

# Population Interference In Panel Experiments

Kevin Wu Han

Department of Statistics, Stanford University

and

Iavor Bojinov

Harvard Business School

and

Guillaume Basse

Department of MS&E and Department of Statistics, Stanford University

June 6, 2023

## Abstract

The phenomenon of population interference, where a treatment assigned to one experimental unit affects another experimental unit's outcome, has received considerable attention in standard randomized experiments. The complications produced by population interference in this setting are now readily recognized, and partial remedies are well known. Less understood is the impact of population interference in panel experiments where treatment is sequentially randomized in the population, and the outcomes are observed at each time step. This paper proposes a general framework for studying population interference in panel experiments and presents new finite population estimation and inference results. Our findings suggest that, under mild assumptions, the addition of a temporal dimension to an experiment alleviates some of the challenges of population interference for certain estimands. In contrast, we show that the presence of carryover effects — that is, when past treatments may affect future outcomes — exacerbates the problem. Our results are illustrated through both an empirical analysis and an extensive simulation study.

*Keywords:* Finite-population Inference, Randomization Distribution, Potential Outcomes, Dynamic Causal Effects

# 1 Introduction

When researchers estimate causal effects from randomized experiments, they almost always make assumptions that restrict the number of counterfactual outcomes to simplify the subsequent inference. In standard experiments, where units are randomly assigned to either a treatment or control, researchers commonly assume that one unit’s assignment does not affect another unit’s response; this is usually referred to as no interference (Cox 1958, Chapter 2). In panel experiments, where units are exposed to different interventions over time (Bojinov et al. 2021), in addition to no interference, researchers regularly assume that the observed outcomes were not impacted by past assignments; this is often called the no carryover assumption (Cox 1958, Chapter 13). Although these two assumptions are useful, there are numerous empirical examples where they are violated. This mismatch between practical applications and theoretical assumptions has catalyzed a growing amount of literature dedicated to studying relaxations of these stringent conditions for either standard or panel experiments, but not both.

In standard experiments without evoking the no interference assumption, each unit’s outcome depends on the assignments received by all other experimental units. Allowing for such arbitrary population interference<sup>1</sup> makes causal inference challenging (Basse & Airoidi 2018). In practice, researchers look for an underlying structure that limits the scope of interference. For example, when studying electoral participation during a special election in 2009 in Chicago, Sinclair et al. (2012) assumed that interference occurred within-household but not across; more broadly, this type of interference has been found in many other applications, including education (Hong & Raudenbush (2006); Rosenbaum (2007)), economics (Sobel (2006); Manski (2013)) and public health (Halloran & Struchiner (1995)). Inference in this setting is challenging because interference increases the number of potential outcomes and makes observations dependent. Aronow & Samii (2017) introduce a general framework for studying causal inference with interference: they introduce the concept of exposure mapping, define useful estimands, and construct asymptotically valid confidence intervals based on the Horvitz-Thompson estimator.

The literature on panel experiments has similarly shifted towards relaxing the no carryover effects assumption that precludes outcomes from being impacted by past assignments. For example, in the most extreme case, Bojinov & Shephard (2019) allows for arbitrary carryover effects when studying whether algorithms or humans are better at executing large financial orders; more generally, these types of relaxations have also been studied in economics (Angrist & Kuersteiner 2011, Rambachan & Shephard 2019), epidemiology (Robins 1986, Robins et al. 1999a), public health (Boruvka et al. 2018), and political science (Blackwell & Glynn 2018). Similarly to relaxing the no interference assumption, removing the no carryover assumption enables researchers to develop and explore a richer class of causal estimands that capture both the contemporaneous and

---

<sup>1</sup>We use the term *population interference* to emphasize that the interference occurred across units.

delayed causal effects (Bojinov & Shephard 2019). The latter is particularly important for technology companies seeking to understand the long-term impact of their interventions (Basse, Ding & Toulis 2019, Hohnhold et al. 2015). Similarly to the population interference setting, researchers use the analogous Horvitz-Thompson type estimator estimators to analyze experiments with carryover effects.

An apparent gap in the literature is an understanding of whether the possibility of running a panel experiment alleviates the challenges posed by population interference or makes them worse. This is particularly important for researchers wishing to run field experiments for two reasons. First, it is often the case the researchers are constrained on the maximal number of experimental units they can recruit, for instance, because of costs or limits in the total population. Second, population interference often leads to increased uncertainty that reduces only by increasing the sample size. Panel experiments can alleviate these as it is often cheaper to change an experimental unit’s treatment than to recruit more subjects, and uncertainty tends to decrease as the sample size and the number of time period increase. However, what happens when there is population interference and carryover effects is unclear.

To address this gap, we introduce a unifying framework for studying panel experiments with population interference. We begin by focusing on panel experiments with population interference but no carryover effects (Section 3). Here, we provide asymptotically valid confidence intervals for estimands defined at specific time periods and estimands that average contrasts over multiple time periods. We also introduce a novel class of assumptions enabling us to leverage past data to improve inference at a given time. Together, our results show that using panel experiments when there is population interference allows us to achieve valid inference under much weaker conditions on the population interference and even drop all restrictions for large time horizons. These results should be particularly encouraging for researchers wishing to run field experiments when the number of experimental subjects is constrained, as is often the case in Economics (for example, Andreoni & Samuelson (2006)) and Management Science (for example, Choudhury et al. (2021)).

We then tackle the most general setting featuring both population and temporal interference (Section 4). Under additional assumptions, we derive a general central limit theorem, which fails to provide the same clear benefit because of the data complexity caused by carryover effects. We also state asymptotic results for a restricted type of mixed interference that generalizes the usual stratified interference to panel experiments and provides a blueprint for deriving additional results in specific contexts. Here we show a clear benefit that incorporating a temporal dimension allows us to relax the main restriction on the maximal cluster size to obtain valid inference. For researchers, these results are slightly less encouraging but, nevertheless, provide an essential next step in understanding how to leverage panel experiments in real-world settings.

Finally, Section 2 details our setup by introducing the potential outcomes framework, our causal

estimands and corresponding estimators, and the randomization-based framework that we leverage for all our results. We conclude with simulations (Section 5), empirical applications (Section 6), and a discussion (Section 7). The Appendix contains a detailed discussion of inference under population interference for standard experiments, all proofs, and additional simulations.

## 2 Setup

### 2.1 Assignments

Consider a randomized experiment occurring over  $T$  periods on a finite population of  $n$  experimental units. At each time step  $t \in \{1, \dots, T\}$ , unit  $i \in \{1, \dots, n\}$  can be assigned to treatment ( $W_{i,t} = 1$ ) or control ( $W_{i,t} = 0$ ); extensions to non-binary treatments are straightforward. We denote by  $W_{i,1:t} = (W_{i,1}, W_{i,2}, \dots, W_{i,t})$  the assignment path up to time  $t$  for unit  $i$ ,  $W_{1:n,t}$  the assignment vector for all  $n$  units at time step  $t$  and  $W_{1:n,1:t} \in \{0, 1\}^{n \times t}$  the assignment matrix. Hence, for each  $i$  and  $t$ ,  $W_{i,1:t}$  is a vector of length  $t$ ,  $W_{1:n,t}$  is a vector of length  $n$  and  $W_{1:n,1:t}$  is a matrix of dimension  $n \times t$ .

We define an assignment mechanism (or design) to be the probability distribution of the assignment matrix  $\mathbb{P}(W_{1:n,1:T})$ .

For example, in a Bernoulli design, the assignment mechanism is independent across time and units such that  $\mathbb{P}(W_{i,t} = 1) = p$  for all  $i, t$ . More generally, we will often work with assignments that are temporally independent.

**Definition 1** (Temporally independent assignment mechanism). *We say an assignment mechanism is temporally independent if  $W_{1:n,t}$  and  $W_{1:n,t'}$  are independent for any  $t$  and  $t'$*

Following much of the literature on analyzing complex experiments, we adopt the randomization-based approach to inference, in which the assignment mechanism is the only source of randomness (Kempthorne 1955); see Abadie et al. (2020) for a recent review. Throughout, we use lower cases  $w$  with the appropriate subscript for realizations of the assignment matrix  $W$ .

### 2.2 Potential outcomes and exposure mappings

The goal of causal inference is to study how an intervention impacts an outcome of interest. Following the potential outcomes formulation, for panel experiments without any assumptions, each unit  $i$  at time  $t$  has  $2^{nT}$  potential outcomes corresponding to the total number of distinct realizations of the assignment matrix, denoted by  $Y_{i,t}(w_{1:n,1:T})$ . For simplicity, we assume that the potential outcomes are one-dimensional, although it is straightforward to relax this assumption.

In randomized experiments, where we control the assignment mechanism, the outcomes at time  $t$  are not impacted by future assignments that have yet to be revealed to the units (Bojinov &

Shephard 2019). This assumption drastically reduces the total number of potential outcomes<sup>2</sup> and will be implicitly made throughout this paper. We now formally define the no carryover effects and no population interference assumption.

**Definition 2** (No carryover effect and population interference). *We say that there is no carryover effect if and only if*

$$Y_{i,t}(w_{1:n,1:t}) = Y_{i,t}(w'_{1:n,1:t}) \text{ whenever } w_{1:n,t} = w'_{1:n,t}.$$

*And we say that there is no population interference if and only if*

$$Y_{i,t}(w_{1:n,1:t}) = Y_{i,t}(w'_{1:n,1:t}) \text{ whenever } w_{i,1:t} = w'_{i,1:t}.$$

If we make both assumptions, inference is relatively straightforward. However, if we only assume that there is population interference and no carryover effects, inference is still impossible without any assumptions on the population interference structure (Basse & Airoidi 2018). One way forward is to assume that the outcomes of unit  $i$  depend only on the treatments assigned to a subset of the population. This intuition extends more generally to the assertion that the outcome of unit  $i$  at time  $t$  depends on a low-dimensional representation of  $w_{1:n,1:t}$ . Formally, for each unique  $i, t$  pair we define the exposure mapping  $f_{i,t} : \{0, 1\}^{n \times t} \rightarrow \Delta$ , where  $\Delta$  is the set of possible exposures<sup>3</sup> (Aronow & Samii 2017). On the other hand, if we assume that there is no population interference, again, we would need additional assumptions to obtain valid randomization-based inference (Bojinov et al. 2021).

Defining exposure mappings in this flexible manner allows us to unify and transparently consider restrictions on the population interference and the duration of the carryover effect. Throughout this paper, we restrict our focus to properly specified time-invariant exposure mappings, which are formally defined below.

**Assumption 1** (Properly specified time-invariant exposure mapping). *The exposure mappings are properly specified if, for all pairs  $i \in \{1, \dots, n\}$  and  $t \in \{1, \dots, T\}$ , and any two assignment matrices  $w_{1:n,1:t}$  and  $w'_{1:n,1:t}$ ,*

$$\begin{aligned} Y_{i,t}(w_{1:n,1:t}) &= Y_{i,t}(w'_{1:n,1:t}) \text{ whenever} \\ f_{i,t}(w_{1:n,1:t}) &= f_{i,t}(w'_{1:n,1:t}). \end{aligned}$$

---

<sup>2</sup>The assumption, known as non-anticipating potential outcomes (Bojinov & Shephard 2019), can be violated if experimental units are told what their future assignments will be and modify their present behavior as a result. For instance, this could occur for shoppers who expect to receive a considerable discount on a subsequent day and may curtail their spending until they receive the discount.

<sup>3</sup>To make exposure mappings useful, we assume the cardinality of  $\Delta$  is (substantially) smaller than  $n \times t$ .

For  $p \in \{1, \dots, T\}$ , we say the exposure mappings are  $p$ -time-invariant if for any  $t, t' \in \{p, \dots, T\}$  and any unit  $i$ ,

$$f_{i,t}(w_{1:n,1:t}) = f_{i,t'}(w_{1:n,1:t'}) \text{ whenever}$$

$$w_{1:n,t-p+1:t} = w_{1:n,t'-p+1:t'}.$$

The exposure mappings are time-invariant if the exposure mappings are  $p$ -time-invariant for some  $p \in \{1, \dots, T\}$ . We say the exposure mappings are properly specified time-invariant exposure mappings if they are both properly specified and time-invariant.

Properly specified exposure mappings can be thought of as defining “effective treatments,” allowing us to write:

$$Y_{i,t}(w_{1:n,1:t}) = Y_{i,t}(f_{i,t}(w_{1:n,1:t})) = Y_{i,t}(h_{i,t}),$$

where  $h_{i,t} = f_{i,t}(w_{1:n,1:t}) \in \Delta$ . Time-invariant exposure mappings constrain the relationship between experimental units to be invariant over time. Specifically, it does not allow the exposure mappings to change across time. For example, if at time  $t = 1$ , the outcomes depend on the fraction of treated neighbors in the graph, then it cannot be the case that at time  $t = 2$  the outcomes now depend on the number of treated neighbors in the graph. We will see why such an invariance assumption is necessary for the next section when we define causal effects. Of course, the validity of Assumption 1 depends on the exact definition of the exposure mapping and should be informed by the empirical context.

Throughout this paper, we consider a special class of exposure mappings that restrict the outcomes of unit  $i$  to depend only on the assignments of a predefined subset of units that we refer to as  $i$ 's neighborhood and index by  $\mathcal{N}_i$ ; note that the index set is not dynamically changing over time. For example, for units connected through a social network,  $\mathcal{N}_i$  indexes the set of nodes connected to  $i$  by an edge; for units organized households,  $\mathcal{N}_i$  indexes the set of units that live in the same household as  $i$ ; and for units located in space,  $\mathcal{N}_i$  indexes the set of units who are at most a certain distance away from unit  $i$ .

**Definition 3** (Locally Effective Assignments (LEA)). *We say the assignments and exposure mappings are locally effective if the exposure mappings are  $p$ -time-invariant for some  $p \in \{1, \dots, T\}$  and*

$$f_{i,t}(w_{1:n,1:t}) = f_{i,t}(w_{\mathcal{N}_i,t-p+1:t}),$$

*with the convention that  $w_{\mathcal{N}_i,t-p+1:t} = w_{\mathcal{N}_i,1:t}$  for  $t - p + 1 \leq 0$ .*

Although LEA imposes further structure, it still provides a great deal of flexibility as it incorporates all notions of traditional population interference and temporal carryover effects as special cases. For example, fixing  $p = 1$  makes the exposure values depend only on current assignments,

which is equivalent to usual population interference. On the other hand, fixing  $\mathcal{N}_i = \{i\}$  is equivalent to the no interference assumption imposed on panel experiments in Bojinov et al. (2021). Of course, these special cases are interesting and extensively studied, but our general formulation’s real benefit is to consider scenarios where there is both population interference and carryover effects.

**Example 1** (Example of Locally Effective Assignments). *We consider an example where the exposure values depend on past assignments. In particular, let*

$$f_{i,t}(w_{1:n,1:t}) = (w_{i,t-1}, w_{i,t}, u_{i,t-1}, u_{i,t})$$

where  $u_{i,t-1} = \frac{1}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} w_{j,t-1}$  and  $u_{i,t} = \frac{1}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} w_{j,t}$ ; we use  $|\mathcal{A}|$  to denote the cardinality of the set  $\mathcal{A}$ . Hence, one unit’s assignment and the fraction of treated neighbors at the previous time step matter as well. This is a special case of LEA with  $p = 2$ . In this example, the exposure mappings are 2-time-invariant: for  $t, t' \geq 2$ , if  $w_{1:n,(t-1):t} = w_{1:n,(t'-1):t'}$  then  $f_{i,t}(w_{1:n,1:t}) = f_{i,t'}(w_{1:n,1:t'})$ .

One limitation of LEA( $p$ ) assumption is that it can not directly capture long-range decaying dependency on past assignments and population interference beyond a unit’s neighborhood. Such long-range decaying dependency on time is uncommon in econometrics literature (Judson & Owen 1999, Wooldridge 2010). For example, if we consider the following parametric model (Wooldridge 2010):

$$Y_{i,t} = \rho Y_{i,t-1} + \beta W_{i,t} + \epsilon_{i,t}, \text{ where } \rho \in (-1, 1) \text{ and } \mathbb{E}[\epsilon_{i,t} | Y_{i,1}, \dots, Y_{i,t}, W_{i,1}, \dots, W_{i,t}] = 0,$$

we have an infinite-long decaying dependency on past assignments on the current outcomes. To capture this, we would need to set  $p = T$  in the LEA( $p$ ) causal effect, which, although possible, is unlikely to be practically useful as we would not be able to estimate this estimand with any reasonable precision. Despite this limitation, the LEA( $p$ ) assumption allows for a great deal of flexibility as it does not require imposing modeling assumptions on the outcomes and is still useful in many real-world situations.

Finally, population interference beyond local interference has also been studied in the econometrics literature (Manski 1993, Bramoullé et al. 2009, Leung 2022). We leave the extensions of our work to this setting as future work.

### 2.3 Causal effects

Causal effects, within the potential outcomes framework, are defined as contrasts of each unit’s potential outcomes under alternate assignments (Imbens & Rubin 2015). As the number of possible contrasts grows exponentially with the number of distinct potential outcomes, we focus on two important special cases.

The first—which is well-defined regardless of the interference structure—compares the difference in the potential outcomes across two extreme scenarios: assigning every unit to treatment,  $W_{1:n,1:t} = 1_{1:n,1:t}$ , as opposed to control,  $W_{1:n,1:t} = 0_{1:n,1:t}$ .

**Definition 4** (Total effect at time  $t$ ). *The total effect at time  $t$  is*

$$\tau_t^{TE} = \frac{1}{n} \sum_{i=1}^n Y_{i,t}(1_{1:n,1:t}) - \frac{1}{n} \sum_{i=1}^n Y_{i,t}(0_{1:n,1:t}).$$

Our total effect at time  $t$  corresponds to the Global Average Treatment Effect sometimes used in single time experiments (Ugander & Yin 2020). In the absence of interference and carryover effects, the total effect at time  $t$  reduces to the usual average treatment effect at time  $t$ .

The second—which requires Assumption 1—provides a much richer class of causal effects with important practical applications. The TEC estimand is the generalization of the usual exposure contrast estimands (Aronow & Samii 2017) to the panel experiment setting. Hereafter, the letter  $k$  will always represent values in  $\Delta$ .

**Definition 5** (Temporal exposure contrast (TEC)). *For any time step  $t$  and exposure values  $k, k' \in \Delta$ , we define the temporal exposure contrast between  $k$  and  $k'$  to be*

$$\tau_t^{k,k'} = \frac{1}{n} \sum_{i=1}^n Y_{i,t}(k) - \frac{1}{n} \sum_{i=1}^n Y_{i,t}(k')$$

Here, we implicitly assume that for every unit, the potential outcome is well defined for all values of  $k \in \Delta$ . This assumption precludes situations where the range of the exposure mappings depends on  $\mathcal{N}_i$ . Further, note that if there exist carryover effects, then TEC may not be well-defined for the first few time steps. In this case, we may assume that all units are in the control group prior to the first time step in the panel experiment.

In panel experiments, researchers are often less interested in the idiosyncratic effects at each point in time and instead focus on the temporal average causal effect that captures the intervention’s average impact across both time and units (Boruvka et al. 2018, Bojinov & Shephard 2019, Bojinov et al. 2021, 2022, Hu & Wager 2022, Xiong et al. 2019). For example, Bojinov & Shephard (2019) are not interested in the relative difference between an algorithm or a human executing a large financial order on an arbitrary day of the experiment but are instead interested in the average difference across multiple trades on the same market. Similarly, technology companies like DoorDash (Tang et al. 2020), Lyft Chamandy (2016), and Uber Farronato et al. (2018) use switchback design for panel experiments and consider average effect across time to make product decisions. Such companies rarely are interested in the effect at time  $t$ —for instance, on Tuesday 2-3 pm—but instead, want to understand the average performance throughout the experiment.



**Definition 6** (Average total effect). *The average total effect is*

$$\overline{\tau^{TE}} = \frac{1}{T} \sum_{t=1}^T \tau_t^{TE}.$$

Similar to the total effect, in many applications, we are interested in the TEC's temporal average.

**Definition 7** (Average temporal exposure contrast (ATEC)). *For any exposure values  $k, k' \in \Delta$ , we define the average temporal exposure contrast between  $k$  and  $k'$  to be*

$$\bar{\tau}^{k,k'} = \frac{1}{T} \sum_{t=1}^T \tau_t^{k,k'}$$

Without assuming that the exposure mappings are time-invariant, the definition of the ATEC becomes more cumbersome as an exposure  $k \in \Delta$  may be in the image of  $f_{i,t}$  for some  $t$ , but not in the image of  $f_{i,t'}$ . That is,  $Y_{i,t}(k)$  might be well-defined while  $Y_{i,t'}(k)$  is not, which makes taking temporal averages difficult.

Of course, our causal estimands are not exhaustive, and there are many other causal estimands of interest. For example, there is a vast literature in econometrics and statistics studying estimation and inference of spillover effects under either different designs or different model assumptions (Robins 1986, Robins et al. 1999b, Leung 2020, Bramoullé et al. 2020, Vazquez-Bare 2022).

## 2.4 Estimation and inference

### 2.4.1 The observed data

For any choice of exposure mappings  $\{f_{i,t}\}$ , the observed assignment path  $W_{1:n,1:t}$  induces the exposure  $H_{i,t} = f_{i,t}(W_{1:n,1:t})$  for each  $i$  and  $t$ ; in particular, the assignment mechanism  $\mathbb{P}(W_{1:n,1:t})$  induces a distribution for the exposures  $\mathbb{P}(H_{i,t})$  for each  $i$  and  $t$ . Under Assumption 1<sup>4</sup>, the observed outcomes  $Y_{i,t}$  for unit  $i$  at time  $t$  can therefore be written:

$$Y_{i,t} = \sum_{k \in \Delta} \mathbf{1}(H_{i,t} = k) Y_{i,t}(k), \quad \forall i \in 1, \dots, n, \forall t \in 1, \dots, T,$$

We use these observed data to estimate the causal effects defined in 2.3.

### 2.4.2 Estimation

For the different interference structures studied in the following sections, we will rely on Horvitz-Thompson estimators (Horvitz & Thompson 1952), or variations of it; e.g., to estimate  $\tau_t^{k,k'}$ , we

---

<sup>4</sup>We additionally assume that each unit fully complies with the assignment, leaving the relaxation of this assumption as future work.

will use:

$$\hat{\tau}_t^{k,k'} = \frac{1}{n} \sum_{i=1}^n \frac{\mathbf{1}(H_{i,t} = k)}{\mathbb{P}(H_{i,t} = k)} Y_{i,t} - \frac{1}{n} \sum_{i=1}^n \frac{\mathbf{1}(H_{i,t} = k')}{\mathbb{P}(H_{i,t} = k')} Y_{i,t}. \quad (1)$$

Taking the temporal average of (1) provides a natural estimator of  $\bar{\tau}^{k,k'}$ ,

$$\hat{\bar{\tau}}^{k,k'} = \frac{1}{T} \sum_{t=1}^T \hat{\tau}_t^{k,k'}. \quad (2)$$

Similarly, if we let  $h_{i,t}^1 := f_{i,t}(1_{1:n,1:t})$  and  $h_{i,t}^0 := f_{i,t}(0_{1:n,1:t})$ , then we can estimate total effect at time  $t$  (c.f. Definition 4) by the following estimator

$$\hat{\tau}_t^{TE} = \frac{1}{n} \sum_{i=1}^n \frac{\mathbf{1}(H_{i,t} = h_{i,t}^1)}{\mathbb{P}(H_{i,t} = h_{i,t}^1)} Y_{i,t} - \frac{1}{n} \sum_{i=1}^n \frac{\mathbf{1}(H_{i,t} = h_{i,t}^0)}{\mathbb{P}(H_{i,t} = h_{i,t}^0)} Y_{i,t}. \quad (3)$$

Again, we have a natural estimator of average total effect induced by the above estimator

$$\hat{\bar{\tau}}^{TE} = \frac{1}{T} \sum_{t=1}^T \hat{\tau}_t^{TE}. \quad (4)$$

The properties of these estimators are discussed in details in the rest of this manuscript.

### 2.4.3 Randomization-based inference

Throughout this paper, we adopt the randomization-based (sometimes called design-based) framework—that is, we consider the potential outcomes as fixed, with the assignment being the only source of randomness. Equivalently, we can view the randomization-based framework as conditioning on the full set of potential outcomes and only using the randomness in the assignment for inference. This framework has seen a recent uptake in causal inference (Lin 2013, Li & Ding 2017, Li et al. 2019, Basse & Bojinov 2020) and has become standard for analyzing experiments with population interference (Aronow & Samii 2017, Sävje et al. 2017, Basse & Feller 2018, Chin 2018) and unbounded carryover effects (Bojinov & Shephard 2019, Rambachan & Shephard 2019, Bojinov et al. 2021, 2022). An additional benefit to adopting a randomization-based approach in the context of population interference is that it explicitly removes the challenges posed by correlations between the potential outcomes and the social relationships, as the potential outcomes are fixed (unknown) constants.

There are two dominant inferential strategies within the randomization framework. The first is to use Fisher (conditional) randomization tests for sharp null hypotheses of no exposure effects, or for pairwise null hypotheses contrasting two exposures. While these tests deliver p-values that are exact and non-asymptotic, they are challenging to run with complex exposure mappings (Athey et al. 2018, Basse, Feller & Toulis 2019, Puelz et al. 2019).

The second, which we focus on in this paper, is to construct confidence intervals based on the asymptotic distribution of our estimators, which can be used for testing if there is an effect on average. Intuitively, the asymptotic distribution represents a sequence of hypothetical randomized experiments in which either the number of units increases, the number of time steps increases, or both (Li & Ding 2017, Bojinov et al. 2021). Within each step, we apply the analogous assignment mechanism, obtain the observed data, and compute our proposed estimand to estimate the causal effect of interest (Aronow & Samii 2017, Chin 2018).

Under the randomization framework, it is easy to show that the Horvitz-Thompson estimators  $\hat{\tau}_t^{k,k'}$ ,  $\hat{\bar{\tau}}^{k,k'}$ ,  $\hat{\tau}_t^{TE}$  and  $\hat{\bar{\tau}}^{TE}$  are unbiased for  $\tau_t^{k,k'}$ ,  $\bar{\tau}^{k,k'}$ ,  $\tau_t^{TE}$  and  $\bar{\tau}^{TE}$ , respectively<sup>5</sup>; obtaining central limit theorems in this setting, however, is notoriously challenging. In the next two sections, we develop such results for the above four estimators under different experiment assumptions.

### 3 Panel experiments with population interference and no carryover effects

Panel experiments are particularly helpful when there is population interference but no carryover effects—a setting we refer to as puper population interference. This situation commonly occurs when the treatment has a relatively short-lived effect; for example, as is the case for most digital experiments on networks and platforms (see the discussion in Kohavi et al. (2020), Bojinov & Gupta (2022)). In this setting, inference for the TEC at a fixed point in time is equivalent to the standard experimental setup (see Appendix A for a full discussion of this setting, including a new central limit theorem that illustrated a fundamental trade-off between the interference structure and the design of the experiment).

We can leverage time and move beyond the standard experimental setup in two ways. First, we can focus on inference for the ATEC that captures the average effect across units and time. Here, we show that varying the treatment over time allows us to handle settings with more expansive population interference. Second, we revisit the TEC and provide a variance reduction technique that leverages a stability assumption that limits the change in the potential outcomes across time for the same unit receiving the same treatment. Together, our results demonstrate the potential of leveraging panel experiments when there are no carryover effects.

#### 3.1 Average temporal exposure contrast

There are three distinct asymptotic regimes when considering inference for the ATEC and its natural estimator  $\hat{\tau}^{k,k'}$ : (1)  $T$  fixed and  $n \rightarrow \infty$ ; (2)  $T \rightarrow \infty$  and  $n \rightarrow \infty$ ; (3)  $T \rightarrow \infty$  and  $n$  fixed. An important insight from this section is that inference in these three regimes requires different

---

<sup>5</sup>For example, see Bojinov & Shephard (2019) and Aronow & Samii (2017) for explicit proof.

constraints on the population interference mechanism. Roughly speaking, the larger  $T$  is relative to  $n$ , the more interference we can tolerate.

### 3.1.1 Assumptions

For most of our central limit theorem results, we require a notion of the dependency graph for a collection of random variables. We define the dependency graph  $G_{n,t}$  for  $H_{1,t}, \dots, H_{n,t}$  to be the graph with vertices  $V_t = \{1, \dots, n\}$  and edges  $E_t$  such that  $(i, j) \in E$  if and only if  $H_{i,t}$  and  $H_{j,t}$  are not independent. The graph  $G_n$  models the dependency relationship among  $n$  random variables  $H_{1,t}, \dots, H_{n,t}$ . Notice that the dependency graph depends both on the interference structure and on the assignment mechanism. Finally, denote by  $d_{n,t}$  the maximal number of dependent exposure values for any unit  $i$  at time step  $t$  and let  $d_n = \limsup_{t \rightarrow \infty} d_{n,t}$  with the convention that for fixed  $T$ ,  $d_n = \max\{d_{n,1}, \dots, d_{n,T}\}$ . See Appendix A.2 for the derivation of  $d_{n,t}$  in several specific contexts.

Throughout this subsection, we work exclusively with temporally independent assignment mechanisms, see Definition 1. Our central limit theorems also require three additional assumptions. The first two assumptions bound the potential outcomes and the inverse probabilities of exposure.

**Assumption 2** (Uniformly bounded potential outcomes). *Assume that all the potential outcomes are uniformly bounded, i.e.,  $|Y_{i,t}(k)| \leq M$  for some  $M$  and for all  $i \in \{1, n\}$ ,  $k \in \Delta$ , and  $t \geq 1$ .*

**Assumption 3** (Overlap). *Assume all the exposure probabilities are bounded away from 0 and 1, i.e.,  $\exists \eta > 0$  such that for all  $i \in \{1, n\}$ ,  $k \in \Delta$ , and  $t \geq 1$ ,  $0 < \eta \leq \pi_i(k) \leq 1 - \eta < 1$ .*

Assumptions 2 and 3 are standard in the causal inference literature (Aronow & Samii (2017); Leung (2022)). Assumption 2 holds in most practical applications as realizations of the outcome variables are almost always bounded. Assumption 3 is necessary as vanishing exposure probabilities make the causal question ill-defined as we cannot observe the associated potential outcomes.

The next assumption rules out the existence of a pathological subsequence  $n_k$  along which the limiting variance of our estimator is zero.

**Assumption 4** (Nondegenerate asymptotic variance). *Assume that  $\liminf_{n \rightarrow \infty} \text{Var}(\sqrt{n}\hat{\tau}_t^{k,k'}) > 0$  for any  $t$ .*

As a consequence of this assumption, for each  $t$ , there exists a constant  $c > 0$  such that  $\text{Var}(\sqrt{n}\hat{\tau}_t^{k,k'}) \geq c$  for all sufficiently large  $n$ . This type of assumption seems unavoidable, even in settings without interference (see, e.g., Corollary 1 in Guo & Basse (2021), and subsequent discussion).

### 3.1.2 Three central limit theorems

We now state and discuss our three new central limit theorems for the ATEC under population interference but no carryover effects.

**Theorem 1.** *Suppose we have pure population interference and a temporally independent assignment mechanism, then for any  $T$ , under Assumption 1-4 and the condition that  $d_n = o(n^{1/4})$ , we have*

$$\frac{\sqrt{nT}(\hat{\tau}^{k,k'} - \bar{\tau}^{k,k'})}{\sqrt{\frac{1}{T} \sum_{t=1}^T \sigma_{n,t}^2}} \xrightarrow{d} \mathcal{N}(0, 1),$$

as  $n \rightarrow \infty$ , where  $\sigma_{n,t}^2 = \text{Var}(\sqrt{n}\hat{\tau}_t^{k,k'})$ .

This first theorem states a central limit theorem for the regime where  $T$  is fixed and  $n \rightarrow \infty$ , making it relevant for applications where  $n$  is much larger than  $T$ . The assumption that  $d_n = o(n^{1/4})$  quantifies the dependence among observations due to interference. If we compare this result to the analogous CLT in the non-temporal experimental setup with interference, Theorem 7 in Appendix A, we have the same requirement, namely  $d_n = o(n^{1/4})$ . Intuitively, this is because this asymptotic regime is closest to the standard setting with no temporal dimension—any finite number of time periods  $T$  is negligible compared with infinitely many observations  $n$ .

At the other extreme, we consider the regime where  $T \rightarrow \infty$  and  $n$  is fixed:

**Theorem 2.** *Suppose we have pure population interference, a temporally independent assignment mechanism, and Assumptions 1, 2, 3 are satisfied. Let  $\sigma_{n,t}^2 = \text{Var}(\sqrt{n}\hat{\tau}_t^{k,k'})$ , we further assume that  $\frac{1}{T} \sum_{t=1}^T \sigma_{n,t}^2$  is bounded away from 0 for any  $T$ . We then have*

$$\frac{\sqrt{nT}(\hat{\tau}^{k,k'} - \bar{\tau}^{k,k'})}{\sqrt{\frac{1}{T} \sum_{t=1}^T \sigma_{n,t}^2}} \xrightarrow{d} \mathcal{N}(0, 1),$$

as  $T \rightarrow \infty$ .

This central limit theorem makes no assumption whatsoever on the interference mechanism, beyond assuming that there are no carryover effects: in particular, we allow a unit's outcome to depend on any other unit's assignment. This perhaps surprising fact sheds some light into the nature of inference for the ATEC, and how it differs from the TEC. Intuitively, a central limit theorem requires enough “nearly independent” observations: this means that even if at any time step  $t$ , the observations are all correlated, we can still have infinitely many independent observations if: (1) observations are uncorrelated across time and (2) we observe infinitely many time periods.

The next theorem formalizes this intuition, by making the trade-off between the growth rates of  $T$  and  $d_n$  explicit:

**Theorem 3.** *Suppose we have pure population interference, a temporally independent assignment mechanism, and Assumption 1-4 are satisfied, then for  $T = T(n)$  such that either*

$$\frac{n}{T} \rightarrow 0 \tag{5}$$

or

$$\frac{\min\{d_n^2, n\}}{\sqrt{nT}} \rightarrow 0 \tag{6}$$

holds, we have

$$\frac{\sqrt{nT}(\hat{\tau}^{k,k'} - \bar{\tau}^{k,k'})}{\sqrt{\frac{1}{T} \sum_{t=1}^T \sigma_{n,t}^2}} \xrightarrow{d} \mathcal{N}(0, 1),$$

as  $n \rightarrow \infty$ , where  $\sigma_{n,t}^2 = \text{Var}(\sqrt{n}\hat{\tau}_t^{k,k'})$ .

Condition (5) is actually a special case of condition (6): if we do not impose any assumptions on the interference,  $\min\{d_n^2, n\}$  is just  $n$ , so we need  $\frac{n}{\sqrt{nT}} \rightarrow 0$ , which is equivalent to  $\frac{n}{T} \rightarrow 0$ . Condition 6 gives us more subtle control over the rate of growth required of  $T$  for any given level of interference. For instance, while for finite  $T$  we would require  $d_n = o(n^{1/4})$ , if  $T$  grows as  $T(n) = \sqrt{n}$  we only require  $d_n = o(n^{1/2})$ . As with the previous theorem, the intuition behind this result is that as  $d_n$  becomes larger, the number of “nearly independent” observations at each time point shrinks — this must be counterbalanced by an increase in the the number of temporal observation, i.e, an increase in the rate of  $T = T(n)$ .

### 3.1.3 Variance bounds and estimation

Unfortunately, as is typical in finite population causal inference,  $\text{Var}(\hat{\tau}_t^{k,k'})$  contains terms that are products of potential outcomes that can never be simultaneously observed from a single experiment, making it non-identifiable (Basse & Bojinov 2020). Instead, researchers derive an upper bound to the variance and compute unbiased estimates for this bound, allowing them to conduct conservative inference (i.e., derive confidence intervals with higher coverage than the nominal level). Without making assumptions on the assignment mechanism, we can obtain a simple bound by replacing all non-observable products of potential outcomes with the sum of their squares, we denote the estimate of the bound by  $\widehat{\text{Var}}(\sqrt{n}\hat{\tau}_t^{k,k'})$ . The specific expression can be found in the following proposition:

**Proposition 1.** (*Estimator of variance*) Let,

$$\begin{aligned}
\widehat{\text{Var}}(\sqrt{n}\hat{\tau}_t^{k,k'}) &= \frac{1}{n} \left\{ \sum_{i=1}^n \mathbf{1}(H_{i,t} = k)(1 - \pi_{i,t}(k)) \left[ \frac{Y_{i,t}}{\pi_{i,t}(k)} \right]^2 + \sum_{i=1}^n \sum_{j \neq i, \pi_{ij}(k)=0} \left[ \frac{\mathbf{1}(H_{i,t} = k)Y_{i,t}^2}{2\pi_{i,t}(k)} + \frac{\mathbf{1}(H_{j,t} = k)Y_{j,t}^2}{2\pi_{j,t}(k)} \right] \right. \\
&\quad + \sum_{i=1}^n \sum_{j \neq i, \pi_{ij}(k) > 0} \mathbf{1}(H_{i,t} = k)\mathbf{1}(H_{j,t} = k) \times \frac{\pi_{ij}(k) - \pi_{i,t}(k)\pi_{j,t}(k)}{\pi_{ij}(k)} \frac{Y_{i,t}}{\pi_{i,t}(k)} \frac{Y_{j,t}}{\pi_{j,t}(k)} \\
&\quad + \sum_{i=1}^n \mathbf{1}(H_{i,t} = k')(1 - \pi_{i,t}(k')) \left[ \frac{Y_{i,t}}{\pi_{i,t}(k')} \right]^2 + \sum_{i=1}^n \sum_{j \neq i, \pi_{ij}(k')=0} \left[ \frac{\mathbf{1}(H_{i,t} = k')Y_{i,t}^2}{2\pi_{i,t}(k')} + \frac{\mathbf{1}(H_{j,t} = k')Y_{j,t}^2}{2\pi_{j,t}(k')} \right] \\
&\quad + \sum_{i=1}^n \sum_{j \neq i, \pi_{ij}(k') > 0} \mathbf{1}(H_{i,t} = k')\mathbf{1}(H_{j,t} = k') \times \frac{\pi_{ij}(k') - \pi_{i,t}(k')\pi_{j,t}(k')}{\pi_{ij}(k')} \frac{Y_{i,t}}{\pi_{i,t}(k')} \frac{Y_{j,t}}{\pi_{j,t}(k')} \\
&\quad - 2 \sum_{i=1}^n \sum_{j \neq i, \pi_{ij}(k,k') > 0} \left( \pi_{ij}(k,k') - \pi_{i,t}(k)\pi_{j,t}(k') \right) \times \frac{\mathbf{1}(H_{i,t} = k)\mathbf{1}(H_{j,t} = k')}{\pi_{ij}(k,k')} \frac{Y_{i,t}}{\pi_{i,t}(k)} \frac{Y_{j,t}}{\pi_{j,t}(k')} \\
&\quad \left. + 2 \sum_{i=1}^n \sum_{j \neq i, \pi_{ij}(k,k')=0} \left[ \frac{\mathbf{1}(H_{i,t} = k)Y_{i,t}^2}{2\pi_{i,t}(k)} + \frac{\mathbf{1}(H_{j,t} = k')Y_{j,t}^2}{2\pi_{j,t}(k')} \right] \right\},
\end{aligned}$$

where  $\pi_i(k) = \mathbb{P}(H_{i,t} = k)$  and  $\pi_{i,j} = \mathbb{P}(H_{i,t} = k \& H_{j,t} = k')$ . Then  $\mathbb{E} \left[ \widehat{\text{Var}}(\sqrt{n}\hat{\tau}_t^{k,k'}) \right] \geq \text{Var}(\sqrt{n}\hat{\tau}_t^{k,k'})$ .

We drop the subscript  $t$  to ease notations. With the above proposition and central limit theorems, inference proceeds as follows:

**Proposition 2.** Suppose Theorem 1 or 3 holds, then for any  $\delta > 0$ ,

$$\mathbb{P} \left( \hat{\tau}^{k,k'} \in \left[ \hat{\tau}^{k,k'} - \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{1-\delta}} \sqrt{\frac{1}{T^2} \sum_{t=1}^T \widehat{\text{Var}}(\hat{\tau}_t^{k,k'})}, \hat{\tau}^{k,k'} + \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{1-\delta}} \sqrt{\frac{1}{T^2} \sum_{t=1}^T \widehat{\text{Var}}(\hat{\tau}_t^{k,k'})} \right] \right) \geq 1 - \alpha$$

for large  $n$ . Moreover, suppose Theorem 2 holds, then for any  $\delta > 0$ ,

$$\mathbb{P} \left( \hat{\tau}^{k,k'} \in \left[ \hat{\tau}^{k,k'} - \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{1-\delta}} \sqrt{\frac{1}{T^2} \sum_{t=1}^T \widehat{\text{Var}}(\hat{\tau}_t^{k,k'})}, \hat{\tau}^{k,k'} + \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{1-\delta}} \sqrt{\frac{1}{T^2} \sum_{t=1}^T \widehat{\text{Var}}(\hat{\tau}_t^{k,k'})} \right] \right) \geq 1 - \alpha$$

for large  $T$ .

The proof of the above proposition builds on the proof of Proposition 7 in Appendix A.

The variance bound and inference for the regime when  $T \rightarrow \infty$  is identical to what is given in Bojinov & Shephard (2019) and is therefore omitted.

### 3.2 Shrinkage estimator under stability assumption

We now focus on deriving a better estimator of the TEC, at a fixed point in time. Suppose a unit receives the same treatment for two consecutive periods. If the potential outcomes are similar across

time, then we can borrow information from past outcomes to reduce the variance of our estimate of the TEC. Intuitively, this section provides a bias-variance trade-off, where we introduce some bias in our inference for a potentially substantial reduction in the variance. For a specific treatment, the similar across time assumption can be formalized as follows.

**Assumption 5** (Weak stability of potential outcomes). *We say the potential outcome matrix  $Y_{i,t}, i = 1, \dots, N, t = 1, \dots, T$  is  $\epsilon$ -weakly stable if for each  $i$  and exposure value  $k$ , we have  $|Y_{i,t}(k) - Y_{i,t+1}(k)| \leq \epsilon, \forall t \in \{1, \dots, T-1\}$ . If we further assume that  $\epsilon = 0$ , we then say that the potential outcome matrix is strongly stable.*

All results in this section easily generalize to the case where the uniform bound  $\epsilon$  is replaced by a time dependent bound  $\epsilon_t$ . Throughout, we focus on the estimation of the total effect at time  $t$  as an example to illustrate how we can leverage temporal information under weak stability.

Under pure population interference and time-invariant exposure mappings,

$$\tau_t^{TE} = \frac{1}{n} \sum_{i=1}^n Y_{i,t}(h_i^1) - \frac{1}{n} \sum_{i=1}^n Y_{i,t}(h_i^0), \quad (7)$$

where  $h_i^1 = f_i(1_{t,1:n})$  and  $h_i^0 = f_i(0_{t,1:n})$ .

To build some intuition, we first describe how to leverage a single past time period,  $t' = t - 1$  to improve estimation at time  $t$ . The idea is that by considering a convex combination  $\hat{\tau}_t^c = \alpha \hat{\tau}_t^{TE} + (1 - \alpha) \hat{\tau}_{t-1}^{TE}$ , for some  $\alpha \in [0, 1]$  as an estimator of  $\tau_t^{TE}$ , we introduce some bias but reduce the variance — the hope being that under weak stability, the bias introduced will be modest compared to the reduction in variance. This is formalized in the following proposition.

**Proposition 3** (Bound on the bias of  $\hat{\tau}_t^c$ ).

$$|\mathbb{E}[\hat{\tau}_t^c] - \tau_t^{TE}| \leq 2(1 - \alpha)\epsilon \quad (8)$$

As we can see, the absolute bias of  $\hat{\tau}_t^c$  is bounded by a quantity that grows linearly with  $\epsilon$ : if  $\epsilon$  is very small, then so will the maximum bias. In particular,  $\hat{\tau}_t^c$  is unbiased for  $\tau_t^{TE}$  if  $\epsilon = 0$ , which corresponds to the somewhat unrealistic assumption that the potential outcomes do not vary across time. Under some conditions, it can be guaranteed that the gain in bias is more than counterbalanced by a reduction in variance, making it a worthwhile trade-off in terms of the mean squared error (MSE).

**Proposition 4.** *Suppose  $\text{Var}(\hat{\tau}_t^{TE}) > \text{Cov}(\hat{\tau}_t^{TE}, \hat{\tau}_{t-1}^{TE})$ , then there exists some  $\alpha \in (0, 1)$  such that  $\hat{\tau}_t^c = \alpha \hat{\tau}_t^{TE} + (1 - \alpha) \hat{\tau}_{t-1}^{TE}$  has lower MSE than  $\hat{\tau}_t^{TE}$ . Moreover, if we have  $\text{Var}(\hat{\tau}_t^{TE}) - \text{Var}(\hat{\tau}_{t-1}^{TE}) > 4\epsilon^2$  then we know that  $\hat{\tau}_t^c = \frac{1}{2} \hat{\tau}_t^{TE} + \frac{1}{2} \hat{\tau}_{t-1}^{TE}$  has lower MSE than  $\hat{\tau}_t^{TE}$ .*



---

**Algorithm 1** Algorithm to estimate  $\epsilon$ 

---

- 1: Initialize  $\hat{\epsilon} = 0$
  - 2: For  $t = 1$  to  $T - 1$ :
    - (a) For  $i = 1, 2, \dots, n$  compute  $h_{i,t}$  and  $h_{i,t+1}$ .
    - (b) If  $h_{i,t} = h_{i,t+1} = k$ , compute  $\epsilon_{i,t} = |y_{i,t} - y_{i,t+1}|$ .
    - (c) If  $\epsilon_{i,t} > \hat{\epsilon}$ , set  $\hat{\epsilon} = \epsilon_{i,t}$ .
  - 3: Output  $\hat{\epsilon}$ .
- 

In other words, if the current variance is larger, by choosing some  $\alpha$ , the convex combination type estimator would give us a better estimator in terms of MSE. Moreover, as the proposition suggests, if we know the difference is bigger than  $4\epsilon^2$ , we know that  $\alpha = \frac{1}{2}$  is sufficient.

If we further assume that the assignment mechanism is temporally independent (Definition 1), then  $\text{Cov}(\hat{\tau}_t^{TE}, \hat{\tau}_{t-1}^{TE}) = 0$ , hence we have the following result.

**Proposition 5.** *Suppose that the assignments mechanism is temporally independent, then there exists some  $\alpha \in (0, 1)$  such that  $\hat{\tau}_t^c = \alpha \hat{\tau}_t^{TE} + (1 - \alpha) \hat{\tau}_{t-1}^{TE}$  has lower MSE than  $\hat{\tau}_t^{TE}$ . The optimal  $\alpha$  is given by  $\alpha = 1 - \frac{\text{Var}(\hat{\tau}_t^{TE})}{4\epsilon^2 + \text{Var}(\hat{\tau}_t^{TE}) + \text{Var}(\hat{\tau}_{t-1}^{TE})}$ .*

Under the  $\epsilon$ -stability assumption, Algorithm 1 provides a data dependent approach to estimate  $\epsilon$  and allows us to obtain estimate  $\hat{\alpha}$  of the weight parameter  $\alpha$ ,

$$\hat{\alpha} = 1 - \frac{\widehat{\text{Var}}(\hat{\tau}_t^{TE})}{\widehat{\text{Var}}(\hat{\tau}_t^{TE}) + \widehat{\text{Var}}(\hat{\tau}_{t-1}^{TE}) + 4(t - t')^2 \hat{\epsilon}^2},$$

where  $\widehat{\text{Var}}(\hat{\tau}_t^{TE})$  can be any estimator of the variance  $\text{Var}(\hat{\tau}_t^{TE})$ : we discuss a few options in Proposition 9 of Appendix B. In addition, under pure population interference and temporally independent assignments,

$$\begin{aligned} \text{Var}(\hat{\tau}_t^c) &= \text{Var}\left(\alpha \hat{\tau}_t^{TE} + (1 - \alpha) \hat{\tau}_{t-1}^{TE}\right) \\ &= \alpha^2 \text{Var}(\hat{\tau}_t^{TE}) + (1 - \alpha)^2 \text{Var}(\hat{\tau}_{t-1}^{TE}), \end{aligned}$$

which suggests the following plug-in estimator of the variance:

$$\widehat{\text{Var}}(\hat{\tau}_t^c) = \hat{\alpha}^2 \widehat{\text{Var}}(\hat{\tau}_t^{TE}) + (1 - \hat{\alpha})^2 \widehat{\text{Var}}(\hat{\tau}_{t-1}^{TE}).$$

We also give the expression of  $\text{Cov}(\hat{\tau}_t^{TE}, \hat{\tau}_{t-1}^{TE})$  and an estimator for it in Proposition 8 and Proposition 10 of Appendix B respectively. Equipped with the variance and the covariance estimators, we can directly check the condition in Proposition 10. The optimal  $\alpha$  is given in the proof and can be estimated in the similar way as in the independent assignment case.

We can easily generalize the above discussion to a general version of this estimator such that we combine  $\hat{\tau}_t^{TE}$  and  $\hat{\tau}_{t'}^{TE}$  for arbitrary  $t' < t$ . The above results trivially generalize and are therefore omitted for brevity.

Based on the variance estimator above, we propose two ways to construct confidence intervals. The first one ignores the bias of  $\hat{\tau}_t^c$  and uses Gaussian confidence interval. The second one takes advantage of Chebyshev's inequality and incorporates the bias. Specifically, note that

$$\mathbb{P}(|\hat{\tau}_t^c - (\mathbb{E}[\hat{\tau}_t^c] - \tau_t^{TE}) - \tau_t^{TE}| \geq \epsilon) \leq \frac{\text{Var}(\hat{\tau}_t^c)}{\epsilon^2},$$

hence  $\forall \delta > 0$ ,

$$\mathbb{P}(\tau_t^{TE} \in [\hat{\tau}_t^c - (\mathbb{E}[\hat{\tau}_t^c] - \tau_t^{TE}) - \epsilon, \hat{\tau}_t^c - (\mathbb{E}[\hat{\tau}_t^c] - \tau_t^{TE}) + \epsilon]) \geq 1 - \delta$$

for  $\epsilon = \sqrt{\frac{\text{Var}(\hat{\tau}_t^c)}{\delta}}$ . Let  $b(\hat{\tau}_t^c) = \mathbb{E}[\hat{\tau}_t^c] - \tau_t^{TE} = (1 - \alpha)(\tau_{t-1}^{TE} - \tau_t^{TE})$  be the bias of our convex combination estimator. If we estimate  $b(\hat{\tau}_t^c)$  by  $\hat{b}(\hat{\tau}_t^c) = (1 - \hat{\alpha})(\hat{\tau}_{t-1}^{TE} - \hat{\tau}_t^{TE})$ , then we can use the following interval as an approximate  $(1 - \delta)$ -level confidence interval of  $\tau_t^{TE}$ :

$$\left[ \hat{\tau}_t^c - \hat{b}(\hat{\tau}_t^c) - \sqrt{\frac{\widehat{\text{Var}}(\hat{\tau}_t^c)}{\delta}}, \hat{\tau}_t^c - \hat{b}(\hat{\tau}_t^c) + \sqrt{\frac{\widehat{\text{Var}}(\hat{\tau}_t^c)}{\delta}} \right].$$

We explore empirically the coverage of the above approximate confidence intervals with a simulation study in Section 5.

The approach we have described in this section naturally extends to using the  $k - 1$  previous time steps, yielding the weighted combination estimator:

$$\hat{\tau}_t^c = \alpha_1 \hat{\tau}_{t-k+1}^{TE} + \dots + \alpha_k \hat{\tau}_t^{TE},$$

where  $\alpha_1, \dots, \alpha_k$  can be estimated by solving a slightly more involved convex optimization problem. We describe this approach in full details in Appendix C.

## 4 Panel experiments with population interference and carryover effects

Section 3 shows that adding a temporal dimension is particularly useful if no carryover effects exist. We now consider the setting where there are both carryover effects and population interference, which we call mixed interference. As we show below, mixed interference affects our ability to draw inference both for the TEC and ATEC, albeit in different ways. For temporal exposure contrasts (TEC), the same theorem as in Section A holds.

**Theorem 4.** *Assume we have a temporally independent assignment mechanism, under Assumption 1-4 and the condition that  $d_n = o(n^{1/4})$ , we have*

$$\frac{\sqrt{n}(\hat{\tau}_t^{k,k'} - \tau_t^{k,k'})}{\text{Var}(\sqrt{n}\hat{\tau}_t^{k,k'})^{1/2}} \xrightarrow{d} \mathcal{N}(0, 1)$$

as  $n \rightarrow \infty$ .

The difference with the pure population setting, discussed in Appendix A, is not mathematical but conceptual: in the mixed setting, the exposures involve the assignments over previous time steps. Consequently, there are generally many more exposures than in the pure population setting, and each unit has a lower probability of receiving each. This leads to Horvitz-Thompson estimators with a much larger variance.

For the average temporal exposure contrast, the difference between population interference and mixed interference is starker. The main difficulty is that mixed interference breaks the temporal independence that powered the results of Section 3.1.

Nevertheless, we can still establish a general central limit theorem, be it under some additional assumptions. In particular, we require a restriction on the rate at which the variance shrinks.

**Assumption 6.** *Assume that*

$$\liminf_{n \rightarrow \infty} \text{Var}(\sqrt{nT}\hat{\tau}^{k,k'}) \geq \epsilon > 0$$

for some  $\epsilon$ .

This technical assumption rules out the pathological case that the variance vanishes as  $n \rightarrow \infty$ . To state our theorem, we also need to introduce the notion of  $s$ -dependent sequence.

**Definition 8** ( $s$ -dependent sequences). *We say that  $\{H_{i,t}\}_{i=1}^n$  is an  $s$ -dependent sequence of random variables if and only if for any index set  $I, J \subseteq 1, \dots, n$ ,  $\{H_{i,t}\}_{i=1}^n$  and  $\{H_{i,t}\}_{i=1}^n$  are independent so long as  $\min_{j \in J} j - \max_{i \in I} i > s$ .*

Intuitively, this assumption limits the depends across units.

**Theorem 5.** *Assume we have a temporally independent assignment mechanism, under Assumption 1-4 and Assumption 6, suppose  $\{H_{i,t}\}_{i=1}^n$  is an  $s$ -dependent sequence of random variables for a fixed  $t$  and LEA( $p$ ) assumption is satisfied with some finite  $p$ . If  $s, n, T$  are such that  $s^5 T^4 = o(n^{1-\alpha})$  for some  $0 < \alpha < 1$ , then we have that*

$$\frac{\sqrt{nT}(\hat{\tau}^{k,k'} - \bar{\tau}^{k,k'})}{\sqrt{\text{Var}(\sqrt{nT}\hat{\tau}^{k,k'})}} \xrightarrow{d} \mathcal{N}(0, 1)$$

as  $n \rightarrow \infty$ .

The above theorem requires general assumptions on exposure values as well as the asymptotic variance though independent assignments are not required. However, it is somewhat difficult to apply this result to practical settings. Therefore, we now focus on a specific setting to illustrate the type of results that can be derived under mixed interference.

#### 4.1 Stratified interference

Consider the following natural temporal extension of the stratified interference (Hudgens & Halloran (2008); Basse & Feller (2018)) setting (Hudgens & Halloran (2008); Basse & Feller (2018)):

$$f_{i,t}(w_{1:n,1:t}) = f(w_{i,t-1}, w_{i,t}, \{w_{j,t-1}\}_{j \in \mathcal{N}_i, j \neq i}, \{w_{j,t}\}_{j \in \mathcal{N}_i, j \neq i})$$

where  $\mathcal{N}_i$  is the group to which unit  $i$  belongs. For convenience, we fix each group to be of size  $r$ <sup>6</sup>.

**Theorem 6.** *With the above setting and temporally independent assignments, under Assumption 1-4 and Assumption 6, suppose  $n, r, T$  are such that  $r = o((nT)^{\frac{1}{4}})$ , then we have that*

$$\frac{\sqrt{nrT}(\hat{\tau}^{k,k'} - \bar{\tau}^{k,k'})}{\sqrt{\text{Var}(\sqrt{nrT}\hat{\tau}^{k,k'})}} \xrightarrow{d} \mathcal{N}(0, 1)$$

as  $n \rightarrow \infty$ .

The theorem holds for heterogeneous group sizes as long as  $\max_i r_i = o((nT)^{\frac{1}{4}})$  where  $r_i = |\mathcal{N}_i|$  is the size of the group unit  $i$  belongs to. To do inference, we consider a specific example of stratified interference:

$$f_{i,t}(w_{1:n,1:t}) = \left( w_{i,t-1}, w_{i,t}, \sum_{j \in \mathcal{N}_i, j \neq i} w_{j,t-1}, \sum_{j \in \mathcal{N}_i, j \neq i} w_{j,t} \right).$$

We focus on the Bernoulli design where each unit is independently assigned to treatment with probability  $\frac{1}{2}$ . We consider the exposures  $k = (1, 1, r-1, r-1)$  and  $k' = (0, 0, 0, 0)$ . Such exposure contrast is exactly the same as the total effect since essentially we are comparing the world of everyone getting treatment to the world of everyone getting control. Notice that in this case,  $r$  cannot be infinite, otherwise the overlap assumption would be violated. To ease notations, we index each unit  $i$  by a tuple  $(l, q)$ , meaning that unit  $i$  is the  $q$ -th unit in the  $l$ -th group<sup>7</sup>.

**Proposition 6.** *Assuming the above setup, let  $X_{n,t} = \sqrt{n}(\hat{\tau}_t^{k,k'} - \tau_t^{k,k'})$ , then we can estimate the asymptotic variance by*

$$\widehat{B}_n^2 = \sum_{t=1}^T \widehat{\text{Var}}(X_{n,t}) + 2 \sum_{t=1}^{T-1} \widehat{\text{Cov}}(X_{n,t}, X_{n,t+1}),$$

<sup>6</sup>This ensures that each unit is associated with exactly the same set of exposure values so that the exposure contrast between two arbitrary exposure values is well-defined.

<sup>7</sup>If we use a tuple  $(l, q)$  to represent the  $q$ -th unit in the  $l$ -th household, then we note by passing that  $0 < C_1 \leq Y_{(l,q),t}(k) \leq C_2$  for all  $l, q, t, k$  for some  $C_1, C_2$  is sufficient for Assumption 6.

where

$$\begin{aligned}
\widehat{\text{Var}}(X_{n,t}) = & \frac{1}{nrT} \left[ \sum_{l=1}^n \sum_{q=1}^r (2^{2r} - 1) \frac{\mathbf{1}(H_{(l,q),t} = k) Y_{(l,q),t}^2}{\mathbb{P}(H_{(l,q),t} = k)} + \sum_{l=1}^n \sum_{q=1}^r (2^{2r} - 1) \frac{\mathbf{1}(H_{(l,q),t} = k') Y_{(l,q),t}^2}{\mathbb{P}(H_{(l,q),t} = k')} \right. \\
& + \sum_{l=1}^n \sum_{q=1}^r \left( \frac{\mathbf{1}(H_{(l,q),t} = k) Y_{(l,q),t}^2}{\mathbb{P}(H_{(l,q),t} = k)} + \frac{\mathbf{1}(H_{(l,q),t} = k') Y_{(l,q),t}^2}{\mathbb{P}(H_{(l,q),t} = k')} \right) \\
& + \sum_{l=1}^n \sum_{q_1=1}^r \sum_{q_2 \neq q_1}^r \left( (2^{2r} - 1) \frac{\mathbf{1}(H_{(l,q_1),t} = k, H_{(l,q_2),t} = k) Y_{(l,q_1),t} Y_{(l,q_2),t}}{\mathbb{P}(H_{(l,q_1),t} = k, H_{(l,q_2),t} = k)} + \right. \\
& \left. + (2^{2r} - 1) \frac{\mathbf{1}(H_{(l,q_1),t} = k', H_{(l,q_2),t} = k') Y_{(l,q_1),t} Y_{(l,q_2),t}}{\mathbb{P}(H_{(l,q_1),t} = k', H_{(l,q_2),t} = k')} \right) \\
& \left. + \sum_{l=1}^n \sum_{q_1=1}^r \sum_{q_2 \neq q_1}^r \frac{\mathbf{1}(H_{(l,q_1),t} = k) Y_{(l,q_1),t}^2}{\mathbb{P}(H_{(l,q_1),t} = k)} + \sum_{l=1}^n \sum_{q_1=1}^r \sum_{q_2 \neq q_1}^r \frac{\mathbf{1}(H_{(l,q_2),t} = k') Y_{(l,q_2),t}^2}{\mathbb{P}(H_{(l,q_2),t} = k')} \right] \quad (9)
\end{aligned}$$

and

$$\begin{aligned}
\widehat{\text{Cov}}(X_{n,t}, X_{n,t+1}) = & \frac{1}{nrT} \sum_{l=1}^n \sum_{q_1=1}^r \sum_{q_2=1}^r \left( (2^r - 1) \frac{\mathbf{1}(H_{(l,q_1),t} = k, H_{(l,q_2),t+1} = k) Y_{(l,q_1),t} Y_{(l,q_2),t+1}}{\mathbb{P}(H_{(l,q_1),t} = k, H_{(l,q_2),t+1} = k)} \right. \\
& + (2^r - 1) \frac{\mathbf{1}(H_{(l,q_1),t} = k', H_{(l,q_2),t+1} = k') Y_{(l,q_1),t} Y_{(l,q_2),t+1}}{\mathbb{P}(H_{(l,q_1),t} = k', H_{(l,q_2),t+1} = k')} \\
& + \frac{\mathbf{1}(H_{(l,q_1),t} = k') Y_{(l,q_1),t}^2}{\mathbb{P}(H_{(l,q_1),t} = k')} + \frac{\mathbf{1}(H_{(l,q_2),t+1} = k) Y_{(l,q_2),t+1}^2}{\mathbb{P}(H_{(l,q_2),t+1} = k)} \\
& \left. + \frac{\mathbf{1}(H_{(l,q_1),t} = k) Y_{(l,q_1),t}^2}{\mathbb{P}(H_{(l,q_1),t} = k)} + \frac{\mathbf{1}(H_{(l,q_2),t+1} = k') Y_{(l,q_2),t+1}^2}{\mathbb{P}(H_{(l,q_2),t+1} = k')} \right) \quad (10)
\end{aligned}$$

The expression of the variance is immediate from the setup, (B.22) and (B.23). The estimator is obtained by replacing the non-identifiable terms with an upper bound and estimating the upper bound accordingly.

The difficulty in doing inference under a more general setting comes from the fact that it is hard to give the explicit expression of the variance. Since there is dependence across time, the variance of  $\hat{\tau}^{k,k'}$  also involves covariance between Horvitz-Thompson estimators across different times. Hence, in this case, we need at least more assumptions on the assignment mechanism in order to express the variance term explicitly.

## 5 Simulations

We now use a simulation to explore some of our theoretical results. Section 5.1 explores some of the finite sample properties of our central limit theorems in different realistic settings. Section 5.2

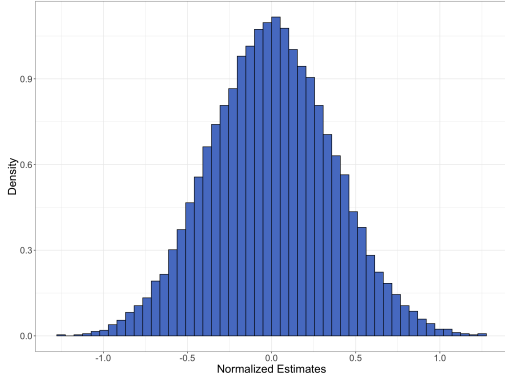


Figure 1: Histogram,  $n = 1000$

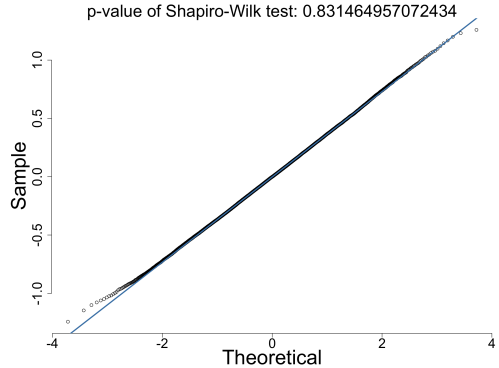


Figure 2: Q-Q normal plot,  $n = 1000$

explores empirically some properties of the convex combination estimator proposed in Section 3.2: in particular, we show that confidence intervals based on normal approximations behave well in our simulation is a reasonable candidate for practical use.

## 5.1 Simulations for central limit theorems

We first explore the finite sample behavior of our central limit theorems. To make our simulations relevant, we consider a version of the popular stratified interference setting (Duflo & Saez (2003); Basse & Feller (2018)), in which individuals are nested in groups of varying sizes, and interference may occur within but not across groups. Specifically, we consider the exposure mapping  $f_{i,t}(w_{1:n,t}) = (w_{i,t}, u_{i,t})$ , where at time  $t$   $u_{i,t} = 1$  if unit  $i$  has at least one treated neighbor and  $u_{i,t} = 0$  otherwise, so each unit may receive one of four exposures:  $(0,0)$ ,  $(0,1)$ ,  $(1,0)$  and  $(1,1)$ . Throughout, we consider a two-stage design whereby each group is assigned independently with probability  $\frac{1}{2}$  to a high-exposure or low-exposure arm, and then each unit is assigned to treatment independently with probability 0.9 in high-exposure groups, and 0.1 in low-exposure groups.

We focus on the central limit theorems for ATEC. Theorem 3 establishes asymptotic results for ATEC under less constraining assumptions on the interference mechanism than for TEC (Theorem 7 in Appendix A). To illustrate this point, we consider the stratified interference setting. We assume that the size of each group is bounded by  $n^{1/3}$ . In this case,  $d_n = n^{1/3}$  and hence  $T = \sqrt{n}$  suffices for Theorem 3. Compared to  $d_n = o(n^{1/4})$  in the cross-sectional setting, we are able to have larger group size. We consider the exposure mapping  $f_{i,t}(w_{1:n}) = (w_{i,t}, u_{i,t})$  where  $u_{i,t} = 0$  if less than 25% of the neighbors are treated;  $u_{i,t} = 1$  if between 25% and 50% of the neighbors are treated;  $u_{i,t} = 2$  if between 50% and 75% of the neighbors are treated and  $u_{i,t} = 3$  if more than 75% of the neighbors are treated. We generate the potential outcomes for unit  $i$  at time step  $t$  according to  $\mathcal{N}(3w_{i,t} + 2u_{i,t} + 5 + \epsilon_{i,t}, 1)$ , where  $\epsilon_{i,t}$  is uniform $\{-1, 1\}$ . Figure 1 and 2 show that  $n = 1000$  suffices for a good approximation. Moreover, the coverage of our 95% confidence interval is 95.4%.

Sample Size	n = 50	n = 100	n = 250	n = 500	n = 750	n = 1000
RMSE for $\hat{\tau}_{20}^{TE}$	64.68	28.98	5.80	1.84	0.95	0.70
RMSE for $\hat{\tau}_{20}^c, k = 2$	14.17	9.18	3.72	1.42	0.68	0.52
RMSE for $\hat{\tau}_{20}^c, k = 5$	4.39	4.58	3.01	1.17	0.58	0.45

Table 1: Root mean squared errors (RMSE) for  $\hat{\tau}_{20}^{TE}$ ,  $\hat{\tau}_{20}^c$  with  $k = 2$  and  $\hat{\tau}_{20}^c$  with  $k = 5$

## 5.2 Estimation under the stability assumption

In Section 3.2, we showed that with an appropriate choice of weights, the family of convex combination estimators outperforms the Horvitz-Thompson estimator. We illustrate this with a simulation study and show that our proposed confidence intervals perform well in our simulated setting.

### 5.2.1 Estimation under stability assumption for total effects

We consider a social network generated according to an Erdős-Rényi model, in which the units are assigned to treatment or control following a Bernoulli(1/2) design at each time step. We assume a local, pure population form of interference, summarized by the following exposure mappings:

$$f_i(w_{1:n,t}) = \left( w_{i,t}, \frac{1}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} w_{j,t} \right) \quad (11)$$

where  $\mathcal{N}_i$  is the neighborhood of the  $i$ -th unit; that is, we assume that only direct neighbors affect one’s potential outcomes. For each unit  $i$ , we generate the potential outcomes at  $t = 1$  randomly from  $\mathcal{N}(10, 1)$ . Then, for each time  $t > 1$ , we generate the potential outcome  $Y_{i,t}(k)$  uniformly from the interval  $(Y_{i,t-1}(k) - \epsilon, Y_{i,t-1}(k) + \epsilon)$ , so  $\epsilon$ -stability holds. Throughout our simulations, we assume that  $T = 20$  and we are interested in the total effect at time step  $t = 20$ . We compare the performance of the standard Horvitz-Thompson estimator and the performance of the convex combination estimator for estimating the total effect  $\tau_T^{TE}$  at time  $t = T = 20$ , varying both the population size  $n$  and the number of time steps  $k$  used in the convex combination. We estimate  $\epsilon$  using Algorithm 1 described in Section 3.2; we use Proposition 5 to estimate  $\alpha$  when  $k = 2$ , and solve the optimization problem introduced in Appendix C for  $k \geq 3$ .

We first fix  $\epsilon$  to be 3 and vary the sample size. To make each unit have the same expected number of neighbors, we scale the probability  $p$  in Erdős-Rényi model accordingly. For each  $n$ , we fix the graph and generate 100 realizations of assignments. Table 1 shows the root mean squared errors for three kinds of estimators for the total effect: the usual Horvitz-Thompson estimator, the convex combination type estimator with  $k = 2$ , and the convex combination estimator with  $k = 5$ . We see that the convex combination type estimators effectively reduce the mean squared error. Moreover, when  $n$  is relatively small, the reduction in mean squared error is significant.

Confidence Interval	Network 1	Network 2	Network 3
Gaussian CI with variance estimated by $\widehat{\text{Var}}^d$	92.9%	98.4%	95.9%
Gaussian CI with variance estimated by $\widehat{\text{Var}}^u$	97.2%	99.8%	100%
Chebyshev CI with variance estimated by $\widehat{\text{Var}}^d$	91.4%	94.1%	96.4%
Chebyshev CI with variance estimated by $\widehat{\text{Var}}^u$	94.6%	95.6%	97.7%

Table 2: Coverage of two approximate confidence intervals for  $\tau_t^{TE}$  with  $k = 2$

### 5.2.2 Coverage of two approximate confidence intervals

Recall that in Section 3.2, we gave two approximate confidence intervals of  $\tau_t^{TE}$  based on our convex combination estimator  $\hat{\tau}_t^c$  and variance estimator. We now provide coverage results of these two approximate confidence intervals. We assume a social network generated from the Erdős-Rényi Model with  $n = 100$  and  $p = 0.05$ . We fix the stability parameter  $\epsilon$  to be 3 and generate the data in the same way as in the previous section. To calculate the coverage, we generate 1000 realizations of the assignments and construct approximate confidence intervals accordingly.

Table 2 shows the two approximate confidence intervals provide reasonable coverage across the three different social networks. Although the Gaussian confidence interval ignores the bias of  $\hat{\tau}_t^c$ , it tends to provide better coverage than the confidence intervals obtained from the Chebyshev approach. Moreover, the Gaussian intervals tend to be shorter, making them practically more useful. Appendix ?? provides an additional table showing the average lengths of the confidence intervals in Table 2.

## 6 Two real data examples

We now apply our methods to two empirical applications. In the first application, we use the convex combination estimator to analyze a panel experiment and show it reduces the variance and leads to more reliable estimates of the temporal exposure contrast. In the second application, we run a semi-synthetic experiment on a social network to demonstrate the necessity of our assumptions for the validity of the analysis and provide further empirical evidence of the advantage of the convex combination estimator.

### 6.1 Rational cooperation

The panel experiment we analyze is from Andreoni & Samuelson (2006). The authors test a game-theoretic model of “rational cooperation” through a panel experiment. Specifically, in each experiment session, they recruited 22 subjects to play 20 twice-played prisoners’ dilemmas, ensuring that no player would meet the same partner twice. The twice-played prisoners’ dilemma consists



of two periods with different pay-off structures, as shown in Table 3. The parameters  $x_1, x_2$  satisfy  $x_1, x_2 \geq 0, x_1 + x_2 = 10$ .

		<i>C</i>	<i>D</i>			<i>C</i>	<i>D</i>
<i>C</i>	$(3x_1, 3x_1)$	$(0, 4x_1)$		<i>C</i>	$(3x_2, 3x_2)$	$(0, 4x_2)$	
<i>D</i>	$(4x_1, 0)$	$(x_1, x_1)$		<i>D</i>	$(4x_2, 0)$	$(x_2, x_2)$	
Period one				Period two			

Table 3: Payoff structure in the experiment conducted by Andreoni & Samuelson (2006). The choice  $C$  denotes “cooperate” and the choice  $D$  “defect.”

Let  $\lambda = \frac{x_1}{x_1+x_2}$ , then for each round of the experiment, 22 subjects were grouped into 11 pairs, and each pair was randomly assigned with a  $\lambda \sim \text{Unif}\{0, 0.1, \dots, 0.9, 1\}$ . The outcomes were the total payoffs. Since there are five sessions in total, we have 110 subjects and 2200 outcomes. We use this panel experiment to illustrate that the convex combination estimator effectively reduces the estimates’ variance and thus produces more reliable estimates. To this end, following Bojinov et al. (2021), we define treatment to be  $\lambda > 0.6$  and control to be  $\lambda \leq 0.6$ . This results in a panel experiment with binary treatments and Bernoulli design with treated probability  $\frac{5}{11}$ . Under this setup, we generally expect a positive treatment affect as the payoffs are more concentrated in period two.

We next build a social network among all subjects in the experiment. If the players have played each other in the first few rounds, then they should have some influence on each other for the later rounds. Hence, we consider any players that played each other in the first five rounds of the game as being connected. We then use the remaining 15 rounds as our experimental data. So, for our panel experiment, we have  $n = 110$  and  $T = 15$ . As Bojinov et al. (2021) showed little evidence of carryover effects, we assume there is only population interference. Then, we use the exposure model in (11) and the temporal exposure contrast we are interested in is the exposure contrast between  $(0, \leq 0.2)$  and  $(1, \geq 0.8)$  for each time step. We now report the Horvitz-Thompson estimates of the temporal exposure contrast for the last 10 time steps, the estimates from the 2-step, and the 5-steps convex combination estimator estimates. Table 4 shows the results.

In general, we do not expect the temporal exposure contrasts to be different for different time steps since all the 15 rounds of games were done together in one session. And as we can see from the table, the convex combination estimator leads to estimates with much smaller variance. Note that the estimates from 2-step and 5-steps convex combination estimators are similar, illustrating that the choice of  $k$  is not crucial since the estimator itself takes care of it. Moreover, as we pointed out earlier, we would expect positive exposure contrast and the estimates from convex combination

Time Step	$T = 6$	$T = 7$	$T = 8$	$T = 9$	$T = 10$	$T = 11$	$T = 12$	$T = 13$	$T = 14$	$T = 15$	Variance
Horvitz-Thompson	-8.08	11.03	29.39	-3.53	57.48	-3.76	10.29	23.40	-13.70	16.19	452.80
2-step	-6.84	9.48	25.40	-2.99	33.81	-2.98	9.83	22.10	-11.20	14.64	226.13
5-steps	-6.83	9.46	25.38	-2.98	33.74	-2.97	9.82	22.09	-11.17	14.63	225.37

Table 4: Estimates for temporal exposure contrasts from the panel experiment in Andreoni & Samuelson (2006)

estimator are more reliable in the sense that it shrinks the estimates towards zero when the Horvitz-Thompson estimator gives a negative value (this is possible since we only have  $n = 110$  subjects which is a small number).

## 6.2 Facebook network semi-synthetic experiment

We now describe a semi-synthetic experiment using the Swarthmore College social network from the Facebook 100 dataset (Traud et al. 2012). All networks in this dataset are complete online friendship networks for one hundred colleges and universities collected from a single-day snapshot of Facebook in September 2005. The network we use is of size 1657 with 61049 edges. We use this network as the graph that describes population interference among units and generate an assignment vector using a Bernoulli design with a success probability of  $1/2$ . We first show mean squared error reduction of using convex combination estimator to estimate temporal exposure contrast between  $(0, 0)$  and  $(1, 1)$  at  $T = 20$ . Let  $\rho_{i,t} = \frac{1}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} w_{j,t}$ , we assume the following exposure mappings:

$$f_i(w_{1:n,t}) = (w_{i,t}, \tilde{\rho}_{i,t}), \quad \text{where } \tilde{\rho}_{i,t} = \begin{cases} 0 & \text{if } \rho_{i,t} \leq 0.3, \\ \rho_{i,t} & \text{if } 0.3 \leq \rho_{i,t} \leq 0.7, \\ 1 & \text{if } \rho_{i,t} > 0.7. \end{cases}$$

Now we make a panel experiment with  $T = 20$ . We generate the outcomes at each time step according to a linear model that is linear in  $w_{i,t}$  and  $\tilde{\rho}_{i,t}$  and add a time-varying component  $\epsilon_t$  that is uniformly distributed on  $[-0.5, 0.5]$ . Table 5 shows the empirical bias, the variance, and the root mean squared errors (RMSE) of the estimates of the temporal exposure contrast at time step  $T = 20$  by using Horvitz-Thompson estimator, 2-step convex combination estimator, and 5-step convex combination estimator. As expected, the convex combination estimator reduces the RMSE significantly. Though the biases seem large compared to the Horvitz-Thompson estimator, as we mentioned previously, we can also control the amount of bias we tolerate, which implicitly accounts for the time effect.

The maximal degree for the network is 577, which is far greater than the  $\sqrt{n}$ . To make Theorem 3 hold approximately for this network, we require having an extremely large  $T$ . Below we illustrate

Estimator of TEC	Horvitz-Thompson	2-step CVX	5-step CVX
RMSE	50.48	3.57	3.31
Empirical bias	0.26	2.75	2.80
Empirical variance	2548.495	5.19	3.09

Table 5: RMSE for different estimators of temporal exposure contrast at  $T = 20$

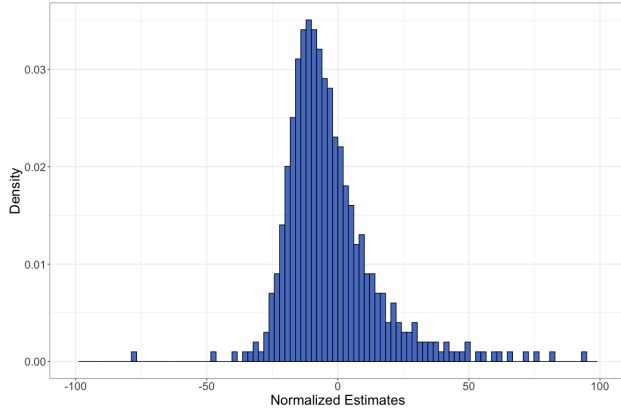


Figure 3: Histogram,  $T = 20$

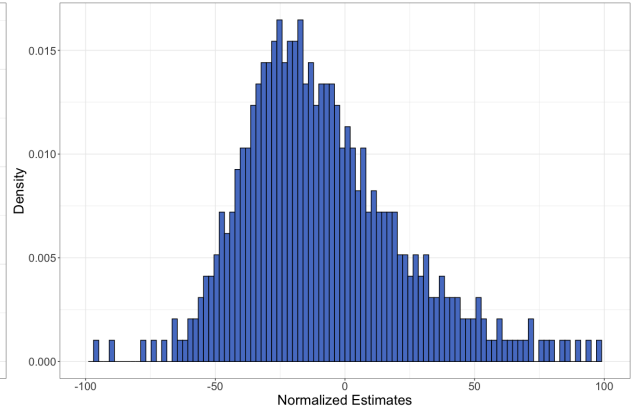


Figure 4: Histogram,  $T = 100$

this empirically through a semi-synthetic experiment. Let

$$f_i(w_{1:n,t}) = (w_{i,t}, \tilde{\rho}_{i,t}), \quad \text{where } \tilde{\rho}_{i,t} = \begin{cases} 0 & \text{if } \rho_{i,t} \leq 0.35, \\ 1 & \text{if } 0.35 < \rho_{i,t} \leq 0.5, \\ 2 & \text{if } 0.5 < \rho_{i,t} \leq 0.65, \\ 3 & \text{if } \rho_{i,t} > 0.65. \end{cases}$$

We are interested in the average temporal exposure contrast between  $(1, 3)$  and  $(0, 0)$ . Since the network is dense, with an average degree of 73.69, we expect the Horvitz-Thompson estimator to be inaccurate since units with exposure values  $(1, 3)$  or  $(0, 0)$  will unlikely to be those units with many neighbors. Figure 3 and 4 show the histograms of Horvitz-Thompson estimates for  $T = 20$  and  $T = 100$  respectively. Here, we calculate ATEC for 10,000 realizations, and since the computation of the variance estimate is time-consuming, we do not rescale the estimates.

Figure 3 shows that when  $T = 20$  the histogram is far from normally distributed. Figure 4 shows that when  $T = 100$ , although the data are much closer to being normally distributed, they still are not. Also, note that the centers of these two histograms are away from 0; as we stated above, since some of the neighborhoods are extremely large, we cannot observe the exposure value we would need for units with large neighborhoods. This illustrates the necessity of condition (6)

— reliable inference requires more experiments if we have a dense network. We also report the coverage using a Gaussian confidence interval here. For both  $T = 20$  and  $T = 100$ , the empirical coverage of naive Gaussian confidence interval is around 80%.

## 7 Conclusion

In this paper, we developed estimation and inference results for panel experiments with population interference. Our work shows that the added temporal dimension in experiments with population interference may either help or hurt our ability to do inference. In the absence of carryover effects, causal effects averaged across time can be estimated more easily than at single time points. Here we also introduced a variance reduction technique that incorporated a novel stability assumption and showed that if the potential outcomes are well-behaved, a simple convex combination type estimator can effectively reduce the mean squared error. In the presence of carryover effects, we can still obtain novel central limit theorems; however, these require additional assumptions and are not as flexible. Finally, our results also explicitly capture the relationship between the growth rate of the dependency graph of the exposure values, allowing us to consider the trade-off between the assignment mechanism and the interference.

Many interesting avenues of investigation around interference in panel experiments have been left unexplored in this manuscript and will be the object of future work. First, our results mostly consider the temporally independent assignment mechanism: this is, of course, limiting as it removes all dynamic designs, but it does present a useful benchmark. Second, while our simulations show that our convex combination estimators seem to behave well, our formal results under this new stability assumption are still limited. Third, we do not explicit discussions hypothesis testing. Although our results do provide a way to test specific hypotheses by inverting the confidence intervals, there has been literature that discusses how to conduct a Fisher randomization test for the sharp null hypothesis in panel experiments (Bojinov et al. 2021).

## References

- Abadie, A., Athey, S., Imbens, G. W. & Wooldridge, J. M. (2020), ‘Sampling-based versus design-based uncertainty in regression analysis’, *Econometrica* **88**(1), 265–296.
- Andreoni, J. & Samuelson, L. (2006), ‘Building rational cooperation’, *Journal of Economic Theory* **127**(1), 117–154.
- Angrist, J. D. & Kuersteiner, G. M. (2011), ‘Causal effects of monetary shocks: Semiparametric conditional independence tests with a multinomial propensity score’, *Review of Economics and Statistics* **93**(3), 725–747.

- Aronow, P. M. & Samii, C. (2017), ‘Estimating average causal effects under general interference, with application to a social network experiment’, *Annals of Applied Statistics* **11**(4), 1912–1947.
- Athey, S., Eckles, D. & Imbens, G. W. (2018), ‘Exact p-values for network interference’, *Journal of the American Statistical Association* **113**(521), 230–240.
- Basse, G. & Bojinov, I. (2020), ‘A general theory of identification’, *arXiv preprint arXiv:2002.06041* .
- Basse, G., Ding, Y. & Toulis, P. (2019), ‘Minimax designs for causal effects in temporal experiments with treatment habituation’, *arXiv e-prints* p. arXiv:1908.03531.
- Basse, G. & Feller, A. (2018), ‘Analyzing two-stage experiments in the presence of interference’, *Journal of the American Statistical Association* **113**(521), 41–55.
- Basse, G., Feller, A. & Toulis, P. (2019), ‘Randomization tests of causal effects under interference’, *Biometrika* **106**(2), 487–494.
- Basse, G. W. & Airoidi, E. M. (2018), ‘Limitations of design-based causal inference and a/b testing under arbitrary and network interference’, *Sociological Methodology* **48**(1), 136–151.
- Blackwell, M. & Glynn, A. N. (2018), ‘How to make causal inferences with time-series cross-sectional data under selection on observables’, *American Political Science Review* **112**(4), 1067–1082.
- Bojinov, I. & Gupta, S. (2022), ‘Online experimentation: Benefits, operational and methodological challenges, and scaling guide’, *Harvard Data Science Review* **4**(3).
- Bojinov, I., Rambachan, A. & Shephard, N. (2021), ‘Panel experiments and dynamic causal effects: A finite population perspective’, *Quantitative Economics* **12**, 1171–1196.
- Bojinov, I. & Shephard, N. (2019), ‘Time series experiments and causal estimands: Exact randomization tests and trading’, *Journal of the American Statistical Association* **114**(528), 1665–1682.
- Bojinov, I., Simchi-Levi, D. & Zhao, J. (2022), ‘Design and analysis of switchback experiments’, *Management Science* .
- Boruvka, A., Almirall, D., Witkiewitz, K. & Murphy, S. A. (2018), ‘Assessing time-varying causal effect moderation in mobile health’, *Journal of the American Statistical Association* **113**(523), 1112–1121.
- Bramoullé, Y., Djebbari, H. & Fortin, B. (2020), ‘Peer effects in networks: A survey’, *Annual Review of Economics* **12**, 603–629.

- Bramoullé, Y., Djebbari, H. & Fortin, B. (2009), ‘Identification of peer effects through social networks’, *Journal of Econometrics* **150**(1), 41–55.
- Chamandy, N. (2016), ‘Experimentation in a ridesharing marketplace lyft engineering’, *URL: <https://eng.lyft.com/experimentation-in-a-ridesharing-marketplace-b39db027a66e>*.
- Chin, A. (2018), ‘Central limit theorems via Stein’s method for randomized experiments under interference’, *arXiv e-prints* p. arXiv:1804.03105.
- Choudhury, P., Lane, J. N. & Bojinov, I. (2021), ‘Virtual water coolers: A field experiment on the role of virtual interactions on organizational newcomer performance’.
- Cox, D. R. (1958), *Planning of experiments*, New York, Wiley.
- Ding, P. (2017), ‘A Paradox from Randomization-Based Causal Inference’, *Statistical Science* **32**(3), 331 – 345.
- Duflo, E. & Saez, E. (2003), ‘The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment’, *The Quarterly Journal of Economics* **118**(3), 815–842.
- Farronato, C., MacCormack, A. & Mehta, S. (2018), ‘Innovation at uber: The launch of express pool’, *Harvard Business School Case* **620**(062).
- Guo, K. & Basse, G. (2021), ‘The generalized oaxaca-blinder estimator’, *Journal of the American Statistical Association* pp. 1–13.
- Halloran, M. E. & Struchiner, C. J. (1995), ‘Causal inference in infectious diseases’, *Epidemiology* **6**(2), 142–151.
- Hohnhold, H., O’Brien, D. & Tang, D. (2015), Focusing on the long-term: It’s good for users and business, *in* ‘Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining’, pp. 1849–1858.
- Hong, G. & Raudenbush, S. W. (2006), ‘Evaluating kindergarten retention policy’, *Journal of the American Statistical Association* **101**(475), 901–910.
- Horvitz, D. G. & Thompson, D. J. (1952), ‘A generalization of sampling without replacement from a finite universe’, *Journal of the American Statistical Association* **47**(260), 663–685.
- Hu, Y. & Wager, S. (2022), ‘Switchback experiments under geometric mixing’, *arXiv preprint arXiv:2209.00197*.

- Hudgens, M. G. & Halloran, M. E. (2008), ‘Toward causal inference with interference’, *Journal of the American Statistical Association* **103**(482), 832–842. PMID: 19081744.
- Imbens, G. W. & Rubin, D. B. (2015), *Causal inference in statistics, social, and biomedical sciences*, Cambridge University Press.
- Judson, R. A. & Owen, A. L. (1999), ‘Estimating dynamic panel data models: a guide for macroeconomists’, *Economics Letters* **65**(1), 9–15.
- Kempthorne, O. (1955), ‘The randomization theory of experimental inference’, *Journal of the American Statistical Association* **50**(271), 946–967.
- Kohavi, R., Tang, D. & Xu, Y. (2020), *Trustworthy online controlled experiments: A practical guide to a/b testing*, Cambridge University Press.
- Leung, M. P. (2020), ‘Treatment and spillover effects under network interference’, *Review of Economics and Statistics* **102**(2), 368–380.
- Leung, M. P. (2022), ‘Causal inference under approximate neighborhood interference’, *Econometrica* **90**(1), 267–293.
- Li, X. & Ding, P. (2017), ‘General forms of finite population central limit theorems with applications to causal inference’, *Journal of the American Statistical Association* **112**(520), 1759–1769.
- Li, X., Ding, P., Lin, Q., Yang, D. & Liu, J. S. (2019), ‘Randomization inference for peer effects’, *Journal of the American Statistical Association* **114**(528), 1651–1664.
- Lin, W. (2013), ‘Agnostic notes on regression adjustments to experimental data: Reexamining freedman’s critique’, *Annals of Applied Statistics* **7**(1), 295–318.
- Liu, L. & Hudgens, M. G. (2014), ‘Large sample randomization inference of causal effects in the presence of interference’, *Journal of the American Statistical Association* **109**(505), 288–301.
- Manski, C. F. (1993), ‘Identification of Endogenous Social Effects: The Reflection Problem’, *The Review of Economic Studies* **60**(3), 531–542.
- Manski, C. F. (2013), ‘Identification of treatment response with social interactions’, *The Econometrics Journal* **16**(1), S1–S23.
- Puelz, D., Basse, G., Feller, A. & Toulis, P. (2019), ‘A graph-theoretic approach to randomization tests of causal effects under general interference’, *arXiv preprint arXiv:1910.10862*.

- Rambachan, A. & Shephard, N. (2019), ‘Econometric analysis of potential outcomes time series: instruments, shocks, linearity and the causal response function’, *arXiv preprint arXiv:1903.01637*.
- Robins, J. (1986), ‘A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect’, *Mathematical modelling* **7**(9-12), 1393–1512.
- Robins, J. M., Greenland, S. & Hu, F.-C. (1999*a*), ‘Estimation of the causal effect of a time-varying exposure on the marginal mean of a repeated binary outcome’, *Journal of the American Statistical Association* **94**(447), 687–700.
- Robins, J. M., Greenland, S. & Hu, F.-C. (1999*b*), ‘Estimation of the causal effect of a time-varying exposure on the marginal mean of a repeated binary outcome’, *Journal of the American Statistical Association* **94**(447), 687–700.
- Romano, J. P. & Wolf, M. (2000), ‘A more general central limit theorem for  $m$ -dependent random variables with unbounded  $m$ ’, *Statistics & Probability Letters* **47**(2), 115 – 124.
- Rosenbaum, P. R. (2007), ‘Interference between units in randomized experiments’, *Journal of the American Statistical Association* **102**(477), 191–200.
- Ross, N. (2011), ‘Fundamentals of stein’s method’, *Probability Surveys* **8**, 210–293.
- Sävje, F., Aronow, P. M. & Hudgens, M. G. (2017), ‘Average treatment effects in the presence of unknown interference’, *arXiv e-prints* p. arXiv:1711.06399.
- Sinclair, B., McConnell, M. & Green, D. P. (2012), ‘Detecting spillover effects: Design and analysis of multilevel experiments’, *American Journal of Political Science* **56**(4), 1055–1069.
- Sobel, M. E. (2006), ‘What do randomized studies of housing mobility demonstrate?’, *Journal of the American Statistical Association* **101**(476), 1398–1407.
- Tang, Y., Huang, C., Kastelman, D. & Bauman, J. (2020), ‘Control using predictions as covariates in switchback experiments’.
- Traud, A. L., Mucha, P. J. & Porter, M. A. (2012), ‘Social structure of facebook networks’, *Physica A: Statistical Mechanics and its Applications* **391**(16), 4165–4180.
- Ugander, J., Karrer, B., Backstrom, L. & Kleinberg, J. (2013), Graph cluster randomization: Network exposure to multiple universes, *in* ‘Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining’, KDD ’13, Association for Computing Machinery, New York, NY, USA, p. 329–337.



Ugander, J. & Yin, H. (2020), ‘Randomized Graph Cluster Randomization’, *arXiv e-prints* p. arXiv:2009.02297.

Vazquez-Bare, G. (2022), ‘Identification and estimation of spillover effects in randomized experiments’, *Journal of Econometrics* .

Wooldridge, J. M. (2010), *Econometric Analysis of Cross Section and Panel Data*, The MIT Press.

Xiong, R., Athey, S., Bayati, M. & Imbens, G. W. (2019), ‘Optimal experimental design for staggered rollouts’, *Available at SSRN 3483934* .

## A Standard population interference

This appendix focuses on estimating TEC under population interference and assumes that either the experiment was conducted over a single time point or that there are no carryover effects. In both cases, we drop the subscript  $t$  for the remainder of the section. Our setup is now equivalent to the one studied in Liu & Hudgens (2014), Aronow & Samii (2017), Chin (2018) and Leung (2022). Our Horvitz-Thompson type estimator  $\hat{\tau}^{k,k'}$  now simplifies to,

$$\hat{\tau}^{k,k'} = \frac{1}{n} \sum_{i=1}^n \left[ \frac{\mathbf{1}(H_i = k)}{\pi_i(k)} Y_i(k) - \frac{\mathbf{1}(H_i = k')}{\pi_i(k')} Y_i(k') \right], \quad (\text{A.1})$$

where  $\pi_i(k) = \mathbb{P}(H_i = k)$  and  $\pi_i(k') = \mathbb{P}(H_i = k')$ .

Aronow & Samii (2017) showed that if the potential outcomes and inverse exposure probabilities are bounded, and the number of dependent pairs of  $H_i$ 's is of order  $o(n^2)$ , then the estimator  $\hat{\tau}^{k,k'}$  is consistent,

$$\left( \hat{\tau}^{k,k'} - \tau^{k,k'} \right) \rightarrow_{\mathbb{P}} 0.$$

In addition, the authors provided an asymptotically conservative confidence interval of  $\hat{\tau}^{k,k'}$  and implicitly outlined a version of a central limit theorem in the proof. However, the conditions stated in their derivations were sufficient but not necessary. Below, we establish a central limit theorem for  $\hat{\tau}^{k,k'}$  under weaker conditions and provide a detailed proof that builds on recent results by Chin (2018). We then illustrate the trade-offs between the strength of the interference structure assumption and the assignment mechanism's flexibility.

### A.1 A central limit theorem

We can now state the following central limit theorem for temporal exposure contrast.

**Theorem 7.** *Under Assumptions 2-4 and the condition that  $d_n = o(n^{1/4})$ , we have*

$$\frac{\sqrt{n}(\hat{\tau}^{k,k'} - \tau^{k,k'})}{\text{Var}(\sqrt{n}\hat{\tau}^{k,k'})^{1/2}} \xrightarrow{d} \mathcal{N}(0, 1)$$

as  $n \rightarrow \infty$ .

The assumption that  $d_n = o(n^{1/4})$  quantifies the dependence among observations due to interference. Let  $d_n$  be the maximal degree in this graph, which is equal to the maximal number of dependent exposure values for each unit; this is different from the definition of  $d_n$  in the main paper, where it is defined as a limit or a maximal over  $t$ .

Theorem 7 strengthens the result of Aronow & Samii (2017) in two ways. First, our Assumption 4 weakens Condition 6 of Aronow & Samii (2017), which requires the convergence of  $\text{Var}(\sqrt{n}\hat{\tau}_t^{k,k'})$ .

Second, we allow for a higher range of dependence ( $d_n = o(n^{1/4})$ ) compared to  $d_n = O(1)$  as in Aronow & Samii (2017)) among exposure values. The proof of this theorem relies on recent results in Chin (2018).

## A.2 Design and interference structure: a trade-off

Intuitively, Theorem 7 asserts that asymptotic normality holds so long as the dependency relations among the  $H_i$ 's are moderate. However, since  $H_i = f_i(W_{1:n})$  is determined by both function  $f_i$  and assignment  $W$ , the dependence structure among the  $H_i$ 's — and therefore the value of  $d_n$  — depends on both the exposure specification and the assignment mechanism.

This suggests that there exists a trade-off between the strength of the dependence in the  $W_i$ 's induced by the assignment mechanism and the dependence induced by the interference structure. The less restricted the interference structure is, the more restricted the assignment mechanism must be; in reverse, the more restricted the interference structure, the more flexible one can be with the design. We illustrate these insights with three special cases of Theorem 7, applied to popular settings. We should also note that our condition on  $d_n$  is not a sufficient condition for the central limit theorem. For example, if we consider  $f_i(W_i) = W_i$  (i.e., there is no interference) and  $W$  follows completely randomized design, then the central limit theorem still holds (see Theorem 1 in Ding (2017)). The discussion here mainly illustrates the entanglement between the assignment mechanism and the interference structure from a general perspective.

**Example 2.** *Suppose that the interference structure among  $n$  units is adequately described by a social network  $\mathcal{A}_n$ , and assume that the exposure mapping is of the form  $f_i(W_{1:n}) = f_i(W_{N_i})$ ; that is, only the neighbors' assignments matter. Let  $\delta_n$  be the maximal number of neighbors a unit can have in the network  $\mathcal{A}_n$  — which is distinct from the dependency graph. Then if  $\delta_n = o(n^{1/8})$  and the  $W_i$ 's are independent (i.e., the design is Bernoulli), then  $d_n = o(n^{1/4})$  as required by Theorem 7.*

This first example explores one extreme end of the trade-off, in which the assignment mechanism is maximally restricted — the  $W_i$ 's are independent — which allows for a comparatively large amount of interference.

**Example 3.** *We consider the graph cluster randomization approach (Ugander et al. (2013)) in which case we group units into clusters and randomize at the cluster level. Following the notations in Ugander et al. (2013), we let the vertices be partitioned into  $n_c$  clusters  $C_1, \dots, C_{n_c}$ . The graph cluster randomization approach assigns either treatment or control to all the units in each cluster. Suppose one's potential outcomes depend only on the assignments of its neighbors. Let  $\delta_n$  be the maximal number of neighbors one can have and  $c_n$  be the maximal size of the cluster. Then  $d_n = o(n^{1/4})$  for  $\delta_n^2 + \delta_n c_n = o(n^{1/4})$ .*

**Example 4.** Another commonly studied scenario is the “household” interference (Basse & Feller (2018); Dufló & Saez (2003)). In household interference, we assume that each unit belongs to a “household” and their potential outcomes depend only on the assignments of the units within the “household”. Suppose we have a two-stage design such that we first assign each household into treatment group or control group independently and then we assign treatments to units in each household depending on the assignment of their associated household. Let  $r_n$  be the maximal size of the “household”, then  $d_n = o(n^{1/4})$  for  $r_n = o(n^{1/4})$ .

Table 6 summarizes the above three examples. In Example 2, to have a general network interference setting with the maximum possible number of neighbors for each unit, we constrain the design to be the Bernoulli design. Further limiting the interference, like in Example 4 where the interference is restricted within households, we can have a more complex two-stage design. In the same spirit, Example 3 shows that for a highly dependent design, we need an even stronger condition on the interference structure, indicated by a stronger rate condition on  $\delta_n$ . In general, a weaker assumption on the interference structure induces a more complex dependence graph for the exposures, which in turn reduces our flexibility in the choice of design.

Interference	Design	Conditions
Network Interference	Bernoulli Design	$\delta_n = o(n^{1/8})$
Network Interference	Graph Cluster Randomization	$\delta_n^2 + \delta_n c_n = o(n^{1/4})$
Group Interference	Two-stage Design	$r_n = o(n^{1/4})$

Table 6: Trade-off between design and interference

### A.3 Inference

The central limit theorem stated in Theorem 7 serves as our basis for inference.

**Proposition 7.** Assuming all the assumptions in Theorem 7, then for any  $\delta > 0$ , we have that

$$\mathbb{P} \left( \frac{\widehat{\text{Var}}(\hat{\tau}^{k,k'})}{\text{Var}(\hat{\tau}^{k,k'})} \geq 1 - \delta \right) \rightarrow 1.$$

where  $\widehat{\text{Var}}(\hat{\tau}^{k,k'}) = n^{-1} \widehat{\text{Var}}(\sqrt{n} \hat{\tau}^{k,k'})$ . Therefore, we can construct asymptotically conservative confidence interval based on the variance estimator: for any  $\delta > 0$ ,

$$\mathbb{P} \left( \hat{\tau}^{k,k'} \in \left[ \hat{\tau}^{k,k'} - \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{1-\delta}} \sqrt{\widehat{\text{Var}}(\hat{\tau}^{k,k'})}, \right. \right. \\ \left. \left. \hat{\tau}^{k,k'} + \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{1-\delta}} \sqrt{\widehat{\text{Var}}(\hat{\tau}^{k,k'})} \right] \right) \geq 1 - \alpha$$

for large  $n$ .

$\widehat{\text{Var}}(\sqrt{n}\hat{\tau}^{k,k'})$  is the same as the one given in Proposition 2. Once again, this result strengthens that of Aronow & Samii (2017) by both removing the requirement that  $n\text{Var}(\hat{\tau}^{k,k'})$  converge, and by relaxing the constraint on the interference mechanism. Note that here  $\delta > 0$  is arbitrary and we present detailed simulations in Section 5 with  $\delta = 0.04$ .

## B Proofs and additional discussions

To begin with, we provide technical tools that we will use in our proofs. We first state a lemma from Ross (2011):

**Lemma 1.** *Let  $X_1, \dots, X_n$  be a collection of random variables such that  $\mathbb{E}[X_i^4] < \infty$  and  $\mathbb{E}[X_i] = 0$ . Let  $\sigma^2 = \text{Var}(\sum_i X_i)$  and  $S = \sum_i X_i$ . Let  $d$  be the maximal degree of the dependency graph of  $(X_1, \dots, X_n)$ . Then for constants  $C_1$  and  $C_2$  which do not depend on  $n, d$  or  $\sigma^2$ ,*

$$d_{\mathcal{W}}(S/\sigma) \leq C_1 \frac{d^{3/2}}{\sigma^2} \left( \sum_{i=1}^n \mathbb{E}[X_i^4] \right)^{1/2} + C_2 \frac{d^2}{\sigma^3} \sum_{i=1}^n \mathbb{E}|X_i|^3, \quad (\text{B.1})$$

where  $d_{\mathcal{W}}(S/\sigma)$  is the Wasserstein distance between  $S/\sigma$  and standard Gaussian.

Second, we provide the expression for the variance of  $\hat{\tau}^{k,k'}$ :

**Lemma 2** (Variance of Horvitz-Thompson estimator). *We have that (Aronow & Samii (2017)):*

$$\begin{aligned} \text{Var}(\sqrt{n}\hat{\tau}^{k,k'}) &= \frac{1}{n} \sum_{i=1}^n \pi_i(k)(1 - \pi_i(k)) \left( \frac{Y_i(k)}{\pi_i(k)} \right)^2 \\ &+ \frac{1}{n} \sum_{i=1}^n \pi_i(k')(1 - \pi_i(k')) \left( \frac{Y_i(k')}{\pi_i(k')} \right)^2 + \frac{2}{n} \sum_{i=1}^n Y_i(k)Y_i(k') \\ &+ \frac{1}{n} \sum_{i=1}^n \sum_{j \neq i} \left\{ [\pi_{ij}(k) - \pi_i(k)\pi_j(k)] \frac{Y_i(k) Y_j(k)}{\pi_i(k) \pi_j(k)} \right. \\ &\quad \left. + [\pi_{ij}(k') - \pi_i(k')\pi_j(k')] \frac{Y_i(k') Y_j(k')}{\pi_i(k') \pi_j(k')} \right\} \\ &- \frac{2}{n} \sum_{i=1}^n \sum_{j \neq i} \left\{ [\pi_{ij}(k, k') - \pi_i(k)\pi_j(k')] \frac{Y_i(k) Y_j(k')}{\pi_i(k) \pi_j(k')} \right\} \end{aligned}$$

Here  $\pi_{ij}(k) = \mathbb{P}(H_i = k \text{ and } H_j = k)$

*Proof of Theorem 7.* Note that  $\hat{\tau}^{k,k'} = \sum_{i=1}^n \tilde{\tau}_i$  where

$$\tilde{\tau}_i = \frac{1}{n} \left[ \frac{\mathbf{1}(H_i = k)}{\pi_i(k)} Y_i(k) - \frac{\mathbf{1}(H_i = k')}{\pi_i(k')} Y_i(k') \right]$$

and  $\mathbb{E}[\tilde{\tau}_i] = \frac{1}{n} [Y_i(k) - Y_i(k')]$ , hence if we let  $X_i = \sqrt{n}(\tilde{\tau}_i - \mathbb{E}[\tilde{\tau}_i])$ , then  $\sqrt{n}(\hat{\tau}^{k,k'} - \tau^{k,k'}) = \sum_{i=1}^n X_i = S$ . By Assumption 2 and Assumption 3, we know that  $X_i = O_p(n^{-1/2})$ , hence there exist some constants  $C_1$  and  $C_2$  such that for sufficiently large  $n$ , both

$$\left( \sum_{i=1}^n \mathbb{E}[X_i^4] \right)^{1/2} \leq C_1 n^{-1/2}$$

and

$$\sum_{i=1}^n \mathbb{E}|X_i|^3 \leq C_2 n^{-1/2}$$

hold. Moreover, by Assumption 4,

$$\sigma^2 = \text{Var}\left(\sum_i X_i\right) = n \text{Var}(\hat{\tau}^{k,k'})$$

is at least  $O(1)$ . Note that  $X_i$  is a function of  $H_i$ , hence  $X_i$  and  $X_j$  are not independent if and only if  $H_i$  and  $H_j$  are not independent. Since  $d_n = o(n^{1/4})$ , we know that the maximal degree of the dependency graph of  $X_i$ 's is  $o(n^{1/4})$ . Now we apply Lemma 1. Since  $\sigma^2$  is at least  $O(1)$ , we get:

$$\text{RHS of (B.1)} = o(n^{-1/8}) + o(1) \rightarrow 0$$

We're done. □

**Remark 1.** *In fact, with the tools in Leung (2022), we can proof this theorem with a weaker condition on  $d_n$ :  $d_n = O(\log n)$ .*

*Proof of Example 2.* Note that  $H_i$  is a function of  $W_i$  and  $W_j$ 's for  $j$  being a neighbor of  $i$ . If  $H_i$  and  $H_j$  are dependent, there must be the case that  $(\{i\} \cup \mathcal{N}_i) \cap (\{j\} \cup \mathcal{N}_j)$  is nonempty since we have the Bernoulli design. Hence, for each fixed unit  $i$ , there are at most  $\delta_n$  units such that the above intersection is nonempty. □

*Proof of Example 4.* We use the same reasoning as in the above proof. The only change is that now we know that each unit is belonged to a group and units in the group are connected. Therefore, for each fixed unit  $i$ , all the units outside the group will not have effect on unit  $i$ . As a result, we can have  $r_n = o(n^{1/4})$ . □

*Proof of Example 3.* Since we do not have Bernoulli design anymore, there might be the case that  $W_i$  and  $W_j$  are dependent, hence except  $(\{i\} \cup \mathcal{N}_i) \cap (\{j\} \cup \mathcal{N}_j)$  is nonempty, there is another case that makes  $H_i$  and  $H_j$  dependent: a neighbor of  $i$  is in the same cluster as a neighbor of  $j$ . For this case, we have at most  $\delta_n c_n$  such  $j$ 's for a fixed unit  $i$ . Hence, in total, there are at most  $\delta_n^2 + \delta_n c_n$   $j$ 's such that  $H_i$  and  $H_j$  are dependent. □

*Proof of Proposition 7.* We first prove the first part of the proposition. The proof is based on A.7 in Aronow & Samii (2017). To start with, for any  $(i, j) \in \{1, \dots, \} \times \{1, \dots, n\}$ , we define  $e_{ij} = 1$  if  $H_i$  and  $H_j$  are dependent and 0 otherwise. Let  $a_{ij}(H_i, H_j)$  be the sum of the elements in  $\widehat{\text{Var}}(\hat{\tau}^{k,k'})$

that incorporate  $i$  and  $j$ , then

$$\begin{aligned} \text{Var}\left(\widehat{\text{Var}}(\hat{\tau}^{k,k'})\right) &\leq n^{-4} \text{Var}\left[\sum_{i=1}^n \sum_{j=1}^n e_{ij} a_{ij}(H_i, H_j)\right] \\ &= n^{-4} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n \text{Cov}[e_{ij} a_{ij}(H_i, H_j), e_{kl} a_{kl}(H_k, H_l)] \end{aligned}$$

Note that  $\text{Cov}[e_{ij} a_{ij}(H_i, H_j), e_{kl} a_{kl}(H_k, H_l)]$  is nonzero if and only if  $e_{ij} = 1, e_{kl} = 1$  and at least one of  $e_{ik}, e_{il}, e_{jk}, e_{jl}$  is 1. In total, there are at most  $4nd_n^3$   $(i, j, k, l)$ 's satisfying this condition. And by Assumption 2 and 3, each covariance term is bounded, so we know that  $\text{Var}\left(\widehat{\text{Var}}(\hat{\tau}^{k,k'})\right) = o(n^{-4} \times n \times n^{3/4}) \rightarrow 0$  as  $n \rightarrow \infty$ . Then by Chebyshev's inequality,

$$\left|\widehat{\text{Var}}(\sqrt{n}\hat{\tau}^{k,k'}) - \mathbb{E}\left[\widehat{\text{Var}}(\sqrt{n}\hat{\tau}^{k,k'})\right]\right| = o_p(1).$$

Since  $\mathbb{E}\left[\widehat{\text{Var}}(\hat{\tau}^{k,k'})\right] \geq \text{Var}(\hat{\tau}^{k,k'})$ ,

$$\mathbb{P}\left(\frac{\widehat{\text{Var}}(\hat{\tau}^{k,k'})}{\text{Var}(\hat{\tau}^{k,k'})} \geq 1 - \delta\right) \rightarrow 1$$

for any  $\delta > 0$ .

Now we can prove the second part of the proposition. We have that

$$\begin{aligned} LHS &= \mathbb{P}\left(\left|\frac{\sqrt{n}(\hat{\tau}^{k,k'} - \tau^{k,k'})}{\sqrt{\text{Var}(\sqrt{n}\hat{\tau}^{k,k'})}}\right| \leq \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{1-\delta}} \sqrt{\frac{\widehat{\text{Var}}(\sqrt{n}\hat{\tau}^{k,k'})}{\text{Var}(\sqrt{n}\hat{\tau}^{k,k'})}}\right) \\ &\geq \mathbb{P}\left(\left|\frac{\sqrt{n}(\hat{\tau}^{k,k'} - \tau^{k,k'})}{\sqrt{\text{Var}(\sqrt{n}\hat{\tau}^{k,k'})}}\right| \leq \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{1-\delta}} \sqrt{\frac{\widehat{\text{Var}}(\sqrt{n}\hat{\tau}^{k,k'})}{\text{Var}(\sqrt{n}\hat{\tau}^{k,k'})}} \text{ and } \frac{\widehat{\text{Var}}(\sqrt{n}\hat{\tau}^{k,k'})}{\text{Var}(\sqrt{n}\hat{\tau}^{k,k'})} \geq 1 - \delta\right) \\ &\geq \mathbb{P}\left(\left|\frac{\sqrt{n}(\hat{\tau}^{k,k'} - \tau^{k,k'})}{\sqrt{\text{Var}(\sqrt{n}\hat{\tau}^{k,k'})}}\right| \leq z_{1-\frac{\alpha}{2}} \text{ and } \frac{\widehat{\text{Var}}(\sqrt{n}\hat{\tau}^{k,k'})}{\text{Var}(\sqrt{n}\hat{\tau}^{k,k'})} \geq 1 - \delta\right) \\ &= \mathbb{P}\left(\left|\frac{\sqrt{n}(\hat{\tau}^{k,k'} - \tau^{k,k'})}{\sqrt{\text{Var}(\sqrt{n}\hat{\tau}^{k,k'})}}\right| \leq z_{1-\frac{\alpha}{2}}\right) - \mathbb{P}\left(\left|\frac{\sqrt{n}(\hat{\tau}^{k,k'} - \tau^{k,k'})}{\sqrt{\text{Var}(\sqrt{n}\hat{\tau}^{k,k'})}}\right| \leq z_{1-\frac{\alpha}{2}} \text{ and } \frac{\widehat{\text{Var}}(\sqrt{n}\hat{\tau}^{k,k'})}{\text{Var}(\sqrt{n}\hat{\tau}^{k,k'})} < 1 - \delta\right) \end{aligned} \tag{B.2}$$



Now,

$$\begin{aligned}
(\text{B.2}) &\geq \mathbb{P} \left( \left| \frac{\sqrt{n}(\hat{\tau}^{k,k'} - \tau^{k,k'})}{\sqrt{\text{Var}(\sqrt{n}\hat{\tau}^{k,k'})}} \right| \leq z_{1-\frac{\alpha}{2}} \right) \\
&\quad - \mathbb{P} \left( \frac{\widehat{\text{Var}}(\sqrt{n}\hat{\tau}^{k,k'})}{\text{Var}(\sqrt{n}\hat{\tau}^{k,k'})} < 1 - \delta \right) \\
&= \mathbb{P} \left( \left| \frac{\sqrt{n}(\hat{\tau}^{k,k'} - \tau^{k,k'})}{\sqrt{\text{Var}(\sqrt{n}\hat{\tau}^{k,k'})}} \right| \leq z_{1-\frac{\alpha}{2}} \right) - \mathbb{P} \left( \frac{\widehat{\text{Var}}(\hat{\tau}^{k,k'})}{\text{Var}(\hat{\tau}^{k,k'})} < 1 - \delta \right) \\
&\rightarrow 1 - \alpha
\end{aligned}$$

as  $n \rightarrow \infty$  by the first part and Theorem 7. □

*Proof of Theorem 1.* We use a characteristic function argument. We first note that

$$\begin{aligned}
\frac{\sqrt{nT}(\hat{\tau}^{k,k'} - \bar{\tau}^{k,k'})}{\sqrt{\frac{1}{T} \sum_{t=1}^T \sigma_{n,t}^2}} &= \frac{\sqrt{nT}(\frac{1}{T} \sum_{t=1}^T \hat{\tau}_t^{k,k'} - \frac{1}{T} \sum_{t=1}^T \tau_t^{k,k'})}{\sqrt{\frac{1}{T} \sum_{t=1}^T \sigma_{n,t}^2}} \\
&= \frac{\sqrt{T} \frac{1}{T} \sum_{t=1}^T \sqrt{n}(\hat{\tau}_t^{k,k'} - \tau_t^{k,k'})}{\sqrt{\frac{1}{T} \sum_{t=1}^T \sigma_{n,t}^2}} \\
&= \frac{\frac{1}{\sqrt{T}} \sum_{t=1}^T X_{n,t}}{\sqrt{\frac{1}{T} \sum_{t=1}^T \sigma_{n,t}^2}},
\end{aligned}$$

where  $X_{n,t} = \sqrt{n}(\hat{\tau}_t^{k,k'} - \tau_t^{k,k'})$ . Now,

$$\mathbb{E} \left[ \exp \left\{ i\lambda \frac{\sqrt{nT}(\hat{\tau}^{k,k'} - \bar{\tau}^{k,k'})}{\sqrt{\frac{1}{T} \sum_{t=1}^T \sigma_{n,t}^2}} \right\} \right] \tag{B.3}$$

$$\begin{aligned}
&= \mathbb{E} \left[ \exp \left\{ i\lambda \frac{\frac{1}{\sqrt{T}} \sum_{t=1}^T X_{n,t}}{\sqrt{\frac{1}{T} \sum_{t=1}^T \sigma_{n,t}^2}} \right\} \right] \\
&= \prod_{t=1}^T \mathbb{E} \left[ \exp \left\{ i\lambda \frac{\frac{1}{\sqrt{T}} X_{n,t}}{\sqrt{\frac{1}{T} \sum_{t=1}^T \sigma_{n,t}^2}} \right\} \right] \\
&= \prod_{t=1}^T \mathbb{E} \left[ \exp \left\{ i \frac{\lambda \sigma_{n,t}}{\sqrt{\sum_{t=1}^T \sigma_{n,t}^2}} \frac{X_{n,t}}{\sigma_{n,t}} \right\} \right] \\
&= \prod_{t=1}^T \phi_{\frac{X_{n,t}}{\sigma_{n,t}}} \left( \frac{\lambda \sigma_{n,t}}{\sqrt{\sum_{t=1}^T \sigma_{n,t}^2}} \right)
\end{aligned} \tag{B.4}$$

The second equality follows from our assumption that assignment vectors are independent across time and  $\phi_X$  denotes the characteristic function of a random variable  $X$ . Pick  $\epsilon > 0$ . Now, to conclude the proof, we note that

$$\phi_{\frac{X_{n,t}}{\sigma_{n,t}}}(\theta) \rightarrow e^{-\frac{\theta^2}{2}}$$

for any  $t \in \{1, \dots, T\}$ . Moreover, for each  $t$ , the convergence is actually uniform on any bounded interval. Therefore, for any  $t \in \{1, \dots, T\}$ ,

$$\phi_{\frac{X_{n,t}}{\sigma_{n,t}}}(\theta) \rightarrow e^{-\frac{\theta^2}{2}} \text{ uniformly on } (0, 1).$$

Note that

$$\frac{\lambda \sigma_{n,t}}{\sqrt{\sum_{t=1}^T \sigma_{n,t}^2}} \in (0, 1),$$

so for any  $t$ ,  $\exists N_t \in \mathbb{N}$  such that for any  $n \geq N_t$ ,

$$\left| \phi_{\frac{X_{n,t}}{\sigma_{n,t}}} \left( \frac{\lambda \sigma_{n,t}}{\sqrt{\sum_{t=1}^T \sigma_{n,t}^2}} \right) - \exp \left\{ -\frac{1}{2} \frac{\lambda^2 \sigma_{n,t}^2}{\sum_{t=1}^T \sigma_{n,t}^2} \right\} \right| = |\epsilon_t| \leq \frac{1}{2K}.$$

Let  $N = \max\{N_1, \dots, N_T\}$ , then for all  $n \geq N$ , and for all  $t \in \{1, \dots, T\}$ ,

$$\left| \phi_{\frac{X_{n,t}}{\sigma_{n,t}}} \left( \frac{\lambda \sigma_{n,t}}{\sqrt{\sum_{t=1}^T \sigma_{n,t}^2}} \right) - \exp \left\{ -\frac{1}{2} \frac{\lambda^2 \sigma_{n,t}^2}{\sum_{t=1}^T \sigma_{n,t}^2} \right\} \right| = |\epsilon_t| \leq \frac{1}{2K},$$

where  $K$  is any big number we want. Now,

$$\begin{aligned} (B.4) &= \prod_{t=1}^T \left( \exp \left\{ -\frac{1}{2} \frac{\lambda^2 \sigma_{n,t}^2}{\sum_{t=1}^T \sigma_{n,t}^2} \right\} + \epsilon_t \right) \\ &= \exp \left\{ -\frac{1}{2} \lambda^2 \right\} + R(\epsilon_t), \end{aligned}$$

where  $R(\epsilon_t)$  is a remainder term that is the sum of several monomial terms of  $\epsilon_t$ 's. Note that  $\exp \left\{ -\frac{1}{2} \frac{\lambda^2 \sigma_{n,t}^2}{\sum_{t=1}^T \sigma_{n,t}^2} \right\}$  is actually bounded by 1, hence by making  $K$  sufficiently large, we can make  $R(\epsilon_t)$  arbitrarily small. Pick such  $K$ , then we know that for sufficiently large  $n$ ,

$$\left| \mathbb{E} \left[ \exp \left\{ i\lambda \frac{\sqrt{nT}(\hat{\tau}^{k,k'} - \bar{\tau}^{k,k'})}{\sqrt{\frac{1}{T} \sum_{t=1}^T \sigma_{n,t}^2}} \right\} \right] - \exp \left\{ -\frac{1}{2} \lambda^2 \right\} \right| \leq \epsilon.$$

Hence, by standard characteristic function argument, we complete the proof of the theorem.  $\square$

To prove Theorem 3, we first state the following version of Lindeberg-Feller central limit theorem.

**Lemma 3** (Lindeberg-Feller CLT). *Let  $\{k_n\}_{n \geq 1}$  be a sequence of positive integers increasing to infinity. For each  $n$ , let  $\{X_{n,i}\}_{1 \leq i \leq k_n}$  is a collection of independent random variables. Let  $\mu_{n,i} := \mathbb{E}(X_{n,i})$  and*

$$s_n^2 := \sum_{i=1}^{k_n} \text{Var}(X_{n,i}).$$

Suppose that for any  $\epsilon > 0$ ,

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^2} \sum_{i=1}^{k_n} \mathbb{E}((X_{n,i} - \mu_{n,i})^2; |X_{n,i} - \mu_{n,i}| \geq \epsilon s_n) = 0. \quad (\text{B.5})$$

Then the random variable

$$\frac{\sum_{i=1}^{k_n} (X_{n,i} - \mu_{n,i})}{s_n} \xrightarrow{d} \mathcal{N}(0, 1)$$

as  $n \rightarrow \infty$ .

*Proof of Theorem 3.* We first prove the theorem with condition (5). We note that  $\sqrt{nT}(\hat{\tau}^{k,k'} - \bar{\tau}^{k,k'}) = \sum_{t=1}^T \sqrt{\frac{n}{T}}(\hat{\tau}_t^{k,k'} - \tau_t^{k,k'})$ . Let  $X_{n,t} = \sqrt{\frac{n}{T}}\hat{\tau}_t^{k,k'}$ , then  $\mu_{n,t} = \sqrt{\frac{n}{T}}\tau_t^{k,k'}$ , so the numerator is exactly  $\sum_{t=1}^T (X_{n,t} - \mu_{n,t})$ . Moreover, note that for any  $n$ ,  $X_{n,1}, \dots, X_{n,T}$  are independent by the pure population interference assumption. Now,

$$\begin{aligned} s_n^2 &= \sum_{t=1}^T \text{Var}(X_{n,t}) \\ &= \sum_{t=1}^T \text{Var}\left(\sqrt{\frac{n}{T}}\hat{\tau}_t^{k,k'}\right) \\ &= \frac{1}{T} \sum_{t=1}^T \text{Var}(\sqrt{n}\hat{\tau}_t^{k,k'}) \\ &= \frac{1}{T} \sum_{t=1}^T \sigma_{n,t}^2. \end{aligned}$$

Hence, to finish the proof, we only need to check (B.5) is satisfied. Notice that for any  $\epsilon > 0$ ,

$$\begin{aligned} |X_{n,t} - \mu_{n,t}| \geq \epsilon s_n &\Leftrightarrow \left| \sqrt{\frac{n}{T}}\hat{\tau}_t^{k,k'} - \sqrt{\frac{n}{T}}\tau_t^{k,k'} \right| \geq \epsilon \sqrt{\frac{1}{T} \sum_{t=1}^T \sigma_{n,t}^2} \\ &\Leftrightarrow \left| \hat{\tau}_t^{k,k'} - \tau_t^{k,k'} \right| \geq \epsilon \sqrt{\frac{1}{n} \sum_{t=1}^T \sigma_{n,t}^2} \end{aligned}$$

By Assumption 4,  $\sigma_{n,t}^2 \geq c$  for some  $c > 0$  and for all  $n$  large. Hence

$$\epsilon \sqrt{\frac{1}{n} \sum_{t=1}^T \sigma_{n,t}^2} \geq \epsilon \sqrt{\frac{T}{n}} c \rightarrow \infty.$$

Note that by Assumptions 2 and 3,  $\left| \hat{\tau}_t^{k,k'} - \tau_t^{k,k'} \right|$  is uniformly bounded. Hence for sufficiently large  $n$ ,  $\left| \hat{\tau}_t^{k,k'} - \tau_t^{k,k'} \right| < \epsilon \sqrt{\frac{1}{n} \sum_{t=1}^T \sigma_{n,t}^2}$  for all  $t$ . Therefore, for sufficiently large  $n$ ,

$$\frac{1}{s_n^2} \sum_{t=1}^T \mathbb{E} \left( (X_{n,t} - \mu_{n,t})^2; |X_{n,t} - \mu_{n,t}| \geq \epsilon s_n \right) = 0.$$

As a result, (B.5) is satisfied. We're done. The proof of this theorem with condition (6) is exactly the same as in single time step case once we notice that the numerator is just a sum of  $nT$  mean 0 dependent random variables.  $\square$

To prove Theorem 2, we need the following version of Lyapunov central limit theorem.

**Lemma 4** (Lyapunov CLT). *Let  $\{X_n\}_{n=1}^\infty$  be a sequence of independent random variables. Let  $\mu_i := \mathbb{E}(X_i)$  and*

$$s_n^2 = \sum_{i=1}^n \text{Var}(X_i).$$

If for some  $\delta > 0$ ,

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^n \mathbb{E} |X_i - \mu_i|^{2+\delta} = 0, \quad (\text{B.6})$$

then the random variable

$$\frac{\sum_{i=1}^n (X_i - \mu_i)}{s_n} \xrightarrow{d} \mathcal{N}(0, 1)$$

*Proof of Theorem 2.* This time, we let  $X_t = \sqrt{\frac{n}{T}} \hat{\tau}_t^{k,k'}$  then the numerator is  $\sum_{t=1}^T (X_t - \mu_t)$ . Since we have pure population interference,  $\{X_t\}_{t=1}^\infty$  are independent. Now,

$$\begin{aligned} s_T^2 &= \sum_{t=1}^T \text{Var}(X_t) \\ &= \frac{1}{T} \sum_{t=1}^T \sigma_{n,t}^2. \end{aligned}$$

Hence, we only need to check (B.6). We have that

$$\begin{aligned} &\lim_{T \rightarrow \infty} \frac{1}{s_T^{2+\delta}} \sum_{t=1}^T \mathbb{E} |X_t - \mu_t|^{2+\delta} \\ &= \lim_{T \rightarrow \infty} \frac{1}{s_T^{2+\delta}} \left( \frac{n}{T} \right)^{1+\frac{\delta}{2}} \sum_{t=1}^T \mathbb{E} \left| \hat{\tau}_t^{k,k'} - \tau_t^{k,k'} \right|^{2+\delta} \end{aligned}$$

Now, by Assumptions 2 and 3,  $\exists M > 0$  such that  $\left| \hat{\tau}_t^{k,k'} - \tau_t^{k,k'} \right| \leq M$  for all  $t$ . Hence,

$$\begin{aligned} & \frac{1}{s_T^{2+\delta}} \left( \frac{n}{T} \right)^{1+\frac{\delta}{2}} \sum_{t=1}^T \mathbb{E} \left| \hat{\tau}_t^{k,k'} - \tau_t^{k,k'} \right|^{2+\delta} \\ & \leq \frac{1}{s_T^{2+\delta}} \left( \frac{n}{T} \right)^{1+\frac{\delta}{2}} T M^{2+\delta} \\ & = \frac{1}{s_T^{2+\delta}} \frac{n^{1+\frac{\delta}{2}}}{T} M^{2+\delta} \end{aligned}$$

If  $T \rightarrow \infty$ ,  $\frac{1}{s_T^{2+\delta}} \frac{n^{1+\frac{\delta}{2}}}{T} M^{2+\delta} \rightarrow 0$ . Therefore, (B.6) is satisfied. We're done.  $\square$

*Proof of Proposition 2.* Now we can prove the second part of the proposition. We have that

$$\begin{aligned} LHS &= \mathbb{P} \left( \left| \frac{\sqrt{nT}(\hat{\tau}^{k,k'} - \bar{\tau}^{k,k'})}{\sqrt{\frac{1}{T} \sum_{t=1}^T \text{Var}(\sqrt{n}\hat{\tau}_t^{k,k'})}} \right| \leq \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{1-\delta}} \sqrt{\frac{\frac{1}{T} \sum_{t=1}^T \widehat{\text{Var}}(\sqrt{n}\hat{\tau}_t^{k,k'})}{\frac{1}{T} \sum_{t=1}^T \text{Var}(\sqrt{n}\hat{\tau}_t^{k,k'})}} \right) \\ &\geq \mathbb{P} \left( \left| \frac{\sqrt{nT}(\hat{\tau}^{k,k'} - \bar{\tau}^{k,k'})}{\sqrt{\frac{1}{T} \sum_{t=1}^T \text{Var}(\sqrt{n}\hat{\tau}_t^{k,k'})}} \right| \leq \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{1-\delta}} \sqrt{\frac{\frac{1}{T} \sum_{t=1}^T \widehat{\text{Var}}(\sqrt{n}\hat{\tau}_t^{k,k'})}{\frac{1}{T} \sum_{t=1}^T \text{Var}(\sqrt{n}\hat{\tau}_t^{k,k'})}} \text{ and } \frac{\frac{1}{T} \sum_{t=1}^T \widehat{\text{Var}}(\hat{\tau}_t^{k,k'})}{\frac{1}{T} \sum_{t=1}^T \text{Var}(\hat{\tau}_t^{k,k'})} \geq 1 - \delta \right) \\ &\geq \mathbb{P} \left( \left| \frac{\sqrt{nT}(\hat{\tau}^{k,k'} - \bar{\tau}^{k,k'})}{\sqrt{\frac{1}{T} \sum_{t=1}^T \text{Var}(\sqrt{n}\hat{\tau}_t^{k,k'})}} \right| \leq z_{1-\frac{\alpha}{2}} \text{ and } \frac{\frac{1}{T} \sum_{t=1}^T \widehat{\text{Var}}(\hat{\tau}_t^{k,k'})}{\frac{1}{T} \sum_{t=1}^T \text{Var}(\hat{\tau}_t^{k,k'})} \geq 1 - \delta \right) \end{aligned} \quad (\text{B.7})$$

Furthermore, we have that

$$\begin{aligned} (\text{B.7}) &= \mathbb{P} \left( \left| \frac{\sqrt{nT}(\hat{\tau}^{k,k'} - \bar{\tau}^{k,k'})}{\sqrt{\frac{1}{T} \sum_{t=1}^T \text{Var}(\sqrt{n}\hat{\tau}_t^{k,k'})}} \right| \leq z_{1-\frac{\alpha}{2}} \right) \\ &\quad - \mathbb{P} \left( \left| \frac{\sqrt{nT}(\hat{\tau}^{k,k'} - \bar{\tau}^{k,k'})}{\sqrt{\frac{1}{T} \sum_{t=1}^T \text{Var}(\sqrt{n}\hat{\tau}_t^{k,k'})}} \right| \leq z_{1-\frac{\alpha}{2}} \text{ and } \frac{\frac{1}{T} \sum_{t=1}^T \widehat{\text{Var}}(\hat{\tau}_t^{k,k'})}{\frac{1}{T} \sum_{t=1}^T \text{Var}(\hat{\tau}_t^{k,k'})} < 1 - \delta \right) \\ &\geq \mathbb{P} \left( \left| \frac{\sqrt{nT}(\hat{\tau}^{k,k'} - \bar{\tau}^{k,k'})}{\sqrt{\frac{1}{T} \sum_{t=1}^T \text{Var}(\sqrt{n}\hat{\tau}_t^{k,k'})}} \right| \leq z_{1-\frac{\alpha}{2}} \right) - \mathbb{P} \left( \frac{\frac{1}{T} \sum_{t=1}^T \widehat{\text{Var}}(\hat{\tau}_t^{k,k'})}{\frac{1}{T} \sum_{t=1}^T \text{Var}(\hat{\tau}_t^{k,k'})} < 1 - \delta \right) \end{aligned}$$

So if we can show  $\mathbb{P} \left( \frac{\frac{1}{T} \sum_{t=1}^T \widehat{\text{Var}}(\hat{\tau}_t^{k,k'})}{\frac{1}{T} \sum_{t=1}^T \text{Var}(\hat{\tau}_t^{k,k'})} \geq 1 - \delta \right) \rightarrow 0$  then we are done. Notice that

$$\text{Var} \left( \frac{1}{T} \sum_{t=1}^T \widehat{\text{Var}}(\hat{\tau}_t^{k,k'}) \right) = \frac{1}{T^2} \sum_{t=1}^T \text{Var} \left( \widehat{\text{Var}}(\hat{\tau}_t^{k,k'}) \right). \quad (\text{B.8})$$

If  $T$  is fixed (i.e., Theorem 1 holds), then by what we have in Proposition 7, we immediately have that (B.8)  $\rightarrow 0$  and we are done. Now suppose Theorem 3 holds. Recall that

$$\begin{aligned} \text{Var} \left( \widehat{\text{Var}}(\hat{\tau}^{k,k'}) \right) &\leq n^{-4} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n \text{Cov} [e_{ij} a_{ij}(H_i, H_j), \\ &\quad e_{kl} a_{kl}(H_k, H_l)], \end{aligned}$$

which implies that  $\text{Var} \left( \widehat{\text{Var}}(\hat{\tau}_t^{k,k'}) \right)$  is uniformly bounded by a constant  $M$  by Assumption 2 and 3. So

$$\frac{1}{T^2} \sum_{t=1}^T \text{Var} \left( \widehat{\text{Var}}(\hat{\tau}_t^{k,k'}) \right) \leq \frac{1}{T^2} \sum_{t=1}^T M = \frac{M}{T} \rightarrow 0$$

as  $T \rightarrow 0$ . So in the regime where both  $n$  and  $T$  go to infinity (i.t. Theorem 3 holds) or  $T$  goes to infinity (i.e., Theorem 2 holds), (B.8)  $\rightarrow 0$  and we are done.  $\square$

*Proof of Theorem 4.* This should be exactly the same as our proof of Theorem 7.  $\square$

*Proof of Proposition 3.*

$$\begin{aligned} |\mathbb{E}[\hat{\tau}_t^c] - \tau_t^{TE}| &= |\mathbb{E}[\alpha \hat{\tau}_t^{TE} + (1 - \alpha) \hat{\tau}_{t-1}^{TE}] - \tau_t^{TE}| \\ &= |\alpha \tau_t^{TE} + (1 - \alpha) \tau_{t-1}^{TE} - \tau_t^{TE}| \\ &= |(1 - \alpha)(\tau_{t-1}^{TE} - \tau_t^{TE})| \\ &= (1 - \alpha) |\tau_{t-1}^{TE} - \tau_t^{TE}| \end{aligned}$$

The second equality follows from unbiasedness of  $\hat{\tau}_t^{TE}$  and  $\hat{\tau}_{t-1}^{TE}$ . To further bound the bias, we need to bound  $|\tau_{t-1}^{TE} - \tau_t^{TE}|$ . We do this below.

$$\begin{aligned} |\tau_{t-1}^{TE} - \tau_t^{TE}| &= \left| \left( \frac{1}{n} \sum_{i=1}^n Y_{i,t}(h_i^1) - \frac{1}{n} \sum_{i=1}^n Y_{i,t}(h_i^0) \right) - \left( \frac{1}{n} \sum_{i=1}^n Y_{i,t-1}(h_i^1) - \frac{1}{n} \sum_{i=1}^n Y_{i,t-1}(h_i^0) \right) \right| \\ &= \left| \frac{1}{n} \sum_{i=1}^n (Y_{i,t}(h_i^1) - Y_{i,t-1}(h_i^1)) - \frac{1}{n} \sum_{i=1}^n (Y_{i,t}(h_i^0) - Y_{i,t-1}(h_i^0)) \right| \\ &\leq \frac{1}{n} \sum_{i=1}^n |Y_{i,t}(h_i^1) - Y_{i,t-1}(h_i^1)| + \frac{1}{n} \sum_{i=1}^n |Y_{i,t}(h_i^0) - Y_{i,t-1}(h_i^0)| \\ &\leq 2\epsilon, \end{aligned}$$

by our  $\epsilon$ -weak-stability assumption. Hence,

$$|\mathbb{E}[\hat{\tau}_t^c] - \tau_t^{TE}| \leq 2(1 - \alpha)\epsilon.$$

$\square$

**Remark 2.** Note that following the exact derivation, we can know that

$$|\tau_t^{TE} - \tau_{t'}^{TE}| \leq 2|t - t'| \epsilon \quad (\text{B.9})$$

**Proposition 8** (Variance and Covariance of Horvitz-Thompson Type Estimators). For each  $i \in \{1, \dots, n\}, t \in \{1, \dots, T\}$ , we let  $\mathbb{P}(H_{i,t} = h_i^1) = \pi_{i,t}^1, \mathbb{P}(H_{i,t} = h_i^0) = \pi_{i,t}^0, \mathbb{P}(H_{j,t} = h_j^1) = \pi_{j,t}^1$  and  $\mathbb{P}(H_{j,t} = h_j^0) = \pi_{j,t}^0$ . Moreover, for each  $i \neq j$  and  $t$ , we let  $\mathbb{P}(H_{i,t} = h_i^1, H_{j,t} = h_j^1) = \pi_{i,j,t}^{1,1}, \mathbb{P}(H_{i,t} = h_i^0, H_{j,t} = h_j^0) = \pi_{i,j,t}^{0,0}, \mathbb{P}(H_{i,t} = h_i^1, H_{j,t} = h_j^0) = \pi_{i,j,t}^{1,0}$  and  $\mathbb{P}(H_{i,t} = h_i^0, H_{j,t} = h_j^1) = \pi_{i,j,t}^{0,1}$ , then

$$\begin{aligned} \text{Var}(\hat{\tau}_t^{TE}) &= \frac{1}{n^2} \sum_{i=1}^n \left[ \frac{Y_{i,t}^2(h_i^1)(1 - \pi_{i,t}^1)}{\pi_{i,t}^1} + \frac{Y_{i,t}^2(h_i^0)(1 - \pi_{i,t}^0)}{\pi_{i,t}^0} + 2Y_{i,t}(h_i^1)Y_{i,t}(h_i^0) \right] \\ &+ \frac{2}{n^2} \sum_{1 \leq i < j \leq n} \left[ \frac{Y_{i,t}(h_i^1)Y_{j,t}(h_j^1)(\pi_{i,j,t}^{1,1} - \pi_{i,t}^1\pi_{j,t}^1)}{\pi_{i,t}^1\pi_{j,t}^1} - \frac{Y_{i,t}(h_i^0)Y_{j,t}(h_j^1)(\pi_{i,j,t}^{0,1} - \pi_{i,t}^0\pi_{j,t}^1)}{\pi_{i,t}^0\pi_{j,t}^1} \right. \\ &\quad \left. - \frac{Y_{i,t}(h_i^1)Y_{j,t}(h_j^0)(\pi_{i,j,t}^{1,0} - \pi_{i,t}^1\pi_{j,t}^0)}{\pi_{i,t}^1\pi_{j,t}^0} + \frac{Y_{i,t}(h_i^0)Y_{j,t}(h_j^0)(\pi_{i,j,t}^{0,0} - \pi_{i,t}^0\pi_{j,t}^0)}{\pi_{i,t}^0\pi_{j,t}^0} \right] \end{aligned} \quad (\text{B.10})$$

As for  $\text{Cov}(\hat{\tau}_t^{TE}, \hat{\tau}_{t'}^{TE})$ , if we let  $\mathbb{P}(H_{i,t} = h_i^1, H_{i,t'} = h_i^1) = \pi_{i,t,t'}^{1,1}, \mathbb{P}(H_{i,t} = h_i^0, H_{i,t'} = h_i^1) = \pi_{i,t,t'}^{0,1}, \mathbb{P}(H_{i,t} = h_i^1, H_{i,t'} = h_i^0) = \pi_{i,t,t'}^{1,0}, \mathbb{P}(H_{i,t} = h_i^0, H_{i,t'} = h_i^0) = \pi_{i,t,t'}^{0,0}$  and  $\mathbb{P}(H_{i,t} = h_i^1, H_{j,t'} = h_j^1) = \pi_{i,t,j,t'}^{1,1}, \mathbb{P}(H_{i,t} = h_i^0, H_{j,t'} = h_j^1) = \pi_{i,t,j,t'}^{0,1}, \mathbb{P}(H_{i,t} = h_i^1, H_{j,t'} = h_j^0) = \pi_{i,t,j,t'}^{1,0}, \mathbb{P}(H_{i,t} = h_i^0, H_{j,t'} = h_j^0) = \pi_{i,t,j,t'}^{0,0}$ , then we have the following expression for  $\text{Cov}(\hat{\tau}_t^{TE}, \hat{\tau}_{t'}^{TE})$ :

$$\begin{aligned} &\frac{1}{n^2} \sum_{i=1}^n \left[ \frac{Y_{i,t}(h_i^1)Y_{i,t'}(h_i^1)(\pi_{i,t,t'}^{1,1} - \pi_{i,t}^1\pi_{i,t'}^1)}{\pi_{i,t}^1\pi_{i,t'}^1} - \frac{Y_{i,t}(h_i^0)Y_{i,t'}(h_i^1)(\pi_{i,t,t'}^{0,1} - \pi_{i,t}^0\pi_{i,t'}^1)}{\pi_{i,t}^0\pi_{i,t'}^1} \right. \\ &\quad \left. - \frac{Y_{i,t}(h_i^1)Y_{i,t'}(h_i^0)(\pi_{i,t,t'}^{1,0} - \pi_{i,t}^1\pi_{i,t'}^0)}{\pi_{i,t}^1\pi_{i,t'}^0} + \frac{Y_{i,t}(h_i^0)Y_{i,t'}(h_i^0)(\pi_{i,t,t'}^{0,0} - \pi_{i,t}^0\pi_{i,t'}^0)}{\pi_{i,t}^0\pi_{i,t'}^0} \right] \\ &+ \frac{2}{n^2} \sum_{1 \leq i < j \leq n} \left[ \frac{Y_{i,t}(h_i^1)Y_{j,t'}(h_j^1)(\pi_{i,t,j,t'}^{1,1} - \pi_{i,t}^1\pi_{j,t'}^1)}{\pi_{i,t}^1\pi_{j,t'}^1} - \frac{Y_{i,t}(h_i^0)Y_{j,t'}(h_j^1)(\pi_{i,t,j,t'}^{0,1} - \pi_{i,t}^0\pi_{j,t'}^1)}{\pi_{i,t}^0\pi_{j,t'}^1} \right. \\ &\quad \left. - \frac{Y_{i,t}(h_i^1)Y_{j,t'}(h_j^0)(\pi_{i,t,j,t'}^{1,0} - \pi_{i,t}^1\pi_{j,t'}^0)}{\pi_{i,t}^1\pi_{j,t'}^0} + \frac{Y_{i,t}(h_i^0)Y_{j,t'}(h_j^0)(\pi_{i,t,j,t'}^{0,0} - \pi_{i,t}^0\pi_{j,t'}^0)}{\pi_{i,t}^0\pi_{j,t'}^0} \right] \end{aligned} \quad (\text{B.11})$$

*Proof of Proposition 8.* This can be done by direct calculations.  $\square$

*Proof of Proposition 4.* We'd like to have reduction in MSE by using  $\hat{\tau}_t^c$ . By the bias-variance decomposition and note that  $\hat{\tau}_t^{TE}$  is unbiased, this boils down to

$$\text{Var}(\hat{\tau}_t^c) + |\mathbb{E}[\hat{\tau}_t^c] - \tau_t^{TE}|^2 \leq \text{Var}(\hat{\tau}_t^{TE})$$

By Proposition 3, it suffices to have

$$\text{Var}(\hat{\tau}_t^c) + 4(1 - \alpha)^2 \epsilon^2 \leq \text{Var}(\hat{\tau}_t^{TE}),$$

which is further equivalent to

$$\begin{aligned} & \alpha^2 \text{Var}(\hat{\tau}_t^{TE}) + (1 - \alpha)^2 \text{Var}(\hat{\tau}_{t-1}^{TE}) \\ & + 2\alpha(1 - \alpha) \text{Cov}(\hat{\tau}_t^{TE}, \hat{\tau}_{t-1}^{TE}) + 4(1 - \alpha)^2 \epsilon^2 \\ & \leq \text{Var}(\hat{\tau}_t^{TE}) \end{aligned} \quad (\text{B.12})$$

Rewrite (B.12), we have

$$\begin{aligned} & (4\epsilon^2 + \text{Var}(\hat{\tau}_t^{TE}) + \text{Var}(\hat{\tau}_{t-1}^{TE}) - 2\text{Cov}(\hat{\tau}_t^{TE}, \hat{\tau}_{t-1}^{TE})) \alpha^2 \\ & - (8\epsilon^2 + 2\text{Var}(\hat{\tau}_{t-1}^{TE}) - 2\text{Cov}(\hat{\tau}_t^{TE}, \hat{\tau}_{t-1}^{TE})) \alpha \\ & + (4\epsilon^2 + \text{Var}(\hat{\tau}_{t-1}^{TE}) - \text{Var}(\hat{\tau}_t^{TE})) \leq 0 \end{aligned} \quad (\text{B.13})$$

Now we look at the left hand side of (B.13), which is quadratic in  $\alpha$ . To ease notations, let  $A = \text{Var}(\hat{\tau}_t^{TE})$ ,  $B = \text{Var}(\hat{\tau}_{t-1}^{TE})$  and  $C = \text{Cov}(\hat{\tau}_t^{TE}, \hat{\tau}_{t-1}^{TE})$ . It's easy to see that the left hand side achieves its minimum at  $\alpha = \delta = 1 - \frac{2(A-C)}{8\epsilon^2 + 2A + 2B - 4C}$  and is 0 at  $\alpha = 1$ . So if we have  $\delta < 1$ , then for some  $\alpha \in (0, 1)$ , we have reduction in MSE. Moreover, if  $\delta < \frac{1}{2}$ , we then know that for  $\alpha = \frac{1}{2}$ , we also have smaller MSE by the property of quadratic functions. And simple algebra shows that  $\delta < \frac{1}{2}$  is equivalent to  $A - B > 4\epsilon^2$ .  $\square$

**Proposition 9** (Estimators of variance). *We define two estimators of the variance:*

$$\begin{aligned} \widehat{\text{Var}}^u(\hat{\tau}_t^{TE}) &= \frac{1}{n^2} \sum_{i=1}^n \left[ \mathbf{1}(H_{i,t} = h_i^1)(1 - \pi_{i,t}^1) \left( \frac{Y_{i,t}}{\pi_{i,t}^1} \right)^2 + \mathbf{1}(H_{i,t} = h_i^0)(1 - \pi_{i,t}^0) \left( \frac{Y_{i,t}}{\pi_{i,t}^0} \right)^2 \right. \\ & \quad \left. + \frac{Y_{i,t}^2}{\pi_{i,t}^1} \mathbf{1}(H_{i,t} = h_i^1) + \frac{Y_{i,t}^2}{\pi_{i,t}^0} \mathbf{1}(H_{i,t} = h_i^0) \right] \\ & + \frac{2}{n^2} \sum_{1 \leq i < j \leq n} \left[ \mathbf{1}(\pi_{i,j,t}^{1,1} \neq 0) \mathbf{1}(H_{i,t} = h_i^1) \mathbf{1}(H_{j,t} = h_j^1) \frac{(\pi_{i,j,t}^{1,1} - \pi_{i,t}^1 \pi_{j,t}^1) Y_{i,t} Y_{j,t}}{\pi_{i,t}^1 \pi_{j,t}^1 \pi_{i,j,t}^{1,1}} \right. \\ & \quad - \left( \mathbf{1}(\pi_{i,j,t}^{0,1} \neq 0) \mathbf{1}(H_{i,t} = h_i^0) \mathbf{1}(H_{j,t} = h_j^1) \frac{(\pi_{i,j,t}^{0,1} - \pi_{i,t}^0 \pi_{j,t}^1) Y_{i,t} Y_{j,t}}{\pi_{i,t}^0 \pi_{j,t}^1 \pi_{i,j,t}^{0,1}} \right. \\ & \quad \left. - \mathbf{1}(\pi_{i,j,t}^{0,1} = 0) \left( \frac{\mathbf{1}(H_{i,t} = h_i^0) Y_{i,t}^2}{2\pi_{i,t}^0} + \frac{\mathbf{1}(H_{j,t} = h_j^1) Y_{j,t}^2}{2\pi_{j,t}^1} \right) \right) \\ & \quad - \left( \mathbf{1}(\pi_{i,j,t}^{1,0} \neq 0) \mathbf{1}(H_{i,t} = h_i^1) \mathbf{1}(H_{j,t} = h_j^0) \times \frac{(\pi_{i,j,t}^{1,0} - \pi_{i,t}^1 \pi_{j,t}^0) Y_{i,t} Y_{j,t}}{\pi_{i,t}^1 \pi_{j,t}^0 \pi_{i,j,t}^{1,0}} \right. \\ & \quad \left. - \mathbf{1}(\pi_{i,j,t}^{1,0} = 0) \left( \frac{\mathbf{1}(H_{i,t} = h_i^1) Y_{i,t}^2}{2\pi_{i,t}^1} + \frac{\mathbf{1}(H_{j,t} = h_j^0) Y_{j,t}^2}{2\pi_{j,t}^0} \right) \right) \\ & \quad \left. + \mathbf{1}(\pi_{i,j,t}^{0,0} \neq 0) \frac{\mathbf{1}(H_{i,t} = h_i^0) \mathbf{1}(H_{j,t} = h_j^0) (\pi_{i,j,t}^{0,0} - \pi_{i,t}^0 \pi_{j,t}^0) Y_{i,t} Y_{j,t}}{\pi_{i,t}^0 \pi_{j,t}^0 \pi_{i,j,t}^{0,0}} \right] \end{aligned} \quad (\text{B.14})$$



and

$$\begin{aligned}
\widehat{Var}^d(\hat{\tau}_t^{TE}) &= \frac{1}{n^2} \sum_{i=1}^n \left[ \mathbf{1}(H_{i,t} = h_i^1)(1 - \pi_{i,t}^1) \left( \frac{Y_{i,t}}{\pi_{i,t}^1} \right)^2 + \mathbf{1}(H_{i,t} = h_i^0)(1 - \pi_{i,t}^0) \left( \frac{Y_{i,t}}{\pi_{i,t}^0} \right)^2 \right] \\
&+ \frac{2}{n^2} \sum_{1 \leq i < j \leq n} \left[ \left( \mathbf{1}(\pi_{i,j,t}^{1,1} \neq 0) \frac{\mathbf{1}(H_{i,t} = h_i^1) \mathbf{1}(H_{j,t} = h_j^1) (\pi_{i,j,t}^{1,1} - \pi_{i,t}^1 \pi_{j,t}^1) Y_{i,t} Y_{j,t}}{\pi_{i,t}^1 \pi_{j,t}^1 \pi_{i,j,t}^{1,1}} \right. \right. \\
&\quad \left. \left. - \mathbf{1}(\pi_{i,j,t}^{1,1} = 0) \left( \frac{\mathbf{1}(H_{i,t} = h_i^1) Y_{i,t}^2}{2\pi_{i,t}^1} + \frac{\mathbf{1}(H_{j,t} = h_j^1) Y_{j,t}^2}{2\pi_{j,t}^1} \right) \right) \right. \\
&\quad \left. - \mathbf{1}(\pi_{i,j,t}^{0,1} \neq 0) \frac{\mathbf{1}(H_{i,t} = h_i^0) \mathbf{1}(H_{j,t} = h_j^1) (\pi_{i,j,t}^{0,1} - \pi_{i,t}^0 \pi_{j,t}^1) Y_{i,t} Y_{j,t}}{\pi_{i,t}^0 \pi_{j,t}^1 \pi_{i,j,t}^{0,1}} \right. \\
&\quad \left. - \mathbf{1}(\pi_{i,j,t}^{1,0} \neq 0) \frac{\mathbf{1}(H_{i,t} = h_i^1) \mathbf{1}(H_{j,t} = h_j^0) (\pi_{i,j,t}^{1,0} - \pi_{i,t}^1 \pi_{j,t}^0) Y_{i,t} Y_{j,t}}{\pi_{i,t}^1 \pi_{j,t}^0 \pi_{i,j,t}^{1,0}} \right. \\
&\quad \left. + \left( \mathbf{1}(\pi_{i,j,t}^{0,0} \neq 0) \frac{\mathbf{1}(H_{i,t} = h_i^0) \mathbf{1}(H_{j,t} = h_j^0) (\pi_{i,j,t}^{0,0} - \pi_{i,t}^0 \pi_{j,t}^0) Y_{i,t} Y_{j,t}}{\pi_{i,t}^0 \pi_{j,t}^0 \pi_{i,j,t}^{0,0}} \right. \right. \\
&\quad \left. \left. - \mathbf{1}(\pi_{i,j,t}^{0,0} = 0) \left( \frac{\mathbf{1}(H_{i,t} = h_i^0) Y_{i,t}^2}{2\pi_{i,t}^0} + \frac{\mathbf{1}(H_{j,t} = h_j^0) Y_{j,t}^2}{2\pi_{j,t}^0} \right) \right) \right]. \quad (\text{B.15})
\end{aligned}$$

Assuming all the potential outcomes are non-negative, we then have that

$$\mathbb{E} \left[ \widehat{Var}^u(\hat{\tau}_t^{TE}) \right] \geq Var(\hat{\tau}_t^{TE})$$

and

$$\mathbb{E} \left[ \widehat{Var}^d(\hat{\tau}_t^{TE}) \right] \leq Var(\hat{\tau}_t^{TE}).$$

**Proposition 10** (Estimator of the covariance). *We have the following unbiased estimator of*

$Cov(\hat{\tau}_t^{TE}, \hat{\tau}_{t'}^{TE})$ :

$$\begin{aligned} \widehat{Cov}(\hat{\tau}_t^{TE}, \hat{\tau}_{t'}^{TE}) &= \frac{1}{n^2} \sum_{i=1}^n \left[ \frac{\mathbf{1}(H_{i,t} = h_i^1) \mathbf{1}(H_{i,t'} = h_i^1) Y_{i,t} Y_{i,t'} (\pi_{i,t,t'}^{1,1} - \pi_{i,t}^1 \pi_{i,t'}^1)}{\pi_{i,t,t'}^{1,1} \pi_{i,t}^1 \pi_{i,t'}^1} \right. \\ &\quad - \frac{\mathbf{1}(H_{i,t} = h_i^0) \mathbf{1}(H_{i,t'} = h_i^1) Y_{i,t} Y_{i,t'} (\pi_{i,t,t'}^{0,1} - \pi_{i,t}^0 \pi_{i,t'}^1)}{\pi_{i,t,t'}^{0,1} \pi_{i,t}^0 \pi_{i,t'}^1} \\ &\quad - \frac{\mathbf{1}(H_{i,t} = h_i^1) \mathbf{1}(H_{i,t'} = h_i^0) Y_{i,t} Y_{i,t'} (\pi_{i,t,t'}^{1,0} - \pi_{i,t}^1 \pi_{i,t'}^0)}{\pi_{i,t,t'}^{1,0} \pi_{i,t}^1 \pi_{i,t'}^0} \\ &\quad \left. + \frac{\mathbf{1}(H_{i,t} = h_i^0) \mathbf{1}(H_{i,t'} = h_i^0) Y_{i,t} Y_{i,t'} (\pi_{i,t,t'}^{0,0} - \pi_{i,t}^0 \pi_{i,t'}^0)}{\pi_{i,t,t'}^{0,0} \pi_{i,t}^0 \pi_{i,t'}^0} \right] \\ &\quad + \frac{2}{n^2} \sum_{1 \leq i < j \leq n} \left[ \frac{\mathbf{1}(H_{i,t} = h_i^1) \mathbf{1}(H_{j,t'} = h_j^1) Y_{i,t} Y_{j,t'} (\pi_{i,t,j,t'}^{1,1} - \pi_{i,t}^1 \pi_{j,t'}^1)}{\pi_{i,t,j,t'}^{1,1} \pi_{i,t}^1 \pi_{j,t'}^1} \right. \\ &\quad - \frac{\mathbf{1}(H_{i,t} = h_i^0) \mathbf{1}(H_{j,t'} = h_j^1) Y_{i,t} Y_{j,t'} (\pi_{i,t,j,t'}^{0,1} - \pi_{i,t}^0 \pi_{j,t'}^1)}{\pi_{i,t,j,t'}^{0,1} \pi_{i,t}^0 \pi_{j,t'}^1} \\ &\quad - \frac{\mathbf{1}(H_{i,t} = h_i^1) \mathbf{1}(H_{j,t'} = h_j^0) Y_{i,t} Y_{j,t'} (\pi_{i,t,j,t'}^{1,0} - \pi_{i,t}^1 \pi_{j,t'}^0)}{\pi_{i,t,j,t'}^{1,0} \pi_{i,t}^1 \pi_{j,t'}^0} \\ &\quad \left. + \frac{\mathbf{1}(H_{i,t} = h_i^0) \mathbf{1}(H_{j,t'} = h_j^0) Y_{i,t} Y_{j,t'} (\pi_{i,t,j,t'}^{0,0} - \pi_{i,t}^0 \pi_{j,t'}^0)}{\pi_{i,t,j,t'}^{0,0} \pi_{i,t}^0 \pi_{j,t'}^0} \right] \end{aligned}$$

Proving Theorem 5 relies on results in  $m$ -dependence central limit theorem. We need the following result as a lemma.

**Lemma 5.** *Let  $\{X_{n,i}\}$  be a triangular array of mean zero random variables. For each  $n = 1, 2, \dots$  let  $d = d_n$ , and suppose  $X_{n,1}, \dots, X_{n,d}$  is an  $m$ -dependent sequence of random variables for some  $m \in \mathbb{N}$ . Define*

$$B_{n,k,a}^2 = \text{Var} \left( \sum_{i=a}^{a+k-1} X_{n,i} \right), B_n^2 = B_{n,d,1} = \text{Var} \left( \sum_{i=1}^d X_{n,i} \right).$$

Assume the following conditions hold. For some  $\delta > 0$ ,  $-1 \leq \gamma < 1$  and  $g = g_n > 2m$  is such that  $\frac{m}{g} \rightarrow 0$ :

$$\mathbb{E}|X_{n,i}|^{2+\delta} \leq \Delta_n \text{ for all } i, \quad (\text{B.16})$$

$$B_{n,k,a}^2 / (k^{1+\gamma}) \leq K_n \text{ for all } a \text{ and for all } k \geq m, \quad (\text{B.17})$$

$$B_n^2 / (dm^\gamma) \geq L_n, \quad (\text{B.18})$$

$$\frac{K_n}{L_n} \cdot \frac{m}{g} \rightarrow 0, \quad (\text{B.19})$$

$$\frac{K_n}{L_n} \cdot \left( \frac{m}{g} \right)^{(1-\gamma)/2} \rightarrow 0, \quad (\text{B.20})$$

$$\Delta_n L_n^{-(2+\delta)/2} g^{\delta/2+(1-\gamma)(2+\delta)/2} d^{-\delta/2} \left(\frac{m}{g}\right)^{(1-\gamma)(2+\delta)/2} \rightarrow 0. \quad (\text{B.21})$$

Then,  $B_n^{-1}(X_{n,1} + \dots + X_{n,d}) \xrightarrow{d} \mathcal{N}(0, 1)$ .

*Proof.* This is essentially Theorem 2.1 in Romano & Wolf (2000). We replace the original conditions 4, 5 and 6 by the last three conditions. In fact, the last three conditions are needed to establish the theorem and the conditions 4, 5 and 6 in Theorem 2.1 in Romano & Wolf (2000) are sufficient conditions.  $\square$

Now, we are ready to prove Theorem 5.

*Proof of Theorem 5.* We define  $\tilde{\tau}_{i,t} = \frac{\mathbf{1}(H_{i,t}=k)}{\mathbb{P}(H_{i,t}=k)} Y_{i,t} - \frac{\mathbf{1}(H_{i,t}=k')}{\mathbb{P}(H_{i,t}=k')} Y_{i,t} = \frac{\mathbf{1}(H_{i,t}=k)}{\mathbb{P}(H_{i,t}=k)} Y_{i,t}(k) - \frac{\mathbf{1}(H_{i,t}=k')}{\mathbb{P}(H_{i,t}=k')} Y_{i,t}(k')$ . Then the ATEC can be written as

$$\hat{\tau}^{k,k'} = \sum_{i=1}^n \sum_{t=1}^T \frac{1}{nT} \tilde{\tau}_{i,t}.$$

Similarly, we define  $\tau_{i,t} = Y_{i,t}(k) - Y_{i,t}(k')$ , which is the true individual exposure contrast. Now,

$$\sqrt{nT}(\hat{\tau}^{k,k'} - \bar{\tau}^{k,k'}) = \sum_{i=1}^n \sum_{t=1}^T \frac{1}{\sqrt{nT}} (\tilde{\tau}_{i,t} - \tau_{i,t}).$$

To proceed, we let  $X_{n,i,t} = \frac{1}{\sqrt{nT}}(\tilde{\tau}_{i,t} - \tau_{i,t})$ . We view  $\{X_{n,i,t}\}$  as a single sequence of random variables by enumerating  $X_{n,i,t}$  following the order  $X_{n,1,1}, \dots, X_{n,1,T}, X_{n,2,1}, \dots, X_{n,2,T}, \dots, X_{n,n,T}$ . Using the language in the lemma,  $d = nT$ . Since  $\{H_{i,t}\}_{i=1}^n$  is a sequence of  $s$ -dependent random variables and  $X_{n,i,t}$  is a function of  $H_{i,t}$ , we know that  $\{X_{n,i,t}\}$  is a  $sT$ -dependent sequence of random variables. In other words,  $m = sT$  in the above lemma. Note that  $|X_{n,i,t}| \leq \frac{C_1}{\sqrt{nT}}$  by uniform boundedness of potential outcomes. Hence, for any  $\delta > 0$ ,  $\Delta_n = C_2(nT)^{-1-\delta/2}$ . Now, we calculate  $B_{n,k,a}^2$  and  $B_n^2$ . We start with  $B_{n,k,a}^2$ . For all  $(i_1, t_1)$  and  $k \geq m$ , let  $(i_2, t_2)$  be the index such that when we order  $X$ 's there are exactly  $k$  indices from  $(i_1, t_1)$  to  $(i_2, t_2)$ .

$$\begin{aligned} B_{n,k,a}^2 &= \text{Var} \left( \sum_{(i,t)=(i_1,t_1)}^{(i,t)=(i_2,t_2)} X_{n,i,t} \right) \\ &= \frac{1}{nT} \text{Var} \left( \sum_{(i,t)=(i_1,t_1)}^{(i,t)=(i_2,t_2)} \tilde{\tau}_{i,t} \right) \\ &= \frac{1}{nT} \left[ \sum_{(i,t)=(i_1,t_1)}^{(i,t)=(i_2,t_2)} \text{Var}(\tilde{\tau}_{i,t}) + 2 \sum_{(u,v) \neq (p,q)} \text{Cov}(\tilde{\tau}_{u,v}, \tilde{\tau}_{p,q}) \right] \end{aligned}$$

Since  $k \geq m = sT$ , we know that at most  $mk$  covariance terms are non-zero. Given uniform boundedness of potential outcomes and overlap, all the variance and covariance terms are upper bounded by constants  $M_1 > 0$  and  $M_2 > 0$  respectively. Hence,

$$B_{n,k,a}^2 \leq \frac{1}{nT}(kM_1 + 2mkM_2) \leq M_3 \frac{mk}{nT} = M_3 \frac{sk}{n}.$$

Therefore,

$$B_{n,k,a}^2/k \leq M_3 \frac{sk}{n}/k = M_3 \frac{s}{n} = K_n.$$

Now we look at  $B_n^2$ . By Assumption 6,  $\text{Var}(\sqrt{nT}\hat{\tau}^{k,k'}) \geq \epsilon > 0$ , hence, for sufficiently large  $n$ ,

$$B_n^2 = \text{Var}(\sqrt{nT}\hat{\tau}^{k,k'}) \geq \epsilon > 0,$$

and

$$B_n^2/d = B_n^2/(nT) \geq \epsilon/(nT) = L_n.$$

We let  $\gamma = 0, \delta = 2$ . Pick  $g = g_n = s^3T^3n^\alpha$ . With such  $g$ ,  $m/g$  obviously goes to 0. Now,

$$\frac{K_n}{L_n} \cdot \frac{m}{g} = \epsilon M_3 sT \cdot \frac{1}{s^2T^2n^\alpha} \rightarrow 0,$$

$$\frac{K_n}{L_n} \cdot \left(\frac{m}{g}\right)^{(1-\gamma)/2} = \epsilon M_3 sT \cdot \frac{1}{sTn^{0.5\alpha}} \rightarrow 0,$$

$$\begin{aligned} & \Delta_n L_n^{-(2+\delta)/2} g^{\delta/2+(1-\gamma)(2+\delta)/2} d^{-\delta/2} \left(\frac{m}{g}\right)^{(1-\gamma)(2+\delta)/2} \\ &= C_2 (nT)^{-1-\delta/2} \epsilon^{-(2+\delta)/2} (nT)^{(2+\delta)/2} g^{\delta/2} (nT)^{-\delta/2} (sT)^{1+\delta/2} \\ &= C_4 g s^2 T / n \quad \text{when } \delta = 2 \\ &= C_4 s^5 T^4 n^\alpha / n. \end{aligned}$$

Since  $s^5 T^4 = o(n^{1-\alpha})$ ,  $s^5 T^4 n^\alpha = o(n)$  and hence  $\Delta_n L_n^{-(2+\delta)/2} g^{\delta/2+(1-\gamma)(2+\delta)/2} d^{-\delta/2} \left(\frac{m}{g}\right)^{(1-\gamma)(2+\delta)/2} = o(1)$ . Having checked all the conditions, by Lemma 5, we are done.  $\square$

*Proof of Theorem 6.* As in the above proof, we check the six conditions in Lemma 5 are satisfied with  $\gamma = 0$  and  $\delta = 2$ . Note that since now  $X_{n,i,t}$  and  $X_{n,j,t}$  are correlated if and only if  $i$  and  $j$  are in the same group, we can reorder  $X_{n,i,t}$ 's as follows:

$$X_{n,1,1}, \dots, X_{n,r,1}, X_{n,1,2}, \dots, X_{n,r,2}, \dots, X_{n,r,T}, X_{n,r+1,1}, \dots, X_{n,nr,T}.$$

Now, this sequence is actually  $(2r)$ -dependent, i.e.,  $m = 2r, s = r$ . Then

$$K_n = M_4/(nT), \quad L_n = \epsilon/(nrT).$$

Hence  $K_n/L_n = M_5 r$ . Pick  $g = g_n$  such that  $g \rightarrow \infty$  and  $g = (nT)^{3/4}$ . Then with  $r = o((nT)^{1/4})$ ,  $r^2/g \rightarrow 0$  and  $r^3/g \rightarrow 0$ .

$$\begin{aligned}\frac{K_n}{L_n} \cdot \frac{m}{g} &= M_5 r \cdot \frac{2r}{g} \rightarrow 0, \\ \frac{K_n}{L_n} \cdot \left(\frac{m}{g}\right)^{(1-\gamma)/2} &= M_5 r \cdot \sqrt{\frac{2r}{g}} \rightarrow 0\end{aligned}$$

and

$$\begin{aligned}\Delta_n L_n^{-(2+\delta)/2} g^{\delta/2 + (1-\gamma)(2+\delta)/2} d^{-\delta/2} \left(\frac{m}{g}\right)^{(1-\gamma)(2+\delta)/2} \\ &= M_6 g^3 (nrT)^{-1} \left(\frac{r}{g}\right)^2 \\ &= M_6 r g / (nT) \\ &= o(nT) / (nT) \rightarrow 0\end{aligned}$$

Hence all the conditions are satisfied. It is also easy to see that instead of just 2 time steps, any finite  $p$  time steps would work.  $\square$

*Proof of Proposition 6.* Let  $X_{n,t} = \sqrt{\frac{nr}{T}}(\hat{\tau}_t^{k,k'} - \tau_t^{k,k'})$ . The key ingredients are the following two expressions:

$$\begin{aligned}\text{Var}(X_{n,t}) &= \frac{1}{nrT} \left[ \sum_{l=1}^n \sum_{q=1}^r (2^{2r} - 1) Y_{(l,q),t}(k)^2 \right. \\ &\quad + \sum_{l=1}^n \sum_{q=1}^r (2^{2r} - 1) Y_{(l,q),t}(k')^2 + 2 \sum_{l=1}^n \sum_{q=1}^r Y_{(l,q),t}(k) Y_{(l,q),t}(k') \\ &\quad + \sum_{l=1}^n \sum_{q_1=1}^r \sum_{q_2 \neq q_1}^r \left( (2^{2r} - 1) Y_{(l,q_1),t}(k) Y_{(l,q_2),t}(k) + (2^{2r} - 1) Y_{(l,q_1),t}(k') Y_{(l,q_2),t}(k') \right) \\ &\quad \left. + 2 \sum_{l=1}^n \sum_{q_1=1}^r \sum_{q_2 \neq q_1}^r Y_{(l,q_1),t}(k) Y_{(l,q_2),t}(k') \right] \quad (\text{B.22})\end{aligned}$$

and

$$\begin{aligned}\text{Cov}(X_{n,t}, X_{n,t+1}) &= \frac{1}{nrT} \sum_{l=1}^n \sum_{q_1=1}^r \sum_{q_2=1}^r \\ &\quad \left( (2^r - 1) Y_{(l,q_1),t}(k) Y_{(l,q_2),t+1}(k) \right. \\ &\quad + (2^r - 1) Y_{(l,q_1),t}(k') Y_{(l,q_2),t+1}(k') \\ &\quad \left. + Y_{(l,q_1),t}(k') Y_{(l,q_2),t+1}(k) + Y_{(l,q_1),t}(k) Y_{(l,q_2),t+1}(k') \right) \quad (\text{B.23})\end{aligned}$$

We have that

$$B_n^2 = \sum_{t=1}^T \text{Var}(X_{n,t}) + 2 \sum_{t=1}^{T-1} \text{Cov}(X_{n,t}, X_{n,t+1})$$

Plugging in (B.22) and (B.23), we have the expression of  $B_n^2$ . The estimator is obtained by replacing the non-identifiable terms by corresponding upper bound.  $\square$

## C $k$ -steps convex estimator

The approach we have described in Section 3.2 naturally extends to using the  $k - 1$  previous time steps, yielding the weighted combination estimator:

$$\hat{\tau}_t^c = \alpha_1 \hat{\tau}_{t-k+1}^{TE} + \cdots + \alpha_k \hat{\tau}_t^{TE},$$

which exhibits the following absolute bias bound:

**Proposition 11** (Bound on the bias of  $\hat{\tau}_t^c$ ).

$$|\mathbb{E}[\hat{\tau}_t^c] - \tau_t^{TE}| \leq 2[(k-1)\alpha_1 + (k-2)\alpha_2 + \cdots + \alpha_{k-1}] \epsilon$$

As in the previous section, we can estimate  $\alpha_1, \dots, \alpha_k$  by solving the following convex optimization problem:

$$\begin{aligned} \arg \min_{\alpha_1, \dots, \alpha_k} \quad & \alpha_1^2 \widehat{\text{Var}}(\hat{\tau}_{t-k+1}^{TE}) + \cdots + \alpha_k^2 \widehat{\text{Var}}(\hat{\tau}_t^{TE}) \\ & + 4[(k-1)\alpha_1 + \cdots + \alpha_{k-1}]^2 \epsilon^2 \\ \text{subject to} \quad & \alpha_1 + \cdots + \alpha_k = 1, \end{aligned}$$

where  $\widehat{\text{Var}}(\hat{\tau}_{t-k+1}^{TE}), \dots, \widehat{\text{Var}}(\hat{\tau}_t^{TE})$  are estimators of the associated variance terms, and are provided in Appendix B. This then suggests the following plug-in estimator:

$$\hat{\tau}_t^c = \hat{\alpha}_1 \hat{\tau}_{t-k+1}^{TE} + \cdots + \hat{\alpha}_k \hat{\tau}_t^{TE}.$$

We can assert stronger control over the bias of  $\hat{\tau}^c$  by incorporating an additional constraint to the optimization problem:

$$\begin{aligned} \arg \min_{\alpha_1, \dots, \alpha_k} \quad & \alpha_1^2 \widehat{\text{Var}}(\hat{\tau}_{t-k+1}^{TE}) + \cdots + \alpha_k^2 \widehat{\text{Var}}(\hat{\tau}_t^{TE}) \\ & + 4[(k-1)\alpha_1 + \cdots + \alpha_{k-1}]^2 \epsilon^2 \\ \text{subject to} \quad & \alpha_1 + \cdots + \alpha_k = 1, \\ & 2[(k-1)\alpha_1 + (k-2)\alpha_2 + \cdots + \alpha_{k-1}] \epsilon \leq \delta. \end{aligned}$$

Numerical solutions for either optimization problem are straightforward to obtain using standard numerical solvers. Variance estimator and confidence interval of  $\tau_t^{TE}$  can be constructed in exactly the same way as in the case  $k = 2$ .

*Proof of Proposition 11.*

$$\begin{aligned}
& |\mathbb{E}[\hat{\tau}_t^c] - \tau_t| \\
&= |\alpha_1 \tau_{t-k+1}^{TE} + \cdots + \alpha_k \tau_t^{TE} - \tau_t^{TE}| \\
&= |\alpha_1 \tau_{t-k+1}^{TE} + \cdots + \alpha_{k-1} \tau_{t-1}^{TE} - (1 - \alpha_k) \tau_t^{TE}| \\
&= |\alpha_1 \tau_{t-k+1}^{TE} + \cdots + \alpha_{k-1} \tau_{t-1}^{TE} - (\alpha_1 + \cdots + \alpha_{k-1}) \tau_t^{TE}| \\
&= |\alpha_1 (\tau_{t-k+1}^{TE} - \tau_t^{TE}) + \cdots + \alpha_{k-1} (\tau_{t-1}^{TE} - \tau_t^{TE})| \\
&\leq \alpha_1 |\tau_{t-k+1}^{TE} - \tau_t^{TE}| + \cdots + \alpha_{k-1} |\tau_{t-1}^{TE} - \tau_t^{TE}| \\
&\leq 2\alpha_1(k-1)\epsilon + \cdots + 2\alpha_{k-1}\epsilon \\
&= 2[(k-1)\alpha_1 + \cdots + \alpha_{k-1}]\epsilon
\end{aligned}$$

□

We first give the optimization problem for the general case that assignments may be correlated across time:

$$\begin{aligned}
& \arg \min_{\alpha_1, \dots, \alpha_k} \alpha_1^2 \widehat{\text{Var}}(\hat{\tau}_{t-k+1}^{TE}) + \cdots + \alpha_k^2 \widehat{\text{Var}}(\hat{\tau}_t^{TE}) \\
& \quad + 2\alpha_i \alpha_j \sum_{1 \leq i < j \leq k} \widehat{\text{Cov}}(\hat{\tau}_{t-k+i}^{TE}, \hat{\tau}_{t-k+j}^{TE}) \\
& \quad + 4[(k-1)\alpha_1 + \cdots + \alpha_{k-1}]^2 \epsilon^2 \\
& \text{subject to } \alpha_1 + \cdots + \alpha_k = 1,
\end{aligned}$$

where  $\widehat{\text{Var}}(\hat{\tau}_{t-k+1}^{TE}), \dots, \widehat{\text{Var}}(\hat{\tau}_t^{TE})$  and  $\widehat{\text{Cov}}(\hat{\tau}_{t-k+i}^{TE}, \hat{\tau}_{t-k+j}^{TE})$  can be any estimator in Proposition 9 and 10. Moreover, suppose that the assignments are independent across time, we know that  $\text{Cov}(\hat{\tau}_{t-k+i}^{TE}, \hat{\tau}_{t-k+j}^{TE}) = 0$ , hence we have an even simpler optimization problem as stated in the main text.

*Derivation of the optimization problem.* We first calculate the variance.

$$\begin{aligned}
\text{Var}(\hat{\tau}_t^c) &= \text{Var}(\alpha_1 \hat{\tau}_{t-k+1}^{TE} + \cdots + \alpha_k \hat{\tau}_t^{TE}) \\
&= \alpha_1^2 \text{Var}(\hat{\tau}_{t-k+1}^{TE}) + \cdots + \alpha_k^2 \text{Var}(\hat{\tau}_t^{TE}) \\
& \quad + 2\alpha_i \alpha_j \sum_{1 \leq i < j \leq k} \text{Cov}(\hat{\tau}_{t-k+i}^{TE}, \hat{\tau}_{t-k+j}^{TE})
\end{aligned}$$

Suppose we want to have smaller MSE by using  $\hat{\tau}_t^c$ , we need to have

$$\text{Var}(\hat{\tau}_t^c) + |\mathbb{E}[\hat{\tau}_t^c] - \tau_t^{TE}|^2 \leq \text{Var}(\hat{\tau}_t^{TE})$$

By Proposition 11, it suffices to have

$$\begin{aligned}
& \alpha_1^2 \text{Var}(\hat{\tau}_{t-k+1}^{TE}) + \cdots + \alpha_k^2 \text{Var}(\hat{\tau}_t^{TE}) \\
& + 2\alpha_i \alpha_j \sum_{1 \leq i < j \leq k} \text{Cov}(\hat{\tau}_{t-k+i}^{TE}, \hat{\tau}_{t-k+j}^{TE}) \\
& + 4[(k-1)\alpha_1 + \cdots + \alpha_{k-1}]^2 \epsilon^2 \leq \text{Var}(\hat{\tau}_t^{TE})
\end{aligned} \tag{C.1}$$

Now, the left hand side of (C.1) is convex in  $\alpha_1, \dots, \alpha_k$ . □



Estimate of $\epsilon$	$\hat{\epsilon}$	$1.5\hat{\epsilon}$	$2\hat{\epsilon}$	$2.5\hat{\epsilon}$	$3\hat{\epsilon}$
RMSE for $\hat{\tau}_{20}^{TE}$	33.27	33.27	33.27	33.27	33.27
RMSE for $\hat{\tau}_{20}^c, k = 2$	8.93	8.81	8.69	8.55	8.42
RMSE for $\hat{\tau}_{20}^c, k = 5$	5.12	6.20	7.11	7.91	8.64

Table 7: Root mean squared errors (RMSE) for  $\hat{\tau}_{20}^{TE}$ ,  $\hat{\tau}_{20}^c$  with  $k = 2$  and  $\hat{\tau}_{20}^c$  with  $k = 5$

## D Additional simulation results

### D.1 Simulation results for estimation under stability assumption

#### D.1.1 Parameters for Erdős-Rényi Model

For the simulation study in Section 5.2.1, we use  $p = 0.1$  for  $n = 50$  and then scale the probability  $p$  accordingly for larger  $n$  so that each unit has the same expected number of neighbors.

#### D.1.2 The effect of estimated stability parameter

Recall that our  $\hat{\epsilon}$  is only a lower bound of the true  $\epsilon$ , hence may underestimate  $\epsilon$ . To investigate how our estimate of  $\epsilon$  affects the results, we fix  $n = 50$  and generate the social network according to Erdős-Rényi Model with  $p = 0.1$ . We generate 500 realizations of assignments and plug in  $\hat{\epsilon}$ ,  $1.5\hat{\epsilon}$ ,  $2\hat{\epsilon}$ ,  $2.5\hat{\epsilon}$  and  $3\hat{\epsilon}$  for three kinds of estimators considered above. Table 7 shows the results. We see that the convex combination type estimator with  $k = 2$  is not sensitive to the estimate of  $\epsilon$  while the convex combination type estimator with  $k = 5$  is. Even we use  $3\hat{\epsilon}$ , two convex combination type estimators still show better performance in terms of root mean squared error.

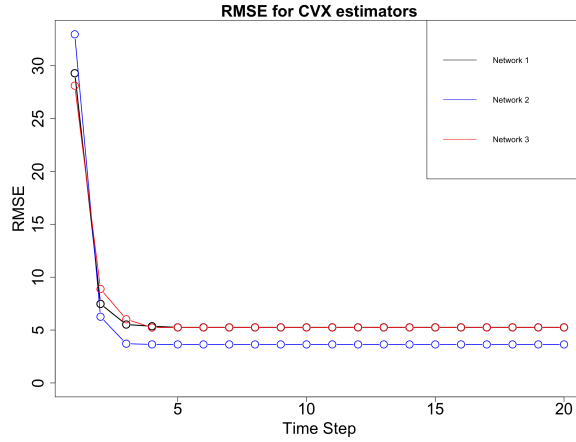


Figure 5: Root mean squared errors (RMSE) for  $\hat{\tau}_{20}^{TE}$  and  $\hat{\tau}_{20}^c$

Confidence Interval	Network 1	Network 2	Network 3
Gaussian CI with variance estimated by $\widehat{\text{Var}}^d$	27.38	26.62	27.02
Gaussian CI with variance estimated by $\widehat{\text{Var}}^u$	34.04	32.34	33.33
Chebyshev CI with variance estimated by $\widehat{\text{Var}}^d$	62.47	60.75	61.66
Chebyshev CI with variance estimated by $\widehat{\text{Var}}^u$	77.67	73.79	76.04

Table 8: Lengths of two approximate confidence intervals for  $\tau_t^{TE}$  with  $k = 2$

### D.1.3 The effect of the number of time steps

Finally, we investigate how  $k$  affects the results. We generate three different social networks, and for each one, we plot the root mean squared errors of using 1 time step (i.e., the Horvitz-Thompson type estimator) to 20 time steps (i.e., we use all time steps to estimate the total effect at time step 20). From Figure 5 we can see that the RMSE curves stay flat after a certain value of  $k$ . Hence, we do not need to worry about using too many time steps as the optimization problem intrinsically pick the right  $k$ .

### D.1.4 Lengths of approximate confidence intervals

Table 8 shows the average lengths of approximate confidence intervals. As expected, Gaussian confidence intervals are shorter.