HARVARD | BUSINESS | SCHOOL

Population Interference in Panel Experiments

Kevin Wu Han
Iavor Bojinov
Guillaume Basse

Working Paper 21-100

Population Interference in Panel Experiments

Kevin Wu Han
Stanford University

Iavor Bojinov
Harvard Business School

Guillaume Basse
Stanford University

Working Paper 21-100

Funding for this research was provided in part by Harvard Business School.

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

March 1, 2021

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. Much 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. Revisiting the special case of standard experiments with population interference, we prove a central limit theorem under weaker conditions than previous results in the literature and highlight the trade-off between flexibility in the design and the interference structure.

Keywords: Finite Population, 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 to assignment; this is usually referred to as no interference (Cox, 1958, Chapter 2). In panel experiments, where units are exposed to different interventions over time, 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[1] makes causal inference very challenging (Basse and Airoldi, 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 and Raudenbush (2006); Rosenbaum (2007)), economics (Sobel (2006); Manski (2013)) and public health (Halloran and Struchiner (1995)). Inference in this setting is challenging because interference increases the number of potential outcomes and makes observations dependent. Aronow and Samii (2017) introduce a general framework for studying causal inference with interference: they introduce the concept of exposure mapping, defined useful estimands, and construct asymptotically valid confidence

[1] We use the term *population interference* to emphasize that the interference occurred across units.

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 and Shephard (2019) allows for arbitrary carryover effects when studying whether algorithms or humans are better at executing large financial trades. 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 delayed causal effects. The latter is particularly important for technology companies seeking to understand the long-term impact of their interventions (Basse et al., 2019; Hohnhold et al., 2015). Researchers use analogous Horvitz-Thompson type estimator estimators to analyze experiments with carryover effects.

In this article, we introduce a unifying framework for studying panel experiments with population interference. Our framework allows us to consider the following settings in order of increasing generality. First, we revisit the standard population interference setting (Aronow and Samii, 2017) as a straightforward special case of our framework (Section 3); we state and prove a stronger central limit theorem than that of Aronow and Samii (2017) and provide insights on a fundamental trade-off between the interference structure and the design of the experiment. We then consider panel experiments with population interference but no carryover effects (Section 4). We provide asymptotically valid confidence intervals for estimands defined at specific time periods, as well as estimands that average contrasts over multiple time periods. We also introduce a novel class of assumptions that enable us to leverage past data to improve inference at a given time point. Finally, we tackle the most general setting featuring both population and temporal interference (Section 5): we state asymptotic results for a restricted type of mixed interference and provide a blueprint for deriving more general results. Section 2 introduces our extension of the potential outcomes framework in detail. We conclude with simulations and a discussion.

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 $t \in \{1, \cdots, T\}$, unit $i \in \{1, \cdots, n\}$ 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 $W_{i,1:t} = (W_{i,1}, W_{i,2}, \cdots, W_{i,t})$ 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 $\mathbb{P}(W_{1:n,1:T})$. 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 (see Kempthorne (1955) and Abadie et al. (2020) for extended discussions). 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 relaxing 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 and Shephard, 2019). This assumption drastically reduces the total number of potential outcomes[2] and will be implicitly made throughout this paper.

Unfortunately, inference is still impossible without any assumptions on the population interference structure (Basse and Airoldi, 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 $f_{i,t} : \{0,1\}^{n\times t} \to \Delta$, where Δ is the set of possible exposures[3] (Aronow and Samii, 2017).

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 $i \in \{1, \cdots, n\}$ and $t \in \{1, \cdots, T\}$, and any two assignment matrices $w_{1:n,1:t}$ and $w^{'}_{1:n,1:t}$,*

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

For $p \in \{1, \cdots, T\}$, we say the exposure mappings are p-time-invariant if for any $t, t^{'} \in \{p, \cdots, T\}$ 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 $p \in \{1, \cdots, T\}$. We say the exposure mappings are properly specified time-invariant exposure mappings if they are both properly specified and time-invariant.

[2]The assumption, known as non-anticipating potential outcomes (Bojinov and 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.

[3]To make exposure mappings useful, we assume the cardinality of Δ is (substantially) smaller than $n \times t$.

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. 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 $\mathcal{N}_i$. 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 1 (Locally Effective Assignments (LEA))**.** *We say the assignments and exposure mappings are locally effective if the exposure mappings are p-time-invariant for some $p \in \{1, \cdots, T\}$ 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. (2020). 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, we 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* $u_{i,t} = \frac{1}{|\mathcal{N}_i|}\sum_{j\in\mathcal{N}_i} w_{j,t}$*. Hence, one unit's assignment and the fraction of treated neighbors at previous time step matter as well. This is a special case of LEA with* $p = 2$*. Notice that 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* $f_{i,t}(w_{1:n,1:t}) = f_{i,t'}(w_{1:n,1:t'})$*.*

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 and Rubin, 2015). As the number of possible contrasts grows exponentially with the number of distinct potential outcomes, we will 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 as opposed to control.

Definition 2 (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 and 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.

In panel experiments, researchers are often less interested in the idiosyncratic effects at each point in time and instead focus on the temporal average total effect that captures the intervention's average impact across both time and units[4].

[4]For example, Bojinov and Shephard (2019) are not interested in the relative difference between an

Definition 3 (Average total effect)**.** *The average total effect is*

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

The second—which requires Assumption 1—provides a much richer class of causal effects with important practical applications.

Definition 4 (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*

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

The TEC estimand is the generalization of the usual exposure contrast estimands (Aronow and Samii, 2017) to the panel experiment setting. Hereafter, the letter k will always represent values in Δ. Similar to the total effect, in many applications, we are interested in the TEC's temporal average.

Definition 5 (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*

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

Remark 1. *Without assuming that the exposure mappings are time-invariant, the definition of the ATEC becomes more cumbersome because 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.*

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

algorithm or a human executing a large financial order on an arbitrary day of the experiment but is interested in the average difference across multiple trades on the same market.

$\mathbb{P}(W_{1:n,1:t})$ induces a distribution for the exposures $\mathbb{P}(H_{i,t})$ for each i and t. Under Assumption 1, the observed outcomes $Y_{i,t}$ for unit i at time t can therefore be written:

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

We will use the 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 and Thompson, 1952), or variations of it; e.g. to estimate $\tau_t^{k,k'}$, we will use:

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

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

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

Similarly, if we let $f_{i,t}(1_{1:n,1:t}) = h_{i,t}^1$ and $f_{i,t}(0_{1:n,1:t}) = h_{i,t}^0$, then we can estimate total effect at time t (c.f. Definition 2) by the following estimator

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

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

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

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

2.4.3 Randomization-based inference

As mentioned earlier, we adopt the randomization-based framework— that is, we consider the potential outcomes as fixed, with the assignment being the only source of randomness.

This framework has seen a recent uptake in causal inference (Lin, 2013; Li and Ding, 2017; Li et al., 2019) and has become the standard for analyzing experiments with population interference (Aronow and Samii, 2017; Sävje et al., 2017; Basse and Feller, 2018; Chin, 2018) and unbounded carryover effects (Bojinov and Shephard, 2019; Rambachan and Shephard, 2019; Bojinov et al., 2020; Bojinov et al., 2020).

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 et al., 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. 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 and Ding, 2017; Bojinov et al., 2020). 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 and 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 $\tau_t^{k,k'}$, $\bar{\tau}^{k,k'}$, τ_t^{TE} and $\bar{\tau}^{TE}$, respectively[5]; 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 Standard population interference

This section 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

[5]For example, see Bojinov and Shephard (2019) and Aronow and Samii (2017) for explicit proof.

is now equivalent to the one studied in Liu and Hudgens (2014), Aronow and Samii (2017), Chin (2018) and Leung (2019). Our Horvitz-Thompson type estimator $\hat{\tau}^{k,k'}$ now simplifies to,

$$\hat{\tau}^{k,k'} = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\mathbf{1}(H_i = k)}{\pi_i(k)}Y_i(k) - \frac{\mathbf{1}(H_i = k')}{\pi_i(k')}Y_i(k')\right], \tag{5}$$

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

Aronow and 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,

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

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.

3.1 A central limit theorem

Our central limit theorem requires four 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.,* $|Y_i(k)| \leq M$ *for some* M *and for all* i *and* k.

Assumption 3 (Overlap)**.** *Assume all the exposure probabilities are bounded away from 0 and 1, i.e.,* $\exists \eta > 0$ *such that* $\forall k$ *and* i, $0 < \eta \leq \pi_i(k) \leq 1 - \eta < 1$.

Assumptions 2 and 3 are standard in the causal inference literature (Aronow and Samii (2017); Leung (2019)). Assumption 2 holds in most practical applications as 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* $\liminf_{n\to\infty} Var(\sqrt{n}\hat{\tau}_t^{k,k'}) > 0$ *for any* t*.*

As a consequence of this assumption, for each t, there exists a constant $c > 0$ such that $\mathrm{Var}(\sqrt{n}\hat{\tau}_t^{k,k'}) \geq c$ for all sufficiently large n. This type of assumption seems unavoidable, even in settings without interference (see, e.g., Corollary 1 in Guo and Basse (2020), and subsequent discussion).

The fourth assumption quantifies the dependence among observations due to interference; to define it, we require a notion of the dependency graph for a collection of random variables. We define the dependency graph G_n for $H_1, \cdots, H_n$ to be the graph with vertices $V = \{1, \cdots, n\}$ and edges E such that $(i, j) \in E$ if and only if H_i and H_j are not independent. The graph G_n models the dependency relationship among n random variables $H_1, \cdots, H_n$. Let d_n be the maximal degree in this graph, which is equal to the maximal number of dependent exposure values for each unit. Notice that the dependency graph depends both on the interference structure and on the assignment mechanism.

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

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

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

as $n \to \infty$*.*

Theorem 1 strengthens the result of Aronow and Samii (2017) in two ways. First, our Assumption 4 weakens Condition 6 of Aronow and Samii (2017), which requires the convergence of $\mathrm{Var}(\sqrt{n}\hat{\tau}_t^{k,k'})$. Second, we allow for a higher range of dependence ($d_n = o(n^{1/4})$

compared to $d_n = O(1)$ as in Aronow and Samii (2017)) among exposure values. The proof of this theorem relies on recent results in Chin (2018).

3.2 Design and interference structure: a trade-off

Intuitively, Theorem 1 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 1, applied to popular settings.

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 $f_i(W_{1:n}) = f_i(W_{\mathcal{N}_i})$; that is, only the neighbors' assignments matter. Let δ_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 1.*

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 $C_1, \cdots, C_{n_c}$. 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 δ_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 and Feller (2018); Duflo and 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 1 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 δ_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 1: Trade-off between design and interference

3.3 Inference

The central limit theorem stated in Theorem 1 serves as our basis for inference. Unfortunately, as is typical in finite population causal inference, our estimator's variance contains terms that are products of potential outcomes that can never be simultaneously observed from a single experiment, making it non-identifiable. 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 (Aronow and Samii, 2017), we denote the estimate of the bound[6] by $\widehat{\mathrm{Var}}(\sqrt{n}\hat{\tau}^{k,k'})$.

Proposition 1. *Assuming all the assumptions in Theorem 1, then for any $\delta > 0$, we have that*

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

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

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

for large n.

Once again, this result strengthens that of Aronow and Samii (2017) by both removing the requirement that $n\mathrm{Var}(\hat{\tau}^{k,k'})$ converge, and by relaxing the constraint on the interference mechanism. Note that here $\delta > 0$ is arbitrary and we present detailed simulations in Section 6 with $\delta = 0.04$.

[6]The explicit form of the estimator is given in the Supplementary Material.

4 Panel experiments with population interference and no carryover effects

Depending on their structure and on the researcher's goals, panel experiments with multiple treatment periods may be a blessing or a curse. Suppose the temporal dimension does not interact with the interference mechanism, which occurs when there is only purely population interference. In that case, the inference is equivalent to the standard experimental setup (Section 3), or it may benefit from the additional information if we are willing to consider a different estimand (Section 4.1) or additional assumptions (Section 4.2). In contrast, the presence of population spillovers in addition to the spatial spillovers from the previous section — a setting we call "mixed interference" — significantly compromises our ability to draw inference (see Section 5).

We work exclusively with temporally independent assignment mechanisms in this section, i.e., $W_{1:n,t}$ and $W_{1:n,t'}$ are independent for any t and $t^{'}$.

4.1 Average temporal exposure contrast

In Section 3, we showed that under pure population interference, inference on the TEC at time t could strictly be reduced to the cross-sectional setting, where the only relevant asymptotic regime takes $n \to \infty$. When considering the ATEC and its natural estimator $\hat{\bar{\tau}}^{k,k'}$, however, the asymptotic picture changes and we may now consider, broadly speaking, three regimes: (1) T fixed and $n \to \infty$; (2) $T \to \infty$ and $n \to \infty$; (3) $T \to \infty$ and n fixed. An important insight that we will emphasize in this section is that inference in these three regimes requires different constraints on the population interference mechanism. Roughly speaking, the larger T is relative to n, the more interference we can tolerate.

To make this more formal, denote by $d_n^{(t)}$ the maximal number of dependent exposure values for any unit i at time step t. Let $d_n = \limsup_{t\to\infty} d_n^{(t)}$ with the convention that for fixed T, $d_n = \max\{d_n^{(1)}, \cdots, d_n^{(T)}\}$. Hence, d_n in this section is different from d_n in the previous section in the sense that it is a bound on all time steps. Our first result establishes a central limit theorem in the fixed T regime.

Theorem 2. *Suppose we have pure population interference, then for any T, under Assumptions 2-4 and the condition that $d_n = o(n^{1/4})$, we have*

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

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

This first theorem states a central limit theorem for the regime where T is fixed and $n \to \infty$, making it relevant for applications where N is much larger than T. Notice that, like Theorem 1, it requires $d_n = o(n^{1/4})$. Intuitively, this is because this asymptotic regime is closest to that of the previous section: any finite number of time periods T is negligible compared with infinitely many observations n.

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

Theorem 3. *Suppose we have pure population interference and Assumptions 2, 3 are satisfied. Let $\sigma_{n,t}^2 = Var(\sqrt{n}\hat{\tau}_t^{k,k'})$, 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*

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

as $T \to \infty$.

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 point 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 4. *Suppose we have pure population interference and Assumptions 2-4 are satisfied, then for $T = T(n)$ such that either*

$$\frac{n}{T} \to 0 \tag{6}$$

or

$$\frac{\min\{d_n^2, n\}}{\sqrt{nT}} \to 0 \tag{7}$$

holds, we have

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

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

Condition (6) is actually a special case of condition (7): if we do not impose any assumptions on the interference, $\min\{d_n^2, n\}$ is just n, so we need $\frac{n}{\sqrt{nT}} \to 0$, which is equivalent to $\frac{n}{T} \to 0$. Condition 7 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 $T = T(n)$.

With the above central limit theorems, inference proceeds as in Section 3.3 (see Proposition 5 in the supplement).

4.2 Stability assumption

In Section 3 and Section 4.1, we considered inference on the TEC and ATEC under population interference. In this section, we focus on the following question: can we leverage the temporal information to improve inference on the TEC? Our results here are weaker than in the previous section — indeed, we provide neither central limit theorem nor asymptotic confidence interval —, but we believe they are an exciting avenue for future work.

So far, we have considered only Horvitz-Thompson estimators which, while analytically tractable, are known to have large variance when exposure probabilities are small. The key idea of this section is that if the potential outcomes do not vary too much across time, then estimates of the TEC at time $t' < t$ can be used to improve our estimate of the TEC at time t. This assumption can be formalized as follows:

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

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,

$$\tau_t^{TE} = \frac{1}{n}\sum_{i=1}^{n} Y_{i,t}(h_i^1) - \frac{1}{n}\sum_{i=1}^{n} Y_{i,t}(h_i^0), \tag{8}$$

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 investigate how to leverage just 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 $\hat{\tau}_t^{TE}$, 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 2 (Bound on the bias of $\hat{\tau}_t^c$)**.**

$$|\mathbb{E}[\hat{\tau}_t^c] - \tau_t^{TE}| \leq 2(1-\alpha)\epsilon \tag{9}$$

As we can see, the absolute bias of $\hat{\tau}_t^c$ is bounded by a quantity that grows linearly with ϵ: if ϵ is very small, then so will the maximum bias. In particular, $\hat{\tau}_t^c$ is unbiased for τ_t^{TE}

if $\epsilon = 0$ — which corresponds to the 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 3. *Suppose that the assignments are independent across time, 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 α is given by $\alpha = 1 - \frac{Var(\hat{\tau}_t^{TE})}{4\epsilon^2 + Var(\hat{\tau}_t^{TE}) + Var(\hat{\tau}_{t-1}^{TE})}$.*

We show in the simulation section that the reduction in mean squared error is significant when n is small. In the Supplementary Material, we extend this proposition to the setting with temporally dependent assignments.

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

Algorithm 1 Algorithm to estimate ϵ

1: Initialize $\hat{\epsilon} = 0$

2: For $t = 1$ to $T-1$:

(a) For $i = 1, 2, \cdots, n$ 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 $\epsilon_{i,t} > \hat{\epsilon}$, update $\hat{\epsilon}$ by $\hat{\epsilon} = \epsilon_{i,t}$.

3: Output $\hat{\epsilon}$.

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

where $\widehat{\mathrm{Var}}(\hat{\tau}_t^{TE})$ can be any estimator of the variance $\mathrm{Var}(\hat{\tau}_t^{TE})$: we discuss a few options in in Proposition 10 of the Supplementary Material. In addition, under pure population interference and temporally independent assignments,

$$\mathrm{Var}(\hat{\tau}_t^c) = \mathrm{Var}\bigg(\alpha\hat{\tau}_t^{TE} + (1-\alpha)\hat{\tau}_{t-1}^{TE}\bigg) = \alpha^2\mathrm{Var}(\hat{\tau}_t^{TE}) + (1-\alpha)^2\mathrm{Var}(\hat{\tau}_{t-1}^{TE}),$$

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

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

As mentioned in the introduction to this section, we do not have formal inferential results at the moment — this is an open area for future work. However, based on the variance estimator above, we do have two crude 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. Specifically, note that

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

hence $\forall \delta > 0$, $\mathbb{P}\left(\tau_t^{TE} \in \left[\hat{\tau}_t^c - (\mathbb{E}[\hat{\tau}_t^c] - \tau_t^{TE}) - \epsilon, \hat{\tau}_t^c - (\mathbb{E}[\hat{\tau}_t^c] - \tau_t^{TE}) + \epsilon\right]\right) \geq 1 - \delta$ for $\epsilon = \sqrt{\frac{\mathrm{Var}(\hat{\tau}_t^c)}{\delta}}$. 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 τ_t^{TE}:

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

We explore empirically the coverage of the above approximate confidence intervals with a simulation study in Section 6.

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

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

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

5 Panel experiments with population interference and carryover effects

Section 4.1 shows that adding a temporal dimension does not hurt inference and may even help if interference remains confined to the spatial dimension. Mixed interference, in contrast, 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 3 holds (recall that d_n is the maximal degree of the dependency graph of $H_1, \cdots, H_n$):

Theorem 5. *Under Assumptions 2-4 and the condition that $d_n = o(n^{1/4})$, we have*

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

The difference with the pure population setting 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 4.1. Stating general theorems is difficult in this setting for the reasons we will discuss in Remark 2 below. Instead, we focus on a specific setting, to illustrate the type of results that can be derived under mixed interference, and emphasize the fact that strong assumptions are required even in a relatively simple scenario.

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

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

where $\mathcal{N}_i$ is the group to which unit i belongs. For convenience, we fix each group to be of size r and focus on the Bernoulli design with $p = \frac{1}{2}$. We consider the exposures $k = (1, 1, r-1, r-1)$ and $k' = (0, 0, 0, 0)$. To ease notations, from now on we index each unit i by a tuple (l, q), meaning that unit i is the q-th unit in the l-th group. We consider n households and $T = T(n)$ time periods. As in the cross-sectional setting, we impose a condition that controls the rate at which the variance shrinks:

Assumption 6. *Assume that* $\liminf_{n\to\infty} Var(\sqrt{nrT}\hat{\bar{\tau}}^{k,k'}) \geq \epsilon > 0$ *for some* ϵ.

This technical assumption (we are generally more worried about the variance not shrinking fast enough) rules out the pathological case that the variance vanishes as $n \to \infty$ [7].

Theorem 6. *With the above setting and under Assumption 6, suppose* T *is such that* $T = O(n^{1+\epsilon})$ *for some* $\epsilon > 0$*, then we have that*

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

as $n \to \infty$.

Here, ϵ can be arbitrarily small but $T = O(n)$ is not sufficient. Similar results can be obtained for general p and general exposure contrasts by doing a more tedious calculation of the variance.

Remark 2. *The difficulty for proving a central limit theorem under more general setting comes from the fact that it is hard to control the variance. If we want to prove a more general version, we need to further make more non-trivial assumptions. For example, since there is dependence across time, the variance of* $\hat{\bar{\tau}}^{k,k'}$ *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 control the variance.*

6 Simulations

In this section, we use simulations to supplement some of our theoretical results, and to provide empirical guidance when theory is lacking. Section 6.1 explores some of the finite sample properties of our central limit theorems in different realistic settings. Section 6.2 explores empirically some properties of the convex combination estimator proposed in Section 4.2: in particular, although we do not prove central limit theorems for this estimator, we show

[7] 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 the assumption.

that confidence intervals based on normal approximations behave well in our simulation, and could therefore be reasonable candidates for practical use.

6.1 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 and Saez (2003); Basse and 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 $f_i(w_{1:n}) = (w_i, u_i)$, where $u_i = 1$ if unit i has at least one treated neighbor and $u_i = 0$ 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.

6.1.1 Household interference in the cross-sectional setting

We first focus on the cross-sectional setting. We consider groups of relatively small sizes (betwen 2 and 4) and, since the finite sample behavior of our results depends largely on the full schedule of potential outcomes, we report simulations based on three different data generating processes: (1) Normal potential outcomes $Y_i(w_i, u_i) \sim \mathcal{N}(3w_i + 2u_i + 5, 1)$, (2) Poisson potential outcomes $Y_i(w_i, u_i) \sim \text{Pois}(3w_i + 2u_i + 5)$ and (3) Bernoulli potential outcomes $Y_i(w_i, u_i) \sim \text{Bern}\left(\frac{3w_i+2u_i+2}{18}\right)$

The conditions of Theorem 1 apply in this setting so we expect, for instance, that our confidence intervals will be conservative for large values of n. The question we seek to answer empirically is "how large" n needs to be for our inference to be reliable? We simulate populations with sizes ranging from $n = 100$ to $n = 5000$ and, for each size n, we draw a realization of the potential outcomes from one of the data generating processes, then draw $50,000$ assignments from the design and compute $\hat{\tau}$ for each.

Sample Size	n = 100	n = 250	n = 500	n = 1000	n = 1500	n = 2000	n = 3000	n = 5000
p-value	1.04e-30	2.53e-20	1.02e-12	1.21e-07	6.95e-06	8.88e-05	0.014	0.210%
Coverage	93.6%	92.7%	95.3%	96.2%	96.6%	96.9%	97.0%	97.2%

Table 2: Simulation results for Theorem 1, Bernoulli distributed potential outcomes

We find that although the quality of the normal approximation (as measured by the Shapiro-Wilk test and Q-Q plots) for a given sample size n varies significantly between the three data generating processes, the coverage of our 95% confidence intervals (constructed using $\delta = 0.04$) is excellent in all three cases, even for small values of n. Table 2 reports the results for the Bernoulli data generating process, which is the most adversarial. The results for the Gaussian and Poisson data generating processes are in the Supplementary Materials. We see that even the adversarial case of the Bernoulli data generating process, coverage is close to nominal even for $n = 100$, and is actually above the nominal level for $n \geq 500$ households— a sample size far smaller than the 3876 households studied in Basse and Feller (2018).

6.1.2 Panel setting

Next, we turn our attention to the central limit theorems for ATEC. Theorem 4 establishes asymptotic results for ATEC under less constraining assumptions on the interference mechanism than or TEC. To illustrate this point, we still consider the stratified interference setting. Now instead of group sizes between 2 and 4, 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 4. 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 $f_i(w_{1:n}) = (w_i, u_i)$ where $u_i = 0$ if less than 25% of the neighbors are treated; $u_i = 1$ if between 25% and 50% of the neighbors are treated; $u_i = 2$ if between 50% and 75% of the neighbors are treated and $u_i = 3$ if more than 75% of the neighbors are treated. We generate the potential outcomes for unit i at time step t according to $\mathcal{N}(3w_i + 2u_i + 5 + \epsilon_t, 1)$, where ϵ_t is uniform$\{-1, 1\}$. Figure 1 and 2 show that

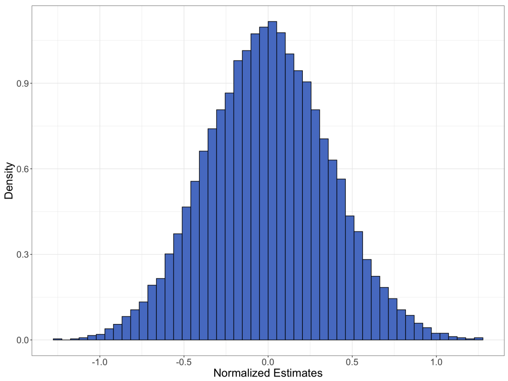


Figure 1: Histogram, $n = 1000$

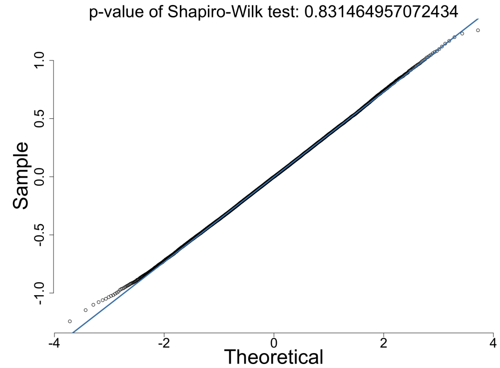


Figure 2: Q-Q normal plot, $n = 1000$

$n = 1000$ suffices for a good approximation. Moreover, the coverage of our 95% confidence interval is 95.4%.

6.2 Estimation under the stability assumption

In Section 4.2, 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. We also show that although not supported by theoretical results, naively constructed confidence intervals perform well in our simulated setting.

6.2.1 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:

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

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 $\mathcal{N}(10, 1)$ and 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 ϵ-stability holds. Throughout

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

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 τ_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 ϵ using Algorithm 1 described in Section 4.2; we use Proposition 3 to estimate α when $k = 2$, and solve the optimization problem introduced in Section B in the Supplementary Material for $k \geq 3$.

We first fix ϵ to be 3 and vary the sample size. To make each unit have the same number of 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 3 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.

6.2.2 Coverage of two approximate confidence intervals

Recall that in Section 4.2 we gave two approximate confidence intervals of τ_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 ϵ to be 3 and generate the data in the same way as in the previous section. To calculate the coverage, we

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

Table 4: Coverage of two approximate confidence intervals for τ_t^{TE} with $k = 2$

generate 1000 realizations of the assignments and construct approximate confidence intervals accordingly. Table 4 shows the coverage of the two approximate confidence intervals for three different social networks. We see that they all provide reasonable coverage. And even if we ignore the bias of Gaussian confidence interval and it is much shorter than the Chebyshev one, it actually provides better coverage. In the supplement, we also provide an additional table showing the average lengths of the confidence intervals in Table 4.

7 Conclusion

In this paper, we have developed estimation and inference results for panel experiments with population interference. In the standard setting with pure population interference, we prove a central limit theorem under weaker conditions than previous results in the existing literature and highlight the trade-off between flexibility in the design and the interference structure. When population interference and carryover effects co-exist, we propose a novel central limit theorem. Finally, we introduce a new type of assumptions —stability assumptions — as an alternative to (or complement of) exposure mappings for controlling interference in temporal settings.

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 only consider the Bernoulli design: this is, of course, limiting, but it does present a useful benchmark. We are particularly interested in exploring how to design panel experiments in the presence of population interference and carryover effects. Basse et al. (2019) study minimax designs with carryover effects, but the symmetries they exploit break under population interference, so new approaches are required. 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. In particular, we plan to study the asymptotic properties of these estimators and provide a firmer theoretical grounding for their inferential properties.

References

Abadie, A., S. Athey, G. W. Imbens, and J. M. Wooldridge (2020). Sampling-based versus design-based uncertainty in regression analysis. *Econometrica 88*(1), 265–296.

Aronow, P. M. and C. Samii (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., D. Eckles, and G. W. Imbens (2018). Exact p-values for network interference. *Journal of the American Statistical Association 113*(521), 230–240.

Basse, G., Y. Ding, and P. Toulis (2019). Minimax designs for causal effects in temporal experiments with treatment habituation. *arXiv preprint arXiv:1908.03531*.

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

Basse, G., A. Feller, and P. Toulis (2019). Randomization tests of causal effects under interference. *Biometrika 106*(2), 487–494.

Basse, G. W. and E. M. Airoldi (2018). Limitations of design-based causal inference and a/b testing under arbitrary and network interference. *Sociological Methodology 48*(1), 136–151.

Bojinov, I., A. Rambachan, and N. Shephard (2020). Panel Experiments and Dynamic Causal Effects: A Finite Population Perspective. *arXiv e-prints*, arXiv:2003.09915.

Bojinov, I. and N. Shephard (2019). Time series experiments and causal estimands: Exact randomization tests and trading. *Journal of the American Statistical Association 114*(528), 1665–1682.

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

Chin, A. (2018). Central limit theorems via Stein's method for randomized experiments under interference. *arXiv e-prints*, arXiv:1804.03105.

Cox, D. R. (1958). *Planning of experiments.* New York, Wiley.

Duflo, E. and E. Saez (2003, 08). 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.

Guo, K. and G. Basse (2020). The generalized oaxaca-blinder estimator. *arXiv preprint arXiv:2004.11615*.

Halloran, M. E. and C. J. Struchiner (1995). Causal inference in infectious diseases. *Epidemiology 6*(2), 142–151.

Hohnhold, H., D. O'Brien, and D. Tang (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. and S. W. Raudenbush (2006). Evaluating kindergarten retention policy. *Journal of the American Statistical Association 101*(475), 901–910.

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

Hudgens, M. G. and M. E. Halloran (2008). Toward causal inference with interference. *Journal of the American Statistical Association 103*(482), 832–842. PMID: 19081744.

Imbens, G. W. and D. B. Rubin (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Kempthorne, O. (1955). The randomization theory of experimental inference. *Journal of the American Statistical Association 50*(271), 946–967.

Leung, M. P. (2019). Causal Inference Under Approximate Neighborhood Interference. *arXiv e-prints*, arXiv:1911.07085.

Li, X. and P. Ding (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., P. Ding, Q. Lin, D. Yang, and J. S. Liu (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. and M. G. Hudgens (2014). Large sample randomization inference of causal effects in the presence of interference. *Journal of the American Statistical Association 109*(505), 288–301.

Manski, C. F. (2013). Identification of treatment response with social interactions. *The Econometrics Journal 16*(1), S1–S23.

Puelz, D., G. Basse, A. Feller, and P. Toulis (2019). A graph-theoretic approach to randomization tests of causal effects under general interference. *arXiv preprint arXiv:1910.10862*.

Rambachan, A. and N. Shephard (2019). Econometric analysis of potential outcomes time series: instruments, shocks, linearity and the causal response function. *arXiv preprint arXiv:1903.01637*.

Romano, J. P. and M. Wolf (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., P. M. Aronow, and M. G. Hudgens (2017). Average treatment effects in the presence of unknown interference. *arXiv e-prints*, arXiv:1711.06399.

Sinclair, B., M. McConnell, and D. P. Green (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.

Ugander, J., B. Karrer, L. Backstrom, and J. Kleinberg (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, New York, NY, USA, pp. 329–337. Association for Computing Machinery.

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

A 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 $X_1, \cdots, X_n$ be a collection of random variables such that $\mathbb{E}\left[X_i^4\right] < \infty$ and $\mathbb{E}\left[X_i\right] = 0$. 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 $(X_1, \cdots, X_n)$. Then for constants C_1 and C_2 which do not depend on n, d or σ^2,*

$$d_{\mathcal{W}}(S/\sigma) \leq C_1 \frac{d^{3/2}}{\sigma^2} \left(\sum_{i=1}^{n} \mathbb{E}\left[X_i^4\right] \right)^{1/2} + C_2 \frac{d^2}{\sigma^3} \sum_{i=1}^{n} \mathbb{E}|X_i|^3, \tag{10}$$

where $d_{\mathcal{W}}(S/\sigma)$ is the Wasserstein distance between S/σ 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 and Samii (2017)):*

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

Proof of Theorem 1. 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 $\mathbb{E}\left[\tilde{\tau}_i\right] = \frac{1}{n}\left[Y_i(k) - Y_i(k')\right]$, hence if we let $X_i = \sqrt{n}(\tilde{\tau}_i - \mathbb{E}\left[\tilde{\tau}_i\right])$, 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

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

and

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

hold. Moreover, by Assumption 4,

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

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 σ^2 is at least $O(1)$, we get:

$$\text{RHS of (10)} = o(n^{-1/8}) + o(1) \to 0$$

We're done. □

Remark 3. *In fact, with the tools in Leung (2019), 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 δ_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. $\square$

Proposition 4. *(Estimator of variance) We let*

$$
\begin{aligned}
\widehat{Var}(\sqrt{n}\hat{\tau}^{k,k'}) = \frac{1}{n}\Bigg\{ & \sum_{i=1}^{n} \mathbf{1}(H_i = k)(1-\pi_i(k)) \left[\frac{Y_i}{\pi_i(k)}\right]^2 + \sum_{i=1}^{n} \sum_{j \neq i, \pi_{ij}(k)=0} \left[\frac{\mathbf{1}(H_i = k)Y_i^2}{2\pi_i(k)} + \frac{\mathbf{1}(H_j = k)Y_j^2}{2\pi_j(k)}\right] \\
& + \sum_{i=1}^{n} \sum_{j \neq i, \pi_{ij}(k)>0} \mathbf{1}(H_i = k)\mathbf{1}(H_j = k) \frac{\pi_{ij}(k) - \pi_i(k)\pi_j(k)}{\pi_{ij}(k)} \frac{Y_i}{\pi_i(k)} \frac{Y_j}{\pi_j(k)} \\
& + \sum_{i=1}^{n} \mathbf{1}(H_i = k')(1-\pi_i(k')) \left[\frac{Y_i}{\pi_i(k')}\right]^2 + \sum_{i=1}^{n} \sum_{j \neq i, \pi_{ij}(k')=0} \left[\frac{\mathbf{1}(H_i = k')Y_i^2}{2\pi_i(k')} + \frac{\mathbf{1}(H_j = k')Y_j^2}{2\pi_j(k')}\right] \\
& + \sum_{i=1}^{n} \sum_{j \neq i, \pi_{ij}(k')>0} \mathbf{1}(H_i = k')\mathbf{1}(H_j = k') \frac{\pi_{ij}(k') - \pi_i(k')\pi_j(k')}{\pi_{ij}(k')} \frac{Y_i}{\pi_i(k')} \frac{Y_j}{\pi_j(k')} \\
& - 2\sum_{i=1}^{n} \sum_{j \neq i, \pi_{ij}(k,k')>0} \left[\left(\pi_{ij}(k,k') - \pi_i(k)\pi_j(k')\right) \frac{\mathbf{1}(H_i = k)\mathbf{1}(H_j = k')}{\pi_{ij}(k,k')} \frac{Y_i}{\pi_i(k)} \frac{Y_j}{\pi_j(k')}\right] \\
& + 2\sum_{i=1}^{n} \sum_{j \neq i, \pi_{ij}(k,k')=0} \left[\frac{\mathbf{1}(H_i = k)Y_i^2}{2\pi_i(k)} + \frac{\mathbf{1}(H_j = k')Y_j^2}{2\pi_j(k')}\right] \Bigg\},
\end{aligned}
$$

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

Proof of Proposition 1. We first prove the first part of the proposition. The proof is based on A.7 in Aronow and Samii (2017). To start with, for any $(i,j) \in \{1,\cdots,\} \times \{1,\cdots,n\}$, 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 $\widehat{\mathrm{Var}}(\hat{\tau}^{k,k'})$ that incorporate i and j, then

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

Note that $\mathrm{Cov}\left[e_{ij}a_{ij}(H_i,H_j), e_{kl}a_{kl}(H_k,H_l)\right]$ 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 $\mathrm{Var}\left(\widehat{\mathrm{Var}}(\hat{\tau}^{k,k'})\right) = o(n^{-4} \times n \times n^{3/4}) \to 0$ as $n \to \infty$. Then by Chebyshev's inequality, $\left|\widehat{\mathrm{Var}}(\sqrt{n}\hat{\tau}^{k,k'}) - \mathbb{E}\left[\widehat{\mathrm{Var}}(\sqrt{n}\hat{\tau}^{k,k'})\right]\right| = o_p(1)$. Since $\mathbb{E}\left[\widehat{\mathrm{Var}}(\hat{\tau}^{k,k'})\right] \geq \mathrm{Var}(\hat{\tau}^{k,k'})$, $\mathbb{P}\left(\frac{\widehat{\mathrm{Var}}(\hat{\tau}^{k,k'})}{\mathrm{Var}(\hat{\tau}^{k,k'})} \geq 1-\delta\right) \to 1$ 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{\mathrm{Var}(\sqrt{n}\hat{\tau}^{k,k'})}}\right| \leq \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{1-\delta}}\sqrt{\frac{\widehat{\mathrm{Var}}(\sqrt{n}\hat{\tau}^{k,k'})}{\mathrm{Var}(\sqrt{n}\hat{\tau}^{k,k'})}}\right) \\
&\geq \mathbb{P}\left(\left|\frac{\sqrt{n}(\hat{\tau}^{k,k'} - \tau^{k,k'})}{\sqrt{\mathrm{Var}(\sqrt{n}\hat{\tau}^{k,k'})}}\right| \leq \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{1-\delta}}\sqrt{\frac{\widehat{\mathrm{Var}}(\sqrt{n}\hat{\tau}^{k,k'})}{\mathrm{Var}(\sqrt{n}\hat{\tau}^{k,k'})}} \text{ and } \frac{\widehat{\mathrm{Var}}(\sqrt{n}\hat{\tau}^{k,k'})}{\mathrm{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{\mathrm{Var}(\sqrt{n}\hat{\tau}^{k,k'})}}\right| \leq z_{1-\frac{\alpha}{2}} \text{ and } \frac{\widehat{\mathrm{Var}}(\sqrt{n}\hat{\tau}^{k,k'})}{\mathrm{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{\mathrm{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{\mathrm{Var}(\sqrt{n}\hat{\tau}^{k,k'})}}\right| \leq z_{1-\frac{\alpha}{2}} \text{ and } \frac{\widehat{\mathrm{Var}}(\sqrt{n}\hat{\tau}^{k,k'})}{\mathrm{Var}(\sqrt{n}\hat{\tau}^{k,k'})} < 1-\delta\right) \\
&\geq \mathbb{P}\left(\left|\frac{\sqrt{n}(\hat{\tau}^{k,k'} - \tau^{k,k'})}{\sqrt{\mathrm{Var}(\sqrt{n}\hat{\tau}^{k,k'})}}\right| \leq z_{1-\frac{\alpha}{2}}\right) - \mathbb{P}\left(\frac{\widehat{\mathrm{Var}}(\sqrt{n}\hat{\tau}^{k,k'})}{\mathrm{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{\mathrm{Var}(\sqrt{n}\hat{\tau}^{k,k'})}}\right| \leq z_{1-\frac{\alpha}{2}}\right) - \mathbb{P}\left(\frac{\widehat{\mathrm{Var}}(\hat{\tau}^{k,k'})}{\mathrm{Var}(\hat{\tau}^{k,k'})} < 1-\delta\right) \\
&\to 1-\alpha
\end{aligned}
$$

as $n \to \infty$ by the first part and Theorem 1. $\square$

Proof of Theorem 2. We use a characteristic function argument. We first note that

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

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

$$\begin{aligned}
\mathbb{E}\left[\exp\left\{i\lambda\frac{\sqrt{nT}(\hat{\bar{\tau}}^{k,k'} - \bar{\tau}^{k,k'})}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}\sigma_{n,t}^2}}\right\}\right] &= \mathbb{E}\left[\exp\left\{i\lambda\frac{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}X_{n,t}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}\sigma_{n,t}^2}}\right\}\right] \\
&= \prod_{t=1}^{T}\mathbb{E}\left[\exp\left\{i\lambda\frac{\frac{1}{\sqrt{T}}X_{n,t}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}\sigma_{n,t}^2}}\right\}\right] \\
&= \prod_{t=1}^{T}\mathbb{E}\left[\exp\left\{i\frac{\lambda\sigma_{n,t}}{\sqrt{\sum_{t=1}^{T}\sigma_{n,t}^2}}\frac{X_{n,t}}{\sigma_{n,t}}\right\}\right] \\
&= \prod_{t=1}^{T}\phi_{\frac{X_{n,t}}{\sigma_{n,t}}}\left(\frac{\lambda\sigma_{n,t}}{\sqrt{\sum_{t=1}^{T}\sigma_{n,t}^2}}\right) \qquad (11)
\end{aligned}$$

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

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

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

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

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 $n \geq N_t$,

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

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

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

where K is any big number we want. Now,

$$\begin{aligned} (11) &= \prod_{t=1}^{T} \left(\exp\left\{ -\frac{1}{2} \frac{\lambda^2 \sigma_{n,t}^2}{\sum_{t=1}^{T} \sigma_{n,t}^2} \right\} + \epsilon_t \right) \\ &= \exp\left\{ -\frac{1}{2} \lambda^2 \right\} + R(\epsilon_t), \end{aligned}$$

where $R(\epsilon_t)$ is a remainder term that is the sum of several monomial terms of ϵ_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,

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

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

□

To prove Theorem 4, 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*

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

Suppose that for any $\epsilon > 0$,

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{i=1}^{k_n} \mathbb{E}\left((X_{n,i} - \mu_{n,i})^2; |X_{n,i} - \mu_{n,i}| \geq \epsilon s_n\right) = 0. \tag{12}$$

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 $n \to \infty$.

Proof of Theorem 4. We first prove the theorem with condition (6). We note that $\sqrt{nT}(\hat{\bar{\tau}}^{k,k'} - \bar{\tau}^{k,k'}) = \sum_{t=1}^{T} \sqrt{\frac{n}{T}}(\hat{\tau}_t^{k,k'} - \tau_t^{k,k'})$. 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, $X_{n,1}, \cdots, X_{n,T}$ are independent by the pure population interference assumption. Now,

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

Hence, to finish the proof, we only need to check (12) 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

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

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, (12) is satisfied. We're done. The proof of this theorem with condition (7) 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 3, 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*

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

If for some $\delta > 0$,

$$\lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^n \mathbb{E}|X_i - \mu_i|^{2+\delta} = 0, \tag{13}$$

then the random variable

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

Proof of Theorem 3. 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,

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

Hence, we only need to check (13). We have that

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

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

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

If $T \to \infty$, $\frac{1}{s_T^{2+\delta}}\frac{n^{1+\frac{\delta}{2}}}{T}M^{2+\delta} \to 0$. Therefore, (13) is satisfied. We're done. $\square$

Proposition 5. *Suppose Theorem 2 or 4 holds, then for any $\delta > 0$,*

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

for large n. Moreover, suppose Theorem 3 holds, then for any $\delta > 0$,

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

for large T.

Proof of Proposition 5. Now we can prove the second part of the proposition. We have that

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

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

$$\mathrm{Var}\left(\frac{1}{T}\sum_{t=1}^{T}\widehat{\mathrm{Var}}(\hat{\tau}_t^{k,k'})\right) = \frac{1}{T^2}\sum_{t=1}^{T}\mathrm{Var}\left(\widehat{\mathrm{Var}}(\hat{\tau}_t^{k,k'})\right). \tag{14}$$

If T is fixed (i.e., Theorem 2 holds), then by what we have in Proposition 1, we immediately have that (14) $\to 0$ and we are done. Now suppose Theorem 4 holds. Recall that

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

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

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

as $T \to 0$. So in the regime where both n and T go to infinity (i.t. Theorem 4 holds) or T goes to infinity (i.e., Theorem 3 holds), (14) $\to 0$ and we are done. □

Proof of Theorem 5. This should be exactly the same as our proof of Theorem 1. □

Proof of Proposition 2.

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

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.

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

by our ϵ-weak-stability assumption. Hence,

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

$\square$

Remark 4. *Note that following the exact derivation, we can know that*

$$|\tau_t^{TE} - \tau_{t'}^{TE}| \le 2|t - t^{'}|\epsilon \tag{15}$$

Proposition 6 (Variance and Covariance of Horvitz-Thompson Type Estimators)**.** *For each $i \in \{1, \cdots, n\}, t \in \{1, \cdots, T\}$, 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 $i \ne j$ and t, we let $\mathbb{P}(H_{i,t} = h_i^1, H_{j,t} = h_j^1) = \pi_{i,j,t}^{1,1}$, $\mathbb{P}(H_{i,t} = h_i^0, H_{j,t} = h_j^1) = \pi_{i,j,t}^{0,1}$, $\mathbb{P}(H_{i,t} = h_i^1, H_{j,t} = h_j^0) = \pi_{i,j,t}^{1,0}$ and $\mathbb{P}(H_{i,t} = h_i^0, H_{j,t} = h_j^0) = \pi_{i,j,t}^{0,0}$, then*

$$\begin{aligned}
Var(\hat{\tau}_t^{TE}) = \frac{1}{n^2}\sum_{i=1}^n &\left[\frac{Y_{i,t}^2(h_i^1)(1-\pi_{i,t}^1)}{\pi_{i,t}^1} + \frac{Y_{i,t}^2(h_i^0)(1-\pi_{i,t}^0)}{\pi_{i,t}^0} + 2Y_{i,t}(h_i^1)Y_{i,t}(h_i^0) \right] \\
+ \frac{2}{n^2}\sum_{1\le i<j\le n} &\left[\frac{Y_{i,t}(h_i^1)Y_{j,t}(h_j^1)(\pi_{i,j,t}^{1,1} - \pi_{i,t}^1\pi_{j,t}^1)}{\pi_{i,t}^1\pi_{j,t}^1} - \frac{Y_{i,t}(h_i^0)Y_{j,t}(h_j^1)(\pi_{i,j,t}^{0,1} - \pi_{i,t}^0\pi_{j,t}^1)}{\pi_{i,t}^0\pi_{j,t}^1} \right. \\
&\left. - \frac{Y_{i,t}(h_i^1)Y_{j,t}(h_j^0)(\pi_{i,j,t}^{1,0} - \pi_{i,t}^1\pi_{j,t}^0)}{\pi_{i,t}^1\pi_{j,t}^0} + \frac{Y_{i,t}(h_i^0)Y_{j,t}(h_j^0)(\pi_{i,j,t}^{0,0} - \pi_{i,t}^0\pi_{j,t}^0)}{\pi_{i,t}^0\pi_{j,t}^0} \right]
\end{aligned} \tag{16}$$

As for $Cov(\hat{\tau}_t^{TE}, \hat{\tau}_{t'}^{TE})$, 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 $\mathbb{P}(H_{i,t} =$

$h_i^1, H_{j,t'} = h_j^1) = \pi^{1,1}_{i,t,j,t'}$, $\mathbb{P}(H_{i,t} = h_i^0, H_{j,t'} = h_j^1) = \pi^{0,1}_{i,t,j,t'}$, $\mathbb{P}(H_{i,t} = h_i^1, H_{j,t'} = h_j^0) = \pi^{1,0}_{i,t,j,t'}$, $\mathbb{P}(H_{i,t} = h_i^0, H_{j,t'} = h_j^0) = \pi^{0,0}_{i,t,j,t'}$, *then we have that*

$$
\begin{aligned}
Cov(\hat{\tau}_t^{TE}, \hat{\tau}_{t'}^{TE}) = \frac{1}{n^2}\sum_{i=1}^{n}\Bigg[&\frac{Y_{i,t}(h_i^1)Y_{i,t'}(h_i^1)(\pi^{1,1}_{i,t,t'} - \pi^1_{i,t}\pi^1_{i,t'})}{\pi^1_{i,t}\pi^1_{i,t'}} - \frac{Y_{i,t}(h_i^0)Y_{i,t'}(h_i^1)(\pi^{0,1}_{i,t,t'} - \pi^0_{i,t}\pi^1_{i,t'})}{\pi^0_{i,t}\pi^1_{i,t'}} \\
&- \frac{Y_{i,t}(h_i^1)Y_{i,t'}(h_i^0)(\pi^{1,0}_{i,t,t'} - \pi^1_{i,t}\pi^0_{i,t'})}{\pi^1_{i,t}\pi^0_{i,t'}} + \frac{Y_{i,t}(h_i^0)Y_{i,t'}(h_i^0)(\pi^{0,0}_{i,t,t'} - \pi^0_{i,t}\pi^0_{i,t'})}{\pi^0_{i,t}\pi^0_{i,t'}}\Bigg] \\
+ \frac{2}{n^2}\sum_{1\le i<j\le n}\Bigg[&\frac{Y_{i,t}(h_i^1)Y_{j,t'}(h_j^1)(\pi^{1,1}_{i,t,j,t'} - \pi^1_{i,t}\pi^1_{j,t'})}{\pi^1_{i,t}\pi^1_{j,t'}} - \frac{Y_{i,t}(h_i^0)Y_{j,t'}(h_j^1)(\pi^{0,1}_{i,t,j,t'} - \pi^0_{i,t}\pi^1_{j,t'})}{\pi^0_{i,t}\pi^1_{j,t'}} \\
&- \frac{Y_{i,t}(h_i^1)Y_{j,t'}(h_j^0)(\pi^{1,0}_{i,t,j,t'} - \pi^1_{i,t}\pi^0_{j,t'})}{\pi^1_{i,t}\pi^0_{j,t'}} + \frac{Y_{i,t}(h_i^0)Y_{j,t'}(h_j^0)(\pi^{0,0}_{i,t,j,t'} - \pi^0_{i,t}\pi^0_{j,t'})}{\pi^0_{i,t}\pi^0_{j,t'}}\Bigg]
\end{aligned}
\tag{17}
$$

Proof of Proposition 6. This can be done by direct calculations. □

Proposition 7. *Suppose $Var(\hat{\tau}_t^{TE}) > Cov(\hat{\tau}_t^{TE}, \hat{\tau}_{t-1}^{TE})$, 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 $Var(\hat{\tau}_t^{TE}) - Var(\hat{\tau}_{t-1}^{TE}) > 4\epsilon^2$ 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 $\hat{\tau}_t^{TE}$.*

By Cauchy-Schwartz inequality,

$$\mathrm{Cov}(\hat{\tau}_t^{TE}, \hat{\tau}_{t-1}^{TE}) \le \sqrt{\mathrm{Var}(\hat{\tau}_t^{TE})}\sqrt{\mathrm{Var}(\hat{\tau}_{t-1}^{TE})},$$

hence if $\mathrm{Var}(\hat{\tau}_t^{TE}) > \mathrm{Var}(\hat{\tau}_{t-1}^{TE})$, then $Var(\hat{\tau}_t^{TE}) > Cov(\hat{\tau}_t^{TE}, \hat{\tau}_{t-1}^{TE})$. Therefore, as long as the current variance is larger, by choosing some α, 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.

Proof of Proposition 7. 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

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

By Proposition 2, it suffices to have

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

which is further equivalent to

$$\alpha^2 Var(\hat{\tau}_t^{TE}) + (1-\alpha)^2 Var(\hat{\tau}_{t-1}^{TE}) + 2\alpha(1-\alpha)Cov(\hat{\tau}_t^{TE}, \hat{\tau}_{t-1}^{TE}) + 4(1-\alpha)^2\epsilon^2 \le Var(\hat{\tau}_t^{TE}) \quad (18)$$

Rewrite (18), we have

$$\begin{aligned}&\left(4\epsilon^2 + Var(\hat{\tau}_t^{TE}) + Var(\hat{\tau}_{t-1}^{TE}) - 2Cov(\hat{\tau}_t^{TE}, \hat{\tau}_{t-1}^{TE})\right)\alpha^2 - \\ &\quad \left(8\epsilon^2 + 2Var(\hat{\tau}_{t-1}^{TE}) - 2Cov(\hat{\tau}_t^{TE}, \hat{\tau}_{t-1}^{TE})\right)\alpha + \left(4\epsilon^2 + Var(\hat{\tau}_{t-1}^{TE}) - Var(\hat{\tau}_t^{TE})\right) \le 0 \quad (19)\end{aligned}$$

Now we look at the left hand side of (19), which is quadratic in α. To ease notations, let $A = Var(\hat{\tau}_t^{TE})$, $B = Var(\hat{\tau}_{t-1}^{TE})$ and $C = Cov(\hat{\tau}_t^{TE}, \hat{\tau}_{t-1}^{TE})$. 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 $A - B > 4\epsilon^2$. □

Note that if we further assume that assignments are also independent across time, then $\mathrm{Cov}(\hat{\tau}_t^{TE}, \hat{\tau}_{t-1}^{TE}) = 0$, hence we have the Proposition 3. Up to now, the type of the estimator we consider is a convex combination of $\hat{\tau}_t^{TE}$ and $\hat{\tau}_{t-1}^{TE}$. We now consider a general version of this estimator such that we combine $\hat{\tau}_t^{TE}$ and $\hat{\tau}_{t'}^{TE}$ for arbitrary $t' < t$. We now give the analogous results.

Proposition 8. *Suppose $Var(\hat{\tau}_t^{TE}) > Cov(\hat{\tau}_t^{TE}, \hat{\tau}_{t'}^{TE})$, then there exists some $\alpha \in (0,1)$ such that $\hat{\tau}_t^c = \alpha\hat{\tau}_t^{TE} + (1-\alpha)\hat{\tau}_{t'}^{TE}$ has lower MSE than $\hat{\tau}_t^{TE}$. Moreover, if we have $Var(\hat{\tau}_t^{TE}) - Var(\hat{\tau}_{t'}^{TE}) > 4(t-t')^2\epsilon^2$, then $\hat{\tau}_t^c = \frac{1}{2}\hat{\tau}_t^{TE} + \frac{1}{2}\hat{\tau}_{t'}^{TE}$ has lower MSE than $\hat{\tau}_t^{TE}$.*

Proposition 9. *Suppose that the assignments are independent across time, then there exists some $\alpha \in (0,1)$ such that $\hat{\tau}_t^c = \alpha\hat{\tau}_t^{TE} + (1-\alpha)\hat{\tau}_{t'}^{TE}$ has lower MSE than $\hat{\tau}_t^{TE}$. The optimal α is given by $\alpha = 1 - \frac{Var(\hat{\tau}_t^{TE})}{4(t-t')^2\epsilon^2 + Var(\hat{\tau}_t^{TE}) + Var(\hat{\tau}_{t'}^{TE})}$.*

Proposition 10 (Estimators of variance)**.** *We define two estimators of the variance:*

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

and

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

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 11 (Estimator of the covariance)**.** *We have the following unbiased estimator of $Cov(\hat{\tau}_t^{TE}, \hat{\tau}_{t'}^{TE})$:*

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

We notice that under LEA(p) assumption, we are essentially having $(p-1)$-dependent sequence. Hence, we can utilize results from literature in $m-$dependence central limit theorem to prove Theorem 6. We first need a lemma.

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

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

Assume the following conditions hold. For some $\delta > 0$, some $-1 \leq \gamma < 1$ and some $p = p_n > 2m$ such that $p \to \infty$ as $n \to \infty$:

$$\mathbb{E}|X_{n,i}|^{2+\delta} \leq \Delta_n \text{ for all } i, \tag{20}$$

$$B_{n,k,a}^2/(k^{1+\gamma}) \leq K_n \text{ for all } a \text{ and for all } k \geq m, \tag{21}$$

$$B_n^2/(dm^\gamma) \geq L_n, \tag{22}$$

$$K_n/L_n = O(1), \tag{23}$$

$$\Delta_n L_n^{-(2+\delta)/2} p^{\delta/2+(1-\gamma)(2+\delta)/2} d^{-\delta/2} = o(1). \tag{24}$$

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

Remark 5. *This is essentially fixed m version of Theorem 2.1 in Romano and Wolf (2000).*

Proof. We modify the proof in Romano and Wolf (2000). We first note that assumption 6 in the original theorem is naturally satisfied since m is fixed. Moreover, the only place we need assumption 5 in the original theorem is to prove

$$\Delta_n L_n^{-(2+\delta)/2} p^{\delta/2+(1-\gamma)(2+\delta)/2} d^{-\delta/2} \left(\frac{m}{p}\right)^{(1-\gamma)(2+\delta)/2} \to 0$$

for $p = p_n > 2m$ so that

$$\lim_{n\to\infty} m/p = 0.$$

By replacing assumption 5 and 6 in the original theorem with (24), the original proof goes through. $\square$

Proof of Theorem 6. We prove this theorem with the sufficient condition we mentioned in the main text. In fact, as we will see in the proof, there is no difference. We check the five conditions in Lemma 5 are satisfied with $m = 1$, $\gamma = 0$, $p = 2 + n^{\epsilon/4}$ and $\delta = 2 + \frac{4}{\epsilon}$. 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:

$$\begin{aligned}\mathrm{Var}(X_{n,t}) = \frac{1}{nrT}\Bigg[&\sum_{l=1}^{n}\sum_{q=1}^{r}(2^{2r}-1)Y_{(l,q),t}(k)^2 + \sum_{l=1}^{n}\sum_{q=1}^{r}(2^{2r}-1)Y_{(l,q),t}(k')^2 + 2\sum_{l=1}^{n}\sum_{q=1}^{r}Y_{(l,q),t}(k)Y_{(l,q),t}(k') \\ &+\sum_{l=1}^{n}\sum_{q_1=1}^{r}\sum_{q_2\neq q_1}\Big((2^{2r}-1)Y_{(l,q_1),t}(k)Y_{(l,q_2),t}(k) + (2^{2r}-1)Y_{(l,q_1),t}(k')Y_{(l,q_2),t}(k')\Big) \\ &+2\sum_{l=1}^{n}\sum_{q_1=1}^{r}\sum_{q_2\neq q_1}Y_{(l,q_1),t}(k)Y_{(l,q_2),t}(k')\Bigg] \end{aligned} \tag{25}$$

and

$$\begin{aligned}\mathrm{Cov}(X_{n,t}, X_{n,t+1}) = \frac{1}{nrT}\sum_{l=1}^{n}\sum_{q_1=1}^{r}\sum_{q_2=1}^{r}\Big(&(2^{r}-1)Y_{(l,q_1),t}(k)Y_{(l,q_2),t+1}(k) + (2^{r}-1)Y_{(l,q_1),t}(k')Y_{(l,q_2),t+1}(k') \\ &+ Y_{(l,q_1),t}(k')Y_{(l,q_2),t+1}(k) + Y_{(l,q_1),t}(k)Y_{(l,q_2),t+1}(k')\Big)\end{aligned} \tag{26}$$

First, we know that $|\hat{\tau}_t^{k,k'} - \tau_t^{k,k'}| \leq M_1$ for all n, t under our current setting. Hence, $|X_{n,t}| \leq \sqrt{\frac{nr}{T}}M_1$ for all t. As a result, $\mathbb{E}|X_{n,t}|^{2+\delta} \leq \left(\frac{nr}{T}\right)^{1+\delta/2} M_1^{2+\delta}$. We let $\Delta_n = \left(\frac{nr}{T}\right)^{1+\delta/2} M_1^{2+\delta}$. With the sufficient condition, we have that

$$\frac{2r2^{2r}C_1^2}{T} \leq \mathrm{Var}(X_{n,t}) \leq \frac{2r2^{2r}C_2^2}{T} \tag{27}$$

and

$$\frac{2^{r+1}rC_1^2}{T} \leq \mathrm{Cov}(X_{n,t}, X_{n,t+1}) \leq \frac{2^{r+1}rC_2^2}{T} \tag{28}$$

Next, we calculate $B_{n,k,a}^2$ and B_n^2. We have that

$$\begin{aligned}B_{n,k,a}^2 &= \mathrm{Var}\left(\sum_{t=a}^{a+k-1} X_{n,t}\right) \\ &= \sum_{t=a}^{a+k-1}\mathrm{Var}(X_{n,t}) + 2\sum_{t=a}^{a+k-2}\mathrm{Cov}(X_{n,t}, X_{n,t+1})\end{aligned}$$

and

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

Therefore, by (27) and (28), we further have that

$$B_{n,k,a}^2 \leq \frac{2rk2^{2r}C_2^2}{T} + \frac{2(k-1)r2^{r+1}C_2^2}{T}$$

and

$$B_n^2 \geq 2r2^{2r}C_1^2 + \frac{2(T-1)r2^{r+1}C_1^2}{T}$$

For $m = 1$ and $\gamma = 0$, we know that for $T \geq 2$

$$B_{n,k,a}^2/(k^{1+\gamma}) = B_{n,k,a}^2/k \leq \frac{2r2^{2r}C_2^2}{T} + \frac{2(k-1)r2^{r+1}C_2^2}{Tk} \leq \frac{2r2^{2r}C_2^2}{T} + \frac{2r2^{r+1}C_2^2}{T} = K_n \quad (29)$$

and

$$B_n^2/(Tm^{\gamma}) = B_n^2/T \geq \frac{2r2^{2r}C_1^2}{T} + \frac{2(T-1)r2^{r+1}C_1^2}{T^2} \geq \frac{2r2^{2r}C_1^2}{T} + \frac{r2^{r+1}C_1^2}{T} = L_n \quad (30)$$

Now,

$$\begin{aligned} K_n/L_n &= \left(\frac{2r2^{2r}C_2^2}{T} + \frac{2r2^{r+1}C_2^2}{T}\right) \Big/ \left(\frac{2r2^{2r}C_1^2}{T} + \frac{r2^{r+1}C_1^2}{T}\right) \\ &= \frac{2^{2r+1} + 2^{r+2}}{2^{2r+1} + 2^{r+1}} \frac{C_2^2}{C_1^2} = O(1). \end{aligned}$$

Here, if we just assume Assumption 6, then K_n/L_n is still $O(1)$. So we only need to check (24). We plug in $\Delta_n, L_n, p_n, \delta$, for large n:

$$\begin{aligned} \text{LHS of (24)} &= \frac{n^{1+\delta/2}r^{1+\delta/2}M_1^{2+\delta}(2+n^{\frac{\epsilon}{4}})^{1+\delta}}{(2r2^{2r}+r2^{r+1})^{1+\frac{\delta}{2}}C_1^{2+\delta}T^{\frac{\delta}{2}}} \\ &= C\frac{n^{1+\delta/2}(2+n^{\frac{\epsilon}{4}})^{1+\delta}}{T^{\frac{\delta}{2}}} \\ &\leq C' \frac{n^{1+\delta/2}n^{\frac{\epsilon}{4}(1+\delta)}}{T^{\frac{\delta}{2}}} \\ &= C' \frac{n^{3+\frac{3}{4}\epsilon+\frac{2}{\epsilon}}}{T^{1+\frac{2}{\epsilon}}} \end{aligned}$$

So for $T = O(n^{1+\epsilon})$, we have that

$$\begin{aligned} \text{LHS of (24)} &= C' \frac{n^{3+\frac{3}{4}\epsilon+\frac{2}{\epsilon}}}{T^{1+\frac{2}{\epsilon}}} \\ &\leq C'' \frac{n^{3+\frac{3}{4}\epsilon+\frac{2}{\epsilon}}}{n^{3+\epsilon+\frac{2}{\epsilon}}} \\ &= C'' \frac{1}{n^{\frac{\epsilon}{4}}} = o(1) \end{aligned}$$

$\square$

B $k-$steps convex estimator

The approach we have described in Section 4.2 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 12 (Bound on the bias of $\hat{\tau}_t^c$)**.**

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

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

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

where $\widehat{\mathrm{Var}}(\hat{\tau}_{t-k+1}^{TE}), \cdots, \widehat{\mathrm{Var}}(\hat{\tau}_t^{TE})$ are estimators of the associated variance terms, and are provided in the Supplementary Material. 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:

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

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

Proof of Proposition 12.

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

□

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

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

where $\widehat{\mathrm{Var}}(\hat{\tau}_{t-k+1}^{TE}), \cdots, \widehat{\mathrm{Var}}(\hat{\tau}_t^{TE})$ and $\widehat{\mathrm{Cov}}(\hat{\tau}_{t-k+i}^{TE}, \hat{\tau}_{t-k+j}^{TE})$ can be any estimator in Proposition 10 and 11. Moreover, suppose that the assignments are independent across time, we know that $\mathrm{Cov}(\hat{\tau}_{t-k+i}^{TE}, \hat{\tau}_{t-k+j}^{TE}) = 0$, hence we have an even simpler optimization problem as stated in the main text.

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

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

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

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

By Proposition 12, it suffices to have

$$\begin{aligned}\alpha_1^2 \mathrm{Var}(\hat{\tau}_{t-k+1}^{TE}) + \cdots + \alpha_k^2 \mathrm{Var}(\hat{\tau}_t^{TE}) + 2\alpha_i\alpha_j \sum_{1\le i<j\le k} \mathrm{Cov}(\hat{\tau}_{t-k+i}^{TE}, \hat{\tau}_{t-k+j}^{TE}) \\ + 4\left[(k-1)\alpha_1 + \cdots + \alpha_{k-1}\right]^2 \epsilon^2 \le \mathrm{Var}(\hat{\tau}_t^{TE})\end{aligned} \tag{31}$$

Now, the left hand side of (31) is convex in $\alpha_1, \cdots, \alpha_k$. □

C General Framework

In this paper, we have a two-dimensional indexing set: one dimension for indexing the multiple units, one dimension for indexing the time. This can be generalized to two arbitrary indices. For example, each place on earth can be indexed by latitude and longitude. We can talk about causal inference in this general case.

We start with an arbitrary indexing set $\mathcal{A}$. Corresponding to each element $a \in \mathcal{A}$, we have an assignment w_a. Hence, there is an assignment array $\underline{w} = (w_a)_{a\in\mathcal{A}}$ associated with $\mathcal{A}$. For each element $a \in \mathcal{A}$, we associate an exposure mapping $f_a : \Omega(\mathcal{A}) \to \Delta$ with it, where $\Omega(\mathcal{A})$ represents all the possible assignment arrays on our indexing set $\mathcal{A}$. Note that although we restrict all the exposure mappings $(f_a)_{a\in\mathcal{A}}$ to have the same range, we do not restrict them to have the same image. We adopt the potential outcome framework and associate each $a \in \mathcal{A}$ a set of potential outcomes $\{Y_a(\underline{w})\}_{\underline{w}\in\Omega(\mathcal{A})}$. Under this general setting, we have the following definition of properly specified exposures:

Definition 6 ($\mathcal{A}$-Properly Specified Exposures)**.** *We say that $(f_a)_{a\in\mathcal{A}}$ is $\mathcal{A}$-properly specified if $\forall a \in \mathcal{A}$, $\forall \underline{w}, \underline{w}' \in \Omega(\mathcal{A})$, we have*

$$f_a(\underline{w}) = f_a(\underline{w}') \implies Y_a(\underline{w}) = Y_a(\underline{w}')$$

In the common causal inference literature, such exposure mappings induce interference and thus quantify our belief of the interference mechanism. On the other hand, properly specified exposure mappings reduce the number of possible potential outcomes and hence make inference possible. Two familiar examples are:

Example 5 (Traditional Causal Inference with Interference)**.** *This is the setting discussed in Aronow and Samii (2017). Under this setting, we have that $\mathcal{A} = \mathcal{I} = \{1, \cdots, n\}$, where n is the number of total units in the experiment.*

Example 6 (Time Series Experiments)**.** *This is the setting discussed in Bojinov and Shephard (2019). Under this setting, we have that $\mathcal{A} = \mathcal{T} = \{1, \cdots, T\}$, where we have only one unit participating the experiment and we assign treatment or control to this unit at T time points.*

The most general causal estimand we are interested in is the following exposure contrast:

Definition 7 (General Exposure Contrast)**.** *For $k, k^{'} \in \Delta$ and $\mathcal{A}_0 \subseteq \mathcal{A}$, we define the exposure contrast between k and $k^{'}$ on $\mathcal{A}_0$ as*

$$\tau_{k,k^{'}}(\mathcal{A}_0) = \frac{1}{|\mathcal{A}_0|} \sum_{a \in \mathcal{A}_0} (Y_a(k) - Y_a(k^{'}))$$

We have two remarks here. First, this may not be well-defined for all k and $k^{'}$ since the f_a's are not constrained to have the same image. Second, some choices of $\mathcal{A}_0$ do not make sense. We continue our two examples above here. For the traditional causal inference with interference, we average over $\mathcal{A} = \mathcal{I}$ and for time series experiment with one unit, we average over time.

Now, consider the case that $\mathcal{A} = \mathcal{I}_1 \times \mathcal{I}_2$, i.e., we have a two dimensional indexing set. In this case, we have two symmetric parts of the problem: fixing $i \in \mathcal{I}_1$ and do inference on $\mathcal{A}_i = \{i\} \times \mathcal{I}_2$, fixing $j \in \mathcal{I}_2$ and do inference on $\mathcal{A}_j = \mathcal{I}_1 \times \{j\}$. We define two special interference structures on two dimensional indexing sets.

Definition 8 (Purely $\mathcal{I}_1$-level Interference)**.** *$\forall t \in \mathcal{I}_2$, $\forall \underline{w}, \underline{w}^{'} \in \Omega(\mathcal{I}_1 \times \mathcal{I}_2)$, we have*

$$(w_{i,t})_{i \in \mathcal{I}_1} = (w^{'}_{i,t})_{i \in \mathcal{I}_1} \implies f_{i,t}(\underline{w}) = f_{i,t}(\underline{w}^{'})$$

We can define purely $\mathcal{I}_2$-level interference similarly. We also have two invariant properties of exposure mappings.

Sample Size	n = 100	n = 250	n = 500	n = 1000	n = 1500	n = 2000	n = 3000	n = 5000
lower and upper quartile of $\frac{\widehat{\text{Var}}}{\text{Var}}$	0.20, 1.27	0.90, 1.23	0.95, 1.17	0.98, 1.14	1.00, 1.13	1.01, 1.13	1.02, 1.12	1.03, 1.10
p-value	9.87e-24	1.09e-06	0.006	0.129	0.653	0.803	0.864	0.922
Coverage	94.3%	95.4%	95.8%	96.0%	96.1%	96.0%	96.0%	96.1%

Table 5: Simulation results for Theorem 1, normal distributed potential outcomes

Sample Size	n = 100	n = 250	n = 500	n = 1000	n = 1500	n = 2000	n = 3000	n = 5000
lower and upper quartile of $\frac{\widehat{\text{Var}}}{\text{Var}}$	0.75, 1.34	0.88, 1.25	0.94, 1.23	0.98, 1.18	1.00, 1.16	1.01, 1.15	1.02, 1.14	1.04, 1.13
p-value	4.20e-14	7.33e-07	1.46e-04	0.013	0.552	0.382	0.701	0.790
Coverage	93.7%	95.1%	95.5%	96.2%	96.1%	96.0%	96.2%	96.4%

Table 6: Simulation results for Theorem 1, Poisson distributed potential outcomes

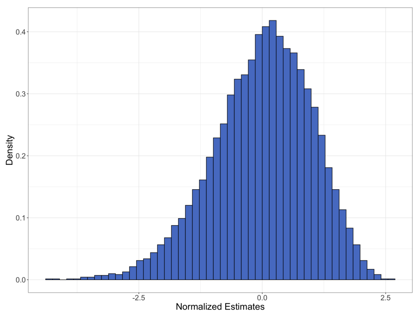
Definition 9 ($\mathcal{I}_1$-invariant Exposure Mappings)**.** *We say $f_{i,t}, (i,t) \in \mathcal{I}_1 \times \mathcal{I}_2$ is $\mathcal{I}_1$-invariant if $\forall t \in \mathcal{I}_2$, $\forall i, i^{'} \in \mathcal{I}_1$, $\forall \underline{w} \in \Omega(\mathcal{I}_1 \times \mathcal{I}_2)$,*

$$(w_{i,t})_{t \in \mathcal{I}_2} = (w_{i^{'},t})_{t \in \mathcal{I}_2} \implies f_{i,t}(\underline{w}) = f_{i^{'},t}(\underline{w})$$

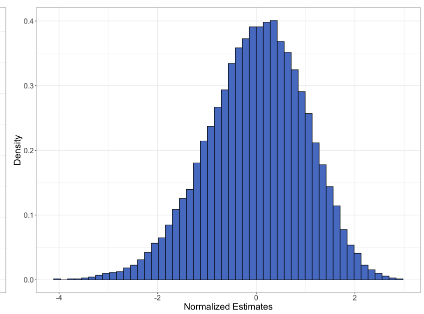
Similarly for $\mathcal{I}_2$-invariant exposure mappings.

Sample Size	n = 100	n = 250	n = 500	n = 1000	n = 1500	n = 2000	n = 3000	n = 5000
lower and upper quartile of $\frac{\widehat{\text{Var}}}{\text{Var}}$	0.46, 1.85	0.77, 1.61	0.92, 1.46	1.01, 1.40	1.06, 1.38	1.07, 1.36	1.10, 1.33	1.13, 1.31
p-value	1.04e-30	2.53e-20	1.02e-12	1.21e-07	6.95e-06	8.88e-05	0.014	0.210%
Coverage	93.6%	92.7%	95.3%	96.2%	96.6%	96.9%	97.0%	97.2%

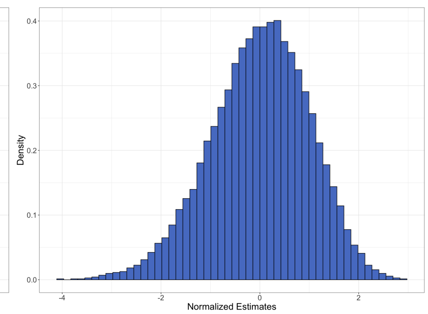
Table 7: Simulation results for Theorem 1, Bernoulli distributed potential outcomes



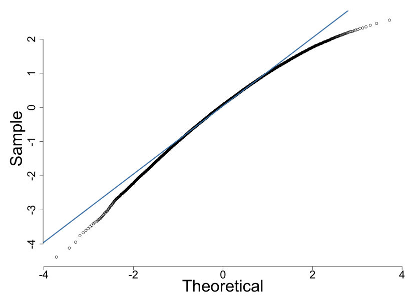
(a) $n = 250$, histogram



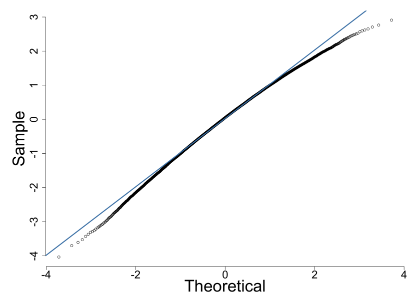
(b) $n = 500$, histogram



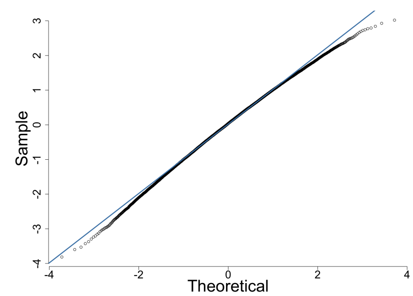
(c) $n = 1000$, histogram



(d) $n = 250$, Q-Q normal plot



(e) $n = 500$, Q-Q normal plot



(f) $n = 1000$, Q-Q normal plot

Figure 3: Histograms and Q-Q normal plots of $\frac{\sqrt{n}(\hat{\tau}^{k,k'}-\tau^{k,k'})}{\mathrm{Var}(\sqrt{n}\hat{\tau}^{k,k'})^{1/2}}$ for $n = 250, 500, 1000$

D Additional simulation results

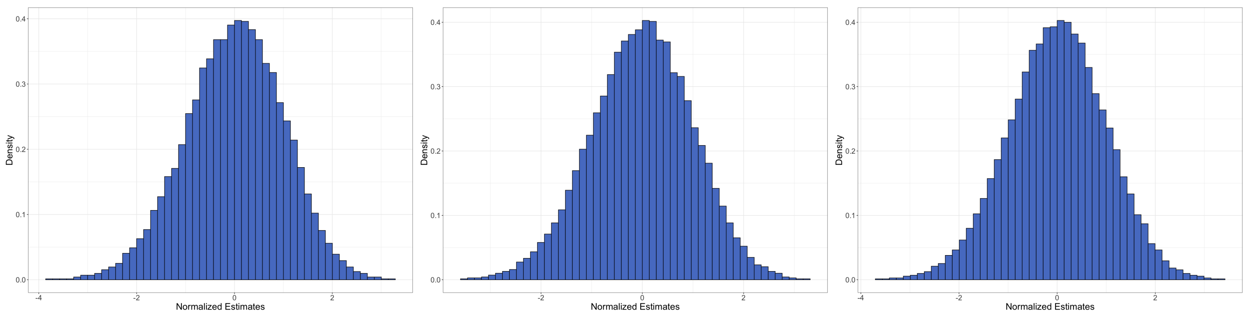
D.1 Simulation results for central limit theorems

D.1.1 Household interference in cross-sectional setting

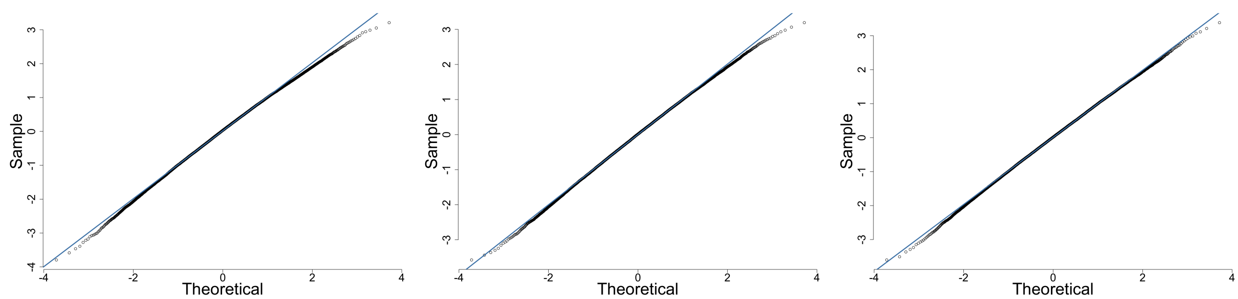
We first provide full simulation results for the household interference example in Section 6.1. Table 5, 6 and 7 show the full results of three data generating processes. As mentioned in the main text, we show the quality of the normal approximation for a given sample size n in Figure 3 and 4.

D.1.2 The effect of group size

We investigate how the group size affects the finite-sample behavior of Theorem 1. To do so, we still assume the group interference, but this time we work with a different set of exposure



(a) $n = 2000$, histogram (b) $n = 3000$, histogram (c) $n = 5000$, histogram



(d) $n = 2000$, Q-Q normal plot (e) $n = 3000$, Q-Q normal plot (f) $n = 5000$, Q-Q normal plot

Figure 4: Histograms and Q-Q normal plots of $\frac{\sqrt{n}(\hat{\tau}^{k,k'} - \tau^{k,k'})}{\mathrm{Var}(\sqrt{n}\hat{\tau}^{k,k'})^{1/2}}$ for $n = 2000, 3000, 5000$

mappings:

$$f_i(w_{1:n}) = (w_i, u_i),$$

where $u_i = k$ if the fraction of treated neighbors is in $[\frac{k}{4}, \frac{k+1}{4})$ (for $k = 3$, the associated range is $[\frac{3}{4}, 1]$ to include 1). Hence, for each unit i, there are 8 possible exposure values: $(0,0), (0,1), (0,2), (0,3), (1,0), (1,1), (1,2), (1,3)$. We generate the potential outcomes according to $\mathcal{N}(w_i + 0.5u_i + 5, 1)$ and we are interested in the exposure contrast between $(1,3)$ and $(0,0)$. We ensure that each household has exactly the same size and compare two household sizes r: 4 and 8. Moreover, for the assignment mechanism, we use Bernoulli design where each unit is treated or not treated with probability one half. Again, it's easy to see that all the assumptions we need for Theorem 1 are satisfied. Table 8 shows the p-values for running Shapiro-Wilk test on normalized estimates with $r = 4$ and $r = 8$. As is shown in the table, the larger the household size the more samples we need for a good

Sample Size	n = 160	n = 480	n = 800	n = 1120	n = 1600	n = 3200
$r = 4$	2.14e-07	0.001	0.237	0.311	0.339	0.640
$r = 8$	5.59e-30	2.63e-10	1.72e-07	1.86e-05	6.08e-05	0.064

Table 8: p-values for running Shapiro-Wilk test on normalized estimates

Estimate of ϵ	$\hat{\epsilon}$	$1.5\hat{\epsilon}$	$2\hat{\epsilon}$	$2.5\hat{\epsilon}$	$3\hat{\epsilon}$
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 9: Root mean squared errors (RMSE) for $\hat{\tau}_{20}^{TE}$, $\hat{\tau}_{20}^{c}$ with $k = 2$ and $\hat{\tau}_{20}^{c}$ with $k = 5$

normal approximation. The intuition is that for larger households, each one's exposure value depends on more units' assignments.

D.2 Simulation results for estimation under stability assumption

D.2.1 Parameters for Erdős-Rényi Model

For the simulation study in Section 6.2.1, 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.2.2 The effect of estimated stability parameter

Recall that our $\hat{\epsilon}$ is only a lower bound of the true ϵ, hence may underestimate ϵ. To investigate how our estimate of ϵ 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 9 shows the results. We see that the convex combination type estimator with $k = 2$ is not sensitive to the estimate of ϵ while the convex combination type estimator

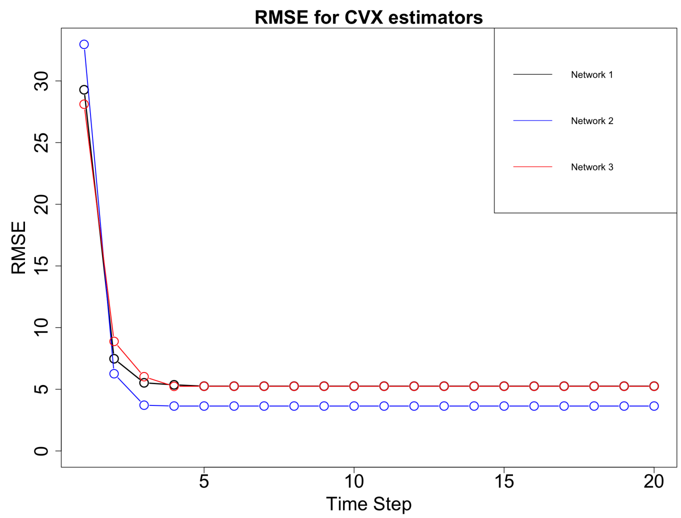


Figure 5: Root mean squared errors (RMSE) for $\hat{\tau}_{20}^{TE}$ and $\hat{\tau}^{c}$

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.

D.2.3 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.2.4 Lengths of approximate confidence intervals

Table 10 shows the average lengths of approximate confidence intervals. As expected, Gaussian confidence intervals are shorter.

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

Table 10: Lengths of two approximate confidence intervals for τ_t^{TE} with $k = 2$