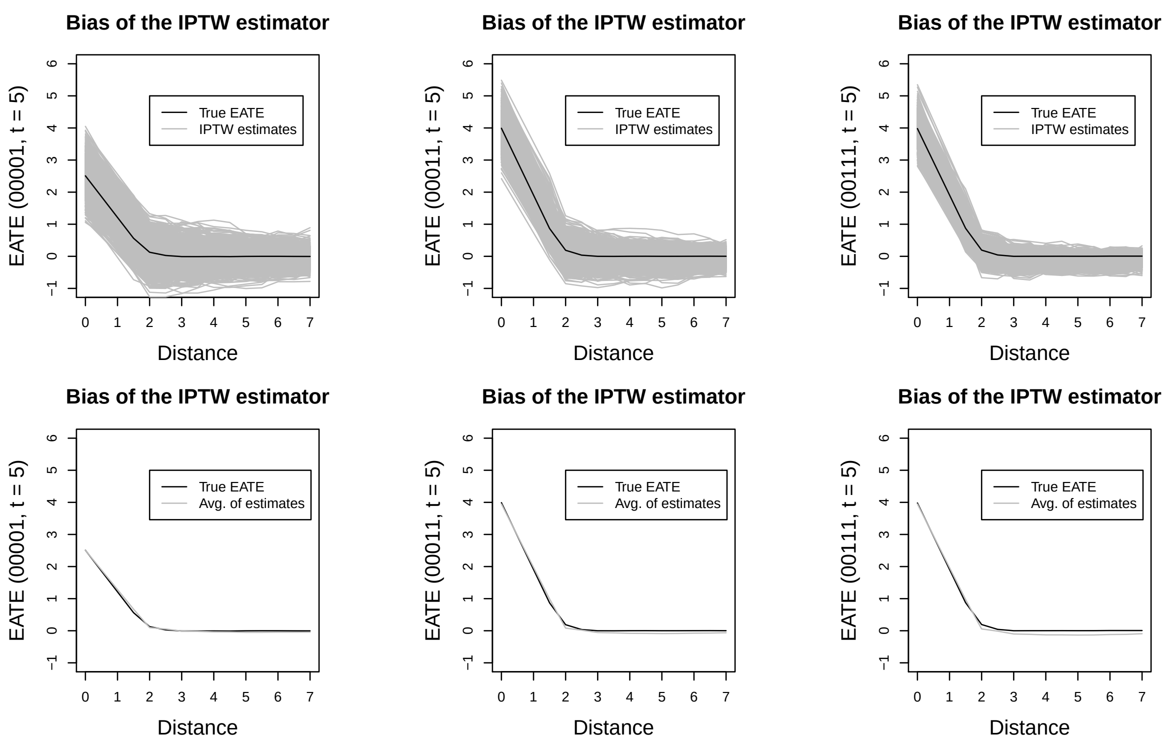


Notes: Figures on the top show the estimates from the DID estimator for all the 1000 assignments and figures on the bottom compare the averages of estimates against the EATEs. The difference between the gray and black curves on the bottom is the bias of the estimator. The effect function is monotonic and sequential ignorability does not hold.

9 Applications

We illustrate the application of the proposed method via two examples. The first example comes from Wang and Wong (2019). In the paper, the authors analyze the impact of Hong

Figure 6: The bias of the IPTW estimator when unit fixed effects are confounders



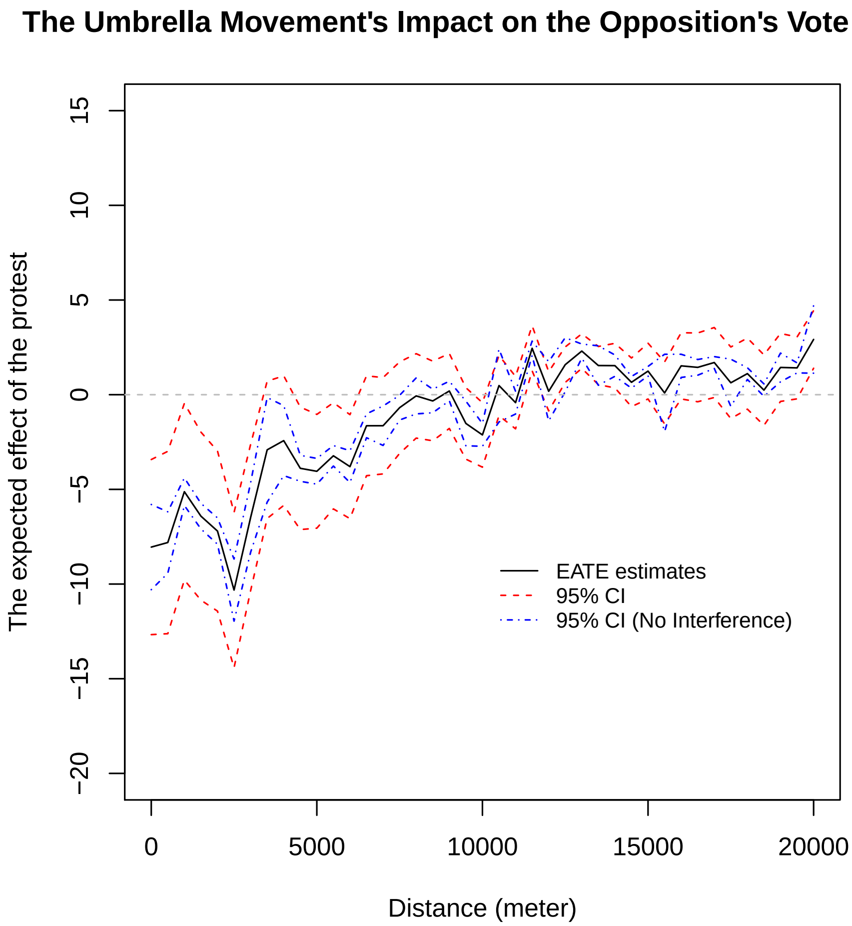
Notes: Figures on the top show the estimates from the Hajek estimator where a polynomial of geographic coordinates is controlled when estimating the propensity scores for all the 1000 assignments and figures on the bottom compare the averages of estimates against the EATEs. The difference between the gray and black curves on the bottom is the bias of the estimator. The effect function is monotonic and sequential ignorability does not hold.

Kong's Umbrella Movement on the result of the following parliamentary election. The pro-democracy movement broke out in 2014, during which protesters occupied public space in 7 out of the city's 401 constituencies. The dataset has a classic two-period DID structure. The outcome is the pro-democracy opposition's vote share in the 2012 and 2016 parliamentary elections. The treatment equals to 0 in 2012 for all the constituencies and equals to 1 only for the occupied constituencies in 2016. It is not surprising that the protest's impact may bleed out of the occupied constituencies, hence spatial interference is a problem. As the dataset has only two periods, we do not need to worry about temporal interference. We use both a logistic model and CBPS to predict the propensity score for each constituency. Since there are only 7 treated units, the estimates of propensity scores may have a large variance and needs stabilization. Thus, we choose the Hajek estimator over the Horvitz-Thompson estimator. The covariates we select when estimating the propensity scores include the opposition's vote share in 2012, percentages of mandarin speakers, male residents, young residents (below 24), old residents (above 65), married residents, college students, trade and financial industry practitioners, rich people (monthly income above 60,000 HKD[25]), poor people (monthly income below 6,000 HKD), and a second-order polynomial of geographic coordinates for each constituency.

The results are presented in Figure 7. The black curve is the estimated EATE of the protest at each distance d. The red dotted curves mark the 95% confidence intervals calculated using the spatial HAC variance estimator and the blue dotted curves represent the wrong 95% confidence intervals when interference is not taken into consideration. Clearly, interference increases the estimates' uncertainty. The result suggests that the protest generated a strong negative impact on the opposition's vote share in the 2016 election. Its magnitude is as large as 6% in constituencies close to the protest sites. The negative effect becomes insignificant only in constituencies that are more than 5 km away from the occupied ones. The findings are similar if we use CBPS to predict the propensity

[25] Roughly \$7740.

Figure 7: Replication of Wang and Wong (2019)



Notes: The black curve is the estimated EATE of the protest at each distance d. The red dotted curves mark the 95% confidence interval at each point when interference is taken into account. The blue dotted curves mark the 95% confidence interval at each point when there is no interference.

scores or apply the augmented estimator[26].

In the original analysis, the authors adopted a model-based approach. They first calculated the minimum distance of each constituency to the occupied ones. Then, they estimated a regression equation of the opposition's vote share change from 2012 to 2016 on the distance and covariates. Although the same negative pattern is detected, such an approach imposes strict assumptions on the data generating process than the proposed estimator. It assumes that the minimum distance is a sufficient statistic for the true effect function and the effect declines at a constant speed with the distance. By contrast, the proposed estimator reaches similar conclusions with more flexibility and fewer assumptions.

The second example is based on Sances (2016). The author examines the effect of an institutional reform in the State of New York. Between 1980 and 2011, many New York towns changed their way to select the real estate assessor, who is in charge of assessing the value of local real estates, from election to appointment. The author argues that appointed assessors have stronger incentives to conduct reassessment. We may suspect that this impact on their incentives spreads to assessors in nearby towns if there is peer effect among them. The proposed method can help us check this possibility.

The dataset has a staggered adoption structure. There are 917 units, 25 periods, and 26 different types of history. In each year, some towns switch from election to appointment. 408 towns use appointment since the first period and 117 towns never change to appointment. We estimate the propensity score for each period after removing units that have been treated in the previous period[27]. The outcome is whether reassessment happened in town i, year t. Covariates we use include the average outcome in the pre-treatment period as well as a second-order polynomial of geographic coordinates for each town. We focus on the cohort with the largest number of units whose treatment starts since period 4. There are 95 units in the cohort. The distance range of interest is set to be between 0 and 100

[26] It seems reasonable to assume that $d(i; d)$ does not grow with N here as constituencies in a city can hardly become denser.

[27] Staggered adoption indicates that the propensity scores are "right censored" in the sense that a unit's propensity score stays at 1 once it is treated.

kilometers.

The upper-left plot in Figure 8 displays how the estimated effects vary across distances and time periods using a heatmap. Each cell represents the estimate for the cumulative average indirect effect, $\tau_t((0^{1:4}, 1^{5:t}), 0^{1:t}; d)$[28], at distance d and period t (when $d = 0$ it is just the estimate of the cumulative average direct effect). Most estimates are close to zero when $d > 0$. But when $d = 0$ and t is large, we see some large positive effects. The same finding can be seen from the upper-right plot of Figure 8, which displays how the estimates vary over time at each distance d. In the bottom plot of Figure 8, we present the average of estimates over periods at each distance d. The 95% confidence intervals are calculated via the two-way HAC variance estimator that accounts for the correlation among observations in both time and space[29]. Only the estimate at $d = 0$ is statistically significant. The coefficient indicates that the probability for having real estate reassessment is around 12% higher in towns with appointed assessors.

The results suggest that there is no salient evidence of spatial interference in the example. It is not really surprising given the setting. Spatial spillover will only appear when assessors do pay attention to each other's action, which may not happen in reality. But the estimates of cumulative effects due to temporal interference are similar to the original estimates[30], indicating that the proposed method does not differ greatly from conventional ones when spatial interference is actually not very strong.

10 Conclusion

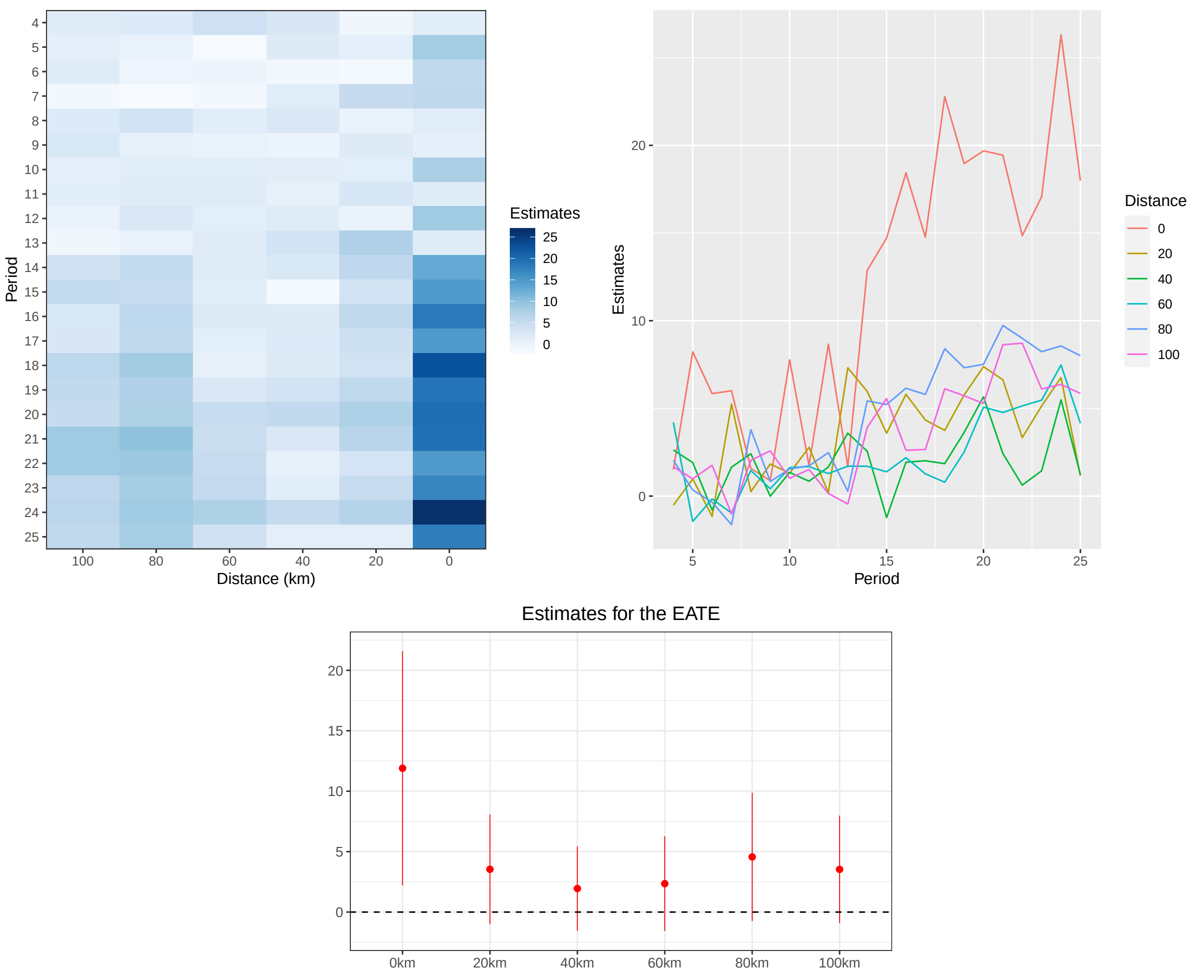
This paper studies the problem of causal inference in TSCS data when both temporal and spatial interference are present. We show that interference between units (spatial

[28] Recall that $(0^{1:4}, 1^{5:t})$ represents the treatment assignment history which equals to 0 in the first four periods and 1 in the following periods.

[29] We define two observations as correlated if their distance is smaller than 100km or the absolute difference in their periods is smaller than 10.

[30] The original estimate is 8.78 for all cohorts while ours is 11.90 for cohort 4. The author actually examines the effect of switching from appointment to election. We recode the treatment to facilitate interpretation.

Figure 8: Replication of Sances (2016)



Notes: The upper-left plot is a heatmap on which the x-axis represents various distance values (from 0 to 100km) and the y-axis is the different time periods. The color of each cell marks the value of the estimate. The upper-right plot presents the dynamics of estimates in each period for various distance values. The bottom plot shows the average of estimates over periods for each distance value. The CIs account for both within- and between-unit correlation among the observations.

interference) makes it improper to only model the outcome variable. Conventional approaches that rely on strict exogeneity, such as DID and the two-way fixed effects model, no longer generate meaningful estimates of the causal effects under interference. The bias of the DID estimator is proportional to the correlation between propensity scores and the individualistic treatment effects.

Instead, under sequential ignorability, we show that there exist a series of IPTW estimators that allow researchers to obtain consistent estimates of the expected average treatment effects generated by any treatment assignment history on the neighboring units. The estimators take the form of either the Horvitz-Thompson estimator or the Hajek estimator with user-chosen spillover mappings as the outcome. They can be augmented for higher efficiency via predicting the spillover mappings more accurately with a diffusion model. The asymptotic distributions of the estimators are derived via combining the theory of semi-parametric models with Stein's method. We further demonstrate that the spatial/two-way HAC variance estimator in econometrics provides a convenient tool to calculate the analytic variances of the estimators.

We illustrate the application of the method via two examples in political science. The first one studies the spillover effect of a protest (Hong Kong's Umbrella Movement) on election results, while the second one examines whether a policy reform in the towns of New York State affects the outcome in nearby areas over periods. We detect the impact of spatial interference in the former and temporal interference in the latter. When spatial interference is not strong, our estimates do not deviate greatly from results generated by conventional approaches.

We expect the proposed method to have wide applications in social sciences. It provides researchers a simple tool to estimate the expected direct effect when both types of interference exist, as well as to examine the spillover effects in both time and space. Compared with the conventional techniques in TSCS data analysis, this method requires stronger assumptions— We must be able to estimate the propensity scores of each observation using its history. Yet the requirement is plausible when the area of interest is not vast and the unobservable attributes vary continuously over it. Future researchers may further explore how to relax the assumptions in larger geographic spaces or social networks with dense ties. It may also worth investigating how to estimate spillover effects under the framework of design-based panel data analysis (Arkhangelsky and Imbens, 2019).

References

Acemoglu, Daron, Camilo García-Jimeno and James A Robinson. 2015. "State capacity and economic development: A network approach." *American Economic Review* 105(8):2364–2409.

An, Weihua and Tyler J VanderWeele. 2019. "Opening the Blackbox of Treatment Interference: Tracing Treatment Diffusion through Network Analysis." *Sociological Methods & Research* p. 0049124119852384.

Angrist, Joshua D. 2014. "The perils of peer effects." *Labour Economics* 30:98–108.

Arkhangelsky, Dmitry and Guido W Imbens. 2019. "Double-robust identification for causal panel data models." *arXiv preprint arXiv:1909.09412* .

Aronow, Peter M and Cyrus Samii. 2016. "Does regression produce representative estimates of causal effects?" *American Journal of Political Science* 60(1):250–267.

Aronow, Peter M and Cyrus Samii. 2017. "Estimating average causal effects under general interference, with application to a social network experiment." *The Annals of Applied Statistics* 11(4):1912–1947.

Aronow, Peter M., Cyrus Samii and Ye Wang. 2019. "Design Based Inference for Spatial Experiments.".

Athey, Susan and Guido W Imbens. 2018. Design-based analysis in difference-in-differences settings with staggered adoption. Technical report National Bureau of Economic Research.

Athey, Susan, Guido W Imbens and Stefan Wager. 2018. "Approximate residual balancing: debiased inference of average treatment effects in high dimensions." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 80(4):597–623.

Athey, Susan, Mohsen Bayati, Nikolay Doudchenko, Guido Imbens and Khashayar Khosravi. 2018. Matrix completion methods for causal panel data models. Technical report National Bureau of Economic Research.

Baird, Sarah, J. Aislinn Bohren, Craig McIntosh and Berk Ozler. 2016. "Optimal Design of Experiments in the Presence of Interference." Typescript, George Washington University, University of Pennsylvania, University of California-San Diego, and the World Bank.

Bang, Heejung and James M. Robins. 2005. "Doubly Robust Estimation in Missing Data and Causal Inference Models." *Biometrics* 61:962–972.

Basse, Guillaume and Avi Feller. 2018. "Analyzing two-stage experiments in the presence of interference." *Journal of the American Statistical Association* 113(521):41–55.

Beck, Nathanial, Kristian Skrede Gleditsch and Kyle Beardsley. 2006. "Space is More than Geography: Using Spatial Ecometrics in the Study of Political Economy." *International Studies Quarterly* 50:27–44.

Blackwell, Matthew. 2013. "A framework for dynamic causal inference in political science." *American Journal of Political Science* 57(2):504–520.

Blackwell, Matthew and Adam N Glynn. 2018. "How to make causal inferences with time-series cross-sectional data under selection on observables." *American Political Science Review* 112(4):1067–1082.

Blume, Lawrence E, William A Brock, Steven N Durlauf and Rajshri Jayaraman. 2015. "Linear social interactions models." *Journal of Political Economy* 123(2):444–496.

Bowers, Jake, Mark M. Fredrickson and Costas Panagopolous. 2013. "Reasoning about Interference Between Units: A General Framework." *Political Analysis* 21(1):97–124.

Bramoullé, Yann, Habiba Djebbari and Bernard Fortin. 2009. "Identification of peer effects through social networks." *Journal of econometrics* 150(1):41–55.

Cameron, A Colin, Jonah B Gelbach and Douglas L Miller. 2012. "Robust inference with multiway clustering." *Journal of Business & Economic Statistics* .

Chatterjee, Sourav. 2014. "A short survey of Stein's method." *preprint arXiv:1404.1392* .

Chatterjee, Sourav et al. 2008. "A new method of normal approximation." *The Annals of Probability* 36(4):1584–1610.

Chin, Alex. 2019. "Central Limit Theorems via Stein's Method for Randomized Experiments Under Interference." *arXiv:1804.03105 [math.ST]* .

Conley, Timothy G. 1999. "GMM estimation with cross sectional dependence." *Journal of Econometrics* 92:1–45.

De Chaisemartin, Clement and Xavier d'Haultfoeuille. 2020. "Two-way fixed effects estimators with heterogeneous treatment effects." *American Economic Review* 110(9):2964–96.

Eckles, Dean, Brian Karrer and Johan Ugander. 2017. "Design and analysis of experiments in networks: Reducing bias from interference." *Journal of Causal Inference* 5(1).

Egami, Naoki. 2018. "Identification of Causal Diffusion Effects Using Stationary Causal Directed Acyclic Graphs." *arXiv preprint arXiv:1810.07858* .

Ertefaie, Ashkan, Nima S Hejazi and Mark J van der Laan. 2020. "Nonparametric inverse probability weighted estimators based on the highly adaptive lasso." *arXiv preprint arXiv:2005.11303* .

Fan, Jianqing, Kosuke Imai, Han Liu, Yang Ning and Xiaolin Yang. 2016. Improving covariate balancing propensity score: A doubly robust and efficient approach. Technical report Technical report, Princeton Univ.

Glynn, Adam N and Kevin M Quinn. 2010. "An introduction to the augmented inverse propensity weighted estimator." *Political analysis* 18(1):36–56.

Goldsmith-Pinkham, Paul and Guido W Imbens. 2013. "Social networks and the identification of peer effects." *Journal of Business & Economic Statistics* 31(3):253–264.

Graham, Bryan S. 2008. "Identifying social interactions through conditional variance restrictions." *Econometrica* 76(3):643–660.

Hirano, Keisuke, Guido W Imbens and Geert Ridder. 2003. "Efficient estimation of average treatment effects using the estimated propensity score." *Econometrica* 71(4):1161–1189.

Hudgens, Michael G. and M. Elizabeth Halloran. 2008. "Toward causal inference with interference." *Journal of the American Statistical Association* 103(482):832–842.

Imai, Kosuke and In Song Kim. 2018. "When Should We Use Linear Fixed Effects Regression Models for Causal Inference with Longitudinal Data." Mimeo, Massachusetts Institute of Technology.

Imai, Kosuke and Marc Ratkovic. 2015. "Robust estimation of inverse probability weights for marginal structural models." *Journal of the American Statistical Association* 110(511):1013–1023.

Kallus, Nathan and Michele Santacatterina. 2018. "Optimal balancing of time-dependent confounders for marginal structural models." *arXiv preprint arXiv:1806.01083* .

LeSage, James and Robert Kelley Pace. 2009. *Introduction to spatial econometrics*. Chapman and Hall/CRC.

Leung, Michael P. 2020. "Treatment and spillover effects under network interference." *Review of Economics and Statistics* 102(2):368–380.

Li, Shuangning and Stefan Wager. 2020. "Random Graph Asymptotics for Treatment Effect Estimation under Network Interference." *arXiv preprint arXiv:2007.13302* .

Liu, Lan and Michael G Hudgens. 2014. "Large sample randomization inference of causal effects in the presence of interference." *Journal of the american statistical association* 109(505):288–301.

Liu, Lan, Michael G Hudgens, Bradley Saul, John D Clemens, Mohammad Ali and Michael E Emch. 2019. "Doubly robust estimation in observational studies with partial interference." *Stat* 8(1):e214.

Liu, Lan, Michael G Hudgens and Sylvia Becker-Dreps. 2016. "On inverse probability-weighted estimators in the presence of interference." *Biometrika* 103(4):829–842.

Liu, Licheng, Ye Wang and Yiqing Xu. 2019. A Practical Guide to Counterfacutal Estimators for Causal Inference with Time-Series Cross-Sectional Data. Technical report Working Paper, Stanford University.

Manski, Charles F. 1995. *Identification Problems in the Social Sciences*. Cambridge, MA: Harvard University Press.

Manski, Charles F. 2012. "Identification of treatment response with social interactions." *The Econometrics Journal* (In press).

Neyman, Jerzy. 1923. "On the Application of Probability Theory to Agricultural Experiments: Essay on Principles." *Statistical Science* 5:465–80. Section 9 (translated in 1990).

Ogburn, Elizabeth L, Oleg Sofrygin, Ivan Diaz and Mark J van der Laan. 2020. "Causal inference for social network data." *arXiv preprint arXiv:1705.08527* .

Ogburn, Elizabeth L and Tyler J VanderWeele. 2017. "Vaccines, contagion, and social networks." *The Annals of Applied Statistics* 11(2):919–948.

Paluck, Elizabeth Levy, Hana Shepherd and Peter M. Aronow. 2016. "Changing climates of conflict: A social network experiment in 56 schools." *Proceedings of the National Academy of Science* 113(3):566–571.

Papadogeorgou, Georgia, Kosuke Imai, Jason Lyall and Fan Li. 2020. "Causal Inference with Spatio-temporal Data: Estimating the Effects of Airstrikes on Insurgent Violence in Iraq." *arXiv preprint arXiv:2003.13555* .

Robins, James. 1986. "A new approach to causal inference in mortality studies with a sustained exposure period?application to control of the healthy worker survivor effect." *Mathematical modelling* 7(9-12):1393–1512.

Robins, James M., Miguel A. Hernan and Babette Brumback. 2000. "Marginal structural models and causal inference in epidemiology." *Epidemiology* 11(5):550–560.

Ross, Nathan. 2011. "Fundamentals of Stein?s method." *Probability Surveys* 8:210–293.

Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments n Randomized and nonrandomized Studies." *Journal of Educational Psychology* 5(66):688–701.

Sances, Michael W. 2016. "The distributional impact of greater responsiveness: Evidence from New York towns." *The Journal of Politics* 78(1):105–119.

Savje, Fredrik, Peter M. Aronow and Michael G. Hudgens. 2018. "Average treatment effects in the presence of unknown interference." *arXiv:1711.06399 [math.ST]* .

Sinclair, Betsy, Margaret McConnell and Donald P Green. 2012. "Detecting spillover effects: Design and analysis of multilevel experiments." *American Journal of Political Science* 56(4):1055–1069.

Stein, Charles. 1972. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Probability Theory*. The Regents of the University of California.

Strezhnev, Anton. 2018. "Semiparametric weighting estimators for multi-period difference-in-differences designs." Working Paper, Harvard University.

Tchetgen-Tchetgen, Eric J. and Tyler J VanderWeele. 2010. "On causal inference in the presence of interference." *Statistical Methods in Medical Research* 21(1):55–75.

van der Laan, Mark J. 2014. "Causal inference for a population of causally connected units." *Journal of Causal Inference J. Causal Infer.* 2(1):13–74.

Van der Vaart, Aad W. 2000. *Asymptotic statistics*. Vol. 3 Cambridge university press.

Wang, Ye and Stan Hok-Wui Wong. 2019. "Electoral Impacts of Failed Revolutions: Evidence from Hong Kong?s Umbrella Movement." unpublished manuscript.

Xu, Yiqing. 2017. "Generalized synthetic control method: Causal inference with interactive fixed effects models." *Political Analysis* 25(1):57–76.

Zigler, Corwin M. and Georgia Papadogeorgou. 2018. "Bipartite Causal Inference with Interference." *arXiv:1807.08660 [stat.ME]* .