Data-Driven Switchback Design

Ruoxuan Xiong
Department of Quantitative Theory and Methods, Emory University, ruoxuan.xiong@emory.edu

Alex Chin
Motif Analytics, alexchin12@gmail.com

Sean Taylor
Motif Analytics, seanjtaylor@gmail.com

We study the design and analysis of experiments conducted on an aggregate unit over time, and outcomes are measured on a sequence of events. The design problem is to partition the continuous time space into intervals and switch treatment between intervals, in order to reduce the estimation error of the treatment effect. Prior work has studied this problem when the effect of treatment carries over to future outcomes. We observe that besides carryover effects, the estimation error depends on three other factors: nonuniform event density, correlated event outcomes, and interference from simultaneous experiments. To account for these factors, we propose a new approach to design experiments using a meta-analysis of historical data. This approach is illustrated through a case study of a large corpus of experiments on a ride-sharing platform. The case study shows balancing time heterogeneity to be especially beneficial for reducing estimation error, followed by selecting interval lengths and endpoints. We then provide a careful bias-variance decomposition of the estimated treatment effects, accounting for all four factors. The decomposition reveals that balancing reduces variance by offsetting time heterogeneity between treated and control intervals. Selecting interval lengths can find the best tradeoff between bias from carryover effects and variance from correlated event outcomes. Fine-tuning interval endpoints can reduce both bias and variance from simultaneous experiments. Finally, simulations are conducted to generalize our findings from the case study to other settings with varying specifications.

Key words: Time-Based Experiment, Carryover Effect, Simultaneous Intervention, Event Analysis, Treatment Effect Estimation, Ride-Sharing Platform

1. Introduction

Experimentation has become an increasingly popular and effective tool for testing and improving social and business policy in digitally mediated economics and social settings. However, the scale and complexity of modern digital settings create many scientific and statistical challenges for the design and analysis of experiments.

Consider a ride-hailing platform where multiple teams would like to measure the effects of their product changes through experiments, while only a limited number of markets are available for experimentation. In this setting, designing and analyzing experiments can be challenging for four prominent reasons. First, users exposed to new interventions may vary their behavior in ways that create interference and affect outcomes for other users on both the rider and driver side of the marketplace (Chamandy 2016). Second, the interventions usually take time to be effective and change the marketplace to a new equilibrium state, resulting in time-varying effects of the intervention. Third, the significant variations and serial correlations in rider demand and driver availability throughout the day and week complicate the analysis of the intervention effect. Fourth, interventions tested simultaneously may interact with one another, complicating the measurement of the marginal effect of each one. Due to these four reasons, developing reliable approaches that allow for precise measurement of the intervention effects can be quite challenging, but is crucial when deciding whether to roll out the new intervention.

A commonly used solution that makes it possible to measure the intervention effects is to run experiments via time-based or *temporal* experimental designs rather than (the far more common) cross-sectional designs. Time-based designs operate on a (few) aggregated units and randomly switch between treatment and control over time for each unit, which are colloquially known as "switchback designs" (Bojinov et al. 2023). Units' longitudinal observations are then used to estimate the treatment effects. These designs help to mitigate interference and have become popular due to their applications in digital marketplaces. 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).

In this paper, we analyze switchback designs 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. We study how to partition the temporal space into intervals with alternating treatments in these designs, in anticipation of precisely estimating a quantity called global average treatment effects (GATE). GATE is an important estimand for decision-making that captures the difference in average outcomes between when an intervention is deployed indefinitely (global treatment) versus when the intervention is absent indefinitely (global control).

Our analysis captures the realistic data properties in digital marketplaces that complicate the design and analysis of switchback experiments. We novelly do so through four distinct lenses. First, we factor in the carryover effects of treatments on future event outcomes. Second, we consider the nonuniform density of observed events, where an event can be a session of a rider checking price. Third, we account for correlation in event outcomes stemming from unobserved (or unmodeled) factors that create nuisance dependence among measurements; outcomes of events close in time can be similar due to weather, traffic, or other external factors. Lastly, we consider the presence of simultaneous experiments run by other teams on the same sequence of events, which may confound effect estimates in finite samples.

1.1. Summary of Contributions

We theoretically and empirically show how the four realistic data properties affect the estimation efficiency from switchback experiments. We propose *data-driven switchback designs* that use prior experiments in designing new experiments, which can effectively account for the realistic data properties in digital marketplaces. Practitioners can use similar data-driven switchback designs tailored to their empirical settings, when armed with prior experiments.

We start with a case study of a corpus of 890 prior experiments run in various markets on a ride-sharing platform. We estimate a curve between cumulative effect and treatment time for each experiment-market pair.[1] To summarize the information in all the estimated curves, we use interpolative decomposition, in which the top 10 curves explain 75% variation of all the curves. Interestingly, the top curves show that the cumulative effect is not always monotonic in treatment time and can switch signs. In these cases, the estimation of GATE is particularly challenging and is prone to be biased.

We then illustrate the data-driven switchback designs by running synthetic experiments on real data. The cumulative effect of synthetic intervention follows one of the top 10 curves. We compare the estimation error of GATE for various heuristic designs when the cumulative effect varies and other data properties are present. We identify a *hierarchical structure* in the effectiveness of design principles in this case study: (a) *balancing* time heterogeneity (such as time-of-day effect) is the most effective; (b) carefully *selecting average switching periods* is moderately effective; (c) *fine-tuning exact switching times* is mildly effective.

To understand the hierarchical structure, we provide a rigorous decomposition of bias and mean-squared error (MSE) of the estimated GATE from the standard Horvitz-Thompson estimators (Horvitz and Thompson 1952). The bias is decomposed into two sources of errors: (a) carryover effects from treatment at earlier times; (b) confounding effects from simultaneous interventions. The MSE is decomposed into squared bias and variance, where the variance is affected by three sources of randomness: (a) the measurement errors of event outcomes and their covariance, determined by their distance in time; (b) the randomness of treatment assignments of focal and simultaneous interventions; (c) the randomness in event occurrence times.

The decomposition together with a careful simulation study explains how the design principles in the case study reduce the estimation error: (a) balancing time heterogeneities reduces all sources of variance; (b) longer switching periods reduce bias from carryovers; (c) shorter switching periods reduce variance from the randomness in measurement errors and treatment assignments; (d) properly staggering switching times of simultaneous interventions reduces both bias and variance

[1] GATE equals the cumulative effect when the treatment time grows to infinity.

from simultaneous interventions; (e) fine-tuning switching times based on event density reduces variance from irregular event occurrences. These insights manifest that the hierarchical structure identified in the ride-sharing setting is due to the low signal-to-noise ratio in most cases; once the variance is reduced by balancing, bias dominates, and carefully choosing switching periods becomes important.

1.2. Related Work

Our data-driven switchback design is most closely related to the recent literature on switchback designs. Bojinov et al. (2023) first study the minimax optimal design of switchback experiments on an experimental unit in the presence of temporal interference. Ni et al. (2023) then advances the study to multiple experimental units in the presence of both temporal and spatial interference. Hu and Wager (2022) concern the case of never fully vanishing carryovers and propose to use burn-in periods in the estimator to reduce estimation error. Chen and Simchi-Levi (2023) propose a new importance sampling estimator to improve statistical efficiency. We complement this literature and show that, besides the carryover effect, nonuniform event density, correlation event outcomes, and simultaneous interventions can affect the performance of switchback design. Due to these factors, we show the value of using prior data and propose new designs to improve efficiency.

Our design is also related to a number of other designs in time-based experiments. One design is the staggered rollout design for panel experiments (Xiong et al. 2023), where the design selects an initial (and possibly different) treatment time for each unit. The objective is to precisely estimate the treatment effect either under temporal interference (Xiong et al. 2023, Basse et al. 2023) or network interference (Cortez et al. 2022, Han et al. 2022, Boyarsky et al. 2023). Another design is the synthetic control design for panel experiments, where the design selects units to be treated, allocates treatment to all of them in a single period, and forms a synthetic treated and control unit for treatment effect estimation (Doudchenko et al. 2019, 2021, Abadie and Zhao 2021). Concerning time heterogeneity, Wu et al. (2022) propose a design that groups sequentially arriving users into consecutive pairs, and randomly treats one user in each pair. Our balanced design complements Wu et al. (2022) and provides a solution to balance time heterogeneity when experimenting on an aggregate unit.

We consider the data in the form of a stream of events and design the experiment for units that are aggregated to a level where interference between users can be abstracted away. But we note a growing literature directly tackles interference using novel experimental ideas. For example, on network data, cluster-randomized designs are commonly used for mitigating interference (Ugander et al. 2013, Eckles et al. 2017, Candogan et al. 2021, Holtz et al. 2023), where the clusters are chosen to minimize edges that cut across clusters. Another popular method is the two-stage or multi-stage randomization, which has been used in public health (Hudgens and Halloran 2008, Liu and Hudgens 2014), digital platforms (Ye et al. 2023a), political science (Sinclair et al. 2012), and social science (Crépon et al. 2013, Baird et al. 2018, Basse and Feller 2018). In the two-sided marketplace, multiple randomization (Bajari et al. 2023, Johari et al. 2022) that randomizes at multiple sides are used and designs that perturb treatments near equilibrium outcomes (Wager and Xu 2021). Li et al. (2021) characterize the bias and variance of such experiments and describe how the design can be optimized in such settings. Besides using novel designs, there is a growing literature on developing new treatment effect estimators and inferential theory accounting for the interference (Chin (2018, 2019), Forastiere et al. (2021), Qu et al. (2021), Yuan et al. (2021), Leung (2022, 2023), Farias et al. (2022) among others). Complementing this literature, our paper takes an agnostic approach to the marketplace interference structure in the design and analysis.

Finally, this paper accounts for the impact of interventions tested simultaneously. When multiple interventions are simultaneously applied to the same units, factorial design (Fisher 1936) is commonly used, which allows for estimating the effect of any treatment combination. However, the design and estimation are complex with a large number of interventions (Dasgupta et al. 2015).[2] Ye et al. (2023b) provide a nice solution to this problem using debiased deep learning. In this paper, we take a different perspective and study the design of the main intervention while being agnostic about the design of simultaneous interventions.

2. Problem Setup

Suppose a decision maker runs an experiment from time 0 to T to study the effect of a new intervention. For example, the intervention can be a new pricing, matching, or routing algorithm. Let $w_t \in \{0,1\}$ be the treatment status at time $t \in [0,T]$, where $w_t = 1$ indicates that the marketplace is exposed to intervention ℓ (treatment) at time t, and $w_t = 0$ indicates otherwise (control).

Before the experiment starts, the decision maker chooses the treatment design for the whole experiment horizon, i.e., $\boldsymbol{W} = \{W_t \in \{0,1\}, \forall t \in [0,T]\}$. Since the treatment decisions need to be made in a continuous time interval, the decision maker first partitions experimental horizon $[0,T]$ into M disjoint intervals and then randomly chooses the treatment assignment of each interval. Let $0 \le t_0 \le t_1 \le \cdots \le t_{M-1} \le t_M = T$ be the endpoints that define the M intervals, let $\mathcal{I}_m = [t_{m-1}, t_m]$ be the m-th interval, and let $|\mathcal{I}_m| = t_m - t_{m-1}$ be the length of the m-th interval.

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

$$w_t = w_{t'}, \qquad \text{for all } t, t' \in \mathcal{I}_m, \quad \text{for all } m.$$

[2] This is because the number of treatment combinations increases exponentially in the number of interventions.

When the treatment assignments vary by interval, the design is referred to as the switchback design (Bojinov et al. 2023). We provide a few examples of switchback designs in Section 2.3.

The raw data available for analyzing the effect of intervention are at the event level, where each event could be a rider opening the app and checking the price. Suppose there are n events occurring in the marketplace between time 0 and time T. Let $Y^{(i)}$ be the outcome of event i that occurred at time t_i, where we assume the occurred time t_i is a random variable. For example, $Y^{(i)}$ could be a binary variable indicating whether the rider requests a ride or not. Let $f(t):[0,T]\to\mathbb{R}^+$ be the density function from which events are sampled. We assume that $f(t)$ is bounded from below and from above for all t. For simplicity, it is possible to consider the uniform event density as in Example 2.1. However, in many realistic settings, the density of events will 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. Then it is possible to consider periodic event density as in Example 2.2.

Example 2.1 (Uniform event density). If events are equally likely to occur at any time in the experiment, then $f(t)=1/T$ for all $t\in[0,T]$.

Example 2.2 (Periodic event density). If there is a periodic pattern in event density, then it is possible to use the periodic function, such as $f(t)=a_1\sin(a_2t+a_3)+a_4$ for some constants a_1, a_2, a_3 and a_4 and $t\in[0,T]$, to capture the periodic event density. See Figure 1 for an example.

Besides the event outcome, we additionally define the marketplace outcome at time t as Y_t. The marketplace outcome Y_t can be viewed as the average outcome of all users in the marketplace, such as the average request rate at time t. Then the event outcome is a noisy measurement of the marketplace outcome, i.e., for all i,

$$Y^{(i)}=Y_{t_i}+\varepsilon^{(i)},$$

where the measurement error $\varepsilon^{(i)}$ has mean zero and bounded variance. For example, when $Y^{(i)}$ is binary indicating whether rider i requests a ride, we can model $Y^{(i)}$ as a random draw from the Bernoulli distribution with probability $\mathbf{P}(Y^{(i)}=1)=Y_{t_i}$ of being 1.

Importantly, measurement errors of events that are close in time can be correlated:

$$\mathrm{Cov}[\varepsilon^{(i)},\varepsilon^{(j)}]\neq 0 \qquad \text{for} \quad t_i\neq t_j\,.$$

The correlation can be caused by external factors like weather, supply conditions, and traffic. This correlation creates a nuisance dependence between event outcomes, which can affect the resulting variance of treatment effect estimates.

We further define the potential outcomes of the marketplace. Here we account for the possibility that other decision makers may run experiments simultaneously to test the effect of other interventions. Let us refer to the experiment that the main decision maker runs as the main experiment.

Suppose K experiments are run simultaneously in addition to the main experiment. We allow K to be zero or nonzero. When K is zero, no experiment is running simultaneously with the main experiment. Let the treatment designs of the K simultaneous experiments be $\boldsymbol{W}_1^{\mathrm{s}}, \cdots, \boldsymbol{W}_K^{\mathrm{s}}$, where $\boldsymbol{W}_\ell^{\mathrm{s}} = \{W_{\ell,t}^{\mathrm{s}} \in \{0,1\}, \forall t \in [0,T]\}$ for $\ell = 1, \cdots, K$. We assume that the treatment designs of simultaneous experiments are chosen independently of the main experiment and independently of one another. We primarily focus on the case where the main decision maker is *agnostic* to the treatment decisions of simultaneous experiments. We further assume the non-anticipating outcome, i.e., the outcome at time t is only affected by the treatment assignments up to time t (Basse et al. (2023) among others). We use $\boldsymbol{W}_t = \{W_u, \forall u \in [0,t]\}$ and $\boldsymbol{W}_{\ell,t}^{\mathrm{s}} = \{W_{\ell,u}^{\mathrm{s}}, \forall u \in [0,t]\}$ to denote the treatment assignments of the main and ℓ-th simultaneous experiment up to time t.

With simultaneous interventions and non-anticipating outcomes, the potential outcomes of the marketplace at time t are defined as

$$Y_t(\boldsymbol{w}_t, \boldsymbol{w}_{1,t}^{\mathrm{s}}, \cdots, \boldsymbol{w}_{K,t}^{\mathrm{s}}),$$

where $\boldsymbol{w}_t$ is a realization of $\boldsymbol{W}_t$ and $\boldsymbol{w}_{\ell,t}^{\mathrm{s}}$ is a realization of $\boldsymbol{W}_{\ell,t}^{\mathrm{s}}$ for all ℓ.[3] The marketplace outcome satisfies $Y_t = Y_t(\boldsymbol{W}_t, \boldsymbol{W}_{1,t}^{\mathrm{s}}, \cdots, \boldsymbol{W}_{K,t}^{\mathrm{s}})$. Given treatment designs $\boldsymbol{W}_t, \boldsymbol{W}_{1,t}^{\mathrm{s}}, \cdots, \boldsymbol{W}_{K,t}^{\mathrm{s}}$ and event occurrence time t_i, there is no randomness in Y_{t_i} anymore, and the randomness in $Y^{(i)}$ purely comes from the measurement error $\varepsilon^{(i)}$.

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 the main and K simultaneous interventions. Second, this definition allows for the existence of carryover effects: the potential outcome of t is not only affected by the treatment status at t but also the treatment assignments at other times.

2.1. Estimands

Post-experiment, the main decision maker uses the observed event outcomes $\{Y^{(i)}\}_{i\in[n]}$ and treatment assignments $\boldsymbol{W}$ to estimate the effect of the intervention. The objective is to decide whether to deploy the intervention indefinitely. Based on this objective, the estimand of primary interest is the *global average treatment effect* (GATE), which measures the difference in average outcomes over time when an intervention is deployed indefinitely (global treatment) versus when an intervention is absent (global control). We formally define the GATE as

$$\delta^{\mathrm{gate}} = \int \delta_t^{\mathrm{gate}} f(t) dt,$$

[3] Suppose both the main intervention and simultaneous interventions are not applied to times outside of the experiment duration, i.e., $w_t = 0$ and $w_{\ell,t}^{\mathrm{s}} = 0$ for $t \notin [0,T]$. Therefore, there are no carryover effects from times outside of the experiment duration, $\mathbb{R}\backslash[0,T]$, to the experiment duration, $[0,T]$. It is then reasonable to define potential outcomes only using treatment assignments within the experiment duration.

which is the average of the total treatment effect δ_t^{gate} at time t weighted by the event density $f(t)$. The total treatment effect δ_t^{gate} at time t is defined as

$$\delta_{\ell,t}^{\text{gate}} = Y_t(\mathbf{1}_t, \mathbf{0}_t, \cdots, \mathbf{0}_t) - Y_t(\mathbf{0}_t, \mathbf{0}_t, \cdots, \mathbf{0}_t)\,,$$

where $\mathbf{1}_t = \{w_u = 1, \forall u \in [0,t]\}$ and $\mathbf{0}_t = \{w_u = 0, \forall u \in [0,t]\}$ denote being and not being in the treatment state for a time duration of t, respectively. In the definition of GATE, simultaneous interventions are held in the global control state. This definition makes sense as the main decision maker is interested in the effect of the main intervention, while holding other conditions as the status quo.

When the intervention has not been employed indefinitely, the decision maker may also want to learn how the treatment effect varies with the treatment duration. We define the cumulative effect at time t given the treatment duration t' as

$$\delta_{\ell,t}^{\text{cum}}(\mathbf{1}_{t'}) = Y_t((\mathbf{0}_{t-t'}, \mathbf{1}_{t'}), \mathbf{0}_t, \cdots, \mathbf{0}_t) - Y_t(\mathbf{0}_t, \mathbf{0}_t, \cdots, \mathbf{0}_t)\,.$$

Based on this definition, the cumulative effect $\delta_{\ell,t}^{\text{cum}}(\mathbf{1}_{t'})$ converges to the total treatment effect $\delta_{\ell,t}^{\text{gate}}$ as the treatment duration t' grows to infinity, i.e.,

$$\delta_{\ell,t}^{\text{gate}} = \lim_{t' \to \infty} \delta_{\ell,t}^{\text{cum}}(\mathbf{1}_{t'})\,.$$

However, it is worth noting that an infinite treatment may not be necessary for the cumulative effect to stabilize and converge to the total treatment effect. Gaining insights into both the necessary duration for convergence and the dynamics of the cumulative effect is valuable for the decision maker. These insights are helpful for understanding the mechanism of how the treatment affects the outcome and designing more efficient experiments.

2.2. Post-Experiment Estimation

We propose to use the Horvitz-Thompson (HT) estimator (Horvitz and Thompson 1952) that estimates δ^{gate} from the observed event outcomes and treatment design:

$$\hat{\delta}^{\text{gate}} = \frac{1}{n}\sum_i \left(\frac{W_{t_i}}{\pi} - \frac{1-W_{t_i}}{1-\pi}\right) Y^{(i)} = \frac{1}{n}\sum_i \alpha_{t_i} Y^{(i)}, \tag{2.1}$$

where $\alpha_{t_i} = \frac{W_{t_i}-\pi}{\pi(1-\pi)}$ is a normalized weight, and

$$\pi = \int_{t\in[0,T]} \mathbb{E}[W_t] f(t) dt$$

is the fraction of treated times under intervention ℓ.

We use the HT estimator for three reasons. First, it does not rely on an assumption about carryover 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 assignments of simultaneous interventions. Due to these three reasons, the HT estimator is flexible and broadly applicable to a wide range of settings in practice.

However, the flexibility of the HT estimator comes at a cost. First, the HT estimator could be biased due to the carryover effect of the same treatment at other times. The HT estimator approximates the outcomes under global treatment by the event outcomes in treated intervals and approximates the outcomes under global control by the event outcomes in control intervals. When the carryover effect is zero, i.e., $\delta_t^{\mathrm{co}}(\boldsymbol{w}) = 0$, the approximation error is zero. For general cases, the approximation error is non-zero, and the HT estimator is biased. The bias scales with the size of the carryover effect. Second, the HT estimator can have a large variance, since the effective sample size is affected by the correlation of event outcomes at different times and the HT estimator does not optimally weight observations. Third, the HT estimator could have a confounding bias from simultaneous interventions when the treatment designs of two interventions are not orthogonal in finite samples.

It is possible to reduce the estimation error of GATE in two ways. First, we can use a better treatment design, which is the main focus of the remaining sections. We compare the performance of various designs using an empirical study in Section 3, and identify several useful design principles in reducing the estimation error. In Section 4, we provide derive a bias-variance decomposition of the estimation error that shows how different sources of errors trade-off. In Section 5, we conduct a comparative study on simulated data to show how each source of errors can be reduced by making appropriate design choices.

Second, we can use a more efficient estimator for GATE by leveraging prior knowledge of carryover and correlation mechanisms and information on other interventions. Specifically, to reduce the carryover bias, we can specify the structure of the carryover mechanisms and estimate instantaneous and carryover effects, and use these quantities to estimate GATE. To reduce the variance from correlated outcomes, we can specify the structure of the correlation mechanisms and use the structure to reweight the event outcomes. If we are aware of the treatment designs of experiments, we can simultaneously estimate the treatment effects of all interventions and reduce confounding bias from simultaneous experiments. For example, we can use the generalized least squares (GLS) estimators that simultaneously estimate the instantaneous and carryover effects for all interventions, while taking advantage of the inverse error covariance weighting. GLS can be more efficient when the model specification in GLS is accurate. In the following sections, our results are based on the HT estimator for the simplicity of exposition and for its benefit of being agnostic to model specification. However, the insights from our analysis generally carry over to alternative estimators.[4]

2.3. Switchback Design

Before the experiment starts, the decision maker chooses the number of intervals M and the interval switching points $0 \le t_1 \le \cdots \le t_{M-1} \le T$, aiming to reduce the estimation error of GATE, post-experiment. An important metric for the decision maker is the MSE of $\hat{\delta}^{\text{gate}}$, defined as

$$\mathbb{E}_{W,\varepsilon,t}\left[\left(\hat{\delta}^{\text{gate}} - \delta^{\text{gate}}\right)^2\right] \tag{2.2}$$

where the expectation is taken with respect to the treatment designs $\boldsymbol{W}, \boldsymbol{W}_1^{\text{s}}, \cdots, \boldsymbol{W}_K^{\text{s}}$, the measurement errors in event outcomes $\varepsilon^{(1)}, \cdots, \varepsilon^{(n)}$, and the event occurrence times $t_1, \cdots, t_n$. Here we focus on the randomized designs, where each time period is equally likely to be treated or untreated, i.e., $\mathbf{P}(W_t = 1) = 1/2$ and $\mathbf{P}(W_{\ell,t}^{\text{s}} = 1) = 1/2$ for all ℓ and t.

In Section 4, we provide the expression of MSE as a function of the interval endpoints. The function is highly complex and nonconvex, so finding the global optimal solution is generally infeasible. Instead we focus on the evaluation of the following three types of heuristic designs, in which cases the decision maker only needs to choose two parameters, and the design problem is substantially simplified. The first type is the fixed duration switchback (Example 2.3), which has constant interval lengths and is most commonly used in practice. The second type is the Poisson duration switchback (Example 2.4), where the length length is random and generated by the Poisson duration. We use Poisson duration switchback to explore the effect of randomizing interval lengths and switching times on the estimation error of GATE. The third type is the change-of-measure switchback (Example 2.5), which has constant interval lengths after changing the measure of nonuniform event density to uniform density. We use change-of-measure switchback to explore the effect of accounting for nonuniform density in the design.

Example 2.3 (Fixed duration switchback). The first interval starts at time $t_0 = q$ for some $q < T/M$, and the length of all the intervals besides the last one is $p = T/M$. The endpoints are then equal to $t_m = m \cdot p + q$ for all m.

Example 2.4 (Poisson duration switchback). The first interval starts at time $t_0 = q$. The length of each interval $t_m - t_{m-1}$ is randomly drawn from the Poisson distribution with the mean parameter $\lambda = T/M$. We sum the lengths of the first to the m-th intervals to obtain the value of the endpoint $t_{\ell m}$.[5]

[4] See Xiong et al. (2023) for a few examples showing that the effective design principles are robust to the choice of the estimator of treatment effects.

[5] If the endpoints of some intervals are bigger than the experiment duration (i.e., there exists some $\bar{M}$ such that $t_{m'} > T$ for $m' \ge \bar{M}$), then we set the endpoints of these intervals to T (i.e., set $t_{m'}$ to T for $m' \ge \bar{M}$ and then the lengths of the last $M - \bar{M}$ are zero).

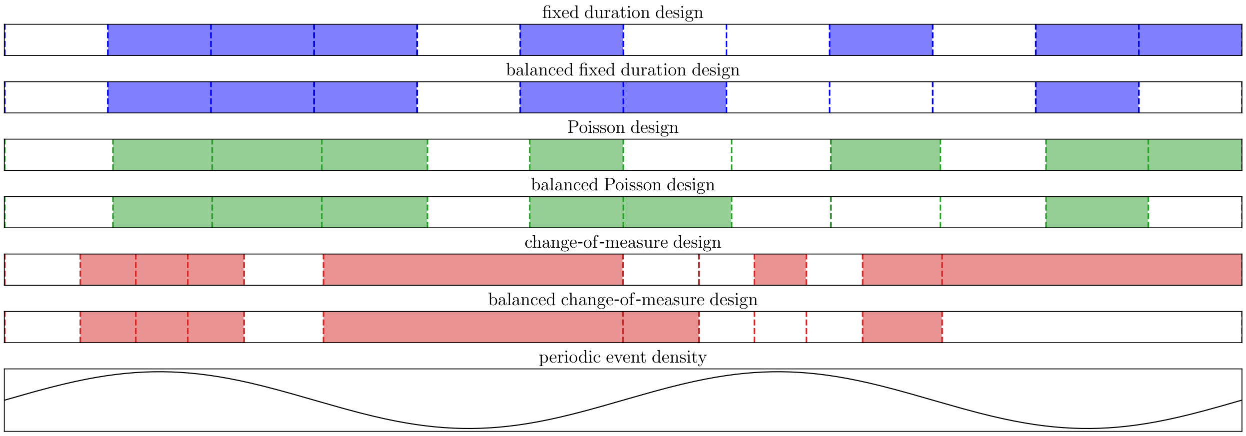


Figure 1 Illustration of various designs and periodic event density (dash lines are switching points, and treated intervals are shaded).

Example 2.5 (Change-of-Measure Switchback). The first interval starts at time $t_0 = q$ for some q that satisfies $\int_0^q f(t)dt < 1/M$. For the remaining endpoints, they are chosen in a way that the event occurrence probability is the same across intervals, i.e., $\int_{t_m}^{t_{m+1}} f(t)dt = 1/M$.

Figure 1 illustrates the switching points of the three designs under the periodic event density. The Poisson duration switchback has similar, but slightly different, interval lengths compared to the fixed duration switchback, due to the randomness in switching times of the Poisson duration switchback. The change-of-measure switchback is the same as the fixed duration switchback under the uniform event density, but they are different under irregular event density. Under the periodic density, the change-of-measure design has much shorter interval lengths in times of high density and much longer interval lengths in times of low density, as compared to the other two designs.

Besides the three types of heuristic designs for choosing interval endpoints, we consider the following balanced design that imposes restrictions on the randomness of treatment assignments.

Example 2.6 (Balanced Randomized Design). Suppose for $t \leq T/2$, $f(t) = f(t+T/2)$ and $Y_t(\boldsymbol{w}_t, \boldsymbol{w}^{\mathrm{s}}_{1,t}, \cdots, \boldsymbol{w}^{\mathrm{s}}_{K,t}) = Y_{t+T/2}(\boldsymbol{w}_{t+T/2}, \boldsymbol{w}^{\mathrm{s}}_{1,t+T/2}, \cdots, \boldsymbol{w}^{\mathrm{s}}_{K,t+T/2})$. For $t > T/2$, the treatment assignment at time t is opposite to that at time $t - T/2$, i.e., $W_{\ell t} = 1 - W_{\ell, t-T/2}$.

For example, when a balanced randomized switchback is used in a two-week experiment, the treatment assignments of the second week are opposite to those of the first week. As the heterogeneities in the outcomes tend to have a periodic pattern by week, the balanced design in a two-week experiment creates matched pairs for the same time in a week.[6] As shown in the following sections, balancing is quite useful for the reduction of MSE.

In addition, we show in the following sections the performance of various designs the assumptions on carryovers, outcome covariance, event density, and simultaneous interventions. The optimality

[6] Balancing the time heterogeneity is found effective in reducing variance in the nonstationary a/b tests (Wu et al. 2022).

of designs also depends on these assumptions. Therefore, making appropriate assumptions is crucial for choosing a switchback design before the experiment starts. In Section 3, we analyze the historical control and experimental data on a ride-sharing platform. The analysis provides guidance on how to make reasonable assumptions, and shows how the assumptions imposed impact the performance of various designs.

3. A Case Study on Ride-Sharing Platform

In this section, we analyze historical data from a ride-sharing platform and explore strategies for designing more efficient experiments. We have access to two sets of historical data. The first data set consists of the event-level data of the top 50 regions between June 2022 and March 2023, referred to as the historical control data thereafter. The second data set includes the event-level data of a large corpus of experiments conducted between June 2021 and March 2023, referred to as the historical experimental data thereafter. In both data sets, each event represents a rider session started from opening the app. The outcome is binary denoting whether the rider requested the ride ($Y_{t_i} = 1$) or not ($Y_{t_i} = 0$).

In Section 3.1, we perform an exploratory analysis on both data sets and conduct a meta-analysis of the dynamics of cumulative effects using historical experimental data. In Section 3.2, we conduct synthetic experiments using the estimated cumulative effects from meta-analysis and evaluate the performance of various designs. We identify a hierarchical structure of the effectiveness of design principles in reducing the estimation error of GATE.

3.1. Analysis of Historical Data

The setup introduced in Section 2 encompasses considerations such as event density, measurement errors, and carryovers. In this subsection, we show the estimates of these components from the historical data. The insights gained can be used as guidance on imposing reasonable assumptions on these components when designing a new experiment.

3.1.1. Event Density Figure 2 shows the estimated event density $f(t)$ from the historical control data across different minutes in a week. There are two main observations. First, the event density has a periodic pattern, with high density during the peak hours, such as 6 PM, and low density during the off-peak times, such as 3 AM. Second, during peak hours on weekends (Fridays, Saturdays, and Sundays), the event density is higher than that during the peak hours on weekdays.

3.1.2. Global Control Outcomes and Heterogeneous Measurement Errors Figure 2 shows the standardized global control outcome (i.e., $Y_t(\mathbf{0}_t, \cdots, \mathbf{0}_t)$ subtracted by its mean and divided by its standard deviation) across various minutes in a week. The average global control outcome has a periodic pattern and is generally higher during the daytime.

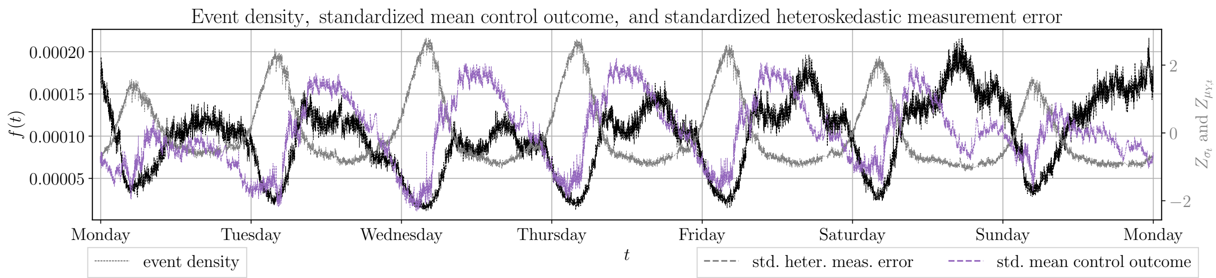


Figure 2 Event density, standardized mean control outcome (denoted by $Z_{\mu_{Y,t}}$), and standardized heteroscedastic measurement error (denoted by Z_{σ_t}) from Monday 12 AM to Sunday 11:59 PM.

Figure 2 further shows the standardized estimated standard error of measurement errors across various minutes in a week. We estimate the standard errors by assuming that the event outcome is drawn from a Bernoulli distribution with the probability of the average conversation rate (i.e., global control outcome) being the value of one. The standard error is heteroscedastic and has a periodic pattern, similar to that of the event density and global control outcome. An interesting observation is that the standard error of measurement errors tends to be lower in times of high event density, and higher in times of low event density. The intuition is that during periods of high event density, there are more events to be averaged over in a minute, resulting in a lower standard error; conversely, during times of low event density, the standard error tends to be higher.

3.1.3. Meta-Analysis of Historical Experiments We analyze a large corpus of the experiments run between June 2021 and March 2023 to construct a prior on the treatment mechanism. Some experiments are run on multiple regions and over different durations. To obtain high-quality experimental data, we focus on the experiments run for the full two weeks (were not stopped early) on larger regions, were effective, and were not an airport test. These filter criteria leave us with 149 two-week experiments run on 114 regions with 890 distinct experiment-region pairs in total.

These experiments employ a fixed-duration switchback design with a constant interval length of 56 minutes. We fit a curve of cumulative effects $\delta^{\text{cum}}_{\ell,t}(\mathbf{1}_{t'})$ versus treatment duration t' for each experiment-region pair. To achieve this, we only use the intervals whose preceding interval is the control interval, to prevent carryover effects from the previous treatment interval. We then calculate the difference in outcomes between the treated and control intervals for every minute since the switch. For each experiment-region pair, we obtain a cumulative effect curve – a 56-dimensional vector where the j-th entry represents the cumulative effect of treating the region for j minutes.

We then obtain 890 cumulative effect curves, one for each experiment-region pair. To summarize the information and reduce dimensionality in these curves, we apply interpolative decomposition. Figure 3 shows the top 10 curves from the decomposition. These 10 curves can explain approximately 75% variation across all 890 curves, demonstrating a substantial reduction in dimensionality.

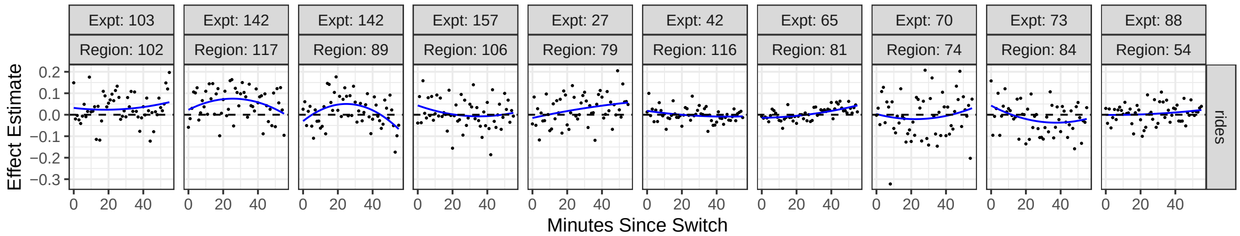


Figure 3 Estimated cumulative effects $\delta^{\mathrm{cum}}_{\ell,t}(\mathbf{1}_{t'})$ (black dots) for $t' \in \{1, \cdots, 56\}$ minutes and their smooth quadratic curves (blue curves).

Moreover, these 10 curves can be considered as basis functions to generate a new cumulative effect curve for a new treatment. To locally smooth the fitted (nonsmooth) curve, we subsequently a quadratic fit.

Two important observations arise from the smoothed cumulative effect curves. First, it is quite common (8 out of 10 curves) that the sign of cumulative effects switches as the treatment duration varies. This suggests that the treatment may first be effective in increasing the conversion rate and then become ineffective. The opposite case, where the treatment is first ineffective and then becomes effective, is also likely to happen. Second, in many cases, the cumulative effects do not vary monotonically with the treatment duration. For example, the cumulative effects can first increase and then decrease, or vice versa, with the treatment duration. In summary, the stabilization and convergence of cumulative effects to the GATE can take dozens of minutes. This observation is robust to the polynomial degree in the smoothed curve. See Figures 11 and 12 for the linear and cubic fits of cumulative effects, respectively.

3.2. Synthetic Experiments

We run synthetic experiments on historical data using various heuristic switchback designs. Results from synthetic experiments illustrate the efficacy of different design principles in reducing the estimation error of GATE in realistic (ride-sharing) settings. Based on the insights from synthetic experiments, decision makers can select specific switchback designs tailored to their specific setting. The designs selected from this approach are referred to as the *data-driven switchback designs*.

3.2.1. Setup of Synthetic Experiments We consider the following six switchback designs to run synthetic experiments: fixed duration (FD), balanced fixed duration (bal. FD), change-of-measure (CM), balanced change-of-measure (bal. CM), Poisson duration (Poisson), and balanced Poisson duration (bal. Poisson) switchback designs. For each design, we vary the "average interval length" across three settings: 28, 56, and 112 minutes. In the Poisson duration switchback, the "average interval length" corresponds to the mean parameter λ. In the change-of-measure design, the "average interval length" represents the event occurrence probability in an interval multiplied by the experiment duration.

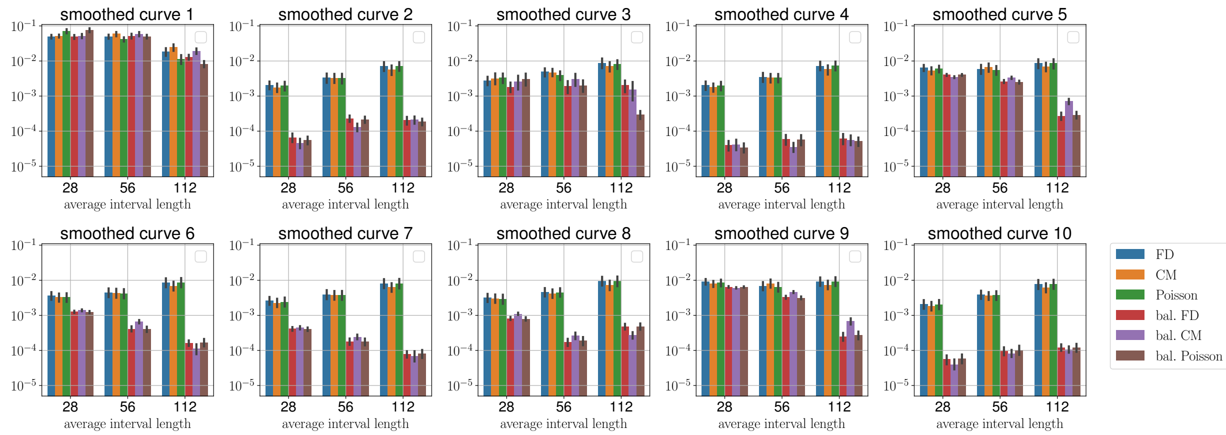


Figure 4 Estimation error of various switchback designs when cumulative effects follow one of the top 10 smoothed cumulative effect curves and when one experiment runs simultaneously with the main synthetic experiment.

We conduct two types of synthetic experiments. In the first type, no other experiments run simultaneously with the main synthetic experiment. Here, we randomly draw a region from the 50 regions in the historical control data, and then randomly draw a consecutive two weeks of historical control data for this region. In the second type, one experiment runs simultaneously with the main synthetic experiment. Here, we randomly select one experiment-region pair and use its two-week historical experimental data. In both types, we assume that the main synthetic intervention has not been applied to the historical (control or experimental) data.

We then use the sampled historical data along with each of the six switchback designs to generate synthetic experimental data. This process requires the specification of the cumulative effects of the treatment. For a practically relevant specification, we use the smoothed curves of cumulative effects in Figure 3. Specifically, we sample one curve, which determines the cumulative effects for each treatment duration. Using this curve and a chosen switchback design, we calculate the cumulative effect at every time period. Finally, we add the cumulative effect to the sampled historical data to obtain the synthetic experimental data.

Next, we apply the HT estimator to synthetic experimental data to estimate GATE and compute the estimation error of GATE using the sampled curve. To obtain the MSE of GATE, we sample 100 distinct two-week-region pairs of historical data. For each curve and each switchback design, we use these 100 pairs to generate 100 synthetic experimental data sets, which are then used to calculate the MSE of GATE.

3.2.2. Results Figure 4 shows the MSE of the estimated GATE for the six switchback designs when one experiment runs simultaneously. The cumulative effects follow one of the ten smoothed curves shown in Figure 3. Figure 5 uses a hierarchical structure to summarize the effectiveness of various design principles in reducing the MSE from all our simulation results.

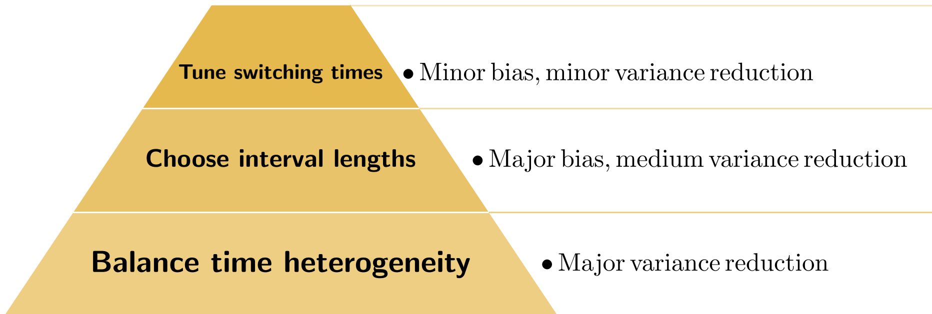


Figure 5 Hierarchical structure in the effectiveness of design principles in reducing MSE in the case study on a ride-sharing platform.

As shown in Figure 4, balancing time heterogeneity, achieved by using balanced switchback designs in Example 2.6, is the most effective in reducing the MSE. This efficacy stems from the inherent heterogeneity in weekly time patterns within the data. Balancing time heterogeneity operates to cancel out the variance term associated with the mean control outcome in the MSE, a point that will be formally explained in Section 4 below. When the treatment only marginally affects the mean outcome, this balancing act is particularly effective for variance reduction.

Carefully choosing the interval lengths can also effectively reduce the MSE. When switchback designs are not balanced, shortening interval lengths helps to reduce the MSE, because fast switching can reduce variance, a point that will be formally shown in Section 4 below. When switchback designs are balanced, lengthening interval lengths helps to reduce the MSE, because bias is a major concern once designs are balanced, and less frequent switching can reduce bias, a point that will also be shown in Section 4 below.

After the interval lengths are selected, tuning the switching times can further reduce the MSE. A heuristic tuning approach is to choose between fixed duration, change-of-measure, and Poisson duration. Depending on different scenarios of cumulative effects, a judicious choice among these three designs can reduce the MSE. Averaging the MSE across different scenarios, Figure 13 in Appendix A shows that the balanced Poisson switchback with long interval lengths has the lowest average error. A key factor contributing to this is that the randomization of switching times in the Poisson switchback reduces the confounding effect from simultaneous experiments. This factor is supported by comparing results with and without a simultaneous experiment in Figure 13 – in the absence of a simultaneous experiment, the improvement of the balanced Poisson design is not as apparent as in the case with one simultaneous experiment. This point will be rigorously shown through theoretical analysis in Section 4 below.

We further conduct a robustness check and run synthetic experiments using both the smoothed linear and cubic curves of cumulative effects. As shown in Figures 14 and 16 in Appendix A, the results are robust to the polynomial degrees of smoothed cumulative effect curve. However, we note that the MSE is generally larger when the smoothed curve has a higher polynomial degree. This observation is due to the increased bias when fitting a flexible curve of cumulative effects that allows for more variation with treatment duration.

4. Analysis of Switchback Design

In this section, we provide a precise decomposition of the bias and MSE for any switchback design. The decomposition provides insights into how effects from treatments at earlier times, correlations in event outcomes, and effects of simultaneous treatments affect the MSE. These insights explain the mechanisms through which different designs affect the MSE.

We first lay out the assumptions required for the identification of treatment effects and for the decomposition in Section 4.1. Subsequently, we introduce several interval-level statistics in Section 4.2. Finally, we present the decomposition in terms of these interval-level statistics in Section 4.3.

4.1. Assumptions

We begin by assuming that the sampling of events is exogenous and independent of the treatment assignments of both the main and simultaneous interventions. This assumption holds true for interventions that potential riders cannot discern a difference before opening the app and checking prices, and consequently do not affect the event density, such as surge pricing algorithms or matching algorithms.

Assumption 4.1 **(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, $\boldsymbol{W}, \boldsymbol{W}_1^{\mathrm{s}}, \cdots, \boldsymbol{W}_K^{\mathrm{s}}$.*

Moreover, we impose an assumption on the structure of the treatment effect for exposition. We first decompose the total treatment effect into the sum of the instantaneous treatment effect and the carryover effect:

$$\delta_t^{\mathrm{gate}}(\boldsymbol{w}_t) = w_t \cdot \underbrace{(Y_t(e_t, \mathbf{0}_t, \cdots, \mathbf{0}_t) - Y_t(\mathbf{0}_t, \mathbf{0}_t, \cdots, \mathbf{0}_t))}_{\delta_t^{\mathrm{inst}}} + \underbrace{Y_t(\boldsymbol{w}_t, \mathbf{0}_t, \cdots, \mathbf{0}_t) - Y_t(e_t, \mathbf{0}_t, \cdots, \mathbf{0}_t)}_{\delta_t^{\mathrm{co}}(\boldsymbol{w}_t)},$$

where the δ_t^{inst} is the *instantaneous effect* of treatment at time t on the outcome at time t, and $\delta_t^{\mathrm{co}}(\boldsymbol{w}_t)$ is the *carryover effect* of treatment assignments at earlier times on the outcome at time t. For notation simplicity, we let $\delta_t^{\mathrm{co}} := \delta_t^{\mathrm{co}}(\mathbf{1}_t)$ be the carryover effect under global treatment. We then assume carryover effects from the treatments at other times are additive and can be parameterized by a carryover kernel. Although this assumption can be relaxed, it comes at the expense of cumbersome notations in the main results, though the insights are generally the same.

Assumption 4.2 **(Carryover effects)**. *For every t, there exists a carryover kernel $d_t^{\mathrm{co}}(t')$ that measures the intensity of carryover effect from t' to t and satisfies $\int d_t^{\mathrm{co}}(t')f(t')dt' = 1$, such that*

$$\delta_t^{\mathrm{co}}(\boldsymbol{w}_t) = \delta_t^{\mathrm{co}} \cdot \int w_{t'} \cdot d_t^{\mathrm{co}}(t') f(t') dt' .$$

The carryover kernel $d_t^{\mathrm{co}}(t')$ can be quite general. Below, we provide two examples of carryover kernels commonly used in practice. In the case of the non-anticipating outcome, $d_t(t') = 0$ for all $t' > t$. Moreover, if the treatment assignments can only affect the outcomes for a duration of $h < \infty$ in the future, then $d_t(t') = 0$ for all $t' < t + h$, corresponding to the assumption of fixed duration carryover effect imposed in the literature (Bojinov et al. (2023) among others).

Example 4.1 (Uniform carryover kernel). If the carryover intensity is uniform in $t' \in [t-h, t]$, but is zero outside this interval, then $d_t^{\mathrm{co}}(t') \propto 1/h$ for all $t' \in [t-h, t]$ and $d_t^{\mathrm{co}}(t') = 0$ for all $t' \notin [t-h, t]$.

Example 4.2 (Linear decay carryover kernel). If the carryover intensity decays linearly in $t - t'$ for $t' \in [t-h, t]$, and is zero outside this interval, then $d_t^{\mathrm{co}}(t') \propto t - t'$ for all $t' \in [t-h, t]$ and $d_t^{\mathrm{co}}(t') = 0$ for all $t' \notin [t-h, t]$. See Examples in Figure 6.

Below we introduce an additive condition for the effects of main and simultaneous interventions. This represents a special case of the confounding between the main and simultaneous interventions. Under this condition, the decomposition of MSE is substantially simplified, making it more interpretable.

Condition 1 (Additivity of Intervention Effects). The effects of main and simultaneous interventions are additive, i.e.,

$$Y_t(\boldsymbol{w}_t', \boldsymbol{w}_{1,t}^{\mathrm{s}}, \cdots, \boldsymbol{w}_{K,t}^{\mathrm{s}}) - Y_t(\boldsymbol{w}_t, \boldsymbol{w}_{1,t}^{\mathrm{s}}, \cdots, \boldsymbol{w}_{K,t}^{\mathrm{s}}) = Y_t(\boldsymbol{w}_t', \boldsymbol{w}_{1,t}^{\mathrm{s}\prime}, \cdots, \boldsymbol{w}_{K,t}^{\mathrm{s}\prime}) - Y_t(\boldsymbol{w}_t, \boldsymbol{w}_{1,t}^{\mathrm{s}\prime}, \cdots, \boldsymbol{w}_{K,t}^{\mathrm{s}\prime}),$$

where $\boldsymbol{w}'$ and $\boldsymbol{w}$ are two treatment assignments of the main intervention, and $\boldsymbol{w}_\ell^{\mathrm{s}}$ and $\boldsymbol{w}_\ell^{\mathrm{s}\prime}$ are two treatment assignments of simultaneous intervention ℓ for $\ell = 1, \cdots, K$.

When $K = 1$, Condition 1 always holds. For $K > 1$, this condition excludes intervention effects from being synergistic (combining two interventions leads to a larger effect than expected) or antagonistic (combining two interventions leads to a smaller effect than expected). Condition 1 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.

4.2. Interval-Level Statistics

We introduce several interval-level statistics that quantify carryover effects, correlations in measurement errors, confounding effects from simultaneous interventions, and other components at the interval level. These interval-level statistics serve as building blocks of the bias and MSE decomposition in Section 4.3, and are crucial for interval partitioning in the design.

Fraction of events. Let

$$\mu^{(m)} = \int_{t \in \mathcal{I}_m} f(t)dt$$

represent the fraction of events occurring in the interval $\mathcal{I}_m$. $\mu^{(m)}$ ranges from 0 to 1 and the sum of $\mu^{(m)}$ over m equals to 1.

EXAMPLE 4.3. If event density $f(t)$ is uniform in t, then $\mu^{(m)} = |\mathcal{I}_m|/T$. Moreover, if the fixed duration switchback is used, then $\mu^{(m)} = 1/M$.

Mean control outcome. Let

$$\mu^{(m)}_{Y^{\text{ctrl}}} = \int_{t \in \mathcal{I}_m} Y_t(\mathbf{0}, \cdots, \mathbf{0}) f(t)dt$$

be the integrated global control outcome, $Y_t(\mathbf{0}, \cdots, \mathbf{0})$, over times t in the interval $\mathcal{I}_m$.

Variance and covariance of measurement errors. Let the integrated variance of the measurement error for events in the interval $\mathcal{I}_m$ be

$$V^{(m)} = \int_{t_i \in \mathcal{I}_m} \mathbb{E}_\varepsilon \left[(\varepsilon^{(i)})^2 \mid t_i \right] f(t_i) dt_i \,,$$

where $\mathbb{E}_\varepsilon \left[(\varepsilon^{(i)})^2 \mid t_i \right]$ represents the variance of measurement error for event i occurring at time t_i (as measurement error has mean zero).

EXAMPLE 4.4. Suppose measurement errors are homoscedastic, that is, $\mathbb{E}_\varepsilon \left[(\varepsilon^{(i)})^2 \mid t_i \right] = \sigma^2$ for all t_i. If the event density $f(t)$ is uniform in t, then $V^{(m)} = \sigma^2 |I_m|/T$. Additionally, if the fixed duration switchback is used, then $V^{(m)} = \sigma^2/M$.

Next, let the integrated covariance between measurement errors of events in the interval $\mathcal{I}_{\ell m}$ be

$$C^{(m)} = \int_{t_i, t_j \in \mathcal{I}_m} \mathbb{E}_\varepsilon \left[\varepsilon^{(i)} \varepsilon^{(j)} \mid t_i, t_j \right] f(t_i) f(t_j) dt_i dt_j \,,$$

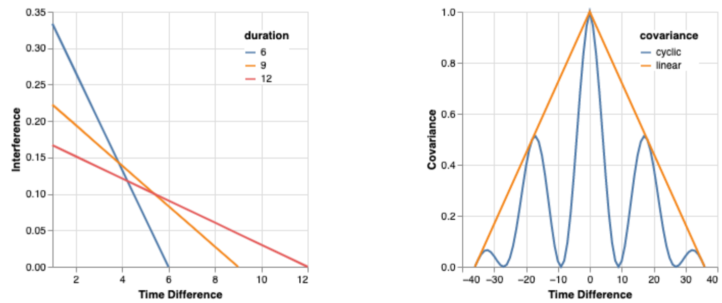
where $\mathbb{E}_\varepsilon \left[\varepsilon^{(i)} \varepsilon^{(j)} \mid t_i, t_j \right]$ is the covariance between the measurement errors of event i occurring at time t_i and event j occurred at time t_j.

In practical settings, patterns often exist in how the covariance $\mathbb{E}_\varepsilon \left[\varepsilon^{(i)} \varepsilon^{(j)} \mid t_i, t_j \right]$ varies with t_i and t_j, e.g., decays monotonically or periodically with the distance between t_i and t_j. Therefore, a kernel function can be used to parameterize and capture the patterns in $\mathbb{E}_\varepsilon \left[\varepsilon^{(i)} \varepsilon^{(j)} \mid t_i, t_j \right]$. See two examples in Figure 6.

Integrated total treatment effects. Let

$$\Xi^{(m)} = \int_{t \in \mathcal{I}_m} \delta^{\text{gate}}_t f(t)dt$$

be the integrated total treatment effect δ^{gate}_t over times t in the interval $\mathcal{I}_m$. Following the definition of δ^{gate}, the sum of $\Xi^{(m)}$ over m equals δ^{gate}. Moreover, if δ^{gate}_t is constant in t, then $\Xi^{(m)} = \delta^{\text{gate}} \mu^{(m)}$ for any m.



(a) Carryover kernel (b) Covariance kernel

Figure 6 Illustration of carryover and covariance kernels. Time difference denotes $t' - t$ in the carryover kernel $d_t^{\text{co}}(t')$. If $t' - t < 0$, then $d_t^{\text{co}}(t') = 0$. The interpretation of time difference is analogous to the covariance kernel.

Integrated carryover effects. Let

$$I^{(m,k)} = \int_{t \in \mathcal{I}_m} \left[\delta_t^{\text{co}} \int_{t' \in \mathcal{I}_k} d_t^{\text{co}}(t') f(t') dt' \right] f(t) dt$$

be integrated carryover effect of treatments at times in the interval $\mathcal{I}_k$ on outcomes at times in the interval $\mathcal{I}_m$. For simplicity in notation, we let $I^{(m)} = I^{(m,m)}$ be the integrated carryover effect of treatments on outcomes in the same interval. The integrated carryover effect $I^{(m,k)}$ increases with the length of both $\mathcal{I}_m$ and $\mathcal{I}_k$, and increases with the size of carryover effect δ_t^{co} for $t \in \mathcal{I}_m$. The sum of $I^{(m,k)}$ over both m and k, which is the integrated carryover effect of the treatment of all intervals on the outcomes of all intervals, is equal to δ^{co}, the average of δ_t^{co} over t. Moreover, if the carryover effect δ_t^{co} is constant in t, then the sum of $I^{(m,k)}$ over k, which is the integrated carryover effect of the treatment of all intervals on the outcomes in the interval $\mathcal{I}_m$, is equal to $\delta^{\text{co}} \mu^{(m)}$. Therefore, we can view $I^{(m,k)}$ as the "building blocks" of δ^{co}.

Confounding effects from simultaneous interventions. For any time t, let

$$\delta_t^{\text{simul}}(\boldsymbol{W}_t) = \mathbb{E}_{\boldsymbol{W}_{1,t}^{\text{s}}, \cdots, \boldsymbol{W}_{K,t}^{\text{s}}} \left[Y_t(\boldsymbol{W}_t, \boldsymbol{W}_{1,t}^{\text{s}}, \cdots, \boldsymbol{W}_{K,t}^{\text{s}}) - Y_t(\boldsymbol{W}_t, \boldsymbol{0}_t, \cdots, \boldsymbol{0}_t) \mid \boldsymbol{W}_t, t \right]$$

be the expected treatment effects from the simultaneous interventions at time t, conditional on $\boldsymbol{W}_t$. Here the expectation is taken with respect to the distribution from which the treatment designs of simultaneous interventions are drawn. If the simultaneous interventions have nonzero treatment effects, then $\delta_t^{\text{simul}}(\boldsymbol{W}_t)$ is generally nonzero, which can then bias the HT estimator.

We introduce a quantity below that measures the integrated bias from the treatment effects of simultaneous interventions

$$S^{(m)} = \int_{t \in \mathcal{I}_m} \Phi_t^{\text{simul}} f(t) dt \,,$$

where Φ_t^{simul} is defined as

$$\Phi_t^{\text{simul}} = \mathbb{E}_{\boldsymbol{W}^{(-m)}} \left[\delta_t^{\text{simul}}(\boldsymbol{W}^{(-m)}, W^{(m)} = 1) \right] - \mathbb{E}_{\boldsymbol{W}^{(-m)}} \left[\delta_t^{\text{simul}}(\boldsymbol{W}^{(-m)}, W^{(m)} = 0) \right].$$

Here $W^{(m)}$ and $\boldsymbol{W}^{(-m)}$ are the treatment assignment of interval $\mathcal{I}_m$ and of all the intervals excluding $\mathcal{I}_m$, respectively. Analogously, let $W_\ell^{\mathrm{s}(m)}$ be the treatment assignment of the m-th interval of the ℓ-th simultaneous intervention. As treatments are assigned at the interval level, given $(W^{(m)}, \boldsymbol{W}^{(-m)})$, $\boldsymbol{W}_t$ is uniquely determined, so we can also write $\delta_t^{\mathrm{simul}}(\boldsymbol{W}_t)$ as $\delta_t^{\mathrm{simul}}(\boldsymbol{W}^{(-m)}, W^{(m)} = 1)$.

The quantity Φ_t^{simul} measures the imbalance of expected treatment effects from simultaneous interventions at time t between when the m-th interval is treated versus when this interval is not treated. Here the expectation is taken with respect to the distribution from which $\boldsymbol{W}^{(-m)}$ is drawn. If the imbalance Φ_t^{simul} is larger, then the integrated bias from simultaneous interventions $S^{(m)}$ is larger.

Note that Φ_t^{simul} is zero in some special cases, such as when the effects of main and simultaneous interventions are additive (Condition 1 holds), which is illustrated in the following example.

Example 4.5 (Additive effects). When Condition 1 holds, we have

$$
\begin{aligned}
&Y_t((\boldsymbol{W}^{(-m)}, W^{(m)} = 1), \boldsymbol{W}_{1,t}^{\mathrm{s}}, \cdots, \boldsymbol{W}_{K,t}^{\mathrm{s}}) - Y_t((\boldsymbol{W}^{(-m)}, W^{(m)} = 0), \boldsymbol{W}_{1,t}^{\mathrm{s}}, \cdots, \boldsymbol{W}_{K,t}^{\mathrm{s}}) \\
=&Y_t((\boldsymbol{W}^{(-m)}, W^{(m)} = 1), \mathbf{0}_t, \cdots, \mathbf{0}_t) - Y_t((\boldsymbol{W}^{(-m)}, W^{(m)} = 0), \mathbf{0}_t, \cdots, \mathbf{0}_t)\,.
\end{aligned}
$$

We subtract both sides by the term on the right-hand side and take the expectation over $\boldsymbol{W}_{1,t}^{\mathrm{s}}, \cdots, \boldsymbol{W}_{K,t}^{\mathrm{s}}$. We then obtain

$$
\delta_t^{\mathrm{simul}}(\boldsymbol{W}^{(-m)}, W^{(m)} = 1) - \delta_t^{\mathrm{simul}}(\boldsymbol{W}^{(-m)}, W^{(m)} = 0) = 0
$$

which implies that Φ_t^{simul} is zero.

However, when the effects of main and simultaneous interventions are not additive (e.g., two interventions are useful only when both are present), the HT estimator is generally biased and we quantify the bias in the subsection below.

4.3. Main Results

We provide the decomposition of the bias and MSE of $\hat{\delta}^{\mathrm{gate}}$ from the HT estimator in terms of the interval-level statistics in Theorems 4.1 and 4.2 below. The decomposition shows how different components in the outcome affect the estimation error of $\hat{\delta}^{\mathrm{gate}}$.

Theorem 4.1 **(Estimation bias)**. *Suppose Assumptions 4.1-4.2 hold, $W^{(m)}$ is independent of m with $\mathbf{P}(W^{(m)} = 1) = 1/2$, and $W_\ell^{\mathrm{s}(m)}$ is independent of m with $\mathbf{P}(W_\ell^{\mathrm{s}(m)} = 1) = 1/2$ for ℓ. The estimation bias of $\hat{\delta}^{\mathrm{gate}}$ is*

$$
\mathbb{E}_{W,\varepsilon,t}\left[\hat{\delta}^{\mathrm{gate}} - \delta^{\mathrm{gate}}\right] = \mathrm{Bias}(\mathcal{E}_{\mathrm{carryover}}) + \mathrm{Bias}(\mathcal{E}_{\mathrm{simul}})\,,
$$

where

$$\mathrm{Bias}(\mathcal{E}_{\mathrm{carryover}}) = \sum_{m=1}^{M} I^{(m)} - \delta^{\mathrm{co}}$$
$$\mathrm{Bias}(\mathcal{E}_{\mathrm{simul}}) = \sum_{m=1}^{M} S^{(m)}\,.$$

Theorem 4.1 shows there are two sources of bias. The first source of bias comes from the carryover effects and is measured by $\mathrm{Bias}(\mathcal{E}_{\mathrm{carryover}})$. If carryover effects are zero, (i.e., $\delta_t^{\mathrm{co}} = 0$ for all t), then $\mathrm{Bias}(\mathcal{E}_{\mathrm{carryover}}) = 0$. If carryover effects are nonzero, then $\mathrm{Bias}(\mathcal{E}_{\mathrm{carryover}})$ quantifies the bias in the HT estimator that arises from using direct treated and control outcomes to approximate globally treated and control outcomes, respectively.

The second source of bias comes from the confounding effects of simultaneous interventions and is measured by $\mathrm{Bias}(\mathcal{E}_{\mathrm{simul}})$. In cases where the effects of main and simultaneous interventions are additive, as illustrated in Example 4.5, $S^{(m)}$ is zero, resulting in $\mathrm{Bias}(\mathcal{E}_{\mathrm{simul}})$ being zero as well. However, when effects are not additive, $\mathrm{Bias}(\mathcal{E}_{\mathrm{simul}})$ is generally nonzero and tends to increase with the number of simultaneous experiments K and the size of treatment effects from simultaneous interventions.

Both sources of bias can be reduced by properly choosing a switchback design. To mitigate the bias from carryover effects, it is helpful to switch less frequently. This is evident in the scenario of uniform event density and fixed-duration switchback. In this case, $I^{(m)} = \delta^{\mathrm{co}}(1/M - h/(2T))$ as shown in Example A.2, resulting in the carryover bias equaling $|\mathrm{Bias}(\mathcal{E}_{\mathrm{carryover}})| = \delta^{\mathrm{co}} Mh/(2T)$, which increases with the number of intervals M. To reduce the bias from simultaneous experiments, randomizing the switching times, such as using Poisson duration switchback, is helpful. See Figure 9c below for an illustration, and Example A.3 in Appendix A.4 for a toy numerical example.

Next, we show the decomposition of the MSE of $\hat{\delta}^{\mathrm{gate}}$.

Theorem 4.2 **(Mean-Squared Error)**. *Suppose Assumptions 4.1-4.2 hold, $W^{(m)}$ is independent of m with $\mathbf{P}(W^{(m)} = 1) = 1/2$, and $W_\ell^{\mathrm{s}(m)}$ is independent of m with $\mathbf{P}(W_\ell^{\mathrm{s}(m)} = 1) = 1/2$ for ℓ. The mean-squared error of $\hat{\delta}^{\mathrm{gate}}$ is*

$$\mathbb{E}_{W,\varepsilon,t}\left[\left(\hat{\delta}^{\mathrm{gate}} - \delta^{\mathrm{gate}}\right)^2\right] = \mathrm{Var}(\mathcal{E}_{\mathrm{meas}}) + \mathrm{Bias}(\mathcal{E}_{\mathrm{carryover}})^2 + \mathrm{Var}(\mathcal{E}_{\mathrm{inst}} + \mathcal{E}_{\mathrm{carryover}}) + \mathbb{E}[\mathcal{E}_{\mathrm{simul}}^2] + 2\mathbb{E}[(\mathcal{E}_{\mathrm{inst}} + \mathcal{E}_{\mathrm{carryover}}) \cdot \mathcal{E}_{\mathrm{simul}}]\,,$$

where

$$\mathrm{Var}(\mathcal{E}_{\mathrm{meas}}) = 4\sum_{m=1}^{M} \left(V^{(m)}/n + C^{(m)} \cdot (n-1)/n\right)$$
$$\mathrm{Var}(\mathcal{E}_{\mathrm{inst}} + \mathcal{E}_{\mathrm{carryover}}) = \sum_{m=1}^{M} \left(\Xi^{(m)} + 2\mu_{Y^{\mathrm{ctrl}}}^{(m)}\right)^2 + \sum_{m=1}^{M} \sum_{m' \neq m} \left(\left[I^{(m,m')}\right]^2 + I^{(m,m')} I^{(m',m)}\right)$$

and

$$\mathbb{E}[\mathcal{E}_{\text{simul}}^2] = \sum_{m=1}^{M}\sum_{m'=1}^{M} S_{\text{var}}^{(m,m')}$$

$$\mathbb{E}[(\mathcal{E}_{\text{inst}} + \mathcal{E}_{\text{carryover}}) \cdot \mathcal{E}_{\text{simul}}] = \sum_{m=1}^{M}\sum_{m'=1}^{M} S_{\text{cov}}^{(m,m')}$$

with $S_{\text{var}}^{(m,m')}$ *and* $S_{\text{cov}}^{(m,m')}$ *defined in Equations* (A.1) *and* (A.2)*, respectively, in Appendix A.3.*

Theorem 4.2 demonstrates that, in addition to the bias terms, the MSE is affected by three sources of variance. The first source of variance arises from the randomness in event measurement errors and is quantified by $\text{Var}(\mathcal{E}_{\text{meas}})$. It is worth noting that $\text{Var}(\mathcal{E}_{\text{meas}})$ consists of two parts: the first part $V^{(m)}$, which measures the variance of event measurement error, and the second part $C^{(m)}$, which measures the covariance between measurement errors of two events. As the number of events grows, the first part diminishes and the second part dominates. If the correlation in measurement errors is persistent, then $C^{(m)}$ is larger, leading to a larger MSE.

The second source of variance stems from the randomness in treatment assignments of the main intervention and is measured by $\text{Var}(\mathcal{E}_{\text{inst}} + \mathcal{E}_{\text{carryover}})$. The expression of $\text{Var}(\mathcal{E}_{\text{inst}} + \mathcal{E}_{\text{carryover}})$ shows that: (1) it increases with the size of the instantaneous effect, as the term $\Xi^{(m)}$ increases with the instantaneous effect; (2) it increases with the size of the carryover effect, as both $\Xi^{(m)}$ and $I^{(m,m')}$ increase with the carryover effect; and (3) it increases with the scale of the mean outcome, as the term $\mu_{Y^{\text{ctrl}}}^{(m)}$ increases with the mean outcome.

The third source of variance arises from the randomness in treatment assignments of simultaneous interventions and affects $\mathbb{E}[\mathcal{E}_{\text{simul}}^2]$. In Proposition 4.1 below, we present the expression for $\mathbb{E}[\mathcal{E}_{\text{simul}}^2]$ under the condition of additive main and simultaneous effects. This expression shows that this source of variance increases with the magnitude of instantaneous and carryover effects of simultaneous interventions. Furthermore, this term increases with the overlap between intervals of the main intervention and simultaneous interventions.

In addition to the bias and variance terms, the MSE includes a cross term $\mathbb{E}[(\mathcal{E}_{\text{inst}} + \mathcal{E}_{\text{carryover}}) \cdot \mathcal{E}_{\text{simul}}]$. Proposition 4.1 below also provides the expression for this term under the condition of additive main and simultaneous effects. Generally, if the variance terms from the main intervention and simultaneous interventions are larger, then this cross term is larger.

PROPOSITION 4.1. *Under the assumptions in Theorem 4.2, if Condition 1 holds, then the bias from simultaneous interventions* $\text{Bias}(\mathcal{E}_{\text{simul}})$ *is zero and the variance from simultaneous interventions is*

$$\mathbb{E}[\mathcal{E}_{\text{simul}}^2] = \sum_{m=1}^{M} \left(\int_{t\in\mathcal{I}_m} \left[\sum_{\ell=1}^{K} \delta_{\ell,t}^{\text{s.gate}} \right] f(t)dt \right)^2$$

$$+\sum_{m=1}^{M}\sum_{m'=1}^{M}\sum_{\ell=1}^{K}\left(\int_{t\in\mathcal{I}_m\cap\mathcal{I}^{\mathrm{s}}_{\ell m'}}\delta^{\mathrm{s.inst}}_{\ell,t}f(t)dt+\int_{t\in\mathcal{I}_m,t'\in\mathcal{I}^{\mathrm{s}}_{\ell m'}}\delta^{\mathrm{s.co}}_{\ell,t}d^{\mathrm{s.co}}_{\ell,t}(t')f(t)f(t')dtdt'\right)^2,$$

and

$$\mathbb{E}[(\mathcal{E}_{\mathrm{inst}}+\mathcal{E}_{\mathrm{carryover}})\cdot\mathcal{E}_{\mathrm{simul}}]=\sum_{m=1}^{M}\left(\Xi^{(m)}+2\mu^{(m)}_{Y^{\mathrm{ctrl}}}\right)\left(\int_{t\in\mathcal{I}_m}\left[\sum_{\ell=1}^{K}\delta^{\mathrm{s.gate}}_{\ell,t}\right]f(t)dt\right),$$

where $\delta^{\mathrm{s.gate}}_{\ell,t}$, $\delta^{\mathrm{s.inst}}_{\ell,t}$, $\delta^{\mathrm{s.co}}_{\ell,t}$, $d^{\mathrm{s.co}}_{\ell,t}$ *are the total treatment effect, instantaneous effect, carryover effect, and carryover kernel of simultaneous intervention* ℓ *at time* t*, respectively, and* $\mathcal{I}^{\mathrm{s}}_{\ell m}$ *is the* m*-th interval of simultaneous intervention* ℓ*.*

The MSE decomposition provides insights into how to design switchback experiments to reduce the MSE of GATE. First and foremost, the switching interval length has a mixed effect in reducing the MSE. Switching frequently reduces most variance and cross terms. The intuition is that switching more frequently increases the number of "interval-level" observations, thereby increasing the effective sample size.

To illustrate this more clearly, consider the setting of uniform event density and fixed-duration switchback. In this context, $\mathrm{Var}(\mathcal{E}_{\mathrm{meas}})=O(1/M)$ decreases with M, following that $C^{(m)}=O(1/M^2)$ as shown in Example A.1. Moreover, the term $\sum_{m=1}^{M}\left(\Xi^{(m)}+2\mu^{(m)}_{Y^{\mathrm{ctrl}}}\right)^2=O(1/M)$ also decreases with M, given that $\Xi^{(m)}=O(1/M)$ and $\mu^{(m)}_{Y^{\mathrm{ctrl}}}=O(1/M)$. Analogously, for the variance from simultaneous interventions, both $\mathbb{E}[\mathcal{E}^2_{\mathrm{simul}}]$ and $\mathbb{E}[(\mathcal{E}_{\mathrm{inst}}+\mathcal{E}_{\mathrm{carryover}})\cdot\mathrm{simul}]$ are $O(1/M)$ under the additivity condition.

On the other hand, switching frequently increases the carryover bias and some of the other variance terms. This is because the carryover effects across intervals increase with switching frequency. Therefore, with the tradeoff involved, the optimal value of M depends on the relative size of instantaneous and carryover effects, the scale of the global control outcomes, and the duration and mechanism of carryover effects.

Second, the choice of interval endpoints also matters for the MSE. To see this more clearly, consider the setting of constant δ^{gate}_t in t and constant $Y_t(\mathbf{0},\cdots,\mathbf{0})$ in t (equals to $\bar{Y}^{\mathrm{ctrl}}$). Then $\sum_{m=1}^{M}\left(\Xi^{(m)}\right)^2=(\delta^{\mathrm{gate}}+2\bar{Y}^{\mathrm{ctrl}})^2\sum_{m=1}^{M}\left(\mu^{(m)}\right)^2$, which is minimized at $\mu^{(m)}=1/M$ for all m, i.e., equalizing the fraction of events in each interval. This implies that when event density $f(t)$ varies with t, this term can be reduced by switching more frequently in times of high event density and less frequently in times of low event density.

In addition, the second term of $\mathbb{E}[\mathcal{E}^2_{\mathrm{simul}}]$ is affected by how much the intervals of the main intervention overlap with the intervals of simultaneous interventions, and it is the largest when the interval endpoints of the main intervention and simultaneous interventions are the same. This implies that it is useful to stagger the switching times of different interventions to reduce $\mathbb{E}[\mathcal{E}^2_{\mathrm{simul}}]$.

Last but not least, balancing time heterogeneity can reduce the MSE. Specifically, consider a two-week experiment where potential outcomes in the second week are the same as those in the first week, and the balanced design in Example 2.6 is used. Then under mild assumptions, the mean control outcome $\mu^{(m)}_{Y^{\mathrm{ctrl}}}$ can be canceled out in both the variance term $\mathrm{Var}(\mathcal{E}_{\mathrm{inst}} + \mathcal{E}_{\mathrm{carryover}})$ and the cross term $\mathbb{E}[(\mathcal{E}_{\mathrm{inst}} + \mathcal{E}_{\mathrm{carryover}}) \cdot \mathcal{E}_{\mathrm{simul}}]$. The variance term is then equal to

$$\mathrm{Var}(\mathcal{E}_{\mathrm{inst}} + \mathcal{E}_{\mathrm{carryover}}) = \sum_{m=1}^{M} \left(\Xi^{(m)}\right)^2 + \sum_{m=1}^{M} \sum_{m' \neq m} \left(\left[I^{(m,m')} \right]^2 + I^{(m,m')} I^{(m',m)} \right).$$

Furthermore, under Condition 1, the cross term is equal to

$$\mathbb{E}[(\mathcal{E}_{\mathrm{inst}} + \mathcal{E}_{\mathrm{carryover}}) \cdot \mathcal{E}_{\mathrm{simul}}] = \sum_{m=1}^{M} \Xi^{(m)} \left(\int_{t \in \mathcal{I}_m} \left[\sum_{\ell=1}^{K} \delta^{\mathrm{s.gate}}_{\ell,t} \right] f(t) dt \right).$$

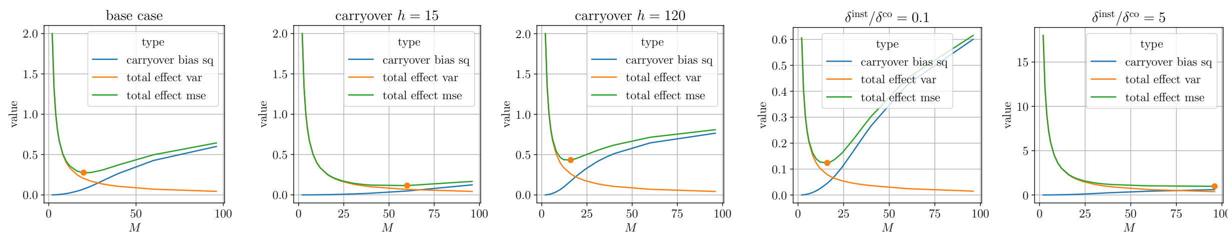
Canceling out $\mu^{(m)}_{Y^{\mathrm{ctrl}}}$ can substantially reduce the variance when the treatment effects are not more than a few percent of the control outcome. This is exactly the case in our case study on a ride-sharing platform. Therefore, Figure 4 shows that balancing time heterogeneity is particularly useful for error reduction. In fact, the HT estimator coincides with the Hajek estimator when the balanced designs are used, but this is not the case for unbalanced designs. The Hajek estimator is generally more efficient than the HT estimator, from another perspective, explaining why balanced designs are more efficient than unbalanced designs.

5. Simulation

In this section, we compute values for error components in the MSE, in the settings where error components are on a similar scale, resulting in an interesting tradeoff in designing a switchback experiment. We use fixed-duration designs in the base case to characterize the tradeoffs involved, and the general insights carry over to other heuristic designs. In the base case, a linear decay carryover kernel is used with a bandwidth of $h_{\mathrm{carryover}} = 60$ (i.e., the duration of carryover effects is 60 minutes). Moreover, a linear decay covariance kernel is used with a bandwidth of $h_{\mathrm{covariance}} = 60$ (i.e., two event outcomes are correlated only if they are within 60 minutes apart). Both the instantaneous and carryover effects are constant in time and equal to $\delta^{\mathrm{inst}}_t = 1$ and $\delta^{\mathrm{co}}_t = 1$ for all t. In the base case, only one intervention is tested, and the experiment lasts one day, i.e., $T = 1,440$ minutes. We vary the parameters in the base case and illustrate how error components vary and affect the performance of various switchback designs.

5.1. Instantaneous and Carryover Effects Only

Figure 7 illustrates the tradeoff between the bias from carryover effects, $\mathrm{Bias}(\mathcal{E}_{\mathrm{carryover}})$, and variance from instantaneous and carryover effects, $\mathrm{Var}(\mathcal{E}_{\mathrm{inst}} + \mathcal{E}_{\mathrm{carryover}})$. If the carryover effects last longer, i.e., $h_{\mathrm{carryover}}$ is larger, then the bias $\mathrm{Bias}(\mathcal{E}_{\mathrm{carryover}})$ tends to be larger, and switching less frequently reduces the MSE from instantaneous and carryover effects, as shown in the comparison between Figures 7a and 7b. The relative size between instantaneous and carryover effects also matters for the tradeoff. If the instantaneous effect is relatively larger than the carryover effect, then switching more frequently reduces the MSE, as shown in the comparison between Figures 7a and 7c.

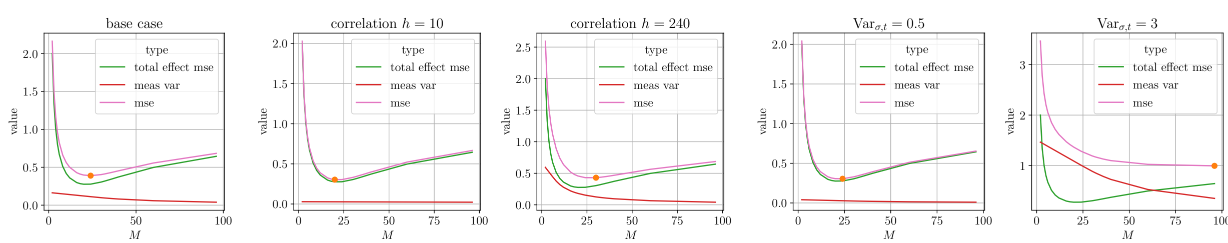


(a) base case (b) vary $h_{\text{carryover}}$ (c) vary $\delta^{\text{inst}}/\delta^{\text{co}}$

Figure 7 Tradeoff between error components from instantaneous and carryover effects. "carryover bias sq" denotes $\text{Bias}(\mathcal{E}_{\text{carryover}})^2$, "total effect var" denotes $\text{Var}(\mathcal{E}_{\text{inst}} + \mathcal{E}_{\text{carryover}})$, and "total effect mse" denotes the sum of the two. Figure 7b varies $h_{\text{carryover}}$ and Figure 7c varies $\delta^{\text{inst}}/\delta^{\text{co}}$, while holding other parameters at the base level. The orange dot denotes the minimum "total effect mse".

5.2. Total Treatment Effects with Measurement Errors

Figure 8 summarizes the tradeoffs between error components from total treatment effects and measurement errors. Switching frequently generates more comparisons, which reduces variance from measurement errors but also increases carryover bias from previous intervals. If correlation in measurement errors are persistent (i.e., large $h_{\text{covariance}}$), then $\text{Var}(\mathcal{E}_{\text{meas}})$ has a larger impact on the MSE, and switching more frequently is helpful to reduce the MSE, as shown in the comparison between Figures 8a and 8b. Similarly, if event outcomes are noisier with a larger $\text{Var}_{\sigma,t}$, then switching more frequently helps, as shown in the comparison between Figures 8a and 8c.

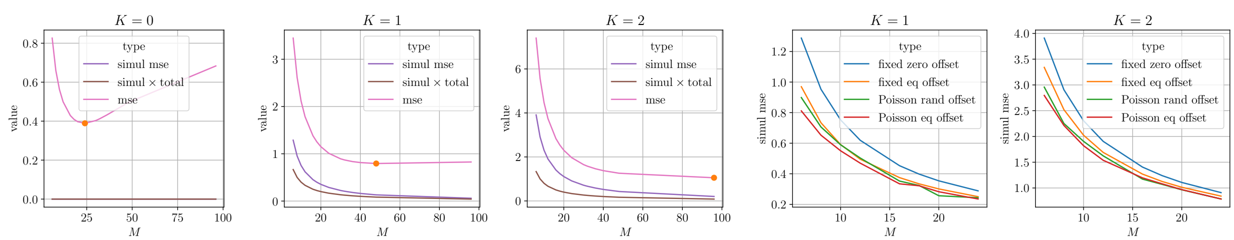


(a) base case (b) vary $h_{\text{covariance}}$ (c) vary $\text{Var}_{\sigma,t}$

Figure 8 Tradeoffs between error components from treatment effects and measurement errors. "total effect mse" denotes the sum of $\text{Bias}(\mathcal{E}_{\text{carryover}})^2$ and $\text{Var}(\mathcal{E}_{\text{inst}} + \mathcal{E}_{\text{carryover}})$, "meas var" denotes $\text{Var}(\mathcal{E}_{\text{meas}})$, and "mse" denotes the sum of "total effect mse" and "meas var". Figure 8b varies $h_{\text{covariance}}$ and Figure 8c varies $\text{Var}_{\sigma,t}$, while holding other parameters at the base level. The orange dot denotes the minimum "mse".

5.3. Simultaneous Interventions

Figure 9 shows the tradeoffs involved in the presence of simultaneous interventions. The MSE of the main intervention is affected by three factors: the number of simultaneous interventions, the interval duration, and the offset in switching times between simultaneous experiments. When more experiments are run simultaneously, the optimal switching frequency increases. This is because switching more frequently helps reduce the confounding effects of simultaneous interventions. Moreover, properly staggering the switching times of the main and simultaneous interventions also decreases the MSE of the main intervention, with the effect being more pronounced when the interval duration is longer due to the increased finite-sample correlation between the designs. As shown in Figure 9c, Poisson switchbacks, which implicitly stagger through randomizing switching times, can be more effective unless the fixed duration designs are staggered well.



(a) base case (b) vary K (c) compare designs (simul mse)

Figure 9 MSE with simultaneous interventions. "simul mse" denotes $\mathbb{E}[\mathcal{E}_{\text{simul}}^2]$, "simul×total" denotes $\mathbb{E}[(\mathcal{E}_{\text{inst}} + \mathcal{E}_{\text{carryover}}) \cdot \mathcal{E}_{\text{simul}}]$, and "mse" denotes $\mathbb{E}_{W,\varepsilon,t}[(\hat{\delta}^{\text{gate}} - \delta^{\text{gate}})^2]$. Figure 9b varies K, while holding other parameters at the base level and switching at the same times for all interventions. Figure 9c compares four designs: fixed duration switchbacks with offset $q = 0$ for all interventions (i.e., fixed zero offset), fixed duration switchbacks with offset $q = p \cdot j/(K+1)$ for the j-th simultaneous intervention (i.e., fixed eq offset), Poisson switchbacks with offset $q \sim \text{Poisson}(T/M)$ (i.e., Poisson rand offset), and Poisson switchbacks with offset $q = p \cdot j/(K+1)$ for the j-th simultaneous intervention (i.e., Poisson eq offset).

5.4. Periodic Event Density

In many realistic settings, the density of events will exhibit periodic patterns due to the seasonality of human behavior. Figure 10 shows results from a periodic density using a fixed duration switchback. When the design has a period that aligns with density, the offset parameter q determines how the alignment alters the bias and variance. Switching at times of high density yields a design with low variance from measurement errors $\text{Var}(\mathcal{E}_{\text{meas}})$. On the other hand, switching at times of low density reduces carryover bias from the preceding interval $\text{Bias}(\mathcal{E}_{\text{carryover}})$ and reduces MSE from treatment effects. Combining measurement errors and treatment effects, the optimal switching times are somewhere in between high and low density times. Therefore, 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.

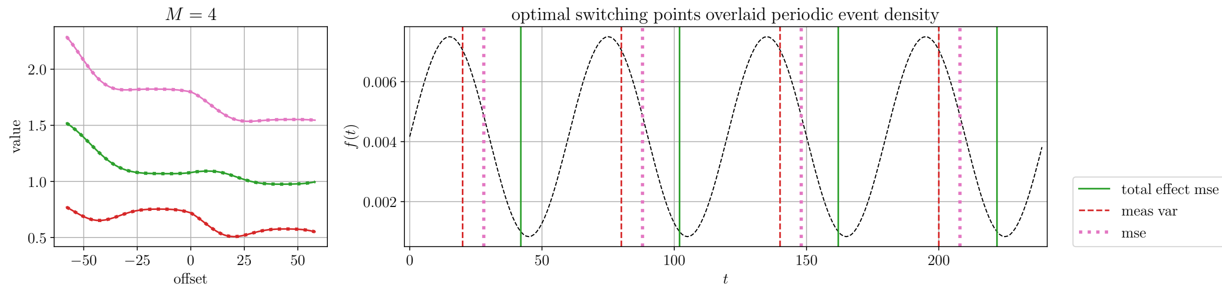


Figure 10 Role of offset parameter q in periodic event density. For the simplicity of exposition, $T = 240$ and $M = 4$, while other parameters are set at the base level.

6. Discussion and Conclusion

This paper studies the design and analysis of simultaneous switchback experiments. We provide a theoretical analysis of how the bias and variance of the Horvitz-Thompson estimator of the GATE are affected by four factors: carryovers from interventions at earlier times, nonuniform event density, correlations in event outcomes, and effects of interventions tested concurrently. Simulation and empirical studies show how these factors trade off each other and provide insights into how one can design efficient switchback experiments.

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 experimenters would make using prior knowledge. We illustrate this point with a case study on a ride-sharing platform, showcasing the value of leveraging prior experiments to make adequate assumptions when designing new experiments.

While we motivate this study by applications in the ride-sharing setting, the theory and practical guidelines presented can find broader applications in other contexts. Indeed, various settings exist where cross-sectional interventions are not possible or outcomes cannot be easily attributed to treatment decisions. Estimating the effectiveness of traditional media advertising aligns well with our problem setup, and a privacy-friendly approach to online advertising might employ temporal variation in campaign spending linked to sales through timestamps only. Prior work has explored 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 expand the use of temporal experiments to settings where they are not currently used.

References

Abadie, Alberto, Jinglong Zhao. 2021. Synthetic controls for experimental design. *arXiv preprint arXiv:2108.02196* .

Baird, Sarah, J Aislinn Bohren, Craig McIntosh, Berk Özler. 2018. Optimal design of experiments in the presence of interference. *Review of Economics and Statistics* **100**(5) 844–860.

Bajari, Patrick, Brian Burdick, Guido W Imbens, Lorenzo Masoero, James McQueen, Thomas S Richardson, Ido M Rosen. 2023. Experimental design in marketplaces. *Statistical Science* **1**(1) 1–19.

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

Basse, Guillaume W, Yi Ding, Panos Toulis. 2023. Minimax designs for causal effects in temporal experiments with treatment habituation. *Biometrika* **110**(1) 155–168.

Bojinov, Iavor, David Simchi-Levi, Jinglong Zhao. 2023. Design and analysis of switchback experiments. *Management Science* **69**(7) 3759–3777.

Boyarsky, Ariel, Hongseok Namkoong, Jean Pouget-Abadie. 2023. Modeling interference using experiment roll-out. *arXiv preprint arXiv:2305.10728* .

Candogan, Ozan, Chen Chen, Rad Niazadeh. 2021. Near-optimal experimental design for networks: Independent block randomization. *Available at SSRN* .

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

Chen, Hongyu, David Simchi-Levi. 2023. Switchback experiments in a reactive environment. *Available at SSRN 4436643* .

Chin, Alex. 2018. Central limit theorems via stein's method for randomized experiments under interference. *arXiv preprint arXiv:1804.03105* .

Chin, Alex. 2019. Regression adjustments for estimating the global treatment effect in experiments with interference. *Journal of Causal Inference* **7**(2).

Cortez, Mayleen, Matthew Eichhorn, Christina Yu. 2022. Staggered rollout designs enable causal inference under interference without network knowledge. *Advances in Neural Information Processing Systems* **35** 7437–7449.

Crépon, Bruno, Esther Duflo, Marc Gurgand, Roland Rathelot, Philippe Zamora. 2013. Do labor market policies have displacement effects? evidence from a clustered randomized experiment. *The quarterly journal of economics* **128**(2) 531–580.

Dasgupta, Tirthankar, Natesh S Pillai, Donald B Rubin. 2015. Causal inference from 2k factorial designs by using potential outcomes. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **77**(4) 727–753.

Doudchenko, Nick, David Gilinson, Sean Taylor, Nils Wernerfelt. 2019. Designing experiments with synthetic controls. Tech. rep., Working paper.

Doudchenko, Nick, Khashayar Khosravi, Jean Pouget-Abadie, Sebastien Lahaie, Miles Lubin, Vahab Mirrokni, Jann Spiess, et al. 2021. Synthetic design: An optimization approach to experimental design with synthetic controls. *Advances in Neural Information Processing Systems* **34**.

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

Farias, Vivek, Andrew Li, Tianyi Peng, Andrew Zheng. 2022. Markovian interference in experiments. *Advances in Neural Information Processing Systems* **35** 535–549.

Fisher, Ronald Aylmer. 1936. Design of experiments. *British Medical Journal* **1**(3923) 554.

Forastiere, Laura, Edoardo M Airoldi, Fabrizia Mealli. 2021. Identification and estimation of treatment and interference effects in observational studies on networks. *Journal of the American Statistical Association* **116**(534) 901–918.

Han, Kevin, Shuangning Li, Jialiang Mao, Han Wu. 2022. Detecting interference in a/b testing with increasing allocation. *arXiv preprint arXiv:2211.03262* .

Holtz, David, Felipe Lobel, Ruben Lobel, Inessa Liskovich, Sinan Aral. 2023. Reducing interference bias in online marketplace experiments using cluster randomization: Evidence from a pricing meta-experiment on airbnb. *Management Science* .

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

Hu, Yuchen, Stefan Wager. 2022. Switchback experiments under geometric mixing. *arXiv preprint arXiv:2209.00197* .

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

Johari, Ramesh, Hannah Li, Inessa Liskovich, Gabriel Y Weintraub. 2022. Experimental design in two-sided platforms: An analysis of bias. *Management Science* **68**(10) 7069–7089.

Karkar, Ravi, Jasmine Zia, Roger Vilardaga, Sonali R Mishra, James Fogarty, Sean A Munson, Julie A Kientz. 2016. A framework for self-experimentation in personalized health. *Journal of the American Medical Informatics Association* **23**(3) 440–448.

Krafft, Peter M, Nicolás Della Penna, Alex Sandy Pentland. 2018. An experimental study of cryptocurrency market dynamics. *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–13.

Leung, Michael P. 2022. Causal inference under approximate neighborhood interference. *Econometrica* **90**(1) 267–293.

Leung, Michael P. 2023. Network cluster-robust inference. *Econometrica* **91**(2) 641–667.

Li, Hannah, Geng Zhao, Ramesh Johari, Gabriel Y Weintraub. 2021. Interference, bias, and variance in two-sided marketplace experimentation: Guidance for platforms. *arXiv preprint arXiv:2104.12222* .

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

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

Ni, Tu, Iavor Bojinov, Jinglong Zhao. 2023. Design of panel experiments with spatial and temporal interference. *Available at SSRN 4466598* .

Qu, Zhaonan, Ruoxuan Xiong, Jizhou Liu, Guido Imbens. 2021. Efficient treatment effect estimation in observational studies under heterogeneous partial interference. *arXiv preprint arXiv:2107.12420* .

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

Ugander, Johan, Brian Karrer, Lars Backstrom, Jon Kleinberg. 2013. Graph cluster randomization: Network exposure to multiple universes. *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. 329–337.

Wager, Stefan, Kuang Xu. 2021. Experimenting in equilibrium. *Management Science* **67**(11) 6694–6715.

Wu, Yuhang, Zeyu Zheng, Guangyu Zhang, Zuohua Zhang, Chu Wang. 2022. Non-stationary a/b tests. *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2079–2089.

Xiong, Ruoxuan, Susan Athey, Mohsen Bayati, Guido W Imbens. 2023. Optimal experimental design for staggered rollouts. *Management Science* .

Ye, Zikun, Dennis J Zhang, Heng Zhang, Renyu Zhang, Xin Chen, Zhiwei Xu. 2023a. Cold start to improve market thickness on online advertising platforms: Data-driven algorithms and field experiments. *Management Science* **69**(7) 3838–3860.

Ye, Zikun, Zhiqi Zhang, Dennis Zhang, Heng Zhang, Renyu Philip Zhang. 2023b. Deep learning based causal inference for large-scale combinatorial experiments: Theory and empirical evidence. *Available at SSRN 4375327* .

Yuan, Yuan, Kristen Altenburger, Farshad Kooti. 2021. Causal network motifs: identifying heterogeneous spillover effects in a/b tests. *Proceedings of the Web Conference 2021*. 3359–3370.

Appendix A: Supplementary Results

A.1. Supplementary Empirical Results

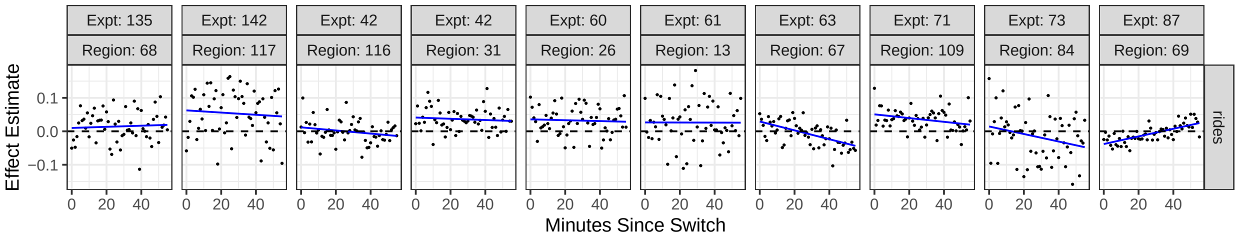


Figure 11 Fitted impulse response function of polynomial **degree 1**

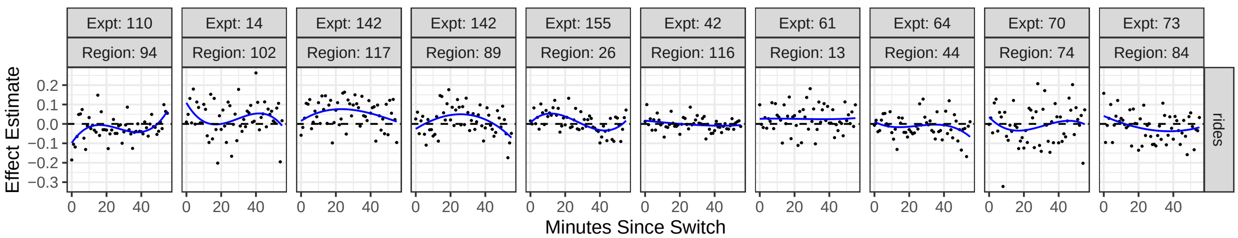
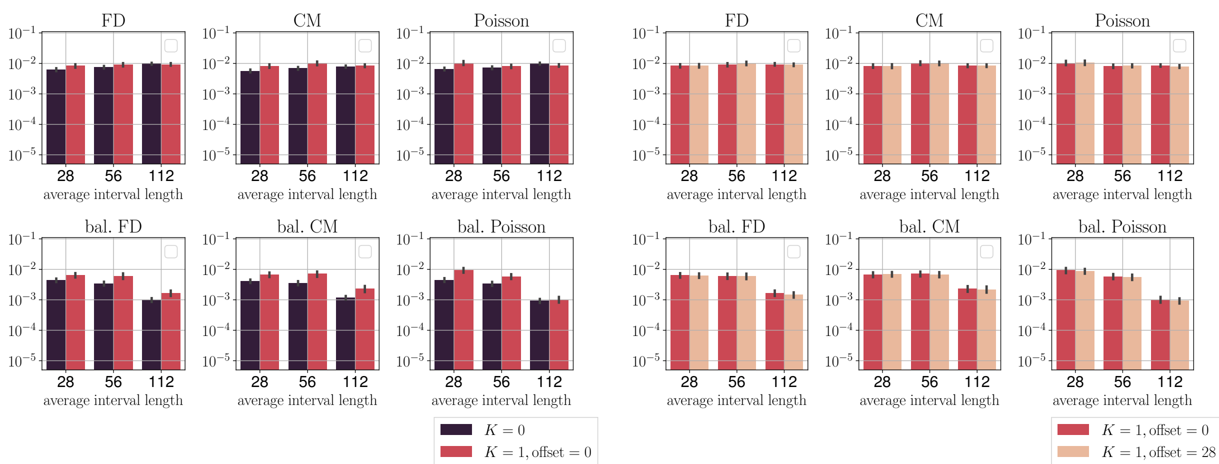


Figure 12 Fitted impulse response function of polynomial **degree 3**



(a) no vs. one simultaneous experiment

(b) zero- vs. 28-minute offset

Figure 13 Effect of simultaneous experiment on estimation error and effect of offset of switching points on estimation error

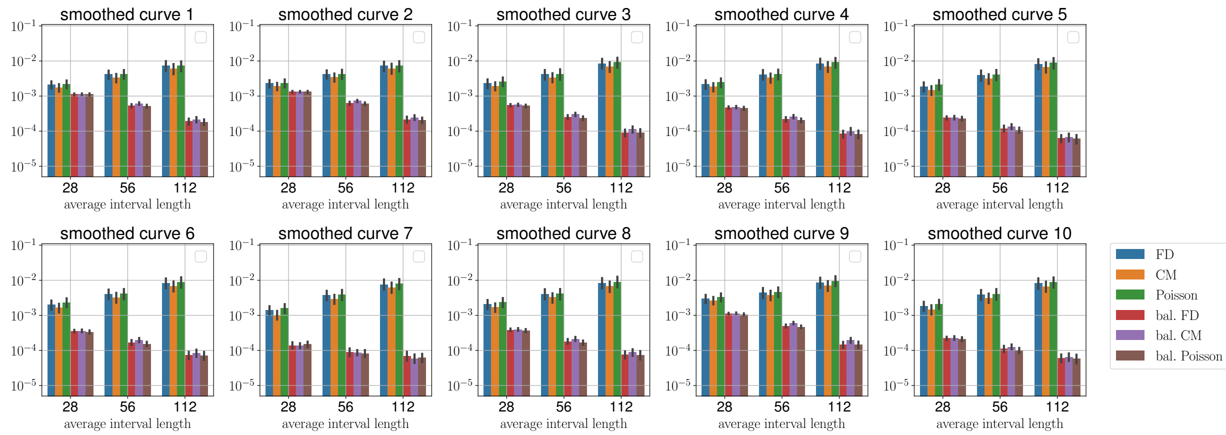


Figure 14 MSE of various designs in simulated experiments on **control data** using fitted impulse response function of polynomial **degree 1**

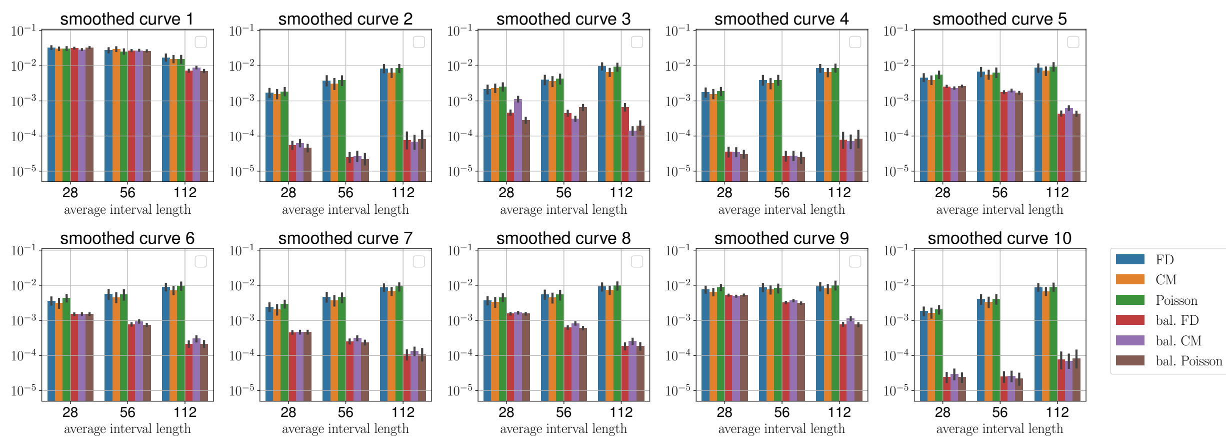


Figure 15 MSE of various designs in simulated experiments on **control data** using fitted impulse response function of polynomial **degree 2**

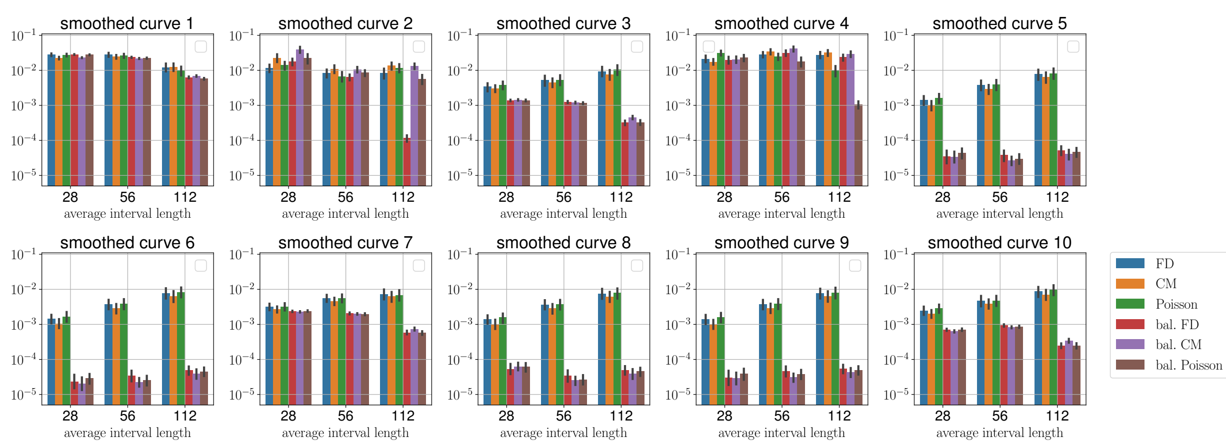


Figure 16 MSE of various designs in simulated experiments on **control data** using fitted impulse response function of polynomial **degree 3**

A.2. Additional examples

Below we show an example of the value of $C^{(m)}$ when the covariance decays linearly in the distance between t_i and t_j.

Example A.1. Suppose the covariance $\mathbb{E}_\varepsilon\left[\varepsilon^{(i)}\varepsilon^{(j)} \mid t_i, t_j\right]$ decays linearly in $|t_i - t_j|$ for all $t_j \in [t_i - h, t_i + h]$, and is zero outside this interval (i.e., $\mathbb{E}_\varepsilon\left[\varepsilon^{(i)}\varepsilon^{(j)} \mid t_i, t_j\right] = \sigma^2(h - |t_i - t_j|)/h$). Suppose the event density $f(t)$ is uniform in t. If $h < |\mathcal{I}_m|$, then $C^{(m)} = \sigma^2\big(|\mathcal{I}_m|^2 - |\mathcal{I}_m|h + 2h^2/3\big)/T^2$; otherwise, $C^{(m)} = \sigma^2\big(|\mathcal{I}_m|^2 - |\mathcal{I}_m|^3/(3h)\big)/T^2$.

Example A.2. Suppose event density $f(t)$ is uniform in t and the fixed-duration design is used. Furthermore, suppose the carryover effect δ_t^{co} is constant in t and carryover intensity is constant for $t' \in [t-h, t]$ for any t and for $h < T/M$. Then $I^{(m)} = \delta^{\mathrm{co}}(1/M - h/(2T))$.

A.3. Notations

Additional treatment effect estimands We additionally define the average instantaneous and carryover effects, which are building blocks of GATE. The average instantaneous effect δ^{inst} is defined as

$$\delta^{\mathrm{inst}} = \int \delta_t^{\mathrm{inst}} f(t)dt\,,$$

where δ_t^{inst} is the instantaneous treatment effect at time t that is defined as

$$\delta_t^{\mathrm{inst}} = Y_t(e_t, \mathbf{0}_t, \cdots, \mathbf{0}_t) - Y_t(\mathbf{0}_t, \mathbf{0}_t, \cdots, \mathbf{0}_t)$$

and $e_t = (0, \cdots, 0, \underbrace{1}_{\text{time } t}, 0, \cdots, 0)$ is a one-hot-encoded vector with the entry of time t to be 1 and all the remaining entries to be 0.

The average carryover effect $\delta_\ell^{\mathrm{co}}(\boldsymbol{w})$, given treatment assignments $\boldsymbol{w}$, is defined as

$$\delta^{\mathrm{co}}(\boldsymbol{w}) = \int \delta_t^{\mathrm{co}}(\boldsymbol{w}_t) f(t)dt\,,$$

where $\delta_t^{\mathrm{co}}(\boldsymbol{w}_t)$ is the carryover effect at time t that is defined as

$$\delta_t^{\mathrm{co}}(\boldsymbol{w}_t) = Y_t(\boldsymbol{w}_t, \mathbf{0}_t, \cdots, \mathbf{0}_t) - Y_t(\boldsymbol{w}_t \circ e_t, \mathbf{0}_t, \cdots, \mathbf{0}_t)$$

and "$\circ$" denotes the entry-wise product. Let $\delta^{\mathrm{co}} := \delta^{\mathrm{co}}(\mathbf{1})$ be the average carryover effect under global treatment. Then we can decompose the GATE as

$$\delta^{\mathrm{gate}} = \delta^{\mathrm{inst}} + \delta^{\mathrm{co}}.$$

The expression of $S_{\mathrm{var}}^{(m,m')}$ in $\mathbb{E}[\mathcal{E}_{\mathrm{simul}}^2]$ in Theorem 4.2

$S_{\mathrm{var}}^{(m,m')}$ is defined as

$$S_{\mathrm{var}}^{(m,m')} = 4\int_{t_i \in \mathcal{I}_m, t_j \in \mathcal{I}_{m'}} \Big(\mathbf{1}\,(m = m')\,\mathbb{E}_W[\delta_{t_i,t_j}^{\mathrm{simul},2}(\boldsymbol{W}) \mid t_i, t_j] + \mathbf{1}\,(m \neq m')\,\Phi_{t_i,t_j}^{2\dagger}\Big) f(t_i)f(t_j)dt_i dt_j\,, \tag{A.1}$$

where $\delta_{t_i,t_j}^{\mathrm{simul},2}(\boldsymbol{W})$ is equal to

$$\begin{aligned}\delta_{t_i,t_j}^{\mathrm{simul},2}(\boldsymbol{W}) = &\mathbb{E}_{\boldsymbol{W}_1^{\mathrm{s}},\cdots,\boldsymbol{W}_K^{\mathrm{s}}}\left[(Y_{t_i}(\boldsymbol{W}, \boldsymbol{W}_1^{\mathrm{s}}, \cdots, \boldsymbol{W}_K^{\mathrm{s}}) - Y_{t_i}(\boldsymbol{W}, \mathbf{0}, \cdots, \mathbf{0})) \times\right.\\ &\left.\big(Y_{t_j}(\boldsymbol{W}, \boldsymbol{W}_1^{\mathrm{s}}, \cdots, \boldsymbol{W}_K^{\mathrm{s}}) - Y_{t_j}(\boldsymbol{W}, \mathbf{0}, \cdots, \mathbf{0})\big) \mid \boldsymbol{W}, t_i, t_j\right]\,,\end{aligned}$$

and for $t_i \in \mathcal{I}_m$ and $t_j \in \mathcal{I}_{m'}$ with $m \neq m'$, $\Phi^{2\dagger}_{t_i,t_j}$ is equal to

$$\begin{aligned}\Phi^{2\dagger}_{t_i,t_j} =& \frac{1}{4}\Bigg(\mathbb{E}_{\boldsymbol{W}^{(-m,-m')}}\left[\delta^{\text{simul},2}_{t_i,t_j}(\boldsymbol{W}^{(-m,-m')}, \boldsymbol{W}^{(m,m')} = (1,1))\right] \\ &- \mathbb{E}_{\boldsymbol{W}^{(-m,-m')}}\left[\delta^{\text{simul},2}_{t_i,t_j}(\boldsymbol{W}^{(-m,-m')}, \boldsymbol{W}^{(m,m')} = (1,0))\right] \\ &- \mathbb{E}_{\boldsymbol{W}^{(-m,-m')}}\left[\delta^{\text{simul},2}_{t_i,t_j}(\boldsymbol{W}^{(-m,-m')}, \boldsymbol{W}^{(m,m')} = (0,1))\right] \\ &+ \mathbb{E}_{\boldsymbol{W}^{(-m,-m')}}\left[\delta^{\text{simul},2}_{t_i,t_j}(\boldsymbol{W}^{(-m,-m')}, \boldsymbol{W}^{(m,m')} = (0,0))\right]\Bigg).\end{aligned}$$

The term $\delta^{\text{simul},2}_{t_i,t_j}(\boldsymbol{W})$ is the expected product of simultaneous effects at time t_i and at time t_j, conditional on $\boldsymbol{W}$. The term $\Phi^{2\dagger}_{t_i,t_j}$ then measures the discrepancy in $\delta^{\text{simul},2}_{t_i,t_j}(\boldsymbol{W})$ by varying the values of $W^{(m)}$ and $W^{(m')}$ and marginalizing over $\boldsymbol{W}^{(-m,-m')}$.

The expression of $S^{(m,m')}_{\text{cov}}$ *in* $\mathbb{E}[(\mathcal{E}_{\text{inst}} + \mathcal{E}_{\text{carryover}}) \cdot \mathcal{E}_{\text{simul}}]$ *in Theorem 4.2*

$S^{(m,m')}_{\text{cov}}$ is defined as

$$S^{(m,m')}_{\text{cov}} = (\delta^{\text{gate}}\mu^{(m')} + 2\mu^{(m')}_{Y^{\text{ctrl}}}) \cdot S^{(m,m')}_1 + \left(\Xi^{\text{inst},(m)} - \delta^{\text{co}}\mu^{(m)}\right) S^{(m,m')}_2 + S^{(m,m')}_3 . \tag{A.2}$$

The term $S^{(m,m')}_1$ is defined as

$$S^{(m,m')}_1 = 2\int_{t_i \in \mathcal{I}_m} \left(\mathbf{1}(m = m')\,\mathbb{E}_W\left[\delta^{\text{simul}}_{t_i}(\boldsymbol{W})\right] + \mathbf{1}(m \neq m')\,\Phi^{\text{simul},(-m')}_{t_i}\right) f(t_i)dt_i , \tag{A.3}$$

where $\mathbb{E}_W\left[\delta^{\text{simul}}_{t_i}(\boldsymbol{W})\right]$ is defined in Section 4.2 and, for $t_i \in \mathcal{I}_m$ and $t_j \in \mathcal{I}_{m'}$ with $m \neq m'$, $\Phi^{\text{simul},(-m')}_{t_i}$ is defined as

$$\begin{aligned}\Phi^{\text{simul},(-m')}_{t_i} =& \frac{1}{4}\Bigg(\mathbb{E}_{\boldsymbol{W}^{(-m,-m')}}\left[\delta^{\text{simul}}_{t_i}(\boldsymbol{W}^{(-m,-m')}, \boldsymbol{W}^{(m,m')} = (1,1))\right] \\ &- \mathbb{E}_{\boldsymbol{W}^{(-m,-m')}}\left[\delta^{\text{simul}}_{t_i}(\boldsymbol{W}^{(-m,-m')}, \boldsymbol{W}^{(m,m')} = (1,0))\right] \\ &- \mathbb{E}_{\boldsymbol{W}^{(-m,-m')}}\left[\delta^{\text{simul}}_{t_i}(\boldsymbol{W}^{(-m,-m')}, \boldsymbol{W}^{(m,m')} = (0,1))\right] \\ &+ \mathbb{E}_{\boldsymbol{W}^{(-m,-m')}}\left[\delta^{\text{simul}}_{t_i}(\boldsymbol{W}^{(-m,-m')}, \boldsymbol{W}^{(m,m')} = (0,0))\right]\Bigg).\end{aligned}$$

Recall that $\delta^{\text{simul}}_{t_i}(\boldsymbol{W})$ is the expected simultaneous effects at time t_i, conditional on $\boldsymbol{W}$. Then $\mathbb{E}_W\left[\delta^{\text{simul}}_{t_i}(\boldsymbol{W})\right]$ is the expected simultaneous effects at time t_i, marginalized over $\boldsymbol{W}$.

The term $\mathbb{E}_{\boldsymbol{W}^{(-m,-m')}}\left[\delta^{\text{simul}}_{t_i}(\boldsymbol{W}^{(-m,-m')}, \boldsymbol{W}^{(m,m')})\right]$ is the simultaneous effects at time t_i, marginalized over $\boldsymbol{W}^{(-m,-m')}$, but conditional on $\boldsymbol{W}^{(m,m')}$, where

$$\boldsymbol{W}^{(m,m')} = (W^{(m)}, W^{(m')}), \qquad \boldsymbol{W}^{(-m,-m')} = \boldsymbol{W} \backslash \boldsymbol{W}^{(m,m')} .$$

$\boldsymbol{W}^{(-m,-m')}$ is an $M-2$ dimensional vector denoting the treatment status of the main intervention for all intervals excluding the m-th and m'-th intervals.

$\Phi^{\text{simul},(-m')}_{t_i}$ is a measure of the discrepancy in simultaneous effects by varying the values of $W^{(m)}$ and $W^{(m')}$. $\Phi^{\text{simul},(-m')}_{t_i}$ is closely connected to $\Phi^{\text{simul}}_{t_i}$ defined in Section 4.2 in that $\Phi^{\text{simul}}_{t_i}$ measures the discrepancy in simultaneous effects by varying the value of $W^{(m)}$, while both $W^{(m)}$ and $W^{(m')}$ are varied in the definition of $\Phi^{\text{simul},(-m')}_{t_i}$.

In addition, the term $S_2^{(m,m')}$ is defined as

$$S_2^{(m,m')} = 2\int_{t_j \in \mathcal{I}_{m'}} \left(\mathbf{1}\,(m=m')\,\mathbb{E}_{\boldsymbol{W}^{(-m)}}\left[\delta_{t_j}^{\text{simul}}(\boldsymbol{W}^{(-m)}, W^{(m)}=1)\right] + \mathbf{1}\,(m \neq m')\,\Phi_{t_j}^{\text{simul},(-m')\dagger} \right) f(t_j)dt_j\,, \tag{A.4}$$

where for $t_j \in \mathcal{I}_{m'}$, $\Phi_{t_j}^{\text{simul},(-m')\dagger}$ is equal to

$$\begin{aligned}\Phi_{t_j}^{\text{simul},(-m')\dagger} =& \frac{1}{2}\Bigg(\mathbb{E}_{\boldsymbol{W}^{(-m,-m')}}\left[\delta_{t_j}^{\dagger}(\boldsymbol{W}^{(-m,-m')}, \boldsymbol{W}^{(m,m')}=(1,1))\right] \\ &- \mathbb{E}_{\boldsymbol{W}^{(-m,-m')}}\left[\delta_{t_j}^{\dagger}(\boldsymbol{W}^{(-m,-m')}, \boldsymbol{W}^{(m,m')}=(1,0))\right]\Bigg).\end{aligned}$$

$S_2^{(m,m')}$ is conceptually very similar to $S_1^{(m,m')}$, but is applied to the simultaneous effects when $W^{(m)}=1$.

Lastly, $S_3^{(m,m')}$ is defined as

$$S_3^{(m,m')} = 4\int_{t_i \in \mathcal{I}_m, t_j \in \mathcal{I}_{m'}} \left(\mathbf{1}\,(m=m')\,\mathbb{E}_{W}\left[\delta_{t_i}^{\text{co}}(\boldsymbol{W})\delta_{t_j}^{\text{simul}}(\boldsymbol{W})\right] + \mathbf{1}\,(m \neq m')\,\Phi_{t_i,t_j}^{\text{co,simul}} \right) f(t_i)f(t_j)dt_i dt_j \tag{A.5}$$

where for $t_i \in \mathcal{I}_m$ and $t_j \in \mathcal{I}_{m'}$, $\Phi_{t_i,t_j}^{\text{co,simul}}$ is equal to

$$\begin{aligned}\Phi_{t_i,t_j}^{\text{co,simul}} =& \frac{1}{4}\Bigg(\mathbb{E}_{\boldsymbol{W}^{(-m,-m')}}\left[\delta_{t_i}^{\text{co}}(\boldsymbol{W}^{(-m,-m')}, \boldsymbol{W}^{(m,m')}=(1,1)) \cdot \delta_{t_j}^{\text{simul}}(\boldsymbol{W}^{(-m,-m')}, \boldsymbol{W}^{(m,m')}=(1,1))\right] \\ &- \mathbb{E}_{\boldsymbol{W}^{(-m,-m')}}\left[\delta_{t_i}^{\text{co}}(\boldsymbol{W}^{(-m,-m')}, \boldsymbol{W}^{(m,m')}=(1,0)) \cdot \delta_{t_j}^{\text{simul}}(\boldsymbol{W}^{(-m,-m')}, \boldsymbol{W}^{(m,m')}=(1,0))\right] \\ &- \mathbb{E}_{\boldsymbol{W}^{(-m,-m')}}\left[\delta_{t_i}^{\text{co}}(\boldsymbol{W}^{(-m,-m')}, \boldsymbol{W}^{(m,m')}=(0,1)) \cdot \delta_{t_j}^{\text{simul}}(\boldsymbol{W}^{(-m,-m')}, \boldsymbol{W}^{(m,m')}=(0,1))\right] \\ &+ \mathbb{E}_{\boldsymbol{W}^{(-m,-m')}}\left[\delta_{t_i}^{\text{co}}(\boldsymbol{W}^{(-m,-m')}, \boldsymbol{W}^{(m,m')}=(0,0)) \cdot \delta_{t_j}^{\text{simul}}(\boldsymbol{W}^{(-m,-m')}, \boldsymbol{W}^{(m,m')}=(0,0))\right]\Bigg)\end{aligned}$$

$S_3^{(m,m')}$ measures the expected value of the product of carryover effect at time t_i and simultaneous effect at time t_j.

A.4. Additional Examples

Example A.3 (Misalignment of switching times for simultaneous interventions). This example illustrates why the bias from simultaneous interventions can be reduced by misaligning the switching times of different interventions. This happens when the confounding effects increase nonlinearly with the times that interventions are jointly treated. Consider a simple example with two interventions, time-invariant treatment effects for the first intervention, and zero treatment effects for the second intervention unless jointly treated with the first intervention. Consider a simple design where intervals are of the same length. Each interval of the first intervention is partitioned into two sub-intervals a and b of the same length, and the treatment effects depend on the treatment assignments of sub-intervals a and b as follows

$$\begin{aligned}Y_t\left((W_{1a}, W_{1b}), (W_{2a}, W_{2b})\right) - Y_t\left((0,0),(0,0)\right) =& \frac{\delta_1}{2}(W_{1a}+W_{1b}) \\ &+ \frac{\delta_{12}}{4}(W_{1a}W_{2a}+W_{1b}W_{2b})^2\end{aligned}$$

for t either in sub-interval a or b. Below we use two designs to illustrate why randomizing switching times can be helpful. For the first design, the switching times of two interventions are aligned so that $((W_{1a}, W_{1b}), (W_{2a}, W_{2b}))$ is equal to each of the following realizations with probability 1/4

$$((1,1),(1,1)) \qquad ((1,1),(0,0)) \qquad ((0,0),(1,1)) \qquad ((0,0),(0,0))$$

The bias of the HT estimator $\hat{\delta}_1$ using the first design is

$$\mathbb{E}_{W,\varepsilon,t}[\hat{\delta}_1 - \delta_1] = \left(\frac{1}{2}\left[(\delta_1 + \delta_{12}) + \delta_1\right]\right) - \delta_1 = \frac{\delta_{12}}{2}$$

For the second design, the switching times of two interventions are not aligned, and the second intervention switches at the end of sub-interval a. Therefore, $((W_{1a}, W_{1b}), (W_{2a}, W_{2b}))$ is equal to each of the following realizations with probability 1/8

$$((1,1),(1,1)) \qquad ((1,1),(0,1)) \qquad ((1,1),(0,1)) \qquad ((1,1),(0,0))$$
$$((0,0),(1,1)) \qquad ((0,0),(0,1)) \qquad ((0,0),(1,0)) \qquad ((0,0),(0,0))$$

The bias of the HT estimator $\hat{\delta}_1$ using the second design is

$$\mathbb{E}_{W,\varepsilon,t}[\hat{\delta}_1 - \delta_1] = \left(\frac{1}{4}\left[(\delta_1 + \delta_{12}) + (\delta_1 + \delta_{12}/4) + (\delta_1 + \delta_{12}/4) + \delta_1\right]\right) - \delta_1 = \frac{3\delta_{12}}{8}$$

which is smaller than the bias by using the first design. The second design reduces the bias because the bias incurred by $(W_{2a}, W_{2b}) = (1,1)$ is much larger than that incurred by $(W_{2a}, W_{2b}) = (0,1)$ or $(1,0)$.

Appendix B: Proof of Main Results

B.1. Proof of Proposition 4.1

Proof of Proposition 4.1 The bias $\mathrm{Bias}(\mathcal{E}_{\mathrm{simul}})$ is zero following Example 4.5.

The expression of $S_{\mathrm{cov}}^{(m,m')}$ *in* $\mathbb{E}_\ell[(\mathcal{E}_{\mathrm{inst}}+\mathcal{E}_{\mathrm{carryover}})\cdot\mathcal{E}_{\mathrm{simul}}]$ *under Condition 1*

LEMMA B.1. *Under the assumptions in Theorem 4.2 and Condition 1,*

$$\mathrm{E}(\mathcal{E}_{\mathrm{simul}}^2)=\sum_{m=1}^{M} S_{\mathrm{var}}^{(m,m)},$$

where

$$\begin{aligned}S_{\mathrm{var}}^{(m,m)}=&\left(\int_{t_i\in\mathcal{I}_m}\left[\sum_{\ell=1}^{K}\delta_{\ell,t_i}^{\mathrm{s.gate}}\right]f(t_i)dt_i\right)^2\\&+\sum_{m'=1}^{M}\sum_{\ell=1}^{K}\left(\int_{t_i\in\mathcal{I}_m\cap\mathcal{I}_{\ell m'}^{\mathrm{s}}}\delta_{\ell,t_i}^{\mathrm{s.inst}}f(t_i)dt_i+\int_{t_j\in\mathcal{I}_m,t'\in\mathcal{I}_{\ell m'}^{\mathrm{s}}}\delta_{\ell,t_j}^{\mathrm{s.co}}d_{\ell,t_j}^{\mathrm{s.co}}(t')f(t')f(t_j)dt_jdt'\right)^2,\end{aligned}$$

and $S_{\mathrm{var}}^{(m,m')}=0$ *for* $m'\neq m$.

Proof of Lemma B.1 When treatment effects of simultaneous interventions are additive, the term $\delta_{t_i,t_j}^{\mathrm{simul},2}(\boldsymbol{W})$ in Equation (A.1) is

$$\begin{aligned}\delta_{t_i,t_j}^{\mathrm{simul},2}(\boldsymbol{W})=&\mathbb{E}_{\boldsymbol{W}_1^{\mathrm{s}},\cdots,\boldsymbol{W}_K^{\mathrm{s}}}\left[\left(\sum_{\ell=1}^{K}\left[W_{\ell,t_i}^{\mathrm{s}}\delta_{\ell,t_i}^{\mathrm{s.inst}}+\delta_{\ell,t_i}^{\mathrm{s.co}}\cdot\sum_{k=1}^{M}W_\ell^{\mathrm{s}(k)}\int_{t'\in\mathcal{I}_{\ell k}^{\mathrm{s}}}d_{\ell,t_i}^{\mathrm{s.co}}(t')f(t')dt'\right]\right)\right.\\&\left.\left(\sum_{\ell=1}^{K}\left[W_{\ell,t_j}^{\mathrm{s}}\delta_{\ell,t_j}^{\mathrm{s.inst}}+\delta_{\ell,t_j}^{\mathrm{s.co}}\cdot\sum_{k=1}^{M}W_\ell^{\mathrm{s}(k)}\int_{t'\in\mathcal{I}_{\ell k}^{\mathrm{s}}}d_{\ell,t_j}^{\mathrm{s.co}}(t')f(t')dt'\right]\right)\mid\boldsymbol{W},t_i,t_j\right]\\=&\frac{1}{4}\left(\sum_{\ell=1}^{K}\delta_{\ell,t_i}^{\mathrm{s.gate}}\right)\left(\sum_{\ell=1}^{K}\delta_{\ell,t_j}^{\mathrm{s.gate}}\right)\\&+\frac{1}{4}\sum_{\ell=1}^{K}\delta_{\ell,t_i}^{\mathrm{s.co}}\delta_{\ell,t_j}^{\mathrm{s.co}}\sum_{k=1}^{M}\left(\int_{t'\in\mathcal{I}_{\ell k}^{\mathrm{s}}}d_{\ell,t_i}^{\mathrm{s.co}}(t')f(t')dt'\right)\left(\int_{t'\in\mathcal{I}_{\ell k}^{\mathrm{s}}}d_{\ell,t_j}^{\mathrm{s.co}}(t')f(t')dt'\right)\\&+\frac{1}{4}\sum_{\ell=1}^{K}\delta_{\ell,t_i}^{\mathrm{s.inst}}\delta_{\ell,t_j}^{\mathrm{s.co}}\int_{t'\in\mathcal{I}_{\ell m(t_i)}^{\mathrm{s}}}d_{\ell,t_j}^{\mathrm{s.co}}(t')f(t')dt'\\&\qquad(\mathcal{I}_{\ell m(t_i)}^{\mathrm{s}}\text{ denotes the interval of simul. intervention }\ell\text{ to which }t_i\text{ belongs})\\&+\frac{1}{4}\sum_{\ell=1}^{K}\delta_{\ell,t_j}^{\mathrm{s.inst}}\delta_{\ell,t_i}^{\mathrm{s.co}}\int_{t'\in\mathcal{I}_{\ell m(t_j)}^{\mathrm{s}}}d_{\ell,t_i}^{\mathrm{s.co}}(t')f(t')dt'\\&+\frac{1}{4}\sum_{\ell=1}^{K}\mathbf{1}\left(t_i\text{ and }t_j\text{ in the same interval of simul. intervention }\ell\right)\delta_{\ell,t_i}^{\mathrm{s.inst}}\delta_{\ell,t_j}^{\mathrm{s.inst}}.\end{aligned}$$

We can see that $\delta_{t_i,t_j}^{\mathrm{simul},2}(\boldsymbol{W})$ does not depend on $\boldsymbol{W}$ for any t_i and t_j. Then the term $\Phi_{t_i,t_j}^{2\dagger}$ in Equation (A.1) is 0, and we have

$$S_{\mathrm{var}}^{(m,m')}=0\qquad\text{for } m\neq m'.$$

For $m=m'$, we have

$$S_{\mathrm{var}}^{(m,m)}=4\int_{t_i,t_j\in\mathcal{I}_m}\mathbb{E}_W[\delta_{t_i,t_j}^{\mathrm{simul},2}(\boldsymbol{W})\mid t_i,t_j]f(t_i)f(t_j)dt_idt_j$$

$$
\begin{aligned}
&= \int_{t_i,t_j\in\mathcal{I}_m} \left(\sum_{\ell=1}^{K} \delta^{\text{s.gate}}_{\ell,t_i}\right)\left(\sum_{\ell=1}^{K} \delta^{\text{s.gate}}_{\ell,t_j}\right) f(t_i)f(t_j)dt_idt_j \\
&\quad + \sum_{\ell=1}^{K}\sum_{k=1}^{M} \int_{t_i,t_j\in\mathcal{I}_m} \delta^{\text{s.co}}_{\ell,t_i}\delta^{\text{s.co}}_{\ell,t_j} \left(\int_{t'\in\mathcal{I}^{\text{s}}_{\ell k}} d^{\text{s.co}}_{\ell,t_i}(t')f(t')dt'\right)\left(\int_{t'\in\mathcal{I}^{\text{s}}_{\ell k}} d^{\text{s.co}}_{\ell,t_j}(t')f(t')dt'\right) f(t_i)f(t_j)dt_idt_j \\
&\quad + 2\sum_{\ell=1}^{K}\sum_{m'=1}^{M} \int_{t_i\in\mathcal{I}_m\cap\mathcal{I}^{\text{s}}_{\ell m'},\, t_j\in\mathcal{I}_m} \delta^{\text{s.inst}}_{\ell,t_i}\delta^{\text{s.co}}_{\ell,t_j} \left(\int_{t'\in\mathcal{I}^{\text{s}}_{\ell m'}} d^{\text{s.co}}_{\ell,t_j}(t')f(t')dt'\right) f(t_i)f(t_j)dt_idt_j \\
&\quad + \sum_{\ell=1}^{K}\sum_{m'=1}^{M} \int_{t_i,t_j\in\mathcal{I}_m\cap\mathcal{I}^{\text{s}}_{\ell m'}} \delta^{\text{s.inst}}_{\ell,t_i}\delta^{\text{s.inst}}_{\ell,t_j} f(t_i)f(t_j)dt_idt_j\,.
\end{aligned}
$$

We can further simplify $S^{(m,m)}_{\text{var}}$ to

$$
\begin{aligned}
S^{(m,m)}_{\text{var}} &= \left(\int_{t_i\in\mathcal{I}_m} \left[\sum_{\ell=1}^{K} \delta^{\text{s.gate}}_{\ell,t_i}\right] f(t_i)dt_i\right)^2 + \sum_{\ell=1}^{K}\sum_{m'=1}^{M} \left(\int_{t_i\in\mathcal{I}_m,\, t'\in\mathcal{I}^{\text{s}}_{\ell m'}} \delta^{\text{s.co}}_{\ell,t_i} d^{\text{s.co}}_{\ell,t_i}(t')f(t')f(t_i)dt_idt'\right)^2 \\
&\quad + 2\sum_{\ell=1}^{K}\sum_{m'=1}^{M} \left(\int_{t_i\in\mathcal{I}_m\cap\mathcal{I}^{\text{s}}_{\ell m'}} \delta^{\text{s.inst}}_{\ell,t_i} f(t_i)dt_i\right)\left(\int_{t_j\in\mathcal{I}_m,\, t'\in\mathcal{I}^{\text{s}}_{\ell m'}} \delta^{\text{s.co}}_{\ell,t_j} d^{\text{s.co}}_{\ell,t_j}(t')f(t')f(t_j)dt_jdt'\right) \\
&\quad + \sum_{\ell=1}^{K}\sum_{m'=1}^{M} \left(\int_{t_i\in\mathcal{I}_m\cap\mathcal{I}^{\text{s}}_{\ell m'}} \delta^{\text{s.inst}}_{\ell,t_i} f(t_i)dt_i\right)^2 \\
&= \left(\int_{t_i\in\mathcal{I}_m} \left[\sum_{\ell=1}^{K} \delta^{\text{s.gate}}_{\ell,t_i}\right] f(t_i)dt_i\right)^2 \\
&\quad + \sum_{\ell=1}^{K}\sum_{m'=1}^{M} \left(\int_{t_i\in\mathcal{I}_m\cap\mathcal{I}^{\text{s}}_{\ell m'}} \delta^{\text{s.inst}}_{\ell,t_i} f(t_i)dt_i + \int_{t_j\in\mathcal{I}_m,\, t'\in\mathcal{I}^{\text{s}}_{\ell m'}} \delta^{\text{s.co}}_{\ell,t_j} d^{\text{s.co}}_{\ell,t_j}(t')f(t')f(t_j)dt_jdt'\right)^2.
\end{aligned}
$$

In summary, $\mathrm{E}(\mathcal{E}^2_{\text{simul}})$ is equal to

$$
\mathrm{E}(\mathcal{E}^2_{\text{simul}}) = \sum_{m=1}^{M}\sum_{m'=1}^{M} S^{(m,m')}_{\text{var}} = \sum_{m=1}^{M} S^{(m)}_{\text{var}}\,.
$$

The expression of $S^{(m,m')}_{\text{cov}}$ in $\mathbb{E}[(\mathcal{E}_{\text{inst}} + \mathcal{E}_{\text{carryover}})\cdot\mathcal{E}_{\text{simul}}]$ under Condition 1

LEMMA B.2. *Under the assumptions in Theorem 4.2 and Condition 1,*

$$
\mathbb{E}[(\mathcal{E}_{\text{inst}} + \mathcal{E}_{\text{carryover}})\cdot\mathcal{E}_{\text{simul}}] = \sum_{m=1}^{M} \left(\Xi^{(m)} + 2\mu^{(m)}_{Y^{\text{ctrl}}}\right)\left(\int_{t_i\in\mathcal{I}_m} \left[\sum_{\ell=1}^{K} \delta^{\text{s.gate}}_{\ell,t_i}\right] f(t_i)dt_i\right).
$$

Proof of Lemma B.2 As shown in Theorem 4.2,

$$
\begin{aligned}
\mathbb{E}_\ell[(\mathcal{E}_{\text{inst}} + \mathcal{E}_{\text{carryover}})\cdot\mathcal{E}_{\text{simul}}] &= \sum_{m=1}^{M}\sum_{m'=1}^{M} \\
&\quad \left[(\delta^{\text{gate}}\mu^{(m')} + 2\mu^{(m)}_{Y^{\text{ctrl}}})\cdot S^{(m,m')}_1 + \left(\Xi^{\text{inst},(m)} - \delta^{\text{co}}\mu^{(m)}\right) S^{(m,m')}_2 + S^{(m,m')}_3\right]
\end{aligned}
$$

In Lemmas B.3, B.4, and B.5 below, we show the expression of $S^{(m,m')}_1$, $S^{(m,m')}_2$, and $S^{(m,m')}_3$ under Condition 1.

LEMMA B.3. *Under the assumptions in Theorem 4.2 and Condition 1,*

$$
S^{(m,m)}_1 = \sum_{\ell=1}^{K} \left(\int_{t_i\in\mathcal{I}_m} \delta^{\text{s.gate}}_{\ell,t_i} f(t_i)dt_i\right).
$$

Proof of Lemma B.3 If the effects of main and simultaneous interventions are additive, the term $\delta_{t_i}^{\mathrm{simul}}(\boldsymbol{W})$ in Equation (A.3) is equal to

$$\begin{aligned}\delta_{t_i}^{\mathrm{simul}}(\boldsymbol{W}) =& \mathbb{E}_{\boldsymbol{W}_1^{\mathrm{s}},\cdots,\boldsymbol{W}_K^{\mathrm{s}}}\left[Y_{t_i}(\boldsymbol{W},\boldsymbol{W}_1^{\mathrm{s}},\cdots,\boldsymbol{W}_K^{\mathrm{s}}) - Y_{t_i}(\boldsymbol{W},\boldsymbol{0},\cdots,\boldsymbol{0}) \mid \boldsymbol{W}, t_i\right] \\ =& \mathbb{E}_{\boldsymbol{W}_1^{\mathrm{s}},\cdots,\boldsymbol{W}_K^{\mathrm{s}}}\left[\sum_{\ell=1}^{K}\left[W_{\ell,t_i}^{\mathrm{s}}\delta_{\ell,t_i}^{\mathrm{s.inst}} + \delta_{\ell,t_i}^{\mathrm{s.co}}\cdot\sum_{k=1}^{M}W_{\ell}^{\mathrm{s}(k)}\int_{t'\in\mathcal{I}_{\ell k}^{\mathrm{s}}} d_{\ell,t_i}^{\mathrm{s.co}}(t')f(t')dt'\right] \mid \boldsymbol{W}, t_i, t_j\right] \\ =& \frac{1}{2}\sum_{\ell=1}^{K}\left[\delta_{\ell,t_i}^{\mathrm{s.inst}} + \delta_{\ell,t_i}^{\mathrm{s.co}}\cdot\sum_{k=1}^{M}\int_{t'\in\mathcal{I}_{\ell k}^{\mathrm{s}}} d_{\ell,t_i}^{\mathrm{s.co}}(t')f(t')dt'\right] = \frac{1}{2}\sum_{\ell=1}^{K}\delta_{\ell,t_i}^{\mathrm{s.gate}}\end{aligned}$$

$\delta_{t_i}^{\mathrm{simul}}(\boldsymbol{W})$ does not depend on the value of $\boldsymbol{W}$. Then for $m \neq m'$, the term $\Phi_{t_i}^{\mathrm{simul},(-m')}$ in Equation (A.3) is 0, and

$$S_1^{(m,m')} = 0 \qquad m \neq m' .$$

For $m = m'$, we have

$$S_1^{(m,m)} = 2\int_{t_i\in\mathcal{I}_m}\left(\frac{1}{2}\sum_{\ell=1}^{K}\delta_{\ell,t_i}^{\mathrm{s.gate}}\right)f(t_i)dt_i = \sum_{\ell=1}^{K}\left(\int_{t_i\in\mathcal{I}_m}\delta_{\ell,t_i}^{\mathrm{s.gate}}f(t_i)dt_i\right) .$$

□

Lemma B.4. *Under the assumptions in Theorem 4.2 and Condition 1,*

$$S_2^{(m,m)} = \sum_{\ell=1}^{K}\left(\int_{t_j\in\mathcal{I}_m}\delta_{\ell,t_j}^{\mathrm{s.gate}}f(t_j)dt_j\right) .$$

Proof of Lemma B.4 If the effects of main and simultaneous interventions are additive, the term $\mathbb{E}_{\boldsymbol{W}^{(-m)}}\left[\delta_{t_j}^{\mathrm{simul}}(\boldsymbol{W}^{(-m)}, W^{(m)} = 1)\right]$ in Equation (A.4) is equal to

$$\mathbb{E}_{\boldsymbol{W}^{(-m)}}\left[\delta_{t_j}^{\mathrm{simul}}(\boldsymbol{W}^{(-m)}, W^{(m)} = 1)\right] = \frac{1}{2}\sum_{\ell=1}^{K}\delta_{\ell,t_i}^{\mathrm{s.gate}}$$

and the term $\Phi_{t_j}^{\mathrm{simul},(-m')\dagger}$ in Equation (A.4) is 0 for $m \neq m'$, and therefore

$$S_2^{(m,m')} = 0 \qquad m \neq m'$$

and

$$S_2^{(m,m)} = 2\int_{t_j\in\mathcal{I}_m}\left(\frac{1}{2}\sum_{\ell=1}^{K}\delta_{\ell,t_j}^{\mathrm{s.gate}}\right)f(t_j)dt_j = \sum_{\ell=1}^{K}\left(\int_{t_j\in\mathcal{I}_m}\delta_{\ell,t_j}^{\mathrm{s.gate}}f(t_j)dt_j\right) ,$$

which is equal to $S_1^{(m,m)}$. □

Lemma B.5. *Under the assumptions in Theorem 4.2 and Condition 1,*

$$S_3^{(m,m)} = \left(\int_{t_i\in\mathcal{I}_m}\delta_{t_i}^{\mathrm{co}}f(t_i)dt_i\right)\left(\sum_{\ell=1}^{K}\int_{t_j\in\mathcal{I}_m}\delta_{\ell,t_j}^{\mathrm{s.gate}}f(t_j)dt_j\right) .$$

Proof of Lemma B.5 If the effects of main and simultaneous interventions are additive, from Lemma B.3, we have

$$\delta_{t_j}^{\mathrm{simul}}(\boldsymbol{W}) = \frac{1}{2}\sum_{\ell=1}^{K}\delta_{\ell,t_j}^{\mathrm{s.gate}} ,$$

which does not depend on $\boldsymbol{W}$. We then have

$$\mathbb{E}_W\left[\delta^{\text{co}}_{t_i}(\boldsymbol{W})\delta^{\text{simul}}_{t_j}(\boldsymbol{W})\right]=\mathbb{E}_W\left[\delta^{\text{co}}_{t_i}(\boldsymbol{W})\right]\cdot\frac{1}{2}\sum_{\ell=1}^{K}\delta^{\text{s.gate}}_{\ell,t_j}=\frac{1}{4}\delta^{\text{co}}_{t_i}\sum_{\ell=1}^{K}\delta^{\text{s.gate}}_{\ell,t_j}$$

following that

$$\mathbb{E}_W\left[\delta^{\text{co}}_{t_i}(\boldsymbol{W})\right]=\mathbb{E}_W\left[\delta^{\text{co}}_{t_i}\cdot\sum_{k=1}^{M}W^{(k)}\int_{t'\in\mathcal{I}_k}d^{\text{co}}_{t_i}(t')f(t')dt'\right]=\frac{1}{2}\delta^{\text{co}}_{t_i}.$$

Therefore, if $t_i\in\mathcal{I}_m$ and $t_j\in\mathcal{I}_{m'}$ with $m\neq m'$, we have $\Phi^{\text{co,simul}}_{t_i,t_j}=0$, and then

$$S_3^{(m,m')}=0\qquad m\neq m'.$$

When $m=m'$, we have

$$\begin{aligned}
S_3^{(m,m)}&=\int_{t_i,t_j\in\mathcal{I}_m}\delta^{\text{co}}_{t_i}\left(\sum_{\ell=1}^{K}\delta^{\text{s.gate}}_{\ell,t_j}\right)f(t_i)f(t_j)dt_idt_j\\
&=\left(\int_{t_i\in\mathcal{I}_m}\delta^{\text{co}}_{t_i}f(t_i)dt_i\right)\left(\sum_{\ell=1}^{K}\int_{t_j\in\mathcal{I}_m}\delta^{\text{s.gate}}_{\ell,t_j}f(t_j)dt_j\right)\\
&=\tilde{\Xi}^{\text{co},(m)}\left(\sum_{\ell=1}^{K}\int_{t_j\in\mathcal{I}_m}\delta^{\text{s.gate}}_{\ell,t_j}f(t_j)dt_j\right).
\end{aligned}$$

□

From Lemmas B.3, B.4, and B.5, we have for $m\neq m'$,

$$S^{(m,m')}_{\text{cov}}=(\delta^{\text{gate}}\mu^{(m')}+2\mu^{(m)}_{Y\text{ctrl}})\cdot S_1^{(m,m')}+\left(\Xi^{\text{inst},(m)}-\delta^{\text{co}}\mu^{(m)}\right)S_2^{(m,m')}+S_3^{(m,m')}=0$$

For $m=m'$, we have

$$\begin{aligned}
S^{(m,m)}_{\text{cov}}=&(\delta^{\text{gate}}\mu^{(m)}+2\mu^{(m)}_{Y\text{ctrl}})\cdot S_1^{(m,m)}+\left(\Xi^{\text{inst},(m)}-\delta^{\text{co}}\mu^{(m)}\right)S_2^{(m,m)}+S_3^{(m,m)}\\
=&\left(\delta^{\text{gate}}\mu^{(m)}+2\mu^{(m)}_{Y\text{ctrl}}\right)\sum_{\ell=1}^{K}\left(\int_{t_i\in\mathcal{I}_m}\delta^{\text{s.gate}}_{\ell,t_i}f(t_i)dt_i\right)\\
&+\left(\Xi^{\text{inst},(m)}-\delta^{\text{co}}\mu^{(m)}\right)\sum_{\ell=1}^{K}\left(\int_{t_j\in\mathcal{I}_m}\delta^{\text{s.gate}}_{\ell,t_j}f(t_j)dt_j\right)\\
&+\left(\int_{t_i\in\mathcal{I}_m}\delta^{\text{co}}_{t_i}f(t_i)dt_i\right)\left(\sum_{\ell=1}^{K}\int_{t_j\in\mathcal{I}_m}\delta^{\text{s.gate}}_{\ell,t_j}f(t_j)dt_j\right)\\
=&\left(\Xi^{(m)}+2\mu^{(m)}_{Y\text{ctrl}}\right)\left(\int_{t_i\in\mathcal{I}_m}\left[\sum_{\ell=1}^{K}\delta^{\text{s.gate}}_{\ell,t_i}\right]f(t_i)dt_i\right).
\end{aligned}$$

following that $\Xi^{(m)}=\Xi^{\text{inst},(m)}+\Xi^{\text{co},(m)}+\delta^{\text{gate}}\mu^{(m)}$ and $\Xi^{\text{co},(m)}=\tilde{\Xi}^{\text{co},(m)}-\delta^{\text{co}}\mu^{(m)}$.

Then we have

$$\begin{aligned}
\mathbb{E}_\ell[(\mathcal{E}_{\text{inst}}+\mathcal{E}_{\text{carryover}})\cdot\mathcal{E}_{\text{simul}}]=&\sum_{m=1}^{M}\sum_{m'=1}^{M}S^{(m,m')}_{\text{cov}}\\
&=\sum_{m=1}^{M}\left(\Xi^{(m)}+2\mu^{(m)}_{Y\text{ctrl}}\right)\left(\int_{t_i\in\mathcal{I}_m}\left[\sum_{\ell=1}^{K}\delta^{\text{s.gate}}_{\ell,t_i}\right]f(t_i)dt_i\right).
\end{aligned}$$

□

B.2. Proof of Theorem 4.1

The estimation error of $\hat{\delta}^{\text{gate}}$ can be decomposed as

$$\begin{aligned}\hat{\delta}^{\text{gate}} - \delta^{\text{gate}} = & \underbrace{\delta^{\text{gate}}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{W_{t_i}}{\pi} - 1\right)}_{\substack{\text{instantaneous effects}\\ \text{denoted by } \mathcal{E}_{\text{inst}}}} + \underbrace{\frac{1}{n}\sum_{i=1}^{n}\alpha_{t_i}\left(Y_{t_i}(\boldsymbol{W}, \boldsymbol{0}, \cdots, \boldsymbol{0}) - W_{t_i}\delta^{\text{gate}}\right)}_{\substack{\text{carryover effects}\\ \text{denoted by } \mathcal{E}_{\text{carryover}}}} \\ & + \underbrace{\frac{1}{n}\sum_{i=1}^{n}\alpha_{t_i}\left(Y_{t_i}(\boldsymbol{W}, \boldsymbol{W}_1^{\text{s}}, \cdots, \boldsymbol{W}_K^{\text{s}}) - Y_{t_i}(\boldsymbol{W}, \boldsymbol{0}, \cdots, \boldsymbol{0})\right)}_{\text{effects from other interventions, denoted by } \mathcal{E}_{\text{simul}}} \quad . \\ & + \underbrace{\frac{1}{n}\sum_{i=1}^{n}\alpha_{t_i}\varepsilon^{(i)}}_{\substack{\text{measurement errors}\\ \text{denoted by } \mathcal{E}_{\text{meas}}}}\end{aligned} \tag{B.6}$$

In the following three lemmas, we show the expected value of $\mathcal{E}_{\text{meas}}$, $\mathcal{E}_{\text{inst}}$, $\mathcal{E}_{\text{carryover}}$, and $\mathcal{E}_{\text{simul}}$ in Equation (B.6). If the expected value of a term is nonzero, then this term results in an estimation bias of $\hat{\delta}^{\text{gate}}$.

LEMMA B.6 **(Mean of measurement errors and constant term)**. *Under the assumptions in Theorem 4.1, the mean of measurement errors is*

$$\mathbb{E}_{W,\varepsilon,t}\left[\mathcal{E}_{\text{meas}}\right] = \mathbb{E}_{W,\varepsilon,t}\left[\frac{1}{n}\sum_{i=1}^{n}\alpha_{t_i}\varepsilon^{(i)}\right] = 0\,.$$

Proof of Lemma B.6 The mean of the measurement errors is

$$\mathbb{E}_{W,\varepsilon,t}\left[\frac{1}{n}\sum_{i=1}^{n}\alpha_{t_i}\varepsilon^{(i)}\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{W,t_i}\left[\alpha_{t_i}\mathbb{E}_{\varepsilon}\left[\varepsilon^{(i)} \mid \boldsymbol{W}, \boldsymbol{W}_1^{\text{s}}, \cdots, \boldsymbol{W}_K^{\text{s}}, t_i\right]\right] = 0$$

following that $\varepsilon^{(i)}$ has mean zero and is independent of $\boldsymbol{W}, \boldsymbol{W}_1^{\text{s}}, \cdots, \boldsymbol{W}_K^{\text{s}}$. □

LEMMA B.7 **(Instantaneous and Carryover effects)**. *Under the assumptions in Theorem 4.1, the carryover bias equals to*

$$\begin{aligned}\mathbb{E}_{W,\varepsilon,t}\left[\mathcal{E}_{\text{inst}}\right] &= \mathbb{E}_{W,\varepsilon,t}\left[\delta^{\text{gate}}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{W_{t_i}}{\pi} - 1\right)\right] = 0 \\ \mathbb{E}_{W,\varepsilon,t}\left[\mathcal{E}_{\text{carryover}}\right] &= \mathbb{E}_{W,\varepsilon,t}\left[\frac{1}{n}\sum_{i=1}^{n}\alpha_{t_i}\left(Y_{t_i}(\boldsymbol{W}, \boldsymbol{0}, \cdots, \boldsymbol{0}) - W_{t_i}\delta^{\text{gate}}\right)\right] = \sum_{m=1}^{M} I^{(m)} - \delta^{\text{co}}\,.\end{aligned}$$

Proof of Lemma B.7 The expected value of $\mathcal{E}_{\text{inst}}$ is equal to

$$\mathbb{E}_{W,\varepsilon,t}\left[\mathcal{E}_{\text{inst}}\right] = \mathbb{E}_{W,\varepsilon,t}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{W_{t_i}}{\pi} - 1\right] = \frac{1}{n}\sum_{i=1}^{n}\frac{\mathbb{E}_{W,\varepsilon,t}[W_{t_i}]}{\pi} - 1 = 0\,.$$

As events are sampled i.i.d. from distribution $f(t)$, the expected value of $\mathcal{E}_{\text{carryover}}$ is equal to

$$\begin{aligned}&\mathbb{E}_{W,\varepsilon,t}\left[\mathcal{E}_{\text{carryover}}\right] = \mathbb{E}_{W,\varepsilon,t}\left[\frac{1}{n}\sum_{i=1}^{n}\alpha_{t_i}\left(Y_{t_i}(\boldsymbol{W}, \boldsymbol{0}, \cdots, \boldsymbol{0}) - W_{t_i}\delta^{\text{gate}}\right)\right] \\ =&\mathbb{E}_{W,\varepsilon,t}\left[\frac{1}{n}\sum_{i=1}^{n}\alpha_{t_i}\left(Y_{t_i}(\boldsymbol{W}, \boldsymbol{0}, \cdots, \boldsymbol{0}) - Y_{t_i}(\boldsymbol{0}, \boldsymbol{0}, \cdots, \boldsymbol{0}) - W_{t_i}\delta^{\text{gate}}\right)\right] \\ &+ \mathbb{E}_{W,\varepsilon,t}\left[\frac{1}{n}\sum_{i=1}^{n}\alpha_{t_i}Y_{t_i}(\boldsymbol{0}, \boldsymbol{0}, \cdots, \boldsymbol{0})\right]\end{aligned}$$

$$
\begin{aligned}
=&\mathbb{E}_{W,t}\left[\alpha_{t_i}(\delta^{\text{inst}}_{t_i}-\delta^{\text{inst}})W_{t_i}+\alpha_{t_i}(\delta^{\text{co}}_{t_i}(\boldsymbol{W})-\delta^{\text{co}}W_{t_i})\right]+\mathbb{E}_{W,t}\left[\alpha_{t_i}Y_{t_i}(\mathbf{0},\mathbf{0},\cdots,\mathbf{0})\right]\\
&\qquad\text{(the expected over } \varepsilon \text{ is dropped as marketplace outcomes do not depend on } \varepsilon^{(i)})\\
=&\mathbb{E}_t\left[(\delta^{\text{inst}}_{t_i}-\delta^{\text{inst}})\cdot\underbrace{\mathbb{E}_W\left[\frac{W_{t_i}}{\pi}\mid t_i\right]}_{=1}\right]+\mathbb{E}_{W,t}\left[\alpha_{t_i}(\delta^{\text{co}}_{t_i}(\boldsymbol{W})-\delta^{\text{co}}W_{t_i})\right] \qquad \text{(multiple } \alpha_{t_i} \text{ by } W_{t_i})\\
&+\mathbb{E}_{W,t}\big[\underbrace{\mathbb{E}_W[\alpha_{t_i}\mid t_i]}_{=0}\cdot Y_{t_i}(\mathbf{0},\mathbf{0},\cdots,\mathbf{0})\big]\\
=&\mathbb{E}_{W,t}\left[\alpha_{t_i}(\delta^{\text{co}}_{t_i}(\boldsymbol{W})-\delta^{\text{co}}W_{t_i})\right]\\
&\qquad\text{(the first term is zero because } \mathbb{E}_t[\delta^{\text{inst}}_{t_i}]=\delta^{\text{inst}})
\end{aligned}
$$

By Assumption 4.2, the last line can be further simplified to

$$
\begin{aligned}
&\mathbb{E}_{W,t}\left[\alpha_{t_i}(\delta^{\text{co}}_{t_i}(\boldsymbol{W})-\delta^{\text{co}}W_{t_i})\right]=\mathbb{E}_{W,t}\left[\frac{W_{t_i}-\pi}{\pi(1-\pi)}\sum_{k=1}^{M}\delta^{\text{co}}_{t_i}W^{(k)}\int_{t'\in\mathcal{I}_k}d^{\text{co}}_{t_i}(t')f(t')dt'\right]-\delta^{\text{co}}\\
=&\mathbb{E}_{W,\varepsilon,t}\left[\frac{W_{t_i}-\pi}{\pi(1-\pi)}\delta^{\text{co}}_{t_i}\sum_{k=1}^{M}W^{(k)}\mathbb{1}(t_i\in\mathcal{I}_k)\int_{t'\in\mathcal{I}_k}d^{\text{co}}_{t_i}(t')f(t')dt'\right]-\delta^{\text{co}}\\
&\qquad\text{(the treatment assignments of any two intervals are independent and } \mathbb{E}_{W,\varepsilon,t}\left[\tfrac{W_{t_i}-\pi}{\pi(1-\pi)}\right]=0)\\
=&\sum_{m=1}^{M}\int_{t_i\in\mathcal{I}_m}\delta^{\text{co}}_{t_i}\left[\int_{t'\in\mathcal{I}_m}d^{\text{co}}_{t_i}(t')f(t')dt'\right]f(t_i)dt_i-\delta^{\text{co}} \qquad (\mathbb{E}_{W,\varepsilon,t}\left[\tfrac{(W_{t_i}-\pi)W_{t_i}}{\pi(1-\pi)}\right]=1)\\
=&\sum_{m=1}^{M}I^{(m)}-\delta^{\text{co}} \qquad \text{(by definition of } I^{(m)})
\end{aligned}
$$

We then finish the proof of Lemma B.7. □

LEMMA B.8 **(Effects from simultaneous interventions)**. *Under the assumptions in Theorem 4.1, the bias from simultaneous interventions is*

$$
\begin{aligned}
\mathbb{E}_{W,\varepsilon,t}\left[\mathcal{E}_{\text{simul}}\right]=&\mathbb{E}_{W,\varepsilon,t}\left[\frac{1}{n}\sum_{i=1}^{n}\alpha_{t_i}\left(Y_{t_i}(\boldsymbol{W},\boldsymbol{W}^{\text{s}}_1,\cdots,\boldsymbol{W}^{\text{s}}_K)-Y_{t_i}(\boldsymbol{W},\mathbf{0},\cdots,\mathbf{0})\right)\right]\\
=&\sum_{m=1}^{M}\int_{t\in\mathcal{I}_m}\Phi^{\text{simul}}_t f(t)dt\,.
\end{aligned}
$$

Proof of Lemma B.8 As events are sampled i.i.d. from $f(t)$, we have

$$
\begin{aligned}
&\mathbb{E}_{W,\varepsilon,t}\left[\frac{1}{n}\sum_{i=1}^{n}\alpha_{t_i}\left(Y_{t_i}(\boldsymbol{W},\boldsymbol{W}^{\text{s}}_1,\cdots,\boldsymbol{W}^{\text{s}}_K)-Y_{t_i}(\boldsymbol{W},\mathbf{0},\cdots,\mathbf{0})\right)\right]\\
=&\mathbb{E}_{W,t}\left[\alpha_{t_i}\left(Y_{t_i}(\boldsymbol{W},\boldsymbol{W}^{\text{s}}_1,\cdots,\boldsymbol{W}^{\text{s}}_K)-Y_{t_i}(\boldsymbol{W},\mathbf{0},\cdots,\mathbf{0})\right)\right]\\
&\qquad\text{(marketplace potential outcomes do not depend on } \varepsilon^{(i)})\\
=&\mathbb{E}_{W,t}\left[\alpha_{t_i}\mathbb{E}_{\boldsymbol{W}^{\text{s}}_1,\cdots,\boldsymbol{W}^{\text{s}}_K}\left[Y_{t_i}(\boldsymbol{W},\boldsymbol{W}^{\text{s}}_1,\cdots,\boldsymbol{W}^{\text{s}}_K)-Y_{t_i}(\boldsymbol{W},\mathbf{0},\cdots,\mathbf{0})\mid\boldsymbol{W},t_i\right]\right] \quad \text{(by the law of total expectation)}\\
=&\mathbb{E}_{W,t}\left[\alpha_{t_i}\delta^{\text{simul}}_{t_i}(\boldsymbol{W})\right] \qquad \text{(by definition of } \delta^{\text{simul}}_{t_i}(\boldsymbol{W}))\\
=&\sum_{m=1}^{M}\int_{t\in\mathcal{I}_m}\Phi^{\text{simul}}_t f(t)dt\\
&\qquad\text{(first take the expected value over } \boldsymbol{W}^{(-m)} \text{ and then take the expected value over } W^{(m)})
\end{aligned}
$$

where Φ^{simul}_t is defined as

$$
\Phi^{\text{simul}}_t=\mathbb{E}_{\boldsymbol{W}^{(-m)}}\left[\delta^{\dagger}_t(\boldsymbol{W}^{(-m)},W^{(m)}=1)\right]-\mathbb{E}_{\boldsymbol{W}^{(-m)}}\left[\delta^{\dagger}_t(\boldsymbol{W}^{(-m)},W^{(m)}=0)\right].
$$

We then finish the proof of Lemma B.8. □

Proof of Theorem 4.1 Based on the decomposition of the estimation error of $\hat{\delta}^{\text{gate}}$ in Equation (B.6), the bias of $\hat{\delta}^{\text{gate}}$ equals to

$$\begin{aligned}\mathbb{E}_{W,\varepsilon,t}\left[\hat{\delta}^{\text{gate}}-\delta^{\text{gate}}\right] =& \mathbb{E}_{W,\varepsilon,t}\left[\mathcal{E}_{\text{carryover}}\right]+\mathbb{E}_{W,\varepsilon,t}\left[\mathcal{E}_{\text{simul}}\right] \\ =& \underbrace{\delta^{\text{co}}\left[\sum_{m=1}^{M} I^{(m)}-1\right]}_{\text{Bias}_\ell(\text{carryover})}+\underbrace{\sum_{m=1}^{M}\int_{t\in\mathcal{I}_m}\Phi_{t_i}^{\text{simul}}f(t)dt}_{\text{Bias}(\mathcal{E}_{\text{simul}})}\end{aligned}$$

following Lemmas B.6, B.7, and B.8. □

B.3. Proof of Theorem 4.2

We can decompose the mean-squared error of $\hat{\delta}^{\text{gate}}$ as follows.

$$\begin{aligned}\mathbb{E}_{W,\varepsilon,t}&\left[\left(\hat{\delta}^{\text{gate}}-\delta^{\text{gate}}\right)^2\right]=\mathbb{E}_{W,\varepsilon,t}\left[\left(\mathcal{E}_{\text{meas}}+\mathcal{E}_{\text{inst}}+\mathcal{E}_{\text{carryover}}+\mathcal{E}_{\text{simul}}\right)^2\right] \\ =&\mathbb{E}_{W,\varepsilon,t}\left[\left(\mathcal{E}_{\text{meas}}\right)^2\right]+2\mathbb{E}_{W,\varepsilon,t}\left[\mathcal{E}_{\text{meas}}\left(\mathcal{E}_{\text{inst}}+\mathcal{E}_{\text{carryover}}+\mathcal{E}_{\text{simul}}\right)\right] \\ &+\mathbb{E}_{W,\varepsilon,t}\left[\left(\mathcal{E}_{\text{inst}}\right)^2\right]+\mathbb{E}_{W,\varepsilon,t}\left[\left(\mathcal{E}_{\text{carryover}}\right)^2\right]+2\mathbb{E}_{W,\varepsilon,t}\left[\mathcal{E}_{\text{carryover}}\mathcal{E}_{\text{inst}}\right] \\ &+\mathbb{E}_{W,\varepsilon,t}\left[\left(\mathcal{E}_{\text{simul}}\right)^2\right]+2\mathbb{E}_{W,\varepsilon,t}\left[\mathcal{E}_{\text{simul}}\mathcal{E}_{\text{inst}}\right]+2\mathbb{E}_{W,\varepsilon,t}\left[\mathcal{E}_{\text{simul}}\mathcal{E}_{\text{carryover}}\right]\end{aligned}$$

Below we show the value of each term in the decomposition separately.

We first introduce a few more notations to measure the heterogeneity in treatment effects, which will be used in showing the value of the three terms that involve $\mathcal{E}_{\text{carryover}}$. Let

$$\Xi^{\text{inst},(m)}=\int_{t_i\in\mathcal{I}_m}(\delta_{t_i}^{\text{inst}}-\delta^{\text{inst}})f(t_i)dt_i$$

measure the discrepancy between the heterogeneous instantaneous effect $\delta_{t_i}^{\text{inst}}$ and average instantaneous effect δ^{inst} for times t_i in the interval $\mathcal{I}_m$. Analogously, let

$$\Xi^{\text{co},(m)}=\int_{t_i\in\mathcal{I}_m}(\delta_{t_i}^{\text{co}}-\delta^{\text{co}})f(t_i)dt_i$$

measure the discrepancy between the heterogeneous carryover effect $\delta_{t_i}^{\text{inst}}$ and average carryover effect δ^{inst} for times t_i in the interval $\mathcal{I}_m$. Note that $\sum_{m=1}^{M}\Xi^{\text{inst},(m)}=0$ and $\sum_{m=1}^{M}\Xi^{\text{co},(m)}=0$ by the definition of δ^{inst} and δ^{co}. Furthermore,

$$\tilde{\Xi}^{\text{co},(m)}=\int_{t_i\in\mathcal{I}_m}\delta_{t_i}^{\text{co}}f(t_i)dt_i=\Xi^{\text{co},(m)}+\delta^{\text{co}}\mu^{(m)}\,.$$

Lemma B.9 **(Second moment of measurement errors and constant term)**. *Under the assumptions in Theorem 4.2, the second moment of the measurement error is*

$$\mathbb{E}_{W,\varepsilon,t}\left[\left(\mathcal{E}_{\text{meas}}\right)^2\right]=\mathbb{E}_{W,\varepsilon,t}\left[\left(\frac{1}{n}\sum_{i=1}^{n}\alpha_{t_i}\varepsilon^{(i)}\right)^2\right]=\frac{4}{n}\sum_{m=1}^{M}\left(V^{(m)}+(n-1)C^{(m)}\right)\,.$$

Proof of Lemma B.9 The second moment of $\mathcal{E}_{\text{meas}}$ equals

$$\begin{aligned}\mathbb{E}_{W,\varepsilon,t}&\left[\left(\mathcal{E}_{\text{meas}}\right)^2\right]=\frac{1}{n^2}\sum_{i,j}\mathbb{E}_{W,\varepsilon,t}\left[\alpha_{t_i}\alpha_{t_j}\varepsilon^{(i)}\varepsilon^{(j)}\right] \\ =&\frac{1}{n}\mathbb{E}_{W,t}\left[\alpha_{t_i}^2\mathbb{E}_{\varepsilon}\left[\left(\varepsilon^{(i)}\right)^2\mid\boldsymbol{W},\boldsymbol{W}_1^{\text{s}},\cdots,\boldsymbol{W}_K^{\text{s}},t_i\right]\right]+\frac{n-1}{n}\mathbb{E}_{W,t}\left[\mathbb{E}_{\varepsilon}\left[\varepsilon^{(i)}\varepsilon^{(j)}\mid\boldsymbol{W},\boldsymbol{W}_1^{\text{s}},\cdots,\boldsymbol{W}_K^{\text{s}},t_i,t_j\right]\right]\end{aligned}$$

$$
\begin{aligned}
=&\frac{1}{n\pi(1-\pi)}\sum_{m=1}^{M}\underbrace{\int_{t_i\in\mathcal{I}_m}\mathbb{E}_\varepsilon\left[(\varepsilon^{(i)})^2\mid t_i\right]f(t_i)dt_i}_{V^{(m)}}\\
&+\frac{n-1}{n\pi(1-\pi)}\sum_{m=1}^{M}\underbrace{\int_{t_i,t_j\in\mathcal{I}_m}\mathbb{E}_\varepsilon\left[\varepsilon^{(i)}\varepsilon^{(j)}\mid t_i,t_j\right]f(t_i)f(t_j)dt_idt_j}_{C^{(m)}}
\end{aligned}
$$

where we use the following property to show the expression of the second term in the last equation

$$
\begin{aligned}
\mathbb{E}_W\left[\alpha_{t_i}\alpha_{t_j}\mid t_i,t_j\right]=&\mathbb{E}_W\left[\frac{(W_{t_i}-\pi)(W_{t_j}-\pi)}{\pi^2(1-\pi)^2}\mid t_i,t_j\right]\\
=&\begin{cases}\frac{1}{\pi(1-\pi)} & t_i \text{ and } t_j \text{ in the same interval}\\ 0 & \text{otherwise.}\end{cases}
\end{aligned}
$$

Setting π as $1/2$, we have

$$
\mathbb{E}_{W,\varepsilon,t}\left[(\mathcal{E}_{\text{meas}})^2\right]=\frac{4}{n}\sum_{m=1}^{M}\left(V^{(m)}+(n-1)C^{(m)}\right).
$$

We then finish the proof of Lemma B.9. $\square$

Lemma B.10 **(Expected product of $\mathcal{E}_{\text{meas}}$ and $\mathcal{E}_{\text{inst}}+\mathcal{E}_{\text{carryover}}+\mathcal{E}_{\text{simul}}$).** *Under the assumptions in Theorem 4.2, the expected product of $\mathcal{E}_{\text{meas}}$ and $\mathcal{E}_{\text{inst}}+\mathcal{E}_{\text{carryover}}+\mathcal{E}_{\text{simul}}$ is equal to*

$$
\mathbb{E}_{W,\varepsilon,t}\left[\mathcal{E}_{\text{meas}}\left(\mathcal{E}_{\text{inst}}+\mathcal{E}_{\text{carryover}}+\mathcal{E}_{\text{simul}}\right)\right]=0\,.
$$

Proof of Lemma B.10 First, for the expected value of the product of $\mathcal{E}_{\text{meas}}$ and $\mathcal{E}_{\text{inst}}$, we have

$$
\begin{aligned}
\mathbb{E}_{W,\varepsilon,t}\left[\mathcal{E}_{\text{meas}}\mathcal{E}_{\text{inst}}\right]=&\mathbb{E}_{W,\varepsilon,t}\left[\left(\frac{1}{n}\sum_{i=1}^{n}\alpha_{t_i}\varepsilon^{(i)}\right)\left(\delta^{\text{gate}}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{W_{t_i}}{\pi}-1\right]\right)\right]\\
=&\mathbb{E}_{W,\varepsilon,t}\left[\left(\frac{1}{n}\sum_{i=1}^{n}\alpha_{t_i}\mathbb{E}_\varepsilon\left[\varepsilon^{(i)}\mid \boldsymbol{W},\boldsymbol{W}_1^{\text{s}},\cdots,\boldsymbol{W}_K^{\text{s}},t_i\right]\right)\left(\delta^{\text{gate}}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{W_{t_i}}{\pi}-1\right]\right)\right]\\
=&0\,.
\end{aligned}
$$

$$
\left(\mathbb{E}_\varepsilon\left[\varepsilon^{(i)}\mid \boldsymbol{W},\boldsymbol{W}_1^{\text{s}},\cdots,\boldsymbol{W}_K^{\text{s}},t_i\right]=0 \text{ for all } i\right)
$$

Second, the expected value of the product of $\mathcal{E}_{\text{meas}}$ and $\mathcal{E}_{\text{carryover}}$ is equal to

$$
\begin{aligned}
&\mathbb{E}_{W,\varepsilon,t}\left[\mathcal{E}_{\text{meas}}\mathcal{E}_{\text{carryover}}\right]\\
=&\mathbb{E}_{W,\varepsilon,t}\left[\left(\frac{1}{n}\sum_{i=1}^{n}\alpha_{t_i}\mathbb{E}_\varepsilon\left[\varepsilon^{(i)}\mid \boldsymbol{W},\boldsymbol{W}_1^{\text{s}},\cdots,\boldsymbol{W}_K^{\text{s}},t_i\right]\right)\times\right.\\
&\qquad\left.\left(\frac{1}{n}\sum_{i=1}^{n}\alpha_{t_i}\left(Y_{t_i}(\boldsymbol{W},\boldsymbol{0},\cdots,\boldsymbol{0})-W_{t_i}\delta^{\text{gate}}\right)\right)\right]\\
=&0\,.
\end{aligned}
$$

Third, the expected value of the product of $\mathcal{E}_{\text{meas}}$ and $\mathcal{E}_{\text{simul}}$ is equal to

$$
\begin{aligned}
&\mathbb{E}_{W,\varepsilon,t}\left[\mathcal{E}_{\text{meas}}\mathcal{E}_{\text{simul}}\right]\\
=&\mathbb{E}_{W,\varepsilon,t}\left[\left(\frac{1}{n}\sum_{i=1}^{n}\alpha_{t_i}\mathbb{E}_\varepsilon\left[\varepsilon^{(i)}\mid \boldsymbol{W},\boldsymbol{W}_1^{\text{s}},\cdots,\boldsymbol{W}_K^{\text{s}},t_i\right]\right)\cdot\right.\\
&\left.\left(\frac{1}{n}\sum_{i=1}^{n}\alpha_{t_i}\left(Y_{t_i}(\boldsymbol{W},\boldsymbol{W}_1^{\text{s}},\cdots,\boldsymbol{W}_K^{\text{s}})-Y_{t_i}(\boldsymbol{W},\boldsymbol{0},\cdots,\boldsymbol{0})\right)\right)\right]=0\,.
\end{aligned}
$$

Then we have

$$\begin{aligned}&\mathbb{E}_{W,\varepsilon,t}\left[\mathcal{E}_{\text{meas}}\left(\mathcal{E}_{\text{inst}}+\mathcal{E}_{\text{carryover}}+\mathcal{E}_{\text{simul}}\right)\right]\\ =\ &\mathbb{E}_{W,\varepsilon,t}\left[\mathcal{E}_{\text{meas}}\mathcal{E}_{\text{inst}}\right]+\mathbb{E}_{W,\varepsilon,t}\left[\mathcal{E}_{\text{meas}}\mathcal{E}_{\text{carryover}}\right]+\mathbb{E}_{W,\varepsilon,t}\left[\mathcal{E}_{\text{meas}}\mathcal{E}_{\text{simul}}\right]=0\,.\end{aligned}$$

We then finish the proof of Lemma B.10. □

LEMMA B.11 **(Second moment of instantaneous effects)**. *Under the assumptions in Theorem 4.1, the second moment of $\mathcal{E}_{\text{inst}}$ is equal to*

$$\mathbb{E}_{W,\varepsilon,t}\left[(\mathcal{E}_{\text{inst}})^2\right]=(\delta^{\text{gate}})^2\sum_{m=1}^{M}\left[\mu^{(m)}\right]^2\,.$$

Proof of Lemma B.11 The second moment of $\mathcal{E}_{\text{inst}}$ equals

$$\begin{aligned}\mathbb{E}_{W,\varepsilon,t}\left[(\mathcal{E}_{\text{inst}})^2\right]=&\frac{(\delta^{\text{gate}})^2}{n^2}\sum_{i,j}\mathbb{E}_{W,t}\left[\left(\frac{W_{t_i}}{\pi}-1\right)\left(\frac{W_{t_j}}{\pi}-1\right)\right]\\ =&(\delta^{\text{gate}})^2\mathbb{E}_{W,t}\left[\left(\frac{W_{t_i}}{\pi}-1\right)\left(\frac{W_{t_j}}{\pi}-1\right)\right]\\ =&(\delta^{\text{gate}})^2\left(\frac{1}{\pi}-1\right)\sum_{m=1}^{M}\underbrace{\int_{t_i,t_j\in\mathcal{I}_m}f(t_i)f(t_j)dt_idt_j}_{\left[\mu^{(m)}\right]^2}\,,\end{aligned}$$

where we use the following property to show the last equation

$$\mathbb{E}_W\left[\left(\frac{W_{t_i}}{\pi}-1\right)\left(\frac{W_{t_j}}{\pi}-1\right)\mid t_i,t_j\right]=\begin{cases}\frac{\pi-1}{\pi} & t_i\text{ and }t_j\text{ in the same interval}\\ 0 & \text{otherwise.}\end{cases}$$

Setting $\pi=1/2$, we have

$$\mathbb{E}_{W,\varepsilon,t}\left[(\mathcal{E}_{\text{inst}})^2\right]=(\delta^{\text{gate}})^2\sum_{m=1}^{M}\left[\mu^{(m)}\right]^2\,.$$

We then finish the proof of Lemma B.11. □

LEMMA B.12 **(Second moment of carryover effects)**. *Under the assumptions in Theorem 4.2, the second moment of $\mathcal{E}_{\text{carryover}}$ equals to*

$$\begin{aligned}&\mathbb{E}_{W,\varepsilon,t}\left[(\mathcal{E}_{\text{carryover}})^2\right]\\ =&\sum_{m=1}^{M}\left(\Xi^{\text{inst},(m)}+\Xi^{\text{co},(m)}+2\mu^{(m)}_{Y^{\text{ctrl}}}\right)^2+\left(\sum_{m=1}^{M}I^{(m)}-\delta^{\text{co}}\right)^2+\sum_{m=1}^{M}\sum_{m'\neq m}\left(\left[I^{(m,m')}\right]^2+I^{(m,m')}I^{(m',m)}\right)\,.\end{aligned}$$

Proof of Lemma B.12 As events are sampled i.i.d. from distribution $f(t)$, we have

$$\begin{aligned}&\mathbb{E}_{W,\varepsilon,t}\left[(\mathcal{E}_{\text{carryover}})^2\right]\\ =&\mathbb{E}_{W,t}\left[\left(\frac{1}{n}\sum_{i=1}^{n}\left[\alpha_{t_i}(\delta^{\text{inst}}_{t_i}-\delta^{\text{inst}})W_{t_i}+\alpha_{t_i}(\delta^{\text{co}}_{t_i}(\boldsymbol{W})-\delta^{\text{co}}W_{t_i})\right]+\frac{1}{n}\sum_{i=1}^{n}\alpha_{t_i}Y_{t_i}(\mathbf{0},\cdots,\mathbf{0})\right)^2\right]\\ &\qquad\text{(carryover effects do not depend on }\varepsilon\text{)}\\ =&\underbrace{\mathbb{E}_{W,t}\left[\left(\frac{1}{n}\sum_{i=1}^{n}\left[\alpha_{t_i}(\delta^{\text{inst}}_{t_i}-\delta^{\text{inst}})W_{t_i}+\alpha_{t_i}(\delta^{\text{co}}_{t_i}(\boldsymbol{W})-\delta^{\text{co}}W_{t_i})\right]\right)^2\right]}_{A^{\text{treat}}}\end{aligned}$$

$$+ 2\underbrace{\mathbb{E}_{W,t}\left[\left(\frac{1}{n}\sum_{i=1}^{n}\left[\alpha_{t_i}(\delta^{\text{inst}}_{t_i}-\delta^{\text{inst}})W_{t_i}+\alpha_{t_i}(\delta^{\text{co}}_{t_i}(\boldsymbol{W})-\delta^{\text{co}}W_{t_i})\right]\right)\left(\frac{1}{n}\sum_{i=1}^{n}\alpha_{t_i}Y_{t_i}(\mathbf{0},\cdots,\mathbf{0})\right)\right]}_{A^{\text{cross}}}$$

$$+\underbrace{\mathbb{E}_{W,t}\left[\left(\frac{1}{n}\sum_{i=1}^{n}\alpha_{t_i}Y_{t_i}(\mathbf{0},\cdots,\mathbf{0})\right)^2\right]}_{A^{\text{control}}}$$

Let us first consider the simplest term A^{control}. For this term, we have

$$\begin{aligned}
A^{\text{control}} =& \frac{1}{n^2}\sum_{i,j}\mathbb{E}_{W,t}\left[\alpha_{t_i}Y_{t_i}(\mathbf{0},\cdots,\mathbf{0})\cdot\alpha_{t_j}Y_{t_j}(\mathbf{0},\cdots,\mathbf{0})\right]\\
=&\mathbb{E}_{W,t}\left[\alpha_{t_i}Y_{t_i}(\mathbf{0},\cdots,\mathbf{0})\cdot\alpha_{t_j}Y_{t_j}(\mathbf{0},\cdots,\mathbf{0})\right]\\
=&\sum_{m=1}^{M}\int_{t_i,t_j\in\mathcal{I}_m}\mathbb{E}_W\left[\alpha_{t_i}\alpha_{t_j}\mid t_i,t_j\right]Y_{t_i}(\mathbf{0},\cdots,\mathbf{0})Y_{t_j}(\mathbf{0},\cdots,\mathbf{0})f(t_i)f(t_j)dt_idt_j\\
&+\sum_{m=1}^{M}\sum_{m':m'\neq m}\int_{t_i\in\mathcal{I}_m,t_j\in\mathcal{I}_{m'}}\mathbb{E}_W\left[\alpha_{t_i}\alpha_{t_j}\mid t_i,t_j\right]Y_{t_i}(\mathbf{0},\cdots,\mathbf{0})Y_{t_j}(\mathbf{0},\cdots,\mathbf{0})f(t_i)f(t_j)dt_idt_j\\
=&\frac{1}{\pi(1-\pi)}\sum_{m=1}^{M}\int_{t_i,t_j\in\mathcal{I}_m}Y_{t_i}(\mathbf{0},\cdots,\mathbf{0})Y_{t_j}(\mathbf{0},\cdots,\mathbf{0})f(t_i)f(t_j)dt_idt_j\\
=&\frac{1}{\pi(1-\pi)}\sum_{m=1}^{M}\left(\int_{t_i\in\mathcal{I}_m}Y_{t_i}(\mathbf{0},\cdots,\mathbf{0})f(t_i)dt_i\right)^2=\frac{1}{\pi(1-\pi)}\sum_{m=1}^{M}\left[\mu^{(m)}_{Y^{\text{ctrl}}}\right]^2
\end{aligned}$$

following the definition that $\mu^{(m)}_{Y^{\text{ctrl}}}=\int_{t_i\in\mathcal{I}_m}Y_{t_i}(\mathbf{0},\cdots,\mathbf{0})f(t_i)dt_i$.

Next we consider the cross term A^{cross}. For this term, we have

$$\begin{aligned}
A^{\text{cross}} =& \frac{1}{n^2}\sum_{i,j}\mathbb{E}_{W,t}\left[\left(\alpha_{t_i}(\delta^{\text{inst}}_{t_i}-\delta^{\text{inst}})W_{t_i}+\alpha_{t_i}(\delta^{\text{co}}_{t_i}(\boldsymbol{W})-\delta^{\text{co}}W_{t_i})\right)\alpha_{t_j}Y_{t_j}(\mathbf{0},\cdots,\mathbf{0})\right]\\
=&\underbrace{\mathbb{E}_{W,t}\left[\alpha_{t_i}(\delta^{\text{inst}}_{t_i}-\delta^{\text{inst}})W_{t_i}\cdot\alpha_{t_j}Y_{t_j}(\mathbf{0},\cdots,\mathbf{0})\right]}_{B_{1,ij}}\\
&+\underbrace{\mathbb{E}_{W,t}\left[\alpha_{t_i}(\delta^{\text{co}}_{t_i}(\boldsymbol{W})-\delta^{\text{co}}W_{t_i})\cdot\alpha_{t_j}Y_{t_j}(\mathbf{0},\cdots,\mathbf{0})\right]}_{B_{2,ij}}
\end{aligned}$$

For $B_{1,ij}$, we have

$$\begin{aligned}
B_{1,ij} =&\mathbb{E}_{W,t}\left[\frac{W_{t_i}}{\pi}\frac{W_{t_j}-\pi}{\pi(1-\pi)}(\delta^{\text{inst}}_{t_i}-\delta^{\text{inst}})\cdot Y_{t_j}(\mathbf{0},\cdots,\mathbf{0})\right]\\
=&\sum_{m=1}^{M}\int_{t_i,t_j\in\mathcal{I}_m}\mathbb{E}_W\left[\frac{W_{t_i}}{\pi}\frac{W_{t_j}-\pi}{\pi(1-\pi)}\mid t_i,t_j\right](\delta^{\text{inst}}_{t_i}-\delta^{\text{inst}})\cdot Y_{t_j}(\mathbf{0},\cdots,\mathbf{0})f(t_i)f(t_j)dt_idt_j\\
&+\sum_{m=1}^{M}\sum_{m':m'\neq m}\int_{t_i\in\mathcal{I}_m,t_j\in\mathcal{I}_{m'}}\underbrace{\mathbb{E}_W\left[\frac{W_{t_i}}{\pi}\frac{W_{t_j}-\pi}{\pi(1-\pi)}\mid t_i,t_j\right]}_{=0}\times\\
&\qquad(\delta^{\text{inst}}_{t_i}-\delta^{\text{inst}})\cdot Y_{t_j}(\mathbf{0},\cdots,\mathbf{0})f(t_i)f(t_j)dt_idt_j\\
=&\frac{1}{\pi}\sum_{m=1}^{M}\int_{t_i,t_j\in\mathcal{I}_m}(\delta^{\text{inst}}_{t_i}-\delta^{\text{inst}})\cdot Y_{t_j}(\mathbf{0},\cdots,\mathbf{0})f(t_i)f(t_j)dt_idt_j\\
=&\frac{1}{\pi}\sum_{m=1}^{M}\Xi^{\text{inst},(m)}\mu^{(m)}_{Y^{\text{ctrl}}}.
\end{aligned}$$

For $B_{2,ij}$, we have

$$B_{2,ij} = \underbrace{\mathbb{E}_{W,t}\left[\alpha_{t_i}\delta^{\mathrm{co}}_{t_i}(\boldsymbol{W})\cdot\alpha_{t_j}Y_{t_j}(\mathbf{0},\cdots,\mathbf{0})\right]}_{C_{1,ij}} - \underbrace{\delta^{\mathrm{co}}\mathbb{E}_{W,t}\left[\frac{W_{t_i}}{\pi}\alpha_{t_j}Y_{t_j}(\mathbf{0},\cdots,\mathbf{0})\right]}_{C_{2,ij}}.$$

$C_{1,ij}$ equals to

$$\begin{aligned}
C_{1,ij} =& \sum_{m,m'=1}^{M}\int_{t_i\in\mathcal{I}_m,t_j\in\mathcal{I}_{m'}}\mathbb{E}_W\left[\frac{W_{t_i}-\pi}{\pi(1-\pi)}\frac{W_{t_j}-\pi}{\pi(1-\pi)}\left(\sum_{k=1}^{M}W^{(k)}\int_{t'\in\mathcal{I}_k}d^{\mathrm{co}}_{t_i}(t')f(t')dt'\right)\mid t_i,t_j\right]\\
&\qquad\times\delta^{\mathrm{co}}_{t_i}Y_{t_j}(\mathbf{0},\cdots,\mathbf{0})f(t_i)f(t_j)dt_idt_j\\
=&\frac{1}{\pi}\sum_{m=1}^{M}\int_{t_i,t_j\in\mathcal{I}_m}\left[\int_{t'\in\mathcal{I}_m}d^{\mathrm{co}}_{t_i}(t')f(t')dt'\right]\delta^{\mathrm{co}}_{t_i}Y_{t_j}(\mathbf{0},\cdots,\mathbf{0})f(t_i)f(t_j)dt_idt_j\\
&+\frac{1}{1-\pi}\sum_{m=1}^{M}\int_{t_i,t_j\in\mathcal{I}_m}\left[\sum_{k:k\neq m}\int_{t'\in\mathcal{I}_k}d^{\mathrm{co}}_{t_i}(t')f(t')dt'\right]\delta^{\mathrm{co}}_{t_i}Y_{t_j}(\mathbf{0},\cdots,\mathbf{0})f(t_i)f(t_j)dt_idt_j\\
=&2\sum_{m=1}^{M}\int_{t_i,t_j\in\mathcal{I}_m}\delta^{\mathrm{co}}_{t_i}Y_{t_j}(\mathbf{0},\cdots,\mathbf{0})f(t_i)f(t_j)dt_idt_j\\
=&2\sum_{m=1}^{M}\tilde{\Xi}^{\mathrm{co},(m)}\mu^{(m)}_{Y^{\mathrm{ctrl}}}.
\end{aligned}$$

$C_{2,ij}$ equals to

$$\begin{aligned}
C_{2,ij} =&\delta^{\mathrm{co}}\mathbb{E}_{W,t}\left[\frac{W_{t_i}}{\pi}\alpha_{t_j}Y_{t_j}(\mathbf{0},\cdots,\mathbf{0})\right]\\
=&\delta^{\mathrm{co}}\sum_{m=1}^{M}\int_{t_i,t_j\in\mathcal{I}_m}\underbrace{\mathbb{E}_W\left[\frac{W_{t_i}}{\pi}\alpha_{t_j}\mid t_i,t_j\right]}_{=1/\pi}Y_{t_j}(\mathbf{0},\cdots,\mathbf{0})f(t_i)f(t_j)dt_idt_j\\
&+\delta^{\mathrm{co}}\sum_{m=1}^{M}\sum_{m':m'\neq m}\int_{t_i\in\mathcal{I}_m,t_j\in\mathcal{I}_{m'}}\underbrace{\mathbb{E}_W\left[\frac{W_{t_i}}{\pi}\alpha_{t_j}\mid t_i,t_j\right]}_{=0}Y_{t_j}(\mathbf{0},\cdots,\mathbf{0})f(t_i)f(t_j)dt_idt_j\\
=&\frac{\delta^{\mathrm{co}}}{\pi}\sum_{m=1}^{M}\int_{t_i,t_j\in\mathcal{I}_m}Y_{t_j}(\mathbf{0},\cdots,\mathbf{0})f(t_i)f(t_j)dt_idt_j\\
=&2\delta^{\mathrm{co}}\sum_{m=1}^{M}\mu^{(m)}\mu^{(m)}_{Y^{\mathrm{ctrl}}}.
\end{aligned}$$

Therefore,

$$B_{2,ij} =2\sum_{m=1}^{M}\left(\tilde{\Xi}^{\mathrm{co},(m)}-\delta^{\mathrm{co}}\mu^{(m)}\right)\mu^{(m)}_{Y^{\mathrm{ctrl}}} = 2\sum_{m=1}^{M}\Xi^{\mathrm{co},(m)}\mu^{(m)}_{Y^{\mathrm{ctrl}}}.$$

A^{cross} is then equal to

$$A^{\mathrm{cross}} =2\sum_{m=1}^{M}\Xi^{\mathrm{inst},(m)}\mu^{(m)}_{Y^{\mathrm{ctrl}}} + 2\sum_{m=1}^{M}\Xi^{\mathrm{co},(m)}\mu^{(m)}_{Y^{\mathrm{ctrl}}} = 2\sum_{m=1}^{M}\Xi^{(m)}\mu^{(m)}_{Y^{\mathrm{ctrl}}}.$$

Last, we consider the term about treatment effect, A^{treat}. For this term, we have

$$A^{\mathrm{treat}} =\frac{1}{n^2}\sum_{i,j}\underbrace{\mathbb{E}_{W,t}\left[\alpha_{t_i}(\delta^{\mathrm{inst}}_{t_i}-\delta^{\mathrm{inst}})W_{t_i}\cdot\alpha_{t_j}(\delta^{\mathrm{inst}}_{t_j}-\delta^{\mathrm{inst}})W_{t_j}\right]}_{:=A_{1,ij}}$$

$$+\frac{2}{n^2}\sum_{i,j}\underbrace{\mathbb{E}_{W,t}\left[\alpha_{t_i}(\delta^{\text{inst}}_{t_i}-\delta^{\text{inst}})W_{t_i}\cdot\alpha_{t_j}(\delta^{\text{co}}_{t_j}(\boldsymbol{W})-\delta^{\text{co}}W_{t_j})\right]}_{:=A_{2,ij}}$$

$$+\frac{1}{n^2}\sum_{i,j}\underbrace{\mathbb{E}_{W,t}\left[\alpha_{t_i}(\delta^{\text{co}}_{t_i}(\boldsymbol{W})-\delta^{\text{co}}W_{t_i})\alpha_{t_j}(\delta^{\text{co}}_{t_j}(\boldsymbol{W})-\delta^{\text{co}}W_{t_j})\right]}_{:=A_{3,ij}}.$$

Below we compute each of $A_{1,ij}$, $A_{2,ij}$, and $A_{3,ij}$. We first compute $A_{1,ij}$.

$$\begin{aligned}A_{1,ij}=&\mathbb{E}_{W,t}\left[\frac{W_{t_i}}{\pi}\frac{W_{t_j}}{\pi}(\delta^{\text{inst}}_{t_i}-\delta^{\text{inst}})(\delta^{\text{inst}}_{t_j}-\delta^{\text{inst}})\right]\\ =&\int_{t_i,t_j\in[0,T]}(\delta^{\text{inst}}_{t_i}-\delta^{\text{inst}})(\delta^{\text{inst}}_{t_j}-\delta^{\text{inst}})f(t_i)f(t_j)dt_idt_j\\ &+\left(\frac{1}{\pi}-1\right)\sum_{m=1}^{M}\int_{t_i,t_j\in\mathcal{I}_m}(\delta^{\text{inst}}_{t_i}-\delta^{\text{inst}})(\delta^{\text{inst}}_{t_j}-\delta^{\text{inst}})f(t_i)f(t_j)dt_idt_j\\ =&\sum_{m=1}^{M}\int_{t_i,t_j\in\mathcal{I}_m}(\delta^{\text{inst}}_{t_i}-\delta^{\text{inst}})(\delta^{\text{inst}}_{t_j}-\delta^{\text{inst}})f(t_i)f(t_j)dt_idt_j=\sum_{m=1}^{M}\left[\Xi^{\text{inst},(m)}\right]^2.\end{aligned}$$

(the first term is zero following that $\mathbb{E}_t[\delta^{\text{inst}}_{t_j}]=\delta^{\text{inst}}$ and $\pi=1/2$)

We then compute $A_{2,ij}$.

$$A_{2,ij}=\underbrace{\mathbb{E}_{W,t}\left[\frac{W_{t_i}}{\pi}\frac{W_{t_j}-\pi}{\pi(1-\pi)}(\delta^{\text{inst}}_{t_i}-\delta^{\text{inst}})\delta^{\text{co}}_{t_j}(\boldsymbol{W})\right]}_{:=B_{1,ij}}-\underbrace{\delta^{\text{co}}\mathbb{E}_{W,t}\left[\frac{W_{t_i}}{\pi}\frac{W_{t_j}}{\pi}(\delta^{\text{inst}}_{t_j}-\delta^{\text{inst}})\right]}_{:=B_{2,ij}}.$$

For $B_{1,ij}$, we have

$$\begin{aligned}B_{1,ij}=&\sum_{m=1}^{M}\int_{t_i,t_j\in\mathcal{I}_m}(\delta^{\text{inst}}_{t_i}-\delta^{\text{inst}})\delta^{\text{co}}_{t_j}\times\\ &\mathbb{E}_W\left[\frac{W_{t_i}}{\pi}\frac{W_{t_j}-\pi}{\pi(1-\pi)}\left(\sum_{k=1}^{M}W^{(k)}\int_{t'\in\mathcal{I}_k}d^{\text{co}}_{t_j}(t')f(t')dt'\right)\mid t_i,t_j\right]f(t_i)f(t_j)dt_idt_j\\ &+\sum_{m=1}^{M}\sum_{m':m'\neq m}\int_{t_i\in\mathcal{I}_m,t_j\in\mathcal{I}_{m'}}(\delta^{\text{inst}}_{t_i}-\delta^{\text{inst}})\delta^{\text{co}}_{t_j}\times\\ &\mathbb{E}_W\left[\frac{W_{t_i}}{\pi}\frac{W_{t_j}-\pi}{\pi(1-\pi)}\left(\sum_{k=1}^{M}W^{(k)}\int_{t'\in\mathcal{I}_k}d^{\text{co}}_{t_j}(t')f(t')dt'\right)\mid t_i,t_j\right]f(t_i)f(t_j)dt_idt_j\\ =&\sum_{m=1}^{M}\underbrace{\left(\int_{t_i\in\mathcal{I}_m}(\delta^{\text{inst}}_{t_i}-\delta^{\text{inst}})f(t_i)dt_i\right)}_{\Xi^{\text{inst},(m)}}\underbrace{\left(\int_{t_j\in\mathcal{I}_m}\delta^{\text{co}}_{t_j}\left(\sum_{k=1}^{M}\int_{t'\in\mathcal{I}_k}d^{\text{co}}_{t_j}(t')f(t')dt'\right)f(t_j)dt_j\right)}_{\tilde{\Xi}^{\text{co},(m)}}\\ &+\left(\frac{1}{\pi}-1\right)\sum_{m=1}^{M}\left(\int_{t_i\in\mathcal{I}_m}(\delta^{\text{inst}}_{t_i}-\delta^{\text{inst}})f(t_i)dt_i\right)\underbrace{\left(\int_{t_j\in\mathcal{I}_m}\delta^{\text{co}}_{t_j}\left(\int_{t'\in\mathcal{I}_m}d^{\text{co}}_{t_j}(t')f(t')dt'\right)f(t_j)dt_j\right)}_{I^{(m)}}\\ &+\sum_{m=1}^{M}\sum_{m':m'\neq m}\left(\int_{t_i\in\mathcal{I}_m}(\delta^{\text{inst}}_{t_i}-\delta^{\text{inst}})f(t_i)dt_i\right)\underbrace{\left(\int_{t_j\in\mathcal{I}_{m'}}\delta^{\text{co}}_{t_j}\left(\int_{t'\in\mathcal{I}_{m'}}d^{\text{co}}_{t_j}(t')f(t')dt'\right)f(t_j)dt_j\right)}_{I^{(m')}}\\ =&\sum_{m=1}^{M}\Xi^{\text{inst},(m)}\tilde{\Xi}^{\text{co},(m)}+\delta^{\text{co}}\left(\sum_{m'=1}^{M}I^{(m')}\right)\underbrace{\left(\sum_{m=1}^{M}\int_{t_i\in\mathcal{I}_m}(\delta^{\text{inst}}_{t_i}-\delta^{\text{inst}})f(t_i)dt_i\right)}_{=0}\qquad(\pi=1/2)\\ =&\sum_{m=1}^{M}\Xi^{\text{inst},(m)}\tilde{\Xi}^{\text{co},(m)}.\end{aligned}$$

For $B_{2,ij}$, we have

$$\begin{aligned}
B_{2,ij} =&\delta^{\text{co}} \sum_{m=1}^{M} \int_{t_i,t_j \in \mathcal{I}_m} (\delta^{\text{inst}}_{t_j} - \delta^{\text{inst}}) \mathbb{E}_W \left[\frac{W_{t_i}}{\pi} \frac{W_{t_j}}{\pi} \mid t_i, t_j \right] f(t_i) f(t_j) dt_i dt_j \\
&+ \delta^{\text{co}} \sum_{m=1}^{M} \sum_{m':m' \neq m} \int_{t_i \in \mathcal{I}_m, t_j \in \mathcal{I}_{m'}} (\delta^{\text{inst}}_{t_j} - \delta^{\text{inst}}) \mathbb{E}_W \left[\frac{W_{t_i}}{\pi} \frac{W_{t_j}}{\pi} \mid t_i, t_j \right] f(t_i) f(t_j) dt_i dt_j \\
=&\frac{\delta^{\text{co}}}{\pi} \sum_{m=1}^{M} \mu^{(m)} \int_{t_j \in \mathcal{I}_m} (\delta^{\text{inst}}_{t_j} - \delta^{\text{inst}}) f(t_j) dt_j + \delta^{\text{co}} \sum_{m=1}^{M} \sum_{m':m' \neq m} \mu^{(m)} \int_{t_j \in \mathcal{I}_{m'}} (\delta^{\text{inst}}_{t_j} - \delta^{\text{inst}}) f(t_j) dt_j \\
=&\delta^{\text{co}} \sum_{m=1}^{M} \mu^{(m)} \int_{t_j \in \mathcal{I}_m} (\delta^{\text{inst}}_{t_j} - \delta^{\text{inst}}) f(t_j) dt_j = \delta^{\text{co}} \sum_{m=1}^{M} \mu^{(m)} \Xi^{\text{inst},(m)} ,
\end{aligned}$$

by the definition of $\delta^{\text{inst}}_{\ell,t}$ and $\pi = 1/2$. Then $A_{2,ij}$ is equal to

$$A_{2,ij} = B_{1,ij} - B_{2,ij} = \sum_{m=1}^{M} \Xi^{\text{inst},(m)} \left(\tilde{\Xi}^{\text{co},(m)} - \delta^{\text{co}} \mu^{(m)} \right) = \sum_{m=1}^{M} \Xi^{\text{inst},(m)} \Xi^{\text{co},(m)} .$$

Finally we compute $A_{3,ij}$.

$$\begin{aligned}
A_{3,ij} =& \underbrace{\mathbb{E}_{W,t} \left[\frac{W_{t_i} - \pi}{\pi(1-\pi)} \frac{W_{t_j} - \pi}{\pi(1-\pi)} \delta^{\text{co}}_{t_i}(\boldsymbol{W}) \delta^{\text{co}}_{t_j}(\boldsymbol{W}) \right]}_{:=B_{1,ij}} - \underbrace{\delta^{\text{co}} \mathbb{E}_{W,t} \left[\frac{W_{t_i} - \pi}{\pi(1-\pi)} \frac{W_{t_j}}{\pi} \delta^{\text{co}}_{t_i}(\boldsymbol{W}) \right]}_{:=B_{2,ij}} \\
& - \underbrace{\delta^{\text{co}} \mathbb{E}_{W,t} \left[\frac{W_{t_i}}{\pi} \frac{W_{t_j} - \pi}{\pi(1-\pi)} \delta^{\text{co}}_{t_j}(\boldsymbol{W}) \right]}_{:=B_{3,ij}} + \underbrace{(\delta^{\text{co}})^2 \mathbb{E}_{W,t} \left[\frac{W_{t_i}}{\pi} \frac{W_{t_j}}{\pi} \right]}_{:=B_{4,ij}} .
\end{aligned}$$

For $B_{1,ij}$, we have

$$\begin{aligned}
B_{1,ij} = \sum_{m,m'=1}^{M} \int_{t_i \in \mathcal{I}_m, t_j \in \mathcal{I}_{m'}} & \delta^{\text{co}}_{t_i} \delta^{\text{co}}_{t_j} \mathbb{E}_W \left[\frac{W_{t_i} - \pi}{\pi(1-\pi)} \frac{W_{t_j} - \pi}{\pi(1-\pi)} \left(\sum_{k=1}^{M} W^{(k)} \int_{t' \in \mathcal{I}_k} d^{\text{co}}_{t_i}(t') f(t') dt' \right) \right. \\
& \left. \left(\sum_{k'=1}^{M} W^{(k')} \int_{t' \in \mathcal{I}_{k'}} d^{\text{co}}_{t_j}(t') f(t') dt' \right) \mid t_i, t_j \right] f(t_i) f(t_j) dt_i dt_j .
\end{aligned}$$

When $m = m'$, the term $\mathbb{E}_W[\cdot \mid t_i, t_j]$ in the last equation equals to

$$\begin{aligned}
& \mathbb{E}_W \left[\frac{(W_{t_i} - \pi)^2}{\pi^2(1-\pi)^2} \left(\sum_{k=1}^{M} W^{(k)} \int_{t' \in \mathcal{I}_k} d^{\text{co}}_{t_i}(t') f(t') dt' \right) \left(\sum_{k'=1}^{M} W^{(k')} \int_{t' \in \mathcal{I}_{k'}} d^{\text{co}}_{t_j}(t') f(t') dt' \right) \mid t_i, t_j \right] \\
=& \frac{1}{\pi} \left(\int_{t' \in \mathcal{I}_m} d^{\text{co}}_{t_i}(t') f(t') dt' \right) \left(\int_{t' \in \mathcal{I}_m} d^{\text{co}}_{t_j}(t') f(t') dt' \right) && (k = k' = m) \\
&+ \frac{1}{1-\pi} \sum_{k:k \neq m} \left(\int_{t' \in \mathcal{I}_k} d^{\text{co}}_{t_i}(t') f(t') dt' \right) \left(\int_{t' \in \mathcal{I}_k} d^{\text{co}}_{t_j}(t') f(t') dt' \right) && (k = k' \text{ and both do not equal to } m) \\
&+ \sum_{k':k' \neq m} \left(\int_{t' \in \mathcal{I}_m} d^{\text{co}}_{t_i}(t') f(t') dt' \right) \left(\int_{t' \in \mathcal{I}_{k'}} d^{\text{co}}_{t_j}(t') f(t') dt' \right) && (k = m, \text{ but } k' \neq m) \\
&+ \sum_{k:k \neq m} \left(\int_{t' \in \mathcal{I}_k} d^{\text{co}}_{t_i}(t') f(t') dt' \right) \left(\int_{t' \in \mathcal{I}_m} d^{\text{co}}_{t_j}(t') f(t') dt' \right) && (k \neq m, \text{ but } k' = m) \\
&+ \frac{\pi}{1-\pi} \sum_{k,k':k \neq m, k' \neq m, k \neq k'} \left(\int_{t' \in \mathcal{I}_k} d^{\text{co}}_{t_i}(t') f(t') dt' \right) \left(\int_{t' \in \mathcal{I}_{k'}} d^{\text{co}}_{t_j}(t') f(t') dt' \right) && (k \neq m, \, k' \neq m, \, k \neq k') \\
=& 1 + \sum_{k=1}^{M} \left(\int_{t' \in \mathcal{I}_k} d^{\text{co}}_{t_i}(t') f(t') dt' \right) \left(\int_{t' \in \mathcal{I}_k} d^{\text{co}}_{t_j}(t') f(t') dt' \right) && (\pi = 1/2)
\end{aligned}$$

where we use $\sum_{k=1}^{K}\int_{t'\in\mathcal{I}_k} d_{t_i}^{\mathrm{co}}(t')f(t')dt' = 1$ for any t_i.

When $m \neq m'$, the term $\mathbb{E}_W[\cdot \mid t_i, t_j]$ is equal to

$$\mathbb{E}_W[\cdot \mid t_i,t_j] = \left(\int_{t'\in\mathcal{I}_m} d_{t_i}^{\mathrm{co}}(t')f(t')dt'\right)\left(\int_{t'\in\mathcal{I}_{m'}} d_{t_j}^{\mathrm{co}}(t')f(t')dt'\right) + \left(\int_{t'\in\mathcal{I}_{m'}} d_{t_i}^{\mathrm{co}}(t')f(t')dt'\right)\left(\int_{t'\in\mathcal{I}_m} d_{t_j}^{\mathrm{co}}(t')f(t')dt'\right).$$

Then $B_{1,ij}$ is equal to

$$\begin{aligned}
B_{1,ij} =& \sum_{m=1}^{M}\int_{t_i,t_j\in\mathcal{I}_m} \delta_{t_i}^{\mathrm{co}}\delta_{t_j}^{\mathrm{co}}\left[1+\sum_{k=1}^{M}\left(\int_{t'\in\mathcal{I}_k} d_{t_i}^{\mathrm{co}}(t')f(t')dt'\right)\left(\int_{t'\in\mathcal{I}_k} d_{t_j}^{\mathrm{co}}(t')f(t')dt'\right)\right] f(t_i)f(t_j)dt_idt_j \\
&+ \sum_{m,m':m\neq m'}\int_{t_i\in\mathcal{I}_m,t_j\in\mathcal{I}_{m'}} \delta_{t_i}^{\mathrm{co}}\delta_{t_j}^{\mathrm{co}}\left(\int_{t'\in\mathcal{I}_m} d_{t_i}^{\mathrm{co}}(t')f(t')dt'\right)\left(\int_{t'\in\mathcal{I}_{m'}} d_{t_j}^{\mathrm{co}}(t')f(t')dt'\right) \\
&+ \delta_{t_i}^{\mathrm{co}}\delta_{t_j}^{\mathrm{co}}\left(\int_{t'\in\mathcal{I}_{m'}} d_{t_i}^{\mathrm{co}}(t')f(t')dt'\right)\left(\int_{t'\in\mathcal{I}_m} d_{t_j}^{\mathrm{co}}(t')f(t')dt'\right) f(t_i)f(t_j)dt_idt_j \\
=& \sum_{m=1}^{M}\left[\tilde{\Xi}^{\mathrm{co},(m)}\right]^2 + \sum_{m=1}^{M}\sum_{m'=1}^{M}\left[I^{(m,m')}\right]^2 + \sum_{m=1}^{M}\sum_{m':m'\neq m}\left(I^{(m)}I^{(m')} + I^{(m,m')}I^{(m',m)}\right),
\end{aligned}$$

where

$$I^{(m,m')} = \int_{t_i\in\mathcal{I}_m,t'\in\mathcal{I}_{m'}} \delta_{t_i}^{\mathrm{co}} d_{t_i}^{\mathrm{co}}(t')f(t_i)f(t')dt_idt'.$$

For $B_{2,ij}$, we have

$$\begin{aligned}
B_{2,ij} =& \delta^{\mathrm{co}}\mathbb{E}_{W,t}\left[\frac{W_{t_i}-\pi}{\pi(1-\pi)}\frac{W_{t_j}}{\pi}\delta_{t_i}^{\mathrm{co}}(\boldsymbol{W})\right] \\
=& \delta^{\mathrm{co}}\mathbb{E}_{W,t}\left[\frac{W_{t_i}-\pi}{\pi(1-\pi)}\frac{W_{t_j}}{\pi}\delta_{t_i}^{\mathrm{co}}\sum_{k=1}^{M}W^{(k)}\int_{t'\in\mathcal{I}_k} d_{t_i}^{\mathrm{co}}(t')f(t')dt'\right] \\
=& \delta^{\mathrm{co}}\sum_{m=1}^{M}\int_{t_i,t_j\in\mathcal{I}_m} \delta_{t_i}^{\mathrm{co}}\mathbb{E}_W\left[\frac{W_{t_i}}{\pi^2}\left(\sum_{k=1}^{M}W^{(k)}\int_{t'\in\mathcal{I}_k} d_{t_i}^{\mathrm{co}}(t')f(t')dt'\right)\mid t_i\right] f(t_i)f(t_j)dt_idt_j \\
& \qquad (t_i \text{ and } t_j \text{ in the same interval}) \\
&+ \delta^{\mathrm{co}}\sum_{m=1}^{M}\sum_{m':m'\neq m}\int_{t_i\in\mathcal{I}_m,t_j\in\mathcal{I}_{m'}} \delta_{t_i}^{\mathrm{co}}\cdot\mathbb{E}_W\left[\frac{W_{t_i}-\pi}{\pi(1-\pi)}\frac{W_{t_j}}{\pi}\left(\sum_{k=1}^{M}W^{(k)}\int_{t'\in\mathcal{I}_k} d_{t_i}^{\mathrm{co}}(t')f(t')dt'\right)\mid t_i\right] f(t_i)f(t_j)dt_idt_j \\
& \qquad (t_i \text{ and } t_j \text{ in different intervals}) \\
=& \delta^{\mathrm{co}}\sum_{m=1}^{M}\mu^{(m)}\int_{t_i\in\mathcal{I}_m}\delta_{t_i}^{\mathrm{co}}\left(1+\int_{t'\in\mathcal{I}_m} d_{t_i}^{\mathrm{co}}(t')f(t')dt'\right)f(t_i)dt_i \\
&+ \delta^{\mathrm{co}}\sum_{m=1}^{M}\sum_{m':m'\neq m}\mu^{(m')}\int_{t_i\in\mathcal{I}_m}\delta_{t_i}^{\mathrm{co}}\left[\int_{t'\in\mathcal{I}_m} d_{t_i}^{\mathrm{co}}(t')f(t')dt'\right]f(t_i)dt_i \\
=& \delta^{\mathrm{co}}\sum_{m=1}^{M}\mu^{(m)}\left(\tilde{\Xi}^{\mathrm{co},(m)} + I^{(m)}\right) + \delta^{\mathrm{co}}\sum_{m=1}^{M}\sum_{m':m'\neq m}\mu^{(m')}I^{(m)} \\
=& \delta^{\mathrm{co}}\left(\sum_{m=1}^{M}I^{(m)} + \sum_{m=1}^{M}\mu^{(m)}\tilde{\Xi}^{\mathrm{co},(m)}\right).
\end{aligned}$$

Similarly we can show that $B_{3,ij} = B_{2,ij}$.

For $B_{4,ij}$, we have

$$
\begin{aligned}
B_{4,ij} =&(\delta^{\text{co}})^2\mathbb{E}_{W,t}\left[\frac{W_{t_i}}{\pi}\frac{W_{t_j}}{\pi}\right] \\
=&(\delta^{\text{co}})^2\int f(t_i)f(t_j)dt_idt_j + (\delta^{\text{co}})^2\left(\frac{1}{\pi}-1\right)\sum_{m=1}^{M}\int_{t_i,t_j\in\mathcal{I}_m} f(t_i)f(t_j)dt_idt_j \\
=&(\delta^{\text{co}})^2 + (\delta^{\text{co}})^2\sum_{m=1}^{M}\left[\mu^{(m)}\right]^2 .
\end{aligned}
$$

In summary, $A_{3,ij}$ equals to

$$
\begin{aligned}
A_{3,ij} =& \sum_{m=1}^{M}\left[\tilde{\Xi}^{\text{co},(m)}\right]^2 + \sum_{m=1}^{M}\sum_{m'=1}^{M}\left[I^{(m,m')}\right]^2 + \sum_{m=1}^{M}\sum_{m':m'\neq m}\left(I^{(m)}I^{(m')} + I^{(m,m')}I^{(m',m)}\right) \\
&- 2\delta^{\text{co}}\left(\sum_{m=1}^{M}I^{(m)} + \sum_{m=1}^{M}\mu^{(m)}\tilde{\Xi}^{\text{co},(m)}\right) + (\delta^{\text{co}})^2 + (\delta^{\text{co}})^2\sum_{m=1}^{M}\left[\mu^{(m)}\right]^2 \\
=& \sum_{m=1}^{M}\left[\Xi^{\text{co},(m)}\right]^2 + \left(\delta^{\text{co}} - \sum_{m=1}^{M}I^{(m)}\right)^2 + \sum_{m=1}^{M}\sum_{m':m'\neq m}\left(\left[I^{(m,m')}\right]^2 + I^{(m,m')}I^{(m',m)}\right)
\end{aligned}
$$

where $\Xi^{\text{co},(m)} = \tilde{\Xi}^{\text{co},(m)} - \delta^{\text{co}}\mu^{(m)}$.

The term A^{treat} is then equal to

$$
\begin{aligned}
&A^{\text{treat}} = A_{1,ij} + 2A_{2,ij} + A_{3,ij} \\
=&\sum_{m=1}^{M}\left[\Xi^{\text{inst},(m)}\right]^2 + 2\sum_{m=1}^{M}\Xi^{\text{inst},(m)}\Xi^{\text{co},(m)} + \sum_{m=1}^{M}\left[\Xi^{\text{co},(m)}\right]^2 + \left(\delta^{\text{co}} - \sum_{m=1}^{M}I^{(m)}\right)^2 \\
&+ \sum_{m=1}^{M}\sum_{m':m'\neq m}\left(\left[I^{(m,m')}\right]^2 + I^{(m,m')}I^{(m',m)}\right) \\
=&\sum_{m=1}^{M}\left(\Xi^{\text{inst},(m)} + \Xi^{\text{co},(m)}\right)^2 + \left(\sum_{m=1}^{M}I^{(m)} - \delta^{\text{co}}\right)^2 + \sum_{m=1}^{M}\sum_{m'\neq m}\left(\left[I^{(m,m')}\right]^2 + I^{(m,m')}I^{(m',m)}\right) .
\end{aligned}
$$

The second moment of $\mathcal{E}_{\text{carryover}}$ is then equal to

$$
\begin{aligned}
&\mathbb{E}_{W,\varepsilon,t}\left[(\mathcal{E}_{\text{carryover}})^2\right] = A^{\text{treat}} + 2A^{\text{cross}} + A^{\text{control}} \\
=&\sum_{m=1}^{M}\left(\Xi^{\text{inst},(m)} + \Xi^{\text{co},(m)}\right)^2 + \left(\sum_{m=1}^{M}I^{(m)} - \delta^{\text{co}}\right)^2 + \sum_{m=1}^{M}\sum_{m'\neq m}\left(\left[I^{(m,m')}\right]^2 + I^{(m,m')}I^{(m',m)}\right) \\
&+ 4\sum_{m=1}^{M}\Xi^{(m)}\mu^{(m)}_{Y^{\text{ctrl}}} + 4\sum_{m=1}^{M}\left[\mu^{(m)}_{Y^{\text{ctrl}}}\right]^2 \\
=&\sum_{m=1}^{M}\left(\Xi^{\text{inst},(m)} + \Xi^{\text{co},(m)} + 2\mu^{(m)}_{Y^{\text{ctrl}}}\right)^2 + \left(\sum_{m=1}^{M}I^{(m)} - \delta^{\text{co}}\right)^2 + \sum_{m=1}^{M}\sum_{m'\neq m}\left(\left[I^{(m,m')}\right]^2 + I^{(m,m')}I^{(m',m)}\right) .
\end{aligned}
$$

We then finish the proof of Lemma B.12. □

LEMMA B.13 **(Expected product of instantaneous effect and carryover effects)**. *Under the assumptions in Theorem 4.1, the expected value of the product of $\mathcal{E}_{\text{inst}}$ and $\mathcal{E}_{\text{carryover}}$ equals to*

$$
\mathbb{E}_{W,\varepsilon,t}\left[\mathcal{E}_{\text{inst}}\mathcal{E}_{\text{carryover}}\right] = \delta^{\text{gate}}\sum_{m=1}^{M}\mu^{(m)}\left(\Xi^{\text{inst},(m)} + \Xi^{\text{co},(m)} + 2\mu^{(m)}_{Y^{\text{ctrl}}}\right) .
$$

Proof of Lemma B.13 Next we compute the expected value of the product of $\mathcal{E}_{\text{carryover}}$ and $\mathcal{E}_{\text{inst}}$.

$$\begin{aligned}
&\mathbb{E}_{W,\varepsilon,t}\left[\mathcal{E}_{\text{inst}}\mathcal{E}_{\text{carryover}}\right] \\
=&\mathbb{E}_{W,\varepsilon,t}\left[\delta^{\text{gate}}\left(\frac{1}{n}\sum_{j=1}^{n}\frac{W_{t_j}}{\pi}-1\right)\left(\frac{1}{n}\sum_{i=1}^{n}\alpha_{t_i}\left(Y_{t_i}(\boldsymbol{W},\mathbf{0},\cdots,\mathbf{0})-Y_{t_i}(\mathbf{0},\cdots,\mathbf{0})-W_{t_i}\delta^{\text{gate}}\right)\right)\right] \\
&+\mathbb{E}_{W,\varepsilon,t}\left[\delta^{\text{gate}}\left(\frac{1}{n}\sum_{j=1}^{n}\frac{W_{t_j}}{\pi}-1\right)\left(\frac{1}{n}\sum_{i=1}^{n}\alpha_{t_i}Y_{t_i}(\mathbf{0},\cdots,\mathbf{0})\right)\right] \\
=&\frac{1}{n^2}\sum_{i,j}\underbrace{\delta^{\text{gate}}\mathbb{E}_t\left[(\delta^{\text{inst}}_{t_i}-\delta^{\text{inst}})\cdot\frac{W_{t_i}}{\pi}\left(\frac{W_{t_j}}{\pi}-1\right)\right]}_{:=A_{1,ij}} \\
&+\frac{1}{n^2}\sum_{i,j}\underbrace{\delta^{\text{gate}}\mathbb{E}_{W,t}\left[\alpha_{t_i}(\delta^{\text{co}}_{t_i}(\boldsymbol{W})-\delta^{\text{co}}W_{t_i})\left(\frac{W_{t_j}}{\pi}-1\right)\right]}_{:=A_{2,ij}} \\
&+\frac{1}{n^2}\sum_{i,j}\underbrace{\delta^{\text{gate}}\mathbb{E}_{W,t}\left[\alpha_{t_i}Y_{t_i}(\mathbf{0},\cdots,\mathbf{0})\left(\frac{W_{t_j}}{\pi}-1\right)\right]}_{:=A_{3,ij}}.
\end{aligned}$$

For $A_{1,ij}$, we have

$$\begin{aligned}
A_{1,ij}=&\delta^{\text{gate}}\sum_{m=1}^{M}\int_{t_i,t_j\in\mathcal{I}_m}(\delta^{\text{inst}}_{t_i}-\delta^{\text{inst}})\cdot\mathbb{E}_W\left[\frac{W_{t_i}}{\pi}\left(\frac{W_{t_j}}{\pi}-1\right)\mid t_i,t_j\right]f(t_i)f(t_j)dt_idt_j \\
&+\delta^{\text{gate}}\sum_{m=1}^{M}\sum_{m':m'\neq m}\int_{t_i\in\mathcal{I}_m,t_j\in\mathcal{I}_{m'}}(\delta^{\text{inst}}_{t_i}-\delta^{\text{inst}})\cdot\underbrace{\mathbb{E}_W\left[\frac{W_{t_i}}{\pi}\left(\frac{W_{t_j}}{\pi}-1\right)\mid t_i,t_j\right]}_{=0\text{ as }t_i\text{ and }t_j\text{ in different interval}}f(t_i)f(t_j)dt_idt_j \\
=&\delta^{\text{gate}}\frac{1-\pi}{\pi}\sum_{m=1}^{M}\int_{t_i,t_j\in\mathcal{I}_m}(\delta^{\text{inst}}_{t_i}-\delta^{\text{inst}})f(t_i)f(t_j)dt_idt_j \\
=&\delta^{\text{gate}}\sum_{m=1}^{M}\Xi^{\text{inst},(m)}\mu^{(m)}. \qquad (\pi=1/2)
\end{aligned}$$

For $A_{2,ij}$, we have

$$A_{2,ij}=\underbrace{\delta^{\text{gate}}\mathbb{E}_{W,t}\left[\alpha_{t_i}\delta^{\text{co}}_{t_i}(\boldsymbol{W})\left(\frac{W_{t_j}}{\pi}-1\right)\right]}_{:=B_{1,ij}}-\underbrace{\delta^{\text{gate}}\delta^{\text{co}}\mathbb{E}_{W,t}\left[\alpha_{t_i}W_{t_i}\left(\frac{W_{t_j}}{\pi}-1\right)\right]}_{:=B_{2,ij}}.$$

For $B_{2,ij}$, we have

$$\begin{aligned}
B_{2,ij}=&\delta^{\text{gate}}\delta^{\text{co}}\mathbb{E}_{W,t}\left[\alpha_{t_i}W_{t_i}\left(\frac{W_{t_j}}{\pi}-1\right)\right]=\delta^{\text{co}}\mathbb{E}_{W,t}\left[\frac{W_{t_i}}{\pi}\left(\frac{W_{t_j}}{\pi}-1\right)\right] \\
=&\delta^{\text{gate}}\delta^{\text{co}}\frac{1-\pi}{\pi}\sum_{m=1}^{M}\int_{t_i,t_j\in\mathcal{I}_m}f(t_i)f(t_j)dt_idt_j \\
=&\delta^{\text{gate}}\delta^{\text{co}}\sum_{m=1}^{M}\left[\mu^{(m)}\right]^2. \qquad (\pi=1/2)
\end{aligned}$$

For $B_{1,ij}$, we have

$$\begin{aligned}
B_{1,ij}=&\delta^{\text{gate}}\mathbb{E}_{W,t}\left[\alpha_{t_i}\delta^{\text{co}}_{t_i}(\boldsymbol{W})\left(\frac{W_{t_j}}{\pi}-1\right)\right] \\
=&\delta^{\text{gate}}\sum_{m=1}^{M}\int_{t_i,t_j\in\mathcal{I}_m}\mathbb{E}_W\left[\frac{W_{t_i}-\pi}{\pi(1-\pi)}\frac{W_{t_j}-\pi}{\pi}\delta^{\text{co}}_{t_i}(\boldsymbol{W})\mid t_i,t_j\right]f(t_i)f(t_j)dt_idt_j
\end{aligned}$$

$$+\,\delta^{\text{gate}} \sum_{m=1}^{M} \sum_{m':m'\neq m} \int_{t_i\in\mathcal{I}_m, t_j\in\mathcal{I}_{m'}} \mathbb{E}_W\left[\frac{W_{t_i}-\pi}{\pi(1-\pi)}\frac{W_{t_j}-\pi}{\pi}\delta^{\text{co}}_{t_i}(\boldsymbol{W}) \mid t_i, t_j\right] f(t_i)f(t_j)dt_idt_j \quad \text{(this term is zero)}$$

$$=\delta^{\text{gate}} \sum_{m=1}^{M} \int_{t_i,t_j\in\mathcal{I}_m} \delta^{\text{co}}_{t_i}\left(\frac{1-\pi}{\pi}\left[\int_{t'\in\mathcal{I}_m} d^{\text{co}}_{t_i}(t')f(t')dt'\right] + \sum_{k:k\neq m}\frac{\pi}{\pi}\left[\int_{t'\in\mathcal{I}_k} d^{\text{co}}_{t_i}(t')f(t')dt'\right]\right) f(t_i)f(t_j)dt_idt_j$$

$$=\delta^{\text{gate}} \sum_{m=1}^{M} \int_{t_i,t_j\in\mathcal{I}_m} \delta^{\text{co}}_{t_i} f(t_i)f(t_j)dt_idt_j = \delta^{\text{gate}} \sum_{m=1}^{M} \mu^{(m)}\tilde{\Gamma}^{(m)}\,.$$

For $A_{3,ij}$, we have

$$\begin{aligned} A_{3,ij} =& \delta^{\text{gate}}\mathbb{E}_{W,t}\left[\alpha_{t_i}Y_{t_i}(\mathbf{0},\cdots,\mathbf{0})\left(\frac{W_{t_j}}{\pi}-1\right)\right] \\ =& 2\delta^{\text{gate}} \sum_{m=1}^{M} \int_{t_i,t_j\in\mathcal{I}_m} Y_{t_i}(\mathbf{0},\cdots,\mathbf{0})f(t_i)f(t_j)dt_idt_j \\ =& 2\delta^{\text{gate}} \sum_{m=1}^{M} \mu^{(m)}\mu^{(m)}_{Y^{\text{ctrl}}}\,. \end{aligned}$$

Therefore, we have

$$\begin{aligned} \mathbb{E}_{W,\varepsilon,t}\left[\mathcal{E}_{\text{inst}}\cdot\mathcal{E}_{\text{carryover}}\right] =& A_{1,ij} + B_{1,ij} - B_{2,ij} + A_{3,ij} \\ =& \delta^{\text{gate}} \sum_{m=1}^{M} \mu^{(m)}\left(\Xi^{\text{inst},(m)} + \Xi^{\text{co},(m)} + 2\mu^{(m)}_{Y^{\text{ctrl}}}\right) \end{aligned}$$

We then finish the proof of Lemma B.13. □

Lemma B.14 **(Second moment of effects from other interventions)**. *Under the assumptions in Theorem 4.1, the bias from simultaneous interventions is*

$$\mathbb{E}_{W,\varepsilon,t}\left[(\mathcal{E}_{\text{simul}})^2\right] = \sum_{m=1}^{M}\sum_{m'=1}^{M} S^{(m,m')}_{\text{var}}\,,$$

where $S^{(m,m')}_{\text{var}}$ is defined in Equation (A.1).

Proof of Lemma B.14 As events are sampled i.i.d. from distribution, we have

$$\begin{aligned} \mathbb{E}_{W,\varepsilon,t}\left[(\mathcal{E}_{\text{simul}})^2\right] =& \mathbb{E}_{W,\varepsilon,t}\left[\left(\frac{1}{n}\sum_{i=1}^{n}\alpha_{t_i}\left(Y_{t_i}(\boldsymbol{W},\boldsymbol{W}^{\text{s}}_1,\cdots,\boldsymbol{W}^{\text{s}}_K) - Y_{t_i}(\boldsymbol{W},\mathbf{0},\cdots,\mathbf{0})\right)\right)^2\right] \\ =& \frac{1}{n^2}\sum_{i,j}\mathbb{E}_{W,t}\left[\frac{W_{t_i}-\pi}{\pi(1-\pi)}\frac{W_{t_j}-\pi}{\pi(1-\pi)}\mathbb{E}_{\boldsymbol{W}^{\text{s}}_1,\cdots,\boldsymbol{W}^{\text{s}}_K}\left[(Y_{t_i}(\boldsymbol{W},\boldsymbol{W}^{\text{s}}_1,\cdots,\boldsymbol{W}^{\text{s}}_K) - Y_{t_i}(\boldsymbol{W},\mathbf{0},\cdots,\mathbf{0}))\cdot\right.\right. \\ & \left.\left.\left(Y_{t_j}(\boldsymbol{W},\boldsymbol{W}^{\text{s}}_1,\cdots,\boldsymbol{W}^{\text{s}}_K) - Y_{t_j}(\boldsymbol{W},\mathbf{0},\cdots,\mathbf{0})\right) \mid \boldsymbol{W}, t_i, t_j\right]\right] \quad \text{(by the law of total expectation)} \\ =& \mathbb{E}_{W,t}\left[\frac{W_{t_i}-\pi}{\pi(1-\pi)}\frac{W_{t_j}-\pi}{\pi(1-\pi)}\cdot\delta^{\text{simul},2}_{t_i,t_j}(\boldsymbol{W})\right] \quad \text{(by definition of } \delta^{\dagger}_{t_i,t_j}(\boldsymbol{W})\text{)} \\ =& 4\sum_{m=1}^{M}\int_{t_i,t_j\in\mathcal{I}_m}\mathbb{E}_W[\delta^{\text{simul},2}_{t_i,t_j}(\boldsymbol{W}) \mid t_i,t_j]f(t_i)f(t_j)dt_idt_j \\ & + 4\sum_{m=1}^{M}\sum_{m':m'\neq m}\int_{t_i\in\mathcal{I}_m,t_j\in\mathcal{I}_{m'}}\Phi^{2\dagger}_{t_i,t_j}f(t_i)f(t_j)dt_idt_j \\ =& \sum_{m=1}^{M}\sum_{m'=1}^{M} S^{(m,m')}_{\text{var}} \end{aligned}$$

following the definition of $S_{\mathrm{var}}^{(m,m')}$ in Equation (A.1) with

$$\begin{aligned}\delta_{t_i,t_j}^{\mathrm{simul},2}(\boldsymbol{W}) =&\mathbb{E}_{\boldsymbol{W}_1^{\mathrm{s}},\cdots,\boldsymbol{W}_K^{\mathrm{s}}}\left[(Y_{t_i}(\boldsymbol{W},\boldsymbol{W}_1^{\mathrm{s}},\cdots,\boldsymbol{W}_K^{\mathrm{s}}) - Y_{t_i}(\boldsymbol{W},\boldsymbol{0},\cdots,\boldsymbol{0}))\cdot\right.\\ &\left.\left(Y_{t_j}(\boldsymbol{W},\boldsymbol{W}_1^{\mathrm{s}},\cdots,\boldsymbol{W}_K^{\mathrm{s}}) - Y_{t_j}(\boldsymbol{W},\boldsymbol{0},\cdots,\boldsymbol{0})\right) \mid \boldsymbol{W},t_i,t_j\right]\end{aligned}$$

and

$$\begin{aligned}\Phi_{\ell,t_i,t_j}^{2\dagger} =&\frac{1}{4}\Bigg(\mathbb{E}_{\boldsymbol{W}^{(-m,-m')}}\left[\delta_{t_i,t_j}^{\mathrm{simul},2}(\boldsymbol{W}^{(-m,-m')},\boldsymbol{W}^{(m,m')}=(1,1))\right]\\ &-\mathbb{E}_{\boldsymbol{W}^{(-m,-m')}}\left[\delta_{t_i,t_j}^{\mathrm{simul},2}(\boldsymbol{W}^{(-m,-m')},\boldsymbol{W}^{(m,m')}=(1,0))\right]\\ &-\mathbb{E}_{\boldsymbol{W}^{(-m,-m')}}\left[\delta_{t_i,t_j}^{\mathrm{simul},2}(\boldsymbol{W}^{(-m,-m')},\boldsymbol{W}^{(m,m')}=(0,1))\right]\\ &+\mathbb{E}_{\boldsymbol{W}^{(-m,-m')}}\left[\delta_{t_i,t_j}^{\mathrm{simul},2}(\boldsymbol{W}^{(-m,-m')},\boldsymbol{W}^{(m,m')}=(0,0))\right]\Bigg).\end{aligned}$$

We then finish the proof of Lemma B.14. □

LEMMA B.15 **(Expected product of $\mathcal{E}_{\mathrm{simul}}$ and $\mathcal{E}_{\mathrm{inst}}$)**. *Under the assumptions in Theorem 4.2, we have*

$$\mathbb{E}_{W,\varepsilon,t}\left[\mathcal{E}_{\mathrm{simul}}\cdot\mathcal{E}_{\mathrm{inst}}\right] = \sum_{m=1}^{M}\sum_{m'=1}^{M}\delta^{\mathrm{gate}}\mu^{(m')}S_1^{(m,m')},$$

where $S_1^{(m,m')}$ is defined in Equation (A.3).

Proof of Lemma B.15 The expected value of the product of $\mathcal{E}_{\mathrm{simul}}$ and $\mathcal{E}_{\mathrm{inst}}$ is equal to

$$\mathbb{E}_{W,\varepsilon,t}\left[\mathcal{E}_{\mathrm{simul}}\mathcal{E}_{\mathrm{inst}}\right] = \frac{1}{n^2}\sum_{i,j}\underbrace{\mathbb{E}_{W,t}\left[\alpha_{t_i}\left(Y_{t_i}(\boldsymbol{W},\boldsymbol{W}_1^{\mathrm{s}},\cdots,\boldsymbol{W}_K^{\mathrm{s}}) - Y_{t_i}(\boldsymbol{W},\boldsymbol{0},\cdots,\boldsymbol{0})\right)\delta^{\mathrm{gate}}\left(\frac{W_{t_j}}{\pi}-1\right)\right]}_{A_{ij}}.$$

For A_{ij}, we have

$$\begin{aligned}A_{ij} &= \delta^{\mathrm{gate}}\mathbb{E}_{W,t}\left[\frac{W_{t_i}-\pi}{\pi(1-\pi)}\left(\frac{W_{t_j}}{\pi}-1\right)\delta_{t_i}^{\mathrm{simul}}(\boldsymbol{W})\right]\\ &=\delta^{\mathrm{gate}}\sum_{m=1}^{M}\int_{t_i,t_j\in\mathcal{I}_m}\mathbb{E}_W\left[\frac{W_{t_i}-\pi}{\pi(1-\pi)}\left(\frac{W_{t_j}}{\pi}-1\right)\delta_{t_i}^{\mathrm{simul}}(\boldsymbol{W})\mid t_i t_j\right]f(t_i)f(t_j)dt_i dt_j\\ &\quad+\delta^{\mathrm{gate}}\sum_{m=1}^{M}\sum_{m':m'\neq m}\int_{t_i\in\mathcal{I}_m,t_j\in\mathcal{I}_{m'}}\mathbb{E}_W\left[\frac{W_{t_i}-\pi}{\pi(1-\pi)}\left(\frac{W_{t_j}}{\pi}-1\right)\delta_{t_i}^{\mathrm{simul}}(\boldsymbol{W})\mid t_i t_j\right]f(t_i)f(t_j)dt_i dt_j\\ &=2\delta^{\mathrm{gate}}\sum_{m=1}^{M}\int_{t_i,t_j\in\mathcal{I}_m}\mathbb{E}_W\left[\delta_{t_i}^{\mathrm{simul}}(\boldsymbol{W})\right]f(t_i)f(t_j)dt_i dt_j\\ &\quad+2\delta^{\mathrm{gate}}\sum_{m=1}^{M}\sum_{m':m'\neq m}\int_{t_i\in\mathcal{I}_m,t_j\in\mathcal{I}_{m'}}\Phi_{t_i}^{\mathrm{simul},(-m')}f(t_i)f(t_j)dt_i dt_j,\end{aligned}$$

where

$$\begin{aligned}\Phi_{t_i}^{\mathrm{simul},(-m')} =&\frac{1}{4}\Bigg(\mathbb{E}_{\boldsymbol{W}^{(-m,-m')}}\left[\delta_{t_i}^{\mathrm{simul}}(\boldsymbol{W}^{(-m,-m')},\boldsymbol{W}^{(m,m')}=(1,1))\right]\\ &-\mathbb{E}_{\boldsymbol{W}^{(-m,-m')}}\left[\delta_{t_i}^{\mathrm{simul}}(\boldsymbol{W}^{(-m,-m')},\boldsymbol{W}^{(m,m')}=(1,0))\right]\\ &-\mathbb{E}_{\boldsymbol{W}^{(-m,-m')}}\left[\delta_{t_i}^{\mathrm{simul}}(\boldsymbol{W}^{(-m,-m')},\boldsymbol{W}^{(m,m')}=(0,1))\right]\\ &+\mathbb{E}_{\boldsymbol{W}^{(-m,-m')}}\left[\delta_{t_i}^{\mathrm{simul}}(\boldsymbol{W}^{(-m,-m')},\boldsymbol{W}^{(m,m')}=(0,0))\right]\Bigg).\end{aligned}$$

Therefore,

$$
\begin{aligned}
\mathbb{E}_{W,\varepsilon,t}\left[\mathcal{E}_{\text{inst}}\mathcal{E}_{\text{simul}}\right] =& 2\delta^{\text{gate}}\sum_{m=1}^{M}\mu^{(m)}\int_{t_i\in\mathcal{I}_m}\mathbb{E}_W\left[\delta_{t_i}^{\text{simul}}(\boldsymbol{W})\right]f(t_i)dt_i \\
&+2\delta^{\text{gate}}\sum_{m=1}^{M}\sum_{m':m'\neq m}\mu^{(m')}\int_{t_i\in\mathcal{I}_m}\Phi_{t_i}^{\text{simul},(-m')}f(t_i)dt_i \\
=&\delta^{\text{gate}}\sum_{m=1}^{M}\sum_{m'=1}^{M}\mu^{(m')}S_1^{(m,m')}
\end{aligned}
$$

following the definition of $S_1^{(m,m')}$ in Equation A.3. We then finish the proof of Lemma B.15. □

LEMMA B.16 **(Expected product of $\mathcal{E}_{\text{simul}}$ and $\mathcal{E}_{\text{carryover}}$)**. *Under the assumptions in Theorem 4.2, we have*

$$
\mathbb{E}_{W,\varepsilon,t}\left[\mathcal{E}_{\text{simul}}\cdot\mathcal{E}_{\text{carryover}}\right]=\sum_{m=1}^{M}\sum_{m'=1}^{M}\left[2\mu_{Y^{\text{ctrl}}}^{(m')}S_1^{(m,m')}+\left(\Xi^{\text{inst},(m)}-\delta^{\text{co}}\mu^{(m)}\right)S_2^{(m,m')}+S_3^{(m,m')}\right],
$$

where $S_1^{(m,m')}$, $S_2^{(m,m')}$, and $S_3^{(m,m')}$ are defined in Equations (A.3), (A.4), *and* (A.5) *respectively.*

Proof of Lemma B.16 The expected value of the product of $\mathcal{E}_{\text{simul}}$ and $\mathcal{E}_{\text{carryover}}$ is equal to

$$
\begin{aligned}
&\mathbb{E}_{W,\varepsilon,t}\left[\mathcal{E}_{\text{carryover}}\mathcal{E}_{\text{simul}}\right] \\
=&\frac{1}{n^2}\sum_{i,j}\underbrace{\mathbb{E}_{W,t}\left[\alpha_{t_i}(\delta_{t_i}^{\text{inst}}-\delta^{\text{inst}})W_{t_i}\alpha_{t_j}\left(Y_{t_j}(\boldsymbol{W},\boldsymbol{W}_1^{\text{s}},\cdots,\boldsymbol{W}_K^{\text{s}})-Y_{t_j}(\boldsymbol{W},\boldsymbol{0},\cdots,\boldsymbol{0})\right)\right]}_{:=A_{1,ij}} \\
+&\frac{1}{n^2}\sum_{i,j}\underbrace{\mathbb{E}_{W,t}\left[\alpha_{t_i}(\delta_{t_i}^{\text{co}}(\boldsymbol{W})-\delta^{\text{co}}W_{t_i})\alpha_{t_j}\left(Y_{t_j}(\boldsymbol{W},\boldsymbol{W}_1^{\text{s}},\cdots,\boldsymbol{W}_K^{\text{s}})-Y_{t_j}(\boldsymbol{W},\boldsymbol{0},\cdots,\boldsymbol{0})\right)\right]}_{:=A_{2,ij}} \\
+&\frac{1}{n^2}\sum_{i,j}\underbrace{\mathbb{E}_{W,t}\left[\alpha_{t_i}Y_{t_i}(\boldsymbol{0},\cdots,\boldsymbol{0})\alpha_{t_j}\left(Y_{t_j}(\boldsymbol{W},\boldsymbol{W}_1^{\text{s}},\cdots,\boldsymbol{W}_K^{\text{s}})-Y_{t_j}(\boldsymbol{W},\boldsymbol{0},\cdots,\boldsymbol{0})\right)\right]}_{:=A_{3,ij}}.
\end{aligned}
$$

For $A_{1,ij}$, we have

$$
\begin{aligned}
A_{1,ij}=&\mathbb{E}_{W,t}\left[\frac{W_{t_i}}{\pi}\frac{W_{t_j}-\pi}{\pi(1-\pi)}(\delta_{t_i}^{\text{inst}}-\delta^{\text{inst}})\left(Y_{t_j}(\boldsymbol{W},\boldsymbol{W}_1^{\text{s}},\cdots,\boldsymbol{W}_K^{\text{s}})-Y_{t_j}(\boldsymbol{W},\boldsymbol{0},\cdots,\boldsymbol{0})\right)\right] \\
=&\mathbb{E}_{W,t}\left[\frac{W_{t_i}}{\pi}\frac{W_{t_j}-\pi}{\pi(1-\pi)}(\delta_{t_i}^{\text{inst}}-\delta^{\text{inst}})\delta_{t_j}^{\text{simul}}(\boldsymbol{W})\right] \\
=&\frac{1}{\pi}\sum_{m=1}^{M}\int_{t_i,t_j\in\mathcal{I}_m}(\delta_{t_i}^{\text{inst}}-\delta^{\text{inst}})\mathbb{E}_{\boldsymbol{W}^{(-m)}}\left[\delta_{t_j}^{\text{simul}}(\boldsymbol{W}^{(-m)},W^{(m)}=1)\right]f(t_i)f(t_j)dt_idt_j \\
&+\sum_{m=1}^{M}\sum_{m':m'\neq m}\int_{t_i\in\mathcal{I}_m,t_j\in\mathcal{I}_{m'}}(\delta_{t_i}^{\text{inst}}-\delta^{\text{inst}})\Phi_{t_j}^{\text{simul},(-m')\dagger}f(t_i)f(t_j)dt_idt_j \\
=&2\sum_{m=1}^{M}\Xi^{\text{inst},(m)}\int_{t_j\in\mathcal{I}_m}\mathbb{E}_{\boldsymbol{W}^{(-m)}}\left[\delta_{t_j}^{\text{simul}}(\boldsymbol{W}^{(-m)},W^{(m)}=1)\right]f(t_j)dt_j \\
&+2\sum_{m=1}^{M}\Xi^{\text{inst},(m)}\sum_{m':m'\neq m}\int_{t_j\in\mathcal{I}_{m'}}\Phi_{t_j}^{\text{simul},(-m')\dagger}f(t_j)dt_j \\
=&\sum_{m=1}^{M}\sum_{m'=1}^{M}\Xi^{\text{inst},(m)}S_2^{(m,m')},
\end{aligned}
$$

following the definition of $S_2^{(m,m')}$ in Equation (A.4) and

$$\Phi_{t_j}^{\text{simul},(-m')\dagger} = \frac{1}{2}\Bigg(\mathbb{E}_{\boldsymbol{W}^{(-m,-m')}}\left[\delta_{t_j}^{\dagger}(\boldsymbol{W}^{(-m,-m')}, \boldsymbol{W}^{(m,m')} = (1,1))\right] - \mathbb{E}_{\boldsymbol{W}^{(-m,-m')}}\left[\delta_{t_j}^{\dagger}(\boldsymbol{W}^{(-m,-m')}, \boldsymbol{W}^{(m,m')} = (1,0))\right]\Bigg).$$

For $A_{2,ij}$, we have

$$A_{2,ij} = \underbrace{\mathbb{E}_{W,t}\left[\frac{W_{t_i}-\pi}{\pi(1-\pi)}\frac{W_{t_j}-\pi}{\pi(1-\pi)}\delta_{t_i}^{\text{co}}(\boldsymbol{W})\delta_{t_j}^{\text{simul}}(\boldsymbol{W})\right]}_{:=B_{1,ij}} - \underbrace{\delta^{\text{co}}\mathbb{E}_{W,t}\left[\frac{W_{t_i}}{\pi}\frac{W_{t_j}-\pi}{\pi(1-\pi)}\delta_{t_j}^{\text{simul}}(\boldsymbol{W})\right]}_{:=B_{2,ij}}.$$

For $B_{1,ij}$, we have

$$\begin{aligned} B_{1,ij} =& \mathbb{E}_{W,t}\left[\frac{W_{t_i}-\pi}{\pi(1-\pi)}\frac{W_{t_j}-\pi}{\pi(1-\pi)}\delta_{t_i}^{\text{co}}(\boldsymbol{W})\delta_{t_j}^{\text{simul}}(\boldsymbol{W})\right] \\ =& 4\sum_{m=1}^{M}\int_{t_i,t_j\in\mathcal{I}_m}\mathbb{E}_W\left[\delta_{t_i}^{\text{co}}(\boldsymbol{W})\delta_{t_j}^{\text{simul}}(\boldsymbol{W})\right]f(t_i)f(t_j)dt_idt_j \qquad (\pi = 1/2) \\ &+ 4\sum_{m=1}\sum_{m':m'\neq m}\int_{t_i\in\mathcal{I}_m, t_j\in\mathcal{I}_{m'}}\Phi_{t_i,t_j}^{\text{co,simul}}f(t_i)f(t_j)dt_idt_j \\ =& \sum_{m=1}^{M}\sum_{m'=1}^{M}S_3^{(m,m')}, \end{aligned}$$

following the definition of $S_3^{(m,m')}$ in Equation (A.5) and

$$\begin{aligned} \Phi_{t_i,t_j}^{\text{co,simul}} =& \frac{1}{4}\Bigg(\mathbb{E}_{\boldsymbol{W}^{(-m,-m')}}\left[\delta_{t_i}^{\text{co}}(\boldsymbol{W}^{(-m,-m')}, \boldsymbol{W}^{(m,m')} = (1,1))\cdot\delta_{t_j}^{\text{simul}}(\boldsymbol{W}^{(-m,-m')}, \boldsymbol{W}^{(m,m')} = (1,1))\right] \\ &- \mathbb{E}_{\boldsymbol{W}^{(-m,-m')}}\left[\delta_{t_i}^{\text{co}}(\boldsymbol{W}^{(-m,-m')}, \boldsymbol{W}^{(m,m')} = (1,0))\cdot\delta_{t_j}^{\text{simul}}(\boldsymbol{W}^{(-m,-m')}, \boldsymbol{W}^{(m,m')} = (1,0))\right] \\ &- \mathbb{E}_{\boldsymbol{W}^{(-m,-m')}}\left[\delta_{t_i}^{\text{co}}(\boldsymbol{W}^{(-m,-m')}, \boldsymbol{W}^{(m,m')} = (0,1))\cdot\delta_{t_j}^{\text{simul}}(\boldsymbol{W}^{(-m,-m')}, \boldsymbol{W}^{(m,m')} = (0,1))\right] \\ &+ \mathbb{E}_{\boldsymbol{W}^{(-m,-m')}}\left[\delta_{t_i}^{\text{co}}(\boldsymbol{W}^{(-m,-m')}, \boldsymbol{W}^{(m,m')} = (0,0))\cdot\delta_{t_j}^{\text{simul}}(\boldsymbol{W}^{(-m,-m')}, \boldsymbol{W}^{(m,m')} = (0,0))\right]\Bigg). \end{aligned}$$

For $B_{2,ij}$, we have

$$\begin{aligned} B_{2,ij} =& \delta^{\text{co}}\mathbb{E}_{W,t}\left[\frac{W_{t_i}}{\pi}\frac{W_{t_j}-\pi}{\pi(1-\pi)}\delta_{t_j}^{\text{simul}}(\boldsymbol{W})\right] \\ =& 2\delta^{\text{co}}\sum_{m=1}^{M}\mu^{(m)}\int_{t_j\in\mathcal{I}_m}\mathbb{E}_{\boldsymbol{W}^{(-m)}}\left[\delta_{t_j}^{\text{simul}}(\boldsymbol{W}^{(-m)}, W^{(m)} = 1)\right]f(t_j)dt_j \\ &+ 2\delta^{\text{co}}\sum_{m=1}^{M}\mu^{(m)}\sum_{m':m'\neq m}\int_{t_j\in\mathcal{I}_{m'}}\Phi_{t_j}^{\text{simul},(-m')\dagger}f(t_j)dt_j \\ =& \sum_{m=1}^{M}\sum_{m'=1}^{M}\delta^{\text{co}}\mu^{(m)}S_2^{(m,m')}. \end{aligned}$$

For $A_{3,ij}$, we have

$$\begin{aligned} A_{3,ij} =& \mathbb{E}_{W,t}\left[\alpha_{t_i}Y_{t_i}(\mathbf{0},\cdots,\mathbf{0})\alpha_{t_j}\left(Y_{t_j}(\boldsymbol{W}, \boldsymbol{W}_1^{\text{s}},\cdots,\boldsymbol{W}_K^{\text{s}}) - Y_{t_j}(\boldsymbol{W},\mathbf{0},\cdots,\mathbf{0})\right)\right] \\ =& \mathbb{E}_{W,t}\left[\frac{W_{t_i}-\pi}{\pi(1-\pi)}\frac{W_{t_j}-\pi}{\pi(1-\pi)}Y_{t_i}(\mathbf{0},\cdots,\mathbf{0})\left(Y_{t_j}(\boldsymbol{W}, \boldsymbol{W}_1^{\text{s}},\cdots,\boldsymbol{W}_K^{\text{s}}) - Y_{t_j}(\boldsymbol{W},\mathbf{0},\cdots,\mathbf{0})\right)\right] \end{aligned}$$

$$
\begin{aligned}
=&4\sum_{m=1}^{M}\int_{t_i,t_j\in\mathcal{I}_m} Y_{t_i}(\mathbf{0},\cdots,\mathbf{0})\mathbb{E}_W\left[\delta_{t_j}^{\text{simul}}(\boldsymbol{W})\right]f(t_i)f(t_j)dt_idt_j\\
&+4\sum_{m=1}^{M}\sum_{m':m'\neq m}\int_{t_i\in\mathcal{I}_m,t_j\in\mathcal{I}_{m'}} Y_{t_i}(\mathbf{0},\cdots,\mathbf{0})\Phi_{t_j}^{\text{simul},(-m)}f(t_i)f(t_j)dt_idt_j\\
=&4\sum_{m=1}^{M}\mu_{Y^{\text{ctrl}}}^{(m)}\int_{t_j\in\mathcal{I}_m}\mathbb{E}_W\left[\delta_{t_j}^{\text{simul}}(\boldsymbol{W})\right]f(t_j)dt_j+4\sum_{m=1}^{M}\mu_{Y^{\text{ctrl}}}^{(m)}\sum_{m':m'\neq m}\int_{t_j\in\mathcal{I}_{m'}}\Phi_{t_j}^{\text{simul},(-m)}f(t_j)dt_j\\
=&2\sum_{m=1}^{M}\sum_{m'=1}^{M}\mu_{Y^{\text{ctrl}}}^{(m')}S_1^{(m,m')},
\end{aligned}
$$

following the definition of $S_1^{(m,m')}$ in Equation A.3. Then we have

$$
\begin{aligned}
&\mathbb{E}_{W,\varepsilon,t}\left[\mathcal{E}_{\text{simul}}\mathcal{E}_{\text{carryover}}\right]\\
=&A_{1,ij}+\underbrace{B_{1,ij}-B_{2,ij}}_{A_{2,ij}}+A_{3,ij}\\
=&\sum_{m=1}^{M}\sum_{m'=1}^{M}\Xi^{\text{inst},(m)}S_2^{(m,m')}+\sum_{m=1}^{M}\sum_{m'=1}^{M}S_3^{(m,m')}-\sum_{m=1}^{M}\sum_{m'=1}^{M}\delta^{\text{co}}\mu^{(m)}S_2^{(m,m')}+2\sum_{m=1}^{M}\sum_{m'=1}^{M}\mu_{Y^{\text{ctrl}}}^{(m')}S_1^{(m,m')}\\
=&\sum_{m=1}^{M}\sum_{m'=1}^{M}\left[2\mu_{Y^{\text{ctrl}}}^{(m')}S_1^{(m,m')}+\left(\Xi^{\text{inst},(m)}-\delta^{\text{co}}\mu^{(m)}\right)S_2^{(m,m')}+S_3^{(m,m')}\right].
\end{aligned}
$$

We then finish the proof of Lemma B.16. □

Proof of Theorem 4.2 Following Lemmas B.11, B.12, and B.13, we have

$$
\begin{aligned}
&\mathbb{E}_{W,\varepsilon,t}\left[\left(\mathcal{E}_{\text{inst}}+\mathcal{E}_{\text{carryover}}\right)^2\right]=\mathbb{E}_{W,\varepsilon,t}\left[\left(\mathcal{E}_{\text{inst}}\right)^2\right]+\mathbb{E}_{W,\varepsilon,t}\left[\left(\mathcal{E}_{\text{carryover}}\right)^2\right]+2\mathbb{E}_{W,\varepsilon,t}\left[\mathcal{E}_{\text{inst}}\mathcal{E}_{\text{carryover}}\right]\\
=&\sum_{m=1}^{M}\left(\Xi^{(m)}+2\mu_{Y^{\text{ctrl}}}^{(m)}\right)^2+\left(\sum_{m=1}^{M}I^{(m)}-\delta^{\text{co}}\right)^2+\sum_{m=1}^{M}\sum_{m'\neq m}\left(\left[I^{(m,m')}\right]^2+I^{(m,m')}I^{(m',m)}\right),
\end{aligned}
$$

where $\Xi^{(m)}=\Xi^{\text{inst},(m)}+\Xi^{\text{co},(m)}+\delta^{\text{gate}}\mu^{(m)}$.

Following Lemmas B.6-B.16, the mean-squared error of $\hat{\delta}^{\text{gate}}$ equals to

$$
\begin{aligned}
\mathbb{E}_{W,\varepsilon,t}\left[\left(\hat{\delta}^{\text{gate}}-\delta^{\text{gate}}\right)^2\right]=&\frac{4}{n}\sum_{m=1}^{M}\left(V^{(m)}+(n-1)C^{(m)}\right) && \left(\text{equals to } \mathbb{E}_{W,\varepsilon,t}\left[\left(\mathcal{E}_{\text{meas}}\right)^2\right]\right)\\
&+\sum_{m=1}^{M}\left(\Xi^{(m)}+2\mu_{Y^{\text{ctrl}}}^{(m)}\right)^2+\left(\sum_{m=1}^{M}I^{(m)}-\delta^{\text{co}}\right)^2+\sum_{m=1}^{M}\sum_{m'\neq m}\left(\left[I^{(m,m')}\right]^2+I^{(m,m')}I^{(m',m)}\right)\\
& && \left(\text{equals to } \mathbb{E}_{W,\varepsilon,t}\left[\left(\mathcal{E}_{\text{inst}}+\mathcal{E}_{\text{carryover}}\right)^2\right]\right)\\
&+\sum_{m=1}^{M}\sum_{m'=1}^{M}S_{\text{var}}^{(m,m')} && \left(\text{equals to } \mathbb{E}_{W,\varepsilon,t}\left[\left(\mathcal{E}_{\text{simul}}\right)^2\right]\right)\\
&+2\sum_{m=1}^{M}\sum_{m'=1}^{M}\delta^{\text{gate}}\mu^{(m')}S_1^{(m,m')} && \left(\text{equals to } 2\mathbb{E}_{W,\varepsilon,t}\left[\mathcal{E}_{\text{simul}}\mathcal{E}_{\text{inst}}\right]\right)\\
&+2\sum_{m=1}^{M}\sum_{m'=1}^{M}\left[2\mu_{Y^{\text{ctrl}}}^{(m')}S_1^{(m,m')}+\left(\Xi^{\text{inst},(m)}-\delta^{\text{co}}\mu^{(m)}\right)S_2^{(m,m')}+S_3^{(m,m')}\right]\\
& && \left(\text{equals to } 2\mathbb{E}_{W,\varepsilon,t}\left[\mathcal{E}_{\text{simul}}\mathcal{E}_{\text{carryover}}\right]\right)\\
=&\operatorname{Var}(\mathcal{E}_{\text{meas}})+\operatorname{Bias}(\mathcal{E}_{\text{carryover}})^2+\operatorname{Var}(\mathcal{E}_{\text{inst}}+\mathcal{E}_{\text{carryover}})+\mathbb{E}[\mathcal{E}_{\text{simul}}^2]+2\mathbb{E}[(\mathcal{E}_{\text{inst}}+\mathcal{E}_{\text{carryover}})\cdot\mathcal{E}_{\text{simul}}].
\end{aligned}
$$

We then finish the proof of Theorem 4.2. □