

Incentive-Aware Synthetic Control: Accurate Counterfactual Estimation via Incentivized Exploration

Daniel Ngo^{*1}, Keegan Harris^{*2}, Anish Agarwal³, Vasilis Syrgkanis⁴, and Zhiwei Steven Wu²

¹University of Minnesota

²Carnegie Mellon University

³Columbia University

⁴Stanford University

ngo00054@umn.edu, {keeganh, zstevenwu}@cmu.edu,
aa5194@columbia.edu, vsyrgk@stanford.edu

Abstract

We consider the classic *panel data* setting in which one observes measurements of *units* over *time*, under different *interventions*. Our focus is on the canonical family of *synthetic control methods (SCMs)* which, after a pre-intervention time period when all units are under control, estimate counterfactual outcomes for *test* units in the post-intervention time period under control by using data from *donor* units who have remained under control for the entire post-intervention period. In order for the counterfactual estimate produced by synthetic control for a test unit to be accurate, there must be sufficient overlap between the outcomes of the donor units and the outcomes of the test unit. As a result, a canonical assumption in the literature on SCMs is that the outcomes for the test units lie within either the convex hull or the linear span of the outcomes for the donor units. However despite their ubiquity, such *overlap* assumptions may not always hold, as is the case when e.g. units select their own interventions and different subpopulations of units prefer different interventions *a priori*.

We shed light on this typically overlooked assumption, and we address this issue by *incentivizing* units with different preferences to take interventions they would not normally consider. Specifically, we provide a SCM for incentivizing exploration in panel data settings which provides incentive-compatible intervention recommendations to units by leveraging tools from information design and online learning. Using our algorithm, we show how to obtain valid counterfactual estimates using SCMs without the need for an explicit overlap assumption on the unit outcomes.

^{*}Denotes equal contribution.

Contents

1	Introduction	1
1.1	Related Work	3
2	Setting and Background	4
2.1	Our Panel Data Setting	4
2.2	Background on PCR	6
2.3	Recommendations and Beliefs	7
3	Incentivized Exploration for Synthetic Control	8
3.1	On the Necessity of the Unit Overlap Assumption	10
4	Extension to Synthetic Interventions	11
5	Testing Whether the Unit Overlap Assumption Holds	13
5.1	A Non-Asymptotic Hypothesis Test	13
5.2	An Asymptotic Hypothesis Test	14
6	Numerical Simulations	15
7	Conclusion	17
A	Appendix for Section 3: Incentivized Exploration for Synthetic Control	22
A.1	Causal Parameter Recovery	22
A.2	Proof of Theorem 3.2	24
B	Appendix for Section 4: Extension to Synthetic Interventions	27
C	Appendix for Section 5: Testing Whether the Unit Overlap Assumption Holds	29

1 Introduction

A ubiquitous task in statistics, machine learning, and econometrics is to estimate counterfactual outcomes for a group of *units* (e.g. people, geographic regions, subpopulations) under different *interventions* (e.g. medical treatments, weather patterns, legal regulations) over *time*. Such multi-dimensional data are often referred to as *panel data* (or *longitudinal data*), where the different units may be thought of as rows of a matrix, and the time-steps as columns. A prominent framework for counterfactual inference using panel data is that of *synthetic control* [1, 2]. Synthetic control methods (SCMs) assume access to a *pre-intervention* time period, during which all units are under *control* (i.e. no treatment). After the pre-intervention time period, every unit is given exactly one intervention from a set of possible interventions (which can include the control) and remains under the intervention for the remaining time-steps (i.e. the *post-intervention* time period). In order to estimate unit-specific counterfactuals under control, SCMs use the pre-intervention time period to learn a model to predict the outcomes for the test unit from the outcomes of the units who remained under control (i.e. the *donor* units). Once the model is learned, it is then extrapolated to the post-intervention time period in order to predict the counterfactual outcome for the test unit, had they remained under control. Since first being introduced in the field of economics over two decades ago, SCMs have become a popular tool for counterfactual inference and are routinely used across a variety of domains ranging from public policy [36, 19] to big tech [44, 17]. Additionally, the synthetic control framework has been extended to estimate counterfactual outcomes under different treatments, in addition to control, via the *synthetic interventions* framework [4].

In order for the counterfactual estimate produced by SCMs to be valid, it should be the case that the test unit’s potential outcomes may be expressed “reasonably well” by the observed outcomes of the donor units. When providing statistical guarantees on the performance of SCMs, such intuition has traditionally been made formal by making an *overlap* assumption on the relationship between the donor units and the test unit, for instance of the following form:

Unit Overlap Assumption. *Denote the outcome for unit i under intervention d at time t by $y_{i,t}^{(d)} \in \mathbb{R}$. For a given unit i and intervention d , there exists a set of weights $\omega^{(i,d)} \in \mathbb{R}^{N_d}$ such that*

$$\mathbb{E}[y_{i,t}^{(d)}] = \sum_{j \in [N_d]} \omega^{(i,d)}[j] \cdot \mathbb{E}[y_{j,t}^{(d)}]$$

for all $t \in [T]$, where T is the number of time-steps, N_d is the number of donor units who have received intervention d , and the expectation is taken with respect to the randomness in the unit outcomes.

More broadly, previous work which provides finite sample guarantees for SCMs assumes that there exists some underlying mapping $\omega^{(i,d)}$ (e.g. linear or convex) through which the outcomes of the test unit (unit i) may be expressed by the outcomes of the N_d donor units. Since such a condition appears to be necessary in order to do valid counterfactual inference, assumptions of this nature have become ubiquitous when proving statistical guarantees about SCMs (see, e.g. [2, 6, 4, 5]).¹

However despite their ubiquity, such overlap assumptions may not hold in all domains in which one would like to apply SCMs. For example, consider a streaming service with two service plans: a yearly subscription and a pay-as-you-go model, and suppose that the streaming service wants to determine the effectiveness of its subscription program (the treatment) on user engagement. Under this setting, the subpopulation of streamers who self-select the subscription plan are most likely those who believe they will consume large amounts of content on the platform. In contrast, those who pay-as-they-go most likely believe they will consume less content. This makes drawing

¹Similar assumptions are also prevalent in the literature on other *matching-style* estimators typically used in panel data settings, such as *difference-in-differences* [15] or *clustering-based* methods [46].

conclusions about the counterfactual user engagement levels of the two subpopulations under different business plans difficult, as they may have very different experiences under the two business plans due to their differing tastes. While the streaming service would ideally like to run a randomized controlled trial (RCT) in order to estimate counterfactual engagement levels across different groups, participation in RCTs is voluntary for ethical and legal reasons, and so ensuring compliance is generally not possible.

In this work, our goal is to leverage tools from information design in order to incentivize the exploration of different treatments by non-overlapping unit subpopulations in order to obtain valid counterfactual estimates using synthetic control methods. This is possible because the *principal* (i.e. the person/platform running the synthetic control method) will have access to more information about counterfactual unit outcomes than any one unit in isolation, due to the fact that they get to observe the outcome trajectories of all units using the platform. Specifically, we adopt tools and techniques from the literature on incentivizing exploration in multi-armed bandits (e.g. [31, 33, 41]) to show how the principal can leverage knowledge gained from previous interactions with similar units to persuade the current unit to take an intervention they would not normally take. The principal does this sending a *signal* (or *recommendation*) to the unit with information about which intervention is best for them. The principal’s *recommendation policy* is designed so that it is *incentive-compatible*, i.e. it is in the unit’s best-interests to follow the intervention recommended to them by the policy. In our streaming service example, incentive-compatible signaling may correspond to recommending a service plan for each user based on their usage of the platform, in a way that guarantees that units are better off in expectation when purchasing the recommended service plan. Our procedure ensures that the [Unit Overlap Assumption](#) becomes satisfied over time, which enables the principal to do valid counterfactual inference using off-the-shelf SCMs after they have interacted with sufficiently many units.

Overview of Our Results We overview related work on learning from panel data, synthetic control methods, incentivized exploration in bandits, and algorithmic (Bayesian) persuasion in Section 1.1. In Section 2 we introduce our model and provide relevant technical background on synthetic control methods.

We introduce our algorithm for incentivizing exploration for synthetic control when there are two interventions (treatment and control) in Section 3 (Algorithm 1). At a high level, we adapt the “hidden exploration” paradigm from incentivized exploration in bandits to the panel data setting: First, we randomly divide units into “exploit” units and “explore” units. For all exploit units, we recommend the intervention which we estimate maximizes their expected utility, given the data we have seen so far. For every explore unit, we recommend the intervention for which we would like the [Unit Overlap Assumption](#) to be satisfied. Units are not aware of their exploit/explore designation by the algorithm, and the explore probability is chosen to be low enough such that the units have an incentive to follow the principal’s recommendations. After Algorithm 1 has been used to provide recommendations to sufficiently-many units, we can guarantee that with high probability, the [Unit Overlap Assumption](#) is satisfied for all units for either intervention of our choosing. This enables us to use existing synthetic control methods off-the-shelf in order to obtain finite sample guarantees for counterfactual estimation for all units under control after running Algorithm 1. Along the way, we draw a conceptual connection between decision-making using panel data and linear contextual bandits, which has not been previously observed in the literature to the best of our knowledge. In particular, we show that under the setting of *robust synthetic control* [6, 7, 3], the problem of assigning interventions to units based on their pre-intervention outcomes may be thought of as a generalization of the linear contextual bandit problem *in which there is noise in the observed context*. While we leverage this conceptual connection to design algorithms for incentivizing exploration in panel data settings, we hope that it will aid in the design of contextual bandit algorithms which are more robust to measurement error in the observed context.

We show that the **Unit Overlap Assumption** is indeed a necessary condition to obtain valid counterfactual estimates in our setting in Section 3.1. In Section 4, we extend Algorithm 1 to the setting of synthetic interventions [4], in which the principal may wish to estimate counterfactual outcomes under different treatments, in addition to control. We provide a hypothesis test for checking whether the **Unit Overlap Assumption** holds for a given test unit and set of donor units in Section 5, and we empirically evaluate the performance of our incentive-aware synthetic control estimator in Section 6. When the Section 1 does not hold *a priori*, we show that our methods enable consistent counterfactual outcome estimation, while existing SCMs generally do not.

1.1 Related Work

Causal Inference and Synthetic Control Methods Popular methods for counterfactual inference using panel data include synthetic control methods [1, 2], difference-in-differences [8, 10, 12], and clustering-based methods [46, 16]. Within the literature on synthetic control, our work builds off of the line of work on *robust* synthetic control [6, 7, 3, 4, 5], which assumes outcomes are generated via a *latent factor model* (e.g. [13, 29, 9]) and leverages *principal component regression* (PCR) [23, 35] to estimate unit counterfactual outcomes. Our work falls in the small-but-growing line of work at the intersection of synthetic control methods and online learning [14, 18, 5], although we are the first to consider unit incentives in this setting. Particularly relevant to our work is the model used by Agarwal et al. [5], which extends the finite sample guarantees from PCR in the panel data settings to online settings. While Harris et al. [21] also consider incentives in synthetic control methods (albeit in an offline setting), they consider a principal who can *assign* interventions to units (e.g. can force compliance). As a result, the strategizing they consider is that of units who modify their pre-intervention outcomes in order to be assigned a more desirable intervention. In contrast, we consider a principal who cannot assign interventions to units, but instead must *persuade* units to take different interventions by providing them with incentive-compatible recommendations.

More broadly, there has been recent interest in causal inference on characterizing optimal treatment policies which must satisfy some additional constraint(s). For example, Luedtke and van der Laan [30] study a setting in which the treatment supply is limited. Much like us, Qiu et al. [37] consider a setting in which intervening on the treatment is not possible, but encouraging treatment is feasible. However their focus is on the setting in which the treatment is a limited resource, so their goal is to encourage treatment to those who would benefit from it the most. Since they do not consider panel data and place different behavioral assumptions on the individuals under intervention, the tools and techniques we use to persuade units in our setting differ significantly from theirs. Finally, Qiu et al. [38], Sun et al. [43] consider settings in which there is some uncertain *cost* associated with treatment.

Bayesian Persuasion (BP) and Incentivized Exploration (IE) Bayesian Persuasion [25, 24] is a popular model of information design [11], a branch of theoretical economics which aims to characterize optimal communication between strategic individuals under information asymmetry. In its simplest form, BP is a game between two players: an informed *sender* and an uninformed *receiver*. After observing a payoff-relevant state, the sender sends a *signal* to the receiver, who then takes a payoff-relevant action. The goal of the sender is to design a *signaling policy* which (probabilistically) reveals information about the state in order to incentivize the receiver to take desirable actions.

Our work draws on techniques from the growing literature on incentivizing exploration [26, 32, 34, 40, 22, 39], which falls at the intersection of BP and multi-armed bandits [42, 27]. In IE, a principal interacts with a sequence of myopic agents over time. Each agent takes an action, and the principal can observe the outcome of each action. While each individual agent would prefer to take the action which appears to be utility-maximizing (i.e. to *exploit*), the principal would like to incentivize agents to *explore* different actions in order to benefit the population

in aggregate. Motivation for IE includes online rating platforms and recommendation systems which rely on information collected by users to provide informed recommendations about various products and services. Incentivized exploration may be viewed as a generalization of BP, as each round of IE is an instance of a one-shot BP game between the principal and a myopic agent. Following the framework of Kremer et al. [26], we consider the recommendations given by the principal to be the only incentive for participating units. More precisely, in our model, the principal does not offer monetary payment for units to choose an intervention, which has known disadvantages such as potential selection bias and ethical concerns [20].

Within the literature on IE, the work most related to ours on a conceptual level is that of Li and Slivkins [28], who consider the problem of incentivizing exploration in clinical trial settings. Like us, Li and Slivkins [28] study a setting in which the principal would like to incentivize agents to explore different treatments. However their goal is to estimate *population-level* statistics about each intervention, while we are interested in estimating *unit-specific* counterfactuals under different interventions using panel data. As a result, while our high-level motivations are somewhat similar, the tools and techniques we use to obtain our results differ significantly. On a technical level, our mechanisms are somewhat similar to the initial exploration phase in Mansour et al. [32], although we consider a more general setting where unit outcomes may vary over time, in contrast to the simpler multi-armed bandit setting they consider.

2 Setting and Background

Notation Subscripts are used to index the unit and time-step, while superscripts are reserved for interventions. We use i to index units, t to index time-steps, and d to index interventions. For $x \in \mathbb{N}$, we use the shorthand $\llbracket x \rrbracket := \{1, 2, \dots, x\}$ and $\llbracket x \rrbracket_0 := \{0, 1, \dots, x-1\}$. $\mathbf{y}[i]$ denotes the i -th component of vector \mathbf{y} , where indexing starts at 1. We sometimes use the shorthand $T_1 := T - T_0$ for $T, T_0 \in \mathbb{N}_{>0}$ such that $T > T_0$. $\Delta(\mathcal{X})$ denotes the set of possible probability distributions over \mathcal{X} . $\mathbf{0}_d$ is shorthand for the vector $[0, 0, \dots, 0] \in \mathbb{R}^d$. Finally, we use the notation $a \wedge b := \min\{a, b\}$ and $a \vee b := \max\{a, b\}$.

2.1 Our Panel Data Setting

We consider a setting in which the principal interacts with a sequence of n units for T time steps each. We assume that there is a pre-intervention period of T_0 time-steps, for which each unit is under the same intervention, i.e., under *control*. After the pre-intervention period, the principal either recommends the *treatment* to the unit or suggests that they remain under control. We denote the sender’s *recommended* intervention to unit i by $\hat{d}_i \in \{0, 1\}$, where 1 denotes the treatment and 0 the control. After receiving the recommendation, unit i chooses an intervention $d_i \in \{0, 1\}$ and remains under intervention d_i for the remaining $T - T_0$ time-steps. We use

$$\mathbf{y}_{i,pre} := [y_{i,1}^{(0)}, \dots, y_{i,T_0}^{(0)}]^\top \in \mathbb{R}^{T_0}$$

to denote unit i ’s pre-treatment outcomes under control, and

$$\mathbf{y}_{i,post}^{(d)} := [y_{i,T_0+1}^{(d)}, \dots, y_{i,T}^{(d)}]^\top \in \mathbb{R}^{T-T_0}$$

to refer to unit i ’s post-intervention outcomes under intervention d . We denote the set of possible pre-treatment outcomes by \mathcal{Y}_{pre} . In order to impose structure on how unit outcomes are related for different units and time-steps, we assume that outcomes are generated via the following *latent factor model*, a popular assumption in the literature (see references in Section 1.1).

Assumption 2.1 (Latent Factor Model). *Suppose the outcome for unit i at time t under treatment $d \in \{0, 1\}$ takes the following factorized form*

$$\begin{aligned}\mathbb{E}[y_{i,t}^{(d)}] &= \langle \mathbf{u}_t^{(d)}, \mathbf{v}_i \rangle \\ y_{i,t}^{(d)} &= \mathbb{E}[y_{i,t}^{(d)}] + \varepsilon_{i,t}^{(d)}\end{aligned}$$

where $\mathbf{u}_t^{(d)} \in \mathbb{R}^r$ is a latent vector which depends only on the time-step t and intervention d , $\mathbf{v}_i \in \mathbb{R}^r$ is a latent vector which only depends on unit i , and $\varepsilon_{i,t}^{(d)}$ is zero-mean sub-Gaussian random noise with variance at most σ^2 . For simplicity, we assume that $|\mathbb{E}[y_{i,t}^{(d)}]| \leq 1$, $\forall i \in [n], t \in [T], d \in \{0, 1\}$.

While we assume the existence of such structure in the data, we do *not* assume that the principal knows or gets to observe $\mathbf{u}_t^{(d)}$, \mathbf{v}_i , or $\varepsilon_{i,t}^{(d)}$. In the literature on synthetic control, the goal of the principal is to estimate unit-specific counterfactual outcomes under different interventions. Specifically, our target causal parameter is the counterfactual average expected post-intervention outcome.

Definition 2.2. (Average expected post-intervention outcome) *The average expected post-intervention outcome of unit i under intervention d is*

$$\mathbb{E}[\bar{y}_{i,\text{post}}^{(d)}] := \frac{1}{T_1} \sum_{t=T_0+1}^T \mathbb{E}[y_{i,t}^{(d)}]$$

where the expectation is taken with respect to $(\varepsilon_{i,t}^{(d)})_{T_0 < t \leq T}$.

In order to infer something about unit outcomes in the post-intervention time period from outcomes in the pre-intervention time period, we require the following *linear span inclusion* assumption on the latent factors in the post-intervention time period.²

Assumption 2.3 (Linear Span Inclusion). *For each intervention $d \in \{0, 1\}$ and time $t > T_0$, we assume that*

$$\mathbf{u}_t^{(d)} \in \text{span}\{\mathbf{u}_1^{(0)}, \mathbf{u}_2^{(0)}, \dots, \mathbf{u}_{T_0}^{(0)}\}.$$

Assumption 2.3 is ubiquitous in the literature on robust synthetic control (see, e.g. [4, 5, 21]). Under Assumptions 2.1 and 2.3, previous work (e.g. [5, 21]) observes that the average expected post-treatment outcome for any unit i may be written as a linear combination of the expected pre-intervention outcomes of unit i . The following is an important structural result for our main algorithm, as it will allow us to transform the problem of estimating unit counterfactual outcomes from a regression over donor units to a regression over pre-intervention outcomes.

Proposition 2.4 (Average expected post-intervention outcome reformulation). *Under Assumption 2.1 and 2.3, there exists a slope vector $\boldsymbol{\theta}^{(d)} \in \mathbb{R}^{T_0}$ such that the average expected post-intervention outcome of unit i under intervention d is given by*

$$\mathbb{E}[\bar{y}_{i,\text{post}}^{(d)}] = \frac{1}{T_1} \langle \boldsymbol{\theta}^{(d)}, \mathbb{E}[y_{i,\text{pre}}] \rangle.$$

We assume that the principal has knowledge of a valid upper-bound D on the ℓ_2 -norm of $\boldsymbol{\theta}^{(d)}$, i.e. $\|\boldsymbol{\theta}^{(d)}\|_2 \leq \Gamma$ for $d \in \{0, 1\}$. While we are not the first to make this observation, we are the first to leverage it to design algorithms for assigning interventions based on ideas from the contextual

²Consider the limiting case in which $\mathbf{u}_t^{(0)} = \mathbf{0}_r$ for all $t \leq T_0$. Under such a setting, all expected unit outcomes in the pre-intervention time period will be 0, regardless of the underlying unit latent factors.

bandit literature.³ In particular, it is sometimes helpful to think of $\mathbb{E}[y_{i,pre}]$ as the “context” of unit i , and $\mathbb{E}[\bar{y}_{i,post}^{(d)}]$ as the principal’s expected reward of assigning intervention d given $\mathbb{E}[y_{i,pre}]$. Since we observe $y_{i,pre}$ instead of $\mathbb{E}[y_{i,pre}]$ we cannot apply linear contextual bandit algorithms to our panel data setting out-of-the-box; however as we will show in Section 3, one can combine ideas from incentivizing exploration in contextual bandits with principled ways of handling the noise in the pre-intervention outcomes to incentivize exploration in panel data settings.

2.2 Background on PCR

Given historical data of the form $\{\mathbf{y}_{i,pre}, \mathbf{y}_{i,post}^{(d_i)}\}_{j=1}^{i-1}$, our goal is to estimate $\boldsymbol{\theta}^{(d)}$ as $\hat{\boldsymbol{\theta}}_i^{(d)}$ for $d \in \{0, 1\}$ via PCR [23, 35]. Let $\mathcal{I}_i^{(d)}$ be the set of units who have received intervention d before unit i arrives, and let $n_i^{(d)}$ be the number of such units. We use

$$Y_{pre,i}^{(d)} := [\mathbf{y}_{j,pre}^\top : j \in \mathcal{I}_i^{(d)}] \in \mathbb{R}^{n_i^{(d)} \times T_0}$$

to denote the matrix of pre-treatment outcomes corresponding to the subset of units who have undergone intervention d before unit i arrives, and

$$Y_{post,i}^{(d)} := \left[\sum_{t=T_0+1}^T y_{j,t}^{(d)} : j \in \mathcal{I}_i^{(d)} \right] \in \mathbb{R}^{n_i^{(d)} \times 1}$$

be the vector of the sum of post-intervention outcomes for the subset of units who have undergone intervention d before unit i arrives. We denote the singular value decomposition of $Y_{pre,i}^{(d)}$ as

$$Y_{pre,i}^{(d)} = \sum_{\ell=1}^{n_i^{(d)} \wedge T_0} s_\ell^{(d)} \hat{\mathbf{u}}_\ell^{(d)} (\hat{\mathbf{v}}_\ell^{(d)})^\top,$$

where $\{s_\ell^{(d)}\}_{\ell=1}^{n_i^{(d)} \wedge T_0}$ are the singular values of $Y_{pre,i}^{(d)}$, and $\hat{\mathbf{u}}_\ell^{(d)}$ and $\hat{\mathbf{v}}_\ell^{(d)}$ are orthonormal column vectors. We assume that the singular values are ordered such that $s_1(Y_{pre,i}^{(d)}) \geq \dots \geq s_{n_i^{(d)} \wedge T_0}(Y_{pre,i}^{(d)}) \geq 0$. For some threshold value r , we use

$$\hat{Y}_{pre,i}^{(d)} := \sum_{\ell=1}^r s_\ell^{(d)} \hat{\mathbf{u}}_\ell^{(d)} (\hat{\mathbf{v}}_\ell^{(d)})^\top$$

to refer to the truncation of $Y_{pre,i}^{(d)}$ to its top r singular values. We define the projection matrix onto the subspace spanned by the top r right singular vectors as $\hat{\mathbf{P}}_{i,r}^{(d)} \in \mathbb{R}^{r \times r}$, given by $\hat{\mathbf{P}}_{i,r}^{(d)} := \sum_{\ell=1}^r \hat{\mathbf{v}}_\ell^{(d)} (\hat{\mathbf{v}}_\ell^{(d)})^\top$. Equipped with this notation, we are now ready to define the procedure for estimating $\boldsymbol{\theta}^{(d)}$ using (regularized) principal component regression.

Definition 2.5 (Regularized Principal Component Regression). *Given regularization parameter $\rho \geq 0$ and truncation level $r \in \mathbb{N}$, for $d \in \{0, 1\}$ and $i \geq 1$, let $\mathcal{V}_i^{(d)} := \left(\hat{Y}_{pre,i}^{(d)} \right)^\top \hat{Y}_{pre,i}^{(d)} + \rho \hat{\mathbf{P}}_{i,r}^{(d)}$. Then, regularized PCR estimates $\boldsymbol{\theta}^{(d)}$ as*

$$\hat{\boldsymbol{\theta}}_i^{(d)} := \left(\mathcal{V}_i^{(d)} \right)^{-1} \hat{Y}_{pre,i}^{(d)} Y_{i,post}^{(d)},$$

where $\boldsymbol{\theta}^{(d)}$ is defined as in Proposition 2.4. The average post-intervention outcome for unit i under intervention d may then be estimated as $\hat{\mathbb{E}}[\bar{y}_{i,post}^{(d)}] := \frac{1}{T_1} \langle \hat{\boldsymbol{\theta}}_i^{(d)}, y_{i,pre} \rangle$.

³For the reader unfamiliar with the linear contextual bandit setting, the remainder of this paragraph may be skipped without any loss in continuity.

Under the [Unit Overlap Assumption](#) and Assumption 2.3, work on robust synthetic control (see Section 1.1) uses (regularized) PCR to obtain consistent estimates for counterfactual post-intervention outcomes under different interventions.

2.3 Recommendations and Beliefs

The interaction between the principal and a unit i may be characterized by the tuple $(\mathbf{y}_{i,pre}, \hat{d}_i, d_i, \mathbf{y}_{i,post}^{(d_i)})$, where $\mathbf{y}_{i,pre}$ are unit i 's pre-treatment outcomes, \hat{d}_i is the intervention recommended to unit i , d_i is the intervention taken by unit i , and $\mathbf{y}_{i,post}^{(d_i)}$ are unit i 's post-intervention outcomes under intervention d_i .

Definition 2.6 (Interaction History). *The interaction history at unit i is the sequence of outcomes, recommendations, and interventions for all units $j \in \llbracket i - 1 \rrbracket$. Formally,*

$$H_i := \{(\mathbf{y}_{j,pre}, \hat{d}_j, d_j, \mathbf{y}_{j,post}^{(d_j)})\}_{j=1}^{i-1}.$$

We denote the set of all possible histories at unit i as \mathcal{H}_i .

Definition 2.7 (Recommendation Policy). *A recommendation policy $\pi_i : \mathcal{H}_i \times \mathcal{Y}_{pre} \rightarrow \Delta(\{0, 1\})$ is a (possibly stochastic) mapping from histories and pre-treatment outcomes to interventions.*

We assume that before the first unit arrives, the principal commits to a sequence of recommendation policies $\{\pi_i\}_{i=1}^n$ which are fully known to all units. Whenever π is clear from the context, we use the shorthand $\hat{d}_i = \pi_i(\mathbf{y}_{i,pre})$ to denote the recommendation of policy π_i to unit i .

In addition to having a corresponding latent factor, each unit has an associated *belief* over the effectiveness of each intervention. This is made formal through the following definitions.

Definition 2.8 (Unit Belief). *Unit i has prior belief $\mathcal{P}_{\mathbf{v}_i}$, which is a joint distribution over potential post-intervention outcomes $\{\mathbb{E}[\mathbf{y}_{i,post}^{(0)}], \mathbb{E}[\mathbf{y}_{i,post}^{(1)}]\}$. We use the shorthand $\mu_{v_i}^{(d)} := \mathbb{E}_{\mathcal{P}_{\mathbf{v}_i}}[\bar{y}_{i,post}^{(d)}]$ to refer to a unit's expected average post-intervention outcome, with respect to their prior $\mathcal{P}_{\mathbf{v}_i}$.*

Definition 2.9 (Unit Type). *Unit i is of type $\tau \in \{0, 1\}$ if $\tau = \arg \max_{d \in \{0, 1\}} \mu_{v_i}^{(d)}$. We denote the set of all units of type τ as $\mathcal{T}^{(\tau)}$ and the dimension of the latent subspace spanned by type τ units as $r_\tau := \text{span}(\mathbf{v}_j : j \in \mathcal{T}^{(\tau)})$.*

In other words, a unit's type is the intervention they would take according to their prior without any additional information about the effectiveness of each intervention. We consider the setting in which the possible latent factors associated with each type lie in *mutually orthogonal* subspaces (i.e. [Unit Overlap Assumption](#) is *not* satisfied).⁴⁵

As is standard in the literature on IE, we assume that units are Bayesian-rational and each unit knows its place in the sequence of n units (i.e., their index $i \in \llbracket n \rrbracket$). Thus given recommendation \hat{d}_i , unit i selects their intervention d_i such that $d_i \in \arg \max_{d \in \{0, 1\}} \mathbb{E}_{\mathcal{P}_{\mathbf{v}_i}}[\bar{y}_{i,post}^{(d)} | \hat{d}_i]$, i.e. they select the intervention d_i which maximizes their utility in expectation over their prior $\mathcal{P}_{\mathbf{v}_i}$, conditioned on receiving recommendation \hat{d}_i from the principal.

Definition 2.10 (Bayesian incentive-compatibility). *We say that a recommendation d is Bayesian incentive-compatible (BIC) for unit i if, conditional on receiving intervention recommendation $\hat{d}_i = d$, unit i 's average expected post-intervention outcome under intervention d is at least as large as their average expected post-intervention outcome under any other intervention:*

$$\mathbb{E}_{\mathcal{P}_{\mathbf{v}_i}}[\bar{y}_{i,post}^{(d)} - \bar{y}_{i,post}^{(d')} | \hat{d}_i = d] \geq 0 \quad \text{for every } d' \in \{0, 1\}.$$

A recommendation policy π is BIC if the above condition holds for every intervention d which is recommended by π with positive probability.

⁴⁵If the [Unit Overlap Assumption](#) is known to be satisfied, units do not need to be incentivized to explore, and thus existing synthetic control methods may be used off-the-shelf.

⁵Note that $\sum_\tau r_\tau \leq r$.

3 Incentivized Exploration for Synthetic Control

In this section, we turn our focus to incentivizing type 1 units to remain under control in the post-intervention period. The methods we present may also be applied to incentivize type 0 units to take the treatment. We focus on incentivizing control amongst type 1 units in order to be in line with the literature on synthetic control, which aims to estimate counterfactual unit outcomes under control.

The goal of the principal is to design a recommendation policy that convinces enough units of type 1 to select the control in the post-treatment period, such that the [Unit Overlap Assumption](#) is satisfied for all units of type 1. Our algorithm (Algorithm 1) is inspired by the “detail free” algorithm for incentivizing exploration in multi-armed bandits in Mansour et al. [31]. Furthermore through our discussion in Section 2, one may view Algorithm 1 as an algorithm for incentivizing exploration in linear contextual bandit settings with measurement error in the context.

The recommendation policy in Algorithm 1 is split into two stages. In the first stage, the principal provides no recommendations to the first N_0 units. Left to their own devices, these units take their preferred intervention according to their prior belief: type 0 units take control, and type 1 units take the treatment. We choose N_0 such that it is large enough for the [Unit Overlap Assumption](#) to be satisfied for all units of type 1 under the treatment with high probability.

In the second stage, we use the initial data collected during the first stage to construct an estimator of the average expected post-intervention outcome for type 1 units under treatment using PCR, as described in Section 2.2. At a high level, in order to incentivize a type 1 unit i to try the control, we assume that (1) there is a non-zero chance under unit i ’s prior $\mathcal{P}_{\mathbf{v}_i}$ that $\bar{y}_{i,post}^{(0)} \geq \bar{y}_{i,post}^{(1)}$ and (2) the principal will be able to infer this given the set of observed outcomes for units in the first phase. (This is made formal in Assumption 3.1 below.) Such “fighting chance” assumptions are common (and oftentimes necessary) in the literature on IE.

By dividing the total number of units in the second stage into phases of L rounds each, the principal can randomly “hide” one *explore* recommendation amongst $L - 1$ *exploit* recommendations. When the principal sends an exploit recommendation to unit i , they recommend the intervention which would result in the highest expected average post-intervention outcome for unit i , conditional on the observed outcomes collected during the first stage. On the other hand, the principal recommends that unit i take the control whenever they send an explore recommendation. Thus under Algorithm 1, if a type 1 unit receives a recommendation to take the treatment, they will always follow the recommendation since they can infer they must have received an exploit recommendation. However if a type 1 unit receives a recommendation to take the control, they will be unsure if they have received an explore or an exploit intervention, and can therefore be incentivized to follow the recommendation as long as L is set to be large enough (i.e. the probability of the recommendation being an explore recommendation is sufficiently low).

Our algorithm requires the following assumption on the units’ knowledge about the interaction:

Assumption 3.1. *[Unit Knowledge for Synthetic Control] We assume that the following are common knowledge among all units and the principal:*

1. *Valid upper- and lower-bounds on the fraction of type 1 units in the population, i.e. if the true proportion of type 1 units is $p_1 \in (0, 1)$, the principal and units know $p_L, p_H \in (0, 1)$ such that $p_L \leq p_1 \leq p_H$.*
2. *Valid upper- and lower-bounds on the prior mean for each receiver type and intervention, i.e. values $\mu^{(d,\tau,h)}, \mu^{(d,\tau,l)} \in \mathbb{R}$ such that $\mu^{(d,\tau,l)} \leq \mu_{v_i}^{(d)} \leq \mu^{(d,\tau,h)}$ for all units $i \in \llbracket N \rrbracket$, receiver types $\tau \in \{0, 1\}$, and interventions $d \in \{0, 1\}$.*
3. *For some $C \in (0, 1)$, a lower bound on the smallest probability of the event $\xi_{C,i}$ over the priors and latent factors of type 1 units, denoted by $\min_{i \in \mathcal{T}^{(1)}} \Pr_{\mathcal{P}_i}[\xi_{C,i}] \geq \zeta_C > 0$, where*

$$\xi_{C,i} := \left\{ \mu_i^{(0)} \geq \bar{y}_{i,post}^{(1)} + C \right\}$$

ALGORITHM 1: Incentivizing Exploration for Synthetic Control: Type 1 units

Input: First stage length N_0 , batch size L , number of batches B , failure probability δ , gap $C \in (0, 1)$
Provide no recommendation to first N_0 units.

```
for batch  $b = 1, 2, \dots, B$  do
  Select an explore index  $i_b \in [L]$  uniformly at random.
  for  $j = 1, 2, \dots, L$  do
    if  $j = i_b$  then
      Recommend intervention  $\hat{d}_{N_0+(b-1) \cdot L+j} = 0$  to unit  $N_0 + (b-1) \cdot L + j$ .
    else
      if  $\mu^{(0,1,L)} - \widehat{\mathbb{E}}[\bar{y}_{j,post}^{(1)}] \geq C$  then
        Recommend intervention  $\hat{d}_{N_0+(b-1) \cdot L+j} = 0$  to unit  $N_0 + (b-1) \cdot L + j$ .
      else
        Recommend intervention  $\hat{d}_{N_0+(b-1) \cdot L+j} = 1$  to unit  $N_0 + (b-1) \cdot L + j$ .
      end
    end
  end
end
end
```

where $\bar{y}_{i,post}^{(1)}$ is the average post-intervention outcome for unit i under intervention 1.

4. A sufficient number of observations N_0 needed such that the *Unit Overlap Assumption* is satisfied for type τ units under intervention τ with probability at least $1 - \delta$, for $\tau \in \{0, 1\}$ and some $\delta \in (0, 1)$.

Assumption 3.1 posits that bounds on various aggregate statistics about the underlying unit population are common knowledge across all units. We are now ready to present our main result: a way to initialize Algorithm 1 to ensure that *Unit Overlap Assumption* is satisfied with high probability.

Theorem 3.2 (Informal; detailed version in Theorem A.4). *Suppose that Assumption 3.1 holds for some constant gap $C \in (0, 1)$. If the number of initial units N_0 is chosen to be large enough such that *Unit Overlap Assumption* is satisfied for all units of type 1 under treatment with high probability, then Algorithm 1 is BIC if the batch size L is chosen to be sufficiently large.*

*Moreover, if the number of batches B is chosen to be sufficiently large, then the *Unit Overlap Assumption* will be satisfied simultaneously for all type 1 units under control with high probability.*

Proof Sketch. See Appendix A for complete proof details. At a high level, the proof follows by expressing the compliance condition for type 1 units as different cases depending on the principal's recommendation. In particular, a type 1 unit could receive recommendation $\hat{d}_i = 0$ for two reasons: (1) Under event $\hat{\xi}_{C,i}$ when control is indeed the better intervention according to the unit prior and the observed outcomes of previous units, or (2) when the unit is randomly selected as an explore unit. Using the probabilities of these two events occurring, we can derive a condition on the minimum phase length L such that the expected gain from exploiting (when the event $\hat{\xi}_{C,i}$ happens) exceeds the expected loss from exploring. We then further simplify the condition on the phase length L so that it is computable by the principal by leveraging existing finite sample guarantees for principal component regression using the samples collected in the first stage when no recommendations are given. \square

Theorem 3.2 says that after running Algorithm 1 with optimally-chosen parameters, the *Unit Overlap Assumption* will be satisfied for all type 1 units under control with probability at least $1 - 3\delta$. Therefore after running Algorithm 1 for a sufficiently long time, the principal can use off-the-shelf synthetic control methods (e.g. [4, 5]) to obtain counterfactual estimates with valid confidence intervals for all type 1 units under control with high probability. Consider the following concrete instantiation of Theorem 3.2.

Example 3.3. Consider a setting with two receiver types, each with equal probability: type 1 units prefer the treatment and type 0 units prefer control. Let $v_i \sim \text{Unif}[0.25, 0.75]$. If unit i is a type 1 unit, they have latent factor $\mathbf{v}_i = [v_i \ 0]$. Otherwise they have latent factor $\mathbf{v}_i = [0 \ v_i]$. Suppose that $T_0 = 2$ and $T_1 = 1$. We leave $\mathbf{u}_1^{(0)}, \mathbf{u}_2^{(0)}, \mathbf{u}_3^{(0)}, \mathbf{u}_3^{(1)}$ unspecified.

Since the dimension of the type 1 subpopulation in this setting is 1, we only need to incentivize a single type 1 unit to take the control in order for the [Unit Overlap Assumption](#) to be satisfied. Therefore, for a given confidence level δ , it suffices to set $N_0 = B = \frac{\log(1/\delta)}{\log(2)}$. After the first N_0 time-steps, the principal will know all of the parameters necessary to compute L .

Since $T_1 = 1$, the units' priors are only over $\{\mathbb{E}[y_{i,3}^{(0)}], \mathbb{E}[y_{i,3}^{(1)}]\}$. For simplicity, suppose that (1) all type 1 units believe that $\mathbb{E}[y_{i,3}^{(0)}] \sim \text{Unif}[0, 0.5]$ and $\mathbb{E}[y_{i,3}^{(1)}] \sim \text{Unif}[0, 1]$ and (2) there is no noise in the post-intervention outcome. Under this setting, event $\xi_{C,i}$ simplifies to $\xi_{C,i} = \{0.25 \leq \mathbb{E}[y_{i,3}^{(1)}] + C\}$ for any unit i and constant $C \in (0, 1)$. Therefore

$$\zeta_C = \Pr_{\mathcal{P}_i}[\mathbb{E}[y_{i,3}^{(1)}] \leq 0.25 - C] = \max\{0, 0.25 - C\}.$$

3.1 On the Necessity of the [Unit Overlap Assumption](#)

We conclude this section by showing that the [Unit Overlap Assumption](#) must be satisfied in order to obtain consistent counterfactual estimates under the setting described in Section 2.1.

Theorem 3.4. Fix a time horizon T , pre-intervention time period T_0 , and number of donor units $n^{(0)}$. For any algorithm used to estimate the average post-intervention outcome under control for a test unit, there exists a problem instance such that the produced estimate has constant error whenever the [Unit Overlap Assumption](#) is not satisfied for the test unit under control, even as $T, T_0, n^{(0)} \rightarrow \infty$.

Proof. Consider the setting in which all type 0 units have latent factor $\mathbf{v}_0 = [0 \ 1]^\top$ and all type 1 units have latent factor $\mathbf{v}_1 = [1 \ 0]^\top$. Suppose that $\mathbf{u}_t^{(0)} = [1 \ 0]^\top$ if $t \bmod 2 = 0$ and $\mathbf{u}_t^{(0)} = [0 \ 1]^\top$ if $t \bmod 2 = 1$ if $t \leq T_0$. For all $t > T_0$, let $\mathbf{u}_t^{(0)} = [H \ 1]^\top$ for some H in range $[-c, c]$, where $c > 0$. Furthermore, suppose that there is no noise in the unit outcomes. Under this setting, the (expected) outcomes for type 0 units under control are

$$\mathbb{E}[y_{0,t}^{(0)}] = \begin{cases} 0 & \text{if } t \bmod 2 = 0, t \leq T_0 \\ 1 & \text{if } t \bmod 2 = 1, t \leq T_0 \\ 1 & \text{if } t > T_0 \end{cases}$$

and the outcomes for type 1 units under control are

$$\mathbb{E}[y_{1,t}^{(0)}] = \begin{cases} 1 & \text{if } t \bmod 2 = 0, t \leq T_0 \\ 0 & \text{if } t \bmod 2 = 1, t \leq T_0 \\ H & \text{if } t > T_0. \end{cases}$$

Suppose that $H \sim \text{Unif}[-c, c]$ and suppose that the principal wants to estimate $\mathbb{E}[\bar{y}_{1,post}^{(0)}]$ using just the set of outcomes $\mathbb{E}[\mathbf{y}_{0,pre}]$, $\mathbb{E}[\mathbf{y}_{0,post}]$, and $\mathbb{E}[\mathbf{y}_{1,pre}]$. Since these outcomes do not contain any information about H , any estimator $\hat{\mathbb{E}}[\bar{y}_{1,post}^{(0)}]$ cannot be a function of H and thus will be at least a constant distance away from the true average post-intervention outcome $\mathbb{E}[\bar{y}_{1,post}^{(0)}]$ in expectation over H . That is,

$$\begin{aligned} \mathbb{E}_{H \sim \text{Unif}[-c, c]} \left| \mathbb{E}[\bar{y}_{1,post}^{(0)}] - \hat{\mathbb{E}}[\bar{y}_{1,post}^{(0)}] \right| &= \mathbb{E}_{H \sim \text{Unif}[-c, c]} \left| H - \hat{\mathbb{E}}[\bar{y}_{1,post}^{(0)}] \right| \\ &= \frac{1}{2c} (c^2 + (\hat{\mathbb{E}}[\bar{y}_{1,post}^{(0)}])^2) \geq \frac{c}{2}. \end{aligned}$$

Note that our choice of T and T_0 was arbitrary, and that estimation does not improve as the number of donor units increase since (1) there is no noise in the outcomes and (2) all type 0 units have the same latent factor. Therefore any estimator for $\mathbb{E}\bar{y}_{1,post}^{(0)}$ is inconsistent in expectation over H as $n^{(0)}, T, T_0 \rightarrow \infty$. This implies our desired result because if estimation error is constant in expectation over problem instances, there must exist at least one problem instance with constant estimation error. \square

4 Extension to Synthetic Interventions

Our focus so far has been to incentivize units who *a priori* prefer the (single) treatment to take the control in order to obtain accurate counterfactual estimates for all units under control. In this section, we extend Algorithm 1 to the setting where there are *multiple* treatments. Now our goal is to incentivize enough exploration across all interventions such that the [Unit Overlap Assumption](#) is satisfied and thus our counterfactual estimates are valid simultaneously for all units in the population under every intervention. In order to do so, we use tools from the literature on *synthetic interventions*, which is a generalization of the synthetic control framework which allows for counterfactual estimation under different treatments, in addition to control [4, 5].

Our setup is the same as in Section 3, with the only difference being that each unit may choose one of $k \geq 2$ interventions after the pre-treatment period. We assume that Assumption 2.1, Assumption 2.3, Definition 2.7, and Definition 2.8 are all extended to hold under k interventions. Under this setting, a unit's beliefs induce a preference ordering over different interventions. We capture this through the notion of a unit *subtype*, which is a generalization of our definition of unit type (Definition 2.9) to the setting with k interventions.

Definition 4.1 (Unit Subtype). *A unit subtype $\tau \in [k]_0^k$ is a preference ordering over interventions. Unit i is of subtype τ if for every $\kappa \leq k$,*

$$\tau[\kappa] = \arg \max_{d \in [k]_0 \setminus \{\tau[\kappa']\}_{\kappa' < \kappa}} \mu_{v_i}^{(d)}.$$

Unit i 's type is $\tau[1]$. We denote the set of all units of subtype τ as $\mathcal{T}^{(\tau)}$ and the dimension of the latent subspace spanned by subtype τ units as $r_\tau := \text{span}(\mathbf{v}_j : j \in \mathcal{T}^{(\tau)})$.

Since different unit subtypes have different preferences over interventions, we design our algorithm in this section (Algorithm 2) to be BIC for all units of a given subtype (in contrast to Algorithm 1, which is BIC for all units of a given type). While there are at most $k!$ subtypes, the principal will only have to run Algorithm 2 on at most r different subtypes in order to ensure that the [Unit Overlap Assumption](#) is satisfied for all units under all interventions with high probability. This is because if the [Unit Overlap Assumption](#) is satisfied for a specific unit subtype whose latent factors lie within a particular subspace, it is also satisfied for all other units whose latent factors lie within the same subspace.

For a given subtype τ , the goal of Algorithm 2 is to incentivize sufficient exploration for the [Unit Overlap Assumption](#) to be satisfied for all interventions for subtype τ units with high probability. As was the case in Section 3, the algorithm will still be split into two phases, and units will not be given any recommendation in the first phase. The main difference compared to Algorithm 1 is how the principal chooses the exploit intervention in the second phase.

When there are only two interventions, the principal can compare their estimate of the average counterfactual outcome under treatment to the lower bound on the prior-mean average counterfactual outcome for control for type 1 units (and vice versa for type 0 units). However, a more complicated procedure is required to maintain Bayesian incentive-compatibility when there are more than two interventions. Instead in Algorithm 2, the principal will set the exploit intervention as follows: For receiver subtype τ , Algorithm 2 sets intervention ℓ to be the exploit

ALGORITHM 2: Incentivizing Exploration for Synthetic Interventions: Subtype τ units

Input: Group size L , number of batches B , failure probability δ , fixed gap $C \in (0, 1)$

Provide no recommendation to first N_0 units.

// Loop through “explore” interventions

```
for  $\ell = \tau[k], \tau[k-1], \dots, \tau[2]$  do
  for batch  $b = 1, 2, \dots, B$  do
    Select an explore index  $i_b \in [L]$  uniformly at random.
    for  $j = 1, 2, \dots, L$  do
      if  $j = i_b$  then
        Recommend intervention  $\hat{d}_{N_0+(b-1) \cdot L+j} = \ell$ 
      else
        if  $\mathbb{E}[\bar{y}_{i,post}^{(\tau[1])}] + C < \mathbb{E}[\bar{y}_{i,post}^{(d)}] < \mu^{(\ell,\tau,l)} - C$  for every intervention  $d \in \{\ell-1, \dots, \tau[k]\}$ 
          then
            Recommend intervention  $\hat{d}_{N_0+(b-1) \cdot L+j} = \ell$ 
          else
            Recommend intervention  $\hat{d}_{N_0+(b-1) \cdot L+j} = \tau[1]$ 
        end
      end
    end
  end
end
end
```

intervention for unit i only if the sample average of *all* interventions $d \in \{\ell-1, \dots, \tau[k]\}$ are both (1) larger than the sample average of intervention $\tau[1]$ by some constant gap C and (2) less than the lower bound on the prior-mean average counterfactual outcome $\mu_i^{(\ell)}$ by C . If no such intervention satisfies both conditions, intervention $\tau[1]$ is chosen to be the exploit intervention.

We require the following “common knowledge” assumption for Algorithm 2, which is analogous to Assumption 3.1:

Assumption 4.2. We assume that the following are common knowledge among all units and the principal:

1. Valid upper- and lower-bounds on the fraction of subtype τ units in the population, i.e. if the true proportion of subtype τ units is $p_\tau \in (0, 1)$, the principal knows $p_{\tau,L}, p_{\tau,H} \in (0, 1)$ such that $p_{\tau,L} \leq p_\tau \leq p_{\tau,H}$.
2. Valid upper- and lower-bounds on the prior mean for each receiver subtype and intervention, i.e. values $\mu^{(d,\tau,h)}, \mu^{(d,\tau,l)} \in \mathbb{R}$ such that $\mu^{(d,\tau,l)} \leq \mu_i^{(d)} \leq \mu^{(d,\tau,h)}$ for all receiver subtypes τ , interventions $d \in \llbracket k \rrbracket_0$ and all units $i \in \mathcal{T}^{(\tau)}$.
3. For some $C \in (0, 1)$, a lower bound on the smallest probability of the event $\mathcal{E}_{C,i}$ over the prior and latent factors of subtype τ units, denoted by $\min_{i \in \mathcal{T}^{(\tau)}} \Pr_{\mathcal{P}_i}[\mathcal{E}_{C,i}] \geq \zeta_C > 0$, where

$$\mathcal{E}_{C,i} := \{\forall j \in \llbracket k \rrbracket_0 \setminus \{\tau[1], \tau[k]\} : \bar{y}_{i,post}^{(\tau[1])} + C \leq \bar{y}_{i,post}^{(j)} \leq \mu_i^{(\tau[k])} - C\}$$

where $\bar{y}_{i,post}^{(\ell)}$ is the average post-intervention outcome for unit i under intervention $\ell \in \llbracket k \rrbracket_0$.

4. A sufficient number of observations N_0 needed such that the *Unit Overlap Assumption* is satisfied for subtype τ units under intervention $\tau[1]$ with probability at least $1 - \delta$, for some $\delta \in (0, 1)$.

Theorem 4.3 (Informal; detailed version in Theorem B.1). *Suppose that Assumption 4.2 holds for some constant gap $C \in (0, 1)$. If the number of initial units N_0 is chosen to be large enough such that Unit Overlap Assumption is satisfied for all units of subtype τ under intervention $\tau[1]$ with high probability, then Algorithm 2 is BIC if the batch size L is chosen to be sufficiently large.*

Moreover, if the number of batches B is chosen to be sufficiently large, then the Unit Overlap Assumption will be satisfied for all units of subtype τ under all interventions with high probability.

See Appendix B for the proof, which follows similarly to the analysis of Algorithm 1. The main difference in the analysis compared to that of Algorithm 1 lies in the definition of the “exploit” intervention.

First, when there are k interventions, an intervention ℓ is chosen as the exploit intervention only when the sample average of each intervention $d \in \{\ell - 1, \dots, \tau[k]\}$ is (i) larger than the sample average of intervention $\tau[1]$, and (ii) smaller than the prior-mean average outcome of intervention ℓ (with some margin C). Second, instead of choosing between a “previously best” intervention (i.e., an intervention $\tau[j] < \ell$ with the highest sample average) and intervention ℓ to be the ‘exploit’ intervention, Algorithm 2 always choose between $\tau[1]$ and ℓ . This is because in the former case, conditional on any given intervention being chosen as the exploit intervention, there is no known concentration-based analysis that can show the exploit intervention will have a higher average expected outcome compared to all other interventions.⁶

5 Testing Whether the Unit Overlap Assumption Holds

So far, our focus has been on designing algorithms to incentivize subpopulations of units to take interventions for which the Unit Overlap Assumption does not initially hold. In this section we study the complementary problem of designing a procedure for determining whether this assumption holds for a given test unit and set of donor units. In particular, we provide two simple hypothesis tests for deciding whether the Unit Overlap Assumption holds, using existing finite sample guarantees for principal component regression and an additional assumption on the underlying latent factors associated with the pre-intervention time-steps.

5.1 A Non-Asymptotic Hypothesis Test

Our hypothesis test proceeds as follows: Using the PCR techniques described in Section 2.2, we learn a linear relationship between the first $T_0/2$ time-steps in the pre-intervention time period and the average outcome in the second half of the pre-intervention time period using data from the donor units. Using this learned relationship, we then compare our estimate for the test unit with their *true* average outcome in the second half of the pre-intervention time period. If the difference between the two is larger than the confidence interval given by PCR (plus an additional term to account for the noise), we can conclude that the Unit Overlap Assumption is not satisfied with high probability.

In order for our hypothesis test to be valid we require the following assumption, which is analogous to Assumption 2.3.

Assumption 5.1 (Linear Span Inclusion, revisited). *For every time-step t such that $T_0/2 < t \leq T_0$, we assume that*

$$\mathbf{u}_t^{(0)} \in \text{span}\{\mathbf{u}_1^{(0)}, \dots, \mathbf{u}_{T_0/2}^{(0)}\}.$$

While Assumption 2.3 requires that the post-intervention latent factors for any intervention fall within the linear span of the pre-intervention latent factors, Assumption 5.1 requires that the second half of the pre-intervention latent factors fall within the linear span of the first half.

⁶This observation is in line with work on frequentist-based BIC algorithms for incentivizing exploration in bandits. See Section 6.2 of Mansour et al. [31] for more details.

This assumption allows us to obtain valid confidence bounds when regressing the second half of the pre-intervention outcomes on the first half. Let

$$y_{i,pre'} := [y_{i,1}^{(0)}, \dots, y_{i,T_0/2}^{(0)}], \quad y_{i,pre''} := [y_{i,T_0/2+1}^{(0)}, \dots, y_{i,T_0}^{(0)}], \quad \text{and} \quad \bar{y}_{i,pre''} := \frac{2}{T_0} \sum_{t=T_0/2+1}^{T_0} y_{i,t}^{(0)}.$$

Under Assumption 5.1 we can estimate $\mathbb{E}[\bar{y}_{i,pre''}]$ as $\widehat{\mathbb{E}}[\bar{y}_{i,pre''}] := \langle \widehat{\theta}_i, y_{i,pre'} \rangle$, where $\widehat{\theta}_i$ is computed using (regularized) PCR.

We are now ready to introduce our hypothesis test. In what follows, we will leverage the high-probability confidence interval of Agarwal et al. [5] (Theorem A.1). Under Assumption 5.1, we can apply Theorem A.1 out-of-the-box to estimate $\mathbb{E}[\bar{y}_{i,pre''}]$ using the first $T_0/2$ time-steps as the pre-intervention period and the next $T_0/2$ time-steps as the post-intervention period.

Non-Asymptotic Hypothesis Test Consider the hypothesis that the [Unit Overlap Assumption](#) is not satisfied for a test unit n .

1. Using donor units \mathcal{I} and test unit n , compute $\widehat{\mathbb{E}}[\bar{y}_{n,pre''}]$ via (regularized) PCR.
2. Compute $\alpha(\delta) + 2\sigma\sqrt{\frac{\log(1/\delta)}{T_0}}$ for confidence level δ using Theorem A.1 with the first $T_0/2$ time-steps as the pre-intervention period and the next $T_0/2$ time-steps as the post-intervention period.
3. If $|\widehat{\mathbb{E}}[\bar{y}_{n,pre''}] - \bar{y}_{n,pre''}| > \alpha(\delta) + 2\sigma\sqrt{\frac{\log(1/\delta)}{T_0}}$ then accept the hypothesis. Otherwise, reject.

Theorem 5.2. *Under Assumption 5.1, if [Unit Overlap Assumption](#) is satisfied for unit n , then*

$$|\widehat{\mathbb{E}}[\bar{y}_{n,pre''}] - \bar{y}_{n,pre''}| \leq \alpha(\delta) + 2\sigma\sqrt{\frac{\log(1/\delta)}{T_0}}$$

with probability $1 - \mathcal{O}(\delta)$, where $\alpha(\delta)$ is the high-probability confidence interval which is defined in Theorem A.1 when using the first $T_0/2$ time-steps as the pre-intervention period and the next $T_0/2$ time-steps as the post-intervention period.

Proof.

$$\begin{aligned} |\widehat{\mathbb{E}}[\bar{y}_{i,pre''}] - \bar{y}_{i,pre''}| &\leq |\widehat{\mathbb{E}}[\bar{y}_{i,pre''}] - \mathbb{E}[\bar{y}_{i,pre''}]| + |\mathbb{E}[\bar{y}_{i,pre''}] - \bar{y}_{i,pre''}| \\ &\leq \alpha(\delta) + \left| \frac{2}{T_0} \sum_{t=T_0/2+1}^{T_0} \epsilon_{i,t}^{(0)} \right| \end{aligned}$$

with probability at least $1 - \mathcal{O}(\delta)$, where the second inequality follows from Theorem A.1. The result follows from a Hoeffding bound. \square

Therefore we can conclude that if $|\widehat{\mathbb{E}}[\bar{y}_{i,pre''}] - \bar{y}_{i,pre''}| > \alpha(\delta) + 2\sigma\sqrt{\frac{\log(1/\delta)}{T_0}}$, then the [Unit Overlap Assumption](#) will not be satisfied for unit i , conditioned on the high-probability event in Theorem 5.2 holding.

5.2 An Asymptotic Hypothesis Test

Next we present a hypothesis test which, while only valid in the limit, may be of more practical use when compared to the non-asymptotic hypothesis test of Section 5.1. At a high level, our asymptotic hypothesis test leverages Assumption 5.1 and the guarantees for PCR in Agarwal et al. [4] to determine whether the [Unit Overlap Assumption](#) holds for a given test unit n . In particular, the results of Agarwal et al. imply that under Assumption 5.1 and the [Unit Overlap](#)

Assumption, a rescaled version of $\widehat{\mathbb{E}}[\bar{y}_{n,pre''}] - \mathbb{E}[\bar{y}_{n,pre''}]$ converges in distribution to the standard normal distribution as $|\mathcal{I}|, T_0, T_1 \rightarrow \infty$. While the rescaling amount depends on quantities which are not computable, a relatively simple application of the continuous mapping theorem and Slutsky's theorem imply that we can compute a valid test statistic which converges in distribution to the standard normal.

Notation In what follows, let $\omega^{(n,0)} \in \mathbb{R}^{|\mathcal{I}|}$ be the linear relationship between the test unit n and donor units \mathcal{I} which is known to exist under the **Unit Overlap Assumption**, i.e. $\mathbb{E}[y_{n,t}^{(0)}] = \sum_{i \in \mathcal{I}} \omega^{(n,0)}[i] \cdot \mathbb{E}[y_{i,t}^{(0)}]$. Let $\tilde{\omega}^{(n,0)}$ be the projection of $\omega^{(n,0)}$ onto the subspace spanned by the latent factors of the \mathcal{I} test units. Let $\hat{\omega}^{(n,0)}$ be the estimate of $\tilde{\omega}^{(n,0)}$ given by the synthetic interventions procedure of Agarwal et al. [4]. Let $\hat{\sigma}$ be the estimate of the standard deviation of the noise terms $\{\epsilon_{i,t} : i \in \mathcal{I}, t \in [T]\}$ given by Agarwal et al.⁷ Finally, let z_α be the z score at significance level $\alpha \in (0, 1)$, i.e. $\alpha = \mathbb{P}(x \leq z_\alpha)$, where $x \sim \mathcal{N}(0, 1)$.

Asymptotic Hypothesis Test Consider the hypothesis that the **Unit Overlap Assumption** is not satisfied for a test unit n .

1. Using donor units \mathcal{I} and test unit n , compute $\widehat{\mathbb{E}}[\bar{y}_{n,pre''}]$ and $\hat{\omega}^{(n,0)}$ using the synthetic interventions procedure of Agarwal et al. [4].
2. If $\frac{\sqrt{T_1}}{\hat{\sigma} \|\hat{\omega}^{(n,0)}\|_2} |\widehat{\mathbb{E}}[\bar{y}_{n,pre''}] - \bar{y}_{n,pre''}| > 1 - z_{0.95}$ then accept the hypothesis. Otherwise, reject.

Theorem 5.3 (Informal; detailed version in Theorem C.1). *Under Assumption 5.1, if $\|\tilde{\omega}^{(n,0)}\|_2$ is sufficiently large and $\hat{\omega}^{(n,0)} \rightarrow \tilde{\omega}^{(n,0)}$ at a sufficiently fast rate, then as $|\mathcal{I}|, T_0, T_1 \rightarrow \infty$ the Asymptotic Hypothesis Test falsely accepts the hypothesis with probability at most 5%.*

6 Numerical Simulations

In this section, we complement our theoretical results with a numerical comparison to standard methods for counterfactual estimation which do not take incentives into consideration.

Experiment setup We consider the setting of Section 3: there are two interventions and two types of units. Type 1 units prefer the treatment and type 0 units prefer the control. If unit i is of type 1 (resp. type 0), we generate a latent factor $\mathbf{v}_i = [0 \cdots 0 \ v_i[r/2 + 1] \cdots v_i[r]]$ (resp. $\mathbf{v}_i = [v_i[0] \cdots v_i[r/2] \ 0 \cdots 0]$), where $\forall j \in [r] : v_i[j] \sim \text{Unif}(0, 1)$. We consider a setting where 500 units of alternating types arrive sequentially.⁸ Our goal is to incentivize type 1 units to take the control in order to obtain accurate counterfactual estimates under control over time. We consider a pre-intervention time period of length $T_0 = 100$ with latent factors as follows:

- If $t \bmod 2 = 0$, generate latent factor $\mathbf{u}_t^{(0)} = [0 \cdots 0 \ u_t^{(0)}[r/2 + 1] \cdots u_t^{(0)}[r]]$, where $\forall \ell \in [r/2 + 1, r] : u_t^{(0)}[\ell] \sim \text{Unif}[0.25, 0.75]$.
- If $t \bmod 2 = 1$, generate latent factor $\mathbf{u}_t^{(0)} = [u_t^{(0)}[0] \cdots u_t^{(0)}[r/2] \ 0 \cdots 0]$, where $\forall \ell \in [0, r/2] : u_t^{(0)}[\ell] \sim \text{Unif}[0.25, 0.75]$.

We set $T_1 = 100$ and generate post-intervention outcomes as follows:

- If $d = 0$, we set $\mathbf{u}_t^{(d)} = [u_t^{(d)}[0] \cdots u_t^{(d)}[r]] : t \in [T_0 + 1, T]$, where $\forall \ell \in [r] : u_t^{(d)}[\ell] \sim \text{Unif}[0, 1]$.

⁷Recall that we only assume knowledge of an upper-bound σ^2 on the variance of $\epsilon_{i,t}$.

⁸This is done only for convenience. The order of agent arrival is unknown to the algorithms.

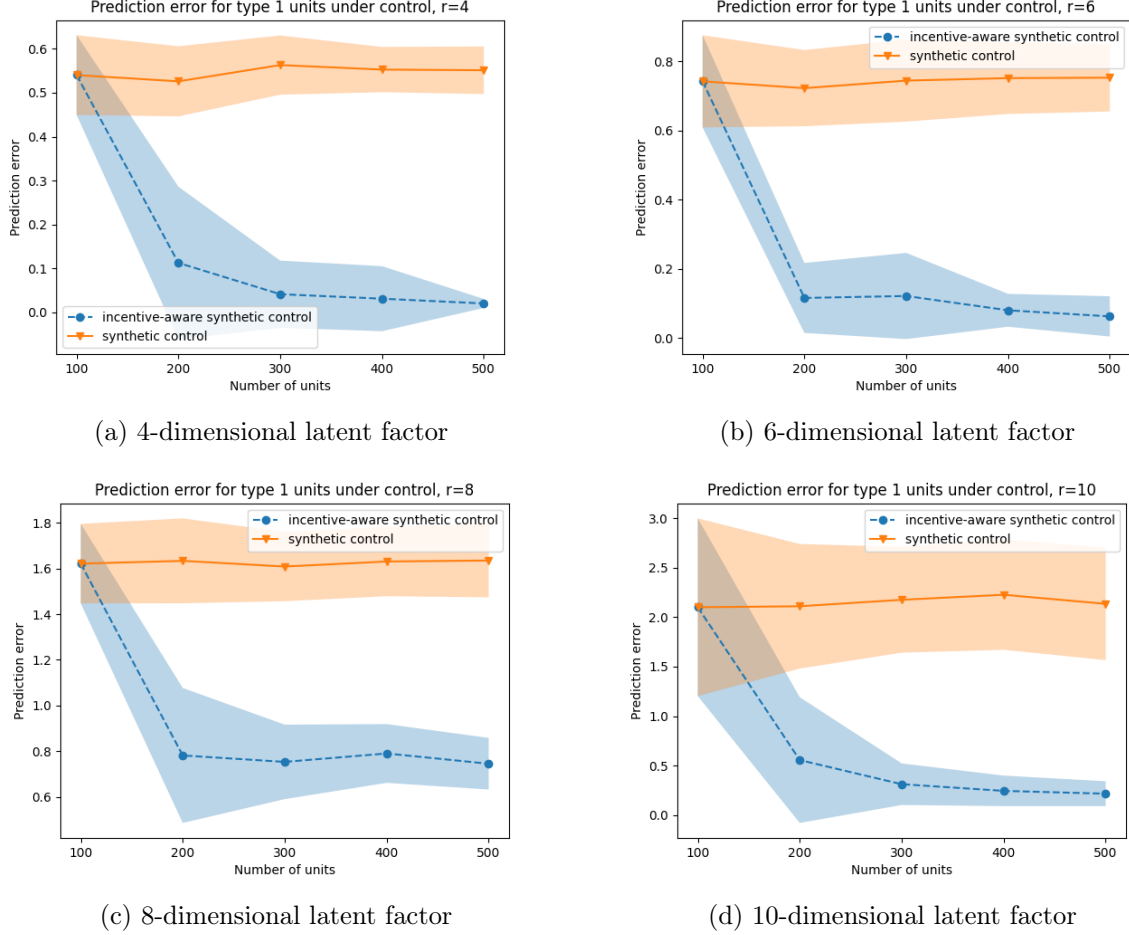


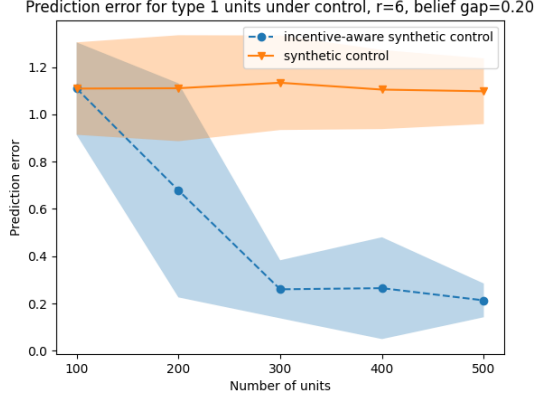
Figure 1: Counterfactual estimation error for units of type 1 under control using Algorithm 1 (blue) and synthetic control without incentives (orange) with increasing number of units and different lengths of the latent factors. Results are averaged over 50 runs, with the shaded regions representing one standard deviation.

- If $d = 1$, we set $\mathbf{u}_t^{(d)} = [u_t^{(d)}[0] \cdots u_t^{(d)}[r]] : t \in [T_0 + 1, T]$, where $\forall \ell \in [r] : u_t^{(d)}[\ell] \sim \text{Unif}[-1, 0]$.

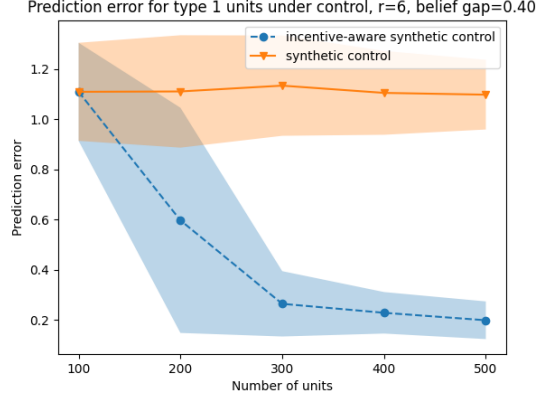
Finally, outcomes are generated by adding independent Gaussian noise $\epsilon_{i,t}^{(d)} \sim \mathcal{N}(0, 0.01)$ to each inner product of latent factors.

Using Theorem A.4, we can calculate a lower bound on the phase length L of Algorithm 1 such that the BIC condition is satisfied for units of type 1 who are recommended the control. We run three different sets of simulations and compare the estimation error $|\hat{\mathbb{E}}\bar{y}_{n,\text{post}}^{(0)} - \mathbb{E}\bar{y}_{n,\text{post}}^{(0)}|$ for a new type 1 unit n under control when incentivizing exploration using Algorithm 1 (blue) to the estimation error without incentivizing exploration (orange). For both our method and the incentive-unaware ablation, we use the adaptive PCR method of Agarwal et al. [5] to estimate counterfactuals. All experiments are repeated 50 times and we report both the average prediction error and the standard deviation for estimating the post-treatment outcome of type 1 units under control.

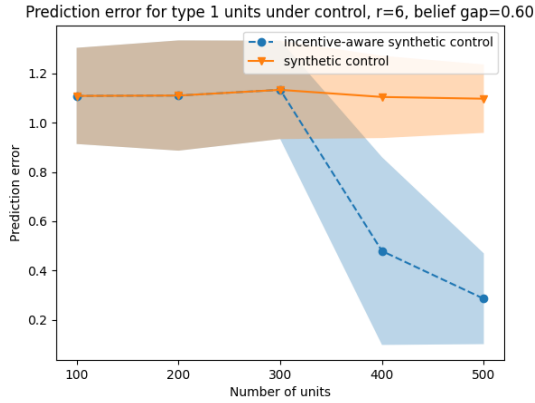
Varying latent factor size In Figure 1, we vary $r \in \{2, 4, 6, 8\}$ while keeping all other parameters of the problem instance constant. Increasing r increases the batch size L , so we see a modest decrease in performance as r increases.



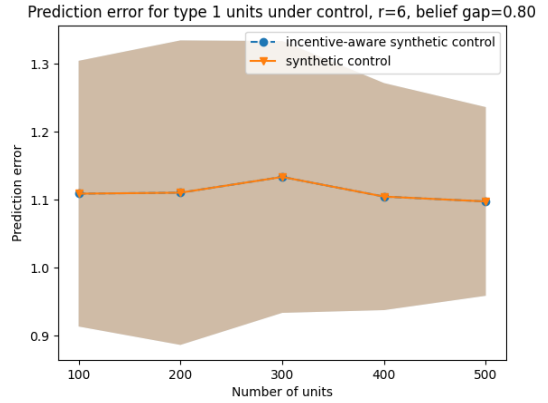
(a) Prior mean gap 0.2



(b) Prior mean gap 0.4



(c) Prior mean gap 0.6



(d) Prior mean gap 0.8

Figure 2: Counterfactual estimation error for units of type 1 under control using Algorithm 1 (blue) and synthetic control without incentives (orange) with different gaps between the prior mean reward of control and treatment. Results are averaged over 50 runs, with the shaded regions representing one standard deviation.

Varying strength of belief In Figure 2, we explore the effects of changing the units’ “strength of beliefs”. Specifically, we vary the gap between the prior mean outcome under control and treatment for type 1 units by setting $\mu_i^{(1)} - \mu_i^{(0)} \in \{0.2, 0.4, 0.6, 0.8\}$. As the gap decreases it becomes easier to persuade units to explore, and therefore our counterfactual estimates improve.

7 Conclusion

We study the problem of non-compliance when performing counterfactual estimation using panel data. Our focus is on synthetic control methods, which canonically require a unit overlap assumption on the donor units in order to provide valid finite sample guarantees. We shed light on this often overlooked assumption, and provide an algorithm for incentivizing units to explore different interventions using tools from information design and online learning. After running our algorithm, the [Unit Overlap Assumption](#) will be satisfied with high probability, which allows for the principal to obtain valid finite sample guarantees for all units when using off-the-shelf synthetic control methods to estimate counterfactuals under control. We also extend our algorithm to satisfy the [Unit Overlap Assumption](#) when there are more than two interventions, and provide a hypothesis test for determining if the [Unit Overlap Assumption](#) hold for a given test unit and set of donor units. Finally, we complement our theoretical findings with numerical

simulations, and observe that our procedure for estimating counterfactual outcomes produces significantly more accurate counterfactual estimates when compared to existing methods which do not take unit incentives into consideration.

There are several exciting directions for future research. Recent work (Sellke and Slivkins [40], Sellke [39]) provides lower bounds on the sample complexity of incentivizing exploration in various multi-armed bandit settings. It would be interesting to provide analogous bounds on the number of units required to incentivize sufficient exploration in our panel data setting. Another avenue for future research is to design difference-in-differences or clustering-based algorithms for incentivizing exploration in other causal inference settings. Finally it would be interesting to relax other assumptions typically needed for synthetic control, such as the analogous linear span inclusion assumption on the time-step latent factors (Assumption 2.3).

References

- [1] Alberto Abadie and Javier Gardeazabal. The economic costs of conflict: A case study of the basque country. *American economic review*, 93(1):113–132, 2003.
- [2] Alberto Abadie, Alexis Diamond, and Jens Hainmueller. Synthetic control methods for comparative case studies: Estimating the effect of california’s tobacco control program. *Journal of the American statistical Association*, 105(490):493–505, 2010.
- [3] Anish Agarwal, Devavrat Shah, and Dennis Shen. On model identification and out-of-sample prediction of principal component regression: Applications to synthetic controls. *arXiv preprint arXiv:2010.14449*, 2020.
- [4] Anish Agarwal, Devavrat Shah, and Dennis Shen. Synthetic interventions. *arXiv preprint arXiv:2006.07691*, 2020.
- [5] Anish Agarwal, Keegan Harris, Justin Whitehouse, and Zhiwei Steven Wu. Adaptive Principal Component Regression with Applications to Panel Data. Papers 2307.01357, arXiv.org, July 2023. URL <https://ideas.repec.org/p/arx/papers/2307.01357.html>.
- [6] Muhammad Amjad, Devavrat Shah, and Dennis Shen. Robust synthetic control. *The Journal of Machine Learning Research*, 19(1):802–852, 2018.
- [7] Muhammad Amjad, Vishal Misra, Devavrat Shah, and Dennis Shen. mrsc: Multi-dimensional robust synthetic control. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 3(2):1–27, 2019.
- [8] Joshua D. Angrist and Jörn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist’s Companion*. Princeton University Press, 2009. ISBN 9780691120348.
- [9] Manuel Arellano and Bo Honore. Panel data models: Some recent developments. *Handbook of Econometrics*, 02 2000.
- [10] Orley C Ashenfelter and David Card. Using the longitudinal structure of earnings to estimate the effect of training programs, 1984.
- [11] Dirk Bergemann and Stephen Morris. Information design: A unified perspective. *Journal of Economic Literature*, 57(1):44–95, March 2019. doi: 10.1257/jel.20181489. URL <https://www.aeaweb.org/articles?id=10.1257/jel.20181489>.
- [12] Marianne Bertrand, Esther Dufo, and Sendhil Mullainathan. How much should we trust differences-in-differences estimates? *The Quarterly journal of economics*, 119(1):249–275, 2004.
- [13] Gary Chamberlain. Panel data. In Z. Griliches† and M. D. Intriligator, editors, *Handbook of Econometrics*, volume 2, chapter 22, pages 1247–1318. Elsevier, 1 edition, 1984. URL <https://EconPapers.repec.org/RePEc:eee:ecochp:2-22>.
- [14] Jiafeng Chen. Synthetic control as online linear regression. *Econometrica*, 91(2):465–491, 2023.
- [15] Stephen G Donald and Kevin Lang. Inference with difference-in-differences and other panel data. *The review of Economics and Statistics*, 89(2):221–233, 2007.
- [16] Raaz Dwivedi, Katherine Tian, Sabina Tomkins, Predrag Klasnja, Susan Murphy, and Devavrat Shah. Counterfactual inference for sequential experiments. *arXiv preprint arXiv:2202.06891*, 2022.

- [17] EBay, Jun 2022. URL https://partnerhelp.ebay.com/helpcenter/s/article/Incrementality-Testing-by-Location-IP?language=en_US.
- [18] Vivek Farias, Ciamac Moallemi, Tianyi Peng, and Andrew Zheng. Synthetically controlled bandits. *arXiv preprint arXiv:2202.07079*, 2022.
- [19] Danilo Freire. Evaluating the effect of homicide prevention strategies in são paulo, brazil: A synthetic control approach. *Latin American Research Review*, 53(2):231–249, 2018.
- [20] Susan Groth. Honorarium or coercion: Use of incentives for participants in clinical research. *The Journal of the New York State Nurses’ Association*, 41:11–3; quiz 22, 03 2010.
- [21] Keegan Harris, Anish Agarwal, Chara Podimata, and Zhiwei Steven Wu. Strategyproof decision-making in panel data settings and beyond. *arXiv preprint arXiv:2211.14236*, 2022.
- [22] Nicole Immorlica, Jieming Mao, Aleksandrs Slivkins, and Zhiwei Steven Wu. Incentivizing exploration with selective data disclosure, 2023.
- [23] Ian T Jolliffe. A note on the use of principal components in regression. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 31(3):300–303, 1982.
- [24] Emir Kamenica. Bayesian persuasion and information design. *Annual Review of Economics*, 11:249–272, 2019.
- [25] Emir Kamenica and Matthew Gentzkow. Bayesian persuasion. *American Economic Review*, 101(6):2590–2615, October 2011. doi: 10.1257/aer.101.6.2590. URL <https://www.aeaweb.org/articles?id=10.1257/aer.101.6.2590>.
- [26] Ilan Kremer, Yishay Mansour, and Motty Perry. Implementing the “wisdom of the crowd”. *Journal of Political Economy*, 122(5):988–1012, 2014. ISSN 00223808, 1537534X. URL <http://www.jstor.org/stable/10.1086/676597>.
- [27] Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.
- [28] Yingkai Li and Aleksandrs Slivkins. Incentivizing participation in clinical trials. *arXiv preprint arXiv:2202.06191*, 2022.
- [29] Kung-Yee Liang and Scott L. Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22, 04 1986. ISSN 0006-3444. doi: 10.1093/biomet/73.1.13. URL <https://doi.org/10.1093/biomet/73.1.13>.
- [30] Alexander R Luedtke and Mark J van der Laan. Optimal individualized treatments in resource-limited settings. *The international journal of biostatistics*, 12(1):283–303, 2016.
- [31] Yishay Mansour, Aleksandrs Slivkins, and Vasilis Syrgkanis. Bayesian incentive-compatible bandit exploration. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, pages 565–582, 2015.
- [32] Yishay Mansour, Aleksandrs Slivkins, and Vasilis Syrgkanis. Bayesian incentive-compatible bandit exploration, 2019.
- [33] Yishay Mansour, Aleksandrs Slivkins, and Vasilis Syrgkanis. Bayesian incentive-compatible bandit exploration. *Operations Research*, 68(4):1132–1161, 2020.
- [34] Yishay Mansour, Aleksandrs Slivkins, Vasilis Syrgkanis, and Zhiwei Steven Wu. Bayesian exploration: Incentivizing exploration in bayesian games, 2021.

- [35] William F Massy. Principal components regression in exploratory statistical research. *Journal of the American Statistical Association*, 60(309):234–256, 1965.
- [36] Scott M Mourtgos, Ian T Adams, and Justin Nix. Elevated police turnover following the summer of george floyd protests: A synthetic control study. *Criminology & Public Policy*, 21(1):9–33, 2022.
- [37] Hongxiang Qiu, Marco Carone, Ekaterina Sadikova, Maria Petukhova, Ronald C Kessler, and Alex Luedtke. Optimal individualized decision rules using instrumental variable methods. *Journal of the American Statistical Association*, 116(533):174–191, 2021.
- [38] Hongxiang Qiu, Marco Carone, and Alex Luedtke. Individualized treatment rules under stochastic treatment cost constraints. *Journal of Causal Inference*, 10(1):480–493, 2022.
- [39] Mark Sellke. Incentivizing exploration with linear contexts and combinatorial actions, 2023.
- [40] Mark Sellke and Aleksandrs Slivkins. The price of incentivizing exploration: A characterization via thompson sampling and sample complexity, 2022.
- [41] Aleksandrs Slivkins. Exploration and persuasion, 2021.
- [42] Aleksandrs Slivkins et al. Introduction to multi-armed bandits. *Foundations and Trends® in Machine Learning*, 12(1-2):1–286, 2019.
- [43] Hao Sun, Evan Munro, Georgy Kalashnov, Shuyang Du, and Stefan Wager. Treatment allocation under uncertain costs. *arXiv preprint arXiv:2103.11066*, 2021.
- [44] Uber, Jun 2019. URL <https://www.uber.com/blog/causal-inference-at-uber/>.
- [45] R. Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2018. ISBN 9781108415194. URL <https://books.google.com/books?id=J-VjswEACAAJ>.
- [46] Yingying Zhang, Huixia Judy Wang, and Zhongyi Zhu. Quantile-regression-based clustering for panel data. *Journal of Econometrics*, 213(1):54–67, 2019.

A Appendix for Section 3: Incentivized Exploration for Synthetic Control

A.1 Causal Parameter Recovery

Theorem A.1 (Theorem G.3 of Agarwal et al. [5]). *Let $\delta \in (0, 1)$ be an arbitrary confidence parameter and $\rho > 0$ be chosen to be sufficiently small. Further, assume that Assumption 2.1 and Unit Overlap Assumption are satisfied, there is some $i_0 \geq 1$ such that $\text{rank}(\mathbf{X}_{i_0}](d)) = r$, and $\text{snr}_i(d) \geq 2$ for all $i \geq i_0$. Then, with probability at least $1 - \mathcal{O}(k\delta)$, simultaneously for all interventions $d \in [k]_0$,*

$$\begin{aligned} \left| \widehat{\mathbb{E}}[\bar{y}_{i,\text{post}}^{(d)}] - \mathbb{E}[y_{i,\text{post}}^{(d)}] \right| &\leq \alpha(\delta), \\ \text{where } \alpha(\delta) &:= \frac{3\sqrt{T_0}}{\widehat{\text{snr}}_i(d)} \left(\frac{L(\sqrt{74} + 12\sqrt{6}\kappa(\mathbf{Z}_i(d)))}{(T - T_0) \cdot \widehat{\text{snr}}_i(d)} + \frac{\sqrt{\text{err}_i(d)}}{\sqrt{T - T_0} \cdot \sigma_r(\mathbf{Z}_i(d))} \right) \\ &\quad + \frac{2L\sqrt{24T_0}}{(T - T_0) \cdot \widehat{\text{snr}}_i(d)} + \frac{12L\kappa(\mathbf{Z}_i(d))\sqrt{3T_0}}{(T - T_0) \cdot \widehat{\text{snr}}_i(d)} + \frac{2\sqrt{\text{err}_i(d)}}{\sqrt{T - T_0} \cdot \sigma_r(\mathbf{Z}_i(d))} \\ &\quad + \frac{L\sigma\sqrt{\log(k/\delta)}}{\sqrt{T - T_0}} + \frac{L\sigma\sqrt{74\log(k/\delta)}}{\widehat{\text{snr}}_i(d)\sqrt{T - T_0}} + \frac{12\sigma\kappa(\mathbf{Z}_i(d))\sqrt{6\log(k/\delta)}}{\widehat{\text{snr}}_i(d)\sqrt{T - T_0}} \\ &\quad + \frac{\sigma\sqrt{2\text{err}_i(d)\log(k/\delta)}}{\sigma_r(\mathbf{Z}_i(d))} \end{aligned}$$

where $\widehat{\mathbb{E}}[\bar{y}_{i,\text{post}}^{(d)}] := \frac{1}{T - T_0} \cdot \langle \hat{\theta}_i(d), y_{i,\text{pre}} \rangle$ is the estimated average post-intervention outcome for unit i under intervention d and $\|\theta_i^{(d)}\| \leq L$.

Corollary A.2. *Given a gap ϵ and the same assumptions as in Theorem A.1, the probability that $\left| \widehat{\mathbb{E}}[\bar{y}_{i,\text{post}}^{(d)}] - \mathbb{E}[\bar{y}_{i,\text{post}}^{(d)}] \right| \leq \epsilon$ is at least $1 - \delta$, where*

$$\delta \leq \frac{\log(n^{(d)}) \vee k}{\exp \left(\left(\sqrt{\frac{\sigma_{r_d}(\mathbf{Z}_i(d))\epsilon - (A+F)(\sqrt{n^{(d)}} + \sqrt{T_0})}{D}} + \frac{\alpha^2}{4D^2} - \frac{\alpha}{2D} \right)^2 \right)}$$

$$\begin{aligned} \text{with } A &= 3\sqrt{T_0} \left(\frac{\|\theta_i^{(d)}\|(\sqrt{74} + 12\sqrt{6}\kappa(\mathbf{Z}_i(d)))}{T - T_0} + \frac{1}{\sqrt{T - T_0}} \right), \quad F = \frac{2\|\theta_i^{(d)}\|\sqrt{24T_0}}{T - T_0} + \frac{12\|\theta_i^{(d)}\|\kappa(\mathbf{Z}_i(d))\sqrt{3T_0}}{T - T_0} + \\ &\quad \frac{2}{\sqrt{T - T_0}}, \quad D = \frac{\|\theta_i^{(d)}\|\sigma\sqrt{74}}{\sqrt{T - T_0}} + \frac{12\sigma\kappa(\mathbf{Z}_i(d))\sqrt{6}}{\sqrt{T - T_0}} + \sigma\sqrt{2}, \quad E = \frac{\|\theta_i^{(d)}\|\sigma}{\sqrt{T - T_0}} \text{ and } \alpha = A + F + D(\sqrt{n^{(d)}} + \sqrt{T_0}) + \\ &\quad \sigma_{r_d}(\mathbf{Z}_i(d))E \end{aligned}$$

Proof. We begin by setting the right-hand side of Theorem A.1 to be ϵ . The goal is to write the failure probability δ as a function of ϵ . Then, using the notations above, we can write

$$\epsilon = \frac{A}{(\widehat{\text{snr}}_i(d))^2} + \frac{F}{\widehat{\text{snr}}_i(d)} + \frac{D\sqrt{\log(k/\delta)}}{\widehat{\text{snr}}_i(d)} + E\sqrt{\log(k/\delta)}$$

First, we take a look at the signal-to-noise ratio $\widehat{\text{snr}}_i(d)$. By definition, we have:

$$\begin{aligned} \widehat{\text{snr}}_i(d) &= \frac{\sigma_r(\mathbf{Z}_i(d))}{U_i} \\ &= \frac{\sigma_r(\mathbf{Z}_i(d))}{\sqrt{n^{(d)}} + \sqrt{T_0} + \sqrt{\log(\log(n^{(d)})/\delta)}} \end{aligned}$$

Hence,

$$\frac{1}{\widehat{\text{snr}}_i(d)} = \frac{\sqrt{n^{(d)}} + \sqrt{T_0} + \sqrt{\log(\log(n^{(d)})/\delta)}}{\sigma_r(\mathbf{Z}_i(d))}$$

Observe that since $\widehat{\text{snr}}_i(d) \geq 2$, we have $\frac{1}{(\widehat{\text{snr}}_i(d))^2} \leq \frac{1}{\widehat{\text{snr}}_i(d)}$. Hence, we can write an upper bound on ϵ as:

$$\begin{aligned} \epsilon &\leq \frac{A+F}{\widehat{\text{snr}}_i(d)} + \frac{D\sqrt{\log(k/\delta)}}{\widehat{\text{snr}}_i(d)} + E\sqrt{\log(k/\delta)} \\ &\leq \frac{(A+F)(\sqrt{n^{(d)}} + \sqrt{T_0})}{\sigma_{r_d}(\mathbf{Z}_i(d))} + \frac{(A+F)\sqrt{\log(\log(n^{(d)})/\delta)}}{\sigma_{r_d}(\mathbf{Z}_i(d))} \\ &\quad + \frac{D\left(\sqrt{n} + \sqrt{d} + \sqrt{\log(\log(n^{(d)})/\delta)}\right)\sqrt{\log(k/\delta)}}{\sigma_r(\mathbf{Z}_i(d))} + E\sqrt{\log(k/\delta)} \end{aligned}$$

Since $\log(x)$ is a strictly increasing function for $x > 0$, we can simplify the above expression as:

$$\begin{aligned} \epsilon &\leq \frac{(A+F)(\sqrt{n^{(d)}} + \sqrt{T_0})}{\sigma_{r_d}(\mathbf{Z}_i(d))} + \frac{(A+F)\sqrt{\log(\log(n^{(d)})\vee k/\delta)}}{\sigma_{r_d}(\mathbf{Z}_i(d))} \\ &\quad + \frac{D(\sqrt{n} + \sqrt{d})\sqrt{\log(\log(n^{(d)})\vee k/\delta)}}{\sigma_{r_d}(\mathbf{Z}_i(d))} + \frac{D\log(\log(n^{(d)})\vee k/\delta)}{\sigma_{r_d}(\mathbf{Z}_i(d))} + E\sqrt{\log(\log(n^{(d)})\vee k/\delta)} \end{aligned}$$

Subtracting the first term from both sides and multiplying by $\sigma_r(\mathbf{Z}_i(d))$, we have:

$$\begin{aligned} &\sigma_{r_d}(\mathbf{Z}_i(d))\epsilon - (A+F)(\sqrt{n^{(d)}} + \sqrt{T_0}) \\ &\leq (A+F)\sqrt{\log(\log(n^{(d)})\vee k/\delta)} + D(\sqrt{n} + \sqrt{d})\sqrt{\log(\log(n^{(d)})\vee k/\delta)} + D\log(\log(n^{(d)})\vee k/\delta) \\ &\quad + E\sigma_{r_d}(\mathbf{Z}_i(d))\sqrt{\log(\log(n^{(d)})\vee k/\delta)} \\ &= (A+F + D(\sqrt{n} + \sqrt{d}) + E\sigma_{r_d}(\mathbf{Z}_i(d)))\sqrt{\log(\log(n^{(d)})\vee k/\delta)} + D\log(\log(n^{(d)})\vee k/\delta) \end{aligned}$$

Let $\alpha = A + F + D(\sqrt{n} + \sqrt{d}) + E\sigma_{r_d}(\mathbf{Z}_i(d))$, we can rewrite the inequality above as:

$$\sigma_r(\mathbf{Z}_i(d))\epsilon - (A+F)(\sqrt{n^{(d)}} + \sqrt{T_0}) \leq \alpha\sqrt{\log(\log(n^{(d)})\vee k/\delta)} + D\log(\log(n^{(d)})\vee k/\delta)$$

Then, we can complete the square and obtain:

$$\begin{aligned}
& \log(\log(n^{(d)}) \vee k / \delta) + \frac{\alpha}{D} \sqrt{\log(\log(n^{(d)}) \vee k / \delta)} + \frac{\alpha^2}{4D^2} \geq \frac{\sigma_{r_d}(\mathbf{Z}_i(d))\epsilon - (A + F)(\sqrt{n^{(d)}} + \sqrt{T_0})}{D} + \frac{\alpha^2}{4D^2} \\
& \iff \left(\sqrt{\log(\log(n^{(d)}) \vee k / \delta)} + \frac{\alpha}{2D} \right)^2 \geq \frac{\sigma_{r_d}(\mathbf{Z}_i(d))\epsilon - (A + F)(\sqrt{n^{(d)}} + \sqrt{T_0})}{D} + \frac{\alpha^2}{4D^2} \\
& \iff \sqrt{\log(\log(n^{(d)}) \vee k / \delta)} + \frac{\alpha}{2D} \geq \sqrt{\frac{\sigma_{r_d}(\mathbf{Z}_i(d))\epsilon - (A + F)(\sqrt{n^{(d)}} + \sqrt{T_0})}{D} + \frac{\alpha^2}{4D^2}} \\
& \iff \sqrt{\log(\log(n^{(d)}) \vee k / \delta)} \geq \sqrt{\frac{\sigma_{r_d}(\mathbf{Z}_i(d))\epsilon - (A + F)(\sqrt{n^{(d)}} + \sqrt{T_0})}{D} + \frac{\alpha^2}{4D^2}} - \frac{\alpha}{2D} \\
& \iff \log(\log(n^{(d)}) \vee k / \delta) \geq \left(\sqrt{\frac{\sigma_{r_d}(\mathbf{Z}_i(d))\epsilon - (A + F)(\sqrt{n^{(d)}} + \sqrt{T_0})}{D} + \frac{\alpha^2}{4D^2}} - \frac{\alpha}{2D} \right)^2 \\
& \iff \frac{\log(n^{(d)}) \vee k}{\delta} \geq \exp \left(\left(\sqrt{\frac{\sigma_{r_d}(\mathbf{Z}_i(d))\epsilon - (A + F)(\sqrt{n^{(d)}} + \sqrt{T_0})}{D} + \frac{\alpha^2}{4D^2}} - \frac{\alpha}{2D} \right)^2 \right) \\
& \iff \delta \leq \frac{\log(n^{(d)}) \vee k}{\exp \left(\left(\sqrt{\frac{\sigma_{r_d}(\mathbf{Z}_i(d))\epsilon - (A + F)(\sqrt{n^{(d)}} + \sqrt{T_0})}{D} + \frac{\alpha^2}{4D^2}} - \frac{\alpha}{2D} \right)^2 \right)}
\end{aligned}$$

□

A.2 Proof of Theorem 3.2

Validity of Assumption 3.1.3 First, we note that in the first stage of Algorithm 1, the principal does not provide any recommendation to the units and instead lets them pick their preferred intervention. The goal of this first stage is to ensure the linear span inclusion assumption ([Unit Overlap Assumption](#)) is satisfied for type 1 units and intervention 1. This condition is equivalent to having enough samples of type 1 units such that the set of latent vectors $\{v_i\}_{i \in \mathcal{I}^{(1)}}$ spans the latent vector space S_1 . We invoke the following theorem from Vershynin [45] that shows $\text{span}(\{v_i\}_{i \in \mathcal{I}^{(1)}}) = S_1$ with high probability:

Theorem A.3 (Theorem 4.6.1 of Vershynin [45]). *Let A be an $m \times n$ matrix whose rows A_i are independent, mean zero, sub-gaussian isotropic random vectors in \mathbb{R}^n . Then for any $t \geq 0$ we have with probability at least $1 - 2 \exp(-t^2)$:*

$$\sqrt{m} - c_{\text{Ver}} K^2 (\sqrt{n} + t) \leq s_n(A) \leq s_1(A) \leq \sqrt{m} + c_{\text{Ver}} K^2 (\sqrt{n} + t) \quad (1)$$

where $K = \max_i \|A_i\|_{\psi_2}$ and c_{Ver} is an absolute constant.

Hence, after observing $N_0^{(1)}$ samples of type 1 units taking intervention 1, the linear span inclusion assumption is satisfied with probability at least $1 - 2 \exp \left(- \left(\frac{\sqrt{N_0^{(1)}}}{c_{\text{Ver}} K^2} - \sqrt{r_1} \right)^2 \right)$.

Theorem A.4. *Suppose there are two interventions, and assume that Assumption 3.1 holds for some constant gap $C \in (0, 1)$. If N_0 is chosen to be large enough such that [Unit Overlap Assumption](#) is satisfied for all units of type 1 under treatment with probability at least $1 - \delta_0$,*

then Algorithm 1 with parameters δ, B, C is BIC for all units of type 1 if

$$L \geq 1 + \max_{i \in \mathcal{T}^{(1)}} \left\{ \frac{\mu_{v_i}^{(1)} - \mu_{v_i}^{(0)}}{\left(C - \alpha(\delta_{PCR}) - \sigma \sqrt{\frac{2 \log(1/\delta_{\epsilon_i})}{T_1}} \right) \Pr \left[\mu_{v_i}^{(0)} - \bar{y}_{i,post}^{(1)} \geq C - \alpha(\delta_{PCR}) - \sigma \sqrt{\frac{2 \log(1/\delta_{\epsilon_i})}{T_1}} \right] - 2\delta} \right\}$$

where $\alpha(\delta_{PCR})$ is the high-probability confidence bound defined in Theorem A.1, $\delta = \delta_{\epsilon_i} + \delta_0 + \delta_{PCR}$ and $\delta_{\epsilon_i} \in (0, 1)$ is the failure probability of the Chernoff-Hoeffding bound on the average of sub-Gaussian random noise $\{\epsilon_{i,t}^{(1)}\}_{t=T_0+1}^T$, and

$$\delta_{PCR} \leq \frac{\log(N_0^{(1)}) \vee k}{\exp \left(\left(\sqrt{\frac{\sigma_r(Y_{pre,N_0}^{(1)})(C/2) - (A+F)(\sqrt{N_0^{(1)}} + \sqrt{T_0})}{D}} + \frac{\alpha^2}{4D^2} - \frac{\alpha}{2D} \right)^2 \right)}$$

and $\kappa(Y_{pre,N_0}^{(1)}) = \sigma_1(Y_{pre,N_0}^{(1)})/\sigma_{r_1}(Y_{pre,N_0}^{(1)})$ is the condition number of the matrix of observed pre-intervention outcomes for the subset of the first N_0 units who have taken the treatment.

The remaining variables in δ_{PCR} are defined as $A = 3\sqrt{T_0} \left(\frac{\Gamma(\sqrt{74} + 12\sqrt{6}\kappa(Y_{pre,N_0}^{(1)}))}{T - T_0} + \frac{1}{\sqrt{T - T_0}} \right)$,

$$F = \frac{2\Gamma\sqrt{24T_0}}{T - T_0} + \frac{12\Gamma\kappa(Y_{pre,N_0}^{(1)})\sqrt{3T_0}}{T - T_0} + \frac{2}{\sqrt{T - T_0}}, \quad D = \frac{\Gamma\sigma\sqrt{74}}{\sqrt{T - T_0}} + \frac{12\sigma\kappa(Y_{pre,N_0}^{(1)})\sqrt{6}}{\sqrt{T - T_0}} + \sigma\sqrt{2}, \quad E = \frac{\Gamma\sigma}{\sqrt{T - T_0}},$$

and $\alpha = A + F + D(\sqrt{N_0^{(1)}} + \sqrt{T_0}) + \sigma_{r_1}(Y_{pre,N_0}^{(1)})E$. Moreover if the number of batches B is chosen to be large enough such that with probability at least $1 - \delta$, $\text{rank}([\mathbb{E}[\mathbf{y}_{i,pre}]^\top : \hat{d}_i = 0, i \in \mathcal{T}^{(1)}]_{i=1}^{N_0+B \cdot L}) = \text{rank}([\mathbb{E}[\mathbf{y}_{j,pre}]^\top]_{j \in \mathcal{T}^{(1)}})$, then the [Unit Overlap Assumption](#) will be satisfied for all type 1 units under control with probability at least $1 - 2\delta$.

Proof. At a particular time step t , unit i of type 1 can be convinced to pick control if $\mathbb{E}_{v_i}[\bar{y}_{i,post}^{(0)} - \bar{y}_{i,post}^{(1)} | \hat{d} = 0] \Pr[\hat{d} = 0] \geq 0$. There are two possible disjoint events under which unit i is recommended intervention 0: either intervention 0 is better empirically, i.e. $\mu_{v_i}^{(0)} \geq \widehat{\mathbb{E}}[\bar{y}_{i,post}^{(1)}] + C$, or intervention 1 is better and unit i is chosen for exploration. Hence, we have

$$\begin{aligned} & \mathbb{E}_{v_i}[\bar{y}_{i,post}^{(0)} - \bar{y}_{i,post}^{(1)} | \hat{d} = 0] \Pr[\hat{d} = 0] \\ &= \mathbb{E}[\bar{y}_{i,post}^{(0)} - \bar{y}_{i,post}^{(1)} | \mu_{v_i}^{(0)} \geq \widehat{\mathbb{E}}[\bar{y}_{i,post}^{(1)}] + C] \Pr[\mu_{v_i}^{(0)} \geq \widehat{\mathbb{E}}[\bar{y}_{i,post}^{(1)}] + C] \left(1 - \frac{1}{L}\right) + \frac{1}{L} \mathbb{E}_{v_i}[\bar{y}_{i,post}^{(0)} - \bar{y}_{i,post}^{(1)}] \\ &= \left(\mu_{v_i}^{(0)} - \mathbb{E}_{v_i}[\bar{y}_{i,post}^{(1)} | \mu_{v_i}^{(0)} \geq \widehat{\mathbb{E}}[\bar{y}_{i,post}^{(1)}] + C] \right) \Pr[\mu_{v_i}^{(0)} \geq \widehat{\mathbb{E}}[\bar{y}_{i,post}^{(1)}] + C] \left(1 - \frac{1}{L}\right) + \frac{1}{L} (\mu_{v_i}^{(0)} - \mu_{v_i}^{(1)}) \end{aligned}$$

Rearranging the terms and taking the maximum over all units of type 1 gives the lower bound on phase length L :

$$L \geq 1 + \frac{\mu_{v_i}^{(1)} - \mu_{v_i}^{(0)}}{(\mu_{v_i}^{(0)} - \mathbb{E}_{v_i}[\widehat{\mathbb{E}}[\bar{y}_{i,post}^{(1)}] | \widehat{\xi}_{C,i}]) \Pr_{v_i}[\widehat{\xi}_{C,i}]}$$

To complete the analysis, we want to find a lower bound for the terms in the denominator. That is, we want to lower bound

$$\left(\mu_{v_i}^{(0)} - \mathbb{E}_{v_i}[\bar{y}_{i,post}^{(1)} | \mu_{v_i}^{(0)} \geq \widehat{\mathbb{E}}[\bar{y}_{i,post}^{(1)}] + C] \right) \Pr_{v_i}[\mu_{v_i}^{(0)} \geq \widehat{\mathbb{E}}[\bar{y}_{i,post}^{(1)}] + C]$$

Let ξ_0 denote the event that [Unit Overlap Assumption](#) is satisfied for type 1 units under treatment. From Assumption 3.1, we know that event ξ_0 occurs with probability at least $1 - \delta_0$ for some $\delta_0 \in (0, 1)$. Let ξ_{PCR} denote the event that the high-probability concentration bound

holds for units of type 1 and intervention 1. Finally, let ξ_{ϵ_i} denote the event where the Chernoff-Hoeffding bound holds on the average of the sub-Gaussian noise $\epsilon_{i,t}^{(1)}$, that is with probability at least $1 - \delta_{\epsilon_i}$, we have:

$$\left| \frac{1}{T_1} \sum_{t=T_0+1}^T \epsilon_{i,t}^{(1)} \right| \leq \sigma \sqrt{\frac{2 \log(1/\delta_{\epsilon_i})}{T_1}}$$

Define another clean event ξ where the events ξ_0 , ξ_{ϵ_i} and ξ_{PCR} happen simultaneously with probability at least $1 - \delta$, where $\delta = \delta_0 + \delta_{PCR} + \delta_{\epsilon_i}$. Then, we have:

$$\begin{aligned} & (\mu_{v_i}^{(0)} - \mathbb{E}_{v_i}[\bar{y}_{i,post}^{(1)} | \hat{\xi}_{C,i}]) \Pr[\hat{\xi}_{C,i}] \\ &= (\mu_{v_i}^{(0)} - \mathbb{E}_{v_i}[\bar{y}_{i,post}^{(1)} | \hat{\xi}_{C,i}, \xi]) \Pr[\hat{\xi}_{C,i}, \xi] + (\mu_{v_i}^{(0)} - \mathbb{E}_{v_i}[\bar{y}_{i,post}^{(1)} | \hat{\xi}_{C,i}, \neg \xi]) \Pr[\hat{\xi}_{C,i}, \neg \xi] \\ &\geq (\mu_{v_i}^{(0)} - \mathbb{E}_{v_i}[\bar{y}_{i,post}^{(1)} | \hat{\xi}_{C,i}, \xi]) \Pr[\hat{\xi}_{C,i}, \xi] - 2\delta \quad (\text{since } \Pr[\neg \xi] < \delta \text{ and } \mu_{v_i}^{(0)} - \mathbb{E}_{v_i}[\bar{y}_{i,post}^{(1)}] \geq -2) \end{aligned}$$

We proceed to lower bound the expression above as follows:

$$\begin{aligned} & \mu_{v_i}^{(0)} - \mathbb{E}_{v_i}[\bar{y}_{i,post}^{(1)} | \hat{\xi}_{C,i}, \xi] \\ &= \mu_{v_i}^{(0)} - \mathbb{E}_{v_i} \left[\bar{y}_{i,post}^{(1)} | \mu_{v_i}^{(0)} \geq \hat{\mathbb{E}}[\bar{y}_{i,post}^{(1)}] + C, \left| \hat{\mathbb{E}}[\bar{y}_{i,post}^{(1)}] - \mathbb{E}[\bar{y}_{i,post}^{(1)}] \right| \leq \alpha(\delta_{PCR}), \left| \frac{1}{T_1} \sum_{t=T_0+1}^T \epsilon_{i,t}^{(1)} \right| \leq \sigma \sqrt{\frac{2 \log(1/\delta_{\epsilon_i})}{T_1}} \right] \\ &= \mu_{v_i}^{(0)} - \mathbb{E}_{v_i} \left[\bar{y}_{i,post}^{(1)} | \mu_{v_i}^{(0)} - \mathbb{E}[\bar{y}_{i,post}^{(1)}] \geq C - \alpha(\delta_{PCR}), \left| \frac{1}{T_1} \sum_{t=T_0+1}^T \epsilon_{i,t}^{(1)} \right| \leq \sigma \sqrt{\frac{2 \log(1/\delta_{\epsilon_i})}{T_1}} \right] \\ &= \mu_{v_i}^{(0)} - \mathbb{E}[v_i] \left[\bar{y}_{i,post}^{(1)} | \mu_{v_i}^{(0)} - \bar{y}_{i,post}^{(1)} \geq C - \alpha(\delta_{PCR}) - \sigma \sqrt{\frac{2 \log(1/\delta_{\epsilon_i})}{T_1}} \right] \\ &\geq C - \alpha(\delta_{PCR}) - \sigma \sqrt{\frac{2 \log(1/\delta_{\epsilon_i})}{T_1}} \end{aligned}$$

Furthermore, we can write the probability of the joint event $\hat{\xi}_{C,i}, \xi$ as:

$$\begin{aligned} & \Pr[\hat{\xi}_{C,i}, \xi] \\ &= \Pr \left[\mu_{v_i}^{(0)} \geq \hat{\mathbb{E}}[\bar{y}_{i,post}^{(1)}] + C, \left| \hat{\mathbb{E}}[\bar{y}_{i,post}^{(1)}] - \mathbb{E}[\bar{y}_{i,post}^{(1)}] \right| \leq \alpha(\delta_{PCR}), \left| \frac{1}{T_1} \sum_{t=T_0+1}^T \epsilon_{i,t}^{(1)} \right| \leq \sigma \sqrt{\frac{2 \log(1/\delta_{\epsilon_i})}{T_1}} \right] \\ &= \Pr \left[\mu_{v_i}^{(0)} - \bar{y}_{i,post}^{(1)} \geq C - \alpha(\delta_{PCR}) - \sigma \sqrt{\frac{2 \log(1/\delta_{\epsilon_i})}{T_1}} \right] \end{aligned}$$

Hence, we can derive the following lower bound on the denominator of L :

$$\begin{aligned} & \mu_{v_i}^{(0)} - \mathbb{E}_{v_i}[\bar{y}_{i,post}^{(1)} | \hat{\xi}_{C,i}] \\ &\geq \left(C - \alpha(\delta_{PCR}) - \sigma \sqrt{\frac{2 \log(1/\delta_{\epsilon_i})}{T_1}} \right) \Pr \left[\mu_{v_i}^{(0)} - \bar{y}_{i,post}^{(1)} \geq C - \alpha(\delta_{PCR}) - \sigma \sqrt{\frac{2 \log(1/\delta_{\epsilon_i})}{T_1}} \right] - 2\delta \end{aligned}$$

Applying this lower bound to the expression of L and taking the maximum over type 1 units, we have:

$$L \geq 1 + \max_{i \in \mathcal{I}^{(1)}} \left\{ \frac{\mu_{v_i}^{(1)} - \mu_{v_i}^{(0)}}{\left(C - \alpha(\delta_{PCR}) - \sigma \sqrt{\frac{2 \log(1/\delta_{\epsilon_i})}{T_1}} \right) \Pr \left[\mu_{v_i}^{(0)} - \bar{y}_{i,post}^{(1)} \geq C - \alpha(\delta_{PCR}) - \sigma \sqrt{\frac{2 \log(1/\delta_{\epsilon_i})}{T_1}} \right] - 2\delta} \right\}$$

□

B Appendix for Section 4: Extension to Synthetic Interventions

Theorem B.1. Suppose that Assumption 4.2 holds for some constant gap $C \in (0, 1)$. If the number of initial units N_0 is chosen to be large enough such that *Unit Overlap Assumption* is satisfied for all units of subtype τ under intervention $\tau[1]$ with probability $1 - \delta$, then Algorithm 2 with parameters δ, L, B, C is BIC for all units of subtype τ if:

$$L \geq 1 + \max_{i \in \mathcal{T}^{(\tau)}} \left\{ \frac{\mu_i^{(\tau[1])} - \mu_i^{(\tau[k])}}{\left(C - 2\alpha(\delta_{PCR}) - 2\sigma \sqrt{\frac{2 \log(1/\delta_{\epsilon_i})}{T_1}} \right) (1 - \delta) \Pr[\mathcal{E}_{C,i}] - 2\delta} \right\}$$

Moreover, if the number of batches B is chosen to be large enough such that with probability at least $1 - \delta$, we have $\text{rank}([\mathbb{E}[\mathbf{y}_{i,pre}]^\top : \hat{d}_i = \ell \neq \tau[1], i \in \mathcal{T}^{(\tau)}]_{i=1}^{N_0+B \cdot L}) = \text{rank}([\mathbb{E}[\mathbf{y}_{j,pre}]^\top]_{j \in \mathcal{T}^{(\tau)}})$, then the *Unit Overlap Assumption* will be satisfied for all units of subtype τ under all interventions with probability at least $1 - \mathcal{O}(k\delta)$.

Proof. According to our recommendation policy, unit i is either recommended intervention $\tau[1]$ or intervention ℓ . We will prove that in either case, unit i will comply with the principal's recommendation.

Let $\hat{\mathcal{E}}_{C,i}^{(\ell)}$ denote the event that the exploit intervention for unit i is intervention ℓ . Formally, we have

$$\hat{\mathcal{E}}_{C,i}^{(\ell)} = \left\{ \hat{\mathbb{E}}[\bar{y}_{i,post}^{(\tau[1])}] \leq \min_{1 < j < \ell} \hat{\mathbb{E}}[\bar{y}_{i,post}^{(\tau[j])}] - C \quad \text{and} \quad \max_{1 \leq j < \ell} \hat{\mathbb{E}}[\bar{y}_{i,post}^{(\tau[j])}] \leq \mu_{v_i}^{(\ell)} - C \right\}$$

When unit i is recommended intervention ℓ : When a unit $i \in \mathcal{T}^{(\tau)}$ is recommended intervention ℓ , we argue that this unit will not switch to any other intervention $j \neq \ell$. Because of our ordering of the prior mean reward, any intervention $j > \ell$ has had no sample collected by a type τ unit and $\mu_{v_i}^{(\tau[j])} \leq \mu_{v_i}^{(\tau[\ell])}$. Hence, we only need to focus on the cases where $j < \ell$. For the recommendation policy to be BIC, we need to show that

$$\mathbb{E}[\mu_{v_i}^{(\ell)} - \bar{y}_{i,post}^{(\tau[j])} | \hat{d} = \ell] \Pr[\hat{d} = \ell] \geq 0$$

There are two possible disjoint events under which unit i is recommended intervention ℓ : either ℓ is determined to be the 'exploit' intervention or unit i is chosen as an 'explore' unit. Since being chosen as an explore unit does not imply any information about the rewards, we can derive that conditional on being in an 'explore' unit, the expected gain for unit i to switch to intervention j is simply $\mu_{v_i}^{(\ell)} - \mu_{v_i}^{(\tau[j])}$. On the other hand, if intervention ℓ is the 'exploit' intervention, then event $\hat{\mathcal{E}}_{C,i}^{(\ell)}$ has happened. Hence, we can rewrite the left-hand side of the BIC condition above as:

$$\mathbb{E}[\mu_{v_i}^{(\ell)} - \bar{y}_{i,post}^{(\tau[j])} | \hat{d} = \ell] \Pr[\hat{d} = \ell] = \mathbb{E}[\mu_{v_i}^{(\ell)} - \bar{y}_{i,post}^{(\tau[j])} | \hat{\mathcal{E}}_{C,i}^{(\ell)}] \Pr[\hat{\mathcal{E}}_{C,i}^{(\ell)}] \left(1 - \frac{1}{L} \right) + \frac{1}{L} (\mu_{v_i}^{(\ell)} - \mu_{v_i}^{(j)})$$

Rearranging the terms, we have the following lower bound on the phase length L for the algorithm to be BIC for type 1 units:

$$L \geq 1 + \frac{\mu_{v_i}^{(\tau[j])} - \mu_{v_i}^{(\ell)}}{\mathbb{E}[\mu_{v_i}^{(\ell)} - \bar{y}_{i,post}^{(\tau[j])} | \hat{\mathcal{E}}_{C,i}^{(\ell)}] \Pr[\hat{\mathcal{E}}_{C,i}^{(\ell)}]}$$

To complete the analysis, we need to lower bound the denominator of the expression above. Consider the following event \mathcal{C} :

$$\mathcal{C} = \{ \forall j \in [k] : |\hat{\mathbb{E}}[\bar{y}_i^{(\tau[j])}] - \mathbb{E}[\bar{y}_{i,post}^{(\tau[j])}]| \leq \alpha(\delta_{PCR}) \}$$

Let \mathcal{E}_0 denote the event that [Unit Overlap Assumption](#) is satisfied for units of type τ under intervention $\tau[1]$. From Assumption 4.2, we know that event \mathcal{E}_0 occurs with probability at least $1 - \delta_0$ for some $\delta_0 \in (0, 1)$. Furthermore, let \mathcal{E}_{ϵ_i} denote the event where the Chernoff-Hoeffding bound on the noise sequences $\{\epsilon_{i,t}^{(\tau[j])}\}_{t=T_0+1}^T$ holds, that is with probability at least $1 - \delta_{\epsilon_i}$, we have:

$$\left| \frac{1}{T_1} \sum_{t=T_0+1}^T \epsilon_{i,t}^{(\tau[j])} \right| \leq \sigma \sqrt{\frac{2 \log(1/\delta_{\epsilon_i})}{T_1}}$$

Then, we can define a clean event \mathcal{E} where the events \mathcal{C} , \mathcal{E}_0 and \mathcal{E}_{ϵ_i} happens simultaneously with probability at least $1 - \delta$, where $\delta = \delta_{\epsilon}^{PCR} + \delta_{\epsilon_i} + \delta_0$.

Note that since $|\bar{y}_{i,post}^{(\tau[j])}| \leq 1$, we can rewrite the denominator as:

$$\begin{aligned} & \mathbb{E}[\mu_{v_i}^{(\ell)} - \bar{y}_{i,post}^{(\tau[j])} | \hat{\mathcal{E}}_{C,i}^{(\ell)}] \Pr[\hat{\mathcal{E}}_{C,i}^{(\ell)}] \\ &= \mathbb{E}[\mu_{v_i}^{(\ell)} - \bar{y}_{i,post}^{(\tau[j])} | \hat{\mathcal{E}}_{C,i}^{(\ell)}, \mathcal{E}] \Pr[\hat{\mathcal{E}}_{C,i}^{(\ell)}, \mathcal{E}] + \mathbb{E}[\mu_{v_i}^{(\ell)} - \bar{y}_{i,post}^{(\tau[j])} | \hat{\mathcal{E}}_{C,i}^{(\ell)}, \neg \mathcal{E}] \Pr[\hat{\mathcal{E}}_{C,i}^{(\ell)}, \neg \mathcal{E}] \\ &\geq \mathbb{E}[\mu_{v_i}^{(\ell)} - \bar{y}_{i,post}^{(\tau[j])} | \hat{\mathcal{E}}_{C,i}^{(\ell)}, \mathcal{E}] \Pr[\hat{\mathcal{E}}_{C,i}^{(\ell)}, \mathcal{E}] - 2\delta \end{aligned}$$

Define an event $\mathcal{E}_i^{(\ell)}$ on the true average expected post-intervention outcome of each intervention as follows:

$$\begin{aligned} \mathcal{E}_i^{(\ell)} := & \left\{ \bar{y}_{i,post}^{(\tau[1])} \leq \min_{1 \leq j < \ell} \bar{y}_{i,post}^{(\tau[j])} - C - 2\alpha(\delta_{PCR}) - 2\sigma \sqrt{\frac{2 \log(1/\delta_{\epsilon_i})}{T_1}} \right. \\ & \left. \text{and } \max_{1 \leq j < \ell} \bar{y}_{i,post}^{(\tau[j])} \leq \mu_{v_i}^{(\ell)} - C - 2\alpha(\delta_{PCR}) - 2\sigma \sqrt{\frac{2 \log(1/\delta_{\epsilon_i})}{T_1}} \right\} \end{aligned}$$

We can observe that under event \mathcal{E} , event $\hat{\mathcal{E}}_{C,i}^{(\ell)}$ is implied by event $\mathcal{E}_i^{(\ell)}$. Hence, we have:

$$\Pr[\mathcal{E}, \hat{\mathcal{E}}_{C,i}^{(\ell)}] \leq \Pr[\mathcal{E}, \mathcal{E}_i^{(\ell)}]$$

We can rewrite the left-hand side as:

$$\Pr[\mathcal{C}, \mathcal{E}_i] = \Pr[\mathcal{C} | \mathcal{E}_i^{(\ell)}] \Pr[\mathcal{E}_i^{(\ell)}] \geq (1 - \delta) \Pr[\mathcal{E}_i^{(\ell)}]$$

Substituting these expressions using $\mathcal{E}_i^{(\ell)}$ for the ones using $\hat{\mathcal{E}}_{C,i}^{(\ell)}$ in the denominator gives:

$$\mathbb{E}[\mu_{v_i}^{(\ell)} - \bar{y}_{i,post}^{(\tau[j])} | \hat{\mathcal{E}}_{C,i}^{(\ell)}] \Pr[\hat{\mathcal{E}}_{C,i}^{(\ell)}] \geq \left(C - 2\alpha(\delta_{PCR}) - 2\sigma \sqrt{\frac{2 \log(1/\delta_{\epsilon_i})}{T_1}} \right) (1 - \delta) \Pr[\mathcal{E}_i^{(\ell)}] - 2\delta$$

Applying this lower bound to the expression of L and taking the maximum over type 1 units, we have:

$$L \geq 1 + \max_{i \in \mathcal{I}^{(1)}} \left\{ \frac{\mu_{v_i}^{(\tau[1])} - \mu_{v_i}^{(\tau[k])}}{\left(C - 2\alpha(\delta_{PCR}) - 2\sigma \sqrt{\frac{2 \log(1/\delta_{\epsilon_i})}{T_1}} \right) (1 - \delta) \Pr[\mathcal{E}_{C,i}^{(\ell)}] - 2\delta} \right\}$$

When unit i is recommended intervention $\tau[1]$: When unit i gets recommended intervention $\tau[1]$, they know that they are not in the explore group. Hence, the event $\hat{\mathcal{E}}_{C,i}$ did not happen, and the BIC condition, in this case, can be written as: for any intervention $\tau[j] \neq \tau[1]$,

$$\mathbb{E}[\bar{y}_{i,post}^{(\tau[1])} - \bar{y}_{i,post}^{(\tau[j])} | \neg \hat{\mathcal{E}}_{C,i}^{(\ell)}] \Pr[\neg \hat{\mathcal{E}}_{C,i}^{(\ell)}] \geq 0$$

Similar to the previous analysis on the recommendation of intervention ℓ , it suffices to only consider interventions $\tau[j] < \ell$. We have:

$$\begin{aligned}\mathbb{E}[\bar{y}_{i,\text{post}}^{(\tau[1])} - \bar{y}_{i,\text{post}}^{(\tau[j])} | \neg \widehat{\mathcal{E}}_{C,i}^{(\ell)}] \Pr[\neg \widehat{\mathcal{E}}_{C,i}^{(\ell)}] &= \mathbb{E}[\bar{y}_{i,\text{post}}^{(\tau[1])} - \bar{y}_{i,\text{post}}^{(\tau[j])}] - \mathbb{E}[\bar{y}_{i,\text{post}}^{(\tau[1])} - \bar{y}_{i,\text{post}}^{(\tau[j])} | \widehat{\mathcal{E}}_{C,i}^{(\ell)}] \Pr[\widehat{\mathcal{E}}_{C,i}^{(\ell)}] \\ &= \mu_{v_i}^{(\tau[1])} - \mu_{v_i}^{(\tau[j])} + \mathbb{E}[\bar{y}_{i,\text{post}}^{(\tau[j])} - \bar{y}_{i,\text{post}}^{(\tau[1])} | \widehat{\mathcal{E}}_{C,i}^{(\ell)}] \Pr[\widehat{\mathcal{E}}_{C,i}^{(\ell)}]\end{aligned}$$

By definition, we have $\mu_{v_i}^{(\tau[1])} \geq \mu_{v_i}^{(\tau[j])}$. Hence, it suffices to show that for any intervention $1 < \tau[j] < \ell$, we have:

$$\mathbb{E}[\bar{y}_{i,\text{post}}^{(\tau[j])} - \bar{y}_{i,\text{post}}^{(\tau[1])} | \widehat{\mathcal{E}}_{C,i}^{(\ell)}] \Pr[\widehat{\mathcal{E}}_{C,i}^{(\ell)}] \geq 0$$

Observe that with the event \mathcal{E} defined above, we can write:

$$\begin{aligned}&\mathbb{E}[\bar{y}_{i,\text{post}}^{(\tau[j])} - \bar{y}_{i,\text{post}}^{(\tau[1])} | \widehat{\mathcal{E}}_{C,i}^{(\ell)}] \Pr[\widehat{\mathcal{E}}_{C,i}^{(\ell)}] \\ &= \mathbb{E}[\bar{y}_{i,\text{post}}^{(\tau[j])} - \bar{y}_{i,\text{post}}^{(\tau[1])} | \widehat{\mathcal{E}}_{C,i}^{(\ell)}, \mathcal{E}] \Pr[\widehat{\mathcal{E}}_{C,i}^{(\ell)}, \mathcal{E}] + \mathbb{E}[\bar{y}_{i,\text{post}}^{(\tau[j])} - \bar{y}_{i,\text{post}}^{(\tau[1])} | \widehat{\mathcal{E}}_{C,i}^{(\ell)}, \neg \mathcal{E}] \Pr[\widehat{\mathcal{E}}_{C,i}^{(\ell)}, \neg \mathcal{E}] \\ &\geq \mathbb{E}[\bar{y}_{i,\text{post}}^{(\tau[j])} - \bar{y}_{i,\text{post}}^{(\tau[1])} | \widehat{\mathcal{E}}_{C,i}^{(\ell)}, \mathcal{E}] \Pr[\widehat{\mathcal{E}}_{C,i}^{(\ell)}, \mathcal{E}] - 2\delta\end{aligned}$$

When event $\widehat{\mathcal{E}}_{C,i}^{(\ell)}$ happens, we know that $\widehat{\mathbb{E}}[\bar{y}_{i,\text{post}}^{(\tau[j])}] \geq \widehat{\mathbb{E}}[\bar{y}_{i,\text{post}}^{(\tau[1])}] + C$. Furthermore, when event \mathcal{E} happens, we know that $\bar{y}_{i,\text{post}}^{(\tau[j])} \geq \widehat{\mathbb{E}}[\bar{y}_{i,\text{post}}^{(\tau[j])}] - \alpha(\delta_{PCR}) - \sigma\sqrt{\frac{2\log(1/\delta_{\epsilon_i})}{T_1}}$ and $\widehat{\mathbb{E}}[\bar{y}_{i,\text{post}}^{(\tau[1])}] \geq \bar{y}_{i,\text{post}}^{(\tau[1])} - \alpha(\delta_{PCR}) - \sigma\sqrt{\frac{2\log(1/\delta_{\epsilon_i})}{T_1}}$. Hence, when these two events $\widehat{\mathcal{E}}_{C,i}^{(\ell)}$ and \mathcal{E} happen simultaneously, we have $\bar{y}_{i,\text{post}}^{(\tau[j])} \geq \bar{y}_{i,\text{post}}^{(\tau[1])} + C - 2\alpha(\delta_{PCR}) - 2\sigma\sqrt{\frac{2\log(1/\delta_{\epsilon_i})}{T_1}}$. Therefore, the lower bound can be written as:

$$\mathbb{E}[\bar{y}_{i,\text{post}}^{(\tau[j])} - \bar{y}_{i,\text{post}}^{(\tau[1])} | \widehat{\mathcal{E}}_{C,i}^{(\ell)}] \Pr[\widehat{\mathcal{E}}_{C,i}^{(\ell)}] \geq \left(C - 2\alpha(\delta_{PCR}) - 2\sigma\sqrt{\frac{2\log(1/\delta_{\epsilon_i})}{T_1}} \right) \Pr[\widehat{\mathcal{E}}_{C,i}^{(\ell)}, \mathcal{E}] - 2\delta$$

Similar to the previous analysis when unit i gets recommended intervention ℓ , we have:

$$\mathbb{E}[\bar{y}_{i,\text{post}}^{(\tau[j])} - \bar{y}_{i,\text{post}}^{(\tau[1])} | \widehat{\mathcal{E}}_{C,i}^{(\ell)}] \Pr[\widehat{\mathcal{E}}_{C,i}^{(\ell)}] \geq \left(C - 2\alpha(\delta_{PCR}) - 2\sigma\sqrt{\frac{2\log(1/\delta_{\epsilon_i})}{T_1}} \right) (1 - \delta) \Pr[\mathcal{E}_i^{(\ell)}] - 2\delta$$

Choosing a large enough C such that the right-hand side is non-negative, we conclude the proof. \square

C Appendix for Section 5: Testing Whether the Unit Overlap Assumption Holds

Theorem C.1. Under Assumption 5.1, if

$$\frac{r^{3/2} \sqrt{\log(T_0 |\mathcal{I}|)}}{\|\tilde{\omega}^{(n,0)}\|_2 \cdot \min\{T_0, |\mathcal{I}|, T_0^{1/4} |\mathcal{I}|^{1/2}\}} = o(1) \quad (2)$$

and

$$\frac{1}{\sqrt{T_1} \|\tilde{\omega}^{(n,0)}\|_2} \sum_{t=T_0+1}^T \sum_{i \in \mathcal{I}} \mathbb{E}[y_{i,t}^{(0)}] \cdot (\hat{\omega}^{(n,0)}[i] - \tilde{\omega}^{(n,0)}[i]) = o_p(1) \quad (3)$$

then as $|\mathcal{I}|, T_0, T_1 \rightarrow \infty$ the Asymptotic Hypothesis Test falsely accepts the hypothesis with probability at most 5%, where $\hat{\sigma}^2$ is defined as in Equation (4) and $\hat{\omega}^{(n,0)}$ is defined as in Equation (5) in Agarwal et al. [4].

Condition (2) requires that the ℓ_2 norm of $\tilde{\omega}^{(n,0)}$ is sufficiently large, and condition (3) requires that the estimation error of $\hat{\omega}^{(n,0)}$ decreases sufficiently fast. See Section 5.3 of Agarwal et al. [4] for more details.

Proof. We use $x \xrightarrow{p} y$ (resp. $x \xrightarrow{d} y$) if x converges in probability (resp. distribution) to y . Let σ' denote the true standard deviation of $\epsilon_{i,t}$. By Theorem 3 of Agarwal et al. [4] we know that as $|\mathcal{I}|, T_0, T_1 \rightarrow \infty$,

1. $\frac{\sqrt{T_1}}{\sigma' \|\hat{\omega}^{(n,0)}\|_2} (\hat{\mathbb{E}}[\bar{y}_{n,pre''}] - \mathbb{E}[\bar{y}_{n,pre''}]) \xrightarrow{d} \mathcal{N}(0, 1)$
2. $\hat{\sigma} \xrightarrow{p} \sigma'$ and
3. $\hat{\omega}^{(n,0)} \xrightarrow{p} \tilde{\omega}^{(n,0)}$.

By the continuous mapping theorem, we know that $\frac{1}{\hat{\sigma}} \xrightarrow{p} \frac{1}{\sigma'}$ and $\frac{1}{\|\hat{\omega}^{(n,0)}\|_2} \xrightarrow{p} \frac{1}{\|\tilde{\omega}^{(n,0)}\|_2}$. We can rewrite $\hat{\mathbb{E}}[\bar{y}_{n,pre''}] - \mathbb{E}[\bar{y}_{n,pre''}]$ as $\hat{\mathbb{E}}[\bar{y}_{n,pre''}] - \bar{y}_{n,pre''} + \bar{\epsilon}_{n,pre''}$, and we know that $\bar{\epsilon}_{n,pre''} \xrightarrow{p} 0$. Applying Slutsky's theorem several times obtains the desired result. \square