Synthetically Controlled Bandits

Vivek F. Farias
Sloan School of Management
Massachusetts Institute of Technology
email: vivekf@mit.edu

Ciamac C. Moallemi
Graduate School of Business
Columbia University
email: ciamac@gsb.columbia.edu

Tianyi Peng
Department of Aeronautics and Astronautics
Massachusetts Institute of Technology
email: tianyi@mit.edu

Andy T. Zheng
Operations Research Center
Massachusetts Institute of Technology
email: atz@mit.edu

Initial Version: February 14, 2022
This Version: December 22, 2022

Abstract

We consider experimentation in settings where, due to interference or other concerns, experimental units are coarse. 'Region-split' experiments on online platforms, where an intervention is applied to a single region over some experimental horizon, are one example of such a setting. Synthetic control is the state-of-the-art approach to inference in such experiments. The cost of these experiments is high since the opportunity cost of a sub-optimal intervention is borne by an entire region over the length of the experiment. More seriously, correct inference requires assumptions limiting the 'non-stationarity' of test and control units that we demonstrate fail in practice. So motivated, we propose a new adaptive approach to experimentation, dubbed Synthetically Controlled Thompson Sampling (SCTS). SCTS is guaranteed to identify the optimal treatment without the attendant non-stationarity assumptions of the status quo, thereby allowing for robust inference. In addition, SCTS minimizes the cost of experimentation by incurring near-optimal, square-root regret in the experimental horizon, as opposed to linear regret for the status quo. Experiments on synthetic and real world data highlight the relative merits of SCTS in regard to both the cost of experimentation and the robustness of inference.

1 Introduction

Experimentation is a crucial tool deployed in the data-driven improvement of modern commerce platforms. On such platforms, a new product feature or algorithmic tweak is often rolled out only after its prospective benefit is understood via an appropriately designed experiment. In some cases, an appropriate unit of experimentation is simply an end user. In such cases, experiment design and inference is relatively well understood. On the other hand, it is often the case that the intervention in question induces interactions among individual users of the platform. Often referred to as 'interference', this effectively violates the Stable Unit Treatment Value Assumption (SUTVA) that is assumed in most designs, and necessary for correct inference. There is an emergent and exciting literature focused on experiment design and inference in the presence of interference.

It remains unclear how to robustly characterize the bias induced by interference. As such, a common strategy used to obviate interference concerns in practice, is simply to pick a sufficiently

coarse unit of experimentation. As a concrete example, a ride hailing platform experimenting with a new payment feature would simply choose the unit of experimentation to be a region or city and then implement the intervention in question in a test city, over some experimental horizon. This is often referred to as a 'region split' experiment. Since the pool of such coarse experimental units (i.e. regions) is by definition smaller, picking the appropriate controls is no longer a simple matter. [1]

The very challenge above arises in program evaluation, a common task in empirical economics. There, the synthetic control method is seen as 'arguably the most important innovation ... in the last 15 years' [16]. The synthetic control method seeks to construct a 'synthetic' control via a linear combination of non-treatment units that best approximates the treatment unit prior to the treatment period. While originally intended primarily for inference given observational data, this method now represents the state-of-the-art for the task of inferring treatment effects in experiments with coarse units, such as the region split experiment described above. While state-of-the-art, this overall approach suffers from two major drawbacks:

- Cost: The opportunity cost (or 'regret') resulting from experimenting with an undesirable intervention is borne at the level of a city or region (as opposed to a substantially smaller group of users) over the entire length of the experiment. Thus, in the parlance of the literature on adaptive experimentation, regret grows linearly with the experiment horizon.

- Fragility: The correctness of the control produced by the synthetic control method depends heavily on assumptions made on the so-called 'shared factors' process, a factor model that is assumed to underlie the evolution of outcomes in the treatment and control units. These assumptions can be seen to relate to how estimable this process is from data available in the pre-treatment period. We will see that these assumptions are not benign and may actually fail in real data. Moreover, we establish that absent these assumptions on the shared factors process, it is impossible to guarantee recovery of the treatment effect under *any* fixed design.

While building on the success of the synthetic control framework, we propose to address the drawbacks above via an *adaptive* approach to experiment design: the Synthetically Controlled Bandit model and an associated algorithm we dub Synthetically Controlled Thompson Sampling (SCTS). Relative to the status-quo approach described above:

- SCTS attains near-optimal, sub-linear regret.

- SCTS is guaranteed to identify the optimal treatment, even when the factor model assumptions required for the synthetic control method fail.

- When the factor model assumptions required for the synthetic control method hold, SCTS estimates the treatment effect at the same rate as that method on the event that the treatment effect is positive. On the event that the treatment effect is negative, SCTS learns that this is the case.

[1] Parenthetically, it is worth noting that the counterfactual value of the outcome being measured for any unit may have slow-mixing, or even non-stationary, temporal effects so that so-called 'switch-back' designs are insufficient.

As such, SCTS eliminates the cost and fragility of the status quo approach to experimentation with coarse units. The price paid by this design is that it can quantify the treatment effect only when the effect is positive while simply learning that the treatment effect is negative when that is the case. Since in practical applications, a quantification of the treatment effect is only of value when this treatment effect is positive (so as to facilitate, for instance, cost-benefit analyses for a roll-out of the intervention), this new approach makes possible an attractive tradeoff.

1.1 The Synthetically Controlled Bandit

We now proceed with describing our contributions in greater detail, focusing in turn on minimizing regret, inferring the treatment effect, and practical considerations.

The Bandit Model and Minimizing the Cost of Exploration: Consider a setting where the decision whether or not to treat the treated unit (city, region, etc.) in any given epoch is a dynamic one. Over some experimentation horizon, a natural goal aligned with minimizing the cost of experimentation, would be to minimize *regret*. That is, over the experimentation horizon, we effectively minimize the expected number of times a sub-optimal treatment option was chosen for the treated unit. It turns out that in the synthetic control setting, this problem is equivalent to a linear contextual bandit, wherein the context at each period is an *unobserved* low-dimensional latent vector (the so-called 'unobserved common factors' in the corresponding synthetic control model). This bandit problem has several salient features that render the application of existing bandit algorithms challenging:

1. Since observations across all units are made contemporaneously, no information about the context vector is available at the time of decision making.
2. We never observe historical context vectors directly either; instead we only ever observe an unknown, noisy, linear transformation of these vectors.
3. This unknown linear transformation can never be recovered exactly even with infinite data: instead, we can only hope to recover it up to an orthogonal transformation.

Put succinctly, the underlying contextual bandit is one where no information regarding the context is available at the time of decision making, and can, post-facto, only be recovered up to an unknown orthogonal transformation and noise. Algorithms for stochastic linear bandits (such as LinUCB or Linear Thompson Sampling) simply do not apply to the setting; these algorithms fundamentally require that the contexts be observed. Adversarial bandit algorithms (such as Exp3) may be applied in this setting, but by virtue of being robust to all possible contexts, these algorithms will essentially disregard the data from the control units. This results in substantially larger regret; in particular, the leading regret constants are arbitrarily larger than what we eventually show possible in this model. As an aside it is worth noting that the synthetically controlled bandit model is interesting beyond the realm of experiments, to settings where the observed outcomes of our actions are confounded by unobserved latent factors.

Beyond proposing this bandit model that extends the synthetic control framework, our primary technical contribution is an algorithm that, despite the challenges above, achieves a regret that

scales like $r\sigma\sqrt{T}$. Here r is the dimension of the latent context, σ is the standard deviation of exogenous additive noise in the test unit observations, and T is the experimentation horizon. We dub our approach *Synthetically Controlled Thompson Sampling* (SCTS). SCTS consists of a Thompson sampling routine with carefully designed 'exploration noise'. Contexts are recovered via principal components analysis (PCA) on historical observations. Importantly, our sampler is robust to the errors in context recovery due to noise and the inability to recover rotations.

Let us put this result in context. Whereas there is no stochastic bandit algorithm to readily benchmark against, it is worth noting that if one were allowed to observe contexts without noise in the period immediately following a decision (an information structure substantially more generous than what we have), Linear Thompson sampling would achieve regret that scales like $r^{3/2}\sigma\sqrt{T}$, so that despite having less information SCTS achieves superior regret; a result of independent interest. An adversarial algorithm such as Exp3 would yield regret that scales like $B\sqrt{T}$ where B is an upper bound on the magnitude of observed outcomes. In contrast, the leading constant in our regret term, $r\sigma$ is independent of the magnitude of observed outcomes, and in fact goes to zero with exogenous noise – this reflects the power of using the control units to control for confounding.

Inference: Turning to inference, we first focus on simply the task of identifying the optimal treatment. There we show that for *any* fixed design, there exists a synthetic control model (i.e. an outcome process for the experimental and control units satisfying the synthetic control model) for which no algorithm can identify the optimal treatment irrespective of horizon. The situation highlighted by this negative result is effectively assumed away in the traditional use of the synthetic control method via assumptions that limit the 'non-stationarity' of the latent shared factors process. We demonstrate through a real world dataset that such assumptions are not benign and that the basic inferential task of identifying the optimal treatment with a fixed design is challenging in real data. In contrast to this situation, we show that SCTS will identify the optimal arm with high probability in the length of the experimental horizon, an observation also reflected in our experimental results.

In settings favorable to fixed designs – i.e., those where non-stationarity in shared factors is limited – SCTS admits estimation of the treatment effect at rates that line up precisely with those achieved by the synthetic control method in the fixed design setting. More precisely, we provide the same quality of inference as in a traditional fixed design experiment with synthetic control inference on the event the treatment effect is positive. When the effect is negative, we are only able to detect that this is the case. Quantification of the treatment effect is typically only relevant when the effect is positive, so as to facilitate, for instance, cost-benefit analyses for a roll-out of the intervention. As such, it would appear that the inability to precisely estimate a negative treatment effect (beyond identifying that it is negative) is a small price to pay for the mitigation of cost and fragility afforded by the approach.

Finally, with an eye to practice, we propose the use of re-randomization based hypothesis tests and confidence intervals derived from inverting these tests. We see, on real world data, that these tests are highly powered even for small treatment effects and that the confidence intervals derived

from them provide near-ideal coverage.

Computational Experience: We present experimental work on both a synthetic data setup (wherein the synthetic control model holds by construction), as well as on real world data, where this setup is, at best, an approximation. In both cases, we see that SCTS employs a sub-optimal intervention for a negligible fraction (typically a single digit percentage) of epochs over the experimentation horizon. Compared with both a fixed and switchback design, SCTS thus materially reduces the cost of exploration. Despite this, we see that our treatment effect estimator correctly identifies whether or not the treatment effect is positive in every single one of our instances. Importantly, on the instances where the treatment effect is positive, the relative RMSE of our estimator is comparable to state-of-the-art estimators for both the switchback and fixed designs, and in fact outperforms these incumbents in the real-data setting. As mentioned previously, re-randomization based hypothesis tests and confidence intervals provide near-ideal coverage, even on real-world data, allowing for effective inference.

1.2 Related Literature

Synthetic Control and Inference The notion of synthetic control was introduced initially in the context of program evaluation: [4, 2] are seminal papers that propose to recover the counterfactual in an observational setting by creating a "synthetic control". Specifically, they proposed constructing a convex combination of control units that matches the treated unit in pre-treatment periods. A series of follow-on studies proposed distinct estimators by employing different constraints and regularizers (e.g, [44, 38, 50, 14, 21]). See [1] for a review of this vibrant literature. Since the underlying generative model justifying the synthetic control framework is in fact a factor model, it is natural to consider using PCA-like techniques in the recovery of a synthetic control; the present work leverages such techniques in a dynamic context. This approach is especially relevant in the setting where the size of the 'donor pool' is large. [15, 63, 12, 20, 11, 9, 40] are all papers in this vein. [40] in particular compute a min-max optimal estimator for a generalization of the synthetic control problem. It is indeed possible to compute limiting distributions for synthetic control estimators in various special cases. For instance, if one were willing to make probabilistic assumptions on the data, it is possible to compute a limiting distribution for the synthetic control estimator (roughly, this distribution is a projection of the OLS limiting distribution to a convex set); [49]. In practice, however, inference for average treatment effects in synthetic control is done via non-parametric methods such as permutation tests; see [33].

Synthetic Control in Experiment Design for Commerce Not surprisingly, synthetic control approaches have gained traction in modern commerce settings; as a relatively early example [27] describes an approach and corresponding software used by Google in the context of marketing attribution. Going further, however, synthetic control has come to be viewed as an important tool in experiment design as well, as opposed to simply in observational settings. For instance, [32, 46] describe practical designs, at Lyft and Uber respectively, that assume a synthetic control model holds across units. In a theoretical direction, [37] and [5] consider the problem of how best to

select an experimental unit assuming the synthetic control model holds, motivated by problems at Facebook and the 'region-split' experiments common to ride-sharing platforms respectively. Like this work, the present paper also actively uses the synthetic control model in experiment design; in our case these designs are 'dynamic'.

Contextual Bandits and Inference The dynamic design that this paper constructs is, in a certain idealized sense, a linear contextual bandit. This sort of bandit is classical, studied at least as early as [17]. The practical constraints around our design necessitate a sampling methodology that draws from recent work on Thompson sampling for such bandits; see [10, 8, 47]. As noted elsewhere, the fundamental challenge we must address is that we never observe contexts directly. There is some limited work discussing the use of dynamic bandit based designs in clinical trials, [62, 22].

Turning to inference, it is well known that naive sample estimates of arm means in bandits are biased (see e.g. [62, 54]). A recent line of work considers 'post-contextual bandit' inference, where certain importance-weighted estimators are shown to be unbiased and asymptotically normal; see [42, 24] and also [36] for a different approach to the problem under a linear reward model. Extending these to our setting is an exciting direction for future work. In addition to the observability of the contexts, such an extension will also need to address the general problem that this line of work requires a type of forced exploration of arms which may not be consistent with our bandit algorithm. The present paper simply constructs high probability confidence intervals via the usual self-normalized martingale concentration bounds. While loose in practice, they already illustrate that we can expect rates that are essentially on par with what is possible for the vanilla synthetic control estimator. In our experimental work, we complement these with bootstrapped confidence intervals and permutation tests for significance that we show work adequately.

Non-stationary bandits A number of existing approaches address the problem of adaptive experiment design with non-stationary outcomes. Foremost among these is the adversarial bandit setting [18], which typically allows for bounded but otherwise arbitrary non-stationarity in outcomes for each arm, and attempts to compete with the best arm on average in hindsight. Related to this literature, [23] analyzes a setting where non-stationarity in outcomes is constrained by a fixed budget, and characterizes regret relative to the best arm at each time. However, the regret guarantees possible in such a setting scale poorly with T, relative to the usual adversarial bandit setting. Compared to these settings, we propose a model which is realistic but significantly more structured, and obtain correspondingly tighter guarantees; see for example the discussion in Section 3.1.

Concurrently to our work, [58] introduces a Thompson sampling variant in a setting in which outcomes are linear functions of contexts, and the goal is to identify the best arm for some fixed context known a priori. Similarly to SCTS, their approach does not require knowledge of the context at time t prior to choosing the action a_t. Their setting differs from ours in several important respects: they provide guarantees for the problem of best-arm identification (i.e., they do not present a regret-minimizing algorithm; in fact their setting has no notion of regret); they assume that contexts are fully observable ex-post, whereas in our setting contexts are latent and only recoverable up to rotation; and they address the Bayesian setting, whereas our results are fully frequentist.

The remainder of this paper is organized as follows:

1. Section 2 formalizes our setting and outlines our key contributions.
2. Section 3 introduces the SCTS algorithm and formalizes our main regret bound.
3. Section 4 presents a proof for the regret bound, which relies on a novel recovery bound for the latent factors which is independent of condition number.
4. Section 5 presents guarantees regarding the recovery of the treatment effect and optimal action, including comparisons against those available under fixed-design synthetic control.
5. Section 6 evaluates SCTS' performance on both synthetic and real-world data, relative to existing alternatives.

2 Model and Results

We measure some quantity of interest for an experimental unit (e.g., a region, in a region-split experiment) over a pre-treatment period of length T_0, and a subsequent treatment period of length T. We denote the measurement made on this experimental unit at any epoch t by $y_t^0 \in \mathbb{R}$. We assume that the pre-treatment period consists of epochs in $\{-T_0+1, \ldots, 0\}$ and that the treatment period consists of epochs in $\{1, 2, \ldots, T\}$. We denote by $a_t \in \{0, 1\}$ the indicator of whether or not the experimental unit is treated at time t, so that $a_t = 0$ for all epochs in the pre-treatment period. We assume that y_t^0 is determined by the structural equation

$$y_t^0 = \tau^* a_t + \langle \lambda^*, \bar{z}_t \rangle + \epsilon_t^0, \tag{1}$$

where $\tau^* \in \mathbb{R}$ is an unknown treatment effect, $\bar{z}_t \in \mathbb{R}^r$ is an (unknown) set of r 'shared common factors' and $\lambda^* \in \mathbb{R}^r$ is a set of (unknown) 'factor loadings' specific to the experimental unit. The noise ϵ_t^0 is assumed to be independent Gaussian with mean zero, and standard deviation σ. The synthetic control paradigm assumes a generative setting where a weighted combination of observations in the pool of donor units closely approximates counterfactual observations on the experimental unit. Specifically we assume a pool of $n \geq r$ donor units, where for the ith such unit

$$y_t^i = \langle \lambda^i, \bar{z}_t \rangle + \epsilon_t^i. \tag{2}$$

Here $\bar{z}_t$ is the same set of shared common factors and $\lambda^i \in \mathbb{R}^r$ is a set of factor loadings specific to the ith donor unit. As before, $\epsilon_t^i \sim \mathcal{N}(0, \sigma^2)$.

The factors $\bar{z}_t$ are assumed to be bounded and adapted to history (precisely, $\{\bar{z}_t\}$ is adapted to the σ-algebra generated by $(a_s, \bar{z}_s, \epsilon_s^i)_{i \in [n], s < t}$), but may otherwise be *arbitrary or even adversarially chosen*; this lack of assumptions on the factors gives the synthetic control model a great deal flexibility.

2.1 Fixed Design and Estimation

For context, it will be useful to review the status-quo approach to experimental design and estimation in this setting, due to [4]. This approach employs a *fixed* experimental design, which

simply sets $a_t = 1$ over the treatment period, (i.e. for $t \in \{1, 2, \ldots, T\}$).

An estimate for τ^* is then computed via the "vanilla"synthetic control method τ^{SC}, as in [2, 3]. [2] gives the following guarantee for the vanilla synthetic control estimator:

Proposition 1. *With probability* $1 - O(\delta)$*,*

$$|\tau^{\mathrm{SC}} - \tau^*| \lesssim \frac{\sigma}{\sqrt{T}}\sqrt{\log(1/\delta)} + \frac{c_2\sigma}{\sqrt{c_1 T_0}}\sqrt{\log(1/\delta) + \log(n)}.$$

where $c_1 = \sigma_r(1/T_0 \sum_{t=-T_0+1,\ldots,0} \bar{z}_t \bar{z}_t^\top)$ and $c_2 \triangleq \max_t \|z_t\|^2$. In particular, c_1 is loosely a metric of non-stationarity in the latent factor process $\{\bar{z}_t\}$: if, for example, some dimension of $\bar{z}_t$ is not present during the pre-treatment period, then c_1 can be arbitrarily small.

As a result, fixed-design synthetic control can utterly fail to learn the treatment effect when c_1 is not bounded away from zero, and is therefore not "safe" to use without strong assumptions on the latent factor process. Existing literature (e.g., [2]) typically simply assumes that c_1 is a constant, but we will see in our experiments (Section 6) that this is often not a benign assumption in the real world.

2.2 Dynamic Design and Estimator

As a solution to these two key weaknesses of fixed-design synthetic control – linear regret and sensitivity to the latent factor process $\{\bar{z}_t\}$ – we will propose a design which allows a_t to be *dynamic*. Specifically, we allow a_t to be selected according to a randomized policy that is adapted to $\mathcal{F}_t = \sigma((a_s, y_s^i)_{i\in[n],s<t})$, i.e. the filtration generated by observations up to and including time $t-1$.

Define the cost of experimentation incurred at time t by the 'regret' incurred in that epoch,

$$R_t \triangleq |\tau^*| \cdot \big(\mathbf{1}\{\tau^* < 0\}\, a_t + \mathbf{1}\{\tau^* \geq 0\}\,(1 - a_t)\big).$$

This definition captures both the negative impact of a sub-optimal treatment, and the opportunity cost of not using the treatment should it be optimal. The total cost incurred by the fixed synthetic control design may then scale linearly with the length of the treatment period, T.

The total expected cost (or *regret*) of experimentation, $R(T)$, is then simply $\mathbb{E}\left[\sum_{t=1}^T R_t\right]$, where the expectation is over the noise in observations and randomization in the design.

We will sometimes refer to this setup as the 'Synthetically Controlled Bandit'.

Finally, an estimator is simply an $\mathcal{F}_{T+1}$ measurable random variable, that ideally provides a good approximation to an estimand of interest: either the treatment effect τ^*, or the optimal action a^*. With this setup, we now state the problems we wish to address:

- First, we would like to produce a dynamic design, i.e., a process $\{a_t\}$, that minimizes the total cost of experimentation $R(T)$.
- Second, on the inferential side, we would like to identify the optimal action, a^*, with high probability – without requiring assumptions on the latent factor process such as those in

[2]We write $A \lesssim B$ if $A \leq cB$ for some absolute constant c.

Proposition 1. In addition, we wish to design an estimator of the treatment effect τ^*, $\hat{\tau}$, for which $|\hat{\tau} - \tau^*|$ is 'small' with high probability on the event where $a^* = 1$ (i.e. $\tau^* > 0$).

2.3 Results

In what follows, we describe our main results, making precise the regret achieved by our proposed dynamic design, and the associated inferential results.

SCTS achieves $\sqrt{T}$ regret: We propose a dynamic design, dubbed *Synthetically Controlled Thompson Sampling* (SCTS), which we show achieves near-optimal experimentation cost:

Theorem 1 (Informal). *Assume that the number of donor units $n = \Omega(T)$. Then, under mild assumptions on the shared common factors, we have that SCTS incurs a cost of experimentation, $R(T) = O\left(r\sqrt{T}\log(T)\right)$.*

In a nutshell, SCTS essentially eliminates the cost of experimentation, which as we noted earlier will scale linearly with T for the fixed design. It is also worth placing the regret guarantee in context. To that end, note that we have no information pertaining to $\bar{z}_t$ (which is essentially arbitrary in the synthetic control model) at the time we decide on a_t. Imagine for a moment, however, that at time t we observed $\bar{z}_s$ for all $s < t$. Of course, this is *substantially* more information than we have, but the problem at hand reduces to a two-armed linear contextual bandit amenable to Linear Thompson sampling [3]. Linear Thompson sampling applied to this setup will achieve $\tilde{O}(r^{3/2}\sqrt{T})$ regret [8].[4]

In our setting, even the history of the shared common factors (i.e. $\bar{z}_s$ for all $s < t$) is not available; rather these must be inferred from our observations over the donor units. There we will see that even with noiseless observations of y_s^i on the donor units we would only succeed in recovering the common factors $\bar{z}_s$ up to a rotation. In light of these salient problem features it is notable that our regret guarantee depends linearly on r, where r is typically much smaller than the ambient number of donor units, n. This guarantee is our main theoretical result.

SCTS estimates the treatment effect as efficiently as vanilla SC: We now turn to inference.

There we know that, under the usual fixed design, the nominal [2, 3] synthetic control estimator τ^{SC} achieves with probability $1 - O(\delta)$,

$$|\tau^{\mathrm{SC}} - \tau^*| \lesssim \frac{\sigma}{\sqrt{T}}\sqrt{\log(1/\delta)} + \frac{c_2\sigma}{c_1\sqrt{T_0}}\sqrt{\log(1/\delta) + \log n}.$$

The regularization implicit in this estimator achieves a rate that is largely independent of n.[5] Loosely, the constant c_1 measures non-stationarity in the latent factor process $\{\bar{z}_t\}$. The constant c_2 depends on the size of the shared common factors. Asymptotic distributions for this estimator without further distributional assumptions on the common factor process are unknown.

[3] this information relaxation continues to remain unamenable to LinUCB however, since the context at time t is unavailable at the time of selecting a_t.

[4] The $\tilde{O}(\cdot)$ notation suppresses dependence on logarithmic factors, $\log(T)$ and $\log(r)$.

[5] We write $A \lesssim B$ if $A \leq cB$ for some absolute constant c.

Our own estimator, τ^{SCTS} achieves similar rates: When $\tau^* > 0$, with probability $1 - O(\delta) - \tilde{O}(1/\sqrt{T})$,

$$|\tau^{\mathrm{SCTS}} - \tau^*| \lesssim \frac{\sigma}{\sqrt{T}}\sqrt{\log(1/\delta)} + \frac{c_2\sigma}{c_1\sqrt{T_0}}\sqrt{\log(1/\delta) + \log n}.$$

On the other hand, when $\tau^* \leq 0$, with probability $1 - \tilde{O}(1/\sqrt{T})$, $\tau^{\mathrm{SCTS}} = 0$ (i.e. we correctly identify that the treatment effect is non-positive).

Contrasting this with the high probability confidence intervals for τ^{SC}, we see that on the event that $\tau^* > 0$, we get the same intervals as the synthetic control estimator. On the event that $\tau^* < 0$, we correctly learn that the treatment effect is non-positive. The result follows from a simple idea expanded on in Section 5. In our computational experiments, we see that re-randomization tests for p-values and the corresponding inverted hypothesis tests [41] for confidence intervals provide adequate power and coverage.

SCTS robustly identifies the optimal treatment: The above guarantees are meaningful in the case where c_1 is large. However, when the shared common factors process $\{\bar{z}_t\}$ is arbitrary, c_1 can in-fact be arbitrarily small. As a result, not only are the guarantees for the synthetic control estimator, τ^{SC} vacuous, but in addition, we are able to show that for *any* fixed design, and *any* estimator $\hat{a}$ of the optimal treatment a^*, there exists a problem instance for which $\mathbb{P}(\hat{a} \neq a^*) = \Theta(1)$ (where expectations are over the randomness in the instance). This is a striking negative result. In fact, we also see that the result is not pathological: a real-world dataset we study exhibits a common factor process for which the constant c_1 is apparently small.

In contrast to this fragility, we show that under the SCTS design, it is simple to construct an estimator of the optimal treatment, $\hat{a}$, for which $\mathbb{P}(\hat{a} = a^*) = 1 - O(\frac{1}{T})$; in other words, SCTS recovers the optimal treatment correctly with probability approaching 1 as T grows large – even when the contexts $\{\bar{z}_t\}$ are chosen by an $\mathcal{F}_t$-adapted adversary. As we see in our experiments, this robustness enables meaningfully better estimation of the treatment effect in real world datasets.

In their totality, these results show that we can largely eliminate the key challenges with typical approaches to experiment design with synthetic controls: the cost of experimentation, and fragility to the latent factor process. We achieve this while paying only a modest cost to inference, in scenarios favorable to fixed-design synthetic controls: when the treatment effect is negative we only learn that this is the case with high probability, as opposed to getting a precise estimate of the effect. Since in practical settings a precise estimate of the treatment effect is typically only needed when the treatment effect is positive (so as to ascertain whether the cost of implementing the intervention is justified), this is perhaps a modest price to pay.

3 Synthetically Controlled Thompson Sampling

This section introduces Synthetically Controlled Thompson Sampling (SCTS), which adapts Thompson Sampling (TS) to the Synthetically Controlled Bandit model introduced in the previous section. The algorithm is conceptually simple: at the start of each epoch, $t+1$, we compute a distribution $\mathcal{D}_t^{\mathrm{TS}}$ over 'plausible' values of τ^*. This distribution may be thought of informally as an

approximation to a posterior over τ^* under a non-informative prior, given the information available up to and including time t. We then sample from this distribution, and pick $a_{t+1} = 1$ if and only if the sampled value, $\tilde{\tau}_t$ is non-negative. To construct $\mathcal{D}_t^{\text{TS}}$, we

1. First, estimate via PCA the (unobserved) shared common factors $\bar{z}_s$ for $s \leq t$.
2. Plugging in the estimates obtained for $\bar{z}_s$ above into the structural equation (1), we compute an estimate of τ^* and λ^* via ridge regression.
3. We use the estimates of τ^* and the precision matrix obtained from the regression in the previous step to construct our approximation to the posterior on τ^*, $\mathcal{D}_t^{\text{TS}}$.

Next, we make precise each of these steps, assuming, simply for notational convenience, that $T_0 = 0$.

Estimating Shared Common Factors Recall from (2), that for each donor unit i and epoch s, we observe $y_s^i = \langle \lambda^i, \bar{z}_s \rangle + \epsilon_s^i$. Define by $Y_t \in \mathbb{R}^{n \times t}$ the matrix with (i, s) entry y_s^i, and similarly, define by $E_t \in \mathbb{R}^{n \times t}$ the noise matrix with (i, s) entry ϵ_s^i. Now, let $\Lambda \in \mathbb{R}^{n \times r}$ be the factor loadings matrix with ith row $\lambda^{i\top}$, and denote by $\bar{Z}_t \in \mathbb{R}^{t \times r}$ the common factors matrix with sth row $\bar{z}_s^\top$. By (2), we then observe at time $t+1$ the outcomes $Y_t = \Lambda \bar{Z}_t^\top + E_t$. We estimate $\bar{Z}_t$ at time $t+1$ by solving

$$\min_{Z \in \mathbb{R}^{t \times r},\ \Lambda \in \mathbb{R}^{n \times r}} \left\| Y_t - \Lambda Z^\top \right\|^2. \tag{3}$$

We fix a specific solution to the above optimization problem via PCA. Specifically, let $Y_t = \hat{U}_t \hat{\Sigma}_t \hat{V}_t^\top$ be any singular value decomposition (SVD) of Y_t. Denote by $\hat{U}_t^r$ and $\hat{V}_t^r$ the matrices obtained from the first r columns of $\hat{U}_t$ and $\hat{V}_t$ respectively. Finally, let $\hat{\Sigma}_t^r$ be the sub-matrix obtained from $\hat{\Sigma}_t$ from its first r rows and columns. By the Young-Eckart theorem, an optimal solution to (3), $(\hat{\Lambda}_t, \hat{Z}_t)$, can be obtained by setting $\hat{\Lambda}_t \triangleq \sqrt{n}\hat{U}_t^r$, and

$$\hat{Z}_t \triangleq \frac{1}{\sqrt{n}} \hat{V}_t^r \hat{\Sigma}_t^r.$$

We recognize $\hat{Z}_t$ as precisely the usual 'PCA loadings'; $\hat{Z}_t$ will serve as our approximation to $\bar{Z}_t$.

Ridge Regression Recall that in our synthetic control model, we have for the treatment unit, $y_s^0 = \tau^* a_s + \langle \lambda^*, \bar{z}_s \rangle + \epsilon_s^0$ at each epoch s. At time $t+1$, we employ this structural equation to estimate τ^* via least squares, using as a plug-in estimator[6] for $\bar{z}_s^\top$, $\hat{z}_{s,t}^\top$, the sth row of $\hat{Z}_t$. Our estimate of τ^* at time $t+1$, $\hat{\tau}_t$, is obtained as the solution to the regularized least squares problem:

$$\min_{\tau \in \mathbb{R},\ \lambda \in \mathbb{R}^r} \sum_{s \leq t} \left(y_s^0 - \tau a_s - \langle \lambda, \hat{z}_{s,t} \rangle \right)^2 + \rho(\tau^2 + \|\lambda\|_2^2). \tag{4}$$

Here, $\rho > 0$ is a regularization penalty.[7]

[6] While the subscript t in $\hat{z}_{s,t}$ makes precise that this is our estimate of $\bar{z}_s$ at time t, we will sometimes drop the t subscript when clear from context.

[7] We fix $\rho \triangleq 1$ throughout the paper.

We find it convenient to define the 'precision matrix' $\Omega_t \in \mathbb{R}^{(r+1)\times(r+1)}$ of the estimator $\hat{\theta}_t^\top \triangleq [\hat{\tau}_t \ \hat{\lambda}_t^\top]$. Specifically, if we denote $x_{s,t}^\top \triangleq [a_s \ \hat{z}_{s,t}^\top]$, then $\Omega_t \triangleq \rho I + \sum_{s\le t} x_{s,t} x_{s,t}^\top$. The 'variance'[8] of our estimator $\hat{\tau}_t$ is simply $(\Omega_t)^{-1}_{1,1} \triangleq \hat{\sigma}_t^2$.

Approximate Posterior For our approximation to the posterior on τ^* at time $t+1$, we take $\mathcal{D}_t^{\mathrm{TS}}$ to be the uniform distribution, $\mathrm{Unif}[\hat{\tau}_t - \beta_t\hat{\sigma}_t, \hat{\tau}_t + \beta_t\hat{\sigma}_t]$. Here $\beta_t > 0$ is a time-dependent 'expansion' factor we make precise later; for now we may simply consider β_t to be an increasing sequence with $\beta_t = O(\sqrt{r\log(rt)})$. As discussed earlier, SCTS draws a sample, $\tilde{\tau}_t$ from $\mathcal{D}_t^{\mathrm{TS}}$ at time $t+1$. Then, SCTS sets $a_{t+1} = 1$ if and only if $\tilde{\tau}_t \ge 0$.

Remark 1 (Inconsistent Designs)**.** *Notice that the design employed in the regression* (4) *is* inconsistent *from period to period in the sense that the estimate for any fixed context $\bar{z}_s$ in the design matrix changes from period to period. Part of this is simply due to noise – as time goes on we hope to compute a more accurate estimate of $\bar{z}_s$ for any fixed s. However, as it turns out even in the absence of noise (i.e., if E_t were identically zero), we would still not expect consistency in the design since even in that case, we would only ever be able to recover the contexts up to a rotation. A priori it is unclear whether this inconsistency will allow for effective recovery of the treatment effect, and as such it is unclear whether we can expect the algorithm we have described to achieve low regret.*

Prior to stating out main result for SCTS, it is worth considering the relative merits of natural alternatives.

3.1 Alternative Approaches

We consider in turn three alternatives to SCTS. The first is a natural UCB variant to SCTS that points to the necessity of randomization in SCTS. Subsequently we consider LinUCB and Linear Thompson Sampling (which we see do not naturally extend) and Exp3 (which we see incurs large regret).

3.1.1 The Failure of UCB

A natural upper confidence bound (UCB) style alternative to the algorithm we have described, might proceed by defining the upper confidence bound $\mathrm{UCB}_t \triangleq \hat{\tau}_t + \beta_t\hat{\sigma}_t$, and then setting $a_{t+1} = 1$ if and only if $\mathrm{UCB}_t \ge 0$. Perhaps surprisingly, this algorithm would incur *linear* regret in general; see Proposition 2 below. We thus see that *the randomization in SCTS plays a crucial role in achieving sub-linear regret.*

We will provide a class of simple and decidedly non-pathological instances wherein the above UCB alternative to SCTS incurs linear regret. To begin, consider the case for $r = 1$ and $\sigma = 0$ (i.e. the noiseless scenario). Suppose the treatment effect is negative: $\tau^* < 0$. We also assume $\lambda^* = 1 - \tau^*$ and $\lambda^i = 1$ for $i \in [n]$. Further, let $\bar{z}_t = 1$. Then the outcome model implies that $y_t^i = 1$ for all $i \in [n]$, and $y_t^0 = 1 - \tau^* + a_t\tau^*$.

[8]While for expositional purposes we use the terminology 'precision matrix' and 'variance', these quantities are of course not a precision matrix or variance since the design of the regression problem is not fixed.

Recall that UCB chooses actions by the following procedure: (i) compute $\hat{z}_t$ by estimating the common factors through SVD; (ii) solve the ridge regression problem to obtain $\hat{\tau}_t$ and its 'variance' estimate $\hat{\sigma}_t$; (iii) play $a_t = 1$ if $\hat{\tau}_t + \beta_t \hat{\sigma}_t \geq 0$, and $a_t = 0$ otherwise. Next, we will show that this algorithm will constantly choose $a_t = 1$, for any ridge regularizer ρ and any sequence of β_t, thereby incurring $O(T|\tau^*|)$ regret.

Estimating shared common factors. By the SVD of Y_t, one has $Y_t = \sqrt{nt}\hat{U}^\top \hat{V}$ with $\hat{U}_i = \sqrt{1/n}$ and $\hat{V}_s = \sqrt{1/t}$. Then the estimator $\hat{z}_t \triangleq \sqrt{1/n}\sqrt{nt}\hat{V}_s = 1$, i.e., in the noiseless setting, $\hat{z}_t = \bar{z}_t = 1$.

Ridge regression. Recall that we will solve the following (regularized) least squares problem at time step $t+1$.

$$\min_{\tau \in \mathbb{R}, \lambda \in \mathbb{R}^r} \sum_{s \leq t} \left(y_s^0 - \tau a_s - \langle \lambda, \hat{z}_s \rangle\right)^2 + \rho(\tau^2 + \|\lambda\|_2^2)$$

Under the constructed setting and the assumption that $a_s = 1$ for $s \leq t$, the problem is equivalent to

$$\min_{\tau \in \mathbb{R}, \lambda \in \mathbb{R}^r} \sum_{s \leq t} (1 - \tau - \lambda)^2 + \rho(\tau^2 + \lambda^2)$$

which has the closed-form solution

$$\hat{\tau}_t = \hat{\lambda}_t = \frac{t}{\rho + 2t}.$$

Action decision. Since $\hat{\tau}_t > 0$, we have $\hat{\tau}_t + \beta_t \hat{\sigma}_t \geq 0$ and hence $a_{t+1} = 1$. This implies the UCB algorithm will play $a_t = 1$ for all t, thereby incurring regret $T|\tau^*|$.

It is remarkable that this example would apply identically to the situation where the common factors $\bar{z}_s$ were observed exactly. The above discussion thus establishes:

Proposition 2. *There exists a class of synthetically controlled bandit problem instances with $r = 1$ such that the UCB algorithm has linear regret, $T|\tau^*|$.*

3.1.2 Inapplicability of contextual bandit algorithms

While the outcome model (1) resembles that of a linear contextual bandit, it is worth emphasizing that *no information* about the 'context' $\bar{z}_t$ is available to the experimenter when choosing the treatment a_t. As discussed earlier, LinUCB requires $\bar{z}_t$ to be observed prior to choosing an action a_t. This information structure can be relaxed: Linear Thomson Sampling requires only that $\bar{z}_s$ for all $s < t$ be observed prior to choosing an action a_t. In our setting, while it is obvious that the contexts $\bar{z}_s$ are never observed, it is worth noting that they can never be recovered either, even as t grows

large. Specifically, as noted earlier, this is due to the fact that (3) allows identification of contexts only up to a rotation even in the absence of noise. SCTS addresses these issues by choosing actions which are simultaneously (1) *invariant* to rotations of context, (2) robust to the 'inconsistency' in designs over time steps and (3) nonetheless enjoy low single-step regret.

3.1.3 Adversarial bandits are overly conservative

'Adversarial' bandit algorithms such as Exp3 [18] are an approach to low-regret experimentation when counter-factual outcomes vary adversarially; these algorithms are guaranteed to incur low regret with respect to the best single action in hindsight. In our setting, the Exp3 algorithm could be applied to the experimental unit, and doing so would obtain $\tilde{O}(\sqrt{T})$ regret. In doing so, we would be entirely discounting all of the information in the control units, which otherwise would intuitively help control for confounding. The price paid for this is that the regret achieved by Exp3 depends linearly on the magnitude of the rewards; here, roughly $B + \sigma$ where B is an upper bound on $\|\bar{z}_t\|_2$. This is problematic since we expect B to be large relative to the treatment effect τ^* in practice; as noted in [55], 'the outside world often has a much larger effect on metrics than product changes do'. In contrast, by virtue of using information from the control units to control for confounding, we will see that the regret of SCTS depends only logarithmically on B, and as the noise level σ goes to zero, the regret of SCTS goes to a constant.

3.2 SCTS Achieves Near-Optimal Regret

We now state a formal regret bound for SCTS (i.e. a formal restatement of Theorem 1). In order to do so, we must first state our assumptions, which concern the expected value of the observation matrix on the donor units, i.e., $\mathbb{E}Y_T \triangleq \bar{Y}_T$, a rank r matrix. Specifically, we make assumptions on the decomposition $\bar{Y}_T = \Lambda \bar{Z}_T^\top$. To do so, we first note that in our model, it is possible to assume, without loss, a canonical version of this decomposition (note that the selection of Λ and $\bar{Z}_T$ is not unique due to the free choice of a rotation). In particular, letting $\bar{Y}_T = \bar{U}\bar{\Sigma}\bar{V}^\top$ be an SVD of $\bar{Y}_T$, we may assume without loss, that $\Lambda = \sqrt{n}\bar{U}$ and $\bar{Z}_T = \bar{V}\bar{\Sigma}/\sqrt{n}$; see Appendix A for details. Given this canonical decomposition, we assume:

Assumption 1. *For all t, $\|\bar{z}_t\|_2$ is upper bounded by a constant, B, and $\|\lambda^*\|_2 = O(\sqrt{r})$.*

This assumption controls the scale of the mean rewards $\mathbf{E}[y_t^i | a_t, z_t] = \tau^* a_t + \langle \lambda^*, \bar{z}_t \rangle$, in particular guaranteeing that they are bounded. The assumption on $\|\lambda^*\|_2$ is satisfied when the mean rewards $\mathbb{E}[y_t^i | a_t, z_t]$ have magnitude constant in n, t (i.e., scaling n actually yields more information about the latent factors $\bar{z}_t$), and is implied by (and is substantially weaker than) the incoherence assumption typically made in the matrix completion and panel data literature [28, 51, 7, 31, 19, 52, 20, 40].

We can now state our main regret bound for SCTS.

Theorem 1 (SCTS Regret). *Let $n = \Omega(T)$. Under Assumption 1, SCTS achieves expected regret $R(T) = O(r\sqrt{T}\log(T))$.*

The $O(\cdot)$ notation in the above regret bound ignores terms that depend polynomially on σ and logarithmically on B. As stated earlier, the linear dependence on r above is of note. We also observe that beyond its rank, our guarantee remarkably has no further dependence on the spectrum of $\bar{Y}_T$ (in particular, this matrix could be arbitrarily badly conditioned). This is encouraging since existing methods such as synthetic control [2] typically make strong assumptions for the spectrum, which are not benign in the real world; see Section 6. We achieve this using a novel matrix recovery bound (Theorem 2; see Section 4.3) which we believe is of general interest for the field of low-rank matrix recovery.

4 Regret Analysis

4.1 Proof Architecture for Theorem 1

The proof of Theorem 1 follows a familiar architecture that decomposes regret over time. We lay out this architecture here and will make precise two key results (Propositions 3 and 4) that enable the proof. Establishing these propositions is the core challenge in establishing a useful regret guarantee. In what follows, we find it convenient to define the 'true context' vector $\bar{x}_t^\top \triangleq [a_t \ \bar{z}_t^\top]$, and the associated precision matrix $\bar{\Omega}_t \triangleq \rho I + \sum_{s=1}^t \bar{x}_s \bar{x}_s^\top$. The Elliptical Potential Lemma [6] then states

Lemma 1 (Elliptical Potential Lemma). *Under Assumption 1, it holds that*

$$\sum_{t=0}^{T-1} \|\bar{x}_{t+1}\|_{\bar{\Omega}_t^{-1}} = O\left(\sqrt{rT \log T}\right).$$

Now, we must control the error in our estimates of the context vectors, $\bar{Z}_t$, and the consequent error in estimating τ^*. Specifically, define C_t^{latent} to be the event that the error in recovering $\bar{Z}_t$ is small:

$$C_t^{\text{latent}} \triangleq \left\{ \inf_{\Phi \in \mathcal{O}_r} \|\bar{Z}_t - \hat{Z}_t \Phi\| \le \alpha \right\}.$$

Here $\mathcal{O}_r$ is the set of r-dimensional rotations, and $\alpha \le c\sigma$ for some universal constant c. Observe that we only control this error up to a rotation. We define the event that the error in estimating τ_t^* is small, C_t^{est}, according to $C_t^{\text{est}} \triangleq \{|\tau^* - \hat{\tau}_t| \le \beta_t \hat{\sigma}_t / 2\}$.

We will control single-step regret on the 'clean' event that both these errors are controlled, $C_t \triangleq C_t^{\text{latent}} \cap C_t^{\text{est}}$; this is a high probability event:

Proposition 3 (clean event). *For all t, under Assumption 1, $\mathbb{P}(C_t) \ge 1 - O(1/t^2)$.*

This result is proved in Appendix C. The result relies on an analysis generalizing the Davis-Kahan theorem (to control C_t^{latent}) and the usual self-normalized martingale concentration bounds (to control C_t^{est}). A key additional ingredient is needed — in controlling C_t^{est}, we must deal with the issue of inconsistent designs in the regression (4). As discussed in Section 3, one issue driving this inconsistency is the fact that $\bar{Z}_t$ can only be recovered up to a rotation. The proof of Proposition 3 overcomes this challenge by showing that the distribution of actions is invariant to rotations of $\bar{Z}_t$,

so that we can assume a canonical rotation without loss of generality. We now state our bound on single-step regret; this is the key result in our regret analysis and will be proved later in this section:

Proposition 4 (Single-step regret). *For some universal constant c_1, we have for all t,*

$$\mathbb{E}\left[R_{t+1} \mid C_t\right] \leq c_1(1+\alpha)\beta_t \mathbb{E}\left[\|\bar{x}_{t+1}\|_{\bar{\Omega}_t^{-1}} \middle| C_t\right].$$

Theorem 1 then follows immediately from summing single-step regret and applying the Elliptical Potential Lemma.

4.2 Bounding the Single-Step Regret

Proposition 4 is a critical enabler of our regret guarantee. The proof is remarkably short and intuitive in the case of two treatments, which we present here. Our results generalize to K treatments (see Section 7.1) via a more complicated geometric argument. We will begin with stating three lemmas key to the proof. To that end, let $\bar{x}_{t+1}^{*\top} \triangleq [a^* \ \bar{z}_{t+1}^\top]$, where $a^* = 1$ if $\tau^* > 0$ and $a^* = 0$ otherwise; and recall that $\bar{x}_{t+1}^\top \triangleq [a_{t+1} \ \bar{z}_{t+1}^\top]$. Then we have:

Lemma 2. *On C_t^{est}, $R_{t+1} \leq 2\beta_t \left(\|\bar{x}_{t+1}^*\|_{\Omega_t^{-1}} + \|\bar{x}_{t+1}\|_{\Omega_t^{-1}}\right)$.*

This lemma is crucial to connecting single-step regret with an appropriate norm of $\bar{x}_t$ so as to eventually facilitate the use of the elliptical potential lemma and will be proved later in this section. It is interesting to note that [8] proves a version of this result which in our setting would eventually yield regret that scaled like $\tilde{O}(r^{3/2}\sqrt{T})$ as opposed to the $\tilde{O}(r\sqrt{T})$ accomplished here; further, the present proof is short. We also note that there is a norm mis-match in the lemma above since we ideally want to measure $\bar{x}_t$ in the $\|\cdot\|_{\bar{\Omega}_t^{-1}}$ norm. To relate these two norms we note that the following is true on the event that $\bar{Z}_t$ is well approximated:

Lemma 3. *On C_t^{latent}, we have $\|x\|_{\Omega_t^{-1}} \leq (1+\alpha)\|x\|_{\bar{\Omega}_t^{-1}}$ for all $x \in \mathbb{R}^{r+1}$.*

The proof crucially uses the fact that the distribution of actions is in fact invariant to rotations of $\hat{Z}_t$, so that we can assume a canonical rotation without loss. Finally, we observe that the probability that the optimal action is selected is lower bounded by a constant:

Lemma 4. *On C_t, SCTS selects the optimal action with constant probability, $\mathbb{P}\left(a_{t+1} = a^* \mid C_t\right) \geq \frac{1}{4}$.*

So equipped, we have

$$\begin{aligned}
R_{t+1} &\leq 2\beta_t \left(\|\bar{x}_{t+1}^*\|_{\Omega_t^{-1}} + \|\bar{x}_{t+1}\|_{\Omega_t^{-1}}\right) \\
&\leq 2\beta_t(1+\alpha) \left(\|\bar{x}_{t+1}^*\|_{\bar{\Omega}_t^{-1}} + \|\bar{x}_{t+1}\|_{\bar{\Omega}_t^{-1}}\right) \\
&\leq 2\beta_t(1+\alpha) \left(\frac{\mathbb{E}[\|\bar{x}_{t+1}\|_{\bar{\Omega}_t^{-1}} \mid C_t]}{\mathbb{P}\left(\bar{x}_{t+1} = \bar{x}_{t+1}^* \mid C_t\right)} + \|\bar{x}_{t+1}\|_{\bar{\Omega}_t^{-1}}\right) \\
&\leq 2\beta_t(1+\alpha) \left(4\mathbb{E}[\|\bar{x}_{t+1}\|_{\bar{\Omega}_t^{-1}} \mid C_t] + \|\bar{x}_{t+1}\|_{\bar{\Omega}_t^{-1}}\right)
\end{aligned}$$

where the first inequality is Lemma 2, the second uses Lemma 3, the third uses the law of total expectation, and the final inequality is via Lemma 4. Taking expectations conditioned on C_t now yields the result of the proposition. In the remainder of this Section, we prove Lemmas 2 and 4.

Proof of Lemma 2 First observe that if $0 \notin [\hat{\tau}_t - \beta_t\hat{\sigma}_t, \hat{\tau}_t + \beta_t\hat{\sigma}_t]$, then $a_{t+1} = a^*$. On the other hand, if $0 \in [\hat{\tau}_t - \beta_t\hat{\sigma}_t, \hat{\tau}_t + \beta_t\hat{\sigma}_t]$, then $|\tau^*| \leq 2\beta_t\hat{\sigma}_t$. Consequently, if $a_{t+1} \neq a^*$, then $R_{t+1} = |\tau^*| \leq 2\beta_t\hat{\sigma}_t$. We then complete the proof: $\hat{\sigma}_t = \|e_1\|_{\Omega_t^{-1}} = \|x^*_{t+1} - \bar{x}_{t+1}\|_{\Omega_t^{-1}} \leq \|x^*_{t+1}\|_{\Omega_t^{-1}} + \|\bar{x}_{t+1}\|_{\Omega_t^{-1}}$

Proof of Lemma 4 Suppose $\tau^* \geq 0$. Then $a_{t+1} = a^*$ whenever $\tilde{\tau}_t \geq 0$. Note that $\mathbb{P}(\tilde{\tau}_t \geq 0) \geq \mathbb{P}(\tilde{\tau}_t \geq \tau^*) \geq \mathbb{P}(\tilde{\tau}_t \geq \hat{\tau}_t + \beta_t\hat{\sigma}_t/2) = \frac{1}{4}$. When $\tau^* < 0$, the bound holds by symmetry.

4.3 A Novel Matrix Recovery Bound Independent of Condition Number

As alluded to earlier, our regret bound suprisingly has no dependence on the spectrum of $\bar{Y}_t$, other than the rank. We achieve this via a novel matrix recovery bound, which guarantees recovery of the latent factors at a rate independent of their conditioning:

Theorem 2 (Scaled Subspace Recovery)**.** *Let $M, \bar{M}, E \in \mathbb{R}^{n\times m}$ be arbitrary matrices such that $M = E + \bar{M}$, and let $\bar{U}\bar{S}\bar{V}^\top, USV^\top$ be the SVDs of $\bar{M}, M$ respectively, truncated to r singular values. Assume $\sigma_r(\bar{M}) > 0$ and $\sigma_{r+1}(\bar{M}) = 0$. Then, there exists an orthogonal matrix $H \in \mathbb{R}^{r\times r}$ such that*

$$\|\bar{V}\bar{S}H - VS\| \leq 3\|E\|.$$

For context, we can compare this bound to the Davis-Kahan Theorem, the classical guarantee for subspace recovery under perturbation [34, 64, 51]. Under the same setup as Theorem 2, the Davis-Kahan Theorem [51] implies the following bound: there exists an orthogonal matrix H such that

$$\max(\|\bar{U}H - U\|, \|\bar{V}H - V\|) \lesssim \frac{\|E\|}{\sigma_r(\bar{M})}. \tag{5}$$

If we apply Eq. (5) directly to prove Theorem 2 (see details below), we would obtain

$$\|\bar{V}\bar{S}H - VS\| \lesssim \frac{\|E\|}{\sigma_r(\bar{M})}\|\bar{M}\| = \kappa\|E\| \tag{6}$$

which incurs an additional dependence on the conditional number $\kappa := \sigma_1(\bar{M})/\sigma_r(\bar{M})$. This condition number is typically unknown in practice, and can be arbitrarily large: in the real-world examples we consider in Section 6, we find that $\kappa \approx 10^3$.

Given that the Davis-Kahan Theorem is central to low-rank matrix recovery, with broad applications in image analysis, recommendation systems, bioinformatics, econometrics, and many other fields [61, 53, 60, 13, 40], our mini-max optimal bound[9] in Theorem 2 is of general interest; for example in optimizing the condition number dependence in matrix completion [51, 7, 31, 29].

Proof. Next, we show the proof of Theorem 2 by constructing H explicitly and using a simple trick (see Lemma 5 below). To begin, note that $\bar{V}\bar{S} = \bar{M}^\top\bar{U}$ and $VS = M^\top U$ simply because $\bar{U}^\top\bar{U} = I_r$

[9]The lower bound can be easily seen by taking $E \propto u_1 v_1^\top$ where u_1 (v_1) is the left (right) singular vector corresponding to the largest singular value of $\bar{M}$.

and $U^\top U = I_r$. Then we can rewrite

$$\|\bar{V}\bar{S}H - VS\| = \|\bar{M}^\top \bar{U}H - M^\top U\|.$$

By the triangle inequality,

$$\begin{aligned}\|\bar{M}^\top \bar{U}H - M^\top U\| &\le \|(\bar{M} - M)^\top U\| + \|\bar{M}^\top(\bar{U}H - U)\| \\ &\le \|E\| + \|\bar{M}^\top(\bar{U}H - U)\|\end{aligned}$$

where the last inequality uses that $\bar{M} - M = E$ and $\|U\| \le 1$. Then it is sufficient to bound $\|\bar{M}^\top(\bar{U}H - U)\|$. Now, if we apply the Davis-Kahan Theorem directly (Eq. (5)), we would obtain Eq. (6), a sub-optimal bound.

Instead, let us analyze $\|\bar{M}^\top(\bar{U}H - U)\|$ more carefully. Note that

$$\begin{aligned}\|\bar{M}^\top(\bar{U}H - U)\| &= \|\bar{V}\bar{S}H - \bar{V}\bar{S}\bar{U}^\top U\| \\ &\le \|\bar{V}\|\|\bar{S}(H - \bar{U}^\top U)\| \\ &\le \|\bar{S}(H - \bar{U}^\top U)\|.\end{aligned}$$

It is sufficient to bound $\|\bar{S}(H - \bar{U}^\top U)\|$. To do so, we construct H to be the optimal orthogonal matrix for minimizing $\|\bar{U}H - U\|_{\mathrm{F}}$ (i.e., aligning $\bar{U}$ and U in Frobenius distance), a common method used in Davis-Kahan-type results. By Orthogonal Procrustes Theorem, we have

$$H \triangleq \tilde{U}\tilde{V}^\top$$

where $\bar{U}^\top U = \tilde{U}\tilde{\Sigma}\tilde{V}^\top$ is the SVD of $\bar{U}^\top U$. Then

$$\begin{aligned}\|\bar{S}(H - \bar{U}^\top U)\| &= \|\bar{S}\tilde{U}(I_r - \tilde{\Sigma})\tilde{V}^\top\| \\ &= \|\bar{S}\tilde{U}(I_r - \tilde{\Sigma})\|\end{aligned}$$

where we use that $\tilde{V} \in \mathbb{R}^{r\times r}$ is an orthogonal matrix (so preserving the operator norm).

Next, we observe the following fact that turns out to be the key inequality of proving our result.

Lemma 5.

$$\|\bar{S}\tilde{U}(I_r - \tilde{\Sigma})\| \le \|\bar{S}\tilde{U}(I_r - \tilde{\Sigma}^2)\| \tag{7}$$

Proof. To see this, note that $\tilde{\Sigma}$ is a diagonal matrix. Further, $0 \le \tilde{\Sigma}_{ii} \le 1$ for $i \in [n]$ since it encodes the singular values of $\bar{U}^\top U$ (the largest singular value $\sigma_1(\bar{U}^\top U) = \|\bar{U}^\top U\| \le \|\bar{U}\|\|U\| = 1$). Therefore, $I_r - \tilde{\Sigma}$ is dominated by $I_r - \tilde{\Sigma}^2$ (for each entry):

$$\forall i \in [n], \quad 0 \le (I_r - \tilde{\Sigma})_{ii} \le (I_r - \tilde{\Sigma}^2)_{ii}.$$

Next, by the property of operator norm,

$$\begin{aligned}
\|\bar{S}\tilde{U}(I_r - \tilde{\Sigma})\| &= \max_{a\in\mathbb{R}^r, \|a\|=1} \|a^\top \bar{S}\tilde{U}(I_r - \tilde{\Sigma})\| \\
&= \max_{b^\top = a^\top \bar{S}\tilde{U}, a\in\mathbb{R}^r, \|a\|=1} \|b^\top (I_r - \tilde{\Sigma})\| \\
&\overset{(i)}{\leq} \max_{b^\top = a^\top \bar{S}\tilde{U}, a\in\mathbb{R}^r, \|a\|=1} \|b^\top (I_r - \tilde{\Sigma^2})\| \\
&= \max_{a\in\mathbb{R}^r, \|a\|=1} \|a^\top \bar{S}\tilde{U}(I_r - \tilde{\Sigma}^2)\| \\
&= \|\bar{S}\tilde{U}(I_r - \tilde{\Sigma}^2)\|.
\end{aligned}$$

Here (i) uses that $\|b^\top(I_r - \tilde{\Sigma})\|^2 = \sum_i b_i^2(1-\Sigma_{ii})^2 \leq \sum_i b_i^2(1-\Sigma_{ii}^2)^2 = \|b^\top(I_r - \tilde{\Sigma}^2)\|^2$. This completes the proof. ■

By Eq. (7), we then have

$$\begin{aligned}
\|\bar{S}(H - \bar{U}^\top U)\| &\leq \|\bar{S}\tilde{U}(I_r - \tilde{\Sigma}^2)\| \\
&= \|\bar{S}\tilde{U}(I_r - \tilde{\Sigma}^2)\tilde{U}^\top\|
\end{aligned}$$

where the last inequality uses that $\tilde{U} \in \mathbb{R}^{r\times r}$ is an orthogonal matrix again. Note that $\tilde{U}\tilde{U}^\top = I_r$ and $\tilde{U}\tilde{\Sigma}^2\tilde{U}^\top = \bar{U}^\top U U^\top \bar{U}$ (this identity inspires us to replace $\tilde{\Sigma}$ by $\tilde{\Sigma}^2$). Then,

$$\begin{aligned}
\|\bar{S}(H - \bar{U}^\top U)\| &\leq \|\bar{S}\tilde{U}(I_r - \tilde{\Sigma}^2)\tilde{U}^\top\| \\
&\leq \|\bar{S} - \bar{S}\bar{U}^\top U U^\top \bar{U}\| \\
&= \|\bar{S}\bar{U}^\top (I_n - UU^\top)\bar{U}\|
\end{aligned}$$

where the last equality is due to $\bar{U}^\top\bar{U} = I_r$. Next, we use $\bar{S}\bar{U}^\top = \bar{V}^\top \bar{M}^\top$, then

$$\begin{aligned}
\|\bar{S}(H - \bar{U}^\top U)\| &\leq \|\bar{V}^\top \bar{M}^\top (I_n - UU^\top)\bar{U}\| \\
&\leq \|\bar{M}^\top (I_n - UU^\top)\| \\
&= \|(M-E)^\top (I_n - UU^\top)\|.
\end{aligned}$$

Finally note that $\|M(I_n - UU^\top)\| = \sigma_{r+1}(M) \leq \|E\|$ due to Weyl's inequality. Then

$$\begin{aligned}
\|\bar{S}(H - \bar{U}^\top U)\| &\leq \|E\| + \|E^\top (I_n - UU^\top)\| \\
&\leq 2\|E\|.
\end{aligned}$$

This completes the proof. ■

5 Inference

Having run our experiment up to time T, we now turn to the problem of recovering two key estimands: the treatment effect τ^*, and the optimal action a^*. We will examine the inferential guarantees possible under SCTS, as compared with fixed design synthetic control, in two settings: (a) when the latent factor process $\{\bar{z}_t\}$ is arbitrary, in which case fixed-design synthetic control utterly fails to learn a^*, while SCTS recovers a^* with probability approaching one; and (b) when $\{\bar{z}_t\}$ is favorable to fixed-design synthetic control, in which case both designs admit recovery of τ^* at the same rate, when τ^* is positive.

5.1 Recovery of the optimal treatment

We first consider the case where the latent factor process $\{\bar{z}_t\}$ is arbitrary, other than being non-anticipative and having bounded norm. As alluded to earlier, without further assumptions, experiments using fixed treatment patterns can fail to yield any information about the optimal treatment. This negative result in fact generalizes to any experiment design such that a_t is deterministic conditional on $\mathcal{F}_t$:

Proposition 5. *Let $\{a_t\}$ be any $\mathcal{F}_t$-adapted treatment pattern, and let $\hat{a}$ be any estimator of the optimal treatment. Then, there exists a problem instance $\{\bar{z}_t\}_{t\in[T]}, \tau^*, \{\lambda^i\}_{i\in[n]}$, where $\{\bar{z}_t\}$ satisfies Assumption 1, such that $\mathbb{P}(\hat{a} \neq a^*) \geq 1/2$.*

In contrast, we show that under the SCTS design (which, one should note, depends on algorithmic noise in addition to $\mathcal{F}_t$), one can always construct a simple hypothesis test, which identifies the optimal treatment with high probability as T grows:

Proposition 6. *Let $\hat{a}$ be the most frequently pulled arm under SCTS. Under Assumption 1, we have $P(\hat{a} \neq a^*) \leq Cr\frac{\log T}{T\tau^{*2}}$ for some absolute constant C.*

That is, taking τ^* to be constant, the probability that SCTS fails to identify the optimal arm is $\tilde{O}(\frac{1}{T})$. In practical terms, this guarantees that by running SCTS, one can estimate a^* correctly with probability approaching one; and furthermore, one can then perform a *power analysis* — prior to the experiment, the analyst can choose an experimentation budget T to guarantee that SCTS identifies the optimal arm with probability $1-\delta$. Clearly, no such analysis is possible a priori under any experimental design for which Proposition 5 holds.

5.2 Estimation of the treatment effect

We now consider estimation of the treatment effect in settings favorable to fixed-design synthetic control. Here, we will develop an estimator τ^{SCTS} for use with the SCTS design, which

- Enjoys identical confidence intervals to the vanilla SC estimator when τ^* is non-negative;
- Is with high probability 0 when the treatment effect τ^* is negative.

As such we make precise our promise of precisely estimating the treatment effect when it matters (i.e., when the treatment effect is non-negative, so as to permit a cost-benefit analysis of implementation, say), while concluding that the treatment is ineffective when it is not.

Fixed-design synthetic control To make this setting precise, we first consider what can be accomplished with synthetic control and a fixed design [2]. In particular, SC seeks to find a linear combination of donor units to match the experimental unit based on the observations from the pre-treatment period. SC assumes the existence of weights $w_1, w_2, \ldots, w_n$ such that $y_t^0 = \sum_{i=1}^n w_i y_t^i$ for all times t in the pre-treatment period $\{-T_0+1, \ldots, 0\}$. These weights are required to be non-negative, and must sum to one.[10] SC estimates the treatment effect τ^{SC} by averaging over the differences between the synthetic control so constructed and the observed y_t^0 over the treatment period, i.e., $\tau^{\mathrm{SC}} := \frac{1}{T}\sum_{t=1}^T \left(y_t^0 - \sum_{i=1}^n w_i y_t^i\right)$.

We restate a proposition which gives high-probability confidence intervals for synthetic controls (see Appendix B in [2]):

Proposition 1. *With probability* $1 - O(\delta)$,

$$|\tau^{\mathrm{SC}} - \tau^*| \lesssim \frac{\sigma}{\sqrt{T}}\sqrt{\log(1/\delta)} + \frac{c_2\sigma}{\sqrt{c_1 T_0}}\sqrt{\log(1/\delta) + \log(n)}.$$

where $c_1 = \sigma_r(1/T_0 \sum_{t=-T_0+1,\ldots,0} \bar{z}_t \bar{z}_t^\top)$ and $c_2 \triangleq \max_t \|z_t\|$. The key term is c_1, which measures the conditioning of the subspace spanned by the latent factors over the pre-treatment period. In full generality c_1 can be arbitrarily small: at one extreme of non-stationarity, where some dimension of $\{\bar{z}_t\}$ is not observed during the pre-experiment period (but subsequently becomes significant during the experiment period), c_1 is then 0. The favorable case, then, is the regime where these bounds are non-vacuous, i.e. $c_1 = \Omega(1)$ and $\delta = 1/\mathrm{poly}(n)$. In this setting, the result (optimal up to logarithmic terms) is: $|\tau^{\mathrm{SC}} - \tau^*| = \tilde{O}\left(1/\sqrt{T_0} + 1/\sqrt{T}\right)$.

Using the synthetic control estimator with an SCTS design We now describe one potential estimator τ^{SCTS} for use with our dynamic SCTS design, built on top of the usual synthetic control estimator [2].[11]

Let $w_1, w_2, \ldots, w_n$ be the same weights used by the vanilla SC estimator; note that these are computed using data available exclusively over the pre-treatment period. Now let $M \triangleq \{t \geq 1 : a_t = 1\}$ be the epochs over which we experiment according to our dynamic SCTS design, and denote by $\tilde{\tau}^{\mathrm{SC}}$, the average difference between observed outcomes and the synthetic control over those epochs: $\tilde{\tau}^{\mathrm{SC}} \triangleq \frac{1}{|M|}\sum_{t\in M} \left(y_t^0 - \sum_{i=1}^n w_i y_t^i\right)$.

We then propose an estimator, τ^{SCTS} for the treatment effect, where $\tau^{\mathrm{SCTS}} = \tau^{\mathrm{SC}}$ if $|M| \geq T/2$, and $\tau^{\mathrm{SCTS}} = 0$ otherwise. This estimator then enjoys high-probability confidence intervals analogous to the vanilla synthetic control setting:

[10] This requirement that the synthetic control be constructed as a *convex* (as opposed to affine) combination of donor units serves effectively as a regularization mechanism.

[11] An alternative would be to use the ridge regression estimator $\hat{\tau}_T$; we instead introduce τ^{SCTS} to enable a better comparison with the "vanilla" synthetic control estimator from [2].

Proposition 7. [12] *Let τ^* be fixed. Then, when $\tau^* \geq 0$, with probability $1 - O(\delta) - \tilde{O}(1/\sqrt{T})$,*

$$|\tau^{\text{SCTS}} - \tau^*| \lesssim \frac{\sigma}{\sqrt{T}}\sqrt{\log(1/\delta)} + \frac{c_2\sigma}{c_1\sqrt{T_0}}\sqrt{\log(1/\delta) + \log(n)}.$$

and when $\tau^* < 0$, with probability at least $1 - \tilde{O}(1/\sqrt{T})$, $\tau^{\text{SCTS}} = 0$.

As outlined at the outset, the result above shows high probability confidence intervals analogous to the vanilla synthetic control setting on the event that the treatment effect is non-negative. On the event where the treatment effect is negative, we see that we only learn that this is the case but do not recover a precise estimate of the effect. As argued earlier, in the practical settings we care about, a precise estimate of the treatment effect is typically not as important when the effect is negative. As a result the regret gains made possible via the use of SCTS likely constitute a beneficial tradeoff relative to the inference possible under τ^{SCTS}.

As discussed earlier, the high-probability confidence intervals described in this section are typically quite conservative in practice. As such, in our experiments, we will explore a re-randomization approach to hypothesis testing, and derive confidence intervals by inverting these tests.

6 Experiments

This section undertakes an experimental evaluation of SCTS using both synthetic and real-world datasets. In the latter datasets, it is unclear that the synthetic control model holds (i.e., it is unclear that the observed data can be explained by a low rank factor model), and the latent factor process exhibits substantial nonstationarity. We compare SCTS against both the standard fixed design (where $a_t = 1$ over the entire treatment period), as well as a switchback design (where a_t is set to 1 with probability 1/2 independently for each epoch in the treatment period) [26]. In the case of these incumbent designs, we estimate the treatment effect using state-of-the-art estimators gleaned from recent advances in 'robust' synthetic control and panel data regression. Our experiments will illustrate the following salient features of SCTS:

1. The fraction of time SCTS chooses a sub-optimal action is small, both in synthetic and real-world environments. In contrast, by definition, the switchback design picks a sub-optimal action half of the time, while the fixed design picks the sub-optimal action all of the time when the treatment effect is negative.
2. In a regime favorable to fixed design synthetic control, where we generate a stationary latent factor process, our estimator of the treatment effect under SCTS achieves relative error comparable to the competing designs when the treatment effect is positive. This same estimator correctly identified that the treatment effect was negative in all experiments where this was the case.
3. In the case of real world data where the latent factor process is non-stationary, and where the fit of the latent factor model is not exact, we observe the same relative merits alluded to above,

[12] We can also show a stronger version of this result that makes precise the dependence of the rate on τ^*, and allows for meaningful confidence intervals provided $|\tau^*| = \omega(1/\sqrt{T})$.

and in addition SCTS estimates the treatment effect much more accurately than the fixed design. This exemplifies the insight that estimation under SCTS is robust to non-stationarity, while in practice fixed designs may not admit accurate estimation of the treatment effect.

4. Re-randomization tests [41] provide a means to construct well-powered hypothesis tests and confidence intervals for SCTS, despite the inferential challenges introduced by adaptive treatment assignment.

6.1 Low Regret and Estimation Error on Synthetic Data

Our first set of experiments seeks to establish that SCTS incurs low regret while recovering the treatment effect accurately. We first consider this on a synthetic set of problems that we describe next, where the latent factor process is stationary and therefore favorable to fixed experimental designs. *Experimental setup:* We experiment with a synthetically generated dataset. We generate the latent factors and loadings $\Lambda, \bar{Z}_T$ with entries distributed i.i.d. as $\mathcal{N}(0,1)$, with $n = 1000, T_0 = 500, T = 500$ and $r = 50$. Similarly, we generate λ^* with i.i.d. $\mathcal{N}(0,1)$ entries. We experiment with a signal-to-noise ratio of 1; i.e., $|\tau^*|/\sigma = 1$, and vary the sign of τ^*. *Estimation:* There are a variety of treatment effect estimators we could use for any given experimental design. For SC, we report the best performance over several estimators from the literature; the best performing estimator on our synthetic examples is the robust synthetic control estimator proposed in [12]. For the switchback design, we use our ridge regression estimator $\hat{\tau}_T$. For SCTS, we estimate the treatment effect as $\hat{\tau}_T$ when $M \geq \frac{N}{2}$, and 0 otherwise. *Results:* Table 1 shows mean regret and estimation error for each algorithm averaged over 50 problem instances. The results comport favorably with the salient features we outlined for SCTS at the outset. Specifically:

1. SCTS picks a sub-optimal action over at most 2% of the available testing epochs, thereby mitigating any cost of experimentation. By construction this number is 50% for the switchback design and 100% for the fixed design when the treatment effective is negative.
2. Despite the above gain, we continue to recover the treatment effect accurately with the SCTS design. The relative RMSE was comparable to switchback and fixed designs when the treatment effect is positive, and SCTS correctly identified that the treatment effect was negative in all instances where this was the case.
3. From the plots in Figure 1, we see that the relative merits of SCTS alluded to above are robust to the choice of experimentation horizon T.

6.2 Real-world data

Our second set of experiments serves the same purpose as the earlier set, except that this time we consider real world data. As such, the latent factor process is non-stationary, and there is no true low-rank factor model describing the data; at best we may hope that such a model provides a good approximation to the observed data. As before, our goal will be to measure regret for SCTS as well as estimation error.

		SCTS	Switchback	SC
$\tau^* < 0$	Regret	0.02	0.50	1.00
$\tau^* > 0$	Regret	0.01	0.50	0.00
$\tau^* > 0$	RMSE	0.06	0.08	0.06

Table 1: Regret and relative RMSE, averaged over 50 random synthetic instances. Regret is normalized between 0 and 1. RMSE is normalized by $|\tau^*|$. SCTS virtually eliminates the cost of experimentation, while providing estimates of τ^* of the same quality as more costly alternatives when $\tau^* > 0$.

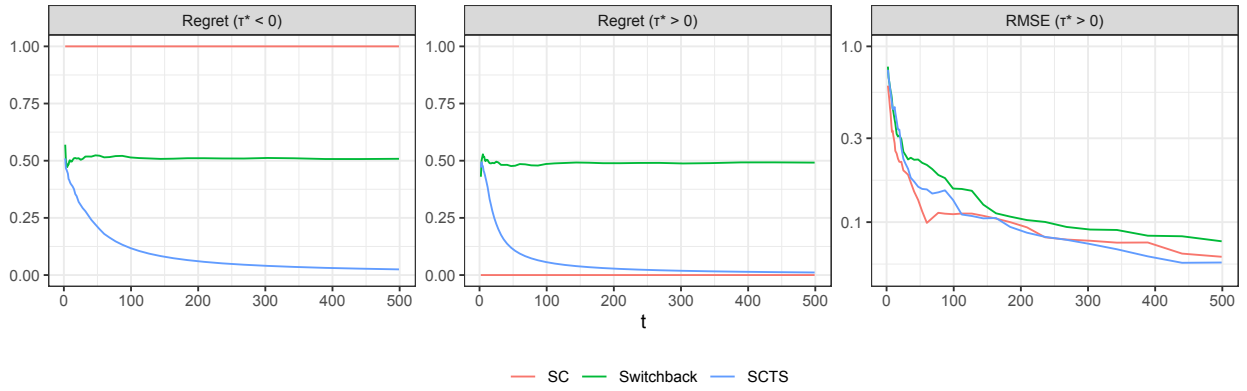

Figure 1: Regret (normalized by $t|\tau^*|$) and RMSE (normalized by $|\tau^*|$) over time, for the synthetic dataset. Unlike SC and switchback, SCTS exhibits regret vanishing over time in addition to a small RMSE. These qualities are robust to the experimentation horizon T.

Experimental Setup: We adapt the Rossman Store Sales dataset[13], which contains daily sales data for $n = 1115$ drug stores over 942 days. We take $T_0 = T = 471$. Letting $O \in \mathbb{R}^{n \times (T+T_0)}$ be the matrix of observations in the dataset, we generate an ensemble of 50 instances as follows. For each instance, we select a random store j to be the experimental unit, with outcomes $y_t^0 = O_{j,t} + \tau^* a_t$. The remaining stores constitute the control units, with observations $y_t^i = O_{i,t}$. Viewing the rank r now as an algorithmic hyper-parameter, we use $r = 70$ in our experiments. This choice was made via cross-validation on the pre-treatment period, as in [56]. As before, we experiment with two values of τ^*: $\tau^* = \sigma$ and $\tau^* = -\sigma$, where σ^2 is estimated as mean squared error of O relative to its best rank 70 approximation. *Estimation:* We use the same set of estimators here as in the previous set of experiments with synthetic data. *Results:* At the outset, we note that the model equations (1)–(2) on which any of our designs or estimation approaches are predicated do not hold exactly in this setup. In particular, approximation error essentially precludes the 'noise' in our rank 70 model from being Gaussian or i.i.d. Referring to Table 2, we observe:

1. SCTS picks a sub-optimal action over at most 13% of the available testing epochs, mitigating the cost of experimentation.
2. Despite the above gain, we continue to recover the treatment effect accurately with the SCTS

[13] https://www.kaggle.com/c/rossmann-store-sales/

design, with relative RMSE is comparable to the switchback design. SCTS correctly identified that the treatment effect was negative in all instances where this was the case.

3. Especially interesting is that relative RMSE is substantially lower than that for the fixed design, exemplifying the robustness of SCTS over fixed designs. This setting is precisely one where, due to non-stationarity, the synthetic controls are poorly estimated during the pre-treatment period. In fact, the constant c_1 of Proposition 1 is roughly 10^{-3}. Here, SCTS estimates the treatment effect well, while the fixed design fails to improve its estimate of the treatment effect over time.

		SCTS	Switchback	SC
$\tau^* < 0$	Regret	0.13	0.50	1.00
$\tau^* > 0$	Regret	0.09	0.50	0.00
$\tau^* > 0$	RMSE	0.12	0.07,0.91*	0.57

Table 2: Regret and relative estimation error, averaged over random instances generated from the Rossman dataset. Regret is normalized between 0 and 1. RMSE is normalized by $|\tau^*|$. In addition to low regret, SCTS even produces better quality estimates of τ^* in this setting, compared to SC. For Switchback, we report RMSE for two estimators: $\hat{\tau}_T$ (RMSE=0.07), and a simple difference in means with no synthetic controls (RMSE=0.91), highlighting the importance of synthetic controls even with switchback designs.

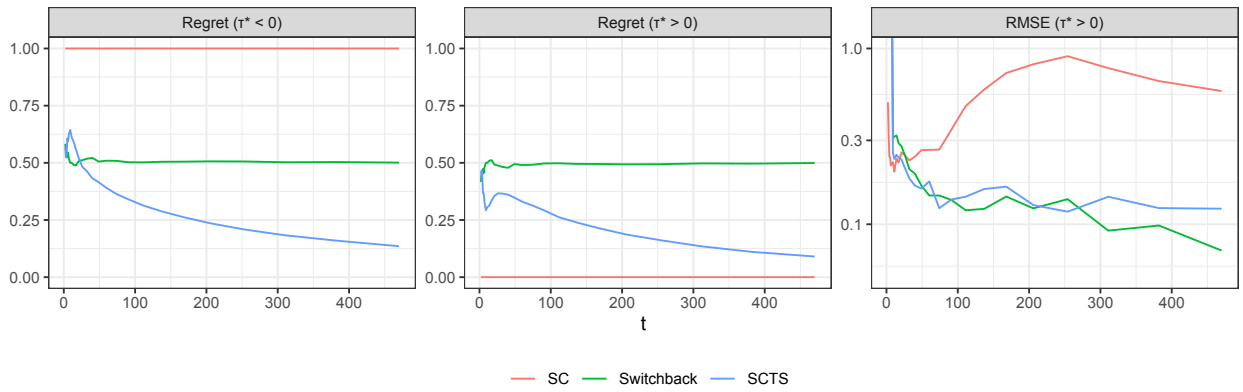

Figure 2: Regret (normalized by $t|\tau^*|$) and RMSE (normalized by $|\tau^*|$) over time, averaged over instances from the Rossman dataset. Unlike SC and Switchback, SCTS exhibits regret vanishing over time in addition to a small RMSE. In particular SCTS and Switchback estimators both display much lower RMSE than SC. These qualities hold essentially for all t in the horizon.

6.3 A Non-Parametric Approach to Inference

Finally, we explore a re-randomization approach to inference for SCTS. In particular, while the high-probability concentration bounds of Lemma 3 can be used to provide confidence intervals for $\hat{\tau}$, they assume that the structural model (1) is realizable, and tend to be conservative in practice. On the other hand, post-bandit inference techniques such as those in [36, 24] could possibly be adapted

	SNR = 0.01	0.02	0.1	0.2	1
Coverage	0.89	0.87	0.87	0.90	0.89
Power	0.14	0.17	0.51	0.81	0.98

Table 3: Performance of the re-randomization test and confidence intervals on Rossman problem instances, as a function of the effect size (normalized as SNR). As expected, coverage attains roughly the nominal level $1 - \alpha$ where $\alpha = 0.1$, while power increases quickly as we increase the effect size.

to our setting, but do not apply immediately (see Section 1.2 for discussion). This latter direction remains an exciting direction for future work.

Here we propose a re-randomization test similar to that of [25] to construct a hypothesis test for a sharp null that the treatment effect is some specific value. We then obtain confidence intervals by inverting this hypothesis test, as described in [45]. The overall conclusion in this section is that (a) our hypothesis test is highly powered for relatively low values of SNR where the treatment effect is dominated by the noise and (b) our confidence intervals attain nearly ideal coverage even for very low SNR. All of these experiments are run on the real data setup described in the preceding section. *A Re-Randomized Hypothesis Test and Confidence Intervals:* We test the sharp null hypothesis H_τ that the treatment effect is some constant τ, for all t. To implement such a test, suppose that we have run SCTS for T time steps, obtaining an estimator $\hat{\tau}_T$. We take this estimator to be our test statistic. We can then construct an approximate hypothesis test, at significance level α, as follows:

1. We are given an observed trajectory of interventions under SCTS, $\{a_t^{\text{hist}}\}$ and the corresponding observations on the experimental unit $\{y_t^{0,\text{hist}}\}$.
2. We next draw k samples $\tau^{(1)} \dots \tau^{(k)}$ of the test statistic under the null hypothesis. We do so by re-running SCTS, but assuming that we observe the sequence of outcomes $y_t^{0,\text{hist}} + \tau a_t - \tau a_t^{\text{hist}}$
3. We can then approximate the p-value of the test statistic as one minus the proportion of the samples $\tau^{(1)} \dots \tau^{(k)}$ which are less than $\hat{\tau}_T$.
4. We reject H_τ if the p-value is less than the significance level α.

We may now construct confidence intervals by 'inverting' the above re-randomized hypothesis test, as described in [45]. Precisely, for every null H_τ for some $\tau \in \mathbb{R}$, we can implement a re-randomization test and decide whether to reject H_τ. The confidence set is then the set of τ values for which we do not reject the corresponding null H_τ. *Results:* We assess our re-randomization test on 100 problem instances generated from the Rossman sales dataset, as above. We draw $k = 100$ samples of the test statistic for each instance and choose the the significance level to be $\alpha = 0.1$. The results in Table 3 show that this test is highly powered even when the treatment effect is dominated by the noise (i.e., at an SNR of 0.2). Power is already nearly 1 at an SNR of 1. Further, we see that the coverage of the test is close to ideal (given the significance level of 0.1, ideal here is 0.9) over a broad range of SNRs from 0.01 to 1. In summary, we conclude that the re-randomization tests and corresponding confidence intervals reported here are adequate for inference even when SNR is low.

7 Discussion

We motivated our dynamic design by the real-world setting where, to mitigate interference, we must select 'coarse' treatment units. By nature, coarse units require synthetic controls for high-quality inference, and also suffer a cost of experimentation (as characterized by regret) that may not be trivially ignored. At the same time, the non-stationarity embodied by the synthetic controls has the potential to confound inference, so as to render any fixed experimental design uninformative.

The SCTS approach is a new dynamic design that addresses these issues. A number of real-world concerns, however, are not yet captured by the present model. Here, we provide extensions that address some of these issues, and outline several more that represent promising directions for future work.

7.1 Extensions

Instance-dependent bounds The bound of Theorem 1 quantifies an instance-independent, worst case regret for SCTS; in particular this entails a worst-case choice of treatment effect. A more refined analysis reveals that regret may indeed be smaller – logarithmic in T – depending on the scaling of the treatment effect:

Theorem 3. *Let $n = \Omega(T)$. Under Assumption 1, SCTS achieves expected regret $R(T) = O\left(r^2 \frac{\log(T)}{|\tau^*|}\right)$.*

As a corollary of this result, we see that the frequency with which SCTS applies a suboptimal treatment – and by extension, the frequency with which SCTS switches treatments – decreases with T when the treatment effect is sufficiently large. This is a major advantage relative to non-adaptive randomized designs (such as "switchbacks", see e.g. [25]) in settings where switching treatments has associated friction.

Early-stopping Another related experimentation objective may be to identify the optimal treatment with high probability, as quickly as possible. To this end, we can define an early-stopping variant of SCTS, which simply stops the experiment when either the confidence set excludes $\tau^* = 0$ (i.e., the sign of τ^* has been identified with high probability), or the horizon T ends. This early-stopping approach has the following guarantee:

Theorem 4. *Let s be the early stopping time, and assume $n = \Omega(T)$. Then, the expected time to stop is $\mathsf{E}[s] = O(r^2 \log(T)/|\tau_*|^2)$, and the probability SCTS fails to identify the optimal treatment is bounded by $O(r^2 \log(T)/(T|\tau^*|^2))$.*

Therefore, taking the treatment effect and rank to be constants, one can expect to identify the optimal treatment with probability $1 - \tilde{O}(1/T)$, with an expected experiment duration of $O(\log(T))$.

7.2 Limitations and Open Directions

Spill-over Effects In dynamic designs — the switchback design is a simple example, SCTS is another — one often cares about 'spill-over' effects wherein a treatment applied in one epoch might

influence outcomes in subsequent epochs. The fix to this issue is to typically allow a 'burn-in' period that ignores epochs impacted by such spill-overs. These burn-in periods typically precede and follow a switch from one type of treatment to another. Whereas we have not posited a formal model, it is reasonable to conjecture that in our setting, one could employ a similar strategy. Since the number of switches in SCTS is small (i.e. $O(\log(T)/|\tau^*|^2)$), the added regret from such burn-in periods will scale sub-linearly with the horizon.

Interventions over Consecutive Epochs For some interventions, it may be practically necessary (for instance, from a consumer experience standpoint) that any intervention be maintained over a certain minimum number of consecutive epochs. We believe this to be an important area for future work, closely related to notions of switching costs and batching in the bandit literature. A number of flavors of this problem have been considered in recent years, including incorporating switching costs [35], and batching that makes the decision to stop using a potential intervention irrevocable [57]. Very recently, [39, 43] have extended the batched bandit formalism to linear contextual bandits. Whereas none of these models precisely address the modeling need above, they provide a very reasonable foundation for a potential extension to SCTS that incorporates the constraint that any intervention must be pursued for a certain minimal number of consecutive epochs.

Post-Bandit Inference As discussed earlier, while we can establish high-probability confidence intervals (that are conservative) and re-randomization tests (that appear to work well practically), we would ideally like to construct estimators with limiting distributions that permit powerful inference. The growing post-bandit inference literature [24, 42, 36] provides an approach to accomplishing this goal. The primary roadblock here is that existing proposals ask for a lower bound on the rate of decay of exploration, and it is not clear that such a lower bound is met by the current proposal.

References

[1] A. Abadie. Using synthetic controls: Feasibility, data requirements, and methodological aspects. *J. of Economic Literature*, 2019.

[2] A. Abadie, A. Diamond, and J. Hainmueller. Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *J. of the American statistical Assoc.*, 105(490):493–505, 2010.

[3] A. Abadie, A. Diamond, and J. Hainmueller. Comparative politics and the synthetic control method. *American J. of Political Science*, 59(2):495–510, 2015.

[4] A. Abadie and J. Gardeazabal. The economic costs of conflict: A case study of the basque country. *American economic review*, 93(1):113–132, 2003.

[5] A. Abadie and J. Zhao. Synthetic controls for experimental design. *arXiv preprint arXiv:2108.02196*, 2021.

[6] Y. Abbasi-Yadkori, D. Pál, and C. Szepesvári. Improved Algorithms for Linear Stochastic Bandits. 2011.

[7] E. Abbe, J. Fan, K. Wang, Y. Zhong, et al. Entrywise eigenvector analysis of random matrices with low expected rank. *Annals of Statistics*, 48(3):1452–1474, 2020.

[8] M. Abeille and A. Lazaric. Linear Thompson Sampling Revisited. In *Proc. of the 20th Intl. Conference on Artificial Intelligence and Statistics, AISTATS 2017, 20-22 April 2017, Fort Lauderdale, FL, USA*, pages 176–184, 2017.

[9] A. Agarwal, D. Shah, D. Shen, and D. Song. On robustness of principal component regression. *J. of the American Statistical Assoc.*, (just-accepted):1–34, 2021.

[10] S. Agrawal and N. Goyal. Thompson Sampling for Contextual Bandits with Linear Payoffs. In *Proc. of the 30th Intl. Conference on Machine Learning, ICML 2013, Atlanta, GA, USA, 16-21 June 2013*, pages 127–135, 2013.

[11] M. Amjad, V. Misra, D. Shah, and D. Shen. mrsc: Multi-dimensional robust synthetic control. *Proc. of the ACM on Measurement and Analysis of Computing Systems*, 3(2):1–27, 2019.

[12] M. Amjad, D. Shah, and D. Shen. Robust synthetic control. *The J. of Machine Learning Research*, 19(1):802–852, 2018.

[13] M. J. Amjad and D. Shah. Censored demand estimation in retail. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 1(2):1–28, 2017.

[14] D. Arkhangelsky, S. Athey, D. A. Hirshberg, G. W. Imbens, and S. Wager. Synthetic difference in differences. Technical report, National Bureau of Economic Research, 2019.

[15] S. Athey, M. Bayati, N. Doudchenko, G. Imbens, and K. Khosravi. Matrix completion methods for causal panel data models. *J. of the American Statistical Assoc.*, pages 1–41, 2021.

[16] S. Athey and G. W. Imbens. The state of applied econometrics: Causality and policy evaluation. *J. of Economic Perspectives*, 31(2):3–32, 2017.

[17] P. Auer. Using Confidence Bounds for Exploitation-Exploration Trade-offs. *J. of Machine Learning Research*, 3(Nov):397–422, 2002.

[18] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire. The nonstochastic multiarmed bandit problem. *SIAM Journal on Computing*, 32(1):48–77, 2002.

[19] J. Bai, K. Li, et al. Theory and methods of panel data models with interactive effects. *Annals of Statistics*, 42(1):142–170, 2014.

[20] J. Bai and S. Ng. Matrix completion, counterfactuals, and factor analysis of missing data. *arXiv preprint arXiv:1910.06677*, 2019.

[21] E. Ben-Michael, A. Feller, and J. Rothstein. The augmented synthetic control method. *J. of the American Statistical Assoc.*, (just-accepted):1–34, 2021.

[22] D. A. Berry. Adaptive clinical trials in oncology. *Nature reviews Clinical oncology*, 9(4):199–207, 2012.

[23] O. Besbes, Y. Gur, and A. Zeevi. Stochastic Multi-Armed-Bandit Problem with Non-stationary Rewards. In *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014.

[24] A. Bibaut, A. Chambaz, M. Dimakopoulou, N. Kallus, and M. J. van der Laan. Post-Contextual-Bandit Inference. *CoRR*, abs/2106.00418, 2021.

[25] I. Bojinov, D. Simchi-Levi, and J. Zhao. Design and Analysis of Switchback Experiments. *Available at SSRN 3684168*, 2020.

[26] A. Brandt. Tests of significance in reversal or switchback trials. 234, 1938.

[27] K. H. Brodersen, F. Gallusser, J. Koehler, N. Remy, and S. L. Scott. Inferring causal impact using bayesian structural time-series models. *The Annals of Applied Statistics*, 9(1):247–274, 2015.

[28] E. J. Candès and B. Recht. Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6):717, 2009.

[29] J. Chen and X. Li. Model-free nonconvex matrix completion: Local minima analysis and applications in memory-efficient kernel pca. *Journal of Machine Learning Research*, 20(142):1–39, 2019.

[30] Y. Chen, Y. Chi, J. Fan, and C. Ma. Spectral methods for data science: A statistical perspective. *arXiv preprint arXiv:2012.08496*, 2020.

[31] Y. Chen, Y. Chi, J. Fan, C. Ma, and Y. Yan. Noisy matrix completion: Understanding statistical guarantees for convex relaxation via nonconvex optimization. *SIAM journal on optimization*, 30(4):3098–3121, 2020.

[32] Y. Chen, M. Loncaric, B. Moallemi, and S. J. Taylor. Synthetic control estimators in practice. Technical report, Lyft, 2020.

[33] V. Chernozhukov, K. Wüthrich, and Y. Zhu. An exact and robust conformal inference method for counterfactual and synthetic controls. *J. of the American Statistical Assoc.*, (just-accepted):1–44, 2021.

[34] C. Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation. iii. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970.

[35] O. Dekel, J. Ding, T. Koren, and Y. Peres. Bandits with switching costs: $T^{2/3}$ regret. In *Proc. of the Forty-Sixth Annual ACM Symposium on Theory of Computing*, STOC '14, pages 459–467, New York, NY, USA, May 2014. Assoc. for Computing Machinery.

[36] Y. Deshpande, L. Mackey, V. Syrgkanis, and M. Taddy. Accurate inference for adaptive linear models. In *Intl. Conference on Machine Learning*, pages 1194–1203. PMLR, 2018.

[37] N. Doudchenko, D. Gilinson, S. Taylor, and N. Wernerfelt. Designing experiments with synthetic controls. Technical report, Working paper, 2019.

[38] N. Doudchenko and G. W. Imbens. Balancing, regression, difference-in-differences and synthetic control methods: A synthesis. Technical report, National Bureau of Economic Research, 2016.

[39] H. Esfandiari, A. Karbasi, A. Mehrabian, and V. Mirrokni. Regret Bounds for Batched Bandits. *Proc. of the AAAI Conf. on Artificial Intelligence*, 35(8):7340–7348, May 2021.

[40] V. F. Farias, A. A. Li, and T. Peng. Learning treatment effects in panels with general intervention patterns. *arXiv preprint arXiv:2106.02780*, 2021.

[41] R. Fisher. *Design of Experiments*. Hafner of Edinburgh, 1966.

[42] V. Hadad, D. A. Hirshberg, R. Zhan, S. Wager, and S. Athey. Confidence intervals for policy evaluation in adaptive experiments. *Proc. of the National Academy of Sciences*, 118(15), 2021.

[43] Y. Han, Z. Zhou, Z. Zhou, J. Blanchet, P. W. Glynn, and Y. Ye. Sequential Batch Learning in Finite-Action Linear Contextual Bandits. Apr. 2020.

[44] C. Hsiao, H. Steve Ching, and S. Ki Wan. A panel data approach for program evaluation: measuring the benefits of political and economic integration of hong kong with mainland china. *J. of Applied Econometrics*, 27(5):705–740, 2012.

[45] G. W. Imbens and D. B. Rubin. *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press, 2015.

[46] N. Jones and Barrows. Uber's synthetic control. `https://www.youtube.com/watch?v=j5DoJV5S2Ao`, 2019.

[47] B. Kveton, M. Konobeev, M. Zaheer, C.-w. Hsu, M. Mladenov, C. Boutilier, and C. Szepesvari. Meta-Thompson Sampling. *arXiv preprint arXiv:2102.06129*, 2021.

[48] T. Lattimore and C. Szepesvári. Bandit Algorithms. July 2020.

[49] K. T. Li. Statistical inference for average treatment effects estimated by synthetic control methods. *J. of the American Statistical Assoc.*, 115(532):2068–2083, 2020.

[50] K. T. Li and D. R. Bell. Estimation of average treatment effects with panel data: Asymptotic theory and implementation. *J. of Econometrics*, 197(1):65–75, 2017.

[51] C. Ma, K. Wang, Y. Chi, and Y. Chen. Implicit regularization in nonconvex statistical estimation: Gradient descent converges linearly for phase retrieval, matrix completion, and blind deconvolution. *Foundations of Computational Mathematics*, pages 1–182, 2018.

[52] H. R. Moon and M. Weidner. Dynamic linear panel regression models with interactive fixed effects. *Econometric Theory*, 33(1):158–195, 2017.

[53] N. Natarajan and I. S. Dhillon. Inductive matrix completion for predicting gene–disease associations. *Bioinformatics*, 30(12):i60–i68, 2014.

[54] X. Nie, X. Tian, J. Taylor, and J. Zou. Why adaptively collected data have negative bias and how to correct for it. In *Intl. Conference on Artificial Intelligence and Statistics*, pages 1261–1269. PMLR, 2018.

[55] J. Overgoor. Experiments at airbnb. `https://medium.com/airbnb-engineering/experiments-at-airbnb-e2db3abf39e7`, 2014.

[56] A. B. Owen and P. O. Perry. Bi-Cross-Validation of the SVD and the Nonnegative Matrix Factorization. *The Annals of Applied Statistics*, 3(2):564–594, 2009.

[57] V. Perchet, P. Rigollet, S. Chassang, and E. Snowberg. Batched bandit problems. *The Annals of Statistics*, 44(2):660–681, Apr. 2016.

[58] C. Qin and D. Russo. Adaptivity and Confounding in Multi-Armed Bandit Experiments. *arXiv:2202.09036 [cs, stat]*, Mar. 2022.

[59] O. Shamir. A variant of azuma's inequality for martingales with subgaussian tails. *arXiv preprint arXiv:1110.2392*, 2011.

[60] A. M.-C. So and Y. Ye. Theory of semidefinite programming for sensor network localization. *Mathematical Programming*, 109(2):367–384, 2007.

[61] X. Su and T. M. Khoshgoftaar. A survey of collaborative filtering techniques. *Advances in artificial intelligence*, 2009, 2009.

[62] S. S. Villar, J. Bowden, and J. Wason. Multi-armed Bandit Models for the Optimal Design of Clinical Trials: Benefits and Challenges. *Statistical Science*, 30(2):199–215, May 2015.

[63] Y. Xu. Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis*, 25(1):57–76, 2017.

[64] Y. Yu, T. Wang, and R. J. Samworth. A useful variant of the davis–kahan theorem for statisticians. *Biometrika*, 102(2):315–323, 2015.

A Canonical Decomposition

Consider the structural model given by parameters $\lambda^{*\prime}, \Lambda', \bar{Z}'_T$

$$\mathbb{E}[y_t^0] = \left\langle \lambda^{*\prime}, \bar{z}'_t \right\rangle + \tau^* a_t \qquad\qquad \mathbb{E}[Y_T] = \Lambda'(\bar{Z}'_T)^\top$$

It is easy to see that the selections of $\lambda^*, \bar{Z}'_T, \Lambda'$ are not unique due to the free rotations.

Let $\mathbb{E}[Y_T] = \bar{U}_T \bar{\Sigma}_T \bar{V}_T^\top$ be an SVD of $\mathbb{E}[Y_T]$ and let $\bar{Z}_T = \frac{1}{\sqrt{n}} \bar{V}_T \bar{\Sigma}_T$. We aim to show that $\bar{Z}_T$ can constitute a canonical representation. In particular, it is sufficient to show that there exist unique λ^* and Λ such that

$$\mathbb{E}[y_t^0] = \langle \lambda^*, \bar{z}_t \rangle + \tau^* a_t \qquad\qquad \mathbb{E}[Y_T] = \Lambda(\bar{Z}_T)^\top$$

Since $\mathbb{E}\left[Y_T^\top\right]$ is exactly rank r, the column spaces of $\bar{Z}_T$ and $\bar{Z}'_T$ must be the same. Therefore there must exist an invertible matrix H such that $\bar{Z}_T = \bar{Z}'_T H$, which induces unique choices of λ^* and Λ:

$$\lambda^* = H\lambda^{*\prime}$$
$$\Lambda = \Lambda' H.$$

B Full Version of Theorem 1

B.1 Statement of Theorem

Here we state the full version of our main theorem, Theorem 1, which includes explicit dependencies on all problem parameters.

Theorem 5. *Under Assumption 1, the regret of SCTS is*

$$R(T) \le 4|\tau^*| + 10\beta_t \left(1 + \frac{1}{\sqrt{\rho}}\alpha\right)(|\tau^*| \vee 1)\sqrt{T \log\left(1 + \frac{TB^2}{r+1}\right)}$$

where we define

$$\alpha = 20\sigma\sqrt{\frac{n \vee T}{n}}$$
$$\beta_t = 2\sigma(r+1)\sqrt{2\log t\,(r + 1 + t(B + 1 + \alpha))} + (\|\lambda^*\| + |\tau^*|)(\sqrt{\rho} + \alpha)$$

C Proof of Proposition 3

We proceed with the proof by showing that $\mathbb{P}(C_t^{\text{latent}})$ and $\mathbb{P}(C_t^{\text{est}})$ are well controlled, separately.

C.1 Controlling $\mathbb{P}(C_t^{\text{latent}})$

Recall the definition of C_t^{latent}:

$$C_t^{\text{latent}} = \left\{ \inf_{\Phi \in \mathcal{O}_r} \|\bar{Z}_t - \hat{Z}_t \Phi\| \le \alpha \right\}.$$

The following lemma, based on a generalization of the Davis-Kahan bound, provides the desired bounds for $\mathbb{P}(C_t^{\text{latent}})$.

Lemma 6. *Let $\alpha \triangleq 20\sigma\sqrt{\frac{n\vee T}{n}}$. Then, for all t, with probability $1 - O(1/t^8)$,*

$$\inf_{\Phi\in\mathcal{O}_r} \|\bar{Z}_t - \hat{Z}_t\Phi\| \le \alpha \tag{8}$$

Proof. We are interested in three SVDs at time t, each of rank r:

- $\bar{Y}_T = \bar{U}_T\bar{S}_T\bar{V}_T^\top = \sqrt{n}\bar{U}_T\bar{Z}_T$, where $\bar{Z}_T$ is the "canonical" representation of the latent covariates.
- $\hat{U}_t\hat{S}_t\hat{V}_t^\top = \sqrt{n}\hat{U}_t\hat{Z}_t$, which is the SVD of Y_t truncated to r singular values.
- $\bar{Y}_t = \bar{U}_t\bar{S}_t\bar{V}_t^\top$, i.e. the SVD of the mean control outcomes up to time t.

First, observe that $\bar{Y}_t = \bar{U}_t\bar{S}_t\bar{V}_t^\top = \bar{U}_T\bar{S}_T(\bar{V}_{T,:t}^\top)$ (where $\bar{V}_{T,:t}$ denotes the first t rows of $\bar{V}_T$), and therefore $\frac{1}{\sqrt{n}}\bar{V}_t\bar{S}_t\bar{U}_t^\top\bar{U}_T = \frac{1}{\sqrt{n}}\bar{V}_{T,:t}\bar{S} = \bar{Z}_t$. This will allow us to relate $\hat{Z}_t$ to $\bar{Z}_t$:

$$\begin{aligned}
\inf_{\Phi\in\mathcal{O}_r} \|\hat{Z}_t\Phi - \bar{Z}_t\| &= \frac{1}{\sqrt{n}} \inf_{\Phi\in\mathcal{O}_r} \|\hat{V}_t\hat{S}_t\Phi - \bar{V}_t\bar{S}_t\bar{U}_t^\top\bar{U}_T\| \\
&\overset{(i)}{=} \frac{1}{\sqrt{n}} \inf_{\Phi\in\mathcal{O}_r} \|\hat{V}_t\hat{S}_t\Phi - \bar{V}_t\bar{S}_t\| \\
&\overset{(ii)}{\le} 4\frac{\|E_t\|}{\sqrt{n}}
\end{aligned}$$

where (i) uses that $\bar{U}_t^\top\bar{U}_T$ is a rotation (see below) and (ii) uses Theorem 2.[14] Then, using Theorem 6, we have $\|E_t\| \le 5\sigma\sqrt{n\vee t}$ with probability $1 - \frac{1}{(n\vee t)^8}$.

It remains to show that $\bar{U}_t^\top\bar{U}_T$ is a rotation. First, we note that $\mathrm{colspan}(\bar{U}_t) = \mathrm{colspan}(\bar{U}_T)$: if $\mathrm{rank}(\bar{Y}_t) = r$ this must hold since $\bar{Y}_t \in \mathrm{colspan}(\bar{U}_t)$ and $\bar{Y}_t \in \mathrm{colspan}(\bar{U}_T)$; if $\mathrm{rank}(\bar{Y}_t) < r$ we can always choose $\bar{U}_t$ such that this holds. As a result, projecting the columns of $\bar{U}_T$ onto $\mathrm{colspan}(\bar{U}_t)$ does not change them – i.e. $\bar{U}_t\bar{U}_t^\top\bar{U}_T = \bar{U}_T$ – and therefore $\bar{U}_T^\top\bar{U}_t\bar{U}_t^\top\bar{U}_T = I$. A symmetric argument gives that $\bar{U}_t^\top\bar{U}_T\bar{U}_T^\top\bar{U}_t = I$, from which it follows that $\bar{U}_t^\top\bar{U}_T$ is a rotation.

C.2 Controlling $\mathbb{P}(C_t^{\mathrm{est}})$

Next, we bound the probability $\mathbb{P}(C_t^{\mathrm{est}})$. Recall that $C_t^{\mathrm{est}} = \{|\tau^* - \hat{\tau}_t| \le \beta_t\hat{\sigma}_t/2\}$, where we define β_t explicitly below.

Lemma 7. *With probability $1 - \frac{2}{t^2}$, it holds that $|\tau^* - \hat{\tau}_t| \le \beta_t\hat{\sigma}_t/2$, where*

$$\beta_t = 2\sigma\sqrt{2(r+1)\log t\left(r+1+t(B+1+20\sigma\sqrt{(n\vee T)/n})\right)} + (\|\lambda^*\| + |\tau^*|)(20\sigma\sqrt{(n\vee T)/n}+1)$$

Note that $\hat{\theta}_t^\top \triangleq [\hat{\tau}_t\ \hat{\lambda}_t^\top]$ is the solution of the following quadratic program:

$$\min_{\tau\in\mathbb{R},\lambda\in\mathbb{R}^r} \sum_{s\le t} \left(y_s^0 - \tau a_s - \langle\lambda, \hat{z}_s\rangle\right)^2 + \rho(\tau^2 + \|\lambda\|_2^2) \tag{9}$$

with $y_s^0 = \tau^* a_s + \langle\lambda^*, \bar{z}_s\rangle + \epsilon_s^0$. Denote $x_s^\top \triangleq [a_s\ \hat{z}_s^\top]$, let $\Omega_t = \rho I + \sum_{s\le t} x_s x_s^\top$ be the 'precision' matrix of $\hat{\theta}_t$. The 'variance' estimator is

[14] A direct application of the Davis-Kahan theorem would have resulted in a dependence on the condition number of $\bar{Y}_t$, $\sigma_1(\bar{Y}_t)/\sigma_r(\bar{Y}_t)$. Theorem 2 represents an improved bound (detailed in Section 4.3) which allows us to avoid dependence on the condition number.

$$\hat{\sigma}_t^2 = (\Omega_t)_{1,1}^{-1}.$$

We address the following two issues to complete the proof.

1. Due to the rotation ambiguity, there is no direct connection between $\hat{z}_s$ and $\bar{z}_s$. That is to say, we can only bound $\|\Phi\hat{z}_s - \bar{z}_s\|$ for some unknown rotation Φ. For addressing this, we show that $\hat{\tau}_t, \hat{\sigma}_t$ are invariant to the rotation Φ, hence one can rewrite $\Phi\hat{z}_s$ as $\hat{z}_s$ without loss, for the purpose of analysis. See Appendix C.2.1.

2. Given the bound $\|\hat{z}_s - \bar{z}_s\|$, the problem reduces to an analysis of ridge regression with errors-in-variables, with an adapted design matrix. We adapt a typical bound from LinUCB [6] for ridge regression without errors-in-variables to our setting. See Appendix C.2.2.

C.2.1 Invariance to Rotation

To begin, we show that under our SCTS algorithm, $\hat{\tau}_t, \hat{\sigma}_t$ are invariant to any rotation of the estimated contexts $\hat{Z}_t$. This then enables us to assume that $\hat{Z}_t$ aligns with $\bar{Z}_t$ by the best rotation, given by Eq. (8), without loss.

Specifically, suppose the observed context were $\Phi\hat{z}_s$ instead $\hat{z}_s$ for some rotation $\Phi \in \mathcal{O}^{r\times r}$, and we wanted to solve the optimization problem:

$$\min_{\tau\in\mathbb{R},\lambda\in\mathbb{R}^r} \sum_{s\le t} \left(y_s^0 - \tau a_s - \langle\lambda, \Phi\hat{z}_s\rangle\right)^2 + \rho(\tau^2 + \|\lambda\|_2^2) \tag{10}$$

Let $\breve{\tau}_t, \breve{\lambda}_t$ be the corresponding optimal solution. One can easily see that

$$\begin{aligned}\breve{\tau}_t &= \hat{\tau}_t\\ \breve{\lambda}_t &= \Phi\hat{\lambda}_t\end{aligned}$$

by the equivalence between Eq. (9) and Eq. (10) through the corresponding transformation. This implies that $\hat{\tau}_t$ is invariant to the rotation Φ. Further, let $\breve{x}_s^\top \triangleq [a_s\ (\Phi\hat{z}_s)^\top]$ and

$$\begin{aligned}\breve{\Omega}_t &= \rho I + \sum_{s\le t} \breve{x}_s\breve{x}_s^\top\\ &= \rho I + \sum_{s\le t} \begin{bmatrix} a_s \\ \Phi\hat{z}_s\end{bmatrix}\begin{bmatrix} a_s \\ \Phi\hat{z}_s\end{bmatrix}^\top\\ &= I_{1,\Phi}\hat{\Omega}_t I_{1,\Phi}^\top\end{aligned}$$

where

$$I_{1,\Phi} \triangleq \begin{pmatrix} 1 & \mathbf{0}_{1\times r}\\ \mathbf{0}_{r\times 1} & \Phi\end{pmatrix}.$$

This implies that

$$\breve{\Omega}_t^{-1} = (I_{1,\Phi}\hat{\Omega}_t I_{1,\Phi}^\top)^{-1} = I_{1,\Phi}\hat{\Omega}_t^{-1} I_{1,\Phi}^\top.$$

Therefore,

$$(\breve{\Omega}_t^{-1})_{11} = e_1^\top I_{1,\Phi}\hat{\Omega}_t^{-1} I_{1,\Phi}^\top e_1 = e_1^\top\hat{\Omega}_t^{-1}e_1 = (\hat{\Omega}_t^{-1})_{11}$$

due to $I_{1,\Phi}^\top e_1 = e_1$. This is to say, $\hat{\sigma}_t$ is also invariant to the rotation Φ.

Hence, since the action chosen by SCTS only depends on $\hat{Z}_t$ via $\hat{\tau}_t$ and $\hat{\sigma}_t$, and these two quantities are invariant to the rotation of $\hat{Z}_t$, then the regret of SCTS is also invariant to the rotation of $\hat{Z}_t$. Without loss, we will rewrite $\hat{Z}_t\Phi$ as $\hat{Z}_t$ for the simplification of the analysis, if there is no ambiguity. Then under C_t^{latent}, we have

$$\|\bar{Z}_t - \hat{Z}_t\| \leq \alpha$$

C.2.2 Confidence Intervals for Ridge Regression.

First, we show a generic confidence interval for ridge regression with errors-in-variables, which adapts the bounds used in LinUCB [6]. For a matrix A, let $\|A\|_{2,\infty} = \max_{\|u\|_2=1} \|Au\|_\infty$ denote the maximum norm of its rows.

Proposition 8. *Let $\{\mathcal{F}_t\}_{t\in\mathbb{N}}$ be a filtration. Let $\{x_t, \bar{x}_t\}_{t\in\mathbb{N}}$ be a $\mathcal{F}_t$ -measurable, $\mathbb{R}^d$ -valued stochastic process. Let $\{\varepsilon_t\}_{t\in\mathbb{N}}$ be an $\mathcal{F}_{t+1}$ -measurable, $\mathbb{R}$ -valued stochastic process. Let ε_t further be σ-subgaussian conditional on $\mathcal{F}_t$. Define $\bar{X} = [\bar{x}_1\ \bar{x}_2 \dots \bar{x}_t]^\top, X = [x_1\ x_2 \dots x_t]^\top, \varepsilon = [\varepsilon_1\ \varepsilon_2 \dots \varepsilon_t]^\top$ and let $y = \bar{X}\theta^* + \varepsilon$.*

Define for some $\rho > 0$ the quantities:

$$\begin{aligned}
\Omega &= \rho I + X^\top X \\
\hat{\theta} &= \Omega^{-1} X^\top Y \\
\beta(\delta) &= \sigma\sqrt{d\log\left(\frac{d\rho + t(\|\bar{X}\|_{2,\infty} + \|X - \bar{X}\|)}{\delta}\right)} + \|\theta^*\|_2 \left(\left\|X - \bar{X}\right\| + \sqrt{\rho}\right)
\end{aligned}$$

Then, with probability $1-\delta$, it holds that

$$\left\|\hat{\theta} - \theta^*\right\|_\Omega \leq \beta(\delta)$$

Proof.

1. We begin by decomposing the error into the the usual ridge regression error, plus a term that depends on the errors in variables.

$$\begin{aligned}
\hat{\theta} &= \Omega^{-1} X^\top y \\
&= \Omega^{-1} X^\top (\bar{X}\theta^* + \varepsilon) \\
&= \Omega^{-1} X^\top ((X + \bar{X} - X)\theta^* + \varepsilon) \\
&= \Omega^{-1} X^\top X\theta^* + \Omega^{-1} X^\top (\bar{X} - X)\theta^* + \Omega^{-1} X^\top \varepsilon \\
&= \Omega^{-1}(\rho I + X^\top X)\theta^* - \rho\Omega^{-1}\theta^* + \Omega^{-1} X^\top (\bar{X} - X)\theta^* + \Omega^{-1} X^\top \varepsilon \\
&= \theta^* - \rho\Omega^{-1}\theta^* + \Omega^{-1} X^\top (\bar{X} - X)\theta^* + \Omega^{-1} X^\top \varepsilon \\
\implies x^\top\hat{\theta} - x^\top\theta^* &= \left\langle x, X^\top\varepsilon\right\rangle_{\Omega^{-1}} + \left\langle x, (X^\top(\bar{X} - X) - \rho I)\theta^*\right\rangle_{\Omega^{-1}} \\
&\leq \|x\|_{\Omega^{-1}} \left(\|X^\top\varepsilon\|_{\Omega^{-1}} + \|\theta^*\|_2 \left\|\bar{X} - X\right\| + \rho\|\theta^*\|_{\Omega^{-1}}\right) \\
&\leq \|x\|_{\Omega^{-1}} \left(\|X^\top\varepsilon\|_{\Omega^{-1}} + \|\theta^*\|_2 \left\|\bar{X} - X\right\| + \sqrt{\rho}\|\theta^*\|_2\right)
\end{aligned}$$

The second inequality uses $\lambda_{\max}(\Omega^{-1}) \geq 1/\rho$, and the first inequality uses Cauchy-Schwarz and the following bound:

$$\begin{aligned} x^\top \Omega^{-1} X^\top (\bar{X} - X)\theta^* &\leq \left\|X\Omega^{-1}x\right\|_2 \left\|(\bar{X} - X)\theta^*\right\|_2 \\ &= \sqrt{\|X\Omega^{-1}x\|_2^2} \left\|(\bar{X} - X)\theta^*\right\|_2 \\ &\leq \sqrt{\|X\Omega^{-1}x\|_2^2 + \rho x^\top \Omega^{-1}\Omega^{-1}x} \left\|(\bar{X} - X)\theta^*\right\|_2 \\ &= \sqrt{x^\top \Omega^{-1}(X^\top X + \rho I)\Omega^{-1}x} \left\|(\bar{X} - X)\theta^*\right\|_2 \\ &= \sqrt{x^\top \Omega^{-1}x} \left\|(\bar{X} - X)\theta^*\right\|_2 \\ &= \|x\|_{\Omega^{-1}} \left\|(\bar{X} - X)\theta^*\right\|_2 \\ &= \|x\|_{\Omega^{-1}} \left\|\bar{X} - X\right\| \|\theta^*\|_2 \end{aligned}$$

2. Choosing $x = \Omega(\hat{\theta} - \theta^*)$, we have:

$$\left\|\hat{\theta} - \theta^*\right\|_\Omega^2 \leq \left\|\hat{\theta} - \theta^*\right\|_\Omega \left(\|X^\top \varepsilon\|_{\Omega^{-1}} + \|\theta^*\|_2 \left\|X - \bar{X}\right\| + \sqrt{\rho}\|\theta^*\|_2\right)$$

3. Theorem 1 of [6] gives that with probability $1 - \delta/2$,

$$\left\|X^\top \varepsilon\right\|_{\Omega^{-1}} \leq \sigma\sqrt{d \log\left(\frac{d\rho + t\|X\|_{2,\infty}}{\delta}\right)}$$

4. Finally, we need only bound $\|X\|_{2,\infty}$ in terms of $\|\bar{X}\|_{2,\infty}$:

$$\|X\|_{2,\infty} \leq \|\bar{X}\|_{2,\infty} + \|X - \bar{X}\|_{2,\infty} \leq \|\bar{X}\|_{2,\infty} + \|X - \bar{X}\|$$

where the first inequality uses the triangle inequality, and the second uses the generic norm inequality $\|A\|_{2,\infty} = \max_{\|u\|_2=1} \|Au\|_\infty \leq \max_{\|u\|_2=1} \|Au\|_2 = \|X\|$

C.2.3 Completing the proof of Lemma 7

Applying Proposition 8, with $\delta = 1/t^2, d = r+1, \rho = 1$, and using that $\|\bar{X}\|_{2,\infty} \leq \|\bar{Z}\|_{2,\infty} + 1 \leq B+1$, we immediately have that

$$\|\hat{\theta}_t - \theta^*\|_{\Omega_t} \leq \sigma\sqrt{2(r+1)\log\left(t\left(r + 1 + t(B + 1 + \|\hat{Z}_t - \bar{Z}_t\|)\right)\right)} + \|\theta^*\|_2 \left(\|\hat{Z}_t - \bar{Z}_t\| + 1\right)$$

On C_t^{latent}, which occurs with probability $1 - \frac{1}{t^8}$, we have the bound $\|\hat{Z}_t - \bar{Z}_t\| \leq 20\sigma\sqrt{\max(n,T)/n}$. Then via a union bound, with probability $1 - \frac{1}{t^2} - \frac{1}{t^8} \geq 1 - \frac{2}{t^2}$ it holds that

$$\begin{aligned}\|\hat{\theta}_t - \theta^*\|_{\Omega_t} &\le \sigma\sqrt{2(r+1)\log\left(t\left(r+1+t(B+1+20\sigma\sqrt{\max(n,T)/n})\right)\right)} \\ &\quad + (\|\lambda^*\| + |\tau^*|)(20\sigma\sqrt{\max(n,T)/n}+1) \\ &= \beta_t/2 = O(\sqrt{r\log(rt)})\end{aligned}$$

Further,

$$|\hat{\tau}_t - \tau^*| = |\langle\hat{\theta}_t - \theta^*, e_1\rangle| \le \|\hat{\theta}_t - \theta^*\|_{\Omega_t}\|e_1\|_{\Omega_t^{-1}} = \beta_t\hat{\sigma}_t/2.$$

This completes the proof.

C.3 Proof of Lemma 3

To begin, we have the following lemma to connect the $\|\cdot\|_{\bar{\Omega}}$ and $\|\cdot\|_{\Omega}$ norms.

Lemma 8. *For some $\bar{X}, X \in \mathbb{R}^{t\times r}, \rho > 0$, let $\bar{\Omega} = \rho I + \bar{X}^\top\bar{X}, \Omega = \rho I + X^\top X$, and $\Xi = X - \bar{X}$. Then for any vector $a \in \mathbb{R}^r$, it holds that*

$$\|a\|_{\bar{\Omega}} \le \|a\|_{\Omega}\left(1 + \rho^{-\frac{1}{2}}\|\Xi\|\right)$$

Proof. Expanding the definition of Ω, we have:

$$\begin{aligned}\|a\|_{\bar{\Omega}-\Omega}^2 &= a^\top(\bar{\Omega} - \Omega)a \\ &= a^\top\left(\Xi^\top X + X^\top\Xi + \Xi^\top\Xi\right)a \\ &\le 2\|a\|_2\|\Xi\|\|Xa\|_2 + \|a\|_2^2\left\|\Xi^\top\Xi\right\| \\ &\le \frac{2}{\sqrt{\rho}}\|a\|_\Omega^2\|\Xi\| + \frac{1}{\rho}\|a\|_\Omega^2\|\Xi\|^2\end{aligned}$$

where the first inequality follows from Cauchy-Schwarz and the triangle inequality, and the second inequality uses the following inequalities:

- $\|Xa\|_2^2 = a^\top X^\top Xa = a^\top(X^\top X + \rho I)a - \rho\|a\|_2^2 = \|a\|_\Omega^2 - \rho\|a\|_2^2 \le \|a\|_\Omega^2$
- $\|a\|_\Omega^2 \ge \sigma_{\mathrm{r}}(\Omega)\|a\|_2^2 \ge \rho\|a\|_2^2$

We then use this to bound the difference between $\|a\|_{\bar{\Omega}}$ and $\|a\|_{\Omega}$:

$$
\begin{aligned}
\|a\|_{\bar{\Omega}}^2 &= \|a\|_{\Omega}^2 + \|a\|_{\bar{\Omega}-\Omega}^2 \\
&\leq \|a\|_{\Omega}^2 + \frac{2}{\rho^{\frac{1}{2}}} \|a\|_{\Omega}^2 \|\Xi\| + \frac{1}{\rho} \|a\|_{\Omega}^2 \|\Xi\|^2 \\
&= \|a\|_{\Omega}^2 \left(1 + \frac{2}{\rho^{\frac{1}{2}}}\|\Xi\| + \frac{1}{\rho} \|\Xi\|^2\right) \\
&= \|a\|_{\Omega}^2 \left(1 + \frac{1}{\rho^{\frac{1}{2}}}\|\Xi\|\right)^2
\end{aligned}
$$

Lemma 8 also implies the same bound on difference between the inverse norms $\|\cdot\|_{\Omega^{-1}}$ and $\|\cdot\|_{\bar{\Omega}^{-1}}$; i.e.

$$
\|a\|_{\bar{\Omega}^{-1}} = \max_{\|u\|_{\bar{\Omega}} \leq 1} \langle a, u \rangle \leq \max_{\|u\|_{\Omega} \leq 1+\rho^{-1/2}\|\Xi\|} \langle a, u \rangle \leq (1 + \rho^{-1/2}\|\Xi\|)\|a\|_{\Omega^{-1}}.
$$

Then, applying Lemma 8 and the rotation invariance discussed in Appendix C.2.1, under C^{latent}, we have

$$
\|a\|_{\bar{\Omega}_t^{-1}} \leq (1 + \rho^{-1/2}\|\Xi\|)\|a\|_{\Omega_t^{-1}} \leq (1 + \|\hat{Z}_t - \bar{Z}_t\|)\|a\|_{\Omega_t^{-1}} \leq (1+\alpha)\|a\|_{\Omega_t^{-1}}.
$$

with $\rho \triangleq 1$. This completes the proof.

D Proofs of Inferential Results

D.1 Proof of proposition 5

Let $\bar{z}_t = a_t/B$, and note that this $\bar{z}_t$ is $\mathcal{F}_t$ -adapted and satisfies Assumption 1. Suppose that $\mathbb{P}(\hat{a} \neq a) \leq 1/2$ for some instance with parameters $\tau_0^*, \{\lambda_0^i\}_{i\in[n]}$. Now consider another instance, with parameters $\tau^* = -\tau_0^*$ and $\lambda^i = \lambda_0^i + 2\tau_0$. For any realization of the exogenous noise $\{\epsilon_t^i\}_{it}$, the rewards will be the same under both instances. By a coupling argument it must then hold that $\mathbb{P}(\hat{a} \neq a^*) \geq \frac{1}{2}$.

D.2 Proof of Proposition 1

Proof. By definition, the estimator τ^{SC} is

$$
\begin{aligned}
\tau^{\text{SC}} &:= \frac{1}{T}\sum_{t=1}^{T}\left(y_t^0 - \sum_{i=1}^{n} w_i y_t^i\right) \\
&= \tau^* + \frac{1}{T}\sum_{t=1}^{T}\left(\lambda^{*\top}\bar{z}_t - \sum_{i=1}^{n} w_i \lambda^{i\top}\bar{z}_t\right) + \underbrace{\frac{1}{T}\sum_{t=1}^{T}\left(\epsilon_t^0 - \sum_{i=1}^{n} w_i \epsilon_t^i\right)}_{R_1} \\
&= \tau^* + \left(\lambda^{*\top} - \sum_{i=1}^{n} w_i \lambda^{i\top}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\bar{z}_t\right) + R_1. \qquad (11)
\end{aligned}
$$

Let $Z = [\bar{z}_1^\top; \bar{z}_2^\top; \cdot; \bar{z}_{T_0}^\top] \in \mathbb{R}^{T_0 \times r}$. Let $E \in \mathbb{R}^{(n+1)\times T_0}$ be noise matrix with entries $E_{ij} = \epsilon^i_{j-T_0}$ for $i = 0, 1, \ldots, n, j = 1, 2, \ldots, T_0$. Let E_i be the i-th row of E.

Then, from $y_t^0 = \sum_{i=1}^n w_i y_t^i, t = -T_0 + 1, \ldots, -1, 0$, we have

$$Z\lambda^* + E_0 = Z\left(\sum_{i=1}^n w_i \lambda^i\right) + \sum_{i=1}^n w_i E_i$$

Let Z^{-1} be the pseudo-inverse of Z. We then have $Z^{-1}Z = I_r$ and

$$\lambda^* - \left(\sum_{i=1}^n w_i \lambda^i\right) = Z^{-1}\left(\sum_{i=1}^n w_i E_i - E_0\right). \tag{12}$$

Plugging Eq. (12) back into Eq. (11), we have

$$\tau^{\mathrm{SC}} - \tau^* = \left(Z^{-1}\left(\sum_{i=1}^n w_i E_i - E_0\right)\right)^\top \left(\frac{1}{T}\sum_{t=1}^T \bar{z}_t\right) + R_1 \tag{13}$$

$$= \underbrace{E_0 Z^{-1\top}\left(\frac{1}{T}\sum_{t=1}^T \bar{z}_t\right)}_{R_2} + \underbrace{w^\top E_{1:n} Z^{-1\top}\left(\frac{1}{T}\sum_{t=1}^T \bar{z}_t\right)}_{R_3} + R_1 \tag{14}$$

where $E_{1:n}$ is the sub-matrix of E consisting of rows indexed by $1, 2, \ldots, n$.

For R_1, note that w and ϵ_t^i for $t > T_0$ are independent, then

$$R_1 \overset{dist}{=} \mathcal{N}\left(0, \sigma^2\left(1 + \sum_{i=1}^n w_i^2\right)\frac{1}{T}\right) \tag{15}$$

Note that $\sum_{i=1}^n w_i^2 \le 1$ since $w_i \ge 0$ and $\sum_i w_i \le 1$. Then with probability $1 - \delta$,

$$|R_1| \lesssim \frac{\sigma}{\sqrt{T}}\sqrt{\log(1/\delta)}.$$

For R_2, let $a := Z^{-1\top}\left(\frac{1}{T}\sum_{t=1}^T \bar{z}_t\right)$. Note that

$$\|a\| \le \frac{\left\|\frac{1}{T}\sum_{t=1}^T \bar{z}_t\right\|}{\sigma_r(Z)} \le \frac{c_2}{\sqrt{c_1 T_0}}.$$

Since E_0 are i.i.d Gaussian and independent from a, with probability $1 - \delta$, we have

$$|R_2| \lesssim \frac{c_2\sigma}{c_1\sqrt{T_0}}\sqrt{\log(1/\delta)}.$$

For R_3, the difficulty is that E and w are not independent. Note that

$$|R_3| = |w^\top E_{1:n} a| \le \|E_{1:n}a\|_\infty = \max_{i\in[1:n]} |E_i^\top a|.$$

By the union bound and $E_i^\top a$ being Gaussian, with probability $1-\delta$, we have

$$|R_3| \lesssim \frac{c_2\sigma}{\sqrt{c_1 T_0}}\sqrt{\log(1/\delta)+\log(n)}.$$

This completes the proof. ■

D.3 Proof of Proposition 7

We will prove a generalized result of Proposition 7 below.

Proposition 9. *The followings hold.*

(a) With probability $1-O(\delta)$, if $M > T/2$, then

$$|\tau^{\mathrm{SCTS}} - \tau^*| \lesssim \frac{\sigma}{\sqrt{T}}\sqrt{\log(1/\delta)} + \frac{c_2\sigma}{c_1\sqrt{T_0}}\sqrt{\log(1/\delta)+\log(n)}.$$

Furthermore,

(b) When $\tau^* \geq 0$, with probability at most $\frac{2R(T)}{T\tau^*}$, $M \leq T/2$ (i.e., $\tau^{\mathrm{SCTS}} = 0$).

(c) When $\tau^* < 0$, with probability at least $1 - \frac{2R(T)}{T|\tau^*|}$, $M \leq T/2$ (i.e., $\tau^{\mathrm{SCTS}} = 0$).

Proof. In order to show (a), we follow the same analysis as the proof of Proposition 1, where the only term that needs to be re-analyzed is

$$R_1 := \frac{1}{M}\sum_{t=1}^{T} \mathbf{1}\{a_t = 1\}\left(\epsilon_t^0 - \sum_{i=1}^{n} w_i \epsilon_t^i\right)$$

Let $z_t := \mathbf{1}\{a_t = 1\}$. Then

$$R_1 = \frac{1}{M}\sum_{i=1}^{T} \epsilon_t^0 z_t - \frac{1}{M}\sum_{t=1}^{T}\sum_{i=1}^{n} w_i \epsilon_t^i z_t \tag{16}$$

Note that z_t is independent from the future ϵ_k^j for $k \geq t$. Hence we can use Azuma's inequality (a sub-Gaussian variant [59]) to obtain, with probability $1-\delta$,

$$\left|\sum_{t=1}^{T} \epsilon_t^0 z_t\right| \lesssim \sqrt{T}\log(1/\delta)\sigma. \tag{17}$$

Similar bounds can be obtained for $\sum_{t=1}^{T}\left(\sum_{i=1}^{n} w_i\epsilon_t^i\right) z_t$. This provides, with probability $1-O(\delta)$,

$$R_1 \lesssim \frac{\sqrt{T}}{M}\log(1/\delta)\sigma.$$

Using $M > \frac{T}{2}$, we finish the proof for (a).

Next, consider the case $\tau^* \geq 0$. The regret incurred for each instance is $R = (T-M)\tau^*$. By Markov inequality, we then have

$$\Pr\left((T-M)\tau^* \geq \frac{T\tau^*}{2}\right) \leq \frac{R(T)}{T\tau^*/2}.$$

Equivalently,

$$\Pr\left(M \leq \frac{T}{2}\right) \leq \frac{2R(T)}{T\tau^*}.$$

Hence, with probability at most $\frac{2R(T)}{T\tau^*}$, we have $M \leq T/2$, i.e., (b) holds.

Finally, consider the case $\tau^* < 0$. The regret incurred is $R = M|\tau^*|$. Then, by Markov inequality,

$$\Pr\left(M|\tau^*| > \frac{T}{2}|\tau^*|\right) \leq \frac{R(T)}{|\tau^*|T/2}.$$

This proves (c). ■

E Technical Lemmas

Theorem 6 (Matrix concentration). *Suppose that* $E_{ij} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$*, and let* $E \in \mathbb{R}^{n\times t}$*. Then, there exists some universal constant* c_3 *such that, with probability* $1 - \frac{1}{(n\vee t)^8}$ *, we have* $\|E\| \leq 4\sigma\sqrt{n \vee t} + c_3\sigma \log(n \vee t)$

For sufficiently large $n \vee t$ this can be further bounded as $\|E\| \leq 5\sigma\sqrt{n \vee t}$. See e.g. [30] Theorem 3.1.4 for reference.

Lemma 9. *Consider a sequence* $\{x_t\}$ *where* $x_t \in \mathbb{R}^r$ *and* $\|x_t\|_2 \leq B\ \forall t$*, and define* $\Omega_t = \rho I + \sum_{t=1}^{T} x_t x_t^\top$*. Then, it holds that*

$$\sum_{t=1}^{T}\left(1 \wedge \|x_t\|^2_{\Omega_{t-1}^{-1}}\right) \leq \sqrt{rT\log\left(\frac{r\rho + TB}{r}\right)}$$

See e.g. [48] for proof.