Bias-Variance Tradeoffs for Designing Simultaneous Temporal Experiments

Ruoxuan Xiong Alex Chin Sean Taylor Susan Athey

May 9, 2022

Abstract

We study the design of simultaneous temporal experiments, where a set of interventions are applied concurrently in continuous time and outcomes are measured on a sequence of events observed in time. The motivating setting is multiple data science teams on a ride-hailing platform simultaneously and independently test changes to marketplace algorithms such as pricing and matching, and estimate effects from observed event outcomes such as the rate at which ride requests are completed. The design problem involves partitioning a continuous space (time) into intervals and assigning treatments at the interval level. Design and analysis must account for three factors: spillovers from interventions at earlier times, correlation in event outcomes, and effects of interventions tested simultaneously. We derive estimators for error components in a highly general setting and build intuition and guidance for practitioners via a careful simulation study.

1 Introduction

In a variety of empirical settings, it is useful to estimate the effects of interventions via time-based or *temporal* experimental designs, rather than (the far more common) cross-sectional designs. Most prominently, heuristic designs colloquially known as "switchbacks" have become popular due to their applications in digital marketplaces. In these modern settings, the interference structure between units is difficult to account for and can cause bias of unknown sign and large magnitude using more traditional approaches. Prior to more recent applications, there is a long history in medicine of designing an experiment using a single unit of observation and leveraging longitudinal observations in medicine where it is known as an "n-of-1" trial (Mirza et al. 2017).

As motivation for the present work, we consider the problem of designing multiple simultaneous temporal experiments, for instance, in a ride-hailing company where multiple teams would like to measure the effects of their product changes with only a small number of available treatment units (e.g., cities or regions). In a dynamic two-sided marketplace, users exposed to new pricing and matching algorithms may change their behavior in ways that affect outcomes for other users on either side of the marketplace. There are a variety of causal mechanisms for these spillovers, such as riders consuming available drivers, relocating drivers, or stimulating drivers to drive for longer or shorter periods of time (Chamandy 2016).

Given the importance of digital marketplaces and the well-acknowledged need to rapidly test new ideas, the design of experiments that provide reliable estimates in the presence of marketplace-mediated interference has drawn increasing attention in recent studies (Holtz et al. 2020; Li et al. 2021; Basse and Feller 2018; Jagadeesan et al. 2020; Johari et al. 2020). A common theme of these approaches is exploiting prior knowledge of the spillover mechanisms, and leveraging this structure to provide alternative analysis procedures or designs.

We study design of experiments in a highly generic setting where interventions are applied in a continuous temporal space, and outcomes are measured on a sequence of events in this space. Good designs in this setting efficiently partition continuous temporal space into intervals with alternating treatments, in anticipation of precisely estimating a quantity we call the global average treatment effects (GATE) of interventions from the observed event outcomes. GATE is an important estimand for decision-making that captures the difference in average outcomes when an intervention is deployed indefinitely (global treatment) versus when the intervention is absent indefinitely (global control).

Our goal is to capture realistic properties of this empirical setting that complicate the design and analysis of temporal experiments. First, we account for spillover effects between treatments and the outcomes of future events. Second, we account for correlation in event outcomes from unobserved (or unmodeled) factors that create nuisance dependence among measurements; outcomes close in time can be similar due to weather, traffic, or other external factors. Correlations do not have to be monotonic in the distance between events, as they can display periodic behavior in weekly or daily cycles. Third, we account for the irregular density of observed events, corresponding to the property that there is strong periodicity in interactions with marketplaces. Finally, we consider the presence of simultaneous experiments run by other teams on the same sequence of events, which can confound effect estimates in finite samples.

Figure 1 introduces the causal structure for our empirical setting. An experimental design is an assignment of $W_\ell(t)$ which are pure parent nodes, affecting the outcome of events occurring at Y_t. A latent variable U causes nuisance correlations in all events, yielding dependence in the observations.

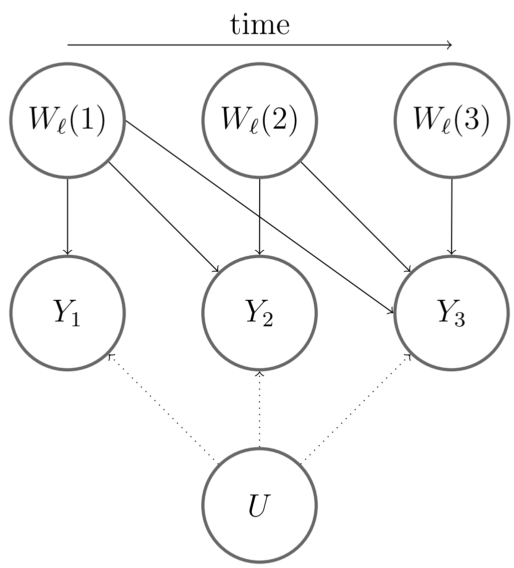


Figure 1: Directed acyclic graph characterizing our empirical setting. $W_\ell(t)$ denote a vector of interventions applied to time t, which affect events Y_t. These direct effects are indicated by solid lines. Past interventions may affect all future events and are observed with nuisance variation (dotted lines) caused by unobserved cause U. The equal spacing of the time of observation is for exposition and is not assumed.

To fix a ride-hailing example, consider $W_p(t)$ as the price charged for a request observed at time t and let Y_t denote whether the rider's session results in a ride request. The direct effect of an intervention on price is intuitive and immediate, but if the ride request occurs, then the effect in a future session may not, due to diminished supply of available drivers, representing the arrows from $W_p(1)$ to Y_2 and Y_3.

The currently widespread approach to temporal experiments is to use a switchback design (Bojinov et al. 2020). Common switchback designs partition time into intervals of equal size and randomize each intervals' treatment assignments. However, fixed-duration switchback designs have two drawbacks. First, the length of time intervals need to be chosen by the practitioner, which we show theoretically in Section 3 and empirically Section 4 has a large impact on the performance of the design. Shorter periods increase spillovers from previous intervals, leading to interference bias. Longer periods decrease precision by decreasing balance in settings with autocorrelation. Second, the absolute time that switching occurs is fixed by choice of period and start time, which limits the ability of the design to exploit information about event density and covariance.

In this paper, we study a setting with multiple interventions and outcomes observed in continuous time. Good designs in this setting will tend to lower the mean square error (MSE) of the estimated GATE estimated using standard Horvitz-Thompson estimators (Horvitz and Thompson 1952), by effectively trading off various sources of bias and variance. We provide two primary contributions: first, a theoretical analysis to decompose sources of expected MSE from any design, and second, a simulation study that helps explore these tradeoffs and build intuition for properties of error-minimizing designs.

From a theoretical perspective, in Section 3 we derive a bias-variance decomposition of the MSE of the estimated GATE of each intervention. The bias is decomposed into three sources of errors: (a) spillover effects across time of a single intervention, (b) imbalance of treatment assignments from simultaneous interventions that cause confounding between interventions, and (c) imbalance of heterogeneous times (for example, peak vs. off-peak times) between the treated and control groups of a single intervention. Variance in our setting is caused by the measurement errors of outcomes and their covariance, determined by their distance in time. The relative contribution of these four sources of error affect the properties of the optimal treatment design. As an extension, we show in Section 5 that for spatiotemporal experiments where treatment decisions can vary with locations, the bias-variance decomposition continues to hold.

To study the temporal experiment design problem empirically, in Section 4 we conduct a simulation study that explores the role of assumptions about spillovers, outcome covariance, and event density in affecting the MSE of heuristic designs. We evaluate the performance of switchbacks with fixed and stochastic periods and characterize the properties of the most efficient designs. Practitioners can use similar simulations with assumptions tailored to their specific design problem in order to design efficient experiments in their empirical settings.

Our results highlight the role of using prior knowledge to select the average period between intervals, the timing of switching, and the role of randomization in improving robustness to simultaneous experiments and interference from spillovers.

1.1 Related Work

Our work is closely connected to several related literatures in the experimental design space. First, there has been extensive work on the design of experiments in temporal or time-series settings, the distinguishing property of which is that outcomes are subject carryover effects from treatments of prior time periods. As discussed above, the most common tool is the switchback design (Bojinov et al. 2020), in which predetermined time intervals are randomly and sequentially exposed to treatment and control variants. Alternative approaches include pulse designs Basse and Feller (2018) where units are treated only for one time period, or designs with irreversible treatment adoption pattern that are based on synthetic control estimators Doudchenko et al. (2019, 2021); Abadie and Zhao (2021) or generalized least squares Xiong et al. (2019).

Designing and analyzing experiments in the presence of interference has been studied in broad settings beyond temporal data. On network data, one common method for mitigating interference is through cluster-randomized designs (Ugander et al. 2013; Eckles et al. 2017; Candogan et al. 2021), where the clusters are chosen to minimize edges that cut across clusters. The cluster size serves an analogous role as the interval length in temporal data, governing the tradeoff between interference bias and estimator variance. Another popular method to mitigate interference is to use two-stage or multi-stage randomization, that has been used in public health Hudgens and Halloran (2008); Liu and Hudgens (2014), political science Sinclair et al. (2012), and social science Crépon et al. (2013); Baird et al. (2018); Basse and Feller (2018). In the spatial setting, a common approach is to conduct experiments at an aggregate level Bojinov et al. (2020); Xiong et al. (2019) or to randomly assign treatments to a set of predetermined spatial intervention points, with a focus on estimating spatial spillover effects Aronow et al. (2020, 2021). Our general approach to the temporal problem suggests that some of these ideas may be useful here as well.

Finally, a body of work has been dedicated to the specific setting of marketplace-mediated interference. (Johari et al. 2020) study how demand-randomized and supply-randomized designs can contribute different types of bias in a manner that is dependent on market balance. (Li et al. 2021) characterize the bias and variance of such experiments and describes how the design can be optimized in such settings. Holtz et al. (2020) compare GATE estimates from a meta experiment on the Airbnb marketplace that contains both cluster randomization and independent randomization. Holtz and Aral (2020) perform simulation studies that show how cluster-randomized experiments can be effective at reducing bias on the Airbnb network. Our paper takes a more agnostic approach to the marketplace by considering data in the form of a stream of events rather than the explicit two-sided setting.

2 Setting

Suppose K decision makers are running experiments on the same time periods simultaneously. For example, each decision maker could be on a different team within the same company. Let $T \in \mathbb{R}$ be the duration of the experiments. Each decision maker runs an experiment to study the effect of one intervention. For example, the interventions could be pricing, matching, or routing algorithms that are all being tested within the same marketplace in a single region or city.

For each intervention $\ell \in [K]$, let $w_\ell : [0, T] \to \{0, 1\}$ denote the binary-valued function where $w_\ell(t) = 1$ indicates at time $t \in [0, T]$ is exposed to intervention ℓ (treatment), and $w_\ell(t) = 0$ indicates otherwise (control).

Each decision maker ℓ makes treatment decisions for all times, i.e., $\boldsymbol{W}_\ell = \{W_\ell(t) : t \in [0, T]\}$, simultaneously, pre-experiment. Since t is continuous, the decision maker first partitions experimental times $[0, T]$ into M disjoint intervals, and then make treatment decisions at the interval level. Let $\mathbf{\Pi}_M : T \to \{t_0, t_1, \cdots, t_{M-1}, t_M\}$ be a function that outputs the endpoints of M intervals and the endpoints satisfy $0 = t_0 \leq t_1 \leq t_2 \leq \cdots \leq t_{M-1} \leq t_M = T$. For example, the temporal switchback designs that are commonly used in practice (Chamandy 2016; Bojinov et al. 2020) partition the time intervals of equal size (i.e., $t_{m+1} - t_m$ is the same for all m). The treatment status is switched randomly between any two consecutive intervals. Our setup is more general than a standard switchback test in that we allow for arbitrary switching times, and intervals of arbitrary length.

Each decision maker ℓ chooses the set of endpoints $\{t_{\ell 0}, t_{\ell 1}, \cdots, t_{\ell M}\}$.[1] It is possible that, interval $[t_{\ell m}, t_{\ell, m+1}]$ chosen by decision maker ℓ overlaps with, but is not identical to, interval $[t_{jk}, t_{j,k+1}]$ chosen by decision maker k.

As the treatment decisions are made at the interval level, the treatment assignments for all times within an interval are the same, i.e.,

$$w_\ell(t) = w_\ell(t'), \qquad \text{for all } t, t' \in [t_{\ell m}, t_{\ell, m+1}], \quad \text{for all } m.$$

Given $\{t_{\ell 0}, t_{\ell 1}, \cdots, t_{\ell M}\}$, the decision making problem simplifies to partitioning the M time intervals into the treatment and control groups of intervention ℓ. Equivalently, decision

[1] Without loss of generality, assume M is the same for all interventions, by allowing for the interval length to be measure zero.

maker ℓ chooses $\mathscr{T}_\ell$ and $\mathscr{C}_\ell$ pre-experiment, where $\mathscr{T}_\ell \in [M]$ is the set of indices of treated time intervals and $\mathscr{C}_\ell = [M] \backslash \mathscr{T}_\ell$ is the set of indices of control time intervals under intervention ℓ.

We define potential outcomes as

$$Y_t(\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_K),$$

where $\boldsymbol{w}_\ell = \{w_\ell(t) : t \in [0,T]\}$ denotes the treatment assignments of intervention ℓ for all times.[2] $Y_t(\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_K)$ is the outcome we would observe at t if treatment assignments satisfy $W_\ell(t') = w_\ell(t')$ for all $t' \in [0,T]$ and $\ell \in [K]$. Let $\boldsymbol{w}_\ell = \mathbf{1} = \{w_k(t) = 1 : t \in [0,T]\}$ and $\boldsymbol{w}_\ell = \mathbf{0} = \{w_k(t) = 0 : t \in [0,T]\}$ be the global treatment and global control of intervention ℓ, respectively.

Note that the definition above generalizes the standard, binary definition of potential outcomes under the stable unit treatment value assumption (SUTVA) in two aspects. First, this definition allows potential outcomes to be jointly affected by K interventions. Second, this definition allows for the existence of temporal spillover effects: the potential outcome of t is not only affected by the treatment status at t, but also the treatment assignments at other times. Note that this definition of potential outcomes can be easily generalized to include the spatial dimension, and similarly for the treatment variables, to account for the heterogeneity in locations for applications on a ride-hailing platform, as discussed in Section 5 below.

In addition, we allow the potential outcomes that are close in time to be correlated:

$$\mathrm{Cov}[Y_t(\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_K), Y_{t'}(\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_K)] \neq 0 \qquad \text{for} \quad t \neq t'.$$

The correlation is not caused by our intervention choices, but rather, captures the similarity of outcomes at nearby locations and times due to external factors like weather, supply conditions, and traffic. The correlation creates nuisance dependence between measurements, which can affect the resulting variance of effect estimates.

We are not able to observe outcomes for every time t. Instead we posit that a sequence of events, indexed by $i = 1, \ldots, n$, occur during the experiment with each taking place at time t_i, and decision makers observe outcomes $Y_{t_i} \coloneqq Y_{t_i}(\boldsymbol{W}_1, \cdots, \boldsymbol{W}_K)$ for all i. For example, the events may be riders checking prices, and Y_{t_i} may be a binary variable indicating whether the rider requests a ride. Let $f(t) : [0,T] \to \mathbb{R}^+$ be the density function from which events are sampled. Post-experiment, each decision maker ℓ use observed event outcomes $\{Y_{t_i}\}_{i \in [n]}$ and

[2]Intervention ℓ is not applied to $t \notin [0,T]$, i.e., $w_\ell(t)$ is always 0 for $t \notin [0,T]$. Therefore, there are no spillover effects of intervention ℓ from $\mathbb{R} \backslash [0,T]$ to $[0,T]$, and it is reasonable to define potential outcomes only by $\boldsymbol{w}_\ell = \{w_\ell(t) : t \in [0,T]\}$.

treatment assignments $\boldsymbol{W}_\ell$ to estimate the effect of intervention ℓ. For notation simplicity, we use Y_i and $W_{\ell i}$ in the following that stand for Y_{t_i} and $W_\ell(t_i)$, respectively.

2.1 Estimand

Our main object of interest is the *global average treatment effect* (GATE). We define the GATE of intervention ℓ as

$$\delta_\ell = \int \delta_{\ell,t} f(t) dt,$$

which is the average of $\delta_{\ell,t}$, the individual treatment effect at t, weighted by the event density $f(t)$, where $\delta_{\ell,t}$ is defined as

$$\delta_{\ell,t} = Y_t(\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \underbrace{\mathbf{1}}_{\text{intervention } \ell}, \cdots, \boldsymbol{w}_K) - Y_t(\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \underbrace{\mathbf{0}}_{\text{intervention } \ell}, \cdots, \boldsymbol{w}_K) \quad \forall t, \forall \{\boldsymbol{w}_k\}_{k\in[K]} \backslash \boldsymbol{w}_\ell.$$

$\delta_{\ell,t}$ measures the difference in outcomes at t between the state where intervention ℓ is deployed indefinitely (global treatment, i.e., $\boldsymbol{w}_\ell = \mathbf{1}$) and the state without intervention ℓ (global control, i.e., $\boldsymbol{w}_\ell = \mathbf{0}$). $\delta_{\ell,t}$ is defined conditional on $\boldsymbol{w}_k$ for $k \neq \ell$. However, we omit $\boldsymbol{w}_k$ in $\delta_{\ell,t}$ for notation simplicity. GATE δ_ℓ is the average of $\delta_{\ell,t}$ weighted by the event density function $f(t)$.

Analogously, we define the average direct effect τ_ℓ as

$$\tau_\ell = \int \tau_{\ell,t} f(t) dt,$$

where $\tau_{\ell,t}$ is the direct treatment effect at t

$$\tau_{\ell,t} = Y_t(\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \underbrace{e_t}_{\boldsymbol{w}_\ell}, \cdots, \boldsymbol{w}_K) - Y_t(\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \underbrace{\mathbf{0}}_{\boldsymbol{w}_\ell}, \cdots, \boldsymbol{w}_K) \quad \forall t, \forall \{\boldsymbol{w}_k\}_{k\in[K]} \backslash \boldsymbol{w}_\ell$$

and

$$e_t = \begin{pmatrix} 0 & \cdots & 0 & \underbrace{1}_{\text{time } t} & 0 & \cdots & 0 \end{pmatrix}$$

is a one-hot-encoded vector with the entry of time t to be 1 and all the remaining entries to be 0.

We define the average spillover effect $\gamma_\ell(\boldsymbol{w}_\ell)$, given treatment assignments $\boldsymbol{w}_\ell$, as

$$\gamma_\ell(\boldsymbol{w}_\ell) = \int \gamma_{\ell,t}(\boldsymbol{w}_\ell) f(t) dt,$$

where $\gamma_{\ell,t}(\boldsymbol{w}_\ell)$ is the spillover effect on t

$$\gamma_{\ell,t}(\boldsymbol{w}_\ell) = Y_t(\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_\ell, \cdots, \boldsymbol{w}_K) - Y_t(\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_\ell \circ e_t, \cdots, \boldsymbol{w}_K) \quad \forall t, \forall \{\boldsymbol{w}_k\}_{k\in[K]},$$

and "$\circ$" denotes the entry-wise product. Both $\tau_{\ell,t}$ and $\gamma_{\ell,t}(\boldsymbol{w}_\ell)$ are defined conditional $\boldsymbol{w}_k$ for all $k \neq \ell$, and we omit $\boldsymbol{w}_k$ in $\tau_{\ell,t}$ and $\gamma_{\ell,t}(\boldsymbol{w}_\ell)$ for notation simplicity.

Let $\gamma_\ell = \gamma_\ell(\mathbf{1})$ be the average treatment effect under global treatment. We can then decompose the GATE δ_ℓ as

$$\delta_\ell = \tau_\ell + \gamma_\ell.$$

2.2 Problem Formulation of Temporal Switchback Designs

Each decision maker ℓ estimates the GATE δ_ℓ, post-experiment, and decides whether to deploy intervention ℓ indefinitely. To make an informed decision, the decision maker seeks to estimate δ_ℓ as precisely as possible. We seek to coordinate treatment assignments of different interventions, such that the estimation error of $\hat{\delta}_\ell$ for all $\ell \in [K]$ can be simultaneously minimized.

We focus on deterministic switchback designs, where the treatment status of each interval is deterministic. We formulate the problem of designing simultaneous temporal switchback experiments as

$$\min_{\{\Pi_{M\ell}, \mathscr{T}_\ell\}_{\ell\in[K]}} \sum_{\ell=1}^{K} \mathbf{E}\left[(\hat{\delta}_\ell - \delta_\ell)^2 \mid \{\boldsymbol{W}_\ell\}_{\ell\in[K]}\right]. \tag{1}$$

The decision variables in (1) are the partition functions $\Pi_{M\ell}$ and treatment assignments $\mathscr{T}_\ell$ based on the partition functions $\Pi_{M\ell}$ for all ℓ. Given $\Pi_{M\ell}$ and $\mathscr{T}_\ell$, $W_\ell(t)$ is uniquely determined for all t, and the corresponding $\{\boldsymbol{W}_\ell\}_{\ell\in[K]}$ minimizes the sum of MSE of $\hat{\delta}_\ell$ for all ℓ.[3]

In addition to deterministic switchback designs, we consider and analyze block randomized designs in Appendix B, where the treatment assignment for each interval is randomized. The general insights of deterministic switchback and block randomized designs are aligned. We focus on the analysis of deterministic switchbacks in Section 3 for exposition, but we make a comparison between deterministic and stochastic designs, and explain the merits of each one via simulation studies in Section 4.

[3]The expected value in (1) is taken with respect the randomness in measurement error in Y_i and in the heterogeneity in treatment effects $\{\tau_{\ell,t} - \tau_\ell\}_{t\in[0,T]}$ and $\{\gamma_{\ell,t} - \gamma_\ell\}_{t\in[0,T]}$.

2.3 Post-experiment Estimation

One non-parametric estimator we could use for δ_ℓ is the Horvitz-Thompson estimator (Horvitz and Thompson 1952):

$$\hat{\delta}_\ell = \frac{1}{n}\sum_i \left(\frac{W_{\ell i}}{\pi_\ell} - \frac{1-W_{\ell i}}{1-\pi_\ell}\right) Y_i = \frac{1}{n}\sum_i \alpha_{\ell i} Y_i, \tag{2}$$

where $\alpha_{\ell i} = \frac{W_{\ell i}-\pi_\ell}{\pi_\ell(1-\pi_\ell)}$ is a normalized weight, and

$$\pi_\ell = \int W_\ell(t) f(t) dt$$

is the fraction of treated time under intervention ℓ.[4]

This estimator is quite general and flexible for three reasons. First, it does not rely on an assumption about spillover mechanisms. Second, it does not rely on assumptions about how event outcomes are correlated in time. Third, it does not require the knowledge of treatment decisions of simultaneous interventions, i.e., $\boldsymbol{W}_k$ for $k \neq \ell$.

However, the flexibility of this estimator comes at a cost. First, the estimator in (2) could be biased due to interference between observations. (2) approximates the outcomes under global treatment by the outcomes of treated units, and approximates the outcomes under global control by the outcomes of control units. When the spillover effect is zero, $\gamma_{\ell,t}(\boldsymbol{w}_\ell) = 0$, the approximation error is zero. For general cases, the approximation error is non-zero and (2) is a biased estimator of δ_ℓ. The interference bias scales with the size of spillover effect γ_ℓ. Second, the estimator in (2) could have a large variance as the effective sample size is affected by the correlation in event outcomes, and (2) does not optimally weight observations. Third, the estimator in (2) could have confounding bias from simultaneous interventions. It is possible that $\boldsymbol{W}_\ell$ is confounded with $\boldsymbol{W}_k$, and $\hat{\delta}_\ell$ is biased by τ_k and γ_k for $k \neq \ell$.

There are two directions that could lower the MSE of the estimated GATE. First, we can use a better treatment design, where we provide some guidance in Sections 3 and 4 below. Specifically, in Section 3, we derive a bias-variance decomposition of the MSE of $\hat{\delta}_\ell$ from (2) that shows how different sources of errors trade off. In Section 4, we conduct a simulation study to show how the MSE of heuristic designs vary with the assumptions on spillovers, outcome covariance, and event density.

Second, we can use a better estimator for GATE by leveraging prior knowledge of spillover and correlation mechanisms, and the information of other interventions. Specifically, to re-

[4] Note that in the Horvitz-Thompson estimator (2), the weights are the same within the treatment group and within the control group. However, the weights differ between the treatment and control groups to adjust for the difference in treated and control fractions of intervention ℓ solved from (1).

duce the interference bias, one could specify the structure of spillover mechanisms, explicitly estimate direct effect τ_ℓ and spillover effect γ_ℓ, and use τ_ℓ and γ_ℓ to estimate δ_ℓ. To reduce the variance from correlated outcomes, one could specify the structure of correlation mechanisms, and use the structure to reweight observations. To reduce confounding bias, one could use the information of $\boldsymbol{W}_k$ for $k \neq \ell$, and simultaneously estimate the treatment effects of all interventions. To simultaneously reduce bias and variance, one could use the generalized least squares (GLS) estimators that simultaneously estimate τ_ℓ and γ_ℓ for all ℓ, by taking advantage of the inverse error covariance weighting.

3 Analysis of Temporal Switchback Designs

In this section, we provide the bias-variance decomposition of the MSE of $\hat{\delta}_\ell$ from the Horvitz-Thompson estimator (2). The decomposition provides insights of how spillovers from interventions at earlier times, correlation in event outcomes, and effects of simultaneous interventions affect the MSE of $\hat{\delta}_\ell$. The insights can then be used as the guidance to optimize $\{\boldsymbol{W}_\ell\}_{\ell \in [K]}$ in practice.

We first lay out the assumptions that are necessary for the identification of treatment effects and bias-variance decomposition in Section 3.1. We then introduce several interval-level statistics that measure the average of time-varying components in the potential outcome model over an interval in Section 3.2. Finally, we provide the bias-variance decomposition of the MSE in terms of interval-level statistics in Section 3.3. The results in this section carry over to the design and analysis of spatiotemporal experiments, as discussed in Section 5.

3.1 Assumptions

We first assume that the sampling of events is independent of the treatment decisions of all interventions.

Assumption 1 (Exogeneity of events)**.** *Events are sampled randomly and independently from the density function $f(t)$, and $f(t)$ is independent of the treatment assignments of all interventions, $\boldsymbol{W}_1, \boldsymbol{W}_2, \cdots, \boldsymbol{W}_K$.*

This assumption makes sense for the interventions that potential riders cannot notice a difference before opening the app and checking prices, such as surge pricing algorithms or matching algorithms. However, this assumption is violated for interventions related to push notifications such as coupons or promo codes that are sent when users are not in the app. These interventions could incentivize more riders to open the app and check prices, consequently changing the event distribution.

In addition, we make a simplifying assumption on the structure of intervention effects.

Assumption 2 (Additivity of Intervention Effects)**.** *For any intervention ℓ, we have*

$$\begin{aligned}&\mathbf{E}\left[Y_t(\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}'_\ell, \cdots, \boldsymbol{w}_K) - Y_t(\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_\ell, \cdots, \boldsymbol{w}_K)\right] \\ =&\mathbf{E}\left[Y_t(\boldsymbol{w}'_1, \boldsymbol{w}'_2, \cdots, \boldsymbol{w}'_\ell, \cdots, \boldsymbol{w}'_K) - Y_t(\boldsymbol{w}'_1, \boldsymbol{w}'_2, \cdots, \boldsymbol{w}_\ell, \cdots, \boldsymbol{w}'_K)\right],\end{aligned}$$

where $\boldsymbol{w}_k$ and $\boldsymbol{w}'_k$ are two treatment assignments of intervention k for all k.[5]

When $K = 1$, Assumption 2 always holds. When $K > 1$, Assumption 2 implies that the effects of K interventions are additive, and this assumption excludes intervention effects to be synergistic (combining two interventions leads to a larger effect than expected) or antagonistic (combining two interventions leads to a smaller effect than expected). Assumption 2 is reasonable for certain classes of distinct interventions; for example, we may often assume that a pricing change and a routing change act via different mechanisms and are thus additive. However, more complex combinations may not satisfy Assumption 2.

We allow direct and spillover effects to be heterogeneous in t. However, for the identification purpose, we restrict the heterogeneity mechanism by assuming that treatment assignments of all interventions are independent of $\tau_{\ell,t}$ and $\gamma_{\ell,t}(\boldsymbol{w})$. In other words, we rule out the scenarios where decision makers allocate treatments to some specific times because the treatment effects at these times are larger. In our experimental design setting, decision makers can ensure this independence assumption to hold when making treatment decisions.[6]

Assumption 3 (Direct Effects)**.** *$\tau_{\ell,t}$ is i.i.d. in t and is independent of $\boldsymbol{W}_1, \cdots, \boldsymbol{W}_K$ for all t.*

Assumption 4 (Spillover Effects)**.** *For every t, there exists a non-negative interference kernel $d_{\ell,t}(t')$ that measures the interference intensity of intervention ℓ from t' to t and satisfies $\int d_{\ell,t}(t')f(t')dt' = 1$, such that*

$$\gamma_{\ell,t}(\boldsymbol{w}_\ell) = \gamma_{\ell,t} \cdot \int w_\ell(t') d_{\ell,t}(t') f(t') dt',$$

where $\gamma_{\ell,t} = \gamma_{\ell,t}(\mathbf{1})$ is i.i.d. in t, and is independent of $\boldsymbol{W}_1, \cdots, \boldsymbol{W}_K$ for all t.

$\gamma_{\ell,t}$ denotes the spillover effect on time t under global treatment of intervention ℓ. Assumption 4 implies that spillover effects from the treatments at other times are additive.

[5] The expected value in Assumption 2 is taken with respect to t.

[6] We can generalize the independence assumption to the conditional independence assumption, where treatment assignments of all interventions are independent of $\tau_{\ell,t}$ and $\gamma_{\ell,t}(\boldsymbol{w})$ conditional on observables. To account for observables in the design of experiments, we can use switchback or randomized designs stratified by observables (Athey and Imbens 2017).

Spillover effects are parametrized by an interference kernel $d_{\ell,t}(t')$ in Assumption 4. $d_{\ell,t}(t')$ can be quite general in t and t'. Below are three examples of interference kernels. A visualization of interference kernels is provided in Figure 2.

Example 1 (No spillover effect). $d_{\ell,t}(t') = 0$ for all $t \neq t'$.

Example 2 (Non-anticipating outcomes). $d_{\ell,t}(t') = 0$ for all $t' > t$. Then the integral is only over t' such that $t' \geq t$.

Example 3 (Bounded spillover effect). There exists $M < \infty$ such that $d_{\ell,t}(t') = 0$ for all $t' > t + M$.

3.2 Interval-level Statistics

We introduce several interval-level statistics that quantify spillover effects, correlation in outcomes, and other components at the interval level. These interval-level statistics are building blocks of the bias-variance decomposition in Section 3.3, and are important quantities to be considered in the partition of intervals and allocation of treatments.

Treated fraction. For any two interventions ℓ and k, let

$$\pi_{\ell k} = \int W_\ell(t) W_k(t) f(t) dt$$

be the fraction of time that is jointly treated under both interventions ℓ and k. If $\pi_{\ell k} \neq \pi_\ell \pi_k$, then the treatment assignments of interventions ℓ and k are correlated.

Mean outcome in the global control state of all interventions. For the m-th interval of intervention ℓ, $[t_{\ell m}, t_{\ell,m+1}]$, let

$$\mu_\ell^{(m)} = \int_{t \in [t_{\ell m}, t_{\ell,m+1}]} \mu_t \cdot f(t) dt$$

be the average control outcome of $t \in [t_{\ell m}, t_{\ell,m+1}]$ weighted by the event density $f(t)$, where

$$\mu_t = \mathbf{E}[Y_t(\mathbf{0}, \mathbf{0}, \cdots, \mathbf{0}) \mid t]$$

is the mean control outcome at t. $\mu_\ell^{(m)}$ for different m measures the heterogeneity in interval average outcomes in the global control state. The heterogeneity could come from the difference in riders'/drivers' behavior on weekdays vs. weekends.

Interference. For any two intervals, $[t_{\ell m}, t_{\ell,m+1}]$ and $[t_{jk}, t_{j,k+1}]$, let

$$I_{\ell j}^{(m,k)} = \int_{t\in[t_{\ell m},t_{\ell,m+1}]} \int_{t'\in[t_{jk},t_{j,k+1}]} d_{j,(t)}(t') f(t) f(t') dt dt'$$

be the intensity of spillover effects from interval $[t_{jk}, t_{j,k+1}]$ to interval $[t_{\ell m}, t_{\ell,m+1}]$. For intervention j, $\gamma_j \cdot I_{\ell j}^{(m,k)}$ is the average spillover effect of treating every $t' \in [t_{jk}, t_{j,k+1}]$ on $t \in [t_{\ell m}, t_{\ell,m+1}]$.

$I_{\ell j}^{(m,k)}$ is bounded between 0 and 1, and $I_{\ell j}^{(m,k)}$ increases with the length of $[t_{\ell m}, t_{\ell,m+1}]$ and $[t_{jk}, t_{j,k+1}]$. For notation simplicity, let $I_{\ell}^{(m,k)} = I_{\ell\ell}^{(m,k)}$.

Correlation in event outcomes. For any two intervals, $[t_{\ell m}, t_{\ell,m+1}]$ and $[t_{jk}, t_{j,k+1}]$, let

$$C_{\ell j}^{(m,k)} = \int_{t\in[t_{\ell m},t_{\ell,m+1}],t'\in[t_{jk},t_{j,k+1}]} \mathbf{E}[\varepsilon_t \varepsilon_{t'} \mid \{\boldsymbol{W}_\ell\}_{\ell\in[K]}] \cdot f(t) f(t') dt dt'$$

be the correlation in event outcomes between interval $[t_{\ell m}, t_{\ell,m+1}]$ and interval $[t_{jk}, t_{j,k+1}]$ given treatment assignments $\{\boldsymbol{W}_\ell\}_{\ell\in[K]}$, where ε_t is defined as

$$\varepsilon_t = \underbrace{Y_t(\mathbf{0}, \mathbf{0}, \cdots, \mathbf{0}) - \mu_t}_{\text{measurement error at } t} + \underbrace{\sum_{\ell=1}^{K} (\tau_{\ell,t} - \tau_\ell) \cdot W_\ell(t) + \sum_{\ell=1}^{K} (\gamma_{\ell,t} - \gamma_\ell) \cdot I_\ell(t)}_{\text{difference between heterogeneous and average treatment effects}}$$

and $I_\ell(t) = \gamma_{\ell,t}(\boldsymbol{W}_\ell)/\gamma_{\ell,t} \in [0, 1]$ is the scaled spillover effect at t given intervention ℓ's treatment assignments $\boldsymbol{W}_\ell$.

Note that ε_t consists of two parts: one is the measurement error in the outcome at t and the other one is the difference between heterogeneous and average treatment effects. Both parts can lead to the correlation between ε_t and $\varepsilon_{t'}$ for distinct t and t'. It is possible to parametrize the conditional covariance between ε_t and $\varepsilon_{t'}$ by a kernel function, depending on how the covariance varies with $t - t'$ (e.g., monotonically or periodically). See two examples in Figure 2.

$C_{\ell j}^{(m,k)}$ measures the average correlation between ε_t and $\varepsilon_{t'}$ for $t \in [t_{\ell m}, t_{\ell,m+1}]$ and $t' \in [t_{jk}, t_{j,k+1}]$, weighted by the event density $f(t)$. $C_{\ell j}^{(m,k)}$ scales with the length of $[t_{\ell m}, t_{\ell,m+1}]$ and $[t_{jk}, t_{j,k+1}]$. For notation simplicity, let $C_{\ell}^{(m,k)} = C_{\ell\ell}^{(m,k)}$.

3.3 Main Results

We provide the decomposition of the bias and mean squared error (MSE) of $\hat{\delta}_\ell$ from the Horvitz-Thompson estimator (2), in terms of the interval-level statistics, in Theorem 1 below.

The decomposition lays out how different components in the potential outcome model affect the estimation error of $\hat{\delta}_\ell$.

Theorem 1. *Suppose Assumptions 1-4 hold and we run switchback experiments. As $n \to \infty$, the bias of $\hat{\delta}_\ell$ estimated from (2) converges to*

$$\mathbf{E}\left[\hat{\delta}_\ell - \delta_\ell \mid \{\boldsymbol{W}_\ell\}_{\ell\in[K]}\right] \xrightarrow{p} \text{Bias}_\ell(\text{interference}) + \text{Bias}_\ell(\text{simultaneous}) + \text{Bias}_\ell(\text{mean})$$

and the MSE of $\hat{\delta}_\ell$ converges to

$$\mathbf{E}\left[(\hat{\delta}_\ell - \delta_\ell)^2 \mid \{\boldsymbol{W}_\ell\}_{\ell\in[K]}\right] \xrightarrow{p} \text{Var}_\ell(\text{error}) + \left[\text{Bias}_\ell(\text{interference}) + \text{Bias}_\ell(\text{simultaneous}) + \text{Bias}_\ell(\text{mean})\right]^2,$$

where

$$\begin{aligned}
\text{Bias}_\ell(\text{interference}) =& -\gamma_\ell \sum_{m\in\mathscr{T}_\ell}\sum_{k\in\mathscr{C}_\ell}\left[\frac{I_\ell^{(m,k)}}{\pi_\ell} + \frac{I_\ell^{(k,m)}}{1-\pi_\ell}\right] \\
\text{Bias}_\ell(\text{simultaneous}) =& \frac{1}{\pi_\ell}\sum_{k\neq\ell}\left[\tau_k\pi_{\ell k} + \gamma_k \sum_{m\in\mathscr{T}_\ell}\sum_{j\in\mathscr{T}_k} I_{\ell k}^{(m,j)}\right] - \frac{1}{1-\pi_\ell}\sum_{k\neq\ell}\left[\tau_k(\pi_k - \pi_{\ell k}) + \gamma_k \sum_{m\in\mathscr{C}_\ell}\sum_{j\in\mathscr{T}_k} I_{\ell k}^{(m,j)}\right] \\
\text{Bias}_\ell(\text{mean}) =& \frac{1}{\pi_\ell}\sum_{m\in\mathscr{T}_\ell}\mu_\ell^{(m)} - \frac{1}{1-\pi_\ell}\sum_{m\in\mathscr{C}_\ell}\mu_\ell^{(m)} \\
\text{Var}_\ell(\text{error}) =& \frac{1}{\pi_\ell^2}\sum_{m,k\in\mathscr{T}_\ell} C_\ell^{(m,k)} - \frac{1}{\pi_\ell(1-\pi_\ell)}\sum_{m\in\mathscr{T}_\ell,k\in\mathscr{C}_\ell} C_\ell^{(m,k)} + \frac{1}{(1-\pi_\ell)^2}\sum_{m,k\in\mathscr{C}_\ell} C_\ell^{(m,k)}.
\end{aligned}$$

Note that the expected value in Theorem 1 conditions on $\{\boldsymbol{W}_\ell\}_{\ell\in[K]}$ given that $\{\boldsymbol{W}_\ell\}_{\ell\in[K]}$ is fixed in the deterministic switchbacks. As the number of events $n \to \infty$, the empirical event distribution better approximates the population event distribution, and then the limit of bias and MSE can be written in terms of interval-level statistics that are defined under the population event distribution.

Theorem 1 shows that the bias consists of three terms, and the MSE consists of four terms. Below we elaborate on each term.

$\text{Bias}_\ell(\text{interference})$ measures the bias from the spillover effects of the treatment of intervention ℓ at other times. If $\gamma_{\ell,t} = 0$ for all t (spillover effects are zero), then $\text{Bias}_\ell(\text{interference}) = 0$; otherwise, $\text{Bias}_\ell(\text{interference})$ is generally nonzero. $\text{Bias}_\ell(\text{interference})$ is defined by $I_\ell^{(m,k)}$ and $I_\ell^{(k,m)}$, where $[t_{\ell m}, t_{\ell,m+1}]$ is a treated interval and $[t_{\ell k}, t_{\ell,k+1}]$ is a control interval. $I_\ell^{(m,k)}$ measures the bias by using the outcomes of treated units to approximate the outcomes in the global treatment state, and symmetrically $I_\ell^{(k,m)}$ measures the bias by using the outcomes of control units to approximate the outcomes in the global control state. All terms in $\text{Bias}_\ell(\text{interference})$ have the same sign. If the interference kernel $d_{\ell,t}(t')$ decays at a faster rate in $|t' - t|$, then both $I_\ell^{(m,k)}$ and $I_\ell^{(k,m)}$ tend to be smaller (interference between two intervals is smaller). Therefore, if treated intervals are lengthened, and same for the control intervals, the size of $\text{Bias}_\ell(\text{interference})$ tends to be smaller, as further discussed in Section 4.1 and shown in Figure 3 below.

$\text{Bias}_\ell(\text{simultaneous})$ comes from the imbalance of treatment assignments of other interventions between the treated and control intervals of intervention ℓ, i.e., $\boldsymbol{W}_\ell$ is confounded with $\boldsymbol{W}_k$ for $k \neq \ell$. If $K = 1$ and $\ell = 1$ (one intervention only), then $\text{Bias}_\ell(\text{simultaneous}) = 0$; otherwise, $\text{Bias}_\ell(\text{simultaneous})$ is generally nonzero. $\text{Bias}_\ell(\text{simultaneous})$ tends to increase with K and the size of τ_k and γ_k for $k \neq \ell$. $\text{Bias}_\ell(\text{simultaneous})$ tends to be smaller when the confounding of treatment assignments of simultaneous experiments is reduced, for example by varying interval lengths, as further discussed in Section 4.3 and shown in Figure 5b below.

$\text{Bias}_\ell(\text{mean})$ comes from the imbalance of mean outcome μ_t in the global control state between the treated and control time intervals of intervention ℓ. If μ_t is the same for all t, then $\text{Bias}_\ell(\text{mean}) = 0$; otherwise, $\text{Bias}_\ell(\text{mean})$ is generally nonzero. $\text{Bias}_\ell(\text{mean})$ tends to be smaller when heterogeneous times (for example, peak vs. off-peak times) are balanced between the treated and control time intervals.

Note that ε_t has mean zero, but affects the MSE of $\hat{\delta}_\ell$ through the term $\text{Var}_\ell(\text{error})$. If $\mathbf{E}[\varepsilon_t \varepsilon_{t'} \mid \{\boldsymbol{W}_\ell\}_{\ell \in [K]}] = 0$ for $t \neq t'$, then $\text{Var}_\ell(\text{error}) = 0$, which is aligned with the standard asymptotic results of the Horvitz-Thompson estimator Horvitz and Thompson (1952).[7] If the following two conditions hold, then $\mathbf{E}[\varepsilon_t \varepsilon_{t'} \mid \{\boldsymbol{W}_\ell\}_{\ell \in [K]}] = 0$ for all $t \neq t'$:

1. constant treatment effect: $\tau_{\ell,t}$ and $\gamma_{\ell,t}$ are the same for all t ;

2. independent measurement error: $Y_t(\mathbf{0}, \mathbf{0}, \cdots, \mathbf{0}) - \mu_t$ is independent in t.

If either of the above two conditions hold, but not both, then $\text{Var}_\ell(\text{error})$ is generally nonzero, but tends to be smaller than the case where both conditions are violated. Note that the middle term in $\text{Var}_\ell(\text{error})$ is negative. When $\mathbf{E}[\varepsilon_t \varepsilon_{t'} \mid \{\boldsymbol{W}_\ell\}_{\ell \in [K]}]$ decreases with $|t - t'|$, then $\text{Var}_\ell(\text{error})$ tends to be smaller when the treatment status switches frequently, and then the middle term tends to be more negative, as further discussed in Section 4.1 and shown in Figure 3 below.

Below we provide three examples of how the MSE is simplified under special cases and what the implication is on the optimal design.

Example 4 (Interference only)**.** Suppose $K = 1$, $\mu_t \equiv \mu$, and $\mathbf{E}[\varepsilon_t \varepsilon_{t'} \mid \{\boldsymbol{W}_\ell\}_{\ell \in [K]}] = 0$ for $t \neq t'$. In this case, $\text{Bias}_\ell(\text{simultaneous}) = 0$, $\text{Bias}_\ell(\text{mean}) = 0$, and $\text{Var}_\ell(\text{error}) = 0$.

[7] For the Horvitz-Thompson estimator under the classical setting where there is no interference and correlation in errors is zero, the Horvitz-Thompson estimator converges to the true value at the rate $O_p(1/\sqrt{n})$. Consequently, the MSE converges to zero at the rate $O_p(1/n)$ and therefore the limit of the MSE is 0.

The MSE of $\hat{\delta}_\ell$ converges to $[\text{Bias}_\ell(\text{interference})]^2$. A design with deterministic and long switchback period is preferable.

Example 5 (Correlation only)**.** Suppose $K = 1$, $\mu_t \equiv \mu$, and $\gamma_{\ell,t} \equiv 0$. In this case, $\text{Bias}_\ell(\text{simultaneous}) = 0$, $\text{Bias}_\ell(\text{mean}) = 0$, and $\text{Bias}_\ell(\text{interference}) = 0$. The MSE of $\hat{\delta}_\ell$ converges to $\text{Var}_\ell(\text{error})$. A design with deterministic and short switchback period is preferable.

Example 6 (Simultaneous interventions only)**.** Suppose $\mu_t \equiv \mu$, $\gamma_{\ell,t} \equiv 0$, and $\mathbf{E}[\varepsilon_t \varepsilon_{t'} \mid \{\boldsymbol{W}_\ell\}_{\ell \in [K]}] = 0$ for $t \neq t'$. In this case, $\text{Bias}_\ell(\text{simultaneous}) = 0$, $\text{Bias}_\ell(\text{interference}) = 0$, and $\text{Var}_\ell(\text{error}) = 0$. The MSE of $\hat{\delta}_\ell$ converges to $[\text{Bias}_\ell(\text{simultaneous})]^2$. A design with randomized switchback frequency is preferable.

For general cases, we need to balance the tradeoffs involved, and the optimal design varies with the relative strength of each component in the decomposition of MSE. In Section 4, we study the performance of heuristic designs under various scenarios, and provide guidance on designing efficient experiments in empirical settings.

4 Simulation Results

In this section, we present estimates of the mean-squared error of heuristic designs under a simulated problem structure to characterize the tradeoffs involved. Evaluating a design through simulation requires the following inputs:

- Spillover kernel $d_{\ell,t}(t')$: we use a finite duration linear kernel in all simulations.
- Covariance kernel $\mathbf{E}[\varepsilon_t \varepsilon_{t'} \mid \{\boldsymbol{W}_\ell\}_{\ell \in [K]}]$: we consider two regimes, a triangular kernel with height 1, and a periodic covariance kernel which is the product of a triangular kernel and cosine function capturing seasonal patterns.
- Event density $f(t)$: we consider two regimes, uniform density of events and a periodic density $f(t) \propto \sin(\alpha t)$ where events are clustered in time according to a known seasonal pattern.

Figure 2 graphically depicts our design choices for the simulations. Additionally, we vary parameters governing the strength of the direct and spillover effect sizes τ_ℓ and γ_ℓ, which affect bias from interference and simultaneous experiments. Since these parameters are arbitrary and must be assumed, we choose them such that the resulting bias is on the same scale as the (fixed) variance.

We restrict our evaluation to two heuristic designs in order to build intuition. The simplest design is a deterministic switchback with period p and offset $q < p$, where we initially switch policies at time q and then every p periods (i.e., the length of $[t_0, t_1]$ is q and the length of $[t_m, t_{m+1}]$ is p for $m \geq 1$). We also consider a natural randomized analog, a random-duration switchback where we switch policies every Poisson(λ) periods (i.e., the length of $[t_m, t_{m+1}]$ is randomized for all m). The λ parameter governs the mean period length and is analogous to the p period parameter for the deterministic policy.

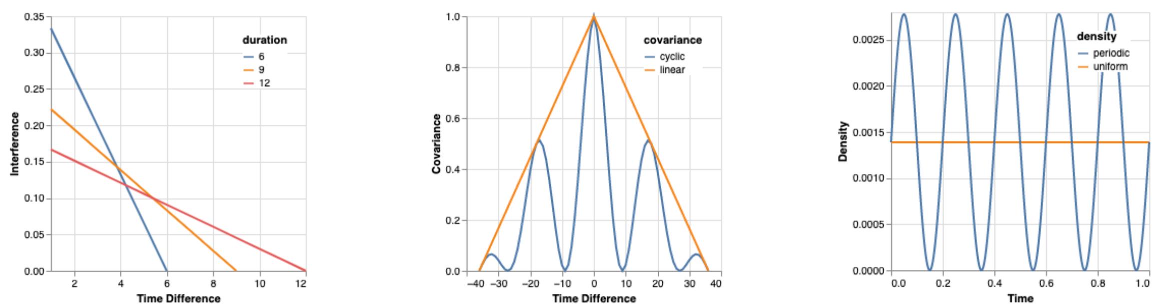


Figure 2: Simulation setup: spillover and covariance kernels, and periodic densities. Time difference denotes $t' - t$ in the spillover kernel $d_{\ell,t}(t')$ and in the covariance kernel $\mathbf{E}[\varepsilon_t \varepsilon_{t'} \mid \{\boldsymbol{W}_\ell\}_{\ell \in [K]}]$. If $t' - t < 0$, then $d_{\ell,t}(t') = 0$.

4.1 Interference and variance tradeoffs

Figure 3 summarizes the most fundamental tradeoff of temporal experiments–policies with shorter periods generate more comparisons that leverage autocorrelation but also increase interference bias from previous intervals in different conditions. When the spillover effect γ_ℓ is small, switching as quickly as possible results in the most efficient design, and when it is large, we improve the design through lengthening the period. We focus most of our ensuing discussion on settings where these two error components are on a similar scale and result in an interesting tradeoff.

4.2 Stochastic versus deterministic designs

Figure 4 compares deterministic designs with various periods to a distribution of errors resulting from different random draws of the stochastic policy. We find that the stochastic switchback generally results in designs with lower bias and increased variance for most values of λ. The randomization generates some longer periods between switching, which helps improve the estimator performance with respect to bias from interference.

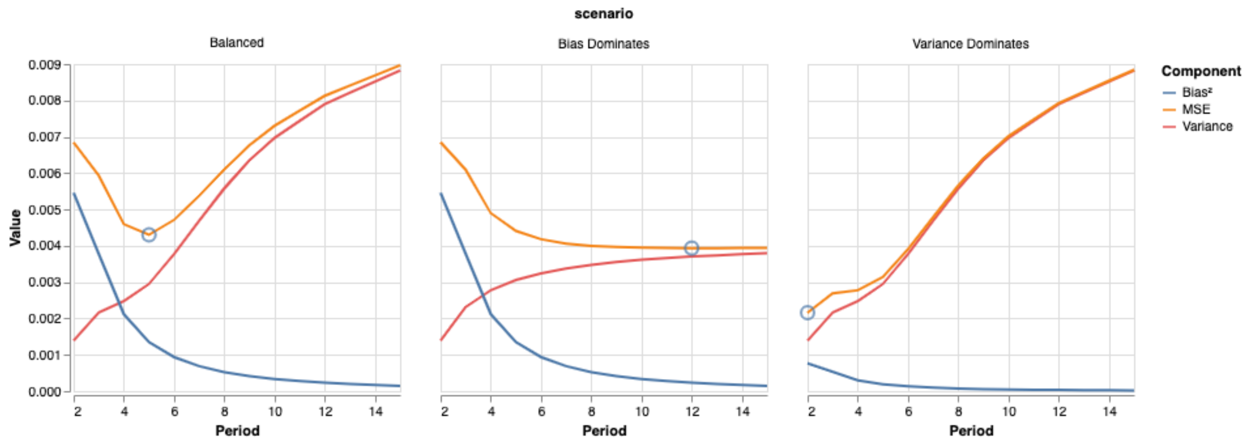


Figure 3: Tradeoffs under different regimes for deterministic switchback. The x-axis denotes period p in the deterministic switchback with offset $q = 0$. The period p with smallest MSE is circled in blue.

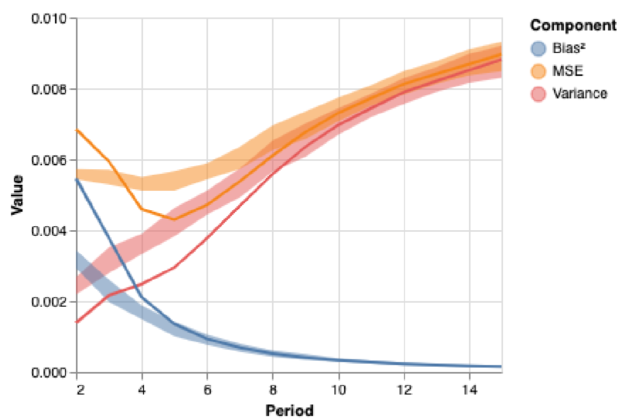
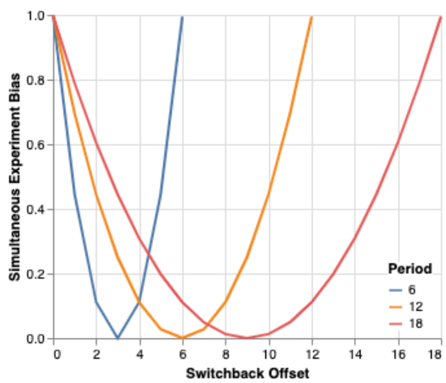


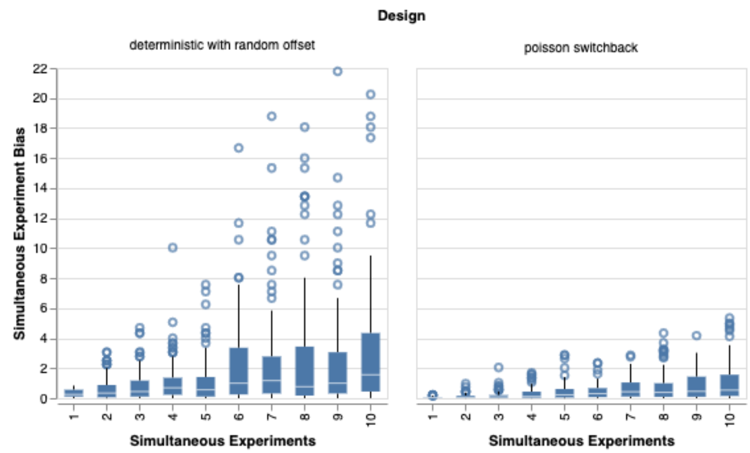
Figure 4: Poisson vs. deterministic switchback. Solid lines denote deterministic switchback (deterministic policy). Shaded bands denote Poisson switchback (stochastic policy).

4.3 Simultaneous experiments

Our main result in Theorem 1 shows that simultaneous experiments may generate additional bias when their effects are large enough, resulting in additional estimation error. In Figure 5a, we show how two experiments are run simultaneously in deterministic switchback designs if their periods are offset correctly. Failing to stagger two tests properly results in a very large bias as one experiment confounds the other. In Figure 5b, we estimate how this bias is affected by stochastic designs with a fixed mean period length. As we increase the number of simultaneous experiments, the average bias increases due to the finite-sample correlation between the treatment periods. The bias depends on the specific randomization, and in some outlier cases, it can be far larger than in the average case. It is not depicted, but nuisance confounding is always worse for simultaneous experiments with longer mean periods due to increased correlation in treatments.



(a) Deterministic policies



(b) Stochastic policies (random offset or Poisson switchback)

Figure 5: Effects of simultaneous experiments on optimal design. In Figure 5a, two simultaneous experiments are run with period $p \in \{6, 12, 18\}$ in both designs, and with offset $q = 0$ in one design and varying offset q (x-axis in Figure 5a) in another design. In Figure 5b we show distributions of bias produced by using random offsets (left panel) or the Poisson switchback (right panel). The Poisson switchback is generally more effective, unless the deterministic designs are staggered perfectly.

4.4 Periodic event density

In many realistic settings, the density of events will be exhibit periodic patterns due to the seasonality of human behavior. For instance, in ride-hailing, many ride requests occur during commute times, and relatively few occur during the late evening on weeknights. These daily and weekly cycles create opportunities for improving the design of temporal experiments, and motivate simulations with a simple periodic density function. Figure 6 shows results from a periodic density using a deterministic switchback. When the design has a period that aligns with density ($p \in \{6, 12\}$), the offset parameter q determines how the alignment alters the bias and variance. For $p = 12$, an offset of 3 (blue dots and lines), yields a design with the lowest variance by switching at an area of maximum density. This results in more events having natural "matches" in an adjacent interval. An offset of 10 (yellow dots and lines) minimizes bias by switching directly after a period with low event density, which minimizes interference from the preceding interval. Knowledge of the density of events can improve the efficiency of the design by leveraging the best absolute times for bias- or variance-minimizing switching points.

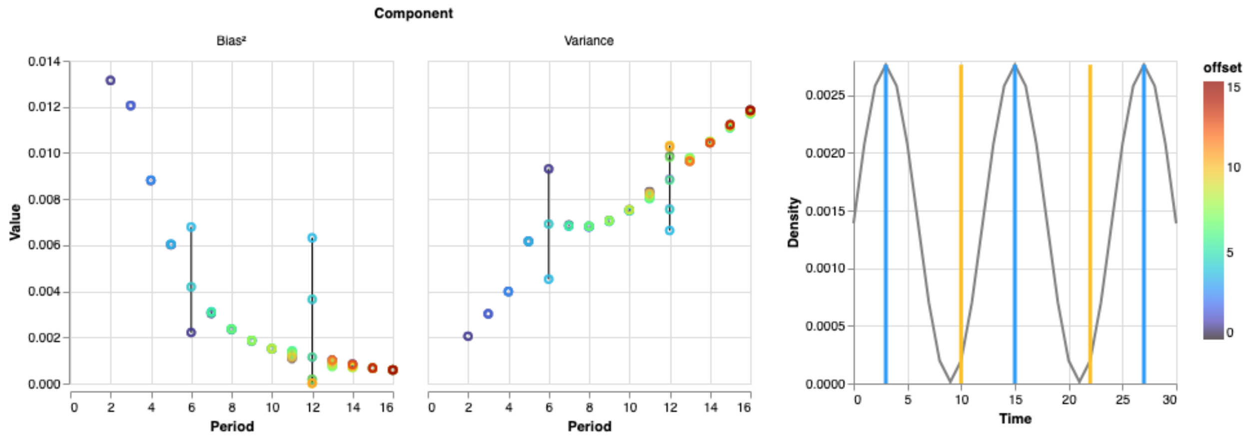


Figure 6: Bias and variance estimates for deterministic switchback in a setting a density with a 12 period cycle. The color of points varies with offset parameter q. In all periods except 6 and 12, the offsets result in almost identical bias and variance.

4.5 Periodic covariance

Another property of realistic settings is that the covariance of errors may exhibit a periodic structure, as events may be exposed to similar unobservable effects despite occurring at different times. In ride-hailing applications, we empirically observe positive correlations between days and weeks, occurring at similar times. For instance, a morning commute may be similarly affected by weather that persists across days. This leads to a covariance structure that both decreases with distance in time, but also increases with *similarity* in time of day. In Figure 7, we compare the variance of deterministic designs when this property is reflected in the covariance. In this setting with a deterministic switchback, the expected variance of the design no longer monotonically increases in the period. Periods that ensure correlated events receive different treatments can achieve lower variance, while other periods increase variance by creating additional correlations between intervals with the same treatment.

4.6 Takeaways

Although our simulation results do not allow us to directly construct an optimal design, they point to the properties that better designs would tend to have and the fundamental constraints implied by the noise and causal structure of the setting.

First, as we learned in Section 3, the mean period of the design trades off variance by increasing correlation and bias by decreasing interference from previous periods. We can see that the MSE-minimizing period can vary substantially depending on assumptions about covariance (which are testable), the magnitude of effects, and spillover structure encoded by $d_{\ell,t}(t')$, γ_ℓ, and τ_ℓ (which must usually be assumed).

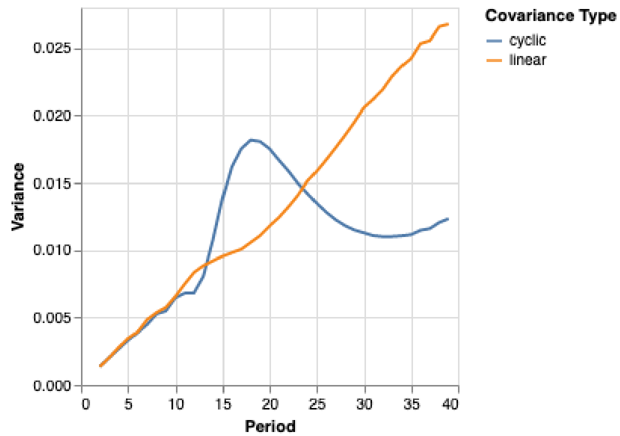


Figure 7: Variance of deterministic switchback under linear and cyclic decay of covariance structure in time. Variance increases monotonically in period p under linear decay, but non-monotonically under cyclic decay.

Second, stochastic designs exhibit lower bias from interference and simultaneous experiments, but do incur some additional variance in order to achieve this. Randomization has the additional benefit that we observe intervals with different lengths, which can help with testing if the treatment causes longer spillovers than were assumed in the design phase.

Third, simultaneous experiments are an important source of error under reasonable assumptions, which is quite a different regime than traditional A/B testing with user-level randomizations, which can generally support many simultaneous tests. In general, throughput of multiple temporal experiments with substantive effects is something a centralized platform should manage in order to prevent a "tragedy of the commons" result. Ensuring that simultaneous experiments have designs that are uncorrelated in finite samples is likely to be valuable and could be validated pre-experiment as proposed in Gupta et al. (2018) ("Seedfinder") or restricted randomizations (Simon 1979).

Fourth, periodic behavior in both event density and in covariance structure implies that there may be benefits and costs to cleverly choosing absolute switching times and periods between switching. A more sophisticated search process could be applied to designing temporal experiments that could leverage estimates of density and the covariance kernel to provide better designs.

5 Extension to Spatiotemporal Experiments

Our framework of designing experiments on the temporal space can be easily generalized to the spatial dimension. First, we show how the definitions of potential outcomes, treatment variables, treatment effects and event density above can be appropriately modified. Specifically, let $f(s,t)$ be the event density at location s and time t that could vary with s (for example, airport vs. residential area) and t (for example, weekdays vs. weekends). The location $s \in \mathbb{R}^2$ could be parametrized, e.g., by latitude and longitude. Let $Y_{(s,t)}(\boldsymbol{w}_1 \cdots, \boldsymbol{w}_K)$ and $w_{\ell,(s,t)}$ be the potential outcome and treatment assignment of intervention ℓ at (s,t). The definition of $Y_{(s,t)}(\boldsymbol{w}_1 \cdots, \boldsymbol{w}_K)$ is analogous to the definitions used in Aronow et al. (2020, 2021), but more general in the sense that, we additionally allow for the temporal dimension and explicitly account for concurrent interventions. Analogously, let $\delta_{\ell,(s,t)}$, $\tau_{\ell,(s,t)}$, and $\gamma_{\ell,(s,t)}(\boldsymbol{w}_\ell)$ be the individual treatment effect, direct treatment effect and spillover effect at (s,t) to account for the heterogeneity of treatment effects in locations and times. The GATE, average direct and spillover effects are the averages of the corresponding effect at (s,t) weighted by event density $f(s,t)$.

Second, the treatments are assigned at the three-dimensional block level. We generalize the definition of partition and let $\boldsymbol{\Pi}_M : \mathcal{S} \times [0,T] \to \{\mathcal{B}_1, \mathcal{B}_2, \cdots, \mathcal{B}_M\}$ be a function that partitions $\mathcal{S} \times [0,T]$ into M disjoint three-dimensional blocks, satisfying $\mathcal{B}_1 \cup \mathcal{B}_2 \cup \cdots \cup \mathcal{B}_M = \mathcal{S} \times [0,T]$ and $\mathcal{B}_k \cap \mathcal{B}_m = \emptyset$ for $k \neq m$, where $\mathcal{S} \in \mathbb{R}^2$ is the set of locations on which the experiments are run. $\mathcal{B}_m$ can be of arbitrary shape. Then the decision making problem is to assign the M blocks to the treatment and control groups of an intervention.[8]

Third, we have exactly the same bias-variance decomposition of MSE as Theorem 1 for spatiotemporal switchback experiments and as Theorem 2 in Appendix B for spatiotemporal block randomized experiments by using block-level statistics (rather than interval-level statistics), as shown in the proof of Theorems 1 and 2. To be more specific about block-level statistics, they are the direct generalizations of $\pi_{\ell k}$, $\mu_\ell^{(m)}$, $I_{\ell j}^{(m,k)}$ and $C_{\ell j}^{(m,k)}$ defined in Section 3.2 to the integrations over (s,t) weighted by $f(s,t)$.[9]

6 Discussion and Conclusion

This paper studies the sources of error in the design and analysis of simultaneous temporal experiments. We provide a theoretical analysis of how the bias and variance of the Horvitz-Thompson estimator of the GATE are affected by three factors: spillovers from interventions at earlier times, correlation in event outcomes, and effects of interventions tested concurrently. We provide simulation examples that show how these three factors trade off each other and provide insights into how one can design efficient temporal experiments.

[8]This design problem of spatiotemporal experiments differs from Aronow et al. (2020, 2021) in the sense that Aronow et al. (2020, 2021) randomly assign treatments to a set of predetermined spatial intervention points, with a focus on estimating spatial spillover effects.

[9]More specifically, $\mu_\ell^{(m)}$ is integrated over $(s,t) \in \mathcal{B}_{\ell m}$ for $\mu_\ell^{(m)}$. Both $I_{\ell j}^{(m,k)}$ and $C_{\ell j}^{(m,k)}$ are integrated over $(s,t) \in \mathcal{B}_{\ell m}, (s',t') \in \mathcal{B}_{jk}$.

Perhaps the most general conclusion we can draw is that designing experiments in this context involves considering a complex set of tradeoffs and critically depends on the assumptions experimentalists would make using prior knowledge. While the expected event density is straightforward to estimate, high-dimensional covariance matrices in event outcomes may pose challenges (Fan et al. 2016). The assumed spillover structure is effectively a causal model for which practitioners may need to use prior experimental evidence to adequately capture.

The wide variation in MSE of designs in various simulation setups highlights that useful theory and priors are important factors in the success of experiments in this setting. This is in contrast to randomized experiments with i.i.d. units, where there are a variety of reliable tools for design and analysis, fewer assumptions are needed in either experiment phase, and bias contributes less prominently to estimation (Lin 2013).

We motivated this study through supporting experiments in a ride-hailing setting where multiple teams share a fixed set of experimental units but can run experiments over long time periods to increase the sample size. These "temporal" experiments are a useful tool in this setting, but we could see broader use in other applications with better development of the theory and practical guidelines.

Indeed, there are a variety of settings where cross-sectional interventions are not possible or where outcomes cannot be easily attributed to treatment decisions. Estimating the effectiveness of traditional media advertising is well suited to our problem setup, and a privacy-friendly approach to online advertising might employ temporal variation in campaign spend linked to sales through timestamps only. There is also prior work using time-varying interventions in financial or cryptocurrency markets (Krafft et al. 2018) or in self-experimentation for personalized medicine (Karkar et al. 2016). An important goal of this work is to broaden the use of temporal experiments to settings where they are not currently used.

There are important questions left unanswered, most importantly a tractable approach for solving the optimal design problem. More sophisticated designs could improve upon the two heuristics we evaluated in Section 4. Solving the globally optimal design that minimizes the MSE is challenging, as conditional on the partitioned time intervals, allocating them to the treatment and control groups is equivalent to the Max-Cut problem that is NP-hard in general graphs. Some heuristic algorithms, such as simulated annealing Van Laarhoven and Aarts (1987), or approximation algorithms for Max-Cut, such as randomized rounding Goemans and Williamson (1995), could be helpful for finding principled designs, and could be another interesting direction to explore for future work.

References

A. Abadie and J. Zhao. Synthetic controls for experimental design. *arXiv preprint arXiv:2108.02196*, 2021.

P. M. Aronow, C. Samii, and Y. Wang. Design-based inference for spatial experiments with interference. *arXiv preprint arXiv:2010.13599*, 2020.

P. M. Aronow, C. Samii, J. Sullivan, and Y. Wang. Inference in spatial experiments with interference using the spatialeffect package. *arXiv preprint arXiv:2106.15081*, 2021.

S. Athey and G. W. Imbens. The econometrics of randomized experiments. In *Handbook of economic field experiments*, volume 1, pages 73–140. Elsevier, 2017.

S. Baird, J. A. Bohren, C. McIntosh, and B. Özler. Optimal design of experiments in the presence of interference. *Review of Economics and Statistics*, 100(5):844–860, 2018.

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

I. Bojinov, D. Simchi-Levi, and J. Zhao. Design and analysis of switchback experiments. *Available at SSRN 3684168*, 2020.

O. Candogan, C. Chen, and R. Niazadeh. Near-optimal experimental design for networks: Independent block randomization. *Available at SSRN*, 2021.

N. Chamandy. Experimentation in a ridesharing marketplace, 2016.

B. Crépon, E. Duflo, M. Gurgand, R. Rathelot, and P. Zamora. Do labor market policies have displacement effects? evidence from a clustered randomized experiment. *The quarterly journal of economics*, 128(2):531–580, 2013.

N. Doudchenko, D. Gilinson, S. Taylor, and N. Wernerfelt. Designing experiments with synthetic controls. Technical report, Working paper, 2019.

N. Doudchenko, K. Khosravi, J. Pouget-Abadie, S. Lahaie, M. Lubin, V. Mirrokni, J. Spiess, et al. Synthetic design: An optimization approach to experimental design with synthetic controls. *Advances in Neural Information Processing Systems*, 34, 2021.

D. Eckles, B. Karrer, and J. Ugander. Design and analysis of experiments in networks: Reducing bias from interference. *Journal of Causal Inference*, 5(1), 2017.

J. Fan, Y. Liao, and H. Liu. An overview of the estimation of large covariance and precision matrices, 2016.

M. X. Goemans and D. P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM (JACM)*, 42(6):1115–1145, 1995.

S. Gupta, L. Ulanova, S. Bhardwaj, P. Dmitriev, P. Raff, and A. Fabijan. The anatomy of a large-scale experimentation platform. In *2018 IEEE International Conference on Software Architecture (ICSA)*, pages 1–109. IEEE, 2018.

D. Holtz and S. Aral. Limiting bias from test-control interference in online marketplace experiments. *Available at SSRN 3583596*, 2020.

D. Holtz, R. Lobel, I. Liskovich, and S. Aral. Reducing interference bias in online marketplace pricing experiments. *Available at SSRN 3583836*, 2020.

D. G. Horvitz and D. J. Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association*, 47(260):663–685, 1952.

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

R. Jagadeesan, N. S. Pillai, and A. Volfovsky. Designs for estimating the treatment effect in networks with interference. *The Annals of Statistics*, 48(2):679–712, 2020.

R. Johari, H. Li, I. Liskovich, and G. Weintraub. Experimental design in two-sided platforms: An analysis of bias. *arXiv preprint arXiv:2002.05670*, 2020.

R. Karkar, J. Zia, R. Vilardaga, S. R. Mishra, J. Fogarty, S. A. Munson, and J. A. Kientz. A framework for self-experimentation in personalized health. *Journal of the American Medical Informatics Association*, 23(3):440–448, 2016.

P. M. Krafft, N. Della Penna, and A. S. Pentland. An experimental study of cryptocurrency market dynamics. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pages 1–13, 2018.

H. Li, G. Zhao, R. Johari, and G. Y. Weintraub. Interference, bias, and variance in two-sided marketplace experimentation: Guidance for platforms. *arXiv preprint arXiv:2104.12222*, 2021.

W. Lin. Agnostic notes on regression adjustments to experimental data: Reexamining freedman's critique. *The Annals of Applied Statistics*, 7(1):295–318, 2013.

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

R. Mirza, S. Punja, S. Vohra, and G. Guyatt. The history and development of n-of-1 trials. *Journal of the Royal Society of Medicine*, 110(8):330–340, 2017.

R. Simon. Restricted randomization designs in clinical trials. *Biometrics*, pages 503–512, 1979.

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

J. Ugander, B. Karrer, L. Backstrom, and J. Kleinberg. Graph cluster randomization: Network exposure to multiple universes. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 329–337, 2013.

P. J. Van Laarhoven and E. H. Aarts. Simulated annealing. In *Simulated annealing: Theory and applications*, pages 7–15. Springer, 1987.

R. Xiong, S. Athey, M. Bayati, and G. W. Imbens. Optimal experimental design for staggered rollouts. *Available at SSRN 3483934*, 2019.

A Theoretical Results for Switchback Designs

To accommodate the extension to spatiotemporal experiments in Section 5, all the results are derived for designs of spatiotemporal experiments. As we treat the space dimension as a whole in temporal designs, so these designs are special cases of spatiotemporal designs. Theorem 1 follows immediately from the derivation in this section.

A.1 Block-level Statistics

To make things clear, we provide the definition of block-level statistics that generalize and encompass the interval-level statistics provided in Section 3.2. All the results and proofs in Appendix A.2 and A.3 are presented in terms of block-level statistics.

$$
\begin{aligned}
\pi_{\ell k} &= \int W_\ell(s,t) W_k(s,t) f(s,t) ds dt \\
\mu_\ell^{(m)} &= \int_{(s,t)\in\mathcal{B}_{\ell m}} \mu_{(s,t)} \cdot f(s,t) ds dt \\
I_{\ell j}^{(m,k)} &= \int_{(s,t)\in\mathcal{B}_{\ell m}} \int_{(s',t')\in\mathcal{B}_{jk}} d_{j,(s,t)}(s',t') f(s,t) f(s',t') ds dt ds' dt' \\
C_{\ell j}^{(m,k)} &= \int_{(s,t)\in\mathcal{B}_{\ell m},(s',t')\in\mathcal{B}_{jk}} \mathbf{E}[\varepsilon_{(s,t)}\varepsilon_{(s',t')} \mid \{\boldsymbol{W}_\ell\}_{\ell\in[K]}] \cdot f(s,t) f(s',t') ds dt ds' dt'
\end{aligned}
$$

A.2 Supplementary Theoretical Results

Let $\delta_{\ell i} = \delta_{\ell,(s_i,t_i)}$, $\tau_{\ell i} = \tau_{\ell,(s_i,t_i)}$, $\gamma_{\ell i} = \gamma_{\ell,(s_i,t_i)}$, $W_{\ell i} = W_\ell(s_i,t_i)$, $I_{\ell i} = \gamma_{\ell,(s_i,t_i)}(\boldsymbol{W}_\ell)/\gamma_{\ell,(s_i,t_i)}$, $Y_i = Y_{(s_i,t_i)}$ and $\varepsilon_i = \varepsilon_{(s_i,t_i)}$. Under Assumptions 1-4, the observed outcome of event i can be written as the reduced-form model with parameters of average direct and spillover effects:

$$
Y_i = \mu_i + \sum_{\ell=1}^{K} \tau_\ell \cdot W_{\ell i} + \sum_{\ell=1}^{K} \gamma_\ell \cdot I_{\ell i} + \varepsilon_i. \tag{3}
$$

The estimation error of $\hat{\delta}_\ell$ can be decomposed as

$$
\begin{aligned}
\hat{\delta}_\ell - \delta_\ell = & \underbrace{\frac{1}{n}\sum_i \alpha_{\ell i} \cdot \gamma_\ell \cdot (I_{\ell i} - W_{\ell i})}_{\text{interference}} \\
& + \underbrace{\frac{1}{n}\sum_i \alpha_{\ell i} \sum_{k \neq \ell} (\tau_k \cdot W_{ki} + \gamma_k \cdot I_{ki})}_{\text{effects of other interventions}} \\
& + \underbrace{\frac{1}{n}\sum_i \alpha_{\ell i} \cdot \varepsilon_i}_{\text{correlation in ``errors''}} \\
& + \underbrace{\frac{1}{n}\sum_i \frac{W_{\ell i} - \pi_\ell}{\pi_\ell}\left(\frac{\mu_i}{1-\pi_\ell} + \delta_\ell\right)}_{\text{constant term}}
\end{aligned}
\tag{4}
$$

The following four lemmas provide the bias and MSE of each of the above four terms. We can then use these four lemmas to show Theorem 1.

Lemma 1 (Correlation in "Errors", Switchback Design)**.** *Under Assumptions 1-4,*

$$
\mathbf{E}\left[\frac{1}{n}\sum_i \alpha_{\ell i} \cdot \varepsilon_i \mid \{\boldsymbol{W}_\ell\}_{\ell \in [K]}\right] = 0
$$

and, as $n \to \infty$,

$$
\begin{aligned}
\mathbf{E}\left[\left(\frac{1}{n}\sum_i \alpha_{\ell i} \cdot \varepsilon_i\right)^2 \mid \{\boldsymbol{W}_\ell\}_{\ell \in [K]}\right] \xrightarrow{p} & \frac{1}{\pi_\ell^2}\sum_{m,k \in \mathscr{T}_\ell} C_\ell^{(m,k)} - \frac{1}{\pi_\ell(1-\pi_\ell)} \sum_{m \in \mathscr{T}_\ell, k \in \mathscr{C}_\ell} C_\ell^{(m,k)} \\
& + \frac{1}{(1-\pi_\ell)^2} \sum_{m,k \in \mathscr{C}_\ell} C_\ell^{(m,k)}.
\end{aligned}
$$

Lemma 2 (Interference, Switchback Design)**.** *Under Assumptions 1-4, as $n \to \infty$,*

$$
\mathbf{E}\left[\frac{1}{n}\sum_i \alpha_{\ell i} \cdot \gamma_\ell \cdot (I_{\ell i} - W_{\ell i}) \mid \{\boldsymbol{W}_\ell\}_{\ell \in [K]}\right] \xrightarrow{p} -\gamma_\ell \sum_{m \in \mathscr{T}_\ell} \sum_{k \in \mathscr{C}_\ell} \left[\frac{I_\ell^{(m,k)}}{\pi_\ell} + \frac{I_\ell^{(k,m)}}{1-\pi_\ell}\right],
$$

Lemma 3 (Constant Term, Switchback Design)**.** *Under Assumptions 1-4, as $n \to \infty$,*

$$
\mathbf{E}\left[\frac{1}{n}\sum_i \frac{W_{\ell i} - \pi_\ell}{\pi_\ell}\left(\frac{\mu_i}{1-\pi_\ell} + \delta_\ell\right) \mid \{\boldsymbol{W}_\ell\}_{\ell \in [K]}\right] \xrightarrow{p} \frac{1}{\pi_\ell}\sum_{m \in \mathscr{T}_\ell} \mu_\ell^{(m)} - \frac{1}{1-\pi_\ell}\sum_{m \in \mathscr{C}_\ell} \mu_\ell^{(m)}.
$$

Lemma 4 (Effects of Other Interventions, Switchback Design)**.** *Under Assumptions 1-4, as $n \to \infty$,*

$$
\begin{aligned}
& \mathbf{E}\left[\frac{1}{n}\sum_i \alpha_{\ell i} \sum_{k \neq \ell} (\tau_k \cdot W_{ki} + \gamma_k \cdot I_{ki}) \mid \{\boldsymbol{W}_\ell\}_{\ell \in [K]}\right] \\
\xrightarrow{p} & \frac{1}{\pi_\ell}\sum_{k \neq \ell}\left[\tau_k \pi_{\ell k} + \gamma_k \sum_{m \in \mathscr{T}_\ell}\sum_{j \in \mathscr{T}_k} I_{\ell k}^{(m,j)}\right] - \frac{1}{1-\pi_\ell}\sum_{k \neq \ell}\left[\tau_k(\pi_\ell - \pi_{\ell k}) + \gamma_k \sum_{m \in \mathscr{C}_\ell}\sum_{j \in \mathscr{T}_k} I_{\ell k}^{(m,j)}\right].
\end{aligned}
$$

A.3 Proof of Lemmas 1-4 and Theorem 1

In this section, we first show Lemmas 1-4, and then we use Lemmas 1-4 to show Theorem 1. Let $m(i) \in [M]$ be the index of block to which unit i belongs.

Proof of Lemma 1, for Switchback Design. The mean of $\frac{1}{n}\sum_i \alpha_{\ell i} \cdot \varepsilon_i$ is zero following that

$$\begin{aligned}
\mathbf{E}\left[\frac{1}{n}\sum_i \alpha_{\ell i} \cdot \varepsilon_i \mid \{\boldsymbol{W}_\ell\}_{\ell\in[K]}\right] =& \frac{1}{n}\sum_i \alpha_{\ell i} \cdot \mathbf{E}\left[\varepsilon_i \mid \{\boldsymbol{W}_\ell\}_{\ell\in[K]}\right] \\
=& \frac{1}{n}\sum_i \left[\sum_{\ell=1}^{K} \underbrace{\mathbf{E}[\tau_{\ell i} - \tau_\ell]}_{=0} \cdot W_{\ell i} + \sum_{\ell=1}^{K} \underbrace{\mathbf{E}[\gamma_{\ell i} - \gamma_\ell]}_{=0} \cdot I_{\ell i} + \underbrace{\mathbf{E}[Y_i(\mathbf{0}, \mathbf{0}, \cdots, \mathbf{0}) - \mu_i]}_{=0}\right] \\
=& 0
\end{aligned}$$

from Assumptions 3 and 4, and the second moment of this term is

$$\begin{aligned}
\mathbf{E}\left[\left(\frac{1}{n}\sum_i \alpha_{\ell i} \cdot \varepsilon_i\right)^2 \mid \{\boldsymbol{W}_\ell\}_{\ell\in[K]}\right] =& \frac{1}{n^2}\sum_i\sum_j \alpha_{\ell i}\alpha_{\ell j}\mathbf{E}\left[\varepsilon_i\varepsilon_j \mid \{\boldsymbol{W}_\ell\}_{\ell\in[K]}\right] \\
=& \frac{1}{n^2}\sum_{m,k\in\mathscr{T}_\ell}\frac{1}{\pi_\ell^2}\sum_{i,j:m_\ell(i)=m,m_\ell(j)=k}\mathbf{E}\left[\varepsilon_i\varepsilon_j \mid \{\boldsymbol{W}_\ell\}_{\ell\in[K]}\right] \\
&- \frac{1}{n^2}\sum_{m\in\mathscr{T}_\ell,k\in\mathscr{C}_\ell}\frac{1}{\pi_\ell(1-\pi_\ell)}\sum_{i,j:m_\ell(i)=m,m_\ell(j)=k}\mathbf{E}\left[\varepsilon_i\varepsilon_j \mid \{\boldsymbol{W}_\ell\}_{\ell\in[K]}\right] \\
&+ \frac{1}{n^2}\sum_{m,k\in\mathscr{C}_\ell}\frac{1}{(1-\pi_\ell)^2}\sum_{i,j:m_\ell(i)=m,m_\ell(j)=k}\mathbf{E}\left[\varepsilon_i\varepsilon_j \mid \{\boldsymbol{W}_\ell\}_{\ell\in[K]}\right]
\end{aligned}$$

From Assumption 1 and Law of Large Numbers (LLN), we have as $n \to \infty$

$$\begin{aligned}
&\frac{1}{n^2}\sum_{i,j:m_\ell(i)=m,m_\ell(j)=k}\mathbf{E}\left[\varepsilon_i\varepsilon_j \mid \{\boldsymbol{W}_\ell\}_{\ell\in[K]}\right] \\
\xrightarrow{p}& \int_{(s,t)\in\mathcal{B}_{\ell m},(s',t')\in\mathcal{B}_{\ell k}}\mathbf{E}\left[\varepsilon_{(s,t)}\varepsilon_{(s',t')} \mid \{\boldsymbol{W}_\ell\}_{\ell\in[K]}\right] f(s,t)f(s',t')dsdtds'dt' = C_\ell^{(m,k)},
\end{aligned}$$

where $C_\ell^{(m,k)}$ measures the correlation in "errors" between $\mathcal{B}_{\ell m}$ and $\mathcal{B}_{\ell k}$, and is defined in Section 3.2. Then we have

$$\mathbf{E}\left[\left(\frac{1}{n}\sum_i \alpha_{\ell i} \cdot \varepsilon_i\right)^2 \mid \{\boldsymbol{W}_\ell\}_{\ell\in[K]}\right] \xrightarrow{p} \frac{1}{\pi_\ell^2}\sum_{m,k\in\mathscr{T}_\ell}C_\ell^{(m,k)} - \frac{1}{\pi_\ell(1-\pi_\ell)}\sum_{m\in\mathscr{T}_\ell,k\in\mathscr{C}_\ell}C_\ell^{(m,k)} + \frac{1}{(1-\pi_\ell)^2}\sum_{m,k\in\mathscr{C}_\ell}C_\ell^{(m,k)}.$$

□

Proof of Lemma 2. From the definition of $I_{\ell i}$ and Assumption 4, we have

$$I_{\ell i} = \int W_\ell(s,t)d_i(s-s_i,t-t_i)f(s,t)dsdt = \sum_m W_\ell^{(m)}\underbrace{\int_{(s,t)\in\mathcal{B}_{\ell m}} d_i(s-s_i,t-t_i)f(s,t)dsdt}_{:=I_{\ell i}^{(m)}} = \sum_m W_\ell^{(m)}I_{\ell i}^{(m)}.$$

By Assumption 4, $\sum_m I_{\ell i}^{(m)} = 1$. If $m(i) \in \mathscr{T}_\ell$, then

$$I_{\ell i} - W_{\ell i} = -\sum_{m \in \mathscr{C}_\ell} I_{\ell i}^{(m)};$$

and if $m(i) \in \mathscr{C}_\ell$, then

$$I_{\ell i} - W_{\ell i} = \sum_{m \in \mathscr{T}_\ell} I_{\ell i}^{(m)}.$$

As $n \to \infty$, from LLN, we have

$$\frac{1}{n}\sum_i \alpha_{\ell i} \cdot \gamma_\ell \cdot (I_{\ell i} - W_{\ell i}) = -\frac{\gamma_\ell}{\pi_\ell}\sum_{m \in \mathscr{T}_\ell} \frac{1}{n} \sum_{i: m_\ell(i)=m} \sum_{k \in \mathscr{C}_\ell} I_{\ell i}^{(k)} - \frac{\gamma_\ell}{1-\pi_\ell}\sum_{m \in \mathscr{C}_\ell} \frac{1}{n} \sum_{i: m_\ell(i)=m} \sum_{k \in \mathscr{T}_\ell} I_{\ell i}^{(k)}$$

$$\xrightarrow{p} -\frac{\gamma_\ell}{\pi_\ell}\sum_{m \in \mathscr{T}_\ell}\sum_{k \in \mathscr{C}_\ell} I_\ell^{(m,k)} - \frac{\gamma_\ell}{1-\pi_\ell}\sum_{m \in \mathscr{C}_\ell}\sum_{k \in \mathscr{T}_\ell} I_\ell^{(m,k)} = -\gamma_\ell \sum_{m \in \mathscr{T}_\ell}\sum_{k \in \mathscr{C}_\ell}\left[\frac{I_\ell^{(m,k)}}{\pi_\ell} + \frac{I_\ell^{(k,m)}}{1-\pi_\ell}\right],$$

where $I_\ell^{(m,k)}$ measures the intensity of spillover effects from $\mathcal{B}_{\ell k}$ to $\mathcal{B}_{\ell m}$, and is defined in Section 3.2.

□

Proof of Lemma 3. As $n \to \infty$, from LLN, the sum of the average of control outcomes and finite-sample approximation error converges to

$$\frac{1}{n}\sum_i \frac{W_{\ell i} - \pi_\ell}{\pi_\ell}\left(\frac{\mu_i}{1-\pi_\ell} + \delta_\ell\right) = \frac{1}{n\pi_\ell}\sum_{i: m_\ell(i) \in \mathscr{T}_\ell} \mu_i - \frac{1}{n(1-\pi_\ell)} \sum_{i: m_\ell(i) \mathscr{C}_\ell} \mu_i + \underbrace{\frac{\delta_\ell}{n}\sum_i \frac{W_{\ell i} - \pi_\ell}{\pi_\ell}}_{\xrightarrow{p} 0}$$

$$\xrightarrow{p} \frac{1}{\pi_\ell}\sum_{m \in \mathscr{T}_\ell} \mu_\ell^{(m)} - \frac{1}{1-\pi_\ell}\sum_{m \in \mathscr{C}_\ell} \mu_\ell^{(m)},$$

where $\mu_\ell^{(m)}$ is $\mathcal{B}_{\ell m}$'s mean outcome in the global control state of all interventions, and is defined in Section 3.2. □

Proof of Lemma 4. As $n \to \infty$, from LLN, the effects of other interventions converge to

$$\begin{aligned}
&\frac{1}{n}\sum_i \alpha_{\ell i} \sum_{k \neq \ell} (\tau_k \cdot W_{ki} + \gamma_k \cdot I_{ki}) \\
=&\frac{1}{n\pi_\ell}\sum_{i: m_\ell(i) \in \mathscr{T}_\ell}\sum_{k \neq \ell}(\tau_k \cdot W_{ki} + \gamma_k \cdot I_{ki}) - \frac{1}{n(1-\pi_\ell)}\sum_{i: m_\ell(i) \in \mathscr{C}_\ell}\sum_{k \neq \ell}(\tau_k \cdot W_{ki} + \gamma_k \cdot I_{ki}) \\
=&\frac{1}{n\pi_\ell}\sum_{k \neq \ell}\left(\tau_k n_{\ell k} + \gamma_k \sum_{i: m_\ell(i) \in \mathscr{T}_\ell}\sum_{j \in \mathscr{T}_k} I_{ki}^{(j)}\right) - \frac{1}{n(1-\pi_\ell)}\sum_{k \neq \ell}\left(\tau_k(n_k - n_{\ell k}) + \gamma_k \sum_{i: m_\ell(i) \in \mathscr{C}_\ell}\sum_{j \in \mathscr{T}_k} I_{ki}^{(j)}\right) \\
\xrightarrow{p}&\frac{1}{\pi_\ell}\sum_{k \neq \ell}\left[\tau_k \pi_{\ell k} + \gamma_k \sum_{m \in \mathscr{T}_\ell}\sum_{j \in \mathscr{T}_k} I_{\ell k}^{(m,j)}\right] - \frac{1}{1-\pi_\ell}\sum_{k \neq \ell}\left[\tau_k(\pi_k - \pi_{\ell k}) + \gamma_k \sum_{m \in \mathscr{C}_\ell}\sum_{j \in \mathscr{T}_k} I_{\ell k}^{(m,j)}\right],
\end{aligned}$$

where $n_{\ell k} = \sum_i \mathbf{1}(m_\ell(i) \in \mathscr{T}_\ell, m_k(i) \in \mathscr{T}_k)$, $\pi_{\ell k}$ and $I_{\ell k}^{(m,j)}$ are defined in Section 3.2. The second equation follows from the proof of Lemma 2.

□

Proof of Theorem 1. From the decomposition (4) and from Lemmas 1-4, as $n \to \infty$

$$\mathbf{E}\left[\hat{\delta}_\ell - \delta_\ell \mid \{\boldsymbol{W}_\ell\}_{\ell \in [K]}\right] \xrightarrow{p} \text{Bias}_\ell(\text{spillover}) + \text{Bias}_\ell(\text{mean}) + \text{Bias}_\ell(\text{simultaneous}),$$

where all of these three terms are defined in Theorem 1. From Lemma 1,

$$\mathbf{E}\left[\frac{1}{n}\sum_i \alpha_{\ell i} \cdot \varepsilon_i \mid \{\boldsymbol{W}_\ell\}_{\ell \in [K]}\right] \cdot [\text{Bias}_\ell(\text{spillover}) + \text{Bias}_\ell(\text{mean}) + \text{Bias}_\ell(\text{simultaneous})] = 0.$$

Then as $n \to \infty$, the MSE converges to

$$\mathbf{E}\left[(\hat{\delta}_\ell - \delta_\ell)^2 \mid \{\boldsymbol{W}_\ell\}_{\ell \in [K]}\right] \xrightarrow{p} \text{Var}_\ell(\text{error}) + [\text{Bias}_\ell(\text{spillover}) + \text{Bias}_\ell(\text{mean}) + \text{Bias}_\ell(\text{simultaneous})]^2 .$$

□

B Theoretical Results for Block Randomized Designs

An alternative design is block randomized design, where the treatment assignments are randomized for every block. Similar to Appendix A, all the results in this section are derived for designs of spatiotemporal experiments. For notation simplicity, in this section, we assume partitions $\Pi_{\ell M}$ are the same for all ℓ and treatment assignments of any two interventions are independent.

Assumption 5 (Treatment assignments)**.** *The treatment assignments of intervention ℓ are independent of the treatment assignments of intervention k for $k \neq \ell$, i.e.,*

$$W_\ell(s,t) \perp W_k(s',t'), \qquad \forall (s,t),(s',t').$$

With more complicated notations, our results can be easily generalized to the case where partitions $\Pi_{\ell m}$ vary with ℓ, and the general insights stay the same. We formulate the optimization problem to solve the optimal partition Π_M for all interventions as:

$$\min_{\Pi_M} \sum_{\ell=1}^{K} \mathbf{E}\left[(\hat{\delta}_\ell - \delta_\ell)^2\right]. \tag{5}$$

In this section, we seek to provide a decomposition of the bias and MSE of $\hat{\delta}_\ell$, where $\hat{\delta}_\ell$ is estimated from the Horvitz-Thompson estimator below

$$\hat{\delta}_\ell = \frac{1}{n}\sum_i \left(\frac{W_{\ell i}}{\pi_{\ell i}} - \frac{1 - W_{\ell i}}{1 - \pi_{\ell i}}\right) Y_i = \frac{1}{n}\sum_i \alpha_{\ell i} Y_i \tag{6}$$

where $\pi_{\ell i} = \mathbf{P}(W_{\ell i} = 1)$ and $\alpha_{\ell i} = \frac{W_{\ell i} - \pi_{\ell i}}{\pi_{\ell i}(1-\pi_{\ell i})}$. The key distinction between (6) and the Horvitz-Thompson estimator for the switchback design is that (6) accounts for different treated probabilities for different units.

B.1 Block-level Statistics

We introduce several more block-level statistics that are defined analogously to those in Section 3.2, but are adapted to block randomized designs.

Event fraction. For any block $\mathcal{B}_m$, let

$$\pi^{(m)} = \int_{(s,t)\in\mathcal{B}_m} f(s,t)dsdt$$

be the fraction of events in $\mathcal{B}_m$.

Treated fraction. For any block $\mathcal{B}_m$, let

$$\pi_\ell^{(m)} = \frac{1}{\pi^{(m)}} \int_{(s,t)\in\mathcal{B}_m} W_\ell(s,t) f(s,t)dsdt$$

be the fraction of treated under intervention ℓ in $\mathcal{B}_m$ weighted by the event density.

Correlation in "errors". For any block $\mathcal{B}_m$, let

$$C^{(m)} = \int_{(s,t)\in\mathcal{B}_m,(s',t')\in\mathcal{B}_m} \mathbf{E}\left[\varepsilon_{(s,t)}\varepsilon_{(s',t')}\right] f(s,t)f(s',t')dsdtds'dt'$$

be the correlation in "errors" between any two units in block $\mathcal{B}_m$. Compared with $C_{\ell\ell}^{(m,m)}$ defined in Section 3.2, the key difference is that the expected value of $\varepsilon_{(s,t)}\varepsilon_{(s',t')}$ in $C^{(m)}$ does not condition on $\{\boldsymbol{W}_\ell\}_{\ell\in[K]}$.

Interference. For any block $\mathcal{B}_m$, let

$$I_\ell^{(m)} = \int_{(s,t)\in\mathcal{B}_m,(s',t')\in\mathcal{B}_m} d_{\ell,(s,t)}(s,t) f(s,t)f(s',t')dsdtds'dt'$$

be the average intensity of spillover effects between any two points in block $\mathcal{B}_m$. $I_\ell^{(m)}$ equals $I_{\ell\ell}^{(m,m)}$ defined in Section 3.2.

B.2 Main Results

The following theorem provides a decomposition of the bias of $\hat{\delta}_\ell$ and an upper bound of MSE, where $\hat{\delta}_\ell$ is estimated from the Horvitz-Thompson estimator and with π_ℓ replaced by $\pi_\ell^{m(i)}$, where $m(i) \in [M]$ denotes the index of block to which unit i belongs, i.e., $(s_i, t_i) \in \mathcal{B}_{\ell m(i)}$.

Theorem 2. *Suppose Assumptions 1-5 hold and we run block randomized experiments. As $n \to \infty$, the bias of $\hat{\delta}_\ell$ estimated from* (2) *converges to*

$$\mathbf{E}\left[\hat{\delta}_\ell - \delta_\ell\right] \xrightarrow{p} \text{Bias}_\ell(\text{interference})$$

and the limit of MSE of $\hat{\delta}_\ell$ can be upper bounded by

$$\lim_{n\to\infty} \mathbf{E}\left[(\hat{\delta}_\ell - \delta_\ell)^2\right] \leq 4\left[\text{Var}(\text{error}) + \text{Var}_\ell(\text{simultaneous}) + \text{Var}_\ell(\text{mean}) + \text{MSE}_\ell(\text{interference})\right]$$

where

$$
\begin{aligned}
\text{Bias}_\ell(\text{interference}) =& -\gamma_\ell \sum_m \left(\pi^{(m)} - I_\ell^{(m)}\right) \\
\text{Var}(\text{error}) =& \sum_m \frac{C^{(m)}}{\pi_\ell^{(m)}(1-\pi_\ell^{(m)})} \\
\text{Var}_\ell(\text{mean}) =& \sum_m \frac{\left(\mu^{(m)} + (1-\pi_\ell^{(m)})\pi^{(m)} \cdot \delta_\ell\right)^2}{\pi_\ell^{(m)}(1-\pi_\ell^{(m)})} \\
\text{Var}_\ell(\text{simultaneous}) =& \sum_m \frac{1}{\pi_\ell^{(m)}(1-\pi_\ell^{(m)})} \left\{ \sum_{k\neq\ell} \left[\pi_k^{(m)}(1-\pi_k^{(m)}) \left(\tau_k \cdot \pi^{(m)} + \gamma_k \cdot I_k^{(m)}\right)^2 \right.\right. \\
& \left.\left. + \gamma_k^2 \sum_{q\neq m} \pi_k^{(q)}(1-\pi_k^{(q)})(I_k^{(q)})^2 \right] + \left[\sum_{k\neq\ell} \left(\tau_k \cdot \pi_k^{(m)}\pi^{(m)} + \gamma_k \cdot \sum_q \pi_k^{(q)} I_k^{(q)} \right) \right]^2 \right\}. \\
\text{MSE}_\ell(\text{interference}) =& \sum_m \left[\frac{1}{\pi_\ell^{(m)}} \left(\sum_{k\neq m} \left(1-\pi_\ell^{(k)}\right) I_\ell^{(m,k)} \right)^2 + \frac{1}{1-\pi_\ell^{(m)}} \left(\sum_{k\neq m} \pi_\ell^{(k)} I_\ell^{(m,k)} \right)^2 \right. \\
& \left. + \frac{1}{\pi_\ell^{(m)}(1-\pi_\ell^{(m)})} \sum_{k\neq m} \pi_\ell^{(k)}(1-\pi_\ell^{(k)}) \left(I_\ell^{(m,k)}\right)^2 \right] \\
& + \sum_{m\neq k} \left[\left(\pi^{(m)} - I_\ell^{(m)}\right)\left(\pi^{(k)} - I_\ell^{(k)}\right) + I_\ell^{(m,k)} I_\ell^{(k,m)} \right].
\end{aligned}
$$

Proposition 1. *In Theorem 2, if* $\pi_\ell^{(m)} \equiv 0.5$*, then the terms in the upper bound of the limit of MSE are simplified to*

$$
\begin{aligned}
\text{Var}(\text{error}) =& 4 \sum_m C^{(m)} \\
\text{Var}_\ell(\text{mean}) =& 4 \sum_m \left(\mu^{(m)} + 0.5\pi^{(m)}\delta_\ell\right)^2 \\
\text{Var}_\ell(\text{simultaneous}) =& \sum_m \left\{ \sum_{k\neq\ell} \left[\left(\tau_k \cdot \pi^{(m)} + \gamma_k \cdot I_k^{(m)}\right)^2 + \gamma_k^2 \sum_{q\neq m} (I_k^{(q)})^2 \right] + \left[\sum_{k\neq\ell} \left(\tau_k \cdot \pi^{(m)} + \gamma_k \cdot \sum_q I_k^{(q)} \right) \right]^2 \right\} \\
\text{MSE}_\ell(\text{interference}) =& \left(1 - \sum_m I_\ell^{(m)}\right)^2 + \sum_{m\neq k} \left[\left(I_\ell^{(m,k)}\right)^2 + I_\ell^{(m,k)} I_\ell^{(k,m)} \right].
\end{aligned}
$$

Given that $C^{(m)}, \mu^{(m)}, \pi^{(m)}, I_\ell^{(m)}$, and $I_\ell^{(m,k)}$ are proportional to the size of $\mathcal{B}_m$ following from their definitions, suppose there exist some constants $\alpha, \beta \in [0,1]$ such that

1. $C^{(m)} = \Theta\left(\frac{1}{M^{1+\alpha}}\right), \forall\ m$
2. $\mu^{(m)} = \Theta\left(\frac{1}{M}\right), \forall\ m$
3. $\pi^{(m)} = \Theta\left(\frac{1}{M}\right), \forall\ m$

4. $I_\ell^{(m)} = O\left(\frac{1}{M^{1+\beta}}\right), \forall\ m$

5. $I_\ell^{(m,k)} = O\left(\frac{1}{M^2}\right), \forall\ m \neq k, \ell$

When $\pi_\ell^{(m)} \equiv 0.5$, the leading terms in the upper bound of MSE are

$$\lim_{n\to\infty} \mathbf{E}\left[(\hat{\delta}_\ell - \delta_\ell)^2\right] = \underbrace{4\sum_m C^{(m)}}_{\Theta\left(\frac{1}{M^\alpha}\right)} + \underbrace{\sum_m \left[\sum_{k\neq\ell}\left(\tau_k \cdot \pi^{(m)} + \gamma_k \cdot \sum_q I_k^{(q)}\right)\right]^2}_{O\left(\frac{1}{M^{2\beta-1}}\right)} + \left(1 - \underbrace{\sum_m I_\ell^{(m)}}_{O\left(\frac{1}{M^\beta}\right)}\right)^2 + O\left(\frac{1}{M}\right)$$

On one hand, if we increase M, then both Var(error) and Var_ℓ(simultaneous) decrease. On the other hand, if we increase M, then MSE_ℓ(interference) decreases. Therefore, we need to optimally choose M balance different terms in MSE.

B.3 Proof of Theorem 2

Proof of Theorem 2. The decomposition of the estimation error of $\hat{\delta}_\ell$ continues to hold. Then the bias decomposition follows directly from Lemmas 5-8 that provide the limit of the expected value of each term in the decomposition of $\hat{\delta}_\ell - \delta_\ell$.

For the upper bound of the limit of the MSE, from Cauchy–Schwarz inequality, we have

$$\begin{aligned}(\hat{\delta}_\ell - \delta_\ell)^2 \leq & 4\left[\left(\frac{1}{n}\sum_i \alpha_{\ell i}\cdot\gamma_\ell\cdot(I_{\ell i} - W_{\ell i})\right)^2 + \left(\frac{1}{n}\sum_i \alpha_{\ell i}\sum_{k\neq\ell}(\tau_k\cdot W_{ki} + \gamma_k\cdot I_{ki})\right)^2\right.\\ & \left. + \left(\frac{1}{n}\sum_i \alpha_{\ell i}\cdot\varepsilon_i\right)^2 + \left(\frac{1}{n}\sum_i \frac{W_{\ell i} - \pi_{\ell i}}{\pi_{\ell i}}\left(\frac{\mu_i}{1-\pi_{\ell i}} + \delta_\ell\right)\right)^2\right]\end{aligned}$$

where $\alpha_{\ell i} = \frac{W_{\ell i} - \pi_{\ell i}}{\pi_{\ell i}(1-\pi_{\ell i})}$ and $\pi_{\ell i} = \mathbf{P}(W_{\ell i} = 1)$ is the treated probability of intervention ℓ for unit i. Lemmas 5-8 provide the limit of the expected value of each term in the RHS of the above inequality. Then the upper bound of the limit of the MSE follows directly from Lemmas 5-8.

□

Lemma 5 (Correlation in Errors). *Under Assumptions 1-5, the mean of $\frac{1}{n}\sum_i \alpha_{\ell i}\cdot\varepsilon_i$,*

$$\mathbf{E}\left[\frac{1}{n}\sum_i \alpha_{\ell i}\cdot\varepsilon_i\right] = 0$$

As $n \to \infty$, the second moment converges to

$$\mathbf{E}\left[\left(\frac{1}{n}\sum_i \alpha_{\ell i}\cdot\varepsilon_i\right)^2\right] \xrightarrow{p} \sum_m \frac{C^{(m)}}{\pi_\ell^{(m)}(1-\pi_\ell^{(m)})}$$

where $C^{(m)} = \int_{(s,t)\in\mathcal{B}_m,(s',t')\in\mathcal{B}_m} \mathbf{E}\left[\varepsilon_{(s,t)}\varepsilon_{(s',t')}\right] f(s,t)f(s',t')dsdtds'dt'$. If $\pi_\ell^{(m)} \equiv 0.5$, then

$$\mathbf{E}\left[\left(\frac{1}{n}\sum_i \alpha_{\ell i}\cdot\varepsilon_i\right)^2\right] \xrightarrow{p} 4\sum_m C^{(m)}.$$

Lemma 6 (Interference). *Under Assumptions 1-5, the mean of* $\frac{1}{n}\sum_i \alpha_{\ell i} \cdot \gamma_\ell \cdot (I_{\ell i} - W_{\ell i})$ *is*

$$\mathbf{E}\left[\frac{1}{n}\sum_i \alpha_{\ell i} \cdot \gamma_\ell \cdot (I_{\ell i} - W_{\ell i})\right] \xrightarrow{p} -\gamma_\ell \sum_m \left(\pi^{(m)} - I_\ell^{(m)}\right).$$

As $n \to \infty$*, the second moment converges to*

$$\mathbf{E}\left[\left(\frac{1}{n}\sum_i \alpha_{\ell i} \cdot \gamma_\ell \cdot (I_{\ell i} - W_{\ell i})\right)^2\right]$$
$$\xrightarrow{p} \sum_m \left[\frac{1}{\pi_\ell^{(m)}}\left(\sum_{k\neq m}\left(1-\pi_\ell^{(k)}\right)I_\ell^{(m,k)}\right)^2 + \frac{1}{1-\pi_\ell^{(m)}}\left(\sum_{k\neq m}\pi_\ell^{(k)} I_\ell^{(m,k)}\right)^2\right.$$
$$\left. + \frac{1}{\pi_\ell^{(m)}(1-\pi_\ell^{(m)})}\sum_{k\neq m}\pi_\ell^{(k)}(1-\pi_\ell^{(k)})\left(I_\ell^{(m,k)}\right)^2\right] + \sum_{m\neq k}\left[\left(\pi^{(m)} - I_\ell^{(m)}\right)\left(\pi^{(k)} - I_\ell^{(k)}\right) + I_\ell^{(m,k)} I_\ell^{(k,m)}\right].$$

If $\pi_\ell^{(m)} \equiv 0.5$*, then*

$$\mathbf{E}\left[\left(\frac{1}{n}\sum_i \alpha_{\ell i} \cdot \gamma_\ell \cdot (I_{\ell i} - W_{\ell i})\right)^2\right] \xrightarrow{p} \left(1 - \sum_m I_\ell^{(m)}\right)^2 + \sum_{m\neq k}\left[\left(I_\ell^{(m,k)}\right)^2 + I_\ell^{(m,k)} I_\ell^{(k,m)}\right].$$

Lemma 7 (Interference from Other Interventions). *Under Assumptions 1-5, the mean of* $\frac{1}{n}\sum_i \alpha_{\ell i}\sum_{k\neq\ell}(\tau_k \cdot W_{ki} + \gamma_k \cdot I_{ki})$ *is* 0*, and the second moment is*

$$\mathbf{E}\left[\left(\frac{1}{n}\sum_i \alpha_{\ell i}\sum_{k\neq\ell}(\tau_k \cdot W_{ki} + \gamma_k \cdot I_{ki})\right)^2\right]$$
$$\xrightarrow{p} \sum_m \frac{1}{\pi_\ell^{(m)}(1-\pi_\ell^{(m)})}\left\{\sum_{k\neq\ell}\left[\pi_k^{(m)}(1-\pi_k^{(m)})\left(\tau_k \cdot \pi^{(m)} + \gamma_k \cdot I_k^{(m)}\right)^2 + \gamma_k^2 \sum_{q\neq m}\pi_k^{(q)}(1-\pi_k^{(q)})(I_k^{(q)})^2\right]\right.$$
$$\left. + \left[\sum_{k\neq\ell}\left(\tau_k \cdot \pi_k^{(m)}\pi^{(m)} + \gamma_k \cdot \sum_q \pi_k^{(q)} I_k^{(q)}\right)\right]^2\right\}.$$

If $\pi_k^{(m)} \equiv 0.5$ *for all* k*, then the second moment converges to*

$$E\left[\left(\frac{1}{n}\sum_i \alpha_{\ell i}\sum_{k\neq\ell}(\tau_k \cdot W_{ki} + \gamma_k \cdot I_{ki})\right)^2\right]$$
$$\xrightarrow{p} \sum_m \left\{\sum_{k\neq\ell}\left[\left(\tau_k \cdot \pi^{(m)} + \gamma_k \cdot I_k^{(m)}\right)^2 + \gamma_k^2\sum_{q\neq m}(I_k^{(q)})^2\right] + \left[\sum_{k\neq\ell}\left(\tau_k \cdot \pi^{(m)} + \gamma_k \cdot \sum_q I_k^{(q)}\right)\right]^2\right\}.$$

Lemma 8 (Constant Term). *Under Assumptions 1-5, the mean of* $\frac{1}{n}\sum_i \frac{W_{\ell i} - \pi_{\ell i}}{\pi_{\ell i}}\left(\frac{\mu_i}{1-\pi_{\ell i}} + \delta_\ell\right)$ *is* 0*, and the second moment equals*

$$\mathbf{E}\left[\left(\frac{1}{n}\sum_i \frac{W_{\ell i} - \pi_{\ell i}}{\pi_{\ell i}}\left(\frac{\mu_i}{1-\pi_{\ell i}} + \delta_\ell\right)\right)^2\right] \xrightarrow{p} \sum_m \frac{\left(\mu^{(m)} + (1-\pi_\ell^{(m)})\pi^{(m)} \cdot \delta_\ell\right)^2}{\pi_\ell^{(m)}(1-\pi_\ell^{(m)})}.$$

If $\pi_k^{(m)} \equiv 0.5$ for all k, then the second moment converges to

$$\mathbf{E}\left[\left(\frac{1}{n}\sum_i \frac{W_{\ell i}-\pi_{\ell i}}{\pi_{\ell i}}\left(\frac{\mu_i}{1-\pi_{\ell i}}+\delta_\ell\right)\right)^2\right] \xrightarrow{p} 4\sum_m \left(\mu^{(m)}+0.5\pi^{(m)}\delta_\ell\right)^2.$$

B.3.1 Proof of Lemma 5

Proof of Lemma 5. The expected value of $\frac{1}{n}\sum_i \alpha_{\ell i}\cdot\varepsilon_i$ is zero following that

$$\mathbf{E}\left[\frac{1}{n}\sum_i \alpha_{\ell i}\cdot\varepsilon_i\right] = \frac{1}{n}\sum_i \mathbf{E}[\alpha_{\ell i}]\cdot\mathbf{E}[\varepsilon_i] = 0$$

and the second moment of this term is

$$\mathbf{E}\left[\left(\frac{1}{n}\sum_i \alpha_{\ell i}\cdot\varepsilon_i\right)^2\right] = \frac{1}{n^2}\sum_i \frac{\mathbf{E}[\varepsilon_i^2]}{\pi_{\ell i}(1-\pi_{\ell i})} + \frac{1}{n^2}\sum_{i\neq j}\frac{\mathbf{E}[W_{\ell i}W_{\ell j}]-\pi_{\ell i}\pi_{\ell j}}{\pi_{\ell i}(1-\pi_{\ell i})\pi_{\ell j}(1-\pi_{\ell j})}\cdot\mathbf{E}[\varepsilon_i\varepsilon_j]$$

If $m(i)=m(j)$, then

$$\frac{\mathbf{E}[W_{\ell i}W_{\ell j}]-\pi_{\ell i}\pi_{\ell j}}{\pi_{\ell i}(1-\pi_{\ell i})\pi_{\ell j}(1-\pi_{\ell j})} = \frac{1}{\pi_{\ell i}(1-\pi_{\ell i})};$$

Otherwise,

$$\frac{\mathbf{E}[W_{\ell i}W_{\ell j}]-\pi_{\ell i}\pi_{\ell j}}{\pi_{\ell i}(1-\pi_{\ell i})\pi_{\ell j}(1-\pi_{\ell j})} = 0.$$

Combining two cases together, as $n\to\infty$, from LLN, we have

$$\mathbf{E}\left[\left(\frac{1}{n}\sum_i \alpha_{\ell i}\cdot\varepsilon_i\right)^2\right] = \frac{1}{n^2}\sum_i\sum_{j:m(j)=m(i)}\frac{\mathbf{E}[\varepsilon_i\varepsilon_j]}{\pi_{\ell i}(1-\pi_{\ell i})} \xrightarrow{p} \sum_m \frac{C_\ell^{(m)}}{\pi_\ell^{(m)}(1-\pi_\ell^{(m)})}.$$

If $\pi_\ell^{(m)}\equiv 0.5$, then the limit is simplified to

$$\mathbf{E}\left[\left(\frac{1}{n}\sum_i \alpha_{\ell i}\cdot\varepsilon_i\right)^2\right] \xrightarrow{p} 4\sum_m C_\ell^{(m)}.$$

□

B.3.2 Proof of Lemma 6

Proof of Lemma 6. The expected value of $\frac{1}{n}\sum_i \alpha_{\ell i}\cdot\gamma_\ell\cdot(I_{\ell i}-W_{\ell i})$ is nonzero and equals

$$\mathbf{E}\left[\frac{1}{n}\sum_i \alpha_{\ell i}\cdot\gamma_\ell\cdot(I_{\ell i}-W_{\ell i})\right] = \frac{\gamma_\ell}{n}\sum_i \mathbf{E}[\alpha_{\ell i}\cdot(I_{\ell i}-W_{\ell i})] = -\frac{\gamma_\ell}{n}\sum_i(1-I_{\ell i}^{m(i)})$$

where $I_{\ell i}^{(m(i))} = \int_{(s,t)\in\mathcal{B}_{m(i)}} d_{\ell,(s_i,t_i)}(s,t)f(s,t)dsdt$. The proof of $\mathbf{E}[\alpha_{\ell i}I_{\ell i}] = I_{\ell i}^{(m(i))}$ is as follows. Note that $\mathbf{E}[\alpha_{\ell i}I_{\ell i}] = \mathbf{E}[\alpha_{\ell i}\mathbf{E}[I_{\ell i} \mid W_{\ell i}]]$ and

$$\begin{aligned}\mathbf{E}[I_{\ell i} \mid W_{\ell i}] =& \mathbf{E}\left[\sum_m W_\ell^{(m)} I_{\ell i}^{(m)} \mid W_{\ell i}\right] = W_{\ell i}\cdot I_{\ell i}^{(m(i))} + \sum_{k\neq m(i)} \mathbf{E}[W_\ell^{(k)}]\cdot I_{\ell i}^{(k)} \\ =& W_{\ell i}\cdot I_{\ell i}^{(m(i))} + \sum_{k\neq m(i)} \pi_\ell^{(k)}\cdot I_{\ell i}^{(k)}.\end{aligned}$$

Since $\mathbf{E}[\alpha_{\ell i}] = 0$ and $\mathbf{E}[\alpha_{\ell i}W_{\ell i}] = 1$, we have

$$\mathbf{E}[\alpha_{\ell i}I_{\ell i}] = \mathbf{E}[\alpha_{\ell i}\mathbf{E}[I_{\ell i} \mid W_{\ell i}]] = I_{\ell i}^{(m(i))}.$$

As $n \to \infty$, from LLN, the expected value converges to

$$\mathbf{E}\left[\frac{1}{n}\sum_i \alpha_{\ell i}\cdot\gamma_\ell\cdot(I_{\ell i} - W_{\ell i})\right] = -\gamma_\ell \sum_m \frac{1}{n}\sum_{i:m(i)=m}(1 - I_{\ell i}^{m(i)}) \xrightarrow{p} -\gamma_\ell \sum_m (\pi^{(m)} - I_\ell^{(m)}),$$

where $I_\ell^{(m)} = \int_{(s,t)\in\mathcal{B}_m,(s',t')\in\mathcal{B}_m} d_{\ell,(s,t)}(s,t)f(s,t)f(s',t')dsdtds'dt'$ that equals $I_{\ell\ell}^{(m,m)}$ defined in Section 3.2. The limit is generally nonzero, implying that $\hat{\delta}_\ell$ is biased and the bias scales with how much a unit is interfered by units in other blocks.

We can consider the second moment of $\frac{1}{n}\sum_i \alpha_{\ell i}\cdot\gamma_\ell\cdot(I_{\ell i} - W_{\ell i})$.

$$\mathbf{E}\left[\left(\frac{1}{n}\sum_i \alpha_{\ell i}\cdot\gamma_\ell\cdot(I_{\ell i} - W_{\ell i})\right)^2\right] = \frac{\gamma_\ell^2}{n^2}\sum_i \mathbf{E}[\alpha_{\ell i}^2\cdot(I_{\ell i} - W_{\ell i})^2] + \frac{\gamma_\ell^2}{n^2}\sum_{i\neq j}\mathbf{E}[\alpha_{\ell i}\alpha_{\ell j}\cdot(I_{\ell i} - W_{\ell i})(I_{\ell j} - W_{\ell j})]$$

Let us first analyze the term $\mathbf{E}[\alpha_{\ell i}^2\cdot(I_{\ell i} - W_{\ell i})^2]$.

$$\mathbf{E}[\alpha_{\ell i}^2\cdot(I_{\ell i} - W_{\ell i})^2] = \mathbf{E}[\alpha_{\ell i}^2(W_{\ell i}^2 - 2W_{\ell i}\mathbf{E}[I_{\ell i} \mid W_{\ell i}] + \mathbf{E}[I_{\ell i}^2 \mid W_{\ell i}])]$$

For the term $\mathbf{E}[I_{\ell i}^2 \mid W_{\ell i}]$, we have

$$\begin{aligned}&\mathbf{E}[I_{\ell i}^2 \mid W_{\ell i}] = \mathbf{E}\left[\left(\sum_m W_\ell^{(m)} I_{\ell i}^{(m)}\right)^2 \mid W_{\ell i}\right] \\ =& W_{\ell i}\cdot(I_{\ell i}^{(m(i))})^2 + 2W_{\ell i}\cdot I_{\ell i}^{(m(i))}\cdot\sum_{k\neq m(i)}\mathbf{E}[W_\ell^{(k)}]\cdot I_{\ell i}^{(k)} + E\left[\left(\sum_{k\neq m(i)} W_\ell^{(k)}\cdot I_{\ell i}^{(k)}\right)^2\right] \\ =& W_{\ell i}\cdot(I_{\ell i}^{(m(i))})^2 + 2W_{\ell i}\cdot I_{\ell i}^{(m(i))}\cdot\sum_{k\neq m(i)}\pi_\ell^{(k)}\cdot I_{\ell i}^{(k)} + \sum_{k\neq m(i)}\pi_\ell^{(k)}\cdot(I_{\ell i}^{(k)})^2 + \sum_{k,j\neq m(i),k\neq j}\pi_\ell^{(k)}\pi_\ell^{(j)}\cdot I_{\ell i}^{(k)} I_{\ell i}^{(j)}.\end{aligned}$$

Since $\mathbf{E}[\alpha_{\ell i}^2] = \frac{1}{\pi_{\ell i}(1-\pi_{\ell i})}$ and $\mathbf{E}[\alpha_{\ell i}^2 W_{\ell i}] = \mathbf{E}[\alpha_{\ell i}^2 W_{\ell i}^2] = \frac{1}{\pi_{\ell i}}$, we have

$$\begin{aligned}
\mathbf{E}[\alpha_{\ell i}^2 \cdot (I_{\ell i} - W_{\ell i})^2] =& \frac{1}{\pi_{\ell i}} - \frac{2}{\pi_{\ell i}} \big[I_{\ell i}^{(m(i))} + \sum_{k \neq m(i)} \pi_\ell^{(k)} \cdot I_{\ell i}^{(k)} \big] \\
&+ \frac{1}{\pi_{\ell i}} \left[(I_{\ell i}^{(m(i))})^2 + 2 \cdot I_{\ell i}^{(m(i))} \cdot \sum_{k \neq m(i)} \pi_\ell^{(k)} \cdot I_{\ell i}^{(k)} \right] \\
&+ \frac{1}{\pi_{\ell i}(1-\pi_{\ell i})} \left[\sum_{k \neq m(i)} \pi_\ell^{(k)} \cdot (I_{\ell i}^{(k)})^2 + \sum_{k,j \neq m(i), k \neq j} \pi_\ell^{(k)} \pi_\ell^{(j)} \cdot I_{\ell i}^{(k)} I_{\ell i}^{(j)} \right] \\
=& \frac{1}{\pi_{\ell i}} \left(1 - I_{\ell i}^{(m(i))} \right)^2 - \frac{2}{\pi_{\ell i}} (1 - I_{\ell i}^{(m(i))}) \cdot \sum_{k \neq m(i)} \pi_\ell^{(k)} \cdot I_{\ell i}^{(k)} \\
&+ \frac{1}{\pi_{\ell i}(1-\pi_{\ell i})} \left[\sum_{k \neq m(i)} \pi_\ell^{(k)} \cdot (I_{\ell i}^{(k)})^2 + \sum_{k,j \neq m(i), k \neq j} \pi_\ell^{(k)} \pi_\ell^{(j)} \cdot I_{\ell i}^{(k)} I_{\ell i}^{(j)} \right].
\end{aligned}$$

Let us consider the cross term $\mathbf{E}[\alpha_{\ell i}\alpha_{\ell j} \cdot (I_{\ell i} - W_{\ell i})(I_{\ell j} - W_{\ell j})]$.

$$\begin{aligned}
&\mathbf{E}[\alpha_{\ell i}\alpha_{\ell j} \cdot (I_{\ell i} - W_{\ell i})(I_{\ell j} - W_{\ell j})] \\
=&\mathbf{E}[\alpha_{\ell i}\alpha_{\ell j} \cdot (\mathbf{E}[I_{\ell i} I_{\ell j} \mid W_{\ell i}, W_{\ell j}] - \mathbf{E}[I_{\ell j} \mid W_{\ell i}, W_{\ell j}] \cdot W_{\ell i} - \mathbf{E}[I_{\ell i} \mid W_{\ell i}, W_{\ell j}] \cdot W_{\ell j} + W_{\ell i} W_{\ell j})]
\end{aligned}$$

This term depends on whether i and j are in the same block or not.
If i and j are in the same block, i.e., $m(i) = m(j)$, then $\mathbf{E}[I_{\ell i} I_{\ell j} \mid W_{\ell i}, W_{\ell j}]$ equals

$$\begin{aligned}
&\mathbf{E}[I_{\ell i} I_{\ell j} \mid W_{\ell i}, W_{\ell j}] = \mathbf{E}\left[\left(\sum_m W_\ell^{(m)} I_{\ell i}^{(m)} \right) \left(\sum_m W_\ell^{(m)} I_{\ell j}^{(m)} \right) \mid W_{\ell i}, W_{\ell j} \right] \\
=& W_{\ell i} \cdot I_{\ell i}^{(m(i))} I_{\ell j}^{(m(i))} + W_{\ell i} \cdot I_{\ell i}^{(m(i))} \cdot \sum_{k \neq m(i)} \underbrace{\mathbf{E}[W_\ell^{(k)}]}_{\pi_\ell^{(k)}} \cdot I_{\ell j}^{(k)} + W_{\ell i} \cdot I_{\ell j}^{(m(i))} \cdot \sum_{k \neq m(i)} \underbrace{\mathbf{E}[W_\ell^{(k)}]}_{\pi_\ell^{(k)}} \cdot I_{\ell i}^{(k)} \\
&+ \sum_{k \neq m(i)} \underbrace{\mathbf{E}[W_\ell^{(k)}]}_{\pi_\ell^{(k)}} \cdot I_{\ell i}^{(k)} I_{\ell j}^{(k)} + \sum_{k,p \neq m(i), k \neq p} \underbrace{\mathbf{E}[W_\ell^{(k)}]}_{\pi_\ell^{(k)}} \underbrace{\mathbf{E}[W_\ell^{(p)}]}_{\pi_\ell^{(p)}} \cdot I_{\ell i}^{(k)} I_{\ell j}^{(p)}.
\end{aligned}$$

$\mathbf{E}[I_{\ell j} \mid W_{\ell i}, W_{\ell j}]$ equals

$$\mathbf{E}[I_{\ell j} \mid W_{\ell i}, W_{\ell j}] = \mathbf{E}\left[\sum_m W_\ell^{(m)} I_{\ell j}^{(m)} \mid W_{\ell i}, W_{\ell j} \right] = W_{\ell i} I_{\ell j}^{(m(i))} + \sum_{k \neq m(i)} \pi_\ell^{(k)} \cdot I_{\ell j}^{(k)},$$

and $\mathbf{E}[I_{\ell i} \mid W_{\ell i}, W_{\ell j}]$ equals

$$\mathbf{E}[I_{\ell i} \mid W_{\ell i}, W_{\ell j}] = W_{\ell i} I_{\ell i}^{(m(i))} + \sum_{k \neq m(i)} \pi_\ell^{(k)} \cdot I_{\ell i}^{(k)}.$$

Combining these terms together, since $\mathbf{E}[\alpha_{\ell i}\alpha_{\ell j}] = \frac{1}{\pi_{\ell i}(1-\pi_{\ell i})}$, $\mathbf{E}[\alpha_{\ell i}\alpha_{\ell j} W_{\ell i}] = \mathbf{E}[\alpha_{\ell i}\alpha_{\ell j} W_{\ell j}] =$

$\mathbf{E}[\alpha_{\ell i}\alpha_{\ell j}W_{\ell i}W_{\ell j}] = \frac{1}{\pi_{\ell i}}$, we have

$$\begin{aligned}
&\mathbf{E}[\alpha_{\ell i}\alpha_{\ell j}\cdot(I_{\ell i}-W_{\ell i})(I_{\ell j}-W_{\ell j})]\\
=&\frac{1}{\pi_{\ell i}}\cdot I_{\ell i}^{(m(i))}I_{\ell j}^{(m(i))}+\frac{1}{\pi_{\ell i}}\cdot I_{\ell i}^{(m(i))}\sum_{k\neq m(i)}\pi_\ell^{(k)}\cdot I_{\ell j}^{(k)}+\frac{1}{\pi_{\ell i}}\cdot I_{\ell j}^{(m(i))}\sum_{k\neq m(i)}\pi_\ell^{(k)}\cdot I_{\ell i}^{(k)}\\
&+\frac{1}{\pi_{\ell i}(1-\pi_{\ell i})}\cdot\sum_{k\neq m(i)}\pi_\ell^{(k)}\cdot I_{\ell i}^{(k)}I_{\ell j}^{(k)}+\frac{1}{\pi_{\ell i}(1-\pi_{\ell i})}\cdot\sum_{k,p\neq m(i),k\neq p}\pi_\ell^{(k)}\pi_\ell^{(p)}\cdot I_{\ell i}^{(k)}I_{\ell j}^{(p)}\\
&-\left(\frac{1}{\pi_{\ell i}}I_{\ell j}^{(m(i))}+\frac{1}{\pi_{\ell i}}\sum_{k\neq m(i)}\pi_\ell^{(k)}\cdot I_{\ell j}^{(k)}\right)-\left(\frac{1}{\pi_{\ell i}}I_{\ell i}^{(m(i))}+\frac{1}{\pi_{\ell i}}\sum_{k\neq m(i)}\pi_\ell^{(k)}\cdot I_{\ell i}^{(k)}\right)+\frac{1}{\pi_{\ell i}}\\
=&\frac{1}{\pi_{\ell i}}\left(1-I_{\ell i}^{(m(i))}\right)\left(1-I_{\ell j}^{(m(i))}\right)-\frac{1}{\pi_{\ell i}}\left(1-I_{\ell i}^{(m(i))}\right)\cdot\sum_{k\neq m(i)}\pi_\ell^{(k)}\cdot I_{\ell j}^{(k)}-\frac{1}{\pi_{\ell i}}\left(1-I_{\ell j}^{(m(i))}\right)\cdot\sum_{k\neq m(i)}\pi_\ell^{(k)}\cdot I_{\ell i}^{(k)}\\
&+\frac{1}{\pi_{\ell i}(1-\pi_{\ell i})}\cdot\left[\sum_{k\neq m(i)}\pi_\ell^{(k)}\cdot I_{\ell i}^{(k)}I_{\ell j}^{(k)}+\sum_{k\neq m(i),p\neq m(i),k\neq p}\pi_\ell^{(k)}\pi_\ell^{(p)}\cdot I_{\ell i}^{(k)}I_{\ell j}^{(p)}\right].
\end{aligned}$$

If i and j are in different blocks, i.e., $m(i)\neq m(j)$, then $\mathbf{E}[I_{\ell i}I_{\ell j}\mid W_{\ell i},W_{\ell j}]$ equals

$$\begin{aligned}
&\mathbf{E}[I_{\ell i}I_{\ell j}\mid W_{\ell i},W_{\ell j}]=\mathbf{E}\left[\left(\sum_m W_\ell^{(m)}I_{\ell i}^{(m)}\right)\left(\sum_m W_\ell^{(m)}I_{\ell j}^{(m)}\right)\mid W_{\ell i},W_{\ell j}\right]\\
=&(W_{\ell i}\cdot I_{\ell i}^{(m(i))}+W_{\ell j}\cdot I_{\ell i}^{(m(j))})(W_{\ell i}\cdot I_{\ell j}^{(m(i))}+W_{\ell j}\cdot I_{\ell j}^{(m(j))})\\
&+(W_{\ell i}\cdot I_{\ell i}^{(m(i))}+W_{\ell j}\cdot I_{\ell i}^{(m(j))})\cdot\sum_{k\neq m(i),m(j)}\underbrace{\mathbf{E}[W_\ell^{(k)}]}_{\pi_\ell^{(k)}}\cdot I_{\ell j}^{(k)}\\
&+(W_{\ell i}\cdot I_{\ell j}^{(m(i))}+W_{\ell j}\cdot I_{\ell j}^{(m(j))})\cdot\sum_{k\neq m(i),m(j)}\underbrace{\mathbf{E}[W_\ell^{(k)}]}_{\pi_\ell^{(k)}}\cdot I_{\ell i}^{(k)}\\
&+\sum_{k\neq m(i),m(j)}\underbrace{\mathbf{E}[W_\ell^{(k)}]}_{\pi_\ell^{(k)}}\cdot I_{\ell i}^{(k)}I_{\ell j}^{(k)}+\sum_{k,p\neq m(i),m(j),k\neq p}\underbrace{\mathbf{E}[W_\ell^{(k)}]}_{\pi_\ell^{(k)}}\underbrace{\mathbf{E}[W_\ell^{(p)}]}_{\pi_\ell^{(p)}}\cdot I_{\ell i}^{(k)}I_{\ell j}^{(p)},
\end{aligned}$$

$\mathbf{E}[I_{\ell j}\mid W_{\ell i},W_{\ell j}]$ equals

$$\begin{aligned}
\mathbf{E}[I_{\ell j}\mid W_{\ell i},W_{\ell j}]=&\mathbf{E}\left[\sum_m W_\ell^{(m)}I_{\ell j}^{(m)}\mid W_{\ell i},W_{\ell j}\right]\\
=&W_{\ell i}I_{\ell j}^{(m(i))}+W_{\ell j}I_{\ell j}^{(m(j))}+\sum_{k\neq m(i),m(j)}\pi_\ell^{(k)}\cdot I_{\ell j}^{(k)},
\end{aligned}$$

and $\mathbf{E}[I_{\ell i}\mid W_{\ell i},W_{\ell j}]$ equals

$$\mathbf{E}[I_{\ell i}\mid W_{\ell i},W_{\ell j}]=W_{\ell i}I_{\ell i}^{(m(i))}+W_{\ell j}I_{\ell i}^{(m(j))}+\sum_{k\neq m(i),m(j)}\pi_\ell^{(k)}\cdot I_{\ell i}^{(k)}.$$

Combining these terms together, since $\mathbf{E}[\alpha_{\ell i}\alpha_{\ell j}]=\mathbf{E}[\alpha_{\ell i}]\mathbf{E}[\alpha_{\ell j}]=0$, $\mathbf{E}[\alpha_{\ell i}\alpha_{\ell j}W_{\ell i}]=\mathbf{E}[\alpha_{\ell i}\alpha_{\ell j}W_{\ell i}^2]=\mathbf{E}[\alpha_{\ell i}\alpha_{\ell j}W_{\ell j}]=\mathbf{E}[\alpha_{\ell i}\alpha_{\ell j}W_{\ell j}^2]=0$, and $\mathbf{E}[\alpha_{\ell i}\alpha_{\ell j}W_{\ell i}W_{\ell j}]=\mathbf{E}[\alpha_{\ell i}W_{\ell i}]\mathbf{E}[\alpha_{\ell j}W_{\ell j}]=1$, we have

$$\begin{aligned}
\mathbf{E}[\alpha_{\ell i}\alpha_{\ell j}\cdot(I_{\ell i}-W_{\ell i})(I_{\ell j}-W_{\ell j})]=&I_{\ell i}^{(m(i))}I_{\ell j}^{(m(j))}+I_{\ell i}^{(m(j))}I_{\ell j}^{(m(i))}-I_{\ell j}^{(m(j))}-I_{\ell i}^{(m(i))}+1\\
=&(1-I_{\ell i}^{(m(i))})(1-I_{\ell j}^{(m(j))})+I_{\ell i}^{(m(j))}I_{\ell j}^{(m(i))}.
\end{aligned}$$

Then as $n \to \infty$, from LLN, we have

$$
\begin{aligned}
&\mathbf{E}\left[\left(\frac{1}{n}\sum_i \alpha_{\ell i}\cdot\gamma_\ell\cdot(I_{\ell i}-W_{\ell i})\right)^2\right] \\
=&\sum_m \frac{1}{n^2}\sum_{i,j:m(i)=m(j)=m}\left\{\frac{1}{\pi_\ell^{(m)}}\left(1-I_{\ell i}^{(m)}\right)\left(1-I_{\ell j}^{(m)}\right)-\frac{1}{\pi_\ell^{(m)}}\left(1-I_{\ell i}^{(m)}\right)\cdot\sum_{k\neq m}\pi_\ell^{(k)}\cdot I_{\ell j}^{(k)}\right. \\
&-\frac{1}{\pi_\ell^{(m)}}\left(1-I_{\ell j}^{(m)}\right)\cdot\sum_{k\neq m}\pi_\ell^{(k)}\cdot I_{\ell i}^{(k)} \\
&\left.+\frac{1}{\pi_\ell^{(m)}(1-\pi_\ell^{(m)})}\cdot\left[\sum_{k\neq m}\pi_\ell^{(k)}\cdot I_{\ell i}^{(k)}I_{\ell j}^{(k)}+\sum_{k\neq m,p\neq m,k\neq p}\pi_\ell^{(k)}\pi_\ell^{(p)}\cdot I_{\ell i}^{(k)}I_{\ell j}^{(p)}\right]\right\} \\
&+\sum_{m\neq k}\frac{1}{n^2}\sum_{i,j:m(i)=m,m(j)=k}\left[(1-I_{\ell i}^{(m)})(1-I_{\ell j}^{(k)})+I_{\ell i}^{(k)}I_{\ell j}^{(m)}\right] \\
\xrightarrow{p}&\sum_m\left[\frac{1}{\pi_\ell^{(m)}}\left(\pi^{(m)}-I_\ell^{(m)}\right)^2-\frac{2}{\pi_\ell^{(m)}}\left(\pi^{(m)}-I_\ell^{(m)}\right)\sum_{k\neq m}\pi_\ell^{(k)}I_\ell^{(m,k)}+\frac{1}{\pi_\ell^{(m)}(1-\pi_\ell^{(m)})}\left(\sum_{k\neq m}\pi_\ell^{(k)}I_\ell^{(m,k)}\right)^2\right. \\
&\left.+\frac{1}{\pi_\ell^{(m)}(1-\pi_\ell^{(m)})}\sum_{k\neq m}\pi_\ell^{(k)}(1-\pi_\ell^{(k)})\left(I_\ell^{(m,k)}\right)^2\right]+\sum_{m\neq k}\left[\left(\pi^{(m)}-I_\ell^{(m)}\right)\left(\pi^{(k)}-I_\ell^{(k)}\right)+I_\ell^{(m,k)}I_\ell^{(k,m)}\right]. \\
=&\sum_m\left[\frac{1}{\pi_\ell^{(m)}}\left(\sum_{k\neq m}\left(1-\pi_\ell^{(k)}\right)I_\ell^{(m,k)}\right)^2+\frac{1}{1-\pi_\ell^{(m)}}\left(\sum_{k\neq m}\pi_\ell^{(k)}I_\ell^{(m,k)}\right)^2\right. \\
&\left.+\frac{1}{\pi_\ell^{(m)}(1-\pi_\ell^{(m)})}\sum_{k\neq m}\pi_\ell^{(k)}(1-\pi_\ell^{(k)})\left(I_\ell^{(m,k)}\right)^2\right]+\sum_{m\neq k}\left[\left(\pi^{(m)}-I_\ell^{(m)}\right)\left(\pi^{(k)}-I_\ell^{(k)}\right)+I_\ell^{(m,k)}I_\ell^{(k,m)}\right].
\end{aligned}
$$

If $\pi_\ell^{(m)} \equiv 0.5$, then the limit is simplified to

$$
\begin{aligned}
&\mathbf{E}\left[\left(\frac{1}{n}\sum_i \alpha_{\ell i}\cdot\gamma_\ell\cdot(I_{\ell i}-W_{\ell i})\right)^2\right] \\
\xrightarrow{p}&\sum_m\left[\left(\sum_{k\neq m}I_\ell^{(m,k)}\right)^2+\sum_{k\neq m}\left(I_\ell^{(m,k)}\right)^2\right]+\sum_{m\neq k}\left[\left(\pi^{(m)}-I_\ell^{(m)}\right)\left(\pi^{(k)}-I_\ell^{(k)}\right)+I_\ell^{(m,k)}I_\ell^{(k,m)}\right] \\
=&\left(1-\sum_m I_\ell^{(m)}\right)^2+\sum_{m\neq k}\left[\left(I_\ell^{(m,k)}\right)^2+I_\ell^{(m,k)}I_\ell^{(k,m)}\right].
\end{aligned}
$$

□

B.3.3 Proof of Lemma 7

Proof of Lemma 7. The expected value of $\frac{1}{n}\sum_i \alpha_{\ell i}\sum_{k\neq\ell}(\tau_k \cdot W_{ki} + \gamma_k \cdot I_{ki})$ is

$$\mathbf{E}\left[\frac{1}{n}\sum_i \alpha_{\ell i}\sum_{k\neq\ell}(\tau_k \cdot W_{ki} + \gamma_k \cdot I_{ki})\right] = \frac{1}{n}\sum_i \mathbf{E}[\alpha_{\ell i}] \cdot \sum_{k\neq\ell}\mathbf{E}\left[\tau_k \cdot W_{ki} + \gamma_k \cdot I_{ki}\right] = 0.$$

The second moment equals

$$\begin{aligned}\mathbf{E}\left[\left(\frac{1}{n}\sum_i \alpha_{\ell i}\sum_{k\neq\ell}(\tau_k \cdot W_{ki} + \gamma_k \cdot I_{ki})\right)^2\right] &= \frac{1}{n^2}\sum_i \mathbf{E}[\alpha_{\ell i}^2] \cdot \mathbf{E}\left[\left(\sum_{k\neq\ell}(\tau_k \cdot W_{ki} + \gamma_k \cdot I_{ki})\right)^2\right] \\ &+ \frac{1}{n^2}\sum_{i\neq j}\mathbf{E}[\alpha_{\ell i}\alpha_{\ell j}] \cdot \mathbf{E}\left[\left(\sum_{k\neq\ell}(\tau_k \cdot W_{ki} + \gamma_k \cdot I_{ki})\right)\left(\sum_{k\neq\ell}(\tau_k \cdot W_{kj} + \gamma_k \cdot I_{kj})\right)\right]\end{aligned}$$

Let us calculate the first term. The expected value of $\tau_k \cdot W_{ki} + \gamma_k \cdot I_{ki}$ is

$$\mathbf{E}[\tau_k \cdot W_{ki} + \gamma_k \cdot I_{ki}] = \tau_k \cdot \pi_{ki} + \gamma_k \cdot \mathbf{E}\left[\sum_m W_k^{(m)} I_{ki}^{(m)}\right] = \tau_k \cdot \pi_{ki} + \gamma_k \cdot \sum_m \pi_k^{(m)} I_{ki}^{(m)}$$

and the second moment equals

$$\begin{aligned}\mathbf{E}[(\tau_k \cdot W_{ki} + \gamma_k \cdot I_{ki})^2] =& \tau_k^2 \cdot \pi_{ki} + 2\tau_k\gamma_k\pi_{ki} \cdot \left(\cdot I_{ki}^{(m(i))} + \sum_{p\neq m(i)} \pi_k^{(p)} \cdot I_{ki}^{(p)}\right) \\ &+ \gamma_k^2\left(\sum_m \pi_k^{(m)}(I_{ki}^{(m)})^2 + \sum_{m\neq p}\pi_k^{(m)}\pi_k^{(p)} I_{ki}^{(m)} I_{ki}^{(p)}\right),\end{aligned}$$

where we use $\mathbf{E}[W_{ki}I_{ki}]$ equals

$$\begin{aligned}\mathbf{E}[W_{ki}I_{ki}] =& \mathbf{E}[W_{ki} \cdot \mathbf{E}[I_{ki} \mid W_{ki}]] = \mathbf{E}\left[W_{ki} \cdot I_{ki}^{(m(i))} + W_{ki}\sum_{p\neq m(i)}\pi_k^{(p)} \cdot I_{ki}^{(p)}\right] \\ =& \pi_{ki} \cdot I_{ki}^{(m(i))} + \pi_{ki}\sum_{p\neq m(i)}\pi_k^{(p)} \cdot I_{ki}^{(p)}\end{aligned}$$

and $\mathbf{E}[I_{ki}^2]$ equals

$$\mathbf{E}[I_{ki}^2] = \mathbf{E}\left[\left(\sum_m W_k^{(m)} I_{ki}^{(m)}\right)^2\right] = \sum_m \pi_k^{(m)}(I_{ki}^{(m)})^2 + \sum_{m\neq p}\pi_k^{(m)}\pi_k^{(p)} I_{ki}^{(m)} I_{ki}^{(p)}.$$

As any two interventions are independent, the cross term of intervention k and h equals

$$\begin{aligned}&\mathbf{E}\left[(\tau_k \cdot W_{ki} + \gamma_k \cdot I_{ki})(\tau_h \cdot W_{hi} + \gamma_h \cdot I_{hi})\right] = \mathbf{E}\left[(\tau_k \cdot W_{ki} + \gamma_k \cdot I_{ki})\right] \cdot \mathbf{E}\left[(\tau_h \cdot W_{hi} + \gamma_h \cdot I_{hi})\right] \\ =& \left(\tau_k \cdot \pi_{ki} + \gamma_k \cdot \sum_m \pi_k^{(m)} I_{ki}^{(m)}\right)\left(\tau_h \cdot \pi_{hi} + \gamma_h \cdot \sum_m \pi_h^{(m)} I_{hi}^{(m)}\right)\end{aligned}$$

Then the first term equals

$$
\begin{aligned}
&\mathbf{E}[\alpha_{\ell i}^2]\cdot\mathbf{E}\left[\left(\sum_{k\neq\ell}(\tau_k\cdot W_{ki}+\gamma_k\cdot I_{ki})\right)^2\right]\\
=&\frac{1}{\pi_{\ell i}(1-\pi_{\ell i})}\sum_{k\neq\ell}\left[\tau_k^2\cdot\pi_{ki}+2\tau_k\gamma_k\pi_{ki}\cdot\left(I_{ki}^{(m(i))}+\sum_{p\neq m(i)}\pi_k^{(p)}\cdot I_{ki}^{(p)}\right)\right.\\
&\left.+\gamma_k^2\left(\sum_m\pi_k^{(m)}(I_{ki}^{(m)})^2+\sum_{m\neq p}\pi_k^{(m)}\pi_k^{(p)}I_{ki}^{(m)}I_{ki}^{(p)}\right)\right]\\
&+\frac{1}{\pi_{\ell i}(1-\pi_{\ell i})}\sum_{k\neq\ell,h\neq\ell,k\neq h}\left(\tau_k\cdot\pi_{ki}+\gamma_k\cdot\sum_m\pi_k^{(m)}I_{ki}^{(m)}\right)\left(\tau_h\cdot\pi_{hi}+\gamma_h\cdot\sum_m\pi_h^{(m)}I_{hj}^{(m)}\right).
\end{aligned}
$$

For the second term, if $m(i)\neq m(j)$, then $\mathbf{E}[\alpha_{\ell i}\alpha_{\ell j}]=0$; if $m(i)=m(j)$, then $\mathbf{E}[\alpha_{\ell i}\alpha_{\ell j}]=\frac{1}{\pi_{\ell i}(1-\pi_{\ell i})}$ and

$$
\begin{aligned}
&\mathbf{E}\left[\left(\sum_{k\neq\ell}(\tau_k\cdot W_{ki}+\gamma_k\cdot I_{ki})\right)\left(\sum_{k\neq\ell}(\tau_k\cdot W_{kj}+\gamma_k\cdot I_{kj})\right)\right]\\
=&\sum_{k\neq\ell}\left[\tau_k^2\cdot\pi_{ki}+\tau_k\gamma_k\pi_{ki}\cdot\left(I_{ki}^{(m(i))}+I_{kj}^{(m(i))}+\sum_{p\neq m(i)}\pi_k^{(p)}\cdot(I_{ki}^{(p)}+I_{kj}^{(p)})\right)\right.\\
&\left.+\gamma_k^2\left(\sum_m\pi_k^{(m)}I_{ki}^{(m)}I_{kj}^{(m)}+\sum_{m\neq p}\pi_k^{(m)}\pi_k^{(p)}I_{ki}^{(m)}I_{kj}^{(p)}\right)\right]\\
&+\sum_{k\neq\ell,h\neq\ell,k\neq h}\left(\tau_k\cdot\pi_{ki}+\gamma_k\cdot\sum_m\pi_k^{(m)}I_{ki}^{(m)}\right)\left(\tau_h\cdot\pi_{hj}+\gamma_h\cdot\sum_m\pi_h^{(m)}I_{hj}^{(m)}\right).
\end{aligned}
$$

In summary, as $n \to \infty$, from LLN, the second moment converges to

$$
\begin{aligned}
&\mathbf{E}\left[\left(\frac{1}{n}\sum_i \alpha_{\ell i}\sum_{k\neq\ell}(\tau_k \cdot W_{ki} + \gamma_k \cdot I_{ki})\right)^2\right] \\
=&\sum_m \frac{1}{\pi_\ell^{(m)}(1-\pi_\ell^{(m)})}\frac{1}{n^2}\sum_{i,j:m(i)=m(j)=m}\left\{\sum_{k\neq\ell}\left[\tau_k^2\cdot\pi_k^{(m)} + \tau_k\gamma_k\pi_k^{(m)}\cdot\left(I_{ki}^{(m)} + I_{kj}^{(m)} + \sum_{p\neq m}\pi_k^{(p)}\cdot(I_{ki}^{(p)} + I_{kj}^{(p)})\right)\right.\right. \\
&\left. + \gamma_k^2\left(\sum_q \pi_k^{(q)} I_{ki}^{(q)} I_{kj}^{(q)} + \sum_{q\neq p}\pi_k^{(q)}\pi_k^{(p)} I_{ki}^{(q)} I_{kj}^{(p)}\right)\right] \\
&\left. + \sum_{k\neq\ell,h\neq\ell,k\neq h}\left(\tau_k\cdot\pi_{ki} + \gamma_k\cdot\sum_q \pi_k^{(q)} I_{ki}^{(q)}\right)\left(\tau_h\cdot\pi_{hj} + \gamma_h\cdot\sum_q \pi_h^{(q)} I_{hj}^{(q)}\right)\right\} \\
\xrightarrow{p}&\sum_m \frac{1}{\pi_\ell^{(m)}(1-\pi_\ell^{(m)})}\left\{\sum_{k\neq\ell}\left[\tau_k^2\cdot\pi_k^{(m)}(\pi^{(m)})^2 + 2\tau_k\gamma_k\cdot\pi_k^{(m)}\pi^{(m)}\left(I_k^{(m)} + \sum_{p\neq m}\pi_k^{(p)}\cdot I_k^{(p)}\right)\right.\right. \\
&\left. + \gamma_k^2\left(\sum_q \pi_k^{(q)}(I_k^{(q)})^2 + \sum_{q\neq p}\pi_k^{(q)}\pi_k^{(p)} I_k^{(q)} I_k^{(p)}\right)\right] \\
&\left. + \sum_{k\neq\ell,h\neq\ell,k\neq h}\left(\tau_k\cdot\pi_k^{(m)}\pi^{(m)} + \gamma_k\cdot\sum_q \pi_k^{(q)} I_k^{(q)}\right)\left(\tau_h\cdot\pi_h^{(m)}\pi^{(m)} + \gamma_h\cdot\sum_q \pi_h^{(q)} I_h^{(q)}\right)\right\}. \\
\xrightarrow{p}&\sum_m \frac{1}{\pi_\ell^{(m)}(1-\pi_\ell^{(m)})}\left\{\sum_{k\neq\ell}\left[\pi_k^{(m)}(1-\pi_k^{(m)})\left(\tau_k\cdot\pi^{(m)} + \gamma_k\cdot I_k^{(m)}\right)^2 + \gamma_k^2\sum_{q\neq m}\pi_k^{(q)}(1-\pi_k^{(q)})(I_k^{(q)})^2\right]\right. \\
&\left. + \left[\sum_{k\neq\ell}\left(\tau_k\cdot\pi_k^{(m)}\pi^{(m)} + \gamma_k\cdot\sum_q \pi_k^{(q)} I_k^{(q)}\right)\right]^2\right\}.
\end{aligned}
$$

If $\pi_k^{(m)} \equiv 0.5$ for all k, then the second moment converges to

$$
\begin{aligned}
&\mathbf{E}\left[\left(\frac{1}{n}\sum_i \alpha_{\ell i}\sum_{k\neq\ell}(\tau_k \cdot W_{ki} + \gamma_k \cdot I_{ki})\right)^2\right] \\
\xrightarrow{p}&\sum_m\left\{\sum_{k\neq\ell}\left[\left(\tau_k\cdot\pi^{(m)} + \gamma_k\cdot I_k^{(m)}\right)^2 + \gamma_k^2\sum_{q\neq m}(I_k^{(q)})^2\right] + \left[\sum_{k\neq\ell}\left(\tau_k\cdot\pi^{(m)} + \gamma_k\cdot\sum_q I_k^{(q)}\right)\right]^2\right\}.
\end{aligned}
$$

□

B.3.4 Proof of Lemma 8

Proof of Lemma 8. Let us consider the term $\frac{1}{n}\sum_i \frac{W_{\ell i}-\pi_{\ell i}}{\pi_{\ell i}}\left(\frac{\mu_i}{1-\pi_{\ell i}}+\delta_\ell\right)$. First the expected value of this term is zero

$$\mathbf{E}\left[\frac{1}{n}\sum_i \frac{W_{\ell i}-\pi_{\ell i}}{\pi_{\ell i}}\left(\frac{\mu_i}{1-\pi_{\ell i}}+\delta_\ell\right)\right]=\frac{1}{n}\sum_i \frac{\mathbf{E}[W_{\ell i}]-\pi_{\ell i}}{\pi_{\ell i}}\left(\frac{\mu_i}{1-\pi_{\ell i}}+\delta_\ell\right)=0$$

and the second moment of this term is

$$\begin{aligned}&\mathbf{E}\left[\left(\frac{1}{n}\sum_i \frac{W_{\ell i}-\pi_{\ell i}}{\pi_{\ell i}}\left(\frac{\mu_i}{1-\pi_{\ell i}}+\delta_\ell\right)\right)^2\right]\\ =&\frac{1}{n^2}\sum_i\Big(\frac{\mu_i^2}{\pi_{\ell i}(1-\pi_{\ell i})}+\frac{2\mu_i\delta_\ell}{\pi_{\ell i}}+\frac{1-\pi_{\ell i}}{\pi_{\ell i}}\cdot\delta_\ell^2\Big)\\ &+\frac{1}{n^2}\sum_{i\neq j}\frac{\mathbf{E}[W_{\ell i}W_{\ell j}]-\pi_{\ell i}\pi_{\ell j}}{\pi_{\ell i}\pi_{\ell j}}\left(\frac{\mu_i}{1-\pi_{\ell i}}+\delta_\ell\right)\left(\frac{\mu_j}{1-\pi_{\ell j}}+\delta_\ell\right).\end{aligned}$$

If $m(i)\neq m(j)$, then $\mathbf{E}[W_{\ell i}W_{\ell j}]=\mathbf{E}[W_{\ell i}]\mathbf{E}W_{\ell j}]=\pi_{\ell i}\pi_{\ell j}$. Otherwise, if$m(i)=m(j)$, then $\frac{\mathbf{E}[W_{\ell i}W_{\ell j}]-\pi_{\ell i}\pi_{\ell j}}{\pi_{\ell i}\pi_{\ell j}}=\frac{1-\pi_{\ell i}}{\pi_{\ell i}}$ and then the second moment is simplified to

$$\begin{aligned}&\mathbf{E}\left[\left(\frac{1}{n}\sum_i \frac{W_{\ell i}-\pi_{\ell i}}{\pi_{\ell i}}\left(\frac{\mu_i}{1-\pi_{\ell i}}+\delta_\ell\right)\right)^2\right]\\ =&\frac{1}{n^2}\sum_i\sum_{j:m(i)=m(j)}\Big(\frac{\mu_i\mu_j}{\pi_{\ell i}(1-\pi_{\ell i})}+\frac{(\mu_i+\mu_j)\delta_\ell}{\pi_{\ell i}}+\frac{1-\pi_{\ell i}}{\pi_{\ell i}}\cdot\delta_\ell^2\Big).\end{aligned}$$

As $n\to\infty$, from LLN, the second moment converges to

$$\mathbf{E}\left[\left(\frac{1}{n}\sum_i \frac{W_{\ell i}-\pi_{\ell i}}{\pi_{\ell i}}\left(\frac{\mu_i}{1-\pi_{\ell i}}+\delta_\ell\right)\right)^2\right]\xrightarrow{p}\sum_m\frac{\left(\mu^{(m)}+(1-\pi_\ell^{(m)})\pi^{(m)}\cdot\delta_\ell\right)^2}{\pi_\ell^{(m)}(1-\pi_\ell^{(m)})}.$$

If $\pi_k^{(m)}\equiv 0.5$ for all k, then the limit is simplified to

$$\mathbf{E}\left[\left(\frac{1}{n}\sum_i \frac{W_{\ell i}-\pi_{\ell i}}{\pi_{\ell i}}\left(\frac{\mu_i}{1-\pi_{\ell i}}+\delta_\ell\right)\right)^2\right]\xrightarrow{p}4\sum_m\left(\mu^{(m)}+0.5\pi^{(m)}\delta_\ell\right)^2.$$

□