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

Daniel Ngo*[1], Keegan Harris*[2], Anish Agarwal[3], Vasilis Syrgkanis[4], and Zhiwei Steven Wu[2]

[1]University of Minnesota
[2]Carnegie Mellon University
[3]Columbia University
[4]Stanford 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 units under different interventions by using data from other units who have received the intervention of interest (i.e. *donor* units). In order for the counterfactual estimate produced by synthetic control for a *test* unit to be accurate, the data from donor units must be sufficiently diverse enough for a reasonable representation of the test unit's outcomes to be learned. 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 assumptions may not hold in practice, as is the case when e.g. units select their own interventions and different subpopulations of units prefer different interventions. We shed light on this typically overlooked assumption, and we address this issue by *incentivizing* units of different types to take interventions they would normally not consider. Specifically, we provide an algorithm for incentivizing exploration in panel data settings using 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 assumption on the variability present in the donor pool.

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*. Synthetic control methods (SCMs) [1, 2] 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 different interventions, 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 received the relevant intervention (i.e. the *donor* units).

*Denotes equal contribution.

Once the model is learned, it is then extrapolated to the post-intervention time period in order to predict the counterfactual outcomes of the test unit. 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 [28, 15] to big tech [32, 13].

In order for the counterfactual estimate produced by synthetic control to be accurate, it should be the case that the test unit's outcomes may be expressed "reasonably well" in terms of the outcomes of the donor units. When providing statistical guarantees on the performance of SCMs, such intuition has traditionally been made formal by making a *overlap* assumption on the relationship between the donor units and the test, 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 any randomness in the unit outcomes.*

In other words, previous work on SCMs usually 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 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 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 than those who pay as they go. This makes drawing conclusions about counterfactual user engagement levels of the subpopulation who did not undergo difficult, as they may not get as much use out of the subscription when compared the subpopulation who subscribed, due to their differing tastes. While the theater chain 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 exploration of different treatments by non-overlapping subpopulations in order to obtain valid counterfactual estimates using synthetic control methods. Specifically, we adopt tools and techniques from the literature on incentivizing exploration in multi-armed bandits [23, 25, 31] to show how the *learner* (e.g. the person running the SCM) can use knowledge gained from previous interactions with participating units to send a credible *signal* to the current units in order to *convince* them to take interventions such that the unit overlap condition becomes satisfied. In our streaming service example, such credible signaling may correspond to messaging select users to recommend that they sign up for the subscription.

Concretely, we introduce a game-theoretic model to study the dynamics of incentivizing compliance when using synthetic control methods for counterfactual estimation. At a high level, we provide a *recommendation policy* (i.e. a mapping from pre-intervention outcomes to interventions) that gradually incentivizes unit compliance over time in order to satisfy the unit overlap condition needed to perform valid counterfactual inference.

[1] More broadly, similar assumptions are also prevalent in the literature on other *matching-style* estimators typically used in panel data settings, such as *difference-in-differences* [12] or *clustering*-based methods [34].

Overview of Our Results We cover related work on learning from panel data and incentivized exploration in Section 1.1. In Section 2 we introduce our model and provide relevant background on synthetic control methods. We introduce our algorithm for incentivizing exploration for synthetic control when there are two interventions in Section 3, and provide finite sample guarantees for counterfacutal estimation when it is used to assign interventions to units. We empirically evaluate the performance of our incentive-aware synthetic control estimator in Section 4, and find that it compares favorably to methods which do not incentivize exploration whenever the unit overlap assumption does not hold *a priori*. Finally, we mention directions for future research in Section 5.

1.1 Related Work

Synthetic Control Methods Within the literature on synthetic control [1, 2], 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. [10, 22, 8]) and leverages *principal component regression* (PCR) [19, 27] 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 [11, 14, 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. [17] 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.

Incentivized exploration (IE) Our work draws on techniques from the growing literature on incentivizing exploration [21, 24, 26, 30, 18, 29], where the goal is to incentivize myopic agents to explore different arms in various bandit settings. Incentivized exploration is related to the literature on Bayesian persuasion [20, 9], as each round of incentivized exploration is an instance of a one-shot Bayesian persuasion game. Following the framework of Kremer et al. [21], 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 [16]. Within the literature on incentivized exploration, the work most related to our work is Mansour et al. [24]. While our mechanisms for the initial exploration phase are technically similar to those in Mansour et al. [24], our work considers a more general setting where unit outcomes may vary over time, compared to the 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 $[\![x]\!] := \{1, 2, \ldots, x\}$ and $[\![x]\!]_0 := \{0, 1, \ldots, x-1\}$. $\mathbf{y}[i]$ 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 $T, T_0 \in \mathbb{N}_{>0}$ and $T > T_0$. $\Delta(\mathcal{X})$ denotes the space of possible probability distributions over the set $\mathcal{X}$. Finally, we use the notation $a \wedge b := \min\{a, b\}$ and $a \vee b := \max\{a, b\}$.

Learning in panel data settings We consider a panel data 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 recommends one of k *interventions* $\widehat{d}_i \in [\![k]\!]_0$ to each unit $i \in [\![n]\!]$. Without loss of generality, we denote the control by 0. After receiving the recommendation, unit i chooses an intervention d_i and remains under intervention d_i for the remaining $T - T_0$ time-steps. We use

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

to refer to unit i's pre-treatment outcomes under control, and

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

to refer to unit i's post-intervention outcomes under intervention d. We denote the set of possible pre-treatment outcomes by $\mathcal{Y}_{pre}$. We assume that unit 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 $d \in [\![k]\!]_0$ takes the following factorized form*

$$y_{i,t}^{(d)} = \langle \mathbf{u}_t^{(d)}, \mathbf{v}_i \rangle + \varepsilon_{i,t}^{(d)}$$

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 σ^2. For simplicity, we assume that $|\mathbb{E}[y_{i,t}^{(d)}]| \leq 1$, $\forall\, i \in [\![n]\!], t \in [\![T]\!], d \in [\![k]\!]_0$.

The goal of the principal is to estimate the unit-specific counterfactual outcomes under different interventions. In line with previous work on synthetic control, the target causal parameter is the counterfactual average expected post-treatment outcome.

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

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

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

In order to infer something about unit outcomes in the post-intervention time period from outcomes in the pre-intervention time period, it should be the case that the time and intervention latent factors in the pre-intervention time period are "sufficiently diverse".[2] A popular way to formalize such intuition in the literature on robust synthetic control (see, e.g. [4, 5, 17]) is through the following *linear span inclusion* assumption on the latent factors in the post-intervention time period.

Assumption 2.3. *For any intervention $d \in [\![k]\!]_0$ and time $t > T_0$, we assume that*

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

Under Assumption 2.1 and 2.3, the average expected post-treatment outcome for any unit i may be written as a linear combination of the expected pre-treatment outcomes of unit i.

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

$$\mathbb{E}[\bar{y}_{i,post}^{(d)}] = \frac{1}{T - T_0} \langle \theta^{(d)}, \mathbb{E}[y_{i,pre}] \rangle$$

[2] Consider 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.

Principal component regression Let $Y^{(d)}_{pre,i} := [\mathbf{y}^\top_{j,pre} \ : \ j \in \mathcal{I}^{(d)}_i] \in \mathbb{R}^{n^{(d)}_i \times T_0}$ be the matrix of pre-treatment outcomes corresponding to the subset of units who have undergone intervention d before unit i arrives. Similarly, let $\bar{Y}^{(d)}_{post,i} := [\bar{y}^\top_{j,post} : j \in \mathcal{I}^{(d)}_i] \in \mathbb{R}^{n^{(d)}_i \times 1}$ be the column of average post-intervention outcomes corresponding to the subset of units who have undergone intervention d before unit i arrives. We denote the singular value decomposition of $Y^{(d)}_{pre,i}$ as $Y^{(d)}_{pre,i} = \sum_{\ell=1}^{n^{(d)}_i \wedge T_0} s^{(d)}_\ell \widehat{\mathbf{u}}^{(d)}_\ell (\widehat{\mathbf{v}}^{(d)}_\ell)^\top$, where $\{s^{(d)}_\ell\}_{\ell=1}^{n^{(d)}_i \wedge T_0}$ are the singular values of $Y^{(d)}_{pre,i}$, and $\widehat{\mathbf{u}}^{(d)}_\ell$ and $\widehat{\mathbf{v}}^{(d)}_\ell$ are orthonormal column vectors. We assume that $s_1(Y^{(d)}_{pre,i}) \geq \cdots \geq s_{n^{(d)}_i \wedge T_0}(Y^{(d)}_{pre,i}) \geq 0$. For some threshold value r, we define the "de-noised" version of $Y^{(d)}_{pre,i}$ as

$$\begin{aligned}\widehat{Y}^{(d)}_{pre,i} &:= \sum_{\ell=1}^{r} s^{(d)}_\ell \widehat{\mathbf{u}}^{(d)}_\ell (\widehat{\mathbf{v}}^{(d)}_\ell)^\top \\ &= [\widehat{\mathbf{y}}^\top_{j,pre} \ : \ j \in \mathcal{I}^{(d)}_i] \in \mathbb{R}^{n^{(d)}_i \times T_0}.\end{aligned}$$

We define the projection matrix onto the subspace spanned by the top r right singular vectors as $\widehat{\mathbf{P}}^{(d)}_{i,r} \in \mathbb{R}^{r \times r}$ given by $\widehat{\mathbf{P}}^{(d)}_{i,r} := \sum_{\ell=1}^{r} \widehat{\mathbf{v}}^{(d)}_\ell \left(\widehat{\mathbf{v}}^{(d)}_\ell\right)^\top$. Equipped with this notation, we are now ready to define the procedure for estimating $\theta^{(d)}$ 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 $d \in [k]_0$ and $i \geq 1$, let $\mathcal{V}^{(d)}_i := \left(\widehat{Y}^{(d)}_{pre,i}\right)^\top \widehat{Y}^{(d)}_{pre,i} + \rho \widehat{\mathbf{P}}^{(d)}_{i,r}$. Then, the regularized principal component regression estimates $\theta^{(d)}$ as:*

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

History and recommendation policy The interaction between the principal and a unit i may be characterized by the tuple $(\mathbf{y}_{i,pre}, \widehat{d}_i, d_i, \mathbf{y}^{(d_i)}_{i,post})$. Recall that $\mathbf{y}_{i,pre}$ are unit i's pre-treatment outcomes, $\widehat{d}_i$ is the intervention recommended to unit i, d_i is the intervention taken by unit i, and $\mathbf{y}^{(d_i)}_{i,post}$ 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 $j \in [\![i-1]\!]$. Formally,*

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

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

Definition 2.7 (Recommendation Policy)**.** *A recommendation policy $\pi_i : \mathcal{H}_i \times \mathcal{Y}_{pre} \to \Delta([\![k]\!]_0)$ is a (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 $\widehat{d}_i = \pi_i(\mathbf{y}_{i,pre})$ to denote the recommendation of policy π_i to unit i.

Beliefs, incentives, and intervention choices In addition to having a corresponding latent factor, we say that the *type* of a unit i is their preferred intervention under no recommendation from the principle. Note that there are at most k unit types. We denote the set of all units of type d as $\mathcal{I}^{(d)}$. Throughout the sequel, 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).[3] We made the following assumptions on the *beliefs* of all units.

[3]If 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.

Assumption 2.8 (Unit Beliefs)**.** *We assume that each unit knows its place in the sequence of n units (i.e., their index $i \in [\![n]\!]$). Furthermore, we assume that (i) agents are Bayesian-rational, (ii) agents aim to maximize their average expected post-intervention outcome, and (iii) each unit's private type $\mathbf{v}_i$ determines their* prior belief *$\mathcal{P}_{\mathbf{v}_i}$, which is a joint distribution over $\{\mathbb{E}[\mathbf{y}_{i,post}^{(d)}]\}_{d=0}^{k}$.*

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 $\mathcal{P}_{\mathbf{v}_i}$. Given recommendation $\widehat{d}_i$, under Assumption 2.8 unit i selects their intervention d_i such that $d_i \in \arg\max_{d \in [\![k]\!]_0} \mathbb{E}_{\mathcal{P}_{\mathbf{v}_i}}[\bar{y}_{i,post}^{(d)} | \widehat{d}_i]$. In other words, they select the intervention d_i in order to maximize their utility in expectation over $\mathcal{P}_{\mathbf{v}_i}$, conditioned on receiving recommendation $\widehat{d}_i$.

Definition 2.9 (Bayesian incentive compatibility)**.** *We say that a recommendation policy π is* Bayesian incentive compatible *for unit i if, conditional on receiving any intervention recommendation $\widehat{d}_i$, the unit's average expected post-intervention outcome under intervention $\widehat{d}_i$ is at least as large as their average expected post-intervention outcome under any other intervention:*

$$\mathbb{E}_{\mathcal{P}_{\mathbf{v}_i}}[\bar{y}_{i,post}^{(d)} - \bar{y}_{i,post}^{(d')} | \widehat{d}_i = d] \geq 0 \quad \forall \text{ interventions } d, d' \in [k] \text{ with } \Pr[\widehat{d}_i = d] > 0.$$

3 Incentivizing Exploration for Synthetic Control

We inherit the canonical synthetic control setting in which there are two interventions: control and a single treatment. Recall that without any additional information, type 0 units prefer control and type 1 units prefer the treatment. The goal of the principal is to provide a recommendation policy that convinces some units of type 1 to select the control in the post-treatment period, such that the Unit Overlap Assumption is satisfied for both interventions for all units of type 1.[4] Before explaining our algorithm, we make the following assumptions on the knowledge of the principal and units:

Assumption 3.1 (Knowledge Assumption for Algorithm 1)**.** *We assume that the following are common knowledge among all units and the principal:*

1. *Initially all type 1 units prefer the treatment and all type 0 units prefer control. The fraction of type 1 units in the population is $p_1 \in (0, 1)$ and the fraction of type 0 units is $1 - p_1$.*
2. *The smallest probability of the event $\hat{\xi}_C$ over the priors of type 1 units, denoted by $\Pr_{v_i : i \in \mathcal{I}^{(1)}}[\hat{\xi}_C] > 0$, where*
$$\hat{\xi}_C = \left\{ \mu_{v_i}^{(0)} \geq \hat{y}_{i,post}^{(1)} + C \right\}$$
for some prior-dependent constant C and the estimated post-treatment outcome $\hat{y}_{i,post}^{(1)}$ for type 1 units under intervention 1.
3. *A sufficient number of observations N_0 needed such that the Unit Overlap Assumption is satisfied with probability at least $1 - \delta$ for some $\delta \in (0, 1)$.*

Our algorithm (Algorithm 1) is inspired by the "detail free' algorithm for incentivizing exploration in multi-armed bandit settings in Mansour et al. [23]. The recommendation policy in Algorithm 1 is split into two stages. In the first stage, the principal provides no recommendations and N_0 units take their preferred intervention according to their prior belief: type 0 units take control, and type 1 units take the treatment. The first stage length N_0 is chosen to be large enough such that the Unit Overlap Assumption is satisfied for all future units of type 1 under

[4] The methods we present may be straightforwardly applied to incentivize type 0 units to take the treatment. We focus on incentivizing control among type 1 units here for simplicity.

treatment with high probability. In the second stage, we use the set of initial samples collected from the first stage to construct a consistent estimator of the average expected outcomes for type 1 units under treatment using *principal component regression*. At a high level, in order to incentivize a type 1 unit i to try the control, we leverage the fact that (1) there is a non-zero chance under unit i's prior $\mathcal{P}_{\mathbf{v}_i}$ that $\bar{y}^{(0)}_{i,post} \geq \bar{y}^{(1)}_{i,post}$ and (2) the principal will be able to infer this given the set of observed outcomes for units in the first phase. By dividing the time horizon of 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 will be incentivized to follow the recommendation as long as L is large enough (i.e. the probability of the recommendation being an explore recommendation is low enough).

ALGORITHM 1: Incentivizing Exploration in Panel Data Settings: Type 1 units

Input: First stage length N_0, batch size L, number of batches B, failure probability δ, gap $C \in (0,1)$
Phase 1: Provide no recommendation to first N_0 units.
Phase 2:
for *batch* $b = 1, 2, \ldots, B$ **do**
  Select an explore index $i_b \in [L]$ uniformly at random.
  for $j = 1, 2, \ldots, L$ **do**
    if $j = i_b$ **then**
      Recommend intervention $\widehat{d}_{N_0+(b-1)\cdot L+j} = 0$ to unit $N_0 + (b-1)\cdot L + j$.
    else
      if $\mu^{(0)}_{v_j} - \hat{y}^{(1)}_{j,post} \geq C$ **then**
        Recommend intervention $\widehat{d}_{N_0+(b-1)\cdot L+j} = 0$ to unit $N_0 + (b-1)\cdot L + j$.
      else
        Recommend intervention $\widehat{d}_{N_0+(b-1)\cdot L+j} = 1$ to unit $N_0 + (b-1)\cdot L + j$.
      end
    end
  end
end

Theorem 3.2. *Suppose the there are two interventions, and assume that Assumption 3.1 holds for some constant gap $C > 0$. If N_0 is large enough such that Unit Overlap Assumption is satisfied for all units of type 1 with probability $1-\delta$, then Algorithm 1 with parameters δ, L, B, C is BIC for all units of type 1 (according to Definition 2.9) with probability at least $1-2\delta$ if*

$$L \geq 1 + \max_{i \in \mathcal{I}^{(1)}} \left\{ \frac{\mu^{(1)}_{v_i} - \mu^{(0)}_{v_i}}{\frac{C}{2}\left(1 - \delta_{PCR}\right) \Pr\left[\mu^{(0)}_{v_i} \geq y^{(1)}_{i,post} + \frac{3C}{2}\right] - 2\delta_{PCR}} \right\}$$

where

$$\delta_{PCR} \leq \frac{\log(n^{(d)}) \vee k}{\exp\left(\left(\sqrt{\frac{\sigma_r(Y^{(1)}_{pre,N_0})C/2 - (A+F)(\sqrt{N^{(1)}_0} + \sqrt{T_0})}{D}} + \frac{\alpha^2}{4D^2} - \frac{\alpha}{2D}\right)^2\right)}$$

where $\kappa\left(Y^{(1)}_{pre,N_0}\right) = \sigma_1\left(Y^{(1)}_{pre,N_0}\right)/\sigma_r\left(Y^{(1)}_{pre,N_0}\right)$ *is the condition number of the matrix of observed pre-intervention outcomes for type* 1 *units in phase* 1*. The remaining variables are defined as* $A = 3\sqrt{T_0}\left(\frac{\left\|\theta_i^{(1)}\right\|(\sqrt{74}+12\sqrt{6}\kappa(Y^{(1)}_{pre,N_0})}{T-T_0} + \frac{1}{\sqrt{T-T_0}}\right)$, $F = \frac{2\left\|\theta_i^{(1)}\right\|\sqrt{24T_0}}{T-T_0} + \frac{12\left\|\theta_i^{(1)}\right\|\kappa(Y^{(1)}_{pre,N_0})\sqrt{3T_0}}{T-T_0} + \frac{2}{\sqrt{T-T_0}}$, $D = \frac{\left\|\theta_i^{(1)}\right\|\sigma\sqrt{74}}{\sqrt{T-T_0}} + \frac{12\sigma\kappa(Y^{(1)}_{pre,N_0})\sqrt{6}}{\sqrt{T-T_0}} + \sigma\sqrt{2}$, $E = \frac{\left\|\theta_i^{(d)}\right\|\sigma}{\sqrt{T-T_0}}$, $\alpha = A+F+D(\sqrt{N_0^{(1)}}+\sqrt{T_0})+\sigma_r(Y^{(1)}_{pre,N_0})E$, *and* $\theta_i^{(1)}$ *is defined as in Proposition 2.4. Moreover if* B *is chosen to be large enough such that with probability at least* $1-\delta$, $\text{rank}([\mathbb{E}[y_{i,pre}]_{i:\widehat{d}_i=0}) = r$*, then the Unit Overlap Assumption will be satisfied for all type* 1 *units under control with probability at least* $1-3\delta$*.*

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 $\widehat{d}_i = 0$ for two reasons: (1) Under event $\hat{\xi}_C$ when intervention 0 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 ocurring, we can derive a condition on the minimum phase length L such that the expected gain from exploiting (when the event $\hat{\xi}_C$ happens) exceeds the expected loss from exploring. We 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. □

Theorem 3.2 says that after running Algorithm 1 with optimally-chosen parameters, the Unit Overlap Assumption will be satisfied for all future type 1 units for both interventions with probability at least $1-3\delta$. Therefore after running Algorithm 1, the principal can use off-the-shelf synthetic control methods (e.g. [4, 5]) to obtain valid finite-sample guarantees for new units with high probability.

4 Simulations

In this section, we complement our theoretical results with a numerical comparison to an ablation which does not take incentives into consideration.

Experimental Description We consider a setting with two interventions and two types of units: *type 1* units who initially prefer the treatment and *type 0* units who initially prefer control. If unit i is of type 1 (resp. type 0), we generate a latent factor $\mathbf{v}_i = [0\ v_i[1]]$ (resp. $\mathbf{v}_i = [v_i[0]\ 0]$), where $v_i[1] \sim \text{Unif}(-1,1)$ (resp. $v_i[0] \sim \text{Unif}(-1,1)$). We consider a setting where 500 units of alternating types arrive sequentially. (This is unknown to the algorithm.) 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 two with latent factors $\mathbf{u}_1^{(0)} = [1\ 10]$, $\mathbf{u}_2^{(0)} = [10\ 1]$ and a post-intervention time period of length one with latent factors $\mathbf{u}_3^{(0)} = [1\ 10]$, $\mathbf{u}_3^{(1)} = [10\ 1]$. Finally, outcomes are generated by adding independent Gaussian noise $\epsilon_{i,t}^{(d)} \sim \mathcal{N}(0, 0.1)$ to each inner product of latent factors.

Using Theorem 3.2, 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. After this, we run Algorithm 1 in batches of increasing size to observe the change in prediction error as we get more samples. The experiment is 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.

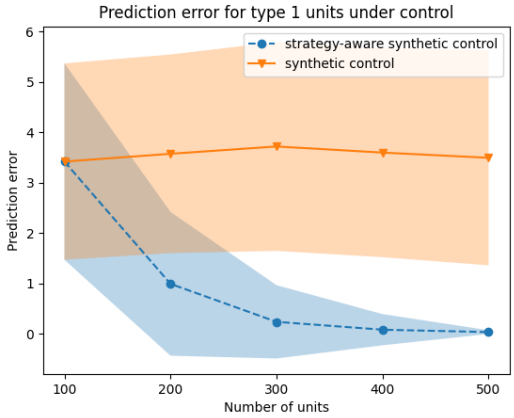

Figure 1: Counterfactual estimation error for units of type 1 under control using Algorithm 1 (blue) and synthetic control without incentives (orange). Results are averaged over 50 runs, with error bars representing one standard deviation.

Results In Figure 1, we compare the performance of Algorithm 1 (blue) with that of the synthetic control method of Agarwal et al. [5] which does not take incentives into consideration (orange). We set N_0 (the number of units to which we provide no recommendation to be 200). Initially, the performance of Algorithm 1 matches that of the synthetic control method which does not take incentives into consideration. However as more and more units of type 1 are incentivized to take the control, the counterfactual estimation error of Algorithm 1 decreases, while the estimation error of the method which does not consider incentives remains constant.

5 Conclusion and Future Work

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 a principled way to remove this assumption using tools from *incentivized exploration*. We complement our theoretical findings with simulation results, and observe that our strategy-aware synthetic control method significantly outperforms methods which do not take incentives into consideration.

Future Work An exciting direction for future work is to extend our results to include a second algorithm that improves the rate of exploration in order to obtain tighter finite sample guarantees. Another avenue for future work is to investigate more efficient algorithms for incentivizing exploration, such as in Sellke and Slivkins [30], Sellke [29].

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] Manuel Arellano and Bo Honore. Panel data models: Some recent developments. *Handbook of Econometrics*, 02 2000.

[9] 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`.

[10] 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`.

[11] Jiafeng Chen. Synthetic control as online linear regression. *Econometrica*, 91(2):465–491, 2023.

[12] 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.

[13] EBay, Jun 2022. URL `https://partnerhelp.ebay.com/helpcenter/s/article/Incrementality-Testing-by-Location-IP?language=en_US`.

[14] Vivek Farias, Ciamac Moallemi, Tianyi Peng, and Andrew Zheng. Synthetically controlled bandits. *arXiv preprint arXiv:2202.07079*, 2022.

[15] 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.

[16] 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.

[17] 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.

[18] Nicole Immorlica, Jieming Mao, Aleksandrs Slivkins, and Zhiwei Steven Wu. Incentivizing exploration with selective data disclosure, 2023.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] Yishay Mansour, Aleksandrs Slivkins, and Vasilis Syrgkanis. Bayesian incentive-compatible bandit exploration, 2019.

[25] Yishay Mansour, Aleksandrs Slivkins, and Vasilis Syrgkanis. Bayesian incentive-compatible bandit exploration. *Operations Research*, 68(4):1132–1161, 2020.

[26] Yishay Mansour, Aleksandrs Slivkins, Vasilis Syrgkanis, and Zhiwei Steven Wu. Bayesian exploration: Incentivizing exploration in bayesian games, 2021.

[27] William F Massy. Principal components regression in exploratory statistical research. *Journal of the American Statistical Association*, 60(309):234–256, 1965.

[28] 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.

[29] Mark Sellke. Incentivizing exploration with linear contexts and combinatorial actions, 2023.

[30] Mark Sellke and Aleksandrs Slivkins. The price of incentivizing exploration: A characterization via thompson sampling and sample complexity, 2022.

[31] Aleksandrs Slivkins. Exploration and persuasion, 2021.

[32] Uber, Jun 2019. URL https://www.uber.com/blog/causal-inference-at-uber/.

[33] 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.

[34] Yingying Zhang, Huixia Judy Wang, and Zhongyi Zhu. Quantile-regression-based clustering for panel data. *Journal of Econometrics*, 213(1):54–67, 2019.

A BIC proofs for two types setting

A.1 Motivating Example

We provide a detailed look at the motivating example in Section 1 where recovery of the counterfactual estimate is impossible.

Remark A.1 (Impossibility result without incentives)**.** *Given our setting with two types of units initially preferring different interventions, without the presence of incentives, all units of type* 1 *will only choose intervention* 1 *for any round* $t \in \mathcal{T}$ *and vice versa. Consider a toy example where we have one unit of each type, the pre-treatment period length is* $\mathcal{T}_0 = 2$*, and the post-treatment period length is* $\mathcal{T}_1 = 1$*. The latent vector for the type* 0 *unit is* $\mathbf{v}_0 = [0\ 1]^\top$ *and the latent vector for the type* 1 *unit is* $\mathbf{v}_1 = [1\ 0]^\top$*. Furthermore, we assume that* $\mathbf{u}_1^{(0)} = [1\ 0]^\top, \mathbf{u}_2^{(0)} = [0\ 1]^\top$ *and* $\mathbf{u}_3^{(0)} = [H\ 1]^\top$ *for some random variable* $H \sim Unif[-c, c]$*. Then, in this scenario, the expected outcomes for type* 0 *unit choosing intervention* 0 *are* $y_{0,1}^{(0)} = 0, y_{0,2}^{(0)} = 1, y_{0,3}^{(0)} = 1$ *and the expected outcomes for type* 1 *unit choosing intervention* 0 *are* $y_{1,1}^{(0)} = 1, y_{1,2}^{(0)} = 0, y_{1,3}^{(0)} = H$*. Suppose the principal wants to estimate* $y_{1,3}^{(0)}$ *using just the set of observed outcomes* $y_{0,1}^{(0)}, y_{0,2}^{(0)}, y_{0,3}^{(0)}, y_{1,1}^{(0)}$ *and* $y_{1,2}^{(0)}$*. Since the history does not contain any information about the random variable* H*, any estimator* $\hat{y}_{1,3}^{(0)}$ *would have a constant distance away from the true outcome* $y_{1,3}^{(0)}$*. That is,* $\mathbb{E}_{H\sim Unif[-c,c]}[|y_{1,3}^{(0)} - \hat{y}_{1,3}^{(0)}|] = c > 0$.

A.2 Causal parameter recovery derivation

Theorem A.2 (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* $\mathrm{rank}(\mathbf{X}_{i_0}](d)) = r$*, and* $\mathrm{snr}_i(d) \geq 2$ *for all* $i \geq i_0$*. Then, with probability at least* $1 - \mathcal{O}(k\delta)$*, simultaneously for all interventions* $d \in [k]_0$*,*

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

where $\hat{\mathbb{E}}[\bar{y}_{i,post}^{(d)}] \coloneqq \frac{1}{T - T_0} \cdot \langle \hat{\theta}_n(d), y_{n,pre} \rangle$ *is the estimated average post-intervention outcome for unit* i *under intervention* d*,* $\left\| \theta_i^{(d)} \right\| \leq L$.

Lemma A.3 (Well-balancing condition)**.** *Let* $A \in \mathbb{R}^{N\times r}$ *be a random matrix whose rows* A_i *are independent, mean zero, sub-gaussian isotropic random vectors in* $\mathbb{R}^r$*. Then,*

1. *with probability at least* $1 - 2\exp(-\frac{1}{2}\sqrt{Nr})$*, we have* $\mathrm{rank}(A) = r$ *and*

$$
\frac{\sigma_1(A)}{\sigma_r(A)} \leq \frac{1 + c_{Ver}K^2\sqrt{r/N} + c_{Ver}K^2(r/N)^{1/4}}{1 - c_{Ver}K^2\sqrt{r/N} - c_{Ver}K^2(r/N)^{1/4}}
$$

2. with probability at least $1-2\exp(\frac{-1}{2}\sqrt{Nr})$,

$$\|A\|_F^2 > Nr+c_{Ver}^2K^4r^2+c_{Ver}^2K^4\sqrt{N}r^{3/2}-2c_{Ver}K^2\sqrt{N}r^{5/4}-2c_{Ver}K^2N^{3/4}r^{5/4}+2c_{Ver}K^2N^{1/4}r^{7/4}$$

Proof. (1) By Theorem 4.6.1 of Vershynin [33], for any $t \geq 0$, we have:

$$\sqrt{N} - c_{\text{Ver}}K^2(\sqrt{r}+t) \leq \sigma_r(A) \leq \sigma_1(A) \leq \sqrt{N} + c_{\text{Ver}}K^2(\sqrt{r}+t)$$

with probability at least $1-\exp(-t^2)$ and $K = \max_i \|A_i\|_{\psi_2}$. Then, we can choose $t = (Nr)^{1/4}$ and gets

$$\frac{\sigma_1(A)}{\sigma_r(A)} \leq \frac{\sqrt{N} + c_{\text{Ver}}K^2(\sqrt{r}+(Nr)^{1/4})}{\sqrt{N} - c_{\text{Ver}}K^2(\sqrt{r}+(Nr)^{1/4})} \leq \frac{1 + c_{\text{Ver}}K^2\sqrt{r/N} + c_{\text{Ver}}K^2(r/N)^{1/4}}{1 - c_{\text{Ver}}K^2\sqrt{r/N} - c_{\text{Ver}}K^2(r/N)^{1/4}}$$

(2) Observe that $\|A\|_F^2 = \sum_{i=1}^r \sigma_i^2(A)$. Hence, with probability at least $1-2\exp(\frac{-1}{2}\sqrt{Nr})$, we have:

$$\begin{aligned}
\|A\|_F^2 &\geq r \cdot \sigma_r^2(A) \\
&\geq r(\sqrt{N} - c_{\text{Ver}}K^2(\sqrt{r}+(Nr)^{1/4}))^2 \\
&= r(N + c_{\text{Ver}}^2K^4r + c_{\text{Ver}}^2K^4\sqrt{Nr} - 2c_{\text{Ver}}K^2\sqrt{Nr} - 2c_{\text{Ver}}K^2N^{3/4}r^{1/4} + 2c_{\text{Ver}}K^2N^{1/4}r^{3/4}) \\
&= Nr + c_{\text{Ver}}^2K^4r^2 + c_{\text{Ver}}^2K^4\sqrt{N}r^{3/2} - 2c_{\text{Ver}}K^2\sqrt{N}r^{5/4} - 2c_{\text{Ver}}K^2N^{3/4}r^{5/4} + 2c_{\text{Ver}}K^2N^{1/4}r^{7/4}
\end{aligned}$$

□

Corollary A.4. *Given a gap* ϵ *and the same assumptions as in Theorem A.2, the probability that* $|\hat{\mathbb{E}}[\bar{Y}_{i,post}^{(d)}] - \mathbb{E}[\bar{Y}_{i,post}^{(d)}]| \leq \epsilon$ *is at least* $1-\delta$*, where*

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

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

Proof. We begin by setting the right-hand side of Theorem A.2 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:

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

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

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

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

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

Let $\alpha = A+F+D(\sqrt{n}+\sqrt{d})+E\sigma_r(\mathbf{Z}_i(d))$, 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:

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

□

A.3 BIC proofs for two-type two-intervention setting

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 [33] that shows $\text{span}(\{v_i\}_{i\in\mathcal{I}^{(1)}}) = S_1$ with high probability:

Theorem A.5 (Theorem 4.6.1 of Vershynin [33]). *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 $1 - 2\exp(-t^2)$:*

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

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 $1 - 2\exp\left(-\left(\frac{\sqrt{N_0^{(1)}}}{c_{\text{Ver}}K^2} - \sqrt{r}\right)^2\right)$.

Concentration bound on number of type 1 units After seeing N_0 units in the first stage of Algorithm 1, let the number of type 1 samples observed be $N_0^{(1)}$. Using Chernoff inequality, we have:

$$\Pr[|N_0^{(1)} - p_1 \cdot N_0| \geq \epsilon p_1 \cdot N_0] \leq 2\exp(-p_1 \cdot N_0\epsilon^2/3) \tag{2}$$

Hence, given the first N_0 observations, the event that we have at least $p \cdot N_0$ samples of type 1 units happens with high probability.

Proof of Theorem 3.2

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

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

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

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

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

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

For some $C > 0$ and $\epsilon \leq C$, let $\xi_{C-\epsilon}$ be the event that $\mu_{v_i}^{(0)} \geq y_{i,post}^{(1)} + C - \epsilon$ and $\hat{\xi}_C$ be the event that $\mu_{v_i}^{(0)} \geq \hat{y}_{i,post}^{(1)} + C$. Then, we have:

$$\begin{aligned}\Pr[\neg\xi_{C-\epsilon}, \hat{\xi}] &= \Pr[\mu_{v_i}^{(0)} \leq y_{i,post}^{(1)} + C - \epsilon, \mu_{v_i}^{(0)} \geq \hat{y}_{i,post}^{(1)} + C] \\ &\leq \Pr[|y_{i,post}^{(1)} - \hat{y}_{i,post}^{(1)}| > \epsilon] \\ &\leq \delta_\epsilon^{PCR}\end{aligned}$$

Furthermore, for any $c \geq 0$, we can also write:

$$\begin{aligned}\Pr[\xi_{-\epsilon}, , \hat{\xi}] &= \Pr[\mu_{v_i}^{(0)} \geq y_{i,post}^{(1)} + C - \epsilon, \mu_{v_i}^{(0)} \geq \hat{y}_{i,post}^{(1)} + C] \\ &\geq \Pr[\mu_{v_i}^{(0)} \geq y_{i,post}^{(1)} + C + c, \mu_{v_i}^{(0)} \geq \hat{y}_{i,post}^{(1)} + C] \\ &= \Pr[\mu_{v_i}^{(0)} \geq \hat{y}_{i,post}^{(1)} + C | \mu_{v_i}^{(0)} \geq y_{i,post}^{(1)} + C + c] \Pr[\mu_{v_i}^{(0)} \geq y_{i,post}^{(1)} + C + c] \\ &= \left(1 - \Pr[\mu_{v_i}^{(0)} \leq \hat{y}_{i,post}^{(1)} + C | \mu_{v_i}^{(0)} \geq y_{i,post}^{(1)} + C + c]\right) \Pr[\mu_{v_i}^{(0)} \geq y_{i,post}^{(1)} + C + c] \\ &\geq \left(1 - \Pr[y_{i,post}^{(1)} - \hat{y}_{i,post}^{(1)} < c | \mu_{v_i}^{(0)} \geq y_{i,post}^{(1)} + C + c]\right) \Pr[\mu_{v_i}^{(0)} \geq y_{i,post}^{(1)} + C + c] \\ &\geq (1 - \delta_c^{PCR}) \Pr[\mu_{v_i}^{(0)} \geq y_{i,post}^{(1)} + C + c]\end{aligned}$$

Then, with the assumption that $\mathbb{E}[y_{i,post}^{(0)} - y_{i,post}^{(1)}] \geq -2$, we can rewrite the conditional expectation in the denominator as:

$$\begin{aligned}&\left(\mu_{v_i}^{(0)} - \mathbb{E}_{v_i}[y_{i,post}^{(1)} | \mu_{v_i}^{(0)} \geq \hat{y}_{i,post}^{(1)} + C]\right) \Pr_{v_i}[\mu_{v_i}^{(0)} \geq \hat{y}_{i,post}^{(1)} + C] \\ &= \mathbb{E}_{v_i}[y_{i,post}^{(0)} - y_{i,post}^{(1)} | \hat{\xi}_C, \xi_{C-\epsilon}] \Pr[\hat{\xi}_C, \xi_{C-\epsilon}] + \mathbb{E}_{v_i}[y_{i,post}^{(0)} - y_{i,post}^{(1)} | \hat{\xi}_C, \neg\xi_{C-\epsilon}] \Pr[\hat{\xi}_C, \neg\xi_{C-\epsilon}] \\ &\geq (C - \epsilon)(1 - \delta_c^{PCR}) \Pr[\mu_{v_i}^{(0)} \geq y_{i,post}^{(1)} + C + c] - 2 \cdot \delta_\epsilon^{PCR} \qquad (3)\end{aligned}$$

Finally, we invoke Corollary A.4 to obtain the values of δ_ϵ^{PCR} and δ_c^{PCR}.

Substituting these values into Equation (3) with $c = \epsilon = C/2$ for some $C \in (0, 1)$ and $\delta_P CR = \delta_c^{PCR} = \delta_\epsilon^{PCR}$, we have:

$$\begin{aligned}&\left(\mu_{v_i}^{(0)} - \mathbb{E}_{v_i}[y_{i,post}^{(1)} | \mu_{v_i}^{(0)} \geq \hat{y}_{i,post}^{(1)} + C]\right) \Pr_{v_i}[\mu_{v_i}^{(0)} \geq \hat{y}_{i,post}^{(1)} + C] \\ &\geq \frac{C}{2}(1 - \delta_{PCR}) \Pr\left[\mu_{v_i}^{(0)} \geq y_{i,post}^{(1)} + \frac{3C}{2}\right] - 2\delta_P CR\end{aligned}$$

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

$$L \geq 1 + \max_{i \in \mathcal{I}^{(1)}} \left\{ \frac{\mu_{v_i}^{(1)} - \mu_{v_i}^{(0)}}{\frac{C}{2}(1 - \delta_{PCR}) \Pr\left[\mu_{v_i}^{(0)} \geq y_{i,post}^{(1)} + \frac{3C}{2}\right] - 2\delta_{PCR}} \right\}$$

□

Sufficient number of samples after BIC algorithm The goal of the Algorithm 1 is to collect sufficient samples of type 1 units taking intervention 0 such that Unit Overlap Assumption is satisfied for this type-intervention pair. This condition is equivalent to having enough samples of type 1 units such that the set of latent vectors for type 1 units span the latent vector space S_0. We invoke Theorem A.5 and obtain the following expression:

$$\begin{aligned}\sqrt{n^{(1)}} &= cK^2(\sqrt{T_0} + t) \\ \iff n^{(1)} &= c^2K^4(\sqrt{T_0} + t)^2\end{aligned}$$

That is, when $N^{(1)} = c^2K^4(\sqrt{T_0} + t)^2$, Unit Overlap Assumption holds with probability at least $1 - \exp(-t^2)$.