A Balanced Design of Time Series Experiments

Anonymous Author(s)
Affiliation
Address
email

Abstract

Time series experiments are a family of experimental designs on a time series. One experimental unit is sequentially exposed to some version of treatment, stays in the version of treatment for a duration of time, and gets exposed to another version of treatment. While this type of experimental designs could handle population interference between units, it typically still needs to account for temporal interference, i.e., a treatment at an earlier period persists in impacting the outcomes of the later periods. Practitioners have widely recognized the applicability of the time series experiments, yet prior work typically requires a long duration to gain enough power. In this paper, we propose a novel randomized design that significantly increases the power of such experiment. We prove the theoretical performance of the novel design, and verify its superior performance by conducting an extensive simulation study.

1 Introduction

Time series experiments — where we expose one aggregated unit to random treatments, measure the responses, and repeat the procedure for some time periods — have rapidly grown in popularity [5]. This growth has been partly propelled by marketplace companies, such as DoorDash [29], Lyft [7], and Uber [9], wishing to run experiments in the presence of population interference (the setting where one unit's treatment assignments impact another's outcomes). To overcome population interference, companies aggregate multiple units together across zip codes, cities, or even states to form a single grouped unit [6, 12, 16, 18]. The aggregation alleviates the population interference, as it ensures that each unit within the aggregated unit receives the same treatment, but it does not remove the temporal interference (where past treatment assignments impact current outcomes). Unfortunately, such a drastic aggregation significantly reduces the sample size as we only collect one data point at each period, leading to much lower statistical power for estimating causal effects. Therefore, it is vital that the design of time series experiments has a high data-efficiency – not too many time periods are required to increase the power of estimating the causal effect of interests.

In this work, we present a new, more powerful, randomized design for time series experiments subject to temporal interference. Our proposed design incorporates a randomization mechanism inspired by the classical completely randomized design to achieve the balancing between treatment and control observations. We analyze the theoretical property of our balanced design, develop efficient inferential methods and evaluate the performance using simulation study.

2 Setups and notations

2.1 The assignment path and the potential outcomes

Consider an experimenter running time series experiments over one large aggregated unit. Let there be a total of $T \in \mathbb{N}$ periods in the experiment. As suggested by earlier works [6, 29], each period is

Submitted to *A causal view on dynamical systems* workshop at NeurIPS 2022. Do not distribute.

typically selected to be the same length as the carryover effect, i.e., the length of time a treatment persists in impacting the outcome. For on-demand service platforms, one period typically ranges from about 30 minutes to several hours [7, 29]. At each time period $t \in [T]$, the experimenter randomly exposes the unit to either treatment $W_t = 1$ or control $W_t = 0$. The assignment path is the collection of treatment assignments over all time periods, $\boldsymbol{W}_{1:T} \in \{0,1\}^T$. We denote a realization of $\boldsymbol{W}_{1:T}$ by $\boldsymbol{w}_{1:T}$.

Once the assignment path $\boldsymbol{w}_{1:T}$ is realized, the experimenter will observe some outcomes of interests. Under the potential outcomes framework [25, 14, 17], we model the observed outcomes to be related to their respective potential outcomes. We denote $Y_t(\boldsymbol{w}_{1:T})$ to be the outcome at time period $t \in [T]$ under the assignment path $\boldsymbol{w}_{1:T}$. As a short-hand notation, denote $\mathbb{Y} = \{Y_t(\boldsymbol{w}_{1:T})\}_{t,\boldsymbol{w}_{1:T}}$ to be the collection of all potential outcomes. We adopt a design-based perspective [23, 11, 19, 26, 17, 1] and treat the potential outcomes as fixed quantities.

The potential outcome $Y_t(\boldsymbol{w}_{1:T})$ is very general in the sense that it depends on the complete assignment path. If experimenter directly observes the outcome $Y_t(\boldsymbol{w}_{1:T})$ under one assignment path $\boldsymbol{w}_{1:T}$, it becomes impossible to observe the potential outcomes under other assignment paths. In other words, all the remaining data are missing. Since we do not assume any structural models for characterizing the outcomes, it is impossible to achieve valid inference with such missing data. Therefore, we need to introduce some assumptions that restrict the dependence of the potential outcomes.

2.2 Temporal interference

In this section we introduce the assumption about the temporal interference, which lays the foundations for our design of time series experiments. We consider a form of temporal interference between time periods, which is also known as the *carryover effects*. We assume that the potential outcome at any time period depends only on the treatment assignments at this time period and the preceding time period, but not the earlier periods.

Assumption 1 (Limited carryover effects)**.** *For any $t \in [T]$ and any two assignment paths $\boldsymbol{w}_{1:T}, \boldsymbol{w}'_{1:T} \in \{0,1\}^T$, we have*

$$Y_t\left(\boldsymbol{w}_{1:T}\right) = Y_t\left(\boldsymbol{w}'_{1:T}\right) \quad \textit{whenever} \quad \boldsymbol{w}_{t-1:t} = \boldsymbol{w}'_{t-1:t}.$$

This assumption is both widely adopted in the literature [20, 27, 2, 5, 6] and viable in many applications. We can relax this assumption by considering general lengths of carryover effects, i.e., the potential outcomes at one time period could depend on the treatment assignments up to $m > 1$ periods ago. The structure of the main results remains the same. In this paper, we focus on the case of $m = 1$ to de-emphasize the importance of the length of carryover effects to our main results, as the existing literature [6, 29] recommends selecting the length of one time period to be the same as the length of the carryover effect. In a ride-sharing platform, the effect of a surge pricing policy quickly vanishes and for instance, the carryover effect lasts 30 - 60 minutes. We would set the length of a period to be one hour so that the length of the carryover effect is one period. We also refer to [**?**] for more discussions about the identification of the length of carryover effects as well as a data driven strategy to select a plausible carryover effect.

In the remaining of this paper, we use the short-hand notation $Y_t\left(\boldsymbol{w}_{t-1:t}\right) = Y_t\left(\boldsymbol{w}_{1:T}\right)$ to focus on the dependence of the potential outcomes at two consecutive periods. Using this short-hand notation, the observed outcomes are related to their respective potential outcomes as follows,

$$Y_t = Y_t(\boldsymbol{w}_{t-1:t}), \quad \text{if} \quad \boldsymbol{W}_{t-1:t} = \boldsymbol{w}_{t-1:t}.$$

2.3 The causal effect

The primary focus of this paper is to understand the time-averaged *Total Treatment Effect*, i.e., the difference between the average outcomes when the aggregated unit is exposed to treatment at all time periods, relative to when it is exposed to control at all time periods [6]. Mathematically, this is defined as

$$\tau(\mathbb{Y}) = \frac{1}{T-1} \sum_{t=2}^{T} \left[Y_t\left(\mathbf{1}\right) - Y_t\left(\mathbf{0}\right) \right], \tag{1}$$

where $\mathbf{1}$ and $\mathbf{0}$ are two assignment paths with all treatments $\boldsymbol{w}_{1:T} = \mathbf{1}_{1:T}$ and controls $\boldsymbol{w}_{1:T} = \mathbf{0}_{1:T}$, respectively. Based on Assumption 1, they can be simplified as $\boldsymbol{w}_{t-1:t} = \mathbf{1}_{t-1:t}$ and $\boldsymbol{w}_{t-1:t} = \mathbf{0}_{t-1:t}$ accordingly.

The causal effect as defined in (1) captures the effect of permanently deploying a new policy. In the Lyft example [7], this could reflect the change in average ride-making rates if the company rolls out a new subsidy policy across all local regions in a market. Such a causal estimands directly helps data scientists evaluate the benefits of permanently deploying such a policy. Since the causal effect $\tau(\mathbb{Y})$ is never directly observable, our goal is to estimate the casual effect using observations from the time series experiments efficiently, which requires a careful design of experiments.

3 Problem formulation

As introduced in Section 2.1, in this paper we focus on a design-based perspective of experimental design, and rely on variations introduced by the random assignment path for doing inference. In this section we study, as a decision-making problem, the design of time series experiments that governs the selection of the random assignment path to gain high statistical power for inference.

Formally, the *design of an experiment* is a discrete probability distribution $\eta(\cdot) : \{0,1\}^T \to [0,1]$ over the assignment paths, such that,

$$\sum_{\boldsymbol{w}_{1:T}} \eta(\boldsymbol{w}_{1:T}) = 1, \quad \eta(\boldsymbol{w}_{1:T}) \geq 0, \quad \forall\, \boldsymbol{w}_{1:T} \in \{0,1\}^T.$$

In practice, the experimenter would sample an assignment path $\boldsymbol{W}_{1:T} = \boldsymbol{w}_{1:T}$ from the distribution $\eta(\cdot)$, and then implements this assignment path to conduct an experiment. Once the experiment has been conducted, the experimenter collects the observed outcomes $\{Y_t\}_{t\in[T]}$, and uses both the realized assignment path and the observed outcomes to estimate the causal effect.

A commonly used estimator is the Inverse Propensity Weighted (IPW) estimator, which is also referred to as the Horvitz-Thompson estimator [15], as follows,

$$\widehat{\tau}(\mathbb{Y}, \eta, \boldsymbol{w}) = \frac{1}{T-1} \sum_{t=2}^{T} \left\{ Y_t \, \frac{\mathbb{1}\{\boldsymbol{w}_{t-1:t} = \mathbf{1}\}}{\Pr(\boldsymbol{W}_{t-1:t} = \mathbf{1})} - Y_t \, \frac{\mathbb{1}\{\boldsymbol{w}_{t-1:t} = \mathbf{0}\}}{\Pr(\boldsymbol{W}_{t-1:t} = \mathbf{0})} \right\}. \tag{2}$$

It is a well-known result that the IPW estimator is unbiased [5], i.e., $\mathbb{E}_\eta[\widehat{\tau}(\mathbb{Y}, \eta, \boldsymbol{w})] = \tau(\mathbb{Y})$.

To evaluate the quality of a specific $\eta(\cdot)$, we adopt the decision-theoretic framework [3, 4] and focus on the variance of the IPW estimator, $\mathrm{Var}_\eta(\widehat{\tau}(\mathbb{Y}, \eta, \boldsymbol{w})) = \mathbb{E}_\eta[(\widehat{\tau}(\mathbb{Y}, \eta, \boldsymbol{w}) - \tau(\mathbb{Y}))^2]$. Since the IPW estimator is unbiased, the variance of the estimator is equivalent to the risk function, or the mean squared error of the estimator. Since the variance of the IPW estimator depends on the potential outcomes $\mathbb{Y}$, there may not exist one design of experiment that uniformly achieves a small variance in all scenarios. To determine the design of experiment, we adopt the minimax decision rule [3, 30, 21, 6] and find the design of experiment that minimizes the worst-case variance against an adversarial selection of potential outcomes, i.e.,

$$\min_{\eta} \max_{\mathbb{Y}\in\mathcal{Y}} \mathrm{Var}_\eta(\widehat{\tau}(\mathbb{Y}, \eta, \boldsymbol{w})) = \min_{\eta} \max_{\mathbb{Y}\in\mathcal{Y}} \mathbb{E}_\eta\left[(\widehat{\tau}(\mathbb{Y}, \eta, \boldsymbol{W}) - \tau(\mathbb{Y}))^2\right]. \tag{3}$$

One compelling reason of adopting the minimax decision rule is that we do not impose any parametric or structural models on the potential outcomes. Yet to make the decision making problem feasible, and to lay the foundation for inference, we impose a bounded support assumption of the potential outcomes.

Assumption 2 (Bounded potential outcomes)**.** *There exists $B > 0$, such that for any $t \in [T], \boldsymbol{w}_{1:T} \in \{0,1\}^T, Y_t(\boldsymbol{w}_{1:t}) \in [0, B]$. Equivalently, $\mathcal{Y} = [0, B]^T$.*

Assumption 2 is typically satisfied in practice as it assumes that the set of potential outcomes is non-negative and upper bounded by some possibly large constant. Note that this assumption is satisfied even if the potential outcomes arise from a stochastic process, as long as the process does not have a point mass at infinity. As we will show in the next section, our experimental design does not require the knowledge of the upper bound B, just its existence.

4 The design of time series experiments

4.1 A balanced design

In this section, we study the design of time series experiments. Solving the optimization problem defined by (3) is typically challenging, as the inner problem is a maximization over a convex function whose optimal solution lies on the exponentially many extreme points of the polyhedron $\mathcal{Y}$, and the outer problem is hindered by the challenges of evaluating the inverse probabilities in the IPW estimator. Even if the optimization problem can be numerically solved, inference and testing typically require a clear understanding of the randomization mechanism of the design, which a numerical solution fails to provide. Furthermore, from a practical perspective, the randomization mechanism needs to be easy to implement when the design of the experiment is brought to a company.

Instead of finding the exact optimal solution, we propose a balanced design that has a sub-optimal theoretical performance and is easy to implement in practice. Suppose T is odd. Consider the following balanced design $\eta^\dagger$ of the time series experiments:

1. During the first $T-1$ periods, conduct a complete randomization with a equal number of treatments and controls. That is, $\sum\limits_{t=1}^{T-1} \mathbb{1}\{W_t = 1\} = \sum\limits_{t=1}^{T-1} \mathbb{1}\{W_t = 0\} = \frac{T-1}{2}$.
2. Set the treatment assignment of the last period to be the same as that of the first period, i.e., $W_T = W_1$.

The first step reflects the key idea of balancing inspired by the classical completely randomized design. The second step simply deals with the boundary conditions of the design. Because $W_T = W_1$, we could specify the precedent period of $t = T$ to be $t = 2$, i.e., $W_{T+1} = W_2$. This simplifies the analysis later a lot without making the design significantly different.

We illustrate in Figure 1 the example of two assignment path realizations with $T = 9$. This design

Periods	1	2	3	4	5	6	7	8	9
Assignment path 1	1	1	0	0	0	1	1	0	1
Assignment path 2	0	1	1	1	0	0	1	0	0

Figure 1: An example of assignment path realizations from the balanced design $\eta^\dagger$ with $T = 9$.

extends the classical completely randomized experiment to the temporal interference setup. A classical completely randomized experiment [10, 22] fixes the number of treatments and controls in advance before randomization is applied, ensuring that the number of valid observations under treatment and control is equal to the number of units who receive treatment and control, respectively. However, under the temporal interference setup that we consider, whether $Y_t(\boldsymbol{w}_{t-1:t})$ is a valid observation at period t depends on the treatment assignments during both periods $t-1$ and t (e.g., Y_6 under the assignment path 1 in Figure 1 is not a valid observation). This suggests a careful analysis for the underlying dependence of the design. To this end, we first introduce some statistics of interest which are involved in our analysis. Define

$$\overline{Y}(\mathbf{1}) = \frac{1}{T-1}\sum_{t=2}^{T} Y_t(\mathbf{1}), \overline{Y}(\mathbf{0}) = \frac{1}{T-1}\sum_{t=2}^{T} Y_t(\mathbf{0})$$

to be the average outcomes of treatment and control, respectively. Given these, we further define

$$S^t = \frac{1}{T-2}\sum_{t=2}^{T}\left(\frac{Y_t(\mathbf{1}) + Y_{t+1}(\mathbf{1})}{2} - \overline{Y}(\mathbf{1})\right)^2, S^c = \frac{1}{T-2}\sum_{t=2}^{T}\left(\frac{Y_t(\mathbf{0}) + Y_{t+1}(\mathbf{0})}{2} - \overline{Y}(\mathbf{0})\right)^2$$

to be the variance of the average outcomes at two consecutive periods. Similarly, we can also measure the variance of the treatment effect using

$$S^\tau = \frac{1}{T-2}\sum_{t=2}^{T}\left(\frac{Y_t(\mathbf{1}) - Y_t(\mathbf{0}) + Y_{t+1}(\mathbf{1}) - Y_{t+1}(\mathbf{0})}{2} - (\overline{Y}(\mathbf{1}) - \overline{Y}(\mathbf{0}))\right)^2.$$

These three statistics help extend the analysis of classical complete randomization by bridging the outcomes at two consecutive periods explicitly. We have the following theorem that characterizes the variance of our balanced design.

Theorem 1. *For any potential outcomes $\mathbb{Y}$, the variance of the balanced design $\eta^\dagger$ can be decomposed as*

$$\begin{aligned}\mathrm{Var}_{\eta^\dagger}(\widehat{\tau}|\mathbb{Y}) = &\frac{8(T-2)^2}{(T-3)^2(T-1)}\left(S^t + S^c\right) - \frac{(4T-10)(T-2)}{(T-3)(T-4)(T-1)}S^\tau \\ &+ \frac{1}{(T-1)(T-4)}\sum_{t=2}^{T}\left(Y_t(\mathbf{1}) - Y_t(\mathbf{0})\right)^2 \\ &- \frac{1}{(T-1)(T-4)(T-3)}\sum_{t=2}^{T}(Y_t(\mathbf{1}) - Y_t(\mathbf{0}))(Y_{t+1}(\mathbf{1}) - Y_{t+1}(\mathbf{0})). \end{aligned} \quad (4)$$

We prove Theorem 1 in Appendix A.1. Later in Section 5, we will further develop a variance upper bound and a corresponding estimate for doing inference.

4.2 Performance analysis

Let $\overline{\mathrm{Var}}_{\eta^*}(\widehat{\tau})$ denote the worst-case variance of an optimal design η^* which solves (3) exactly. In this section, we analyze the performance of the balanced design $\eta^\dagger$ by first comparing its worst-case variance $\overline{\mathrm{Var}}_{\eta^\dagger}(\widehat{\tau})$ to $\overline{\mathrm{Var}}_{\eta^*}(\widehat{\tau})$, the worst-case variance of the optimal design.

The following result shows that the worst-case outcomes against the balanced design $\eta^\dagger$ can be characterized explicitly, as can the worst-case variance.

Theorem 2. *The worst-case outcomes against the balanced design $\eta^\dagger$ can be written as*

$$Y_t(\mathbf{1}) = Y_t(\mathbf{0}) = \begin{cases} B & 2 \le t \le \frac{T+1}{2}, \\ 0 & t > \frac{T+1}{2}. \end{cases} \quad (5)$$

This leads to the worst-case variance

$$\overline{\mathrm{Var}}_{\eta^\dagger}(\widehat{\tau}) = \frac{4(T-2)}{(T-1)(T-3)}B^2. \quad (6)$$

We prove Theorem 2 in Appendix A.2 and outline here the key ideas of the proof. The proof shows that, given any other potential outcomes, one can always transform the outcomes to the ones that match the above structure, obtaining no smaller variance.

Since solving (3) for obtaining the worst-case variance of the optimal design $\overline{\mathrm{Var}}_{\eta^*}(\widehat{\tau})$ is difficult, we then construct its lower bound as follows.

Theorem 3. *The worst-case variance of the optimal design $\overline{\mathrm{Var}}_{\eta^*}(\widehat{\tau})$ has a lower bound*

$$\overline{\mathrm{Var}}_{\eta^*}(\widehat{\tau}) \ge V^{LB} = \frac{2}{T-3}B^2.$$

We prove Theorem 3 in Appendix A.3. This lower bound directly implies a 2-approximation ratio for the balanced design.

Proposition 4. *The balanced design $\eta^\dagger$ has an approximation ratio*

$$\frac{\overline{\mathrm{Var}}_{\eta^\dagger}(\widehat{\tau})}{\overline{\mathrm{Var}}_{\eta^*}(\widehat{\tau})} \le \frac{2(T-2)}{T-1} \le 2.$$

We further compare the balanced design $\eta^\dagger$ to another Bernoulli design studied in [**?**]. The optimal Bernoulli design $\eta^\#$ of time series experiments was developed, where the experimenter can only draw i.i.d. Bernoulli trials at some time periods to determine the random assignments. Theorem 2 in [6] shows that the optimal Bernoulli design $\eta^\#$ has the worst-case variance $\overline{\mathrm{Var}}_{\eta^\#}(\widehat{\tau}) = \frac{16T-56}{(T-1)^2}B^2$.

Proposition 5. *The relative performance of the worst-case variances between the balanced design $\eta^\dagger$ and the optimal Bernoulli design $\eta^\#$ is given by*

$$\lim_{T\to\infty}\frac{\overline{\mathrm{Var}}_{\eta^\dagger}(\widehat{\tau})}{\overline{\mathrm{Var}}_{\eta^\#}(\widehat{\tau})} = \lim_{T\to\infty}\frac{(T-2)(T-1)}{(4T-14)(T-3)} = \frac{1}{4}.$$

4.3 Simulation study

In this section, we conduct a simulation study to investigate the general performance of our balanced design $\eta^\dagger$. First of all, we set the outcomes $\mathbb{Y}$ to follow the worst-case structure in (2). For each numerical experiment, we randomly sample an assignment path, compute the Horvitz-Thompson estimator (2), and repeat the procedure 10000 times to estimate the performance of the design. In Figure 2, the variance of the balanced design is significantly lower than that of the optimal Bernoulli design. Note that both designs are evaluated under the outcomes in (2), which correspond to the worst-case scenario for the balanced design, but do not correspond to the worst-case scenario for the Bernoulli design. This further justifies the robustness of the balanced design.

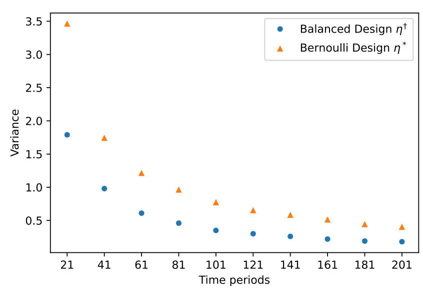


Figure 2: The estimated variance for different experimental duration under the outcomes given in (2) with $B = 3$.

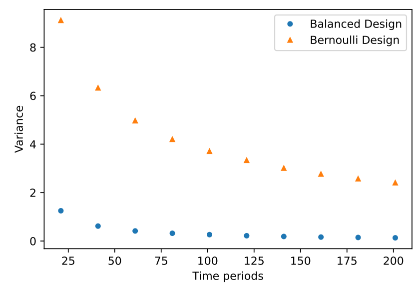


Figure 3: The estimated variance for different experimental duration under the outcomes given in (7) with $\alpha_t = \log(t), \beta_0 = 0.5, \beta_1 = 0.5, \epsilon_t \sim \mathcal{N}(0, 1)$.

Next, we consider the following outcome model that is studied in [6]:

$$Y_t(\boldsymbol{w}_{t-1:t}) = \alpha_t + \beta_0 w_t + \beta_1 w_{t-1} + \epsilon_t \tag{7}$$

where $\epsilon_t \sim \mathcal{N}(0, 1)$. Here α_t depicts the base structure of the time series, β_0 governs the direct causal effect of the treatment, and β_1 governs the carryover effect of the treatment. The causal effect of interests is $\tau = \beta_0 + \beta_1$. We first let $\alpha_t = \log(t), \beta_0 = 0.5, \beta_1 = 0.5$. In Figure 3, the balanced design dominates the Bernoulli design. Our new design is an order of magnitude better than the previous Bernoulli style design; for example, the balanced design with $T = 21$ has a lower variance than the Bernoulli design with $T = 201$. This implies that the benefit of the balanced design could be more significant beyond the worst-case scenario. Moreover, we test different outcome models by changing the parameters in (7) and layout the corresponding variances in Table 1.

Table 1: Variances under different outcome models

α_t	β_0	β_1	Bernoulli design	Balanced design
$\log(t)$	1	1	3.706	0.264(-92.8%)
	1	0	2.970	0.233(-92.1%)
	1	-1	2.334	0.223(-90.5%)
$1 + \sin(\pi t/4)$	1	1	0.852	0.182(-78.6%)
	1	0	0.533	0.152(-71.5%)
	1	-1	0.314	0.143(-54.4%)

To summarize, in this section, we propose a balanced design of time series experiments and study the performance of its variance both in the worse-case perspective analytically, and in a general perspective numerically. This type of design is simple in nature, easy to implement in practice, and more importantly, effective for making inference. To see this, in the next section, we will investigate how the variance reduction is translated to the value for the inference.

5 Inference and testing

After running an experiment, we want to test whether the estimate treatment effect is systematic or due to change. To do that, we consider the following null hypothesis for the time-averaged total treatment effect and the alternative:

$$H_0 : \frac{1}{T-1}\sum_{t=2}^{T}\left[Y_t\left(\mathbf{1}\right) - Y_t\left(\mathbf{0}\right)\right] = 0, \quad H_1 : \frac{1}{T-1}\sum_{t=2}^{T}\left[Y_t\left(\mathbf{1}\right) - Y_t\left(\mathbf{0}\right)\right] \neq 0. \tag{8}$$

To test this null hypothesis, we first derive a finite population central limit theorem to approximate the distribution of the Horvitz-Thompson estimator under our design, together with a conservative variance estimation. We then conduct simulations to examine the effectiveness of inference using different designs.

5.1 Central limit theorem

Before stating our central limit theorem, we must make an assumptions that guarantees that the variance is not dominated by a small number of time periods.

Assumption 3 (Non-negligible variance)**.**

$$\mathrm{Var}_{\eta^\dagger}(\widehat{\tau}|\mathbb{Y}) = \Omega\left(\frac{1}{T}\right).$$

This assumption often holds in practice and is regularly made by researchers [6].

Theorem 6. *Under Assumptions 1 - 3, the limiting distribution of the Horvitz-Thompson estimator has an asymptotic normal distribution. That is, as $T \to +\infty$,*

$$\frac{\widehat{\tau} - \tau}{\sqrt{\mathrm{Var}_{\eta^\dagger}(\widehat{\tau}|\mathbb{Y})}} \xrightarrow{D} \mathcal{N}(0,1) \tag{9}$$

We prove Theorem 6 in Appendix A.4. The balanced design studied in this paper is inspired by the classical complete randomization, where the number of valid observations for treatment and control is fixed and known before the randomization. A generalized central limit theorem has been developed [22] to handle multiple treatments and multi-dimension outcomes. However, the number of valid observations for treatment and control becomes random under temporal interference, which requires more complex and careful analysis. We adopt the framework of the permutation statistics [8, 13, 31] to analyze the fundamental behavior of the random permutation in the balanced design.

Whenever we adopt design-based inference, the variance of the Horvitz-Thompson estimator depends on the potential outcomes of both treatment $Y_t(\mathbf{1})$ and control $Y_t(\mathbf{0})$ at all periods $t \in [T]$, as shown in (4). Therefore, to use the normal approximation for testing, we need to replace the unknown true variance $\mathrm{Var}_{\eta^\dagger}(\hat{\tau})$ by some variance estimate. Although one can always resort to the worst-case variance (2) as a conservative estimate, the observations from the experiment are not well leveraged. We aim to develop a more informative variance estimate $\hat{\sigma}^2_{\mathrm{U}}$. To this end, we introduce several sample estimates. First we define

$$\overline{Y}^{\mathrm{obs}}(\mathbf{1}) = \frac{\sum_{t=2}^{T} Y_t \cdot \mathbb{1}\{\boldsymbol{w}_{t-1:t} = \mathbf{1}\}}{\sum_{t=2}^{T} \mathbb{1}\{\boldsymbol{w}_{t-1:t} = \mathbf{1}\}}, \overline{Y}^{\mathrm{obs}}(\mathbf{0}) = \frac{\sum_{t=2}^{T} Y_t \cdot \mathbb{1}\{\boldsymbol{w}_{t-1:t} = \mathbf{0}\}}{\sum_{t=2}^{T} \mathbb{1}\{\boldsymbol{w}_{t-1:t} = \mathbf{0}\}}$$

to be the sample estimate for $\overline{Y}(\mathbf{1})$ and $\overline{Y}(\mathbf{0})$. Next we define

$$\hat{S}^t = \frac{\sum_{t=2}^{T}\left(\frac{Y_{t-1}+Y_t}{2} - \overline{Y}^{obs}(\mathbf{1})\right)^2 \cdot \mathbb{1}\{\boldsymbol{w}_{t-2:t} = \mathbf{1}\}}{\sum_{t=2}^{T} \mathbb{1}\{\boldsymbol{w}_{t-2:t} = \mathbf{1}\} - 1}, \hat{S}^c = \frac{\sum_{t=2}^{T}\left(\frac{Y_{t-1}+Y_t}{2} - \overline{Y}^{obs}(\mathbf{0})\right)^2 \cdot \mathbb{1}\{\boldsymbol{w}_{t-2:t} = \mathbf{0}\}}{\sum_{t=2}^{T} \mathbb{1}\{\boldsymbol{w}_{t-2:t} = \mathbf{0}\} - 1}$$

to be the sample estimate for S^t and S^c.

Proposition 7. *There exists an upper bound for the variance of the balanced design. That is,*

$$\mathrm{Var}_{\eta^\dagger}(\widehat{\tau}|\mathbb{Y}) \leq \mathrm{Var}^{\mathrm{U}}_{\eta^\dagger}(\widehat{\tau}|\mathbb{Y}) = \frac{8(T-2)^2}{(T-1)(T-3)^2}(S^t + S^c) + \frac{T-2}{(T-1)(T-3)(T-4)} \sum_{t=2}^{T} \left(Y_t^2(\mathbf{1}) + Y_t^2(\mathbf{0})\right). \tag{10}$$

This upper bound has an unbiased estimate

$$\hat{\sigma}^2_{\mathrm{U}} = \frac{8(T-2)^2}{(T-1)(T-3)^2}(\hat{S}^t + \hat{S}^c) + \frac{4(T-2)^2}{(T-1)(T-3)^2(T-4)} \sum_{t=2}^{T} Y_t^2 \mathbb{1}\{w_{t-1} = w_t\}. \tag{11}$$

We prove Proposition 7 in Appendix A.5. Compared to the true variance (4), the upper bound is not tight in general. Nevertheless, we will show that it admits a great inference power for testing the null hypothesis.

5.2 Simulation study

We do the simulation study using the same outcome model in (7) with $\alpha_t = \log(t)$. We aim to justify the normal approximation and test the effectiveness of the inference. Following the simulation procedure in Section 4.3, we not only calculate the Horvitz-Thompson estimator (2) based on the observed outcomes, but also calculate the conservative variance estimate (11).

5.2.1 Asymptotic normality

We first justify the normal approximation using the samples generated from the outcome model with $\beta_0 = \beta_1 = 1$ and $T = 401$. More specifically, we generate 100000 samples of the estimator $\widehat{\tau}$ and conduct a Kolmogorov–Smirnov test [28] for the null hypothesis that the samples come from a normal distribution. The test returns an estimated p-value 0.57, which implies a good normal approximation. Figure 4 shows the histogram and the Q-Q plot that correspond to the distribution induced by $\frac{\widehat{\tau}-\tau}{\sqrt{\mathrm{Var}_{\eta^\ddagger}(\widehat{\tau}|\mathbb{Y})}}$, for which we numerically compute $\mathrm{Var}_{\eta^\ddagger}(\widehat{\tau}|\mathbb{Y})$ using samples from the simulation.

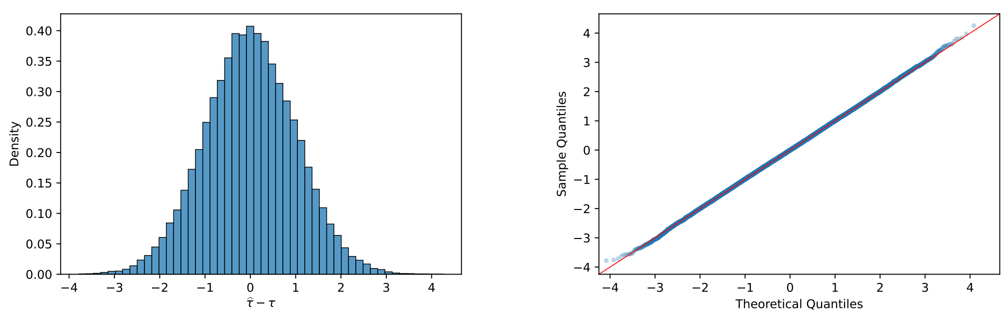


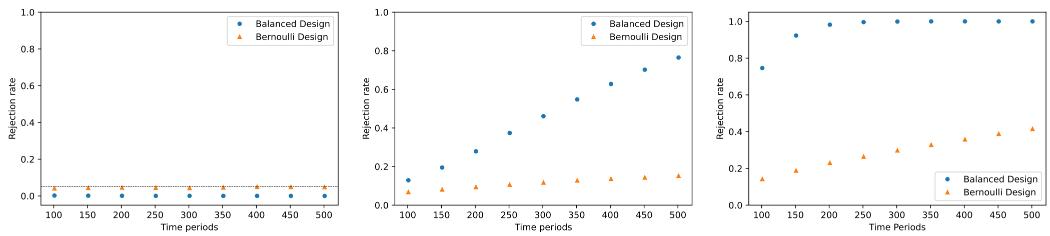
Figure 4: Normal approximation of $\widehat{\tau} - \tau$ using 100000 samples under the outcomes given in (7) with $\alpha_t = \log(t), \beta_0 = \beta_1 = 1$ and $T = 401$.

5.2.2 Rejection rate

We test the null hypothesis (8) using the normal approximation. We plug in the conservative variance estimate to obtain the estimated p-value $\hat{p}$. We reject the null hypothesis if $\hat{p} < 0.05$. By repeating this procedure 10000 times, we summarize the frequency of a null hypothesis being rejected(i.e. rejection rate).

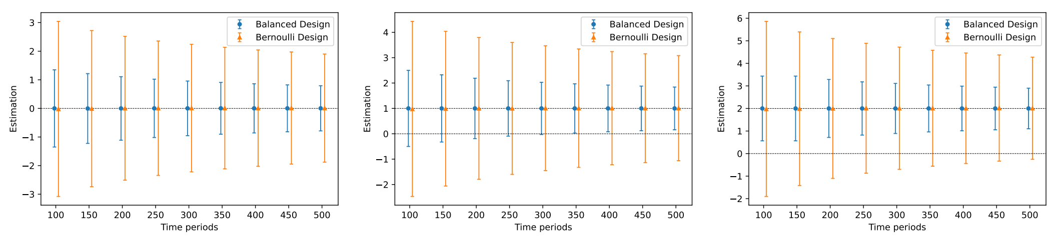
We present the rejection rates as the number of periods T grows under three outcome models in Figure 5. The balanced design leads to the right decision more efficiently than the optimal Bernoulli design in all scenarios. In particular, when there exists some degree of the casual effect $\tau = \beta_0 + \beta_1 > 0$, the balanced design only needs 20% as much time periods as in the optimal Bernoulli design to achieve the same rejection rate.

Furthermore, we plot the average point estimates and the average confidence intervals in Figure 6. We can observe that all the estimates are indeed unbiased and the balanced design consistently achieves much narrower confidence intervals. Specifically, for getting the average confidence interval above 0, the balanced design only needs 350 periods when $\tau = 1$ and less than 100 periods when $\tau = 2$. In contrast, the optimal Bernoulli needs more than 500 periods in both scenarios. This further justifies that the balanced design is more data-efficient.



$\beta_0 = 1, \beta_1 = -1$ $\beta_0 = 1, \beta_1 = 0$ $\beta_0 = 1, \beta_1 = 1$

Figure 5: Rejection rates for testing the null hypothesis (8)



$\beta_0 = 1, \beta_1 = -1$ $\beta_0 = 1, \beta_1 = 0$ $\beta_0 = 1, \beta_1 = 1$

Figure 6: Point estimates and confidence intervals for testing the null hypothesis (8)

6 Concluding remarks

This paper studied the design of time series experiments in the presence of interference. The proposed balanced design improves the casual estimator's variance leading to better inference and testing efficiency. The main results could also be extended to the design of panel experiments, where we conduct time series experiments on multiple units simultaneously.

References

[1] Alberto Abadie, Susan Athey, Guido W Imbens, and Jeffrey M Wooldridge. Sampling-based versus design-based uncertainty in regression analysis. *Econometrica*, 88(1):265–296, 2020.

[2] Guillaume Basse, Yi Ding, and Panos Toulis. Minimax crossover designs. *arXiv preprint arXiv:1908.03531*, 2019.

[3] James O Berger. *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media, 2013.

[4] Peter J Bickel and Kjell A Doksum. *Mathematical statistics: basic ideas and selected topics, volume I*, volume 117. CRC Press, 2015.

[5] Iavor Bojinov and Neil Shephard. Time series experiments and causal estimands: exact randomization tests and trading. *Journal of the American Statistical Association*, 114(528):1665–1682, 2019.

[6] Iavor Bojinov, David Simchi-Levi, and Jinglong Zhao. Design and analysis of switchback experiments. *Management Science*, 2022.

[7] N Chamandy. Experimentation in a ridesharing marketplace lyft engineering. *URL: https://eng. lyft. com/experimentation-in-a-ridesharing-marketplace-b39db027a66e*, 2016.

[8] Henry E Daniels. The relation between measures of correlation in the universe of sample permutations. *Biometrika*, 33(2):129–135, 1944.

[9] Chiara Farronato, Alan MacCormack, and Sarah Mehta. Innovation at uber: The launch of express pool. *Harvard Business School Case*, 620(062), 2018.

[10] Ronald Aylmer Fisher. Design of experiments. *Br Med J*, 1(3923):554–554, 1936.

[11] Ronald Aylmer Fisher et al. The design of experiments. *The design of experiments.*, (2nd Ed), 1937.

[12] Peter W Glynn, Ramesh Johari, and Mohammad Rasouli. Adaptive experimental design with temporal interference: A maximum likelihood approach. *Advances in Neural Information Processing Systems*, 33:15054–15064, 2020.

[13] Wassily Hoeffding. A Combinatorial Central Limit Theorem. *The Annals of Mathematical Statistics*, 22(4):558 – 566, 1951.

[14] Paul W Holland. Statistics and causal inference. *Journal of the American statistical Association*, 81(396):945–960, 1986.

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

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

[17] Guido W Imbens and Donald B Rubin. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press, 2015.

[18] David Kastelman and Raghav Ramesh. Switchback tests and randomized experimentation under network effects at doordash. *URL: https://medium.com/@DoorDash/switchback-tests-and-randomized-experimentation-under-network-effects-at-doordash-f1d938ab7c2a*, 2018.

[19] Oscar Kempthorne. The randomization theory of experimental inference. *Journal of the American Statistical Association*, 50(271):946–967, 1955.

[20] Nan M Laird, Jonathan Skinner, and Michael Kenward. An analysis of two-period crossover designs with carry-over effects. *Statistics in Medicine*, 11(14-15):1967–1979, 1992.

[21] Ker-Chau Li. Minimaxity for randomized designs: some general results. *The Annals of Statistics*, 11(1):225–239, 1983.

[22] Xinran Li and Peng Ding. General forms of finite population central limit theorems with applications to causal inference. *Journal of the American Statistical Association*, 112(520):1759–1769, 2017.

[23] Jerzy Neyman. On the application of probability theory to agricultural experiments. essay on principles. section 9. masters thesis. portion translated by d. dabrowska and t. speed (1990). *Statistical Science*, pages 465–472, 1923.

[24] Gesine Reinert and Adrian Röllin. Multivariate normal approximation with stein's method of exchangeable pairs under a general linearity condition. *The Annals of Probability*, 37(6):2150–2173, 2009.

[25] Donald B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688, 1974.

[26] Donald B Rubin. Randomization analysis of experimental data: The fisher randomization test comment. *Journal of the American statistical association*, 75(371):591–593, 1980.

[27] Stephen Senn and Dimitrios Lambrou. Robust and realistic approaches to carry-over. *Statistics in Medicine*, 17(24):2849–2864, 1998.

[28] Peter Sprent. *Applied nonparametric statistical methods*. Springer Science & Business Media, 2012.

[29] Yixin Tang, Caixia Huang, David Kastelman, and Jared Bauman. Control using predictions as covariates in switchback experiments. 11 2020.

[30] Chien-Fu Wu. On the robustness and efficiency of some randomized designs. *The Annals of Statistics*, pages 1168–1177, 1981.

[31] Lincheng Zhao, Zhidong Bai, Chern-Ching Chao, and Wen-Qi Liang. Error bound in a central limit theorem of double-indexed permutation statistics. *The Annals of Statistics*, 25(5):2210–2227, 1997.

A Proof of main results

A.1 Proof of Theorem 1

Proof. Proof of Theorem 1. We first compute several joint probabilities that will be used later. The propensity score

$$\Pr(\boldsymbol{W}_{t-1:t} = \mathbf{1}) = \Pr(\boldsymbol{W}_{t-1:t} = \mathbf{0}) = \frac{\binom{T-3}{(T-1)/2}}{\binom{T-1}{(T-1)/2}} = \frac{T-3}{4T-8},$$

where $\binom{n}{k}$ stands for the combinatorial number of choosing k items from a total of n items. Similarly, the probabilities that we can observe three and four consecutive treatment/control are

$$\Pr(\boldsymbol{W}_{t-1:t+1} = \mathbf{1}) = \Pr(\boldsymbol{W}_{t-1:t+1} = \mathbf{0}) = \frac{\binom{T-4}{(T-1)/2}}{\binom{T-1}{(T-1)/2}} = \frac{(T-3)(T-5)}{(4T-8)(2T-6)},$$

and

$$\Pr(\boldsymbol{W}_{t-1:t+2} = \mathbf{1}) = \Pr(\boldsymbol{W}_{t-1:t+2} = \mathbf{0}) = \frac{\binom{T-5}{(T-1)/2}}{\binom{T-1}{(T-1)/2}} = \frac{(T-3)(T-5)(T-7)}{(4T-8)(2T-6)(2T-8)}.$$

Furthermore, we have

$$\Pr(\boldsymbol{W}_{t-1:t} = \mathbf{1}, \boldsymbol{W}_{t+1:t+2} = \mathbf{0}) = \Pr(\boldsymbol{W}_{t-1:t} = \mathbf{0}, \boldsymbol{W}_{t+1:t+2} = \mathbf{1}) = \frac{\binom{T-5}{(T-5)/2}}{\binom{T-1}{(T-1)/2}} = \frac{(T-3)(T-1)(T-3)}{(4T-8)(2T-6)(2T-8)}.$$

Now we are ready to analyze the variance. We first write the estimator as follows,

$$\widehat{\tau} = \frac{1}{T-1} \sum_{t=2}^{T} \left[Y_t(\mathbf{1}) \frac{\mathbb{1}\{\boldsymbol{W}_{t-1:t} = \mathbf{1}\}}{\Pr(\boldsymbol{W}_{t-1:t} = \mathbf{1})} - Y_t(\mathbf{0}) \frac{\mathbb{1}\{\boldsymbol{W}_{t-1:t} = \mathbf{0}\}}{\Pr(\boldsymbol{W}_{t-1:t} = \mathbf{0})} \right].$$

The variance of the estimator can be decomposed as

$$\begin{aligned} \operatorname{Var}(\widehat{\tau}) = & \operatorname{Var}\left(\frac{1}{T-1} \sum_{t=2}^{T} Y_t(\mathbf{1}) \frac{\mathbb{1}\{\boldsymbol{W}_{t-1:t} = \mathbf{1}\}}{\Pr(\boldsymbol{W}_{t-1:t} = \mathbf{1})} \right) + \operatorname{Var}\left(\frac{1}{T-1} \sum_{t=2}^{T} Y_t(\mathbf{0}) \frac{\mathbb{1}\{\boldsymbol{W}_{t-1:t} = \mathbf{0}\}}{\Pr(\boldsymbol{W}_{t-1:t} = \mathbf{0})} \right) \\ & + 2 \operatorname{Cov}\left(\frac{1}{T-1} \sum_{t=2}^{T} Y_t(\mathbf{1}) \frac{\mathbb{1}\{\boldsymbol{W}_{t-1:t} = \mathbf{1}\}}{\Pr(\boldsymbol{W}_{t-1:t} = \mathbf{1})}, -\frac{1}{T-1} \sum_{t=2}^{T} -Y_t(\mathbf{0}) \frac{\mathbb{1}\{\boldsymbol{W}_{t-1:t} = \mathbf{0}\}}{\Pr(\boldsymbol{W}_{t-1:t} = \mathbf{0})} \right). \end{aligned}$$

We first examine the first part:

$$\operatorname{Var}\left(\frac{1}{T-1} \sum_{t=2}^{T} Y_t(\mathbf{1}) \frac{\mathbb{1}\{\boldsymbol{W}_{t-1:t} = \mathbf{1}\}}{\Pr(\boldsymbol{W}_{t-1:t} = \mathbf{1})} \right) = \frac{1}{(T-1)^2} \sum_{t=2}^{T} \sum_{t'=2}^{T} \frac{\operatorname{Cov}\left(\mathbb{1}\{\boldsymbol{W}_{t-1:t} = \mathbf{1}\}, \mathbb{1}\{\boldsymbol{W}_{t'-1:t'} = \mathbf{1}\}\right)}{\Pr(\boldsymbol{W}_{t-1:t} = \mathbf{1}) \Pr(\boldsymbol{W}_{t'-1:t'} = \mathbf{1})} Y_t(\mathbf{1}) Y_{t'}(\mathbf{1})$$

Let us define $\delta_{t,t'}$ to be the distance between two periods. Using the joint probabilities we have derived, we can specify the covariance in three scenarios:

$$\frac{\operatorname{Cov}\left(\mathbb{1}\{\boldsymbol{W}_{t-1:t} = \mathbf{1}\}, \mathbb{1}\{\boldsymbol{W}_{t'-1:t'} = \mathbf{1}\}\right)}{\Pr(\boldsymbol{W}_{t-1:t} = \mathbf{1}) \Pr(\boldsymbol{W}_{t'-1:t'} = \mathbf{1})} = \begin{cases} \dfrac{4T-8}{T-3} - 1, & \delta_{t,t'} = 0 \\ \dfrac{(T-5)(4T-8)}{(2T-6)(T-3)} - 1, & \delta_{t,t'} = 1 \\ \dfrac{(T-5)(T-7)(4T-8)}{(2T-6)(2T-8)(T-3)} - 1, & \delta_{t,t'} \geq 2 \end{cases}$$

Therefore, we can calculate the variance as follows,

$$\begin{aligned} (T-1)^2 \operatorname{Var}\left(\frac{1}{T-1} \sum_{t=2}^{T} Y_t(\mathbf{1}) \frac{\mathbb{1}\{\boldsymbol{W}_{t-1:t} = \mathbf{1}\}}{\Pr(\boldsymbol{W}_{t-1:t} = \mathbf{1})} \right) = & \sum_{t=2}^{T} \left(\frac{4T-8}{T-3} - 1 \right) Y_t^2(\mathbf{1}) \\ & + 2 \sum_{t=2}^{T} \left(\frac{(T-5)(4T-8)}{(2T-6)(T-3)} - 1 \right) Y_t(\mathbf{1}) Y_{t+1}(\mathbf{1}) \\ & + \sum_{t=2}^{T} \sum_{\delta_{t,t'} \geq 2} \left(\frac{(T-5)(T-7)(4T-8)}{(2T-6)(2T-8)(T-3)} - 1 \right) Y_t(\mathbf{1}) Y_{t'}(\mathbf{1}). \end{aligned}$$

Because of

$$\sum_{t=2}^{T} (Y_t(\mathbf{1}) - \overline{Y}(\mathbf{1}))^2 = \sum_{t=2}^{T} Y_t^2(\mathbf{1}) - \frac{1}{T-1} \sum_{t=2}^{T} \sum_{t'=2}^{T} Y_t(\mathbf{1}) Y_{t'}(\mathbf{1}),$$

$$\sum_{t=2}^{T} (Y_t(\mathbf{1}) - \overline{Y}(\mathbf{1}))(Y_{t+1}(\mathbf{1}) - \overline{Y}(\mathbf{1})) = \sum_{t=2}^{T} Y_t(\mathbf{1}) Y_{t+1}(\mathbf{1}) - \frac{1}{T-1} \sum_{t=2}^{T} \sum_{t'=2}^{T} Y_t(\mathbf{1}) Y_{t'}(\mathbf{1})$$

the variance is equivalent to

$$\begin{aligned} \operatorname{Var}\left(\frac{1}{T-1} \sum_{t=2}^{T} Y_t(\mathbf{1}) \frac{\mathbb{1}\{\boldsymbol{W}_{t-1:t} = \mathbf{1}\}}{\Pr(\boldsymbol{W}_{t-1:t} = \mathbf{1})} \right) = & \frac{2T^2 - 13T + 17}{(T-1)(T-4)(T-3)^2} \sum_{t=2}^{T} (Y_t(\mathbf{1}) - \overline{Y}(\mathbf{1}))^2 \\ & + \frac{2T^2 - 13T + 17}{(T-1)(T-4)(T-3)^2} \sum_{t=2}^{T} (Y_t(\mathbf{1}) - \overline{Y}(\mathbf{1}))(Y_{t+1}(\mathbf{1}) - \overline{Y}(\mathbf{1})) \\ & + \frac{1}{(T-1)(T-4)} \sum_{t=2}^{T} Y_t^2(\mathbf{1}) - \frac{1}{(T-1)(T-4)(T-3)} \sum_{t=2}^{T} Y_t(\mathbf{1}) Y_{t+1}(\mathbf{1}). \end{aligned} \tag{12}$$

Similarly, we can characterize the variance for outcomes of control:

$$\operatorname{Var}\left(\frac{1}{T-1}\sum_{t=2}^{T} Y_t(\mathbf{0})\frac{\mathbb{1}\{\boldsymbol{W}_{t-1:t}=\mathbf{0}\}}{\Pr(\boldsymbol{W}_{t-1:t}=\mathbf{0})}\right) = \frac{2T^2-13T+17}{(T-1)(T-4)(T-3)^2}\sum_{t=2}^{T}(Y_t(\mathbf{0})-\overline{Y}(\mathbf{0}))^2$$
$$+\frac{2T^2-13T+17}{(T-1)(T-4)(T-3)^2}\sum_{t=2}^{T}(Y_t(\mathbf{0})-\overline{Y}(\mathbf{0}))(Y_{t+1}(\mathbf{0})-\overline{Y}(\mathbf{0}))$$
$$+\frac{1}{(T-1)(T-4)}\sum_{t=2}^{T}Y_t^2(\mathbf{0}) - \frac{1}{(T-1)(T-4)(T-3)}\sum_{t=2}^{T}Y_t(\mathbf{0})Y_{t+1}(\mathbf{0}) \tag{13}$$

Next, we examine the covariance between the outcomes of treatment and control:

$$\operatorname{Cov}\left(\frac{1}{T-1}\sum_{t=2}^{T} Y_t(\mathbf{1})\frac{\mathbb{1}\{\boldsymbol{W}_{t-1:t}=\mathbf{1}\}}{\Pr(\boldsymbol{W}_{t-1:t}=\mathbf{1})}, -\frac{1}{T-1}\sum_{t=2}^{T} Y_t(\mathbf{0})\frac{\mathbb{1}\{\boldsymbol{W}_{t-1:t}=\mathbf{0}\}}{\Pr(\boldsymbol{W}_{t-1:t}=\mathbf{0})}\right)$$
$$=\mathbb{E}\left(\frac{1}{T-1}\sum_{t=2}^{T} Y_t(\mathbf{1})\frac{\mathbb{1}\{\boldsymbol{W}_{t-1:t}=\mathbf{1}\}}{\Pr(\boldsymbol{W}_{t-1:t}=\mathbf{1})}\right)\mathbb{E}\left(\frac{1}{T-1}\sum_{t=2}^{T} Y_t(\mathbf{0})\frac{\mathbb{1}\{\boldsymbol{W}_{t-1:t}=\mathbf{0}\}}{\Pr(\boldsymbol{W}_{t-1:t}=\mathbf{0})}\right)$$
$$-\frac{1}{(T-1)^2}\mathbb{E}\left(\sum_{t=2}^{T} Y_t(\mathbf{1})\frac{\mathbb{1}\{\boldsymbol{W}_{t-1:t}=\mathbf{1}\}}{\Pr(\boldsymbol{W}_{t-1:t}=\mathbf{1})}\sum_{t=2}^{T} Y_t(\mathbf{0})\frac{\mathbb{1}\{\boldsymbol{W}_{t-1:t}=\mathbf{0}\}}{\Pr(\boldsymbol{W}_{t-1:t}=\mathbf{0})}\right).$$

The product of expectations is given by

$$\mathbb{E}\left(\frac{1}{T-1}\sum_{t=2}^{T} Y_t(\mathbf{1})\frac{\mathbb{1}\{\boldsymbol{W}_{t-1:t}=\mathbf{1}\}}{\Pr(\boldsymbol{W}_{t-1:t}=\mathbf{1})}\right)=\overline{Y}(\mathbf{1}), \mathbb{E}\left(\frac{1}{T-1}\sum_{t=2}^{T} Y_t(\mathbf{0})\frac{\mathbb{1}\{\boldsymbol{W}_{t-1:t}=\mathbf{0}\}}{\Pr(\boldsymbol{W}_{t-1:t}=\mathbf{0})}\right)=\overline{Y}(\mathbf{0}).$$

Using the jointly probabilities we have derived again, we can specify the expectation of the product in three scenarios:

$$\frac{\mathbb{E}\left(\mathbb{1}\{\boldsymbol{W}_{t-1:t}=\mathbf{1}\}\mathbb{1}\{\boldsymbol{W}_{t'-1:t'}=\mathbf{0}\}\right)}{\Pr(\boldsymbol{W}_{t-1:t}=\mathbf{1})\Pr(\boldsymbol{W}_{t'-1:t'}=\mathbf{0})}=\begin{cases}0, & \delta_{t,t'}=0\\ 0, & \delta_{t,t'}=1\\ \dfrac{(T-2)(T-1)}{(T-3)(T-4)}, & \delta_{t,t'}\geq 2.\end{cases}$$

Therefore, we can calculate the covariance as follows,

$$\operatorname{Cov}\left(\frac{1}{T-1}\sum_{t=2}^{T} Y_t(\mathbf{1})\frac{\mathbb{1}\{\boldsymbol{W}_{t-1:t}=\mathbf{1}\}}{\Pr(\boldsymbol{W}_{t-1:t}=\mathbf{1})}, -\frac{1}{T-1}\sum_{t=2}^{T} Y_t(\mathbf{0})\frac{\mathbb{1}\{\boldsymbol{W}_{t-1:t}=\mathbf{0}\}}{\Pr(\boldsymbol{W}_{t-1:t}=\mathbf{0})}\right)$$
$$=\overline{Y}(\mathbf{1})\overline{Y}(\mathbf{0})-\frac{(T-2)(T-1)}{(T-3)(T-4)(T-1)^2}\sum_{t=2}^{T}\sum_{\delta_{t,t'}\geq 2}Y_t(\mathbf{1})Y_{t'}(\mathbf{0})$$

Because of

$$\sum_{t=2}^{T}(Y_t(\mathbf{1})-\overline{Y}(\mathbf{1}))(Y_t(\mathbf{0})-\overline{Y}(\mathbf{0}))=\sum_{t=2}^{T}Y_t(\mathbf{1})Y_t(\mathbf{0})-\frac{1}{T-1}\sum_{t=2}^{T}\sum_{t'=2}^{T}Y_t(\mathbf{1})Y_{t'}(\mathbf{0}),$$
$$\sum_{t=2}^{T}(Y_t(\mathbf{1})-\overline{Y}(\mathbf{1}))(Y_{t+1}(\mathbf{0})-\overline{Y}(\mathbf{0}))=\sum_{t=2}^{T}Y_t(\mathbf{1})Y_{t+1}(\mathbf{0})-\frac{1}{T-1}\sum_{t=2}^{T}\sum_{t'=2}^{T}Y_t(\mathbf{1})Y_{t'}(\mathbf{0})$$

the covariance is equivalent to

$$\begin{aligned}
&\frac{2T-5}{(T-4)(T-3)(T-1)}\sum_{t=2}^{T}(Y_t(\mathbf{1})-\overline{Y}(\mathbf{1}))(Y_t(\mathbf{0})-\overline{Y}(\mathbf{0}))\\
&+\frac{2T-5}{2(T-4)(T-3)(T-1)}\sum_{t=2}^{T}(Y_t(\mathbf{1})-\overline{Y}(\mathbf{1}))(Y_{t+1}(\mathbf{0})-\overline{Y}(\mathbf{0}))\\
&+\frac{2T-5}{2(T-4)(T-3)(T-1)}\sum_{t=2}^{T}(Y_t(\mathbf{0})-\overline{Y}(\mathbf{0}))(Y_{t+1}(\mathbf{1})-\overline{Y}(\mathbf{1}))\\
&-\frac{1}{(T-4)(T-1)}\sum_{t=2}^{T}Y_t(\mathbf{1})Y_t(\mathbf{0})\\
&+\frac{1}{2(T-4)(T-1)(T-3)}\sum_{t=2}^{T}Y_t(\mathbf{1})Y_{t+1}(\mathbf{0})+Y_t(\mathbf{0})Y_{t+1}(\mathbf{1})
\end{aligned}$$

Now we are going to reformulate the covariance using previous expressions. Since we have the following two equations

$$\begin{aligned}
&\sum_{t=2}^{T}(Y_t(\mathbf{1})-\overline{Y}(\mathbf{1}))(Y_{t+1}(\mathbf{0})-\overline{Y}(\mathbf{0}))+\sum_{t=2}^{T}(Y_t(\mathbf{0})-\overline{Y}(\mathbf{0}))(Y_{t+1}(\mathbf{1})-\overline{Y}(\mathbf{1}))\\
=&\sum_{t=2}^{T}(Y_t(\mathbf{1})-\overline{Y}(\mathbf{1}))(Y_{t+1}(\mathbf{1})-\overline{Y}(\mathbf{1}))+\sum_{t=2}^{T}(Y_t(\mathbf{0})-\overline{Y}(\mathbf{0}))(Y_{t+1}(\mathbf{0})-\overline{Y}(\mathbf{0}))\\
&-\sum_{t=2}^{T}(Y_t(\mathbf{1})-Y_t(\mathbf{0})-\overline{Y}(\mathbf{1})+\overline{Y}(\mathbf{0}))(Y_{t+1}(\mathbf{1})-Y_{t+1}(\mathbf{0})-\overline{Y}(\mathbf{1})+\overline{Y}(\mathbf{0}))
\end{aligned}$$

and

$$\begin{aligned}
2\sum_{t=2}^{T}(Y_t(\mathbf{1})-\overline{Y}(\mathbf{1}))(Y_t(\mathbf{0})-\overline{Y}(\mathbf{0}))=&\sum_{t=2}^{T}(Y_t(\mathbf{1})-\overline{Y}(\mathbf{1}))^2+\sum_{t=2}^{T}(Y_t(\mathbf{0})-\overline{Y}(\mathbf{0}))^2\\
&-\sum_{t=2}^{T}(Y_t(\mathbf{1})-Y_t(\mathbf{0})-\overline{Y}(\mathbf{1})+\overline{Y}(\mathbf{0}))^2,
\end{aligned}$$

we then rewrite the covariance as

$$\begin{aligned}
&\frac{2T-5}{2(T-4)(T-3)(T-1)}\left(\sum_{t=2}^{T}(Y_t(\mathbf{1})-\overline{Y}(\mathbf{1}))^2+\sum_{t=2}^{T}(Y_t(\mathbf{0})-\overline{Y}(\mathbf{0}))^2-\sum_{t=2}^{T}(Y_t(\mathbf{1})-Y_t(\mathbf{0})-\overline{Y}(\mathbf{1})+\overline{Y}(\mathbf{0}))^2\right)\\
&+\frac{2T-5}{2(T-4)(T-3)(T-1)}\left(\sum_{t=2}^{T}(Y_t(\mathbf{1})-\overline{Y}(\mathbf{1}))(Y_{t+1}(\mathbf{1})-\overline{Y}(\mathbf{1}))+\sum_{t=2}^{T}(Y_t(\mathbf{0})-\overline{Y}(\mathbf{0}))(Y_{t+1}(\mathbf{0})-\overline{Y}(\mathbf{0}))\right.\\
&\left.-\sum_{t=2}^{T}(Y_t(\mathbf{1})-Y_t(\mathbf{0})-\overline{Y}(\mathbf{1})+\overline{Y}(\mathbf{0}))(Y_{t+1}(\mathbf{1})-Y_{t+1}(\mathbf{0})-\overline{Y}(\mathbf{1})+\overline{Y}(\mathbf{0}))\right)\\
&-\frac{1}{(T-4)(T-1)}\sum_{t=2}^{T}Y_t(\mathbf{1})Y_t(\mathbf{0})+\frac{1}{2(T-4)(T-1)(T-3)}\sum_{t=2}^{T}Y_t(\mathbf{1})Y_{t+1}(\mathbf{0})+Y_t(\mathbf{0})Y_{t+1}(\mathbf{1})
\end{aligned} \tag{14}$$

Finally, putting all parts (12), (13), (14) together and using S^c, S^t, S^τ, we derive the variance of the estimator

$$\begin{aligned}\mathrm{Var}(\widehat{\tau}) = &\frac{8(T-2)^2}{(T-3)^2(T-1)}\left(S^t + S^c\right) - \frac{(4T-10)(T-2)}{(T-3)(T-4)(T-1)}S^\tau \\ &+ \frac{1}{(T-1)(T-4)}\sum_{t=2}^{T}\left(Y_t^2(\mathbf{1}) + Y_t^2(\mathbf{0}) - 2Y_t(\mathbf{1})Y_t(\mathbf{0})\right) \\ &- \frac{1}{(T-1)(T-4)(T-3)}\sum_{t=2}^{T}\left(Y_t(\mathbf{1})Y_{t+1}(\mathbf{1}) + Y_t(\mathbf{0})Y_{t+1}(\mathbf{0}) - Y_t(\mathbf{1})Y_{t+1}(\mathbf{0}) - Y_t(\mathbf{0})Y_{t+1}(\mathbf{1})\right).\end{aligned}$$

This can be further simplified as

$$\begin{aligned}\mathrm{Var}(\widehat{\tau}) = &\frac{8(T-2)^2}{(T-3)^2(T-1)}\left(S^t + S^c\right) - \frac{(4T-10)(T-2)}{(T-3)(T-4)(T-1)}S^\tau \\ &+ \frac{1}{(T-1)(T-4)}\sum_{t=2}^{T}\left(Y_t(\mathbf{1}) - Y_t(\mathbf{0})\right)^2 \\ &- \frac{1}{(T-1)(T-4)(T-3)}\sum_{t=2}^{T}(Y_t(\mathbf{1}) - Y_t(\mathbf{0}))(Y_{t+1}(\mathbf{1}) - Y_{t+1}(\mathbf{0})). \end{aligned} \tag{15}$$

□

A.2 Proof of Theorem 2

Lemma 8. *The worst-case outcomes must satisfy the following structure:*

$$Y_t(\mathbf{1}) = Y_t(\mathbf{0}) = \begin{cases} B & 2 \le t \le s, \\ 0 & t > s. \end{cases} \tag{16}$$

Proof. Proof of Lemma 8. We first expand the variance by definition and have the following:

$$\begin{aligned}(T-1)^2 \cdot \mathrm{Var}_{\eta^\dagger}(\widehat{\tau}|\mathbb{Y}) = \sum_{t=2}^{T}\sum_{t'=2}^{T} &\left(\frac{\Pr(\boldsymbol{W}_{t-1:t} = \mathbf{1}, \boldsymbol{W}_{t'-1:t'} = \mathbf{1})}{\Pr(\boldsymbol{W}_{t-1:t} = \mathbf{1})\Pr(\boldsymbol{W}_{t'-1:t'} = \mathbf{1})} - 1\right) Y_t(\mathbf{1})Y_{t'}(\mathbf{1}) \\ &+ \left(\frac{\Pr(\boldsymbol{W}_{t-1:t} = \mathbf{0}, \boldsymbol{W}_{t'-1:t'} = \mathbf{0})}{\Pr(\boldsymbol{W}_{t-1:t} = \mathbf{0})\Pr(\boldsymbol{W}_{t'-1:t'} = \mathbf{0})} - 1\right) Y_t(\mathbf{0})Y_{t'}(\mathbf{0}) \\ &+ \left(1 - \frac{\Pr(\boldsymbol{W}_{t-1:t} = \mathbf{1}, \boldsymbol{W}_{t'-1:t'} = \mathbf{0})}{\Pr(\boldsymbol{W}_{t-1:t} = \mathbf{1})\Pr(\boldsymbol{W}_{t'-1:t'} = \mathbf{0})}\right) Y_t(\mathbf{1})Y_{t'}(\mathbf{0}) \\ &+ \left(1 - \frac{\Pr(\boldsymbol{W}_{t-1:t} = \mathbf{0}, \boldsymbol{W}_{t'-1:t'} = \mathbf{1})}{\Pr(\boldsymbol{W}_{t-1:t} = \mathbf{0})\Pr(\boldsymbol{W}_{t'-1:t'} = \mathbf{1})}\right) Y_t(\mathbf{0})Y_{t'}(\mathbf{1})\end{aligned}$$

This is a quadratic function with variables $Y_t(\mathbf{1}), Y_t(\mathbf{0}), \forall t \in \{2, 3, \ldots, T\}$. To show it is also convex, we can rewrite the summation as $\boldsymbol{y}'\boldsymbol{\Sigma}\boldsymbol{y}$, where $\boldsymbol{y} \in \mathbb{R}^{2(T-1)}$ is the vector of all variables and $\boldsymbol{\Sigma}$ is a symmetric matrix of coefficients. Because variance is non-negative, we know that $\boldsymbol{y}'\boldsymbol{\Sigma}\boldsymbol{y} \ge 0$ for any $\boldsymbol{y}$, which implies that $\boldsymbol{\Sigma}$ is PSD and the function is convex in $\boldsymbol{y}$. Since the inner optimization is a minimization in a bounded feasible region, the worst-case solution can be attained at one of the extreme points. That is,

$$Y_t(\mathbf{1}) \in \{0, B\}, Y_t(\mathbf{0}) \in \{0, B\}, \forall t.$$

Next, given any outcomes at the extreme point, we will argue that transforming into the structure (16) leads to a larger variance. To see this, we need to carefully analyze the coefficients of $\boldsymbol{\Sigma}$. We introduce some shorthand notations for future reference. Due to the symmetry of the design, we set

$$\begin{aligned}q^+(\delta_{t,t'}) &= \frac{\Pr(\boldsymbol{W}_{t-1:t} = \mathbf{1}, \boldsymbol{W}_{t'-1:t'} = \mathbf{1})}{\Pr(\boldsymbol{W}_{t-1:t} = \mathbf{1})\Pr(\boldsymbol{W}_{t'-1:t'} = \mathbf{1})} - 1 = \frac{\Pr(\boldsymbol{W}_{t-1:t} = \mathbf{0}, \boldsymbol{W}_{t'-1:t'} = \mathbf{0})}{\Pr(\boldsymbol{W}_{t-1:t} = \mathbf{0})\Pr(\boldsymbol{W}_{t'-1:t'} = \mathbf{0})} - 1, \\ q^-(\delta_{t,t'}) &= 1 - \frac{\Pr(\boldsymbol{W}_{t-1:t} = \mathbf{1}, \boldsymbol{W}_{t'-1:t'} = \mathbf{0})}{\Pr(\boldsymbol{W}_{t-1:t} = \mathbf{1})\Pr(\boldsymbol{W}_{t'-1:t'} = \mathbf{0})} = 1 - \frac{\Pr(\boldsymbol{W}_{t-1:t} = \mathbf{0}, \boldsymbol{W}_{t'-1:t'} = \mathbf{1})}{\Pr(\boldsymbol{W}_{t-1:t} = \mathbf{0})\Pr(\boldsymbol{W}_{t'-1:t'} = \mathbf{1})}.\end{aligned}$$

Based on the joint probabilities we have derived in the proof of Theorem 1, the coefficients of $\mathbf{\Sigma}$ can be characterized according to $\delta_{t,t'}$:

$$q^+(\delta_{t,t'}) = \begin{cases} \dfrac{4T-8}{T-3} - 1, & \delta_{t,t'} = 0 \\ \dfrac{(T-5)(4T-8)}{(2T-6)(T-3)} - 1, & \delta_{t,t'} = 1 \\ \dfrac{(T-5)(T-7)(4T-8)}{(2T-6)(2T-8)(T-3)} - 1, & \delta_{t,t'} \geq 2 \end{cases}$$

$$q^-(\delta_{t,t'}) = \begin{cases} 1, & \delta_{t,t'} = 0 \\ 1, & \delta_{t,t'} = 1 \\ \dfrac{-4T+10}{(T-3)(T-4)}, & \delta_{t,t'} \geq 2 \end{cases}$$

It is easy to see that both $q^+(\delta_{t,t'})$ and $q^-(\delta_{t,t'})$ are decreasing in $\delta_{t,t'}$. Therefore, the closer two outcomes B, the more they contribute to the variance. Suppose we are given some outcomes at the extreme point and there are T_1 periods whose outcome of treatment is B while T_0 periods whose outcome of control is B. W.L.O.G., assuming $T_1 \geq T_0$, let us consider the following alternative outcomes:

$$Y_t(\mathbf{1}) = \begin{cases} B & 2 \leq t \leq T_1 + 1 \\ 0 & otherwise \end{cases}, \quad Y_t(\mathbf{0}) = \begin{cases} B & \frac{T_1 - T_0}{2} + 2 \leq t \leq \frac{T_1 + T_0}{2} + 1 \\ 0 & otherwise \end{cases} \tag{17}$$

We check the variance of the alternative outcomes using the monotonicity of $q^+(\delta_{t,t'})$ and $q^-(\delta_{t,t'})$. Since the alternative outcomes group B together with the minimal distance, the variance from the outcomes of treatment(control) increases. Moreover, because the alternative outcomes synchronize the outcomes between treatment and control as much as possible, the covariance from the outcomes between treatment and control increases as well. Together, the alternative outcomes achieve a larger variance.

Lastly, it remains to show that further transforming (17) into (16) gives us a larger variance. Essentially, the transformation is doing

$$Y_t(\mathbf{1}) = B \Longrightarrow Y_t(\mathbf{1}) = 0, \quad \frac{T_1 + T_0}{2} + 2 \leq t \leq T_1 + 1$$
$$Y_t(\mathbf{0}) = 0 \Longrightarrow Y_t(\mathbf{0}) = B, \quad 2 \leq t \leq \frac{T_1 - T_0}{2} + 1.$$

This can be illustrated using the examples in Figure 7 with $T = 9$. To see that the variance increases,

Period	2	3	4	5	6	7	8	9
Treatment	B	B	B	B	B	B	B	B
Control	0	0	0	B	B	0	0	0

⇨

Period	2	3	4	5	6	7	8	9
Treatment	B	B	B	B	B	0	0	0
Control	B	B	B	B	B	0	0	0

Figure 7: Outcomes in (17)

due to the symmetry of the design, we need to show that the covariance between blue outcomes and red outcomes is getting larger. Precisely,

$$\sum_{t=2}^{(T_1-T_0)/2+1} \sum_{t'=2}^{(T_1-T_0)/2+1} q^-(\delta_{t,t'}) \geq \sum_{t=(T_1+T_0)/2+2}^{T_1+1} \sum_{t'=2}^{(T_1-T_0)/2+1} q^+(\delta_{t,t'}).$$

It is sufficient to show that

$$\sum_{t'=2}^{(T_1-T_0)/2+1} q^-(\delta_{t,t'}) \geq \sum_{t'=2}^{(T_1-T_0)/2+1} q^+(\delta_{t,t'}), \forall\, 2 \leq t \leq \frac{T_1 - T_0}{2} + 1.$$

Because of the monotonicity of $q^+(\cdot)$ and $q^-(\cdot)$, it suffices to show that

$$\sum_{\delta_{t,t'}=0}^{(T_1-T_0)/2-1} q^-(\delta_{t,t'}) \geq \sum_{\delta_{t,t'}=1}^{(T_1-T_0)/2} q^+(\delta_{t,t'}).$$

Plugging in the expressions, this is equivalent to

$$\begin{aligned}
&\left(\frac{T_1-T_0}{2}-2\right)\left(q^-(2)-q^+(2)\right)+q^-(0)+q^-(1)-q^+(1)-q^+(2)\geq 0\\
\Longleftrightarrow &-\left(\frac{T_1-T_0}{2}-2\right)\frac{4(T-1)}{(T-3)^2(T-4)}-\frac{(T-7)(T-5)(4T-8)}{(2T-8)(2T-6)(T-3)}-\frac{(T-5)(4T-8)}{(2T-6)(T-3)}+4\geq 0.
\end{aligned}$$

Since T_1-T_0 is bounded above by $\frac{T-1}{2}$, it suffices to show that

$$\begin{aligned}
&-\frac{2(T-5)(T-1)}{(T-3)^2(T-4)}-\frac{(T-7)(T-5)(4T-8)}{(2T-8)(2T-6)(T-3)}-\frac{(T-5)(4T-8)}{(2T-6)(T-3)}+4\geq 0\\
\Longleftrightarrow &-\frac{(T-5)(3T-7)}{(T-3)^2}+4\geq 0\\
\Longleftrightarrow &\frac{(T-1)^2}{(T-3)^2}\geq 0.
\end{aligned}$$

Hence, the worst-case outcomes must obey the structure that:

$$Y_t(\mathbf{1})=Y_t(\mathbf{0})=\begin{cases}B & 2\leq t\leq s,\\ 0 & t>s.\end{cases}$$

□

Proof. Proof of Theorem 2 From Lemma 8, we know that the worst-case outcomes obey the following structure

$$Y_t(\mathbf{1})=Y_t(\mathbf{0})=\begin{cases}B & 2\leq t\leq s,\\ 0 & t>s.\end{cases}$$

We need to show that $s=\frac{T+1}{2}$. We do it by contradiction.

Suppose that $s<\frac{T+1}{2}$, then we can set $Y_{s+1}(\mathbf{1})=Y_{s+1}(\mathbf{0})=B$. In this way, we have one more pair of outcomes B which contributes to the variance. To show that the variance increases, it is equivalent to prove that

$$q^+(0)+q^-(0)+2\sum_{\delta_{t,t'}=1}^{s}(q^+(\delta_{t,t'})+q^-(\delta_{t,t'}))\geq 0$$

Since $q^+(\delta_{t,t'})+q^-(\delta_{t,t'})$ takes negative value when $\delta_{t,t'}\geq 2$, it is sufficient to show that

$$q^+(0)+q^-(0)+2\sum_{\delta_{t,t'}=1}^{(T-3)/2}(q^+(\delta_{t,t'})+q^-(\delta_{t,t'}))\geq 0$$

Plugging in the expressions of $\boldsymbol{q}^+$ and $\boldsymbol{q}^-$, we have

$$q^+(0)+q^-(0)+2\sum_{\delta_{t,t'}=1}^{(T-3)/2}(q^+(\delta_{t,t'})+q^-(\delta_{t,t'}))=\frac{8(T-2)}{(T-3)^2}>0.$$

In the other way around when $s>\frac{T+1}{2}$, then we can set $Y_{s-1}(\mathbf{1})=Y_{s-1}(\mathbf{0})=0$. Following the similar argument, it is sufficient to show that

$$-q^+(0)-q^-(0)-2\sum_{\delta_{t,t'}=1}^{(T-1)/2}(q^+(\delta_{t,t'})+q^-(\delta_{t,t'}))\geq 0$$

Plugging in the expressions again, we have

$$-q^+(0) - q^-(0) - 2\sum_{\delta_{t,t'}=1}^{(T-1)/2} (q^+(\delta_{t,t'}) + q^-(\delta_{t,t'})) = \frac{8(T-2)}{(T-3)^2} > 0.$$

Hence, the variance reaches the maximum when $s = \frac{T+1}{2}$.

To further derive the worst-case variance, we can simply use the variance decomposition (15). Note that the last three parts are all zero, the worst-case variance can be calculated by

$$\mathrm{Var}(\hat{\tau}) = \frac{8(T-2)^2}{(T-3)^2(T-1)}\left(S^t + S^c\right) = \frac{4(T-2)}{(T-3)(T-1)}B^2.$$

□

A.3 Proof of Theorem 3

Lemma 9. *Let us consider two consecutive time periods t and $t+1$. For any symmetric design, we have the following inequality:*

$$\begin{aligned}&\frac{1}{\Pr(\boldsymbol{W}_{t-1:t}=\mathbf{1})} - 1 + \frac{\Pr(\boldsymbol{W}_{t-1:t}=\mathbf{1}, \boldsymbol{W}_{t:t+1}=\mathbf{1})}{\Pr(\boldsymbol{W}_{t-1:t}=\mathbf{1})\Pr(\boldsymbol{W}_{t:t+1}=\mathbf{1})} - 1 + \\ &\frac{1}{\Pr(\boldsymbol{W}_{t:t+1}=\mathbf{1})} - 1 + \frac{\Pr(\boldsymbol{W}_{t-1:t}=\mathbf{1}, \boldsymbol{W}_{t:t+1}=\mathbf{1})}{\Pr(\boldsymbol{W}_{t-1:t}=\mathbf{1})\Pr(\boldsymbol{W}_{t:t+1}=\mathbf{1})} - 1 \geq 4\end{aligned} \tag{18}$$

Proof. Proof of Lemma 9. We first reformulate the inequality as

$$\begin{aligned}&\Pr(\boldsymbol{W}_{t-1:t}=1) + 2\Pr(\boldsymbol{W}_{t-1:t}=1, \boldsymbol{W}_{t:t+1}=1) + \Pr(\boldsymbol{W}_{t:t+1}=1) \\ &\geq 8\Pr(\boldsymbol{W}_{t:t+1}=1)\Pr(\boldsymbol{W}_{t:t+1}=1).\end{aligned} \tag{19}$$

Now let us focus on three time periods: $t-1$, t and $t+1$. There are 8 possible assignment paths. We layout 4 of them and the remaining ones are just symmetric:

$$\boldsymbol{W}_{t-1:t+1} \in \{(1,1,1), (1,1,0), (0,1,1), (1,0,1)\}.$$

with their probability mass denoted as $a_3, a_{2,1}, a_{2,2}, a_{2,3}$ respectively. Then we can characterize the probabilities in the inequality using these $\boldsymbol{a}$:

$$a_3 + a_{2,1} + 2a_3 + a_3 + a_{2,2} \geq 8(a_3 + a_{2,1})(a_3 + a_{2,2}).$$

Since $a_3 + a_{2,1} + a_{2,2} + a_{2,3} = 0.5$, it is equivalent to show

$$2(a_3 + a_{2,1} + a_{2,2} + a_{2,3})(4a_3 + a_{2,1} + a_{2,2}) \geq 8(a_3 + a_{2,1})(a_3 + a_{2,2}).$$

Notice that $a_{2,3}$ only appears on the left-hand-side, so it is sufficient to show

$$2(a_3 + a_{2,1} + a_{2,2} + a_{2,3})(4a_3 + a_{2,1} + a_{2,2}) \geq 8(a_3 + a_{2,1})(a_3 + a_{2,2}).$$

which can be further simplified as

$$(a_{2,1} - a_{2,2})^2 + a_3(a_{2,1} + a_{2,2}) \geq 0.$$

This is true for any $\boldsymbol{a}$. □

Proof. Proof of Theorem 3. In the proof of Lemma 8, we rewrite the variance by introducing $\boldsymbol{y} \in \mathbb{R}^{2(T-1)}$ to denote the vector of outcomes. Our original minimax problem is equivalent to

$$(T-1)^2 \cdot \mathrm{Var}(\hat{\tau}) = \min_{\Sigma} \max_{\boldsymbol{y}\in[0,B]} \boldsymbol{y}^T \Sigma \boldsymbol{y},$$

where Σ is some covariance matrix that can be mapped from a feasible design and the adversary finds an outcome vector to maximize the variance. To get a lower bound of the optimal worst-case variance, we consider a randomized feasible solution $\tilde{\boldsymbol{y}}$ regardless of the covariance matrix in the outer optimization. We first combine every two time periods as a group, so we have overall $n = \frac{T-1}{2}$ groups. We then randomly pick half of the groups and set their corresponding outcomes to B and others to 0. Let $h(i)$ denote the group of some outcome $\tilde{y}_i$. In this way, for any two outcomes $\tilde{y}_i$ and $\tilde{y}_j$ from the same group(i.e. $h(i) = h(j)$), $\mathbb{E}[\tilde{y}_i\tilde{y}_j] = \frac{1}{2}B^2$; for any two outcomes from different groups, $\mathbb{E}[\tilde{y}_i\tilde{y}_j] = \frac{n-2}{4(n-1)}B^2$. Now we can bound the inner optimization as follows:

$$\max_{\boldsymbol{y}\in[0,B]} \boldsymbol{y}^T\Sigma\boldsymbol{y} \geq \mathbb{E}\left[\tilde{\boldsymbol{y}}^{\mathrm{T}}\Sigma\tilde{\boldsymbol{y}}\right] = \sum_{h(i)=h(j)} \Sigma_{i,j}\frac{1}{2}B^2 + \sum_{h(i)\neq h(j)} \Sigma_{i,j}\frac{n-2}{4(n-1)}B^2$$
$$= \sum_{h(i)=h(j)} \Sigma_{i,j}\frac{n}{4(n-1)}B^2 + \sum_{\forall i,j}\Sigma_{i,j}\frac{n-2}{4(n-1)}B^2$$

Note that the last term is non-negative, so it implies that

$$\max_{\boldsymbol{y}\in[0,B]} \boldsymbol{y}^T\Sigma\boldsymbol{y} \geq \sum_{h(i)=h(j)} \Sigma_{i,j}\frac{n}{4(n-1)}B^2.$$

Then it remains to investigate $\Sigma_{i,j}$ when two outcomes $\tilde{y}_i$ and $\tilde{y}_j$ are from the same group(i.e. two consecutive periods). Let us focus on what will happen in one group. First of all, it is easy to observe that the following is true for any design:

$$1 - \frac{\Pr(\boldsymbol{W}_{t-1:t} = \boldsymbol{0}, \boldsymbol{W}_{t:t+1} = \boldsymbol{1})}{\Pr(\boldsymbol{W}_{t-1:t} = \boldsymbol{0})\Pr(\boldsymbol{W}_{t:t+1} = \boldsymbol{1})} = 1 - \frac{\Pr(\boldsymbol{W}_{t-1:t} = \boldsymbol{0}, \boldsymbol{W}_{t-1:t} = \boldsymbol{1})}{\Pr(\boldsymbol{W}_{t-1:t} = \boldsymbol{0})\Pr(\boldsymbol{W}_{t-1:t} = \boldsymbol{1})} = 1$$

We have 4 such pairs in one group, so they contribute $\frac{8n}{4(n-1)}B^2$ to the variance. Next, if we set the assignments in the above equation to be jointly $\boldsymbol{0}$ or $\boldsymbol{1}$, we are not able to know the exact values. Fortunately, based on Lemma 9, we can still bound the variance to which they contribute by $\frac{8n}{4(n-1)}B^2$. Lastly, as we have n groups, we get the lower bound

$$\max_{\boldsymbol{y}\in[0,B]} \boldsymbol{y}^T\Sigma\boldsymbol{y} \geq n\frac{(8+8)n}{4(n-1)}B^2 = \frac{4n^2}{n-1}B^2 = \frac{2}{T-3}B^2.$$

Together with the upper bound, the approximation ratio follows easily. □

A.4 Proof of Theorem 6

Proof. Proof of Theorem 6. Let us first define

$$\xi_T(t,t',s,s') = \begin{cases} \frac{4T-8}{(T-1)(T-3)}Y_{t'}(\boldsymbol{1}) & t' = t+1, 2 \leq s \neq s' \leq \frac{T+1}{2}, \\ -\frac{4T-8}{(T-1)(T-3)}Y_{t'}(\boldsymbol{0}) & t' = t+1, \frac{T+1}{2} \leq s \neq s' \leq T, \\ 0 & otherwise. \end{cases} \tag{20}$$

Let π be a random permutation that shuffles the original indices:

$$\{2,3,\ldots,T-1,T\} \to \{\pi(2),\pi(3),\ldots,\pi(T-1),\pi(T)\}.$$

Given these, we can rewrite the estimator as

$$\widehat{\tau} = \sum_{t\neq t'}^{T} \xi_T(t,t',\pi(t),\pi(t')). \tag{21}$$

where $\sum_{t\neq t'}^{T}$ indicates $\sum_{t=2}^{T}\sum_{t'=2:t\neq t'}^{T}$. To derive the normal approximation of this, we adopt Stein's method of exchange pairs for double-index permutation statistics proposed in [24]. Specifically, they construct an exchangeable pair as follows. Let t and t' be distributed uniformly over $1,\ldots,T-1$ conditioned that $t \neq t'$. Define the permutation $\pi' = (\pi(t)\pi(t')) \circ \pi$ so that π' is the permutation where $\pi'(s) = \pi(s)$ for all $k \neq t, t'$, and where $\pi'(t) = \pi(t')$ and $\pi'(t') = \pi(t)$. Let $V_1 = \widehat{\tau}$, and we define the other two random variables for proof purposes:

$$V_2 = \frac{1}{T-1}\sum_{t=2}^{T}\sum_{s,s'}^{T}\xi_T(t,s,\pi(t),s'), V_3 = \frac{1}{T-1}\sum_{t=2}^{T}\sum_{s,s'}^{T}\xi_T(s,t,s',\pi(t)).$$

Then we have $\boldsymbol{V}' = (V_1', V_2', V_3') = \boldsymbol{V}(\pi')$ to be the estimators with the exchange pair. For the random exchange pair (t, t'), we have the following equations:

$$
\begin{aligned}
V_1' - V_1 &= \xi_T(t, t+1, \pi(t'), \pi(t+1)) + \xi_T(t', t'+1, \pi(t), \pi(t'+1)) \\
&\quad + \xi_T(t-1, t, \pi(t-1), \pi(t')) + \xi_T(t'-1, t', \pi(t'-1), \pi(t)) \\
&\quad - \xi_T(t, t+1, \pi(t), \pi(t+1)) - \xi_T(t', t'+1, \pi(t'), \pi(t'+1)) \\
&\quad - \xi_T(t-1, t, \pi(t-1), \pi(t)) - \xi_T(t'-1, t', \pi(t'-1), \pi(t'), \\
V_2' - V_2 &= \frac{1}{T-1}\sum_{s=2}^{T} \xi_T(t, t+1, \pi(t'), s) + \frac{1}{T-1}\sum_{s=2}^{T} \xi_T(t', t'+1, \pi(t), s) \\
&\quad - \frac{1}{T-1}\sum_{s=2}^{T} \xi_T(t, t+1, \pi(t), s) - \frac{1}{T-1}\sum_{s=2}^{T} \xi_T(t', t'+1, \pi(t'), s), \\
V_3' - V_3 &= \frac{1}{T-1}\sum_{s=2}^{T} \xi_T(t-1, t, s, \pi(t')) + \frac{1}{T-1}\sum_{s=2}^{T} \xi_T(t'-1, t', s, \pi(t)) \\
&\quad - \frac{1}{T-1}\sum_{s=2}^{T} \xi_T(t-1, t, s, \pi(t)) - \frac{1}{T-1}\sum_{s=2}^{T} \xi_T(t'-1, t', s, \pi(t')).
\end{aligned}
$$

They further satisfy that

$$
\mathbb{E}^{\boldsymbol{V}}\left(\boldsymbol{V}' - \boldsymbol{V}\right) = -\boldsymbol{\Lambda}\boldsymbol{V} + \boldsymbol{R} \tag{22}
$$

where

$$
\boldsymbol{\Lambda} = \frac{2}{T-2}\begin{pmatrix} \frac{2T-3}{T-1} & -1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{R} = \left(-\frac{2}{(T-1)(T-2)} \sum_{t,t'}^{T} \xi_T(t, t', \pi(t'), \pi(t)), 0, 0 \right).
$$

To be self-contained, we re-state the following theorem to show the asymptotic normality.

Theorem 2 in [24]. Assume that $(\boldsymbol{V}, \boldsymbol{V}')$ is an exchangeable pair of random vectors such that

$$
\mathbb{E}[\boldsymbol{V}] = \boldsymbol{0}, \quad \mathbb{E}[\boldsymbol{V}\boldsymbol{V}^t] = \boldsymbol{\Sigma},
$$

with $\boldsymbol{\Sigma} \in \mathbb{R}^{3\times 3}$ symmetric and positive definite. If (22) holds and $\boldsymbol{Z}$ has a 3-dimensional standard normal distribution, we have for every three times differentiable function h,

$$
\left|\mathbb{E}h(\boldsymbol{V}) - \mathbb{E}h\left(\boldsymbol{\Sigma}^{1/2}\boldsymbol{Z}\right)\right| \le \frac{|h|_2}{4}A + \frac{|h|_3}{12}B + \left(|h|_1 + \frac{3}{2}\|\boldsymbol{\Sigma}\|^{1/2}|h|_2\right)
$$

where

$$
\begin{aligned}
\gamma^{(i)} &= \sum_{m=1}^{3} \left|\left(\boldsymbol{\Lambda}^{-1}\right)_{m,i}\right| \\
A &= \sum_{i,j=1}^{3} \gamma^{(i)} \sqrt{\operatorname{Var} \mathbb{E}^{\boldsymbol{V}} \left(V_i' - V_i\right)\left(V_j' - V_j\right)}, \\
B &= \sum_{i,j,k=1}^{3} \gamma^{(i)} \mathbb{E}\left|\left(V_i' - V_i\right)\left(V_j' - V_j\right)\left(V_k' - V_k\right)\right|, \\
C &= \sum_{i=1}^{3} \gamma^{(i)} \sqrt{\operatorname{Var} R_i}.
\end{aligned}
$$

To apply the theorem, we first note that $\mathbb{E}[\boldsymbol{V}] = \boldsymbol{0}$ may not hold. Nevertheless, we can simply de-mean $\boldsymbol{V}$ by $\mathbb{E}[\boldsymbol{V}]$, and thus the condition is satisfied. Next, it is easy to see $\boldsymbol{\gamma} = O(T)$ and we need to characterize A, B, C using (20):

- A: Let us use the analysis of $\operatorname{Var}\mathbb{E}^{\boldsymbol{V}}(V_1'-V_1)^2$ as instance. First of all, we have

$$\begin{aligned}\mathbb{E}^{\boldsymbol{V}}(V_1'-V_1)^2 &= \frac{1}{(T-1)(T-2)}\sum_{t\neq t'}^{T}(V_1(\pi')-V_1))^2\\ &= \frac{1}{(T-1)(T-2)}\sum_{t\neq t'}^{T}(\xi_T(t,t+1,\pi(t'),\pi(t+1))+\xi_T(t',t'+1,\pi(t),\pi(t'+1))\\ &+\xi_T(t-1,t,\pi(t-1),\pi(t'))+\xi_T(t'-1,t',\pi(t'-1),\pi(t))\\ &-\xi_T(t,t+1,\pi(t),\pi(t+1))-\xi_T(t',t'+1,\pi(t'),\pi(t'+1))\\ &-\xi_T(t-1,t,\pi(t-1),\pi(t))-\xi_T(t'-1,t',\pi(t'-1),\pi(t')))^2.\end{aligned}$$

  Let π' be the permutation with the exchange pair t,t' and π'' be the permutation with the exchange pair s,s'. To analyze the variance of $\mathbb{E}^{\boldsymbol{V}}(V_1'-V_1)^2$, it suffices to see that

$$\operatorname{Cov}((V_1(\pi')-V_1))^2,(V_1(\pi'')-V_1))^2)=O\left(\frac{1}{T^4}\right).$$

  This further leads to

$$\sqrt{\operatorname{Var}\mathbb{E}^{\boldsymbol{V}}(V_1'-V_1)^2}=O\left(\frac{1}{T^2}\right)$$

  Following the same procedure, we can obtain that

$$\sqrt{\operatorname{Var}\mathbb{E}^{\boldsymbol{V}}\left(V_i'-V_i\right)\left(V_j'-V_j\right)}=O\left(\frac{1}{T^2}\right).$$

- B: Let us use the analysis of $\mathbb{E}[|(V_1'-V_1)^3|]$ as instance. We take the conditioning on the exchange pair (t,t'), which gives

$$\begin{aligned}\mathbb{E}[|(V_1'-V_1)^3|] &= \frac{1}{(T-1)(T-2)}\sum_{t\neq t'}^{T}\mathbb{E}[|(\xi_T(t,t+1,\pi(t'),\pi(t+1))+\xi_T(t',t'+1,\pi(t),\pi(t'+1))\\ &+\xi_T(t-1,t,\pi(t-1),\pi(t'))+\xi_T(t'-1,t',\pi(t'-1),\pi(t))\\ &-\xi_T(t,t+1,\pi(t),\pi(t+1))-\xi_T(t',t'+1,\pi(t'),\pi(t'+1))\\ &-\xi_T(t-1,t,\pi(t-1),\pi(t))-\xi_T(t'-1,t',\pi(t'-1),\pi(t')))^3|]\leq\left(\frac{8B}{T}\right)^3=O\left(\frac{1}{T^3}\right)\end{aligned}$$

  Following the same procedure, we can obtain that $\mathbb{E}\left|\left(V_i'-V_i\right)\left(V_j'-V_j\right)\left(V_k'-V_k\right)\right| = O\left(\frac{1}{T^3}\right)$.

- C: Since $R_2=R_3=0$, we simply need to consider R_1.

$$\begin{aligned}\sqrt{\operatorname{Var}R_1} &= \frac{2}{(T-1)(T-2)}\sqrt{\operatorname{Var}\left(\sum_{t\neq t'}^{T}\xi_T(t,t',\pi(t'),\pi(t))\right)}\\ &= \frac{2}{(T-1)(T-2)}\sqrt{\operatorname{Var}(V_1)}=O\left(\frac{1}{T^{2.5}}\right)\end{aligned}$$

Putting A,B,C together, we have

$$\left|\mathbb{E}h(\boldsymbol{V})-\mathbb{E}h\left(\boldsymbol{\Sigma}^{1/2}\boldsymbol{Z}\right)\right|=O\left(\frac{1}{T}\right).$$

Note that $\Sigma_{1,1}^{1/2}=\sqrt{\operatorname{Var}_{\eta^\dagger}(\widehat{\tau})}$ has an order of $\frac{1}{\sqrt{T}}$. If we normalize V_1 by the standard deviation, this leads to the typical rate of convergence $O\left(\frac{1}{\sqrt{T}}\right)$ for asymptotic normality. □

A.5 Proof of Proposition 7

Proof. Proof of Proposition 7 We write the original variance decomposition (4):

$$\begin{aligned}\mathrm{Var}(\widehat{\tau}|\mathbb{Y}) = & \frac{8(T-2)^2}{(T-3)^2(T-1)}\left(S^t+S^c\right) - \frac{(4T-10)(T-2)}{(T-3)(T-4)(T-1)}S^\tau \\ & + \frac{1}{(T-1)(T-4)}\sum_{t=2}^{T}\left(Y_t(\mathbf{1})-Y_t(\mathbf{0})\right)^2 - \frac{1}{(T-1)(T-3)(T-4)}\sum_{t=2}^{T}(Y_t(\mathbf{1})-Y_t(\mathbf{0}))(Y_{t+1}(\mathbf{1})-Y_{t+1}(\mathbf{0})).\end{aligned}$$

First, because of

$$\sum_{t=2}^{T}(Y_t(\mathbf{1})-Y_t(\mathbf{0}))^2 + (Y_t(\mathbf{1})-Y_t(\mathbf{0}))(Y_{t+1}(\mathbf{1})-Y_{t+1}(\mathbf{0})) = \frac{1}{2}\sum_{t=2}^{T}(Y_t(\mathbf{1})+Y_{t+1}(\mathbf{1})-Y_t(\mathbf{0})-Y_{t+1}(\mathbf{0}))^2,$$

we rewrite the variance as

$$\begin{aligned}\mathrm{Var}(\widehat{\tau}|\mathbb{Y}) = & \frac{8(T-2)^2}{(T-3)^2(T-1)}\left(S^t+S^c\right) - \frac{(4T-10)(T-2)}{(T-3)(T-4)(T-1)}S^\tau \\ & + \frac{T-2}{(T-1)(T-3)(T-4)}\sum_{t=2}^{T}\left(Y_t(\mathbf{1})-Y_t(\mathbf{0})\right)^2 \\ & - \frac{1}{2(T-1)(T-3)(T-4)}\sum_{t=2}^{T}(Y_t(\mathbf{1})+Y_{t+1}(\mathbf{1})-Y_t(\mathbf{0})-Y_{t+1}(\mathbf{0}))^2.\end{aligned}$$

Removing the non-positive parts gives

$$\mathrm{Var}(\widehat{\tau}|\mathbb{Y}) \le \frac{8(T-2)^2}{(T-3)^2(T-1)}\left(S^t+S^c\right) + \frac{T-2}{(T-1)(T-3)(T-4)}\sum_{t=2}^{T}\left(Y_t(\mathbf{1})-Y_t(\mathbf{0})\right)^2.$$

Next, because of the non-negative outcomes, we have

$$\left(Y_t(\mathbf{1})-Y_t(\mathbf{0})\right)^2 \le Y_t^2(\mathbf{1})+Y_t^2(\mathbf{0}).$$

This finally leads to the upper bound

$$\mathrm{Var}_{\eta^\dagger}(\widehat{\tau}|\mathbb{Y}) \le \frac{8(T-2)^2}{(T-1)(T-3)^2}(S^t+S^c) + \frac{T-2}{(T-1)(T-3)(T-4)}\sum_{t=2}^{T}\left(Y_t^2(\mathbf{1})+Y_t^2(\mathbf{0})\right).$$

Following the similar argument in [17], we obtain an unbiased estimate for the upper bound

$$\hat{\sigma}_{\mathrm{U}}^2 = \frac{8(T-2)^2}{(T-1)(T-3)^2}(\hat{S}^t+\hat{S}^c) + \frac{4(T-2)^2}{(T-1)(T-3)^2(T-4)}\sum_{t=2}^{T}Y_t^2\mathbb{1}\{w_{t-1}=w_t\}.$$

where $\hat{S}^t, \hat{S}^c$ are the sample estimates and Y_t is the observed outcome. □