Causal Inference under Temporal and Spatial Interference

*Ye Wang**

February 15, 2020

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 exercise, I propose a semi-parametric approach to estimate the direct and indirect/spillover treatment effect of any event or policy under the sequential ignorability assumption, when both temporal and spatial interference are present. The estimator is shown to be unbiased, consistent, and normally distributed if the degree of interference does not grow too fast with the sample size. The conventional difference-in-differences (DID) or two-way fixed effects approach, nevertheless, leads to biased estimates in this scenario. I apply the method to examine the impact of Hong Kong's Umbrella Movement on election results and how an institutional reform in New York state affects real estate assessment.

*PhD Student, Department of Politics, New York University.

1. Introduction

To understand the impact of some social policy or event is often the goal of empirical studies in social sciences. In many cases, the impact is assumed to reside within the unit of analysis. In the conventional time-series cross-sectional (TSCS) data analysis, for example, researchers usually assume that only some units (the treatment group) are affected by the treatment, but not the others (the control group). The validity of such assumptions, nevertheless, is not always guaranteed. If one state cuts its corporation tax, the productivity of companies in the neighboring states will also be influenced. Once a protest breaks out, all the people living nearby might be inspired. Such a "spillover effect" may even persist into the future. In other words, the treatment imposed to unit i at period t can affect the outcome of unit j at period t'. Statistically speaking, we have interference both over time and across units. As the units of interest are often located on a geographical space, I call the latter spatial interference and the former temporal interference.

In this paper, I investigate the problem of identifying causal relationships when both types of interference are present. Approaches that deal with either spatial or temporal interference have been developed by previous studies (Hudgens and Halloran, 2008; Aronow and Samii, 2015; Blackwell and Glynn, 2014). Yet as far as I know, there is little guidance for practitioners on how to conduct causal inference under both of them. In addition, many methods impose stringent restrictions on the structure of spillover— for example, spillover only occur among adjacent units— that are neither practical nor testable. Although progress has been made to handle general interference under which arbitrary spillover is allowed to exist, most works concentrate on experimental designs rather than observational studies (Savje, Aronow and Hudgens, 2018; Aronow, Samii and Wang, 2019; Chin, 2019), not to mention the scenario with the time dimension.

By drawing a connection between the literature on TSCS data analysis and experimental design with interference, this paper provides a systematic discussion on what causal

conclusions could be reached under common assumptions on TSCS datasets. It takes a design-based perspective and imposes no structural restriction on the heterogeneity of treatment effects or the functional form of interference. I start from introducing a series of novel causal quantities to capture the effect generated by a unit's treatment status history on its own outcome as well as on the outcome of other units, based on the idea of marginalization. Generalizing from "circle mean" defined in Aronow, Samii and Wang (2019), I propose that researchers can construct theory-driven "spillover mapping" to aggregate each unit's influence on the others into meaningful estimands. Essentially, these quantities describe the expected difference brought forth by the change of treatment assignment history at a representative unit on its neighborhood in any particular time period.

I then show that these estimands can be non-parametrically identified and unbiasedly estimated via a semi-parametric inverse probability of treatment weighting (IPTW) estimator under the assumption of sequential ignorability, which requires that the propensity score of unit i at period t can be fully predicted by its own history. Estimation proceeds by first predicting the propensity score via statistical models or covariate balancing (Imai and Ratkovic, 2015; Kallus and Santacatterina, 2018), and then plugging the predicted values into a Horvitz-Thompson/Hajek estimator with the spillover mapping's value as the outcome. The estimator can be augmented by employing a diffusion model to predict the spillover mapping with higher accuracy. Unbiasedness is achieved when either the propensity score or the diffusion model is correctly specified.

I further exploit Stein's method to prove that the proposed estimators are consistent and asymptotically normal when the degree of interference does not increase too fast with the sample size (Stein et al., 1972; Chatterjee et al., 2008; Chatterjee, 2014; Chin, 2019). Variance estimates could be obtained either analytically or via resampling techniques. The analytic variance estimate is shown to be asymptotically equivalent to the spatial HAC variance

(Conley, 1999). If unobservable attributes, or "fixed effects" exist, however, the introduced estimands are identifiable only under strong (and often implausible) homogeneity assumptions. Otherwise, the proposed estimator converges to a weighted sum of individualistic causal quantities and leads to bias even with infinite sample.

The framework could be extended to incorporate more possibilities. For example, researchers can account for the contagion of the treatment by revising the propensity score model accordingly. The estimator preserves its properties as long as the model is correctly specified. After adding more restrictions on the outcomes, we may also estimate the lag effect of treatments in previous periods, as in Acharya, Blackwell and Sen (2016). Finally, we can replace the geographical space with a social network, and measure the distance among units with the length of path that connect them. I demonstrate that such a setting requires stronger assumptions as the "small world" phenomenon tends to make the degree of interference grow too rapidly.

The proposed method works well in both simulations and real world examples. For the purpose of illustration, I 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. I use this example to compare the difference between design-based and outcome-based approaches, and show that the former can generate similar results under weaker assumptions. The replication confirms the authors' conclusion that the protest reduces the opposition's vote share in constituencies that are 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 the state of New York. This study has multiple periods and a staggered adoption data structure. I find little evidence of spatial interference in this case, but estimates on the time dimension are consistent with the original results.

The rest of the paper is organized as follows: Section two summarizes extant studies that

are correlated to the current work. Section three describes the basic framework, including causal quantities of interest and identification assumptions. Section four discusses identification under different assumptions. Section five introduces the proposed estimators and their statistical properties. Section six presents some simulation results. Section seven illustrates the application of the method with two empirical studies and section eight concludes.

2. Related studies

Interference, also called "spillover effect," "peer effect," or "diffusion" in social sciences, refers to the phenomenon that one observation's outcome is 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 in period t relies its own treatment in period t'. For simplicity, I use spatial interference to call the former, and temporal interference to call the latter. I also distinguish interference from contagion, which means that one observation's outcome is affected by the outcome of other observations (Ogburn, VanderWeele et al., 2017).

This paper is built upon the recently burgeoning literature on experimental design with interference. Most early works in this field explicitly specify the structure of interference. For example, Hudgens and Halloran (2008) proposed the split-plot design with "partial interference" in which they allow for arbitrary interference within strata but none between strata. They also distinguished the treatment's direct effect from the indirect effect caused by interference and suggested a variance estimator. This setting is 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 of social networks, Aronow and Samii (2015) suggested an inverse probability of

treatment weighting (IPTW) estimator under the assumption that the "exposure mapping" is known to the researcher. Such an approach is also extended by following works (Paluck, Shepherd and Aronow, 2016; Ogburn et al., 2017; An and VanderWeele, 2019).

Realizing that these structural assumptions are not realistic, more and more scholars are now working under the assumption of general interference in which one's outcome may be affected by the entire treatment assignment vector 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) claimed that the difference-in-means estimator converges to the expected average treatment effect (EATE) when the degree of interference satisfies certain conditions in any general experiment design. Chin (2019) exploited Stein's method to show that the same estimator is normally distributed in large samples. Closest to the current work, Aronow, Samii and Wang (2019) investigated general interference effects 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 expected effect generated by the treatment at distance d. Nevertheless, they focused only on experimental designs with a bipartite structure where treatment assignment and outcome measurement occur at two different levels. This paper will generalize their idea to one-partite observational studies with the time dimension.

Most extant attempts to account for spatial interference in observational studies rely on the model-based approach, such as linear and generalized linear models (Manski, 1995, 2012; Bowers, Fredrickson and Panagopolous, 2013; Blume et al., 2015), spatial econometric models (Beck, Gleditsch and Beardsley, 2006; LeSage and Pace, 2009), and network models (Graham, 2008; Leung, 2016; Acemoglu, García-Jimeno and Robinson, 2015). The fundamental problem of the model-based approach is that social interactions

can be too complicated to be captured by any single model, as pointed out by Angrist (2014). Liu, Hudgens and Becker-Dreps (2016) is one of the pioneers to explicate general interference in observational studies. The authors suggested a Hajek estimator and proved its statistical properties. Zigler and Papadogeorgou (2018) discussed the estimation of indirect effect in bipartite designs. The authors applied their method to estimate the effect of pollution reduction technology in US power plants on cardiovascular hospitalization in nearby areas.

Techniques to handle temporal interference are developed almost independently in the literature. Motivated by dynamic experiments in medicine where treatment in the next stage is determined by the treatment and the outcome up to now, biostatisticians have been studying cases where one's outcome is jointly decided by one's treatment assignment in history since Robins (1986). On the basis of sequential ignorability, Robins, Hernan and Brumback (2000) proposed the marginal structural models (MSM), in which researchers can exploit IPTW to balance the probability for a particular history to appear and obtain unbiased estimate of the history's accumulative effect. Bang and Robins (2005) further modified MSM and made it into a doubly robust estimator. These methods have been introduced into social sciences by Blackwell (2013) and Blackwell and Glynn (2018). But model-based approach, especially two-way fixed effects models, is still the first option for social scientists to analyze TSCS data. Differing from sequential ignorability, these models typically assume that all error terms are "strictly exogenous" to the treatment assignment in history after conditioning on some unobservable factors, and generate causal estimates by adjusting the response surface.

But as pointed out by Imai and Kim (2018), the two-way fixed effects model implicitly assumes that the outcome cannot be affected by past history, thus does not allow for temporal interference. Otherwise the estimates will be biased. Strezhnev (2018) further demonstrates 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 for some DID estimators will be violated, which leads to the bias. The only exception is when the data have a generalized DID or staggered adoption structure (Athey and Imbens, 2018). Then the cumulative treatment effect on the treated can be estimated via the idea of counterfactual estimation (Xu, 2017; Strezhnev, 2018; Athey et al., 2018; Liu, Wang and Xu, 2019), where the model is used to impute the missing counterfactual outcome.

This paper aims at building a bridge between these separate branches of literature, and focuses on TSCS data analysis when both temporal and spatial interference exist. 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 sections, I will discuss the construction and identification of causal quantities under either the sequential exogeneity assumption or the strict exogeneity assumption.

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, $\mathcal{X}$. Denote the geographic location of a unit 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 upper-case characters (e.g. Y_{it}) to represent random variables and lower-case ones to represent their specific value (e.g. y_{it}).

For simplicity, we denote a random variable or its value in a particular period with subscript and their past history with superscript. For example, $\mathbf{Y}_i^t = (Y_{i1}, Y_{i2}, \ldots, Y_{it})$

and $\mathbf{Z}_i^t = (Z_{i1}, Z_{i2}, \ldots, Z_{it})$ 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'. To avoid confusion, we sometimes also use $i.$ and $.t$ to represent a vector of variables. For example, $\tau_{.t} = (\tau_{1t}, \tau_{2t}, \ldots, \tau_{Nt})$.

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: $Y_{it} = Y_{it}(\mathbf{Z}^T) = Y_{it}(\mathbf{Z}_1^T, \mathbf{Z}_2^T, \ldots, \mathbf{Z}_N^T)$. 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 we have the classic stable unit treatment value assumption (SUTVA). With only temporal interference, $Y_{it} = Y_{it}(\mathbf{Z}_i^T) = Y_{it}(Z_{i1}, Z_{i2}, \ldots, Z_{iT})$. And with only spatial interference, $Y_{it} = Y_{it}(\mathbf{Z}_t) = Y_{it}(Z_{1t}, Z_{2t}, \ldots, Z_{Nt})$.

We measure the degree of interference between any two units using the dependency graph. Treating each unit in the sample as a node in a graph, then a link between node i and node j, I_{ij}, exists when $Y_{it}(\mathbf{Z}^T) \neq Y_{it}(\tilde{\mathbf{Z}}^T)$ for some t, such that $\mathbf{Z}^T \setminus \mathbf{Z}_j^T = \tilde{\mathbf{Z}}^T \setminus \mathbf{Z}_j^T$. Here $\mathbf{Z}^T \setminus \mathbf{Z}_j^T$ is the treatment history of all the units except j. We also assume that a unit always interferes with itself. In other words,

$$I_{ij} = \begin{cases} 1, \text{ if } i = j; \\ 1, \text{ if } Y_{it}(\mathbf{Z}^T) \neq Y_{it}(\tilde{\mathbf{Z}}^T) \text{ for some } t, \text{ and } \mathbf{Z}^T \setminus \mathbf{Z}_j^T = \tilde{\mathbf{Z}}^T \setminus \mathbf{Z}_j^T; \\ 0, \text{ otherwise.} \end{cases}$$

The dependency graph is often not observable to researchers and cannot be explicitly modeled. There are different ways to measure the average degree of interference. For example, we can define the average interference dependence as:

$$d_{AVG} = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} d_{ij}, \text{ where } d_{ij} = \begin{cases} 1 \text{ if } I_{il}I_{jl} = 1 \text{ for some } l \\ 0 \text{ otherwise} \end{cases}$$

Here $d_{ij} = 1$ if unit i and j interfere with a common unit l and d_{AVG} is the average of d_{ij}. We can also define $c_i = \sum_{j=1}^{N} I_{ij}$, which is the number of units that interfere with unit i, and $C_p = \left[\frac{1}{N}\sum_{i=1}^{N} c_i^p\right]^{\frac{1}{p}}$, which is the $p-$norm of c_i. Savje, Aronow and Hudgens (2018) discuss the relationship among these measures in details.

3.2 Define causal quantities

We are interested in the causal effect generated by the change of the history of treatment assignment. For each unit i in period t, we define the individualistic treatment effect as:

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

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

The existence of spatial interference makes a unit's outcome dependent on the treatment of other units. Therefore, we define the individualistic contemporary indirect effect as $\tau_{it}(\mathbf{Z}^T \setminus Z_{jt}) = Y_{it}(1, \mathbf{Z}^T \setminus Z_{jt}) - Y_{it}(0, \mathbf{Z}^T \setminus Z_{jt})$. It is the effect of observation (j,t)'s treatment on

the outcome of unit i in the same period. Finally, we use the individualistic cumulative indirect effect to describe the cumulative impact of unit j's history on unit i's outcome, which includes the effect of both temporal and spatial interference: $\tau_{it}(\mathbf{z}_j^{s:t}, \tilde{\mathbf{z}}_j^{s:t}; \mathbf{Z}^T \setminus \mathbf{Z}_j^{s:t}) = Y_{it}(\mathbf{z}_j^{s:t}, \mathbf{Z}^T \setminus \mathbf{Z}_j^{s:t}) - Y_{it}(\tilde{\mathbf{z}}_j^{s:t}, \mathbf{Z}^T \setminus \mathbf{Z}_j^{s:t})$. One common choice is $\tilde{\mathbf{z}}_j^{s:t} = \mathbf{0}_j^{s:t}$. Then the quantity measures the effect of $\mathbf{z}_j^{s:t}$ relative to the status quo on unit i, period t.

It is not difficult to notice that all the quantities defined above are random variables since their values rely 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:

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

respectively.

Even after marginalization, none of these individualistic quantities is identifiable from dataset. 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:

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

These quantities are known as expected average treatment effect (EATE) in the literature. They reflect the effect generated by the change of the variable of interest when other relevant variables are fixed at their average value.

To construct average indirect effects, we introduce the concept of "spillover mapping", a mapping μ from $\mathcal{R}^N$ to $\mathcal{R}$ that captures the influence of one unit on the others. The form of μ is chosen by the researcher under theoretical guidance. One example of the spillover mapping is the "circle mean" defined in Aronow, Samii and Wang (2019): $\mu_i(\mathbf{x}; d) = \frac{\sum_{j=1}^{N} \mathbf{1}\{d_{ji}=d\} x_j}{\sum_{j=1}^{N} \mathbf{1}\{d_{ji}=d\}}$. Here μ_i is the value of μ for each unit i. 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 neighbor to unit i. For simplicity, we focus on the circle mean in the following text. The discussion on general spillover mappings can be seen in the appendix. In this case, the average contemporary/cumulative indirect effect can be defined as:

$$\tau_t(d) = \frac{1}{N}\sum_{i=1}^{N} \mu_i(\tau^i_{\cdot t}; d) = \frac{1}{N}\sum_{i=1}^{N} \frac{\sum_{j=1}^{N} \mathbf{1}\{d_{ji}=d\}\tau^i_{jt}}{\sum_{j=1}^{N} \mathbf{1}\{d_{ji}=d\}}$$

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

They describe the average difference brought to units that are d distance units away when the treatment status/history at a unit is changed[1]. The four average causal quantities are the main estimands of this paper, although other estimands are possible. Note that all the other three effects can be considered as special cases of the average cumulative indirect effect with either $s = t$ or $d = 0$. Hence, our analysis focuses on the average cumulative indirect effect. The result naturally generalizes to the other three estimands. We will proceed to discuss when they are identifiable and how to estimate them.

[1]Clearly, the interpretation hinges on the choice of spillover mapping.

3.3 Assumptions

Following the literature on TSCS data analysis, we make the following assumptions:

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

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

for any i and t.

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

Assumption 2 *Sequential ignorability*

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

and

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

for any i and t.

This is the most critical assumption for our analysis. It suggests that the treatment in period t is "as-if" randomly assigned conditional on past treatment assignment, past outcome, 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 common in TSCS analysis. Such an assumption is largely motivated by dynamic experiments in medical science and biostatistics. Its advantage is that we can exploit the information conveyed by history

to estimate the propensity score, $P(\mathbf{Z}_t = \mathbf{z}_t|\mathbf{Z}_i^{t-1}, \mathbf{Y}_i^{t-1}, \mathbf{X}_i^t)$, which plays a crucial role in identification when interference is a serious concern.

But unlike in the difference-in-differences (DID) analysis or fixed effects model, no unobservable confounder is allowed to exist under sequential ignorability. Therefore, if both the outcome and the true propensity are affected by some unobservable variable, the assumption will no longer hold. However, such a concern is partly alleviated in our case as we possess the geo-location of all the units. Suppose the unit fixed effects vary continuously on the surface $\mathcal{X}$, then their influence could be well approximated by a polynomial of each unit's coordinates. We will further discuss this issue in following sections. In addition, we will show that the conventional approaches such as DID or fixed effects models under the strict exogeneity assumption no longer leads to meaningful estimates when spatial interference exists.

Theoretically speaking, Assumption 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 dimension is high and no extra assumption is made. We thus impose the following two assumptions on the treatment assignment mechanism.

Assumption 3 *Bernoulli design*

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

The first part of the assumption is usually implicit in observational studies. We write it down to emphasize that Z_{it} can be dependent on its own history, $\mathbf{Z}_i^{t-1}$, but not on the assignment 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 $\mathbf{Z}_j^{t-1}$ to Z_{it}. Now $\mathbf{Z}_j^{t-1}$ can only affect Z_{it} via $\mathbf{Y}_i^{t-1}$. In other words, the set of confounder excludes $\mathbf{Z}_j^{t-1}$, but may include $\mathbf{Z}_i^{t-1}$ and $\mathbf{Y}_i^{t-1}$. This assumption reduces the dimension of the propensity score. We will explore how to relax it in the extension section.

Under assumptions 1-4, we can write the propensity score for each observation (i, t) as: $e_{it} = P(Z_{it} = 1|\mathbf{Z}_i^{t-1}, \mathbf{Y}_i^{t-1}, \mathbf{X}_i^t)$. The propensity score for the entire history of treatment assignment for unit i up to period t, $\mathbf{Z}_i^t$, can be expressed as $W_{it} = \prod_{s=1}^{t} P(Z_{is}|\mathbf{Z}_i^{s-1}, \mathbf{Y}_i^{s-1}, \mathbf{X}_i^s)$. The propensity score's expression can be further simplified by imposing more restrictions, such as it is only decided by the treatment status and outcome value in the previous period. Then we will have $W_{it} = \prod_{s=1}^{t} P(Z_{is}|\mathbf{Z}_{i,s-1}, \mathbf{Y}_{i,s-1}, \mathbf{X}_{is})$.

4. Identification

4.1 Identification without unobservable confounders

We first discuss identification when all the confounders are observable, that is, when assumptions 1-4 are satisfied. In this case, all the causal quantities defined in the previous section are non-parametrically identifiable from data. For the average contemporary direct effect, it is easy to show that:

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

Notice that all the quantities within the expectation sign can be constructed from data. Intuitively, after reweighting with its propensity score, each outcome value is an unbiased estimate of the expected outcome that marginalizes over the treatment assignment of other

observations. Similar results hold for the other three estimands. To save space, we only present the result for the average cumulative indirect effect at distance d:

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

where W_{it} is the propensity score for the history $\tilde{\mathbf{z}}^{s:t}$ and $Y_{.t}(\mathbf{Z}^t) = (Y_{1t}(\mathbf{Z}^t), Y_{2t}(\mathbf{Z}^t), \ldots, Y_{Nt}(\mathbf{Z}^t))$. Again, all the quantities can be constructed from data. We formally state the results as the following theorem:

Theorem 1

Under assumptions 1-4, average contemporary direct effect, average cumulative direct effect, average contemporary indirect effect, average cumulative indirect effect are all non-parametrically identifiable.

4.2 Identification with unobservable confounders

As discussed, another common approach in TSCS data analysis is the fixed effects models (including DID as a special case). Their identification relies on the strict exogeneity assumption. It typically indicates the mean independence of error terms to the treatment variable conditional on unobservable variables that do not vary with time, the unit fixed effects. Since these confounders are not observable, we are no longer able to estimate the propensity score for each observation. Nevertheless, if we assume that SUTVA holds and the fixed effects influence the outcome variable in a linear way, subtracting the unit-specific average from each observation will completely eliminate the impact of these unobservables

on the response surface. Then the difference between the treated and control observations can be purely attributed to the treatment, which results in an unbiased estimate of the average treatment effect on the treated (ATT).

This is also the idea behind most model-based methods in TSCS data analysis. Instead of considering the assignment mechanism, 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 and mean independence, enable researchers to obtain unbiased estimates for the ATT without knowing the propensity score of each observation. 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 hinges on untreated observations. Yet when spatial interference is also present, the model-based approach can no longer return any result of substantive interest. Without information on the propensity score, we do not know to what extent the response surface of different groups contaminates each other.

To see the point more clearly, let's consider a simple case where there are two groups of units and two periods. The treatment group receives the treatment Z_i 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:

$$
\begin{aligned}
Y_{i1} &= \mu + \alpha_i + \xi_1 + \varepsilon_{i1} \\
Y_{i2} &= \mu + g_i(\mathbf{Z}) + \alpha_i + \xi_2 + \varepsilon_{i2}
\end{aligned}
$$

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 in period 2 on unit i. We assume that $\mathrm{E}\left[\varepsilon_{it}|\alpha_i, \xi_t, \mathbf{Z}\right] = 0$, where the expectation is taken over samplings. 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 DID:

$$\mathrm{E}\left[Y_{i2}(\mathbf{0}^2) - Y_{i1}(\mathbf{0}^2)|Z_i = 1\right] = \mathrm{E}\left[Y_{i2}(\mathbf{0}^2) - Y_{i1}(\mathbf{0}^2)|Z_i = 0\right].$$

The conventional DID estimator in this case is equivalent to the within estimator in two-way fixed effects models, and its value can be expressed as:

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

The quantity in the last line is the estimand we have defined, the average contemporary direct effect. Note that the sample average is replaced by population mean to make the setting consistent with the convention. It is easy to check that the estimate converges to the estimand only if p_i and $\mathrm{E}_{\mathbf{Z}\setminus Z_i} g_i(z_i, \mathbf{Z}\setminus Z_i)$ are uncorrelated, for example, when $\mathrm{E}_{\mathbf{Z}\setminus Z_i} g_i(z_i, \mathbf{Z}\setminus Z_i)$ is homogeneous. In practice, however, both of them could be decided by the unit fixed effect α_i and correlation exists by nature. In this case, we know the bias of the DID estimate equals to:

$$\begin{aligned}
Bias =& \tau_{DID} - \tau_2(d) \\
=& \frac{\mathrm{E}\left[p_i \mathrm{E}_{\mathbf{Z}\setminus Z_i} g_i(1, \mathbf{Z}\setminus Z_i)\right]}{\mathrm{E}\, p_i} - \frac{\mathrm{E}\left[(1-p_i)\mathrm{E}_{\mathbf{Z}\setminus Z_i} g_i(0, \mathbf{Z}\setminus Z_i)\right]}{\mathrm{E}\,(1-p_i)} \\
& - \mathrm{E}\left[\mathrm{E}_{\mathbf{Z}\setminus Z_i} g_i(1, \mathbf{Z}\setminus Z_i)\right] + \mathrm{E}\left[\mathrm{E}_{\mathbf{Z}\setminus Z_i} g_i(0, \mathbf{Z}\setminus Z_i)\right] \\
=& \frac{\mathrm{Cov}\left(p_i, \mathrm{E}_{\mathbf{Z}\setminus Z_i} g_i(1, \mathbf{Z}\setminus Z_i)\right)}{\mathrm{E}\, p_i} - \frac{\mathrm{Cov}\left(1-p_i, \mathrm{E}_{\mathbf{Z}\setminus Z_i} g_i(0, \mathbf{Z}\setminus Z_i)\right)}{\mathrm{E}\,(1-p_i)} \\
=& \frac{\mathrm{Cov}\left(p_i, \mathrm{E}_{\mathbf{Z}\setminus Z_i} g_i(1, \mathbf{Z}\setminus Z_i)\right)}{\mathrm{E}\, p_i} + \frac{\mathrm{Cov}\left(p_i, \mathrm{E}_{\mathbf{Z}\setminus Z_i} g_i(0, \mathbf{Z}\setminus Z_i)\right)}{\mathrm{E}\,(1-p_i)}
\end{aligned}$$

The direction of the bias depends on the correlation between the unobservable propensity

score p_i and the effect function's expectation. When both of them are positively correlated with α_i, for instance, DID will overestimate the average contemporary direct effect. Otherwise it will underestimate the effect.

The problem is similar to the one addressed by Aronow and Samii (2016) in the cross-sectional setting. But unlike in their case, p_i here is unknown, thus we cannot adjust the weights to obtain unbiased estimate. The result 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 response surface $E[g_i(\mathbf{Z})|\mathbf{Z}^t, \mathbf{Y}^{t-1}, \mathbf{X}^t]$. But such a model can hardly be correct in reality.

Another solution to the problem is to strengthen the assumption as in Arkhangelsky and Imbens (2019). The authors argue that when the assignment process belongs to some specific function class, there exist observable sufficient statistics S_i such that $\varepsilon_{it} \perp \mathbf{Z}|S_i$ for any i, t. For example, when $E[Z_{it}|\alpha_i] = \frac{exp(\alpha_i+\xi_t)}{1+exp(\alpha_i+\xi_t)}$, we have $\mathbf{Z}_i^T \perp \mathbf{Y}_{jt}(\mathbf{Z}^T)|\frac{1}{T}\sum_{t=1}^T Z_{it}$. Then we can unbiasedly estimate the propensity score for each unit and obtain correct causal estimates. But I do not explore these results in details here.

5. Estimation and inference

5.1 The IPTW estimator

Under assumptions 1-4, there exists a natural estimator for each of the four estimands—the quantity within the expectation sign in the results on identification. For the average contemporary direct effect and the average cumulative indirect effect, the estimators are:

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

and

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

respectively. Estimators for other estimands can be constructed in a similar way.

Both estimators have the same Horvitzh-Thompson form as the spatial estimator proposed by Aronow, Samii and Wang (2019). The only difference here is that both the propensity score and the spillover mapping's value are unknown and have to be estimated from data. Treating them as nuisance parameters, the estimator clearly belongs to the class of semi-parametric estimators whose statistical properties have been well studied (Van der Vaart, 2000). The spillover mapping's value can be approximated by its sample analogue. For propensity score, we may fit a simple logistic model or employ more advanced techniques such as covariates balancing propensity score proposed by Imai and Ratkovic (2015) and kernel optimal weighting developed by Kallus and Santacatterina (2018).

There is a well-known problem of the IPTW estimator: When the estimation of propensity scores is not very precise, the estimate deviates greatly from the true value in practice. Therefore, researchers may "stabilize" the propensity score estimates using the Hajek estimator to replace the Horvitz-Thompson estimator. For the average cumulative indirect effect, it has the following form:

$$\begin{aligned}\hat{\tau}_{t,HA}(\mathbf{z}^{s:t}, \tilde{\mathbf{z}}^{s:t}; d) = &\frac{\sum_{i=1}^{N} \mathbf{1}\{\mathbf{Z}_i^{s:t} = \mathbf{z}_i^{s:t}\}\hat{\mu}_i(Y_{.t}(\mathbf{Z}^t); d)/\hat{P}(\mathbf{Z}_i^{s:t} = \mathbf{z}_i^{s:t})}{\sum_{i=1}^{N} \mathbf{1}\{\mathbf{Z}_i^{s:t} = \mathbf{z}_i^{s:t}\}/\hat{P}(\mathbf{Z}_i^{s:t} = \mathbf{z}_i^{s:t})} \\ &- \frac{\sum_{i=1}^{N} \mathbf{1}\{\mathbf{Z}_i^{s:t} = \tilde{\mathbf{z}}^{s:t}\}\hat{\mu}_i(Y_{.t}(\mathbf{Z}^t); d)/\hat{P}(\mathbf{Z}_i^{s:t} = \tilde{\mathbf{z}}^{s:t})}{\sum_{i=1}^{N} \mathbf{1}\{\mathbf{Z}_i^{s:t} = \mathbf{z}_i^{s:t}\}/\hat{P}(\mathbf{Z}_i^{s:t} = \tilde{\mathbf{z}}^{s:t})}\end{aligned}$$

From the result on identification, it is straightforward to see that the Horvitzh-Thompson estimator is unbiased when we know the true value of nuisance parameters. Otherwise, the bias diminishes to zero as N grows, if the nuisance parameters can be consistently

estimated. Asymptotic unbiasedness also holds for the Hajek estimator. The consistency of these estimators requires more assumptions. Intuitively, we need their variance to gradually decrease to zero as $N \to \infty$. Yet this may not happen when the degree of interference is too strong. Suppose each unit interferes with all the others, we actually have only one observation and the variance does not change with the sample size. Therefore, consistency is guaranteed only when degree of interference is limited:

Assumption 5 *Local interference*

Denote $\max_i c_i$ *as* $\tilde{c}$*, then* $\frac{\tilde{c}}{N^{\frac{1}{3}}} \to 0$ *as* $N \to \infty$*.*

Remember that c_i is the number of units that interfere with unit i. The assumption says that the maximal number of units each unit interferes with, $\tilde{c}$, grows not as fast as the sample size (only with the speed of $o(N^{1/3})$). In other words, the dependency graph becomes "bigger" with a larger sample, but not much "denser." From the perspective of Chin (2019), we can distinguish strong interference from weak interference. The latter gradually vanishes as N increases while the former does not. Then the assumption suggests that each unit has strong interference with at most $\tilde{c}$ other units. That's why it is called "local interference." A similar assumption is made by Aronow, Samii and Wang (2019). Yet the current one is weaker as it allows for arbitrary weak interference. Finally, as when SUTVA holds, we need some regularity conditions on the population moments:

Assumption 6 *Bounded potential outcomes*

There exist a constant $\tilde{b}$ *such that* $|Y_{it}(\mathbf{z}^t)| < \tilde{b}$ *for all* i*,* t*, and* $\mathbf{z}^t$*.*

Assumption 7 *Lipschitz condition of the spillover mapping*

There exist a constant $L > 0$ *such that* $||\mu(\mathbf{Y}(\mathbf{z}^t); d) - \mu(\tilde{\mathbf{Y}}(\mathbf{z}^t); d)|| < L||\mathbf{Y}(\mathbf{z}^t) - \tilde{\mathbf{Y}}(\mathbf{z}^t)||$*.*

These conditions ensure the existence of all the population moments. Obviously the circle mean satisfies assumption 7.

With these extra assumptions, we are able to show that the variance of both the Horvitz-

Thompson and the Hajek estimator converges to zero in the limit. Intuitively, the variance can be decomposed into three parts: the conventional Neyman variance without interference, the variance induced by spatial interference and the variance from estimating nuisance parameters. As we only care about effects in each period, there is no variance due to temporal interference. It is easy to see that the first part is infinitesimal. The third part converges to zero under the correct model specification. The second is asymptotically equivalent to the spatial HAC variance. When spatial interference is local, it is also infinitesimal. The expression of the variance is given in the next subsection and the proof is in the appendix. We formally state the result below:

Theorem 2

Under assumptions 1-7, when the estimates of nuisance parameters are consistent for their true value, both the Horvitz-Thompson and the Hajek estimator are asymptotically unbiased and consistent.

5.2 The augmented estimator

The estimator proposed above can be "augmented" by approximating the response surface, Y_{it} and μ_i, more precisely. The idea is to fit a "diffusion model" for $E[Y_{it}|\mathbf{Z}^t, \mathbf{Y}^{t-1}, \mathbf{X}^t]$ and plug the fitted value $\hat{Y}_{it}$ into the estimator. For the spillover mapping, we also have the fitted value $\hat{\mu}_i = \mu_i(\hat{Y}_{.t}(\mathbf{Z}^t); d)$. For example, we may guess that the true effect function is homogeneous, additive and proportional to the distance to a treated unit: $g_{it}(\mathbf{Z}_t) = \sum_{d=1}^{\tilde{d}} \beta_d \sum_{j=1}^{N} Z_{jt}\mathbf{1}\{d_{ij} = d\}$. Further assume that the effect is not cumulative in time, we can use the following model to predict Y_{it}:

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

where h is an arbitrary function. After fitting the model (using parametric models or non-parametric techniques), we can obtain the fitted values $\hat{Y}_{it}$.

If the above model captures the true data generating process, then β_d equals to the average contemporary indirect effect, τ_d. 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. As we have discussed in the identification section, though, it is highly unlikely that we can find the true data generating process for $E[Y_{it}|\mathbf{Z}^t, \mathbf{Y}^{t-1}, \mathbf{X}^t]$ (just see how many restrictions we impose on $g_{it}(\mathbf{Z}_t)$ above!). Even if the selected model is close to the true DGP, we still need correct propensity scores to marginalize the effects when the effect function is not additive. Yet suppose the model is wrong, it still improves the efficiency of estimation. One may interpret the estimator from the perspective of residual balancing (Glynn and Quinn, 2010; Liu et al., 2019; Athey, Imbens and Wager, 2018): We use the diffusion model to reduce noises and then apply IPTW to balance out the bias generated by the model. The augmented estimators for the average contemporary direct effect and the average cumulative indirect effect are as follows:

$$\hat{\tau}_{t,Aug} = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{Z_{it}\left(Y_{it}-\hat{Y}_{it}^{1}\right)}{\hat{P}(Z_{it}=1)}+\hat{Y}_{it}^{1}\right]-\frac{1}{N}\left[\sum_{i=1}^{N}\frac{(1-Z_{it})\left(Y_{it}-\hat{Y}_{it}^{0}\right)}{1-\hat{P}(Z_{it}=1)}+\hat{Y}_{it}^{0}\right]$$

$$\begin{aligned}\hat{\tau}_{t,Aug}(\mathbf{z}^{s:t},\tilde{\mathbf{z}}^{s:t};d) = &\frac{1}{N}\sum_{i=1}^{N}\left[\frac{\mathbf{1}\{\mathbf{Z}_i^{s:t}=\mathbf{z}_i^{s:t}\}\left[\mu_i(Y_{.t}(\mathbf{Z}^t);d)-\mu_i(\hat{Y}_{.t}^{1}(\mathbf{Z}^t);d)\right]}{\hat{P}(\mathbf{Z}_i^{s:t}=\mathbf{z}_i^{s:t})}+\mu_i(\hat{Y}_{.t}^{1}(\mathbf{Z}^t);d)\right]\\ &-\frac{1}{N}\sum_{i=1}^{N}\left[\frac{\mathbf{1}\{\mathbf{Z}_i^{s:t}=\tilde{\mathbf{z}}^{s:t}\}\left[\mu_i(Y_{.t}(\mathbf{Z}^t);d)-\mu_i(\hat{Y}_{.t}^{0}(\mathbf{Z}^t);d)\right]}{\hat{P}(\mathbf{Z}_i^{s:t}=\tilde{\mathbf{z}}^{s:t})}+\mu_i(\hat{Y}_{.t}^{0}(\mathbf{Z}^t);d)\right]\end{aligned}$$

When the dataset has a generalized DID or staggered adoption structure, one may simply set $\hat{Y}_{it}$ to be the average of Y_{is} in the pre-treatment periods (which clearly would not be a correct model). Then the augmented estimator has a similar form as the estimator

proposed by Strezhnev (2018) under the conditional parallel trends assumption. We denote the last period before unit i's treatment kicks off as K_i. Units with the same value of K_i are called a cohort and all the units with $K_i = \infty$ belong to the control group. Notice that K_i is a sufficient statistic of the assignment history $\mathbf{Z}_i^t$, thus $P(\mathbf{Z}_i^t = \mathbf{z}_i^t)$ equals to $P(K_i = k)$. Then the estimator is:

$$\begin{aligned}\hat{\tau}_{t,Aug}(\mathbf{z}^{s:t}, \tilde{\mathbf{z}}^{s:t}; d) =& \frac{1}{N}\sum_{i=1}^{N}\frac{\mathbf{1}\{K_i = k\}}{\hat{P}(K_i = k)}\left[\hat{\mu}_i(Y_{.t}(\mathbf{Z}^t); d) - \frac{1}{k}\sum_{s=1}^{k}\hat{\mu}_i(Y_{.s}(\mathbf{Z}^t); d)\right] \\ &- \frac{1}{N}\sum_{i=1}^{N}\frac{\mathbf{1}\{K_i \geq k\}}{P(K_i \geq k)}\left[\hat{\mu}_i(Y_{.t}(\mathbf{Z}^t); d) - \frac{1}{k}\sum_{s=1}^{k}\hat{\mu}_i(Y_{.s}(\mathbf{Z}^t); d)\right] \\ &+ \frac{1}{k}\sum_{s=1}^{k}\mathbf{1}\{K_i = k\}\hat{\mu}_i(Y_{.s}(\mathbf{Z}^t); d) - \frac{1}{k}\sum_{s=1}^{k}\mathbf{1}\{K_i \geq k\}\hat{\mu}_i(Y_{.s}(\mathbf{Z}^t); d).\end{aligned}$$

The estimator proposed by Strezhnev (2018) equals to the one above without the last term. But we should keep in mind that the two estimators are built on different assumptions and they converge to different types of estimand (cohort ATT and marginalized ATE). Compared with the IPTW estimator, this augmented version has more nuisance parameters. But it is also one of the semi-parametric estimators hence the statistical properties are the same:

Theorem 3

Under assumptions 1-7, when the estimates of either the propensity score or the response surface is consistent for its true value, the augmented estimator is asymptotically unbiased and consistent.

5.3 Inference and asymptotic distribution

In this section I present the main analytic result of the paper and show that each $\hat{\tau}$ defined above is normally distributed. The proof is based on the combination of the theory on

semi-parametric models with Stein's method. The method was originally proposed by Stein et al. (1972) to show that a random variable obeys the normal distribution if it satisfies a particular differential equation. It was further developed into the "generalized perturbative approach" by Chatterjee et al. (2008) and Chin (2019) to establish a central limit theorem when there is interdependence among observations. Essentially, asymptotic normality holds as long as the estimator is robust to small perturbation of the treatment assignment, which translates into limited interference among the units. Aronow, Samii and Wang (2019) have utilized the approach to prove the asymptotic normality of the spatial estimator. For simplicity, I only present the result for the Hajek estimator of the average cumulative indirect effect, which is the most general case. The result for the augmented estimator is similar as only the outcome is modified. When the true value of the propensity scores is known, we have the following proposition:

Theorem 4

Under assumptions 1-7, when the true value of the propensity scores is known, the proposed Hajek estimator of the average cumulative indirect effect converges to the normal distribution:

$$\sqrt{N}(\hat{\tau}_{d,HA} - \tau) \to N(0, V_d).$$

where V_d is the asymptotic variance of $\hat{\tau}_{d,HA}$.

I show that the variance consists of two parts, $V_d = V_{1d} + V_{2d}$, each of which can be expressed as follows:

$$V_{1d} = \frac{1}{N^2}\sum_{i=1}^{N}\frac{\mathrm{E}\left[\left(\mu_i(Y_{.t}(\mathbf{z}_i^{s:t}, Z^t \setminus \mathbf{Z}_i^{s:t}); d) - \bar{\mu}^0_{\mathbf{z}_i^{s:t}}\right)^2\right]}{P(\mathbf{Z}_i^{s:t} = \mathbf{z}_i^{s:t})} + \frac{1}{N^2}\sum_{i=1}^{N}\frac{\mathrm{E}\left[\left(\mu_i(Y_{.t}(\tilde{\mathbf{z}}_i^{s:t}, Z^t \setminus \mathbf{Z}_i^{s:t}); d) - \bar{\mu}^0_{\tilde{\mathbf{z}}_i^{s:t}}\right)^2\right]}{P(\mathbf{Z}_i^{s:t} = \tilde{\mathbf{z}}_i^{s:t})}$$

$$
\begin{aligned}
V_{2d} = \frac{1}{N^2} \sum_{i=1} \sum_{j \in \mathbf{B}(i)} \{ & \text{Cov}\left[\mu_i(Y_{.t}(\mathbf{z}_i^{s:t}, Z^t \setminus \mathbf{Z}_i^{s:t}); d), \mu_j(Y_{.t}(\mathbf{z}_i^{s:t}, Z^t \setminus \mathbf{Z}_i^{s:t}); d)\right] \\
& - \text{Cov}\left[\mu_i(Y_{.t}(\mathbf{z}_i^{s:t}, Z^t \setminus \mathbf{Z}_i^{s:t}); d), \mu_j(Y_{.t}(\tilde{\mathbf{z}}_i^{s:t}, Z^t \setminus \mathbf{Z}_i^{s:t}); d)\right] \\
& - \text{Cov}\left[\mu_i(Y_{.t}(\tilde{\mathbf{z}}_i^{s:t}, Z^t \setminus \mathbf{Z}_i^{s:t}); d), \mu_j(Y_{.t}(\mathbf{z}_i^{s:t}, Z^t \setminus \mathbf{Z}_i^{s:t}); d)\right] \\
& + \text{Cov}\left[\mu_i(Y_{.t}(\tilde{\mathbf{z}}_i^{s:t}, Z^t \setminus \mathbf{Z}_i^{s:t}); d), \mu_j(Y_{.t}(\tilde{\mathbf{z}}_i^{s:t}, Z^t \setminus \mathbf{Z}_i^{s:t}); d)\right]\}.
\end{aligned}
$$

where

$$
\begin{aligned}
\bar{\mu}^0_{\mathbf{z}_i^{s:t}} &= \frac{1}{N} \sum_{i=1}^{N} \frac{\text{E}\left[\mu_i(Y_{.t}(\mathbf{z}_i^{s:t}, Z^t \setminus \mathbf{Z}_i^{s:t}); d)\right]}{P(\mathbf{Z}_i^{s:t} = \mathbf{z}_i^{s:t})} \\
\bar{\mu}^0_{\tilde{\mathbf{z}}_i^{s:t}} &= \frac{1}{N} \sum_{i=1}^{N} \frac{\text{E}\left[\mu_i(Y_{.t}(\tilde{\mathbf{z}}_i^{s:t}, Z^t \setminus \mathbf{Z}_i^{s:t}); d)\right]}{P(\mathbf{Z}_i^{s:t} = \tilde{\mathbf{z}}_i^{s:t})}.
\end{aligned}
$$

As mentioned in the previous subsection, the first part is the classic Neyman variance. The second part is the variance caused by spatial interference. To estimate V, we rely on a regression representation of the estimator. Previous studies have noticed that there exists the following equivalent form of the Hajek estimator:

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

I show that OLS estimation of the regression equation returns the same result as the Hajek estimator, and its spatial HAC variance estimate is asymptotically unbiased for the Hajek estimator's variance, V.

When the propensity score has to be estimated, the same conclusion holds when the model for the propensity score is correct. Now we need to estimate the regression form and the propensity score simultaneously. Each of the equations is a moment condition. Therefore, we are solving a moment estimation problem with (possibly) infinite-dimensional parameters. As we already know the asymptotic distribution of the Hajek estimator, the proof can be completed via using the conclusion in Chapter 8 of Newey and McFadden (1994).

Theorem 5

Under assumptions 1-7, when the estimate for the propensity score converges to its true value at a speed faster than or equal to $o_P(N^{\frac{1}{2}})$, the proposed estimators converge to the normal distribution:

$$\sqrt{N}(\hat{\tau}_{d,HA} - \tau_d) \rightarrow N(0, \tilde{V}_d).$$

where $\tilde{V}_d = V_d + G_{d,P}^{-1} V_P$; V_P is the variance of the propensity score's estimate; $G_{d,P}$ is the partial derivative of the Hajek estimator's moment condition on the propensity score.

Researchers can also use the Fisher randomization test or bootstrap to generate confidence interval. Since the asymptotic distribution is normal, these resampling methods are valid.

6. Extensions

6.1 Contagious treatment assignment

In reality, policy or event that appeared in one place may not only affect the outcome, but also the probability for the same policy or event to emerge in another place. States often copy each other's policy innovation, and a small-scale protest sometimes develops into a storm of revolution. In this case, our Z_{it} is a function of not just the unit's own history, but also the past treatment assignment or even past outcome of other units: We have interference from $Z_{jt'}$ to Y_{it}, as well as contagion from $Z_{jt'}$ to Z_{it}. The basic framework does not have to change much to accommodate the existence of contagion. The only part that requires modification is the propensity score, which must account for the contagion of treatment assignment.

Various models have been proposed in both statistics and social sciences to describe contagion. For example, in biostatistics, scholars rely on the SIR (Susceptible-Infected-Recovered) model to approximate the contagion of disease. In political science, Egami

(2018) developed an approach using static causal graphs to identify the causal contagion effect. Researchers may use their substantive knowledge to select the model they prefer. As long as the model satisfies our requirement of the propensity score, the main results are still valid.

6.2 Identify the lag effect

Except for the cumulative effect of the treatment assignment history, researchers may also be interested in the "lag effect" 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 $t' < t$. 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 treatment effect heterogeneity, it is impossible to distinguish the lag effects from the heterogeneous effects in any period, as discussed in Acharya, Blackwell and Sen (2016). In the stagger 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. Assume the first period effect is the same for both cohorts, 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 must be cautious if they try to make claims on the lag effect.

6.3 Apply the method to network setting

Our framework can also be applied to social network analysis, in which the Euclidean or Geodesic distance on the geographic surface is replaced by the length of the shortest path between any pair of nodes. All the analyses will remain the same. The dependency

graph now is just the graph that represents the network. Nevertheless, the assumption on local interference will be harder to satisfy in this case, due to the famous "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 local interference is still ensured. But it may no longer be the case after some weak ties, or "bridges," are added into the graph.

7. Simulation

In this section, we test the performance of the proposed estimator via simulation. Our simulated dataset has 400 units and 5 periods. Treatment starts from the third period. The potential outcome when there is no treatment, $Y_{it}(\mathbf{0}^T)$, is generated as follows:

$$Y_{it}(\mathbf{0}^T) = \mu + X_{1,it} + 3 * X_{2,it} + \alpha_i + \xi_t + \varepsilon_{it}.$$

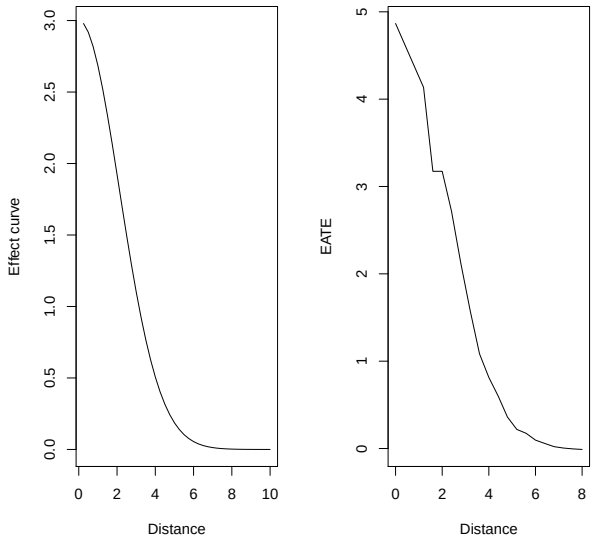
where μ is the grand intercept; $X_{1,it}$ and $X_{2,it}$ are two covariates; α_i and ξ_t represent unit and period fixed effects, respectively; ε_{it} is the idiosyncratic shock. We set $\mu = 5$, and assume that all the other variables are normally distributed except for α_i, which is the absolute value of a normal variable.

We assume that the exists an effect function emanating from each unit whose value changes monotonically with the distance (gradually declines) to that unit. Effects from different units are additive. The total effect each unit receives is positively correlated with its fixed effect α_i. Effects in the previous periods have a lag effect with the discount rate equal to

0.3. Treatment assignment in each period is decided by the outcome and treatment status in the previous period, but not the fixed effects. To be specific, we have:

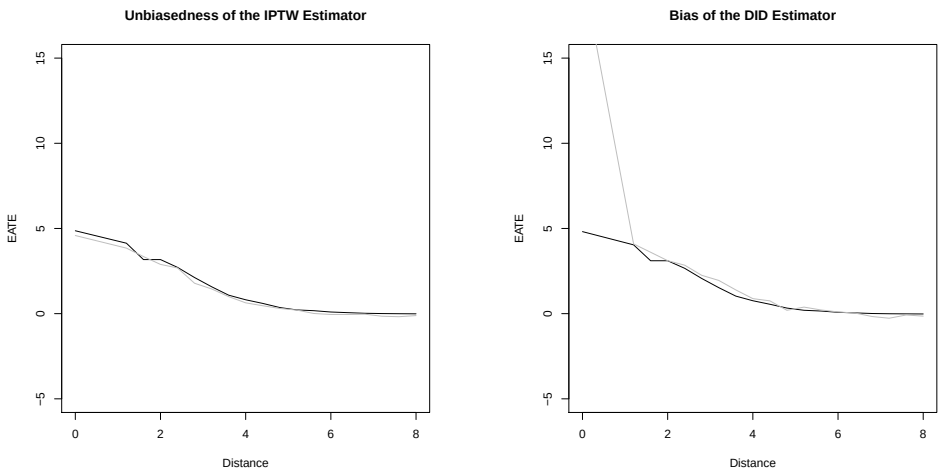
$$P(Z_{it} = 1) = Logit(-2 - X_{1,it} + X_{2,it} + 0.5 * Y_{i,t-1} + 0.2 * Z_{i,t-1}).$$

Figure 1 shows the effect function and the corresponding expected ATE in period 3 at each distance d. The EATE is larger as it is amplified by the value of fixed effects.



(Figure 1)

We first check the unbiasedness of the proposed estimator for the EATE in period 3 and compare it against the DID estimator. We run the simulation for 1000 times and plot the results together with the EATE curve. On the left of Figure we present results from the IPTW estimator and on the right results from the DID estimator. Clearly, the IPTW estimator is unbiased while the DID estimator is not. The bias is large and positive when d is small because both the effect's magnitude and the propensity score are positively correlated with the fixed effects.



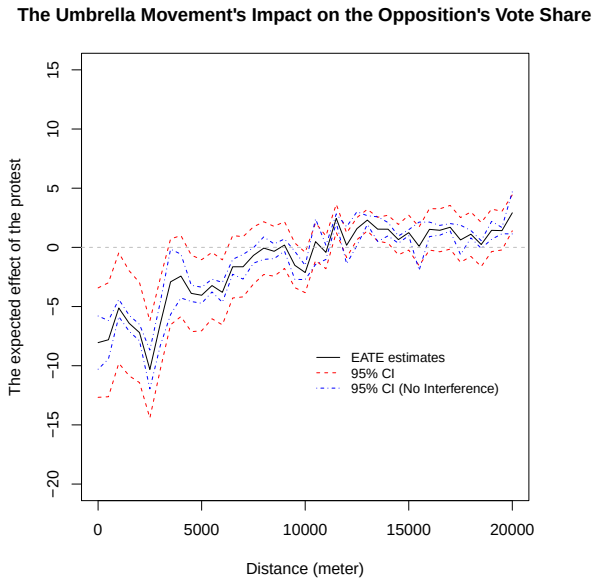
(Figure 2)

The IPTW estimator is also unbiased for the cumulative effect in period 4 and 5. The difference between the estimates and the EATEs gradually vanishes as N becomes bigger, suggesting that the estimator is indeed consistent. We report these results in the appendix.

8. Application

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 Kong's Umbrella Movement on the election result two years later. 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 election. 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 travel beyond the occupied constituencies thus spatial interference is a real concern. As the dataset has only two periods, we do not need to account for temporal interference. We first use a simple logistic model to predict the propensity score for each constituency and

apply the Hajek estimator. Since there are only 7 treated units, the estimate of propensity score may have a large variance and needs stabilization. The covariates we use when estimating the propensity score include the opposition's vote share in 2012, percentage of mandarin speakers, male residents, young residents, old residents, married residents, college students, trade and financial industry practitioners, rich people, poor people, and a second-order polynomial of geographic coordinates for each constituency.



(Figure 3)

The result is presented in Figure 3. The black curve is the estimated effect of the protest at each distance. The red dotted lines mark the 95% confidence interval and the blue dotted lines represent the wrong 95% confidence interval when interference is not taken into account. Clearly, interference increases the estimates' uncertainty. From the graph, we can see that the protest generated a strong negative impact on the opposition's vote share in the 2016 election. Its magnitude is as large as 10% in constituencies near the protest sites. The negative effect becomes no longer significant only in constituencies that are more than 5 km away from the occupied ones. The result is similar if we use CBPS to predict the propensity scores or apply the augmented estimator.

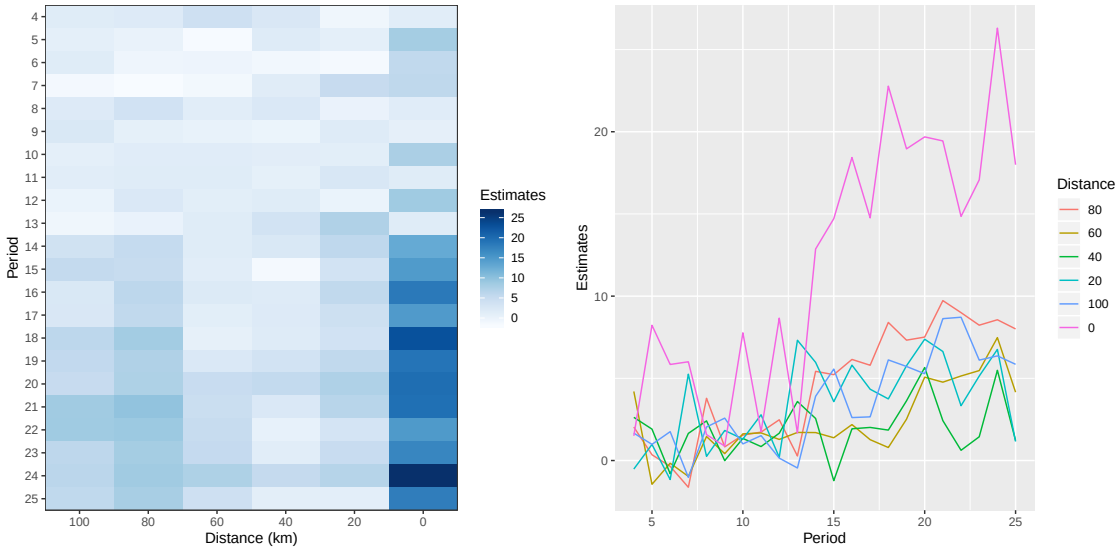
In the original analysis, the authors adopted a model-based approach. They first calculated the minimum distance of each constituency to the occupied ones. Then they estimated a regression equation of the opposition's vote share change from 2012 to 2016 on the distance and covariates. Although the same negative pattern is detected, such an approach imposes more 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 the conclusion 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 any local coordination 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. 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. The distance of interest is set to be between 0 and 100 kilometers.

The first plot in Figure 4 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, $\tau_t((\mathbf{0}^{1:4}, \mathbf{1}^{4:t}), \mathbf{0}^{1:t}; d)$, 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 $d > 0$ and insignificant. But when $d = 0$ and t is large, we see some large positive effects. The same finding can be seen from the second plot of Figure 4, which displays how the estimates vary over time at each distance d.



(Figure 4)

The results suggest that there is no strong evidence of spatial interference in this case. It is not really surprising given the setting. Spatial spillover will only exist when assessors do coordinate, which may not happen in practice. But the estimates of cumulative effects due to temporal interference are similar to the original estimates, indicating that the proposed method does not differ greatly from conventional ones when no spatial interference is actually present.

9. Conclusion

This paper studies causal inference in TSCS data when both temporal and spatial interference are present. I show that interference between units makes it no longer proper to model only the response surface. Conventional methods such difference-in-differences analysis

and two-way fixed effects models thus do not generate meaningful estimate of the average treatment effect in this case. Instead, when sequential inexorability is assumed, there exist semi-parametric estimators that allow researchers to obtain unbiased and consistent estimate of the expected average treatment effect generated by a particular treatment assignment history. The estimators take the form of either the Horvitz-Thompson estimator or the Hajek estimator with theory-driven spillover mappings as the outcome. They can be augmented to obtain higher efficiency by predicting the spillover mappings more accurately using a diffusion model. The asymptotic distribution of the estimators is derived via combining the theory of semi-parametric models with Stein's method. The analytic variance is equivalent to the summation of the Neyman variance and the spatial HAC variance.

The proposed method works well in both simulations and real-world examples. It provides researchers a way to estimate the conventional ATE when both types of interference exist, as well as a tool to examine the spillover effect in both time and space. Compared with the conventional model-based approaches in TSCS data analysis, this method requires stronger assumptions that we must be able to estimate the propensity score using the history of each unit. Yet the requirement is plausible when the area of interest is not vast and the propensity score varies continuously over it. Future researchers may further explore how to relax the assumptions in larger geographic spaces and social networks with dense ties. It may also worth investigating how to estimate spillover effects under the design-based panel data (Arkhangelsky and Imbens, 2019) framework.

Appendix

(In progress)

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.

Acharya, Avidit, Matthew Blackwell and Maya Sen. 2016. "Explaining Causal Findings without Bias: Detecting and Assessing Direct Effects." *American Political Science Review* (In Press).

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. 2015. "Estimating Average Causal Effects Under General Interference." Manuscript, Yale University and New York University.

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., 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 Glynn. 2014. "How to Make Causal Inferences with Time-Series Cross-Sectional Data." Unpublished Manuscript, Harvard University.

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.

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.

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* .

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

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

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. 2016. "Treatment and spillover effects under network interference." *Review of Economics and Statistics* pp. 1–42.

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).

Newey, KW and D McFadden. 1994. "Large sample estimation and hypothesis." *Handbook of Econometrics, IV, Edited by RF Engle and DL McFadden* pp. 2112–2245.

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. 2017. "Causal inference for social network data." *arXiv preprint arXiv:1705.08527* .

Ogburn, Elizabeth L, Tyler J VanderWeele et al. 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.

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.

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 et al. 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 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]* .