# Policy design in experiments with unknown interference*

Davide Viviano†

(Job Market Paper)

First Version: November, 2020;
This Version: November 24, 2021
Please click here for the most recent version

**Abstract**

This paper proposes an experimental design for estimation and inference on welfare-maximizing policies in the presence of spillover effects. I consider a setting where units are organized into a *finite* number of large clusters and interact in unknown ways within each cluster. As a first contribution, I introduce a single-wave experiment which, by carefully varying the randomization across pairs of clusters, estimates the marginal effect of a change in treatment probabilities, taking spillover effects into account. Using the marginal effect, I propose a practical test for policy optimality. The idea is that researchers should report the marginal effect and test for policy optimality: the marginal effect indicates the direction for a welfare improvement, and the test provides evidence on whether it is worth conducting additional experiments to estimate a welfare-improving treatment allocation. As a second contribution, I design a multiple-wave experiment to estimate treatment assignment rules and maximize welfare. I derive small-sample guarantees on the difference between the maximum attainable welfare and the welfare evaluated at the estimated policy (regret). A corollary of such guarantees is that the regret converges to zero linearly in the number of iterations and clusters. Simulations calibrated to existing experiments on information diffusion and cash-transfer programs show welfare improvements up to fifty percentage points.

*Keywords:* Experimental Design, Spillovers, Welfare Maximization, Causal Inference.
*JEL Codes:* C31, C54, C90.

*Previous versions can be found at https://arxiv.org/abs/2011.08174. I am especially grateful to my advisor Graham Elliott, James Fowler, Paul Niehaus, Yixiao Sun, and Kaspar Wüthrich for their continuous advice and support. I wish to thank Karun Adusumilli, Peter Aronow, Timothy Christensen, Tjeerd de Vries, Brian Karrer, Matt Goldman, Craig Mcintosh, Xinwei Ma, Karthik Muralidharan, Cyrus Samii, Fredrik Sävje, Pietro Spini, Ye Wang, and participants at the NYU Methods Workshop seminar, 2021 Young Economists Symposium, EEA-ESEM 2021 summer meeting, EGSC Conference, and several internal seminars at UCSD for helpful comments and discussion. All mistakes are my own.

†Department of Economics, UC San Diego. Correspondence: dviviano@ucsd.edu.

# 1 Introduction

One of the goals of a government or NGO is to estimate the welfare-maximizing policy. Network interference is often a challenge: treating an individual may also generate spillovers and affect the design of the policy. For instance, approximately 40% of experimental papers published in the "top-five" economic journals in 2020 mention spillover effects as a possible threat when estimating the effect of the program.[1] Researchers have become increasingly interested in experimental designs for choosing the treatment rule (policy) which maximizes welfare.[2] But when it comes to experiments on networks, standard approaches are geared towards the estimation of treatment effects. Estimation of treatment effects, on its own, is not sufficient for welfare-maximization.[3] For example, when assigning cash transfers, these may have the largest direct effect when given to people living in remote areas but generate the smallest spillovers. This trade-off has significant policy implications when treating each individual is costly or infeasible.

This paper studies experimental designs in the presence of network interference when the goal is welfare maximization. The main difficulty in these settings is that interactions can be challenging to measure, and collecting network information can be very costly as it may require enumerating all individuals and their connections in the population.[4] We, therefore, focus on a setting with limited information on the network. This is formalized by assuming that units are organized into a *finite* number of large clusters, such as schools, districts, or regions, and interact through an unobserved network (and in unknown ways) within each cluster. In the cash-transfer program, we may expect that treatments generate spillovers to those living in the same or nearby villages[5], but spillovers are negligible between individuals in different regions. We propose the first experimental design to analyze and estimate *welfare-maximizing* treatment rules in the presence of unobserved spillovers on networks.

We make two main contributions. As a first contribution, we introduce a design where researchers randomize treatments and collect outcomes once (*single-wave experiment*) to (i) test whether one or more treatment allocation rules, such as the one currently implemented by the policymaker, maximize welfare; (ii) estimate how we can improve welfare with a (small) change to allocation rules. The experiment is based on a simple idea. With a

[1]This is based on the author's calculation. Top-five economic journals are *American Economic Review*, *Econometrica*, *Journal of Political Economy*, *Quarterly Journal of Economics*, *Review of Economic Studies*.

[2]See Bubeck et al. (2012) and Kasy and Sautmann (2019) for a discussion.

[3]Examples of treatment effects are the direct effects of the treatment and the overall effect, i.e., the effect if we treat all individuals, compared to treating *none*. For welfare maximization, none of these estimands is sufficient. The direct effect ignores spillovers, while the optimal rule may only treat some but not individuals either because of treatment costs or because of decreasing marginal returns from neighbors' treatments.

[4]See Breza et al. (2020) for a discussion on the cost associated with collecting network information.

[5]For instance, Egger et al. (2019) document spillovers from cash-transfers between nearby villages.

small number of clusters, we do not have enough information to precisely estimate the welfare-maximizing treatment rule. However, if we take *two* clusters and assign treatments in each cluster independently with slightly different (locally perturbated) probabilities, we can estimate the marginal effect of a change in the treatment assignment rule (which we will refer to as marginal *policy* effect, MPE). For instance, in the previous example, the MPE defines the marginal effect of treating more people in remote areas, taking spillover effects into account.[6] Using the MPE, we introduce a practical test for whether there exists a welfare-improving treatment allocation rule. As this paper suggests, researchers should report estimates of the MPE and test for welfare-maximizing policies. The MPE indicates the *direction* for a welfare improvement, and the test provides evidence on whether it is worth conducting additional experiments to estimate a welfare-improving treatment allocation.

Specifically, the experiment *pairs* clusters and randomizes treatments independently within clusters, with local perturbations to treatment probabilities within each pair. The difference in treatment probabilities balances the bias and variance of a difference-in-differences estimator. We show that the estimator for each pair converges to the marginal effect and derive properties for inference with finitely many clusters. The experiment separately estimates the direct and spillover effects, which are of independent interest. These are the effect on the recipients and the marginal effect of increasing the neighbors' treatment probability.[7]

As a second contribution, we offer an adaptive (i.e., *multiple-wave*) experiment to estimate welfare-maximizing allocation rules. Our goal here is to adaptively randomize treatments to estimate the welfare-maximizing policy while also improving participants' welfare.[8] We propose an experiment which guarantees tight small-sample upper bounds for *both* the (i) out-of-sample regret, i.e., the difference between the maximum attainable welfare and the welfare evaluated at the estimated policy deployed on a new population, and the (ii) in-sample regret, i.e., the regret of the experimental participants. The experiment groups clusters into pairs, using as many pairs as the number of iterations (or more); every iteration, it randomizes treatments in a cluster and perturbs the treatment probability within each pair; finally, it updates policies sequentially, using the information on the marginal effects from a different pair via gradient descent. We illustrate the existence of a bias in adaptive experiments with repeated sampling and develop a novel algorithm to avoid such a bias.

From a theoretical perspective, a corollary of our sequential experiment's small-sample

[6]This is the derivative of welfare with respect to the policy's parameters, different from what known in observational studies as marginal treatment effect (Carneiro et al., 2010), which instead depends on the selection into treatment mechanism.

[7]Also, between-clusters local perturbations accommodate settings where policymakers cannot allow much variation in how treatments are assigned between different clusters because of exogenous constraints.

[8]Improving participants' welfare is desirable for large-scale experiments, common on online platforms (Karrer et al., 2021), and of increasing interest in development studies (Muralidharan and Niehaus, 2017).

guarantees is that the out-of-sample regret converges linearly in the number of clusters and iterations, and similarly, the in-sample regret, up to a logarithmic factor. We note that there are no regret guarantees tailored to unobserved interference. Existing results for *i.i.d.* data, treating clusters as sampled observations, would instead imply slower (sub-linear) convergence rates in the number of clusters.[9] We achieve this fast rate by (a) exploiting *within*-cluster variation in treatment assignments and *between* clusters' local perturbations; (b) deriving concentration results within each cluster as the cluster's size increases; (c) assuming and leveraging decreasing marginal effects of increasing neighbors' treatment probability.

We illustrate the numerical properties of the method with calibrated experiments that use data from an information diffusion experiment (Cai et al., 2015) and a cash-transfer program (Alatas et al., 2012, 2016). We show that our test can, in expectation, lead to welfare improvements up to fifty percentage points if, upon rejections of the null hypothesis that increasing treatment probabilities does not improve welfare, policy-makers increase treatment probabilities by five percent. When designing an adaptive experiment, the proposed method substantially improves both out-of-sample and in-sample regret, even with few iterations.

Throughout the text, we assume that the maximum degree grows at an appropriate slower rate than the cluster size; covariates and potential outcomes are identically distributed between clusters; treatment effects do not carry over in time. In the Appendix, we relax these assumptions and study three extensions: (a) experimental design with a global interference mechanism; (b) matching clusters with covariates drawn from cluster-specific distributions, and introduce matching via distributional embeddings; (c) experimental design with *dynamic* treatment effects, and propose a novel experimental design in this setting.

Our paper adds to both the literature on single-wave and multiple-wave experiments. In the context of *single-wave* (or two-wave) experiments, existing network experiments include clustered experiments (Eckles et al., 2017; Ugander et al., 2013; Karrer et al., 2021) and saturation designs (Baird et al., 2018; Pouget-Abadie, 2018). References with observed networks include Basse and Airoldi (2018b), Jagadeesan et al. (2020), Viviano (2020) among others.[10] See also Bai (2019); Tabord-Meehan (2018) with *i.i.d.* data. These authors study

[9] In a distribution-free setting, Bubeck et al. (2011) provides distribution-free lower bounds of order $1/\sqrt{n}$ for the stochastic bandit problem. Flaxman et al. (2004) considers a stochastic optimization with a single noisy observation. The authors provide sub-linear rates in the number of iterations (and observations). Shamir (2013) provides sublinear lower bounds of order $1/\sqrt{n}$ for stochastic optimization procedure with strongly-convex functions (allowed here). For offline optimization, Kitagawa and Tetenov (2018); Athey and Wager (2021) establish distribution-free lower bounds of order $1/\sqrt{n}$ in the presence of a single-wave experiment. We note that Wager and Xu (2021) provide linear rates in the number of *iterations* but assume infinite asymptotics in the number of observations and leverage an explicit model for market interactions. Here, we consider a finite number of clusters, and we do not impose assumptions on the interference mechanism. Kasy and Sautmann (2019) instead provide controls of either (but not both) notions of regret in a finite sample.

[10] For the analysis on the bias of average treatment effect estimators with interference see also Basse and

experimental designs for inference on treatment effects only, but not the problem of inference on welfare-maximizing policies. Different from all the above references, we propose a design that allows us to identify the marginal effect under interference (as well as treatment effects). We introduce the first test and design for inference on policy optimality under unobserved interference. This difference motivates us to introduce a novel design consisting of local perturbations between clusters. The idea of using the information on marginal effects connects to the literature on optimal taxation (Saez, 2001; Chetty, 2009), which, differently, considers observational studies with independent units.

In the presence of *multiple-wave* experiments, we introduce a framework for adaptive experimentation with unknown interference. We connect to the literature on adaptive exploration (Bubeck et al., 2012; Russo et al., 2017; Kasy and Sautmann, 2019, among others), and the one on stochastic optimization (Flaxman et al., 2004; Kleinberg, 2005; Shamir, 2013, among others). These references assume *i.i.d.* sampling, with no network interference. Here, we leverage between-clusters perturbations and within-cluster concentration to obtain fast rates of the in-sample and out-of-sample regret. In related work, Wager and Xu (2021) have studied estimation of prices via local experimentation in the different context of a single market, with asymptotically independent agents. They assume infinitely many individuals and an explicit model for market prices. As noted by the authors, the structural assumptions imposed in the above reference do not allow for spillovers on a network (i.e., individuals may depend arbitrarily on neighbors' assignments). Our setting differs due to the presence of network spillovers and the fact that individuals are organized into finitely many independent components (clusters), where such spillovers are unobserved. These differences motivate (i) our design mechanism, which exploits two-level randomization at the cluster and individual level instead of individual-level randomization, (ii) pairing and perturbations between clusters. From a theoretical perspective, network dependence and repeated sampling induce novel challenges for an adaptive experiment which we address in this paper.

We relate to the literature on inference on treatment effects under interference and draw from Hudgens and Halloran (2008) for definitions of potential outcomes. Differently from this paper, this literature considers existing (quasi)experiments and does not study experimental design and welfare-maximization. Aronow and Samii (2017); Manski (2013); Leung (2020); Ogburn et al. (2017); Goldsmith-Pinkham and Imbens (2013); Li and Wager (2020) assume an observed network, while Vazquez-Bare (2017), Hudgens and Halloran (2008), Ibragimov and Müller (2010) consider clusters among others. Sävje et al. (2021) study inference of the direct effect of treatment only. Our focus on policy optimality and experimental design differs

Feller (2018), Johari et al. (2020), Basse and Airoldi (2018a), and Imai et al. (2009) when matching with different-sized clusters for overall average treatment effects.

from the above references. We show that inference on policy-optimality requires information on the MPE, which, we demonstrate, can be estimated with a clusters pair.

More broadly, we connect to the literature on statistical treatment rules, on estimation Manski (2004); Kitagawa and Tetenov (2018); Athey and Wager (2021); Bhattacharya and Dupas (2012); Stoye (2009); Mbakop and Tabord-Meehan (2021); Kitagawa and Wang (2021); Sasaki and Ura (2020); Viviano (2019), among others, and inference Andrews et al. (2019); Rai (2018); Armstrong and Shen (2015);[11] see also the literature on classification (Elliott and Lieli, 2013). This literature considers an existing experiment and does not study experimental design. Here, instead, we leverage an adaptive procedure to maximize out-of-sample and participants' welfare. Also, this literature has not studied policy design with unobserved interference. We broadly relate also to the literature on targeting on networks (see for a review Bloch et al., 2019), which mostly focus on model-based approaches with a single observed network – different from here where we leverage clusters' variations – and the one on inference with externalities (e.g., Bhattacharya et al., 2013, which study inference on welfare associated to a treatment choice, but do not study inference on policy optimality and policy estimation). These references also do not study experimental designs.

The rest of the paper is organized as follows. Section 2 introduces the setup, and an illustration of the method. Section 3 studies the single-wave experiment. Section 4 presents the adaptive experiment. Section 5 collects the numerical experiments. Section 6 presents dynamic treatments. Section 7 gives our conclusions. Appendices A, B present extensions.

# 2 Setup and method's overview

This section introduces conditions, estimands, and a brief overview of the method.

We consider a setting with $K$ clusters, where $K$ is an even number. We assume that each cluster has $N$ individuals, while our framework directly extends to clusters of different but proportional sizes. Observables and unobservables are jointly independent between clusters but not necessarily within clusters. Each cluster $k$ is associated with a vector of outcomes, covariates, treatment assignments, and an adjacency matrix which is different for each cluster. These are $Y_{i,t}^{(k)} \in \mathcal{Y}, D_{i,t}^{(k)} \in \{0,1\}, X_i^{(k)} \in \mathcal{X} \subseteq \mathbb{R}^p, A^{(k)} \in \mathcal{A}$, respectively. Here, $(Y_{i,t}^{(k)}, D_{i,t}^{(k)})$ denote the outcome and treatment assignment of individual $i$ at time $t$ in cluster $k$, respectively, $X_i^{(k)}$ are time-invariant (baseline) covariates, and $A^{(k)}$ is a cluster-specific adjacency matrix. Each period $t$, researchers only observe a random subsample

$$\left(Y_{i,t}^{(k)}, X_i^{(k)}, D_{i,t}^{(k)}\right)_{i=1}^{n}, \quad n = \lambda N, \quad \lambda \in (0,1],$$

[11]See also Kato and Kaneko (2020); Hadad et al. (2019); Imai and Li (2019); Hirano and Porter (2020), which do not allow for testing for policy optimality, but construct confidence bands for the welfare.

where $n$ defines the sample size of observations from each cluster and proportional to the cluster size.[12] There are $T$ periods in total. While units sampled each period may or may not be the same, with abuse of notation, we index sampled units $i \in \{1, \cdots, n\}$. Whenever we provide asymptotic analyses, we let $N$ grow[13] and $K$ be fixed.

## 2.1 Setup: covariates, network and potential outcomes

Next, we introduce conditions on the covariates, network, and potential outcomes. These conditions guarantee that Lemma 2.1 (in Section 2.2) holds; practitioners may skip this subsection and directly refer to Section 2.2 for their implications, keeping in mind the covariates' distribution in Equation (1). We now discuss the network and covariates. We assume that individuals can form a link with a subset of individuals in each cluster. Formally, in each cluster, nodes are spaced under some latent space (Lubold et al., 2020) and can interact with at most the $\gamma_N^{1/2}$ closest nodes under the latent space. We say $1\{i_k \leftrightarrow j_k\} = 1$ if individual $i$ can interact with $j$ in cluster $k$ and zero otherwise. Conditional on the indicators $1\{i_k \leftrightarrow j_k\}$,

$$(X_i^{(k)}, U_i^{(k)}) \sim_{i.i.d.} F_{U|X} F_X, \quad A_{i,j}^{(k)} = l\Big(X_i^{(k)}, X_j^{(k)}, U_i^{(k)}, U_j^{(k)}\Big) 1\{i_k \leftrightarrow j_k\} \tag{1}$$

for an arbitrary and unknown function $l(\cdot)$ and unobservables $U_i^{(k)}$. Equation (1) states that whether two individuals interact depends on: (i) whether they are close enough within a certain latent space (captured by $1\{i_k \leftrightarrow j_k\}$); (ii) their covariates and unobserved individual heterogeneity (i.e., $X_i, U_i$) which capture homophily. Equation (1) also states that covariates are $i.i.d.$ unconditionally on $A^{(k)}$, but not necessarily conditionally. Figure 1 provides an illustration. Here, we condition on the indicators $1\{i_k \leftrightarrow j_k\}$ (which can differ across clusters) to control the network's maximum degree but not on the network $A^{(k)}$. Equivalently, we can interpret such indicators as exogenously drawn from some arbitrary distribution.[14]

**Assumption 2.1** (Network)**.** For $i \in \{1, \cdots, N\}, k \in \{1, \cdots, K\}$, let (i) Equation (1) hold given the indicators $1\{i_k \leftrightarrow j_k\}$, for some unknown $l(\cdot)$; (ii) $\sum_{j=1}^n 1\{i_k \leftrightarrow j_k\} = \gamma_N^{1/2}$.

Assumption 2.1 states the following: before being born, each individual may interact with $\gamma_N^{1/2}$ many other individuals. After the birth, the individual's gender, income, parental status determine her type and the distribution of her and her potential connections' edges.[15] Here

[12]Our framework directly extends when $n = g(N)$ for a generic monotonic function $g(\cdot)$, and when we sample a different but proportional number of individuals from each cluster.

[13]Here, we consider a sequence of data-generating processes.

[14]Formally, $\mathcal{I}_k \sim \mathcal{P}_k, \quad (X_i^{(k)}, U_i^{(k)})|\mathcal{I}_k \sim_{i.i.d.} F_{U|X}F_X, \quad A_{i,j}^{(k)}|\mathcal{I}_k = l\Big(X_i^{(k)}, X_j^{(k)}, U_i^{(k)}, U_j^{(k)}\Big)1\{i_k \leftrightarrow j_k\}$, where $\mathcal{I}_k$ is the matrix of such indicators in cluster $k$ and $\mathcal{P}_k$ is a cluster-specific distribution left unspecified.

[15]Networks formed from pairwise interactions have also been discussed in Jackson and Wolinsky (1996),

$\gamma_N^{1/2}$ captures the degree of dependence. Whenever $\gamma_N^{1/2}$ equals $N$, we impose no restriction on the individual's number of connections. In Theorem 3.1 the maximum degree can be equal to the cluster's size, while subsequent results require stricter restrictions.[16]

We now discuss potential outcomes. Under interference, outcomes depend on individuals' covariates, unobservables and neighbors' treatments. That is, $Y_{i,t}^{(k)}(\mathbf{d}_1^{(k)}, \cdots, \mathbf{d}_t^{(k)})$ denotes the potential outcome of individual $i$ at time $t$. Here $\mathbf{d}_s^{(k)} \in \{0,1\}^N$ denotes the treatment assignments at time $s$ of all individuals in cluster $k$. The following condition is imposed.

**Assumption 2.2** (Potential outcomes)**.** Suppose that for any $i, t, k, \mathbf{d}_s^{(k)} \in \{0,1\}^N, s \leq t$

$$Y_{i,t}^{(k)}(\mathbf{d}_1^{(k)}, \cdots, \mathbf{d}_t^{(k)}) = r\Big(\mathbf{d}_{i,t}^{(k)}, \mathbf{d}_{j:A_{i,j}^{(k)}>0,t}^{(k)}, X_i^{(k)}, X_{j:A_{i,j}^{(k)}>0}^{(k)}, U_i, U_{j:A_{i,j}^{(k)}>0}, A_{i,\cdot}^{(k)}, \nu_{i,t}^{(k)}\Big) + \tau_k + \alpha_t$$

for some unknown function $r(\cdot)$, stationary (but possibly serially dependent) unobservables $\nu_{i,\cdot}^{(k)} | X^{(k)}, U^{(k)} \sim_{i.i.d.} P_\nu$, fixed effects $\tau_k, \alpha_t$.

Assumption 2.2 imposes three conditions. First, treatment effects do not carry-over in time. Second, potential outcomes are stationary up to separable fixed effects (unobservables can depend over time).[17] Third, potential outcomes can depend on neighbors' assignments, neighbors' covariates, and neighbors' identities, with arbitrary and unknown heterogeneity in spillover effects. Note that $A_{i,\cdot}^{(k)}$ denotes the set of connections of individual $i$ in cluster $k$, with $\{j : A_{i,j}^{(k)} > 0\}$ denoting those individuals with some connection to $i$. Assumption 2.2 imposes no condition on the dependence on neighbors' assignments but assumes that cluster fixed effects do not depend on treatment assignments, which is important for identification.

**Remark 1** (Extensions)**.** In Appendix A and Appendix B we present several extensions. In Section A.1 individuals also depend on a (latent) factor that captures general equilibrium effects, and in Section A.3, they also depend on the past assignments. Appendix B.3 presents non-separable fixed effects. Appendix B.5 studies staggered adoption of the treatment. □

## 2.2 Policy and welfare maximization

The goal of this paper is to estimate a policy (treatment assignment rule) that maximizes welfare. We focus on a parametric class of policies, indexed by some parameter $\beta$. Formally,

---

Li and Wager (2020), Leung (2019). Here, we impose such restrictions to obtain easy-to-interpret conditions on the degree, once we average over *unobserved* neighbors' assignments. Assumption 2.1 is sufficient but not necessary.

[16] Assumption 2.1 would not be required if we were to observe neighbors' assignments as in Viviano (2019).

[17] Such condition is often implicitly imposed in studies on experimental design (Kasy and Sautmann, 2019). For a discussion on the no-carry-over assumption, see Athey and Imbens (2018). We relax it in Section A.3.

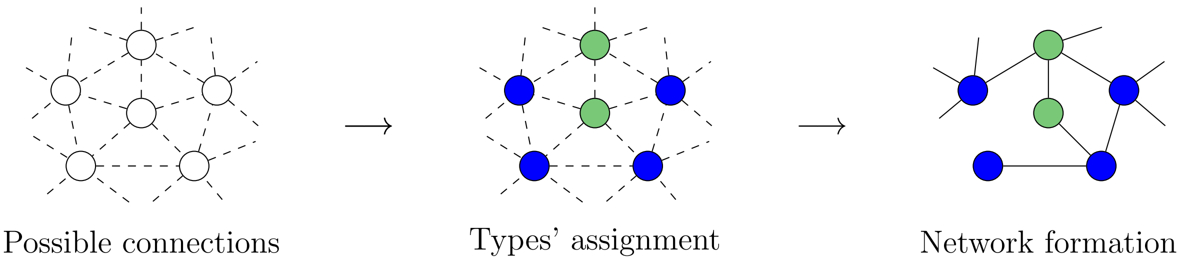

Figure 1: Example of the network formation model, with $\gamma_N = 5$. Individuals' are assigned different types which may or may not be observed by the researcher (corresponding to different colors). Individuals interact based on their types and form links among the possible connections. The possible connections and the realized adjacency matrix remain unobserved.

a policy

$$\pi(\cdot;\beta) : \mathcal{X} \mapsto [0,1], \quad \beta \in \mathcal{B},$$

is a map that prescribes the individual treatment probability based on covariates. Here $\mathcal{B}$ is a compact parameter space, and $\pi(x,\beta)$ is twice differentiable in $\beta$. The experiment assigns treatments independently based on $\pi(\cdot)$, and time/cluster-specific parameters $\beta_{k,t}$.

**Assumption 2.3** (Treatment assignments in the experiment)**.** For given parameters $\beta_{k,t}$

$$D_{i,t}^{(k)} | X^{(k)}, \beta_{k,t} \sim_{i.i.d.} \text{Bern}\Big(\pi(X_i^{(k)};\beta_{k,t})\Big),$$

which, for short of notation, we refer to as $D_{i,t}^{(k)} | X^{(k)}, \beta_{k,t} \sim \pi(X_i^{(k)}, \beta_{k,t})$.

Assumption 2.3 defines a treatment rule in experiments. Treatments are assigned independently based on covariates and time and cluster-specific parameters $\beta_{k,t}$ (whose choice is discussed in the next sections). The assignment in Assumption 2.3 is easy to implement: it can be implemented in an online fashion and does not require network information, which justifies its choice; also, it generalizes assignments in saturation designs studied for inference on treatment effects (Baird et al., 2018). Our goal is to estimate the welfare-maximizing $\beta$ (see Remark 2).[18] Our framework extends to continuous treatments, omitted for brevity.[19]

An example of assignment rule is treating individuals with equal probability (Akbarpour et al., 2018), i.e., $\pi(\cdot;\beta) = \beta \in [0,1]$. We can also *target* treatments, i.e., $\pi(x;\beta) = \beta_x$, indicating the treatment probability for $X_i^{(k)} = x$ (with $\mathcal{X}$ discrete).

Throughout our discussion, whenever we write $\pi(\cdot;\beta)$, omitting the subscripts $(k,t)$, we refer to a generic exogenous (i.e., not data dependent) vector of parameters $\beta$. We define $\mathbb{E}_\beta[\cdot]$ the expectation taken over the distribution of treatments assigned according to $\pi(\cdot;\beta)$.

[18] In Theorem 4.4 we show that the optimum obtained under Assumption 2.3 is asymptotically equivalent to the one with arbitrary dependent assignments under additional conditions on spillovers and costs.

[19] All our results hold for $D_{i,t}^{(k)} | X_i^{(k)}, \beta_{k,t} = \pi(X_i^{(k)}, \beta_{k,t})$, where $\pi(\cdot;\beta)$ is smooth in $\beta$.

**Lemma 2.1** (Outcomes)**.** *Under Assumption 2.1, 2.2, under an assignment in Assumption 2.3 with exogenous (i.e., not data-dependent)* $\beta_{k,t}$ *the following holds:*

$$Y_{i,t}^{(k)} = y\Big(X_i^{(k)}, \beta_{k,t}\Big) + \varepsilon_{i,t}^{(k)} + \alpha_t + \tau_k, \quad \mathbb{E}_{\beta_{k,t}}\Big[\varepsilon_{i,t}^{(k)}|X_i^{(k)}\Big] = 0, \tag{2}$$

*for some function* $y(\cdot)$ *unknown to the researcher. In addition, for some unknown* $m(\cdot)$, $\mathbb{E}_{\beta_{k,t}}\Big[Y_{i,t}^{(k)}|D_{i,t}^{(k)} = d, X_i^{(k)} = x\Big] = m(d, x, \beta_{k,t}) + \alpha_t + \tau_k$.

The proof is in Appendix C.2.3. Equation (2) states that the outcome depends on two components. The first is the conditional expectation given the individual covariates, and the parameter $\beta_{k,t}$, *unconditional* on covariates, adjacency matrix, individual, and neighbors' assignments. We can interpret the function $y(\cdot)$ (and $m(\cdot)$) as predictive, which depends on observables only. The dependence with $\beta_{k,t}$ captures spillover effects, since treatments' distribution depends on $\beta_{k,t}$, while we average over neighbors' treatments and covariates. The second component $\varepsilon_{i,t}$ are unobservables that also depend on the neighbors' assignments and covariates. As shown in Appendix C.2.3, under the above conditions, such unobservables only depend on $\gamma_N$ many others, where $\gamma_N^{1/2}$ is the maximum degree of the network (see Assumption 2.1). Also, note that Lemma 2.1 assumes that $\beta_{k,t}$ is exogenous. We guarantee exogeneity with a careful choice of the experimental design discussed in subsequent sections.

**Example 2.1.** Let $D_{i,t}^{(k)} \sim_{i.i.d.} \text{Bern}(\beta)$ be exogenous, $\mathcal{N}_i = \{j : A_{i,j} = 1\}$, $A_{i,j} \in \{0, 1\}$, and

$$Y_{i,t} = \alpha_t + D_{i,t}\phi_1 + \frac{\sum_{j\in\mathcal{N}_i} D_{j,t}^{(k)}}{|\mathcal{N}_i|}\phi_2 - \Big(\frac{\sum_{j\in\mathcal{N}_i} D_{j,t}}{|\mathcal{N}_i|}\Big)^2\phi_3 + \nu_{i,t}, \quad \mathbb{E}[\nu_{i,t}] = 0. \tag{3}$$

Equation (3) states that outcomes depend on the individual treatment, and the percentage of treated neighbors. With some algebra, taking expectations, $Y_{i,t} = \alpha_t + \beta\phi_1 + \beta\phi_2 - \beta\phi_3\kappa - \beta^2\phi_3(1-\kappa) + \varepsilon_{i,t}$ where $\varepsilon_{i,t}$ also depends on neighbors' assignments and $\kappa = \mathbb{E}[1/|\mathcal{N}_i|]$. □

Lemma 2.1 provides the basis for the definition of (utilitarian) welfare.

**Definition 2.1** (Welfare)**.** For treatments as assigned in Assumption 2.3 with exogenous $\beta$ parameter, let welfare be

$$W(\beta) = \int y(x, \beta) dF_X(x). \tag{4}$$

We define welfare as the expected outcome, had treatments been assigned with policy $\pi(\cdot, \beta)$. The expectation is taken over treatment assignments, covariates, and adjacency matrix. We interpret $y(x, \beta)$ the outcome *net of costs*[20], and incorporate the costs in the

[20]This is standard in the literature (Kitagawa and Tetenov, 2018). However, some applications may not have explicit definitions of costs. For these cases, one possible choice of the cost is the opportunity cost, had the treatment being assigned to a population with no externalities. See Section 4.2 for a discussion.

outcome function.[21] We define the welfare-maximizing policy and the marginal effect[22] as

$$\beta^* \in \arg\sup_{\beta \in \mathcal{B}} W(\beta), \quad V(\beta) = \frac{\partial W(\beta)}{\partial \beta}. \tag{5}$$

The marginal effect defines the derivative of the welfare with respect to the vector of parameters $\beta$. A useful insight is that, under mild regularity conditions, we can write

$$V(\beta) = \int \Big[ \underbrace{\pi(x;\beta)\frac{\partial m(1,x,\beta)}{\partial \beta} + (1-\pi(x;\beta))\frac{\partial m(0,x,\beta)}{\partial \beta}}_{(S)} + \underbrace{\frac{\partial \pi(x;\beta)}{\partial \beta}(m(1,x,\beta) - m(0,x,\beta))}_{(D)} \Big] dF_X(x). \tag{6}$$

The marginal effect depends on the direct effect of the treatment, (D), weighted by the marginal probability of being selected into treatment; the marginal spillover effect (S) denotes the marginal effect of increasing the neighbors' treatment probability. This decomposition follows in the same spirit of the decomposition for the total treatment effects as direct and indirect effects in Hudgens and Halloran (2008) and will be useful for our experiment.[23]

**Remark 2** (Assumption 2.3 and unconstrained optimum)**.** In Theorem 4.4, we compare $W(\beta^*)$ to the unconstrained optimum $W_N^*$, where treatments can be assigned arbitrary. We show, under additional conditions, that $W(\beta^*) - W_N^* \to 0$, as $N \to \infty$, whenever the treatment costs are the opportunity costs of an intervention with no spillovers. In a cash-transfer program these are the opportunity costs had the cash transfers be given to individuals spread out on a large area instead of individuals in nearby villages. □

## 2.3 Method's overview and example

We now give an overview and example. Consider a policy-maker who must allocate cash-transfers to *half* of the population. Let $X_i \in \{0,1\}$, equal to one for households farer from the district's center than the median household and zero otherwise, with $P(X_i = 1) = 1/2$. Due to the constraint, the policy-maker assigns treatments $D_{i,t}|X_i = x \sim \text{Bern}(\pi(x,\beta))$, where $\pi(x,\beta) = x\beta + (1-x)(1-\beta)$ is the treatment probability for $x \in \{0,1\}$.[24] Different treatment probabilities for people in remote areas produce different welfare effects, and assigning all treatments to individuals in remote areas is sub-optimal. This is illustrated in Figure 2, where we report $W(\beta)$, calibrated to data from Alatas et al. (2012, 2016).[25]

[21] Namely, we can parametrize $Y_{i,t} = Y_{i,t} - c(X_{i,t};\beta)$, for a cost function $c(\cdot;\beta)$.

[22] Here, we are assuming differentiability. See Assumption 3.1 for explicit conditions.

[23] We note that in recent work, Hu et al. (2021) have proposed targeting as causal estimand the average indirect effect, which is different from (S) for heterogenous assignments.

[24] This follows from the budget constraint where $\beta_1 P(X_i = 1) + \beta_2 P(X_i = 0) = 1/2$, where $\beta_1 = \beta, \beta_2 = 1 - \beta$ are the treatment probabilities for people in remote and closer areas from the center, respectively.

[25] See Appendix D.3 for details.

### 2.3.1 Single-wave experiment: hypothesis testing

First, we would like to test whether a certain (baseline) intervention $\beta$, such as the one currently implemented by the government or NGO, is welfare-optimal. That is, we test

$$H_0 : W(\beta) = W(\beta^*).$$

Its rejection is informative on whether a (small) change to $\beta$ improves welfare. We use the marginal effect for:

(a) *Hypothesis testing*: assuming that $\beta^*$ is an interior point, if $\frac{\partial W(\beta)}{\partial \beta} \neq 0$, then $H_0$ is false;

(b) *Policy update*: estimate the welfare-improving *direction* (increase or decrease $\beta$).

Here, (a) tests whether the line's slope in Figure 2 is zero. Similarly, a rejection of one-sided hypotheses $\frac{\partial W(\beta)}{\partial \beta} \leq 0$ suggests to increase $\beta$ (treat more people in remote areas).

We proceed to construct estimators of the marginal effect. We start from Equation (6). The direct effect (D) can be identified from a single network, taking the difference between treated and untreated outcomes. However, the spillover effect (S) cannot be identified from a single network when unobserved. We instead exploit *two* clusters' variation.

We take two clusters such as two regions. We collect *baseline* ($t = 0$) outcomes and covariates; we then randomize treatments with slightly different probabilities between the regions. In the first region, we treat individuals in remote areas ($X_i = 1$) with probability $\beta + \eta_n$. Here, $\eta_n$ is a small deterministic number (local perturbation). The remaining individuals are treated with probability $1-\beta-\eta_n$. In the second region, we treat individuals in remote area with probability $\beta - \eta_n$, and the remaining ones with probability $1 - \beta + \eta_n$.

As shown in Figure 2, we can estimate welfare for two different but similar treatment probabilities; the line's slope between the points is approximately equal to the marginal effect. That is, for a suitable choice of $\eta_n$ (see Theorem 3.1), a consistent marginal effect's estimator is

$$\widehat{V}_{(k,k+1)}(\beta) = \frac{1}{2\eta_n}\Big[\bar{Y}_1^{(k)} - \bar{Y}_0^{(k)}\Big] - \frac{1}{2\eta_n}\Big[\bar{Y}_1^{(k+1)} - \bar{Y}_0^{(k+1)}\Big], \tag{7}$$

where $\bar{Y}_t^{(h)}$ is the outcomes' sample average in cluster $h$ at time $t$, $Y_{i,0}$ is the baseline outcome with no experiment in place yet, and $(k, k+1)$ index the two clusters. The above estimator is a difference-in-differences; we subtract baseline outcomes due to fixed effects.

We can then leverage the marginal effect for hypothesis testing. We take $K = 2G, G > 1$ clusters and 1) we match clusters in $G$ pairs; 2) within each pair $g$, we obtain an estimate $\widehat{V}_g(\beta)$ for pair $g$; 3) we construct a (scale-invariant) test statistics using $G$ pairs, which does not depend on the estimator's variance.[26] As we discuss in Remark 5, pairing clusters guarantees finite clusters asymptotics. Formal details of the algorithm are in Section 3.

Algorithm 1 illustrates the experiment in the cash-transfer example with two clusters. With more than two clusters, we pair clusters and implement Algorithm 1 in each pair.

---

**Algorithm 1** Local perturbation with *two* clusters, $\beta$ is a scalar

---

**Require:** Value $\beta$, $K = 2$ with $h \in \{k, k+1\}$, constant $\bar{C}$.

1: $t = 0$ (baseline): either nobody receives treatments or treatments are assigned with $\pi(\cdot; \beta)$ (either case is allowed).

a: Experimenters collect baseline outcomes: for $n$ units in each cluster observe $Y_{i,0}^{(h)}, X_i^{(h)}, h \in \{k, k+1\}$.

2: $t = 1$: experiment starts

a: Based on the target parameter $\beta$, assign treatments for $X_i = 1$ as

$$D_{i,1}^{(k)} | \beta, X_i^{(k)} = x \sim \begin{cases} \text{Bern}(\pi(x, \beta + \eta_n)) \text{ if } h = k \\ \text{Bern}(\pi(x, \beta - \eta_n)) \text{ if } h = k+1 \end{cases}, \quad \bar{C}n^{-1/2} < \eta_n < \bar{C}n^{-1/4}.$$

b: For $n$ units in each cluster $h \in \{k, k+1\}$ observe $Y_{i,1}^{(h)}$.

3: Estimate the marginal effect as in Equation (7).

---

**Remark 3** (Inference on treatment effects)**.** Testing $H_0$ is different from testing hypotheses on treatment effects since these hypotheses do not depend on $\beta^*$. However, our design permits us to estimate separately the direct treatment effect and the spillover effect, which may be of independent interest. We estimate the direct effect with a means' difference between treated and controls, *pooled* across clusters. In Theorem 3.3, we show that the estimated direct effect has a negligible bias of order $o(n^{-1/2})$. We estimate the spillover effect by taking the outcomes' difference between clusters (see Section 3.1).[27] □

### 2.3.2 Multiple-wave experiment: estimation of $\beta^*$

We now show how we can estimate $\beta^*$ using $T$ experimentation periods and $K \geq 2(T+1)$ many clusters. The policy-maker assigns cash transfers and collects outcomes sequentially. Every period $t$, she assigns treatments in cluster $k$ based on parameters $\beta_{k,t}$, with two goals. First, at time $T$, she wants to obtain an estimate $\hat{\beta}^*$ which well approximates $\beta^*$. Second, she wants to improve the experimental participants' welfare.

[26] See Equation (10). Scale-invariance follows from Ibragimov and Müller (2010).

[27] Also, if the experimenter's goal is precise inference on direct effects only, $\beta$ can be chosen based on variance considerations, but we can employ our method – which only requires small perturbations to $\beta$ – to identify spillover effects.

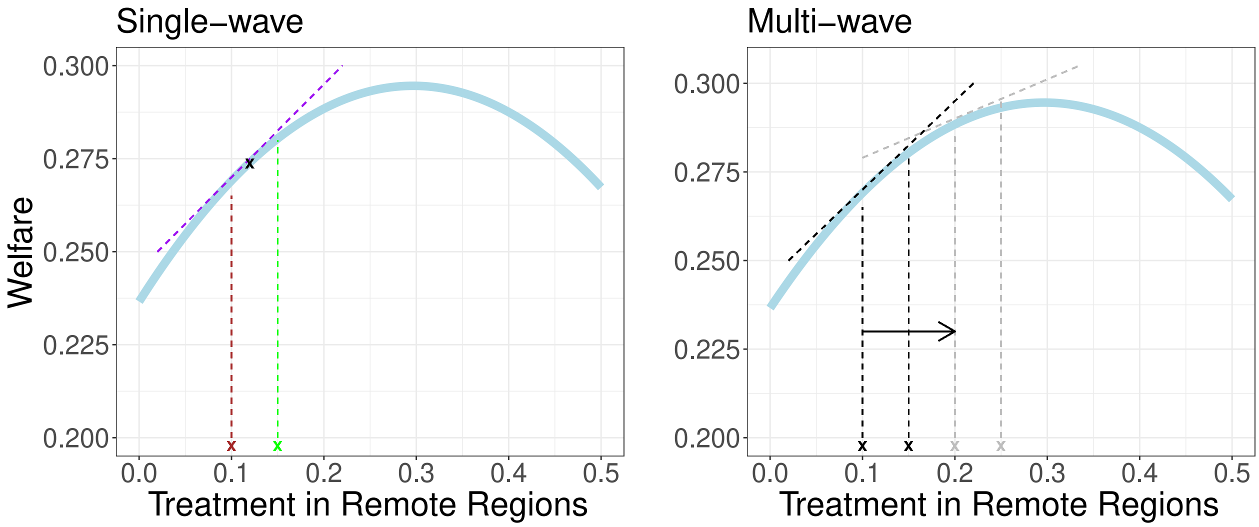


Figure 2: Example of experimental design. The objective function is satisfaction with a cash-transfer program. Half of the population is treated. The left panel is a single-wave experiment with two clusters. In the first cluster, we assign the policy colored in green and the second cluster colored in brown. The right panel is a two-wave experiment. In the first period, we use a pair of clusters to estimate the marginal effect (black color), and we update the policy for a different pair (gray color).

We maximize welfare with the following algorithm: 1) we pair clusters and organize pairs in a circle as in Figure 3; 2) every step $t$, we estimate the marginal effect within each pair; 3) we then update the policy in a given clusters' pair using the information on the marginal effect from the subsequent pair on the circle. We refer to Step 3 as circular cross-fitting.[28] Step 3 guarantees that the experiment is unconfounded. See Section 4 for details.[29]

The right panel in Figure 2 illustrates the procedure for two waves: we use one pair of clusters to estimate the marginal effect, which we then use for the second pair (and vice-versa, see Algorithm 3). The experiment assumes and leverages the concavity of welfare, generally attained under decreasing marginal effects of neighbors' treatments. Under lack of concavity, the experiment returns a *local* optimum. We relax concavity in Appendix B.2.

In Section 4 we measure the method's performance based on the in-sample and out-of-sample regret, and show that with high-probability

$$W(\beta^*) - W(\hat{\beta}) = \mathcal{O}(1/T), \quad \max_k \frac{1}{T}\sum_{t=1}^{T}\Big[W(\beta^*) - W(\beta_{k,t})\Big] = \mathcal{O}(\log(T)/T).$$

where the first equation indicates the out-of-sample regret, and the second equation indicates

[28] Cross-fitting algorithms were used by Chernozhukov et al. (2018) in the different context of double-machine learning with *i.i.d.* data. Our procedure differs due to the adaptive sampling and the circular structure.

[29] Also, note that our method also extends when treatments can only be assigned once. See Appendix B.5.

the in-sample regret. The above equations provide strong guarantees on the method's performance as we deploy the estimated policy on a new population (out-of-sample regret) and the experimental participants (in-sample regret). The regret scales linearly in the number of clusters, as we choose $2T + 2 = K$.

**Remark 4.** An alternative approach we could take is to estimate $y(\cdot)$ assigning different policies to clusters and extrapolating the overall effect. We do not consider this alternative for two reasons. For a generic $p$-dimensional vector $\beta$, the out-of-sample regret is either sensitive to the model used for extrapolation or suffers a curse of dimensionality (e.g., when a grid search is employed). Second, this method does not control the *in-sample* regret, i.e., it must incur significant in-sample welfare loss to estimate the response function $y(\cdot)$.[30] □

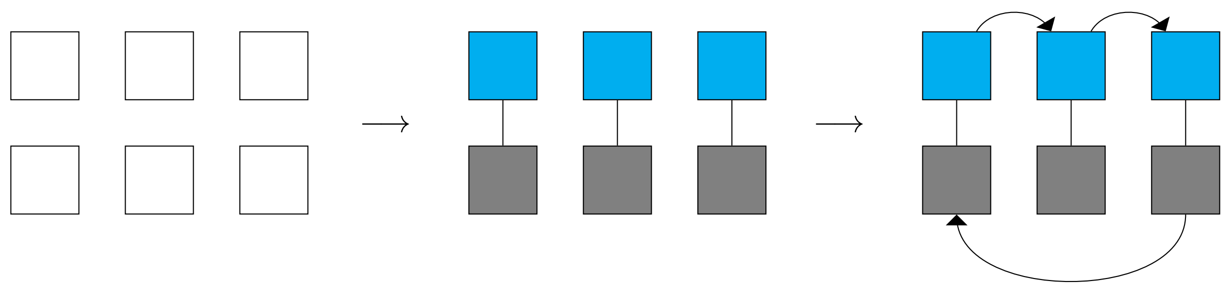

Figure 3: Circular cross-fitting method. Clusters (rectangles) are paired. Within each pair, researchers assign different treatment probabilities to clusters with different colors. Finally, the policy in each pair is updated using information from the consecutive pair.

# 3 Single-wave experiment

In this section, we turn to the design and analysis of a single-wave experiment.

We consider an experiment to test the following hypothesis.

**Definition 3.1** (Testable implication)**.** Let $\beta^* \in \mathcal{B}$ be an interior point. If $W(\beta) = W(\beta^*)$, then

$$H_0 : V^{(j)}(\beta) = 0, \quad \forall j \in \{1, \cdots, l\}, l \leq p. \tag{8}$$

The above implication is at the core of our approach. We can test whether $l$ arbitrary entries of the marginal effect are equal to zero. Rejection implies a lack of global optimality. For expositional convenience, we consider $l = 1$ only (test the first entry being zero). In Appendix B.1 we show how our method generalizes to $l > 1$. We may also test $V^{(j)}(\beta) \leq 0$;

[30] Also, in Section A.3 we show how information on the marginal effect (with a variation of Algorithm 1) is useful for estimation in the absence of a sequential experiment.

for example, for $\pi(x, \beta) = \beta_x$ ($\mathcal{X}$ discrete), one-sided test is informative for whether we should increase treatment probabilities for individuals with $x = j$ (without assuming that $\beta^*$ is in the interior). Finally, it is useful to define the vector

$$\underline{e}_j = \Big[0, \cdots, 0, 1, 0, \cdots, 0\Big], \text{ where } \underline{e}_j \in \{0,1\}^p, \text{ and } \underline{e}_j^{(j)} = 1. \tag{9}$$

Algorithm 2 presents the design. The algorithm pairs clusters. Within each pair, it estimates the first entry of the marginal effect (since here we test $V^{(1)}(\beta) = 0$) using local perturbations – as discussed in Section 2.3. It then constructs a scale-invariant test statistics. Without loss of generality, we index clusters such that each pair contains two consecutive clusters $\{k, k+1\}$ with $k$ being an odd number.

In the following lines, we discuss estimation of marginal and treatment effects and guarantees for inference on $H_0$.

---

**Algorithm 2** One wave experiment for inference with $l = 1$

---

**Require:** Value $\beta \in \mathbb{R}^p$ (exogenous), $K$ clusters, constant $\bar{C}$, size $\alpha$;

1: Organize clusters into $G = K/2$ pairs with consecutive indexes $\{k, k+1\}$;

2: For each pair $g = \{k, k+1\}$, $k$ is odd, run Algorithm 1, with at $t = 1$,

$$D_{i,1}^{(k)} | \beta, X_i^{(k)} = x \sim \begin{cases} \text{Bern}(\pi(x, \beta + \eta_n \underline{e}_1)) \text{ if } h = k \\ \text{Bern}(\pi(x, \beta - \eta_n \underline{e}_1)) \text{ if } h = k+1 \end{cases}, \quad \bar{C} n^{-1/2} < \eta_n < \bar{C} n^{-1/4},$$

and estimate the marginal effect as in Equation (6).

3: Construct the test statistics

$$\mathcal{T}_n = \frac{\sqrt{G}\bar{V}_n(\beta)}{\sqrt{(G-1)^{-1} \sum_g (\widehat{V}_g(\beta) - \bar{V}_n(\beta))^2}}, \quad \bar{V}_n(\beta) = \frac{1}{G} \sum_g \widehat{V}_g(\beta); \tag{10}$$

here $\widehat{V}_g$ is the marginal effect estimated in pair $g$.

4: Construct the test as $1\Big\{|\mathcal{T}_n| > \text{cv}_{G-1}(\alpha)\Big\}$ with size $\alpha$; $\text{cv}_{G-1}(\alpha)$ is the size $\alpha$ t-test's critical quantile with $G - 1$ degrees of freedom.

---

## 3.1 Estimation of marginal and treatment effects

The experiment we just described permits us to estimate three quantities of independent interest: the marginal effect, the direct effect, and the spillover effect. These should be reported by researchers once the experiment is concluded. We describe the estimators below.

Equation (6) provides the marginal effect estimator $\widehat{V}_g(\beta)$ for each pair of clusters $g$. Researchers may report $\bar{V}_n(\beta)$ (Equation 10) in their results – the average across clusters' pairs. We show below that both $\bar{V}_n(\beta)$ and $\widehat{V}_g(\beta)$ provide a consistent estimate of $V^{(1)}(\beta)$.[31]

The experiment also allows us to estimate the direct effect of the treatment and the (marginal) spillover effect separately, respectively defined as:[32]

$$\Delta(\beta) = \int \Big[m(1, x, \beta) - m(0, x, \beta)\Big] dF_X(x), \quad S_1(d, \beta) = \int \frac{\partial m(d, x, \beta)}{\partial \beta^{(1)}} dF_X(x).$$

The direct effect is the treatment effect, keeping fixed the neighbors' treatment probability. $S_1(\cdot)$, the spillover effect, is the marginal effect of a small change in the first entry of $\beta$, keeping fixed individual treatment status.[33] For a given pair of clusters $(k, k+1)$, we estimate

$$\widehat{\Delta}_k(\beta) = \frac{1}{2n} \sum_{h \in \{k, k+1\}} \sum_{i=1}^{n} \Big[\frac{D_{i,1}^{(h)} Y_{i,1}^{(h)}}{\pi(X_i^{(h)}, \beta + \eta_n v_h \underline{e}_1)} - \frac{(1 - D_{i,1}^{(h)}) Y_{i,1}^{(h)}}{1 - \pi(X_i^{(h)}, \beta + \eta_n v_h \underline{e}_1)}\Big], \quad v_h = \begin{cases} 1 \text{ if } h = k \\ -1 \text{ if } h = k+1. \end{cases} \tag{11}$$

The estimator pools observations between the two clusters and takes a difference between treated and control units within each cluster, divided by the probability of treatments. This follows similarly to classical Horvitz-Thompson estimators (Horvitz and Thompson, 1952). Importantly, we divide by the probability of treatments, taking into account the perturbation $\eta_n$. We average direct effects across clusters' pairs to obtain a single measure $\bar{\Delta}_n = \frac{1}{G} \sum_g \widehat{\Delta}_g(\beta)$. The indirect effect is estimated as follows:

$$\widehat{S}_{(k,k+1)}(0, \beta) = \frac{1}{2n} \sum_{h \in \{k, k+1\}} \frac{v_h}{\eta_n} \sum_{i=1}^{n} \Big[\frac{Y_{i,1}^{(h)}(1 - D_{i,1}^{(h)})}{1 - \pi(X_i^{(h)}, \beta + v_h \eta_n \underline{e}_1)} - \bar{Y}_0^{(h)}\Big].$$

The estimator takes a difference between the control units between the two clusters, divided by their corresponding treatment probabilities. Researchers may report the between pairs average $\bar{S}_n(0, \beta) = \frac{1}{G} \sum_g \widehat{S}_g(0, \beta)$,[34] which captures spillovers on the control units.

## 3.2 Theoretical Analysis

Next, we study theoretical guarantees. The following regularity condition is imposed.

**Assumption 3.1** (Regularity 1)**.** Suppose that for all $x \in \mathcal{X}, d \in \{0, 1\}$, $\pi(x, \beta), m(d, x, \beta)$ are uniformly bounded and twice differentiable with bounded derivative.

[31] Note that our discussion directly extends to estimating each entry of $V(\beta)$ as shown in Appendix B.1.

[32] Here, we are implicitly assuming that $m(\cdot)$ is uniformly bounded to invoke the dominated convergence theorem. See the next subsection for formal assumptions.

[33] Similarly to the marginal effect, our setting also extends to estimating $S_j(\cdot)$ for arbitrary entries of $\beta$ as in Appendix B.1.

[34] Here, $\hat{S}(1, \beta)$ follows similarly and omitted for brevity.

Assumption 3.1 imposes smoothness and boundedness restrictions.[35]

**Theorem 3.1** (Consistency of the marginal effect)**.** *Suppose that $\varepsilon_{i,t}^{(k)}$ is sub-gaussian. Let Assumptions 2.1, 2.2, 3.1 hold. Let* $\mathrm{Var}(\sqrt{n}\hat{V}_{(k,k+1)}(\beta)) = \mathcal{O}(\rho_n)$*. Then with probability at least* $1-\delta$*, for any* $\delta \in (0,1)$*,*

$$\Big|\hat{V}_{(k,k+1)}(\beta) - V^{(1)}(\beta)\Big| = \mathcal{O}\Big(\eta_n + \min\Big\{\sqrt{\frac{\gamma_N \log(\gamma_N/\delta)}{n\eta_n^2}}, \sqrt{\frac{\rho_n}{n\eta_n^2\delta}}\Big\}\Big),$$

*where* $\hat{V}_{(k,k+1)}$ *is estimated as in Algorithm 2.*

*For* $\gamma_N \log(\gamma_N)/N^{1/3} = o(1), \eta_n = n^{-1/3}, \hat{V}_{(k,k+1)}(\beta) \rightarrow_p V(\beta)$*.*

The proof is in Appendix C.3.1. Theorem 3.1 shows that we can consistently estimate the marginal effects with two large clusters. Consistency depends on the degree of dependence among unobservables $\varepsilon_{i,t}^{(k)}$ (which also depends on neighbors' assignments). The convergence rate depends on the *minimum* between the maximum degree of the network, which is proportional to $\gamma_N^{1/2}$, and the covariances among unobservables, captured by $\rho_n$. If either the network has a degree that grows at a slower rate than $N$ (recall that $n/N = O(1)$) or a degree equal to $N$ but vanishing covariances, we can consistently estimate the marginal effects. The theorem also illustrates the trade-off in the choice of the deviation parameter $\eta_n$: a larger deviation parameter $\eta_n$ decreases the variance, but it increases the bias. The choice of $\eta_n$ can be based on minimizing the bound in the theorem above.

**Corollary 1.** *Under the conditions in Theorem 3.1, letting* $\gamma_N \log(\gamma_N)/n^{1/3} = o(1), \eta_n = n^{-1/3}$*, for any $K$,* $\bar{V}_n \rightarrow_p V^{(1)}(\beta)$*.*

The above corollary illustrates consistency once we pool information from different clusters (with $K$ being finite). Next, we study inference assuming the following condition.

**Assumption 3.2** (Regularity 2)**.** Assume that for treatments as assigned in Algorithm 2, for all $k \in \{1, \cdots, K\}$, $\varepsilon_{i,t}^{(k)}$ has bounded fourth moment and for some $\bar{C}_k > 0$, $\rho_n \geq 1$,

$$\mathrm{Var}\Big(\frac{1}{\sqrt{n}}\Big[\bar{Y}_1^{(k)} - \bar{Y}_0^{(k)}\Big]\Big) = \bar{C}_k \rho_n. \tag{12}$$

Assumption 3.2 imposes standard moment bounds and a lower bound on the variance of the estimator, attained under independence and positive dependence.

[35]These restrictions hold for a large set of linear and non-linear functions, assuming that $\mathcal{X}$ is compact. Boundedness is often imposed in the literature, see, e.g., Kitagawa and Tetenov (2018).

**Theorem 3.2.** *Let Assumptions 2.1, 2.2, 3.1, 3.2 hold. Let* $n^{1/4}\eta_n = o(1), \gamma_N/N^{1/4} = o(1)$, $K < \infty$*. Then for each pair* $(k, k+1)$*, for* $\widehat{V}_{(k,k+1)}$ *estimated as in Algorithm 2,*

$$\mathrm{Var}\Big(\widehat{V}_{(k,k+1)}\Big)^{-1/2}\Big(\widehat{V}_{(k,k+1)} - V^{(1)}(\beta)\Big) \rightarrow_d \mathcal{N}(0,1).$$

The proof is in Appendix C.3.2. Theorem 3.2 guarantees asymptotic normality. The theorem assumes that the maximum degree $\gamma_N^{1/2}$ grows at a slower rate than the sample size of order $N^{1/8}$ (and hence $n^{1/8}$ since $n$ is proportional to $N$). This condition is stronger than what is required for consistency only.[36] Given Theorem 3.2, we conduct inference with scale-invariant test statistics without necessitating estimation of the (unknown) variance.

**Corollary 2.** *Let the conditions in Theorem 3.2 hold. For* $4 \le K < \infty$*,* $\alpha \le 0.08$*,*

$$\lim_{n\to\infty} P\Big(|\mathcal{T}_n| \le \mathrm{cv}_{K/2-1}(\alpha)\Big| H_0\Big) \ge 1-\alpha, \tag{13}$$

*where* $\mathrm{cv}_{K/2-1}(h)$ *is the size-$h$ critical value of a t-test with* $K/2-1$ *degrees of freedom.*

The proof is in Appendix C.5. The theorem guarantees asymptotically valid inference on $H_0$ as $n \to \infty$ and $K$ is finite. With $l = 1$ the proof is a direct consequence of Theorem 3.2, combined with properties of pivotal statistics in Ibragimov and Müller (2010).[37] In Appendix B.1 we provide expressions for the test statistics and derivations for $l > 1$.

To our knowledge, this is the first set of results for inference on welfare-maximizing policies with unknown interference.

We conclude this section with a study on the estimated direct and spillover effect.

**Theorem 3.3** (Asymptotically neglegible bias of the direct effect)**.** *Let Assumptions 2.1, 2.2, 3.1 hold, and* $\eta_n = o(n^{-1/4})$*. Then for all pairs* $(k, k+1)$*,* $\mathbb{E}\Big[\widehat{\Delta}_{(k,k+1)}(\beta)\Big] = \Delta(\beta) + o(n^{-1/2})$*. Similarly,* $\mathbb{E}\Big[\bar{\Delta}_n(\beta)\Big] = \Delta(\beta) + o(n^{-1/2})$*, where the second term does not depend on* $K$*.*

The proof is in Appendix C.3.3. Theorem 3.3 shows that the bias of the estimated direct effect is asymptotically negligible at a rate faster than the parametric rate $n^{-1/2}$ when *pooling* observations from different clusters. Our insight here is that, with pairing and perturbations of opposite signs, the first-order bias cancels out. This result implies that our experimental design induces a bias that can be ignored for estimation and inference.[38] Given that the bias

[36] We conjecture that weaker restrictions on the degree are possible, as for consistency. We leave their study to future research.

[37] See also Chernozhukov et al. (2018) for a discussion on pivotal inference in the different context of synthetic controls.

[38] Note that $\eta_n = o(n^{-1/4})$ is consistent with requirements in previous theorems.

is asymptotically negligible, we can use existing results for inference on direct effects with a single network (e.g., Sävje et al. 2021). For completeness, we show consistency below.

**Corollary 3.** *Suppose that $\varepsilon_{i,t}^{(k)}$ is sub-gaussian. Let Assumptions 2.1, 2.2, 3.1 hold, and $\pi(x, \beta) \in (\kappa, 1-\kappa), \kappa \in (0,1)$ for all $x \in \mathcal{X}$. Let $\eta_n = o(n^{-1/4})$. Then with probability at least $1-\delta$, for any $\delta \in (0,1)$, for any $K$*

$$\left|\bar{\Delta}_n - \Delta(\beta)\right| = \mathcal{O}\Big(\sqrt{\frac{\gamma_N \log(\gamma_N/\delta)}{Kn}}\Big) + o(n^{-1/2}),$$

The proof is in Appendix C.5. The corollary requires strict overlap (standard in the literature on causal inference) and shows that we can attain consistency for $K < \infty$, $n \to \infty$.

The following result is on the bias of the marginal spillover effects estimators.

**Theorem 3.4** (Marginal Spillover effects)**.** *Let Assumptions 2.1, 2.2, 3.1 hold. Then for all pairs $(k, k+1)$, $\mathbb{E}\Big[\widehat{S}_{(k,k+1)}(0,\beta)\Big] = S_1(0,\beta) + \mathcal{O}(\eta_n)$.*

The proof is in Appendix C.3.4. Theorem 3.4 shows that the bias converges to zero as $\eta_n \to 0$. The rate is slower than the rate of the direct effect's bias (since $\eta_n > n^{-1/2}$). We obtain a slower rate because the marginal spillover effect depends on *between* clusters variations. Consistency and inference follow verbatim as discussed for the marginal effect.

**Remark 5** (Why pairing? Pairing clusters permits finite-clusters asymptotics)**.** Pairing plays a fundamental role in our design with finite $K$. In the absence of pairing, the bias of the marginal and treatment effects would not converge to zero for finite $K$. To gain further intuition, consider a uni-dimensional setting ($X = 1$ almost surely), and $\pi(1,\beta) = \beta$. Then

$$\begin{aligned}\mathbb{E}\Big[\widehat{V}_{(k,k+1)}(\beta)\Big] &= \frac{1}{2}\sum_{h \in \{k,k+1\}} \frac{v_h}{\eta_n} y(1; \beta + v_h \eta_n) \\ &= \underbrace{\frac{1}{2}\sum_{h \in \{k,k+1\}} \frac{v_h}{\eta_n} y(1;\beta)}_{(i)} + \underbrace{\frac{\partial y(1;\beta)}{\partial \beta}}_{(ii)} + \mathcal{O}(\eta_n), \quad v_h = \begin{cases} 1 \text{ if } h = k \\ -1 \text{ if } h = k+1. \end{cases}\end{aligned}$$

Component (i) induces a bias, while (ii) is the target estimand. Observe that (i) equals zero because of the paired design: $v_h$ is one for one cluster and minus one for the other cluster. Suppose instead that a paired design was not implemented, and instead we have $v_h \in \{\pm 1\}$ with equal probability.[39] Then (i) would scale to zero at a slow rate $1/\sqrt{K\eta_n^2}$ after averaging across all the clusters, requiring infinite clusters asymptotics. □

[39]Note that random assignments of probabilities of treatments are common in saturation experiments (Baird et al., 2018; Pouget-Abadie, 2018).

## 4 Multi-wave experiment

In this section, we design the adaptive experiment and derive its theoretical properties.

For illustrative purposes, we provide the algorithm for the one-dimensional case $p = 1$, in Algorithm 3, that is when $\beta \in \mathcal{B} = [\mathcal{B}_1, \mathcal{B}_2]$ is a scalar. In Remark 6 and formally in Appendix E we provide the complete algorithm for the $p$-dimensional case. Theoretical results are for the general $p$-dimensional case ($p$ is finite). A description is below the algorithmic box.

**Algorithm 3** Multiple-wave experiment with $\beta$ scalar

**Require:** Starting value $\beta_0$, $K$ clusters, $T+1$ periods, constant $\bar{C}$.

1: Create pairs of clusters $\{k, k+1\}, k \in \{1, 3, \cdots, K-1\}$;

2: $t = 0$ (initialization):

a: Assign treatments as $D_{i,0}^{(h)} | X_i^{(h)} = x \sim \text{Bern}(\pi(x, \beta_0))$ for all $h \in \{1, \cdots, K\}$.

b: For $n$ units in each cluster observe $Y_{i,0}^{(h)}, h \in \{1, \cdots, K\}$; initalize $\widehat{V}_{k,t} = 0$, $\check{\beta}_k^0 = \beta_0$.

3: **while** $1 \le t \le T$ **do**

a: Define

$$\check{\beta}_h^t = \begin{cases} P_{\mathcal{B}_1, \mathcal{B}_2 - \eta_n}\Big[\check{\beta}_h^{t-1} + \alpha_{h+2,t}\widehat{V}_{h+2,t-1}\Big], & h \in \{1, \cdots, K-2\}, \\ P_{\mathcal{B}_1, \mathcal{B}_2 - \eta_n}\Big[\check{\beta}_h^{t-1} + \alpha_{1,t}\widehat{V}_{1,t-1}\Big], & h \in \{K-1, K\}; \end{cases}$$

where $\alpha_{k,t}$ is the learning rate (see Remark 7), and $P_{\mathcal{B}_1, \mathcal{B}_2 - \eta_n}$ is the projection operator.

b: Assign treatments as (for $\bar{C} n^{-1/2} < \eta_n < \bar{C} n^{-1/4}$)

$$D_{i,t}^{(h)} | X_i^{(h)} = x \sim \text{Bern}(\pi(x, \beta_{h,t})), \quad \beta_{h,t} = \begin{cases} \check{\beta}_h^t + \eta_n \text{ if } h \text{ is odd} \\ \check{\beta}_h^t - \eta_n \text{ if } h \text{ is even} \end{cases} \tag{14}$$

c: For $n$ units in each cluster $h \in \{1, \cdots, K\}$ observe $Y_{i,t}^{(h)}$;

d: For each pair $\{k, k+1\}$, estimate

$$\hat{V}_{k,t} = \hat{V}_{k+1,t} = \frac{1}{2\eta_n}\Big[\bar{Y}_t^{(k)} - \bar{Y}_0^{(k)}\Big] - \frac{1}{2\eta_n}\Big[\bar{Y}_t^{(k+1)} - \bar{Y}_0^{(k+1)}\Big].$$

4: **end while**

5: Return $\hat{\beta}^* = \frac{1}{K}\sum_{k=1}^K \check{\beta}_k^T$

The algorithm pairs clusters, and initializes clusters at the same starting value $\beta_0$, $\check{\beta}_1^1 = \cdots = \check{\beta}_K^1 = \beta_0$. At $t = 0$, it randomizes treatments independently as

$$t = 0: \quad D_{i,t}^{(k)} | X_i^{(k)} = x \sim \pi(x; \beta_0), \quad \text{for all } (i, k).$$

Here, $\beta_0$ is chosen exogenously, e.g., it is the current policy in place. Over each iteration $t$, we assign treatments based on $\beta_{k,t}$ for cluster $k$ at time $t$ which equals the parameter

$\check{\beta}_k^t$ obtained from a previous iteration plus a positive (negative) perturbation $\eta_n$ in the first (second) cluster in a pair. The local perturbation follows similarly to what discussed in the previous section; also, by construction, $\check{\beta}_k^t$ is the same for a given pair $(k, k+1)$, where $k$ is odd. We choose $\check{\beta}_k^{t+1}$ via *circular cross-fitting*: we wrap clusters in a circle and update the parameter in a pair of clusters $(k, k+1)$ using information from the subsequent pair (see Figure 3). The algorithm runs over $T$ periods and returns $\hat{\beta}^* = \frac{1}{K}\sum_{k=1}^K \check{\beta}_k^{T+1}$.

In our experiment, we update the policy in each clusters' pair with information from a subsequent pair. This approach guarantees that the estimated policy used to randomize treatments in cluster $k$ does not depend on observables and unobservables in that same cluster, assuming that the number of clusters is twice as large as the number of iterations.

**Lemma 4.1** (Unconfoundedness)**.** *Let $T/p+1 \leq K/2$. Consider the experimental design in Algorithm E.1 for generic p-dimensions (and Algorithm 3 for $p=1$). Then for any $k$,*

$$\Big(\beta_{k,1}, \cdots, \beta_{k,T}\Big) \perp \Big\{Y_{i,t}^{(k)}(\mathbf{d}), X_i^{(k)}, \mathbf{d} \in \{0,1\}^N\Big\}_{i \in \{1,\cdots,N\}, t \leq T}.$$

The proof is in Appendix C.2.5. Lemma 4.1 shows that the parameters used in the experiment are independent of potential outcomes and covariates in the same cluster. Namely, the circular cross-fitting breaks the dependence due to repeated sampling, which would otherwise confound the experiment.[40] See Remark 8 for a discussion.

**Remark 6** ($p$-dimensional case: Algorithm E.1)**.** The algorithm for the $p$-dimensional case follows similarly to the uni-dimensional case with a minor change: we consider $T/p$ many *waves*/iterations, each consisting of $p$ periods. Within each wave $w$, every period, we perturb a single coordinate of $\check{\beta}_k^w$, compute the marginal effect for that coordinate, and repeat over all coordinates $j \in \{1, \cdots p\}$ before making the next policy update to select $\check{\beta}_k^{w+1}$. □

**Remark 7** (Learning rate)**.** We are now left to discuss how "large" should be the step size, i.e., if we know that the marginal effect is positive, by how much should we increase the treatment probability? Assuming strong concavity of the objective function, the learning rate $\alpha_{k,t}$ should be of order $1/t$. A more robust choice (see Theorem B.2) is

$$\alpha_{k,t} = \begin{cases} \frac{J}{T^{1/2-v/2}||\hat{V}_{k,t}||} \text{ if } ||\hat{V}_{k,t}||_2^2 > \frac{c}{T^{1-v}} - \epsilon_n, \\ 0 \text{ otherwise} \end{cases}, \tag{15}$$

[40]This setting is different from previous literature on adaptive experimentation, which focuses on settings where repeated sampling does not occur. See, for example, Kasy and Sautmann (2019); Wager and Xu (2021).

for a positive $\epsilon_n$, $\epsilon_n \to 0$, and small constants $1 \geq v, c > 0$.[41] Here, the learning rate divides the estimated marginal effect by its norm (known as gradient norm rescaling, Hazan et al. 2015), and guarantees control of the out-of-sample regret under strict quasi-concavity. This choice is appealing since it guarantees comparable step sizes between different clusters. □

**Remark 8** (Why circular cross-fitting? Bias with repeated sampling)**.** Here, we illustrate the source of bias if the circular cross fitting was not employed. Every period, the researcher can only identify the expected outcome of $Y_{i,t}^{(k)}$ conditional on the parameter $\beta_{k,t}$, namely $\widetilde{W}(\beta_{k,t}) = \mathbb{E}_{\beta_{k,t}}[Y_{i,t}^{(k)}|\beta_{k,t}]$. If $\beta_{k,t}$ was chosen exogenously, based on information from a different cluster, then $\mathbb{E}_{\beta_{k,t}}[Y_{i,t}^{(k)}|\beta_{k,t}] = \mathbb{E}_{\beta_{k,t}}[Y_{i,t}^{(k)}] = W(\beta_{k,t})$, where $W(\beta_{k,t})$ defines the expected welfare once we deploy the policy $\beta_{k,t}$ on a new population. However, this is not the case if $\beta_{k,t}$ is estimated using information on $Y_{i,t-1}^{(k)}$. Consider the example where the outcome depends on some auto-correlated unobservables $\nu_{i,t}$ and treatment assignments in Figure 4. The *dependence* structure of Figure 4 implies that: $W(\beta_{k,t}) = \mathbb{E}_{\beta_{k,t}}[Y_{i,t}^{(k)}] \neq \mathbb{E}_{\beta_{k,t}}[Y_{i,t}^{(k)}|\beta_{k,t}] = \widetilde{W}(\beta_{k,t})$, if $\beta_{k,t}$ depends on covariates and unobservables previous outcomes (and so on unobservables $\nu_{i,t}^{(k)}$) in cluster $k$. Here $W(\beta_{k,t})$ captures the estimand of interest. Instead, $\widetilde{W}(\beta_{k,t})$ denotes what we can identify. Our algorithm breaks such dependence and guarantees unconfounded experimentation as shown in Lemma 4.1. □

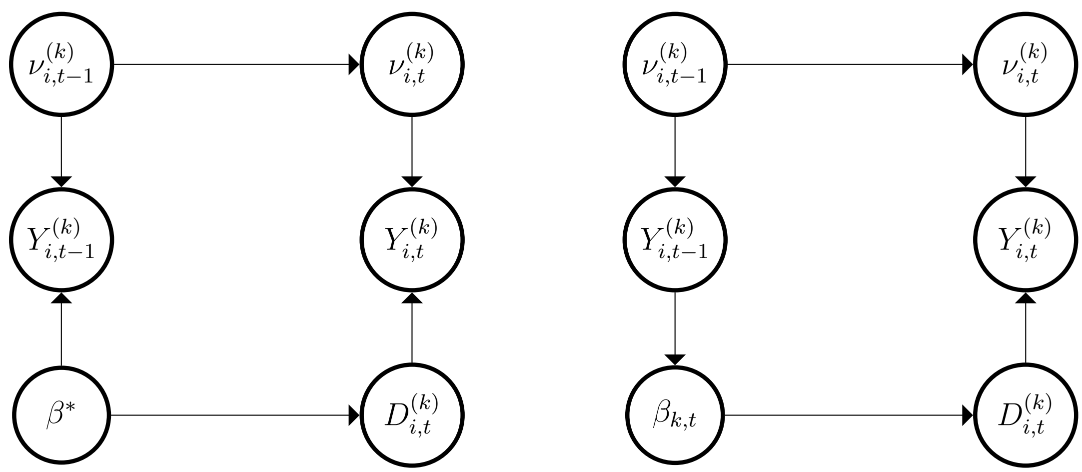


Figure 4: The left-panel shows the dependence structure when a static policy is implemented on a new population (we omit $D_{i,t-1}^{(k)}$ for expositional convenience). The right panel shows the dependence structure of a sequential experiment that uses the same units for policy updates over subsequent periods in the presence of *repeated* sampling.

[41] Formally, we let $\epsilon_n \propto \sqrt{\frac{\gamma_N}{\eta_n^2 n}} + \eta_n$. See Lemma C.12 in the Appendix for further details.

## 4.1 Theoretical guarantees

Next, we derive theoretical properties. Let $\check{T} = T/p$. We assume the following.

**Assumption 4.1.** Let : (A) $\varepsilon_{i,t}^{(k)}$ be sub-gaussian; (B) $K \geq 2(T/p + 1)$.

Condition (A) states that unobservables have sub-gaussian tails (attained, for example, by bounded random variables); (B) assumes that we have at least twice as many clusters as the number of waves, which guarantees that Lemma 4.1 (unconfoundedness) holds.

In the following results, we impose the following restriction.

**Assumption 4.2** (Strong concavity)**.** Assume that $W(\beta)$ is $\sigma$-strongly concave over $\mathcal{B}$, for some arbitrary $\sigma > 0$.

Strong concavity is a common feature of objective functions (Bottou et al., 2018). A simple example of strong concavity is Example 2.1, when parameters are calibrated to real-world data as shown in Figure 5.[42] Strong concavity guarantees uniqueness of the optimum. We relax Assumption 4.2 in Appendix B.2.

**Theorem 4.2.** *Let Assumptions 2.1, 2.2, 3.1, 4.1, 4.2 hold. Take a small* $1/4 > \xi > 0$, $\alpha_{k,w} = J/w$ *for a finite* $J \geq 1/\sigma$. *Let* $n^{1/4-\xi} \geq C\sqrt{p \log(n)\gamma_N T^{B_p} \log(KT)}$, $\eta_n = 1/n^{1/4+\xi}$, *for finite constants* $B_p, C > 0$. *Then with probability at least* $1 - 1/n$, *for a constant* $\bar{C}' < \infty$, *independent of* $(p, n, N, K, T)$,

$$||\beta^* - \hat{\beta}^*||^2 \leq \frac{p\bar{C}'}{\check{T}}.$$

The proof is in Appendix C.3.5. Theorem 4.2 provides a *small sample* upper bound on the distance between the estimated policy and the optimal one. The bound only depends on $T$ (and not $n$) since $n$ is assumed to be sufficiently larger than $T$.

**Corollary 4.** *Let the conditions in Theorem 4.2 hold. Let* $K = 2(T/p + 1)$. *Then with probability at least* $1 - 1/n$, *for a constant* $C' < \infty$ *independent of* $(p, n, N, K, T)$,

$$W(\beta^*) - W(\hat{\beta}^*) \leq \frac{pC'}{K}.$$

The proof is in Appendix C.5. The above corollary formalizes the out-of-sample regret bound scaling *linearly* with the number of periods and clusters, choosing $K = 2(T/p + 1)$. Also, the bound scales polynomially in $p$, for $n$ sufficiently large.[43]

Researchers may wonder whether the procedure is "harmless" also on the in-sample units. We provide in-sample guarantees in the following theorem.

[42] See Section 5 for details.

[43] This is different from grid-search procedures, where the rate in $K$ would decay (exponentially) in $p$.

**Theorem 4.3** (In-sample regret)**.** *Let the conditions in Theorem 4.2 hold. Then with probability at least* $1 - 1/n$*, for a constant* $c < \infty$ *independent of* $(p, n, N, K, T)$*,*

$$\max_{k \in \{1, \cdots, K\}} \frac{1}{\check{T}} \sum_{w=1}^{\check{T}} \Big[ W(\beta^*) - W(\check{\beta}_k^w) \Big] \leq c \frac{p \log(\check{T})}{\check{T}}.$$

The proof is in Appendix C.3.6. Theorem 4.3 guarantees that the cumulative welfare in *each* cluster $k$, incurred by deploying the current policy $\check{\beta}_k^w$ at wave $w$ (recall that in the general $p$-dimensional case we have $\check{T}$ many waves), converges to the largest achievable welfare at a rate $\log(T)/T$, also for those units participating in the experiment.[44] This result guarantees that the proposed design is not *harmful* to experimental participants.

To our knowledge, this is the first set of regret guarantees under unknown (and partial) interference.

We now contrast the rate and theoretical derivations with past literature. In online optimization, the rate $1/T$ is common for stochastic gradient descent methods (Bottou et al., 2018), but assuming a sufficiently large set of independent observations each iteration. Wager and Xu (2021) derive regret guarantees of order $1/T$ in the different setting of market pricing, under mean-field asymptotics (i.e., $n \to \infty$), with units and samples over each wave being independent. Our results provide small sample guarantees without imposing independence or modeling assumptions, other than partial interference. When instead only one or few independent noisy observations are available for each sample, rates are generally sub-linear (e.g., Flaxman et al., 2004). We get linear rates because, in our setting, we can exploit within-cluster concentration (which leverages independent treatments) and between clusters' variation to control large deviations of the estimated gradients. Finally, Viviano (2019) provides a set of guarantees under interference in the presence of a single observational study, where the neighbors of sampled individuals are observed, but there is a single network. Here, we consider a sequential experiment with partial interference and exploit smoothness conditions instead of bounds which depend on the geometric (VC) complexity.

These differences require a different proof technique from previous references. The proof of the theorem (i) uses concentration arguments for locally dependent graphs (Janson, 2004) to derive an exponential rate of convergence, adjusted by the dependence component $\gamma_N$; (ii) it uses the within-cluster and between-clusters variation for consistent estimation of the marginal effect, together with the matching design to guarantee that there is a vanishing bias when estimating marginal spillover effects for $K < \infty$; (iii) it uses a recursive argument

[44] Observe that by a first-order Taylor expansion under Assumption 4.1, a direct corollary is that the bound also holds for policies $\check{\beta}_k^w \pm \eta_n$ up to an additional factor which scales to zero at rate $\eta_n$ (and therefore negligible under the conditions imposed on $n$).

to bound the cumulative error obtained through the estimation and circular cross-fitting.

**Remark 9** (Relaxing strong concavity)**.** In Appendix B.2 (Theorem B.2), we show that under weaker restrictions than Assumption 4.2 (assuming that the function is strictly quasi-concave and locally strong concave), we obtain an almost linear rate of the out-of-sample regret, using the gradient norm rescaling in Equation (15). □

## 4.2 A comparison with the unconstrained optimum

We conclude our discussion with the following question: how does $\beta^*$ compare to the policy that assigns treatments arbitrary (without restrictions on the policy function)? We study this question under more restrictive conditions. Here, we omit the super-script $k$ since our reasoning will apply to any cluster. We want to study:

$$W_N^* - W(\beta^*), \quad W_N^* = \sup_{\mathcal{P}_N(\cdot) \in \mathcal{F}} \frac{1}{N} \sum_{i=1}^N \mathbb{E}\Big[\mathbb{E}_{D \sim \mathcal{P}_N(A,X)}[Y_{i,t}|A, X]\Big] \tag{16}$$

where $\mathcal{F}$ denotes the set of conditional distribution of the vector $D \in \{0,1\}^N$, given the network $A$ and the covariates of all observations $X$. Equation (16) denotes the difference between the expected potential outcomes, evaluated at the global optimum over all possible assignments (conditional on $A, X$), and the welfare evaluated at the optimal policy $\beta^*$.

**Assumption 4.3** (Discrete parameter space, assignment and minimum degree)**.** Assume that $X_i \in \mathcal{X}, \mathcal{X} = \{1, \cdots, |\mathcal{X}|\}, |\mathcal{X}| < \infty$. Let $\pi(x, \beta) = \beta_x$, and $\mathcal{B} = [0,1]^{|\mathcal{X}|}$. Assume in addition that $\inf_{x,x',u'} \int l(x, u, x', u') dF_{U|X=x}(u) \geq \kappa$, for some $\kappa \in (0, 1]$.

Assumption 4.3 states that we assign treatments based on finitely many observable types.[45] Each type $x \in \mathcal{X}$ is assigned a different probability $\beta_x$, which can take any value between zero and one. Assumption 4.3 also states that conditional on individual's type $(X_i, U_i)$, any other unobserved type $U_j$ can form a connection with individual $i$ with some positive probability, conditional on observables $X_j$, provided that $i$ and $j$ are connected under the latent space representation (recall Equation 1).[46] The second restriction is on the potential outcomes. Let

$$\begin{aligned} &Y_{i,t}(\mathbf{d}_t) = \Big[\Delta(X_i) - c(X_i)\Big]\mathbf{d}_{i,t} + \mathcal{S}_{i,t}(\mathbf{d}_t) + \nu_{i,t}, \quad \mathbb{E}[\nu_{i,t}|X, A] = 0 \\ &\mathcal{S}_{i,t}(\mathbf{d}_t) = s\Big(\frac{\sum_{j \neq i} A_{i,j}\mathbf{d}_{j,t} 1\{X_j = 1\}}{\max\{\sum_{j \neq i} A_{i,j} 1\{X_j = 1\}, 1\}}, \cdots, \frac{\sum_{j \neq i} A_{i,j}\mathbf{d}_{j,t} 1\{X_j = |\mathcal{X}|\}}{\max\{\sum_{j \neq i} A_{i,j} 1\{X_j = |\mathcal{X}|\}, 1\}}\Big). \end{aligned} \tag{17}$$

[45] This is often imposed, see e.g., Manski (2004), Graham et al. (2010).

[46] This condition is consistent with Assumption 2.1, since the assumption states that the expected minimum degree is bounded from below by $\kappa\gamma_N^{1/2}$, which is smaller than the maximum degree $\gamma_N^{1/2}$.

Here $\Delta(\cdot)$ is the conditional direct treatment effect, and $c(\cdot)$ is the cost of the treatment. The function $s(\cdot)$ captures the spillover effects. The spillover effects depend on the fraction of neighbors who have received treatments. Spillovers are heterogeneous in the observable neighbors' types (note also that individuals with different types may connect differently to other types as shown in Equation 1) .

**Assumption 4.4** (Costs are opportunity costs of an equal-impact intervention with no spillovers)**.** Assume that $\Delta(x) = c(x)$ for all $x \in \mathcal{X}$.

Assumption 4.4 states that the cost of the treatment is the opportunity cost, had the treatment been assigned to the same individuals which are disconnected. In a cash-transfer program, we may assign treatments to individuals in the same or nearby villages or to individuals spread out on an entire state or continent, without creating spillover effects in the latter case. The cost of the treatment to assign treatments in the same villages is their opportunity cost, which in this case is the direct treatment effects, assuming that the two interventions have the same direct effects.[47]

**Theorem 4.4.** *Let Equation (17) holds, with $s(\cdot)$ twice differentiable with bounded derivatives. Suppose that Assumption 2.1, 2.2, 4.3, 4.4 hold. Then as $N \to \infty, \gamma_N \to \infty$,*

$$\sup_{\mathcal{P}_N(\cdot) \in \mathcal{F}} \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}\Big[\mathbb{E}_{D \sim \mathcal{P}_N(A,X)}[Y_{i,t}|A, X]\Big] - W(\beta^*) \to 0.$$

The proof is in Appendix C.3.7. Theorem 4.4 shows that the assignment mechanism that maximizes welfare, when treatments can be assigned arbitrarily, converges to the optimal policy $\beta^*$ when treatments are individual specific and randomized. The theorem assumes that the maximum degree converges to infinity, but it may converge at a slower rate than $N$, consistently with our conditions in previous theorems. This is a novel result in the context of the literature on targeting networked individuals, and it depends on the particular model of spillovers considered in this subsection and the costs' assumption.[48]

**Corollary 5.** *Let the conditions in Theorem 4.2 and Theorem 4.4 hold. Then for a constant $C < \infty$ independent of $(N, n, T, K)$, $\lim_{N \to \infty, \gamma_N \to \infty} P\Big(W_N^* - W(\hat{\beta}) \leq \frac{C}{T}\Big) = 1$.*

[47]Note that, under such restriction, treating each individual is not optimal only if, for *some* treatment configurations, there are negative spillovers. For instance, giving cash transfers to richer individuals may decrease the average satisfaction with the program, or in an advertisement campaign, treating certain individuals may reduce treatment efficacy.

[48]In a different context, Akbarpour et al. (2018) show instead that for a class of diffusion mechanisms, random seeding is approximately optimal as we choose a few more seeds. Here, we do not study the problem from the perspective of network diffusion but instead focus on an exogenous interference mechanism with heterogeneity. Our result hinges on Assumption 4.4.

# 5 Calibrated Experiments

In this section, we study the properties of the methodology in numerical studies. We calibrate simulations to data from Cai et al. (2015) and Alatas et al. (2012, 2016), while making simplifying assumptions whenever necessary. In the first calibration, the outcome is insurance adoption, and the treatment is whether an individual received an intensive information session in the experiment. In the second calibration, the treatment is whether a household received a cash transfer, and the outcome is program satisfaction.[49]

Throughout these simulations, we study the problem of choosing a univariate parameter $\beta$, which denotes the unconditional treatment probability. In each cluster $k$, we generate data under a quadratic model of the form

$$Y_{i,t} = \phi_0 + \phi_1 D_{i,t} + \phi_2 S_{i,t} + \phi_3 S_{i,t}^2 - c D_{i,t} + \eta_{i,t}, \quad S_{i,t} = \frac{\sum_{j \neq i} A_{i,j} D_{i,t}}{\max\{1, \sum_{j \neq i} A_{i,j}\}}, \quad \eta_{i,t} \sim_{i.i.d.} \mathcal{N}(0, \sigma^2), \tag{18}$$

where $c$ is the cost of the treatment. We consider two sets of parameters $\Big(\phi_0, \phi_1, \phi_2, \phi_3, \sigma^2\Big)$ calibrated to data from Cai et al. (2015) and Alatas et al. (2012, 2016) respectively. We obtain information on neighbors' treatment directly from data from Cai et al. (2015). For the second application, we merge data from Alatas et al. (2012), and Alatas et al. (2016), and use information from approximately one hundred observations whose neighbors' treatments are all observable to estimate the parameters.[50] For either application, we estimate a linear model as in Equation (18) also controlling for additional covariates to guarantee unconfoundedness of the treatment.[51] We consider as cost of treatment $c = \phi_1$, i.e., the opportunity cost of allocating the treatment to a population of disconnected individuals.

We generate clusters with $N = 600$ units, and sample $n \in \{200, 400, 600\}$. We generate

[49] The experiment of Cai et al. (2015) contains multiple arms assigned at the household and village level. Here, we only focus on the treatment effects of intensive information sessions, pooling the remaining arms together for simplicity. The experiment of Alatas et al. (2012) contains different arms assigned at the village level, as well as information on cash transfers assigned at the household level. Here, we study the effect of cash transfers only and control for village-level treatments when estimating the parameters of interest.

[50] This is different from Section 2 (Figure 2) where we use information from individuals whose 80% or more neighbors are observable. We make this choice in Section 2 to increase the sample size and precision to estimate heterogeneous effects. This approach introduces a sampling bias in the estimation procedure, which we ignore for simplicity, given that our goal is not the analysis of the original experiment but only calibrating numerical studies.

[51] For Cai et al. (2015) the covariates are gender, age, rice area, literacy level, a coefficient that captures the risk aversion, the baseline disaster probability, education, and a dummy containing information on whether the individual has one to five friends. For Alatas et al. (2012) we control for the education level, village-level treatments, i.e., how individuals have been targeted in a village (i.e., via a proxy variable for income, a community-based method, or a hybrid), the size of the village, the consumption level, the ranking of the individual poverty level, the gender, marital status, household size, the quality of the roof and top (which are indicators of poverty).

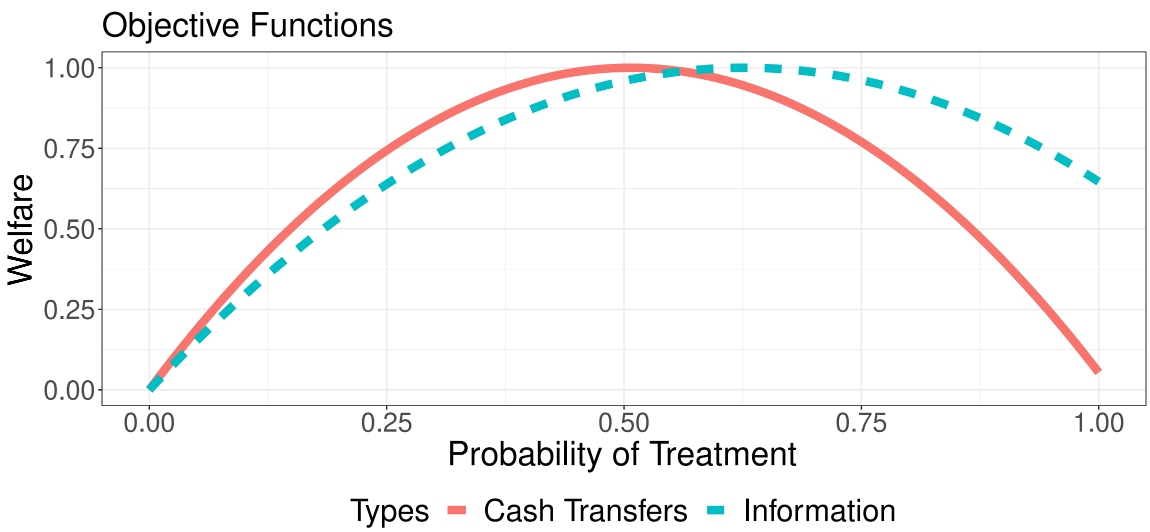


Figure 5: Objective functions (rescaled by $W(\beta^*)$, and minus the intercept $\phi_0$) as functions of unconditional treatment probabilities, with cost of treatments $c = \phi_1$.

a geometric network of the form

$$A_{i,j} = 1\Big\{||U_i - U_j||_1 \leq 2\rho/\sqrt{N}\Big\}, \quad U_i \sim_{i.i.d.} \mathcal{N}(0, I_2),$$

where the parameter $\rho$ governs the density of the network. The geometric formation process and the $1/\sqrt{N}$ follows similarly to simulations in Leung (2020). We consider two networks, a "sparse network" with $\rho = 2$, reported in the main text, and a "dense network", with $\rho = 6$, studied in the Appendix. Throughout our analysis, without loss of generality, we report welfare divided by its maximum $W(\beta^*)$ (i.e., $W(\beta^*) = 1$), and we subtract the intercept $\phi_0$ since $\phi_0$ does not depend on $\beta$.

We conclude with details on estimation. We fix the perturbation parameter $\eta_n = 10\%$; similarly, in the adaptive experiment, we choose the learning rate $10\%/\sqrt{t}$ with gradient norm rescaling as Remark 7. This choice guarantees that for each iteration, we only vary treatment probabilities by at most 10%, and the size of the variation is decreasing over each iteration, in the same spirit of learning rate under strong concavity without norm rescaling.[52] Since the model does not allow for time-varying fixed effects, we estimate marginal effects without baseline outcomes. For the multi-wave experiment, we initialize parameters at a small treatment probability $\beta = 0.2$.[53]

[52]This choice is preferable to $10\%/\sqrt{T}$ because it allows for larger steps in the initial iterations. A valid alternative choice is also $10\%/t$, corresponding to the one under strong concavity. The latter case has a practical drawback: updates become very small after a few iterations. For a comparison, see Figure D.5.

[53]This choice guarantees that no less than 10% of individuals are treated once we impose the negative perturbation.

## 5.1 One-wave experiment

First, we study the properties of the one wave experiment as we vary the number of clusters $K$ and the sample size from each cluster $n$. We are interested in testing the one-sided null of whether we should increase the number of treated individuals to increase welfare, i.e.,

$$H_0 : \frac{\partial W(\beta)}{\partial \beta} \leq 0, \quad H_1 = \frac{\partial W(\beta)}{\partial \beta} > 0 \quad \beta \in [0.1, \cdots, \beta^*]. \tag{19}$$

In Figure D.1, in the Appendix, we report the power of the test as a function of the regret for $\rho = 2$. Power is increasing in the regret, the number of clusters, and sample size. However, the marginal improvement in the power from twenty to thirty clusters is small. This result is suggestive of the benefit of the method even with few clusters and a small sample size.

Figure 6 illustrates the benefit of the method as, upon rejection of $H_0$, we recommend increasing the treatment probability by 5%, as a function of the baseline treatment probability. For example, for $\beta = 0.2$, it indicates the relative welfare increase if we were to increase the treatment probability to 0.25, relative to the status quo where $\beta = 0.2$. The figure shows that the relative improvement can be as large as fifty percentage points compared to the status quo. In addition, while the gain is increasing in the sample size, substantial gains can be obtained even if the sample size from each cluster is as small as $n = 200$. This is particularly relevant for targeting information: in such a case, differences in power across different sample sizes only occur when the regret is small (see Figure D.1), or equivalently, when we are already close to the optimum. As a result, once we take welfare effects into account, larger sample size may lead to small (negligible) improvements in welfare. As this example suggests, when the goal is welfare maximization, power analysis may be complemented by the welfare analysis discussed here.

Results are robust as we increase the density of the network with $\rho = 6$. These are reported in Appendix D for the sake of brevity. Finally, in Table 1 we report the size of the test with size 5%. We observe good size control for $n = 600$, and the test undercovers by no more than seven percentage points under any data generating process.

## 5.2 Multiple-wave experiment

Finally, we study the performance of the adaptive experiment. We let $T \in \{5, 10, 15, 20\}$. In Table 2 we report the welfare improvement of the proposed method with respect to a grid search method that samples observations from an equally spaced grid between $[0.1, 0.9]$ with a size equal to the number of clusters (i.e., $2T$). We consider the best competitor between the one that maximizes the estimated welfare obtained from a correctly specified quadratic function and the one that chooses the treatment with the largest value within the grid. The panel at the top of Table 2 reports the out-of-sample welfare improvement. The improvement is positive and up to three percentage points for targeting information and up to sixty percentage points for targeting cash transfers. Improvements are generally larger for larger $T$. In one instance only, for $T = 5$ and a small sample size $n = 200$, we observe a negative effect for targeting information of two percentage points. The panel at the bottom of Table 2 reports positive and large improvements for the in-sample welfare across all the designs, worst-case across clusters. For the worst-case regret, we fix the number of clusters to $K = 40$ for our method and study the properties as a function of the number of iterations.[54] The improvements are twice as large for targeting information and thirty percentage points larger for targeting cash transfers. These are often increasing in $T$ with a few exceptions.[55]

These results illustrate large benefits for *both* estimating policies to be implemented on a new population and to maximize participants' welfare.

Figure 7 reports the out-of-sample regret of the method (and Figure D.2 the in-sample regret). The regret converges to zero as we increase the number of iterations. The larger sample size guarantees a smaller regret. In Appendix D.1, D.2, we report results for $\rho = 6$, which are consistent with findings in the main text. In Appendix D.3 we provide simulations with covariates using data from Alatas et al. (2012). In Appendix D.4 we discuss simulations calibrated to a small M-Turk experiment on information diffusion to increase vaccination.

Table 1: One wave experiment. 200 replications. Coverage for testing $H_0$ (size is 5%). First panel corresponds to $\rho = 2$, and second panel to $\rho = 6$.

| | Information | | | | Cash Transfer | | | |
|---|---|---|---|---|---|---|---|---|
| $K =$ | 10 | 20 | 30 | 40 | 10 | 20 | 30 | 40 |
| $n = 200$ | 0.905 | 0.950 | 0.905 | 0.900 | 0.920 | 0.940 | 0.915 | 0.895 |
| $n = 400$ | 0.980 | 0.960 | 0.900 | 0.925 | 0.980 | 0.960 | 0.895 | 0.930 |
| $n = 600$ | 0.975 | 0.970 | 0.955 | 0.945 | 0.970 | 0.995 | 0.960 | 0.935 |
| $n = 200$ | 0.925 | 0.880 | 0.880 | 0.900 | 0.925 | 0.940 | 0.905 | 0.905 |
| $n = 400$ | 0.980 | 0.940 | 0.920 | 0.920 | 0.980 | 0.960 | 0.900 | 0.925 |
| $n = 600$ | 0.975 | 0.890 | 0.930 | 0.995 | 0.975 | 0.995 | 0.950 | 0.915 |

[54] Fixing $K = 40$ allows us to change the number of iterations up to $T = 20$ while keeping fixed the number of clusters.

[55] The reason improvements are not always increasing in $T$ is that uniform concentration may deteriorate for large $T$ and small $n$ as we consider the worst-case welfare across clusters.

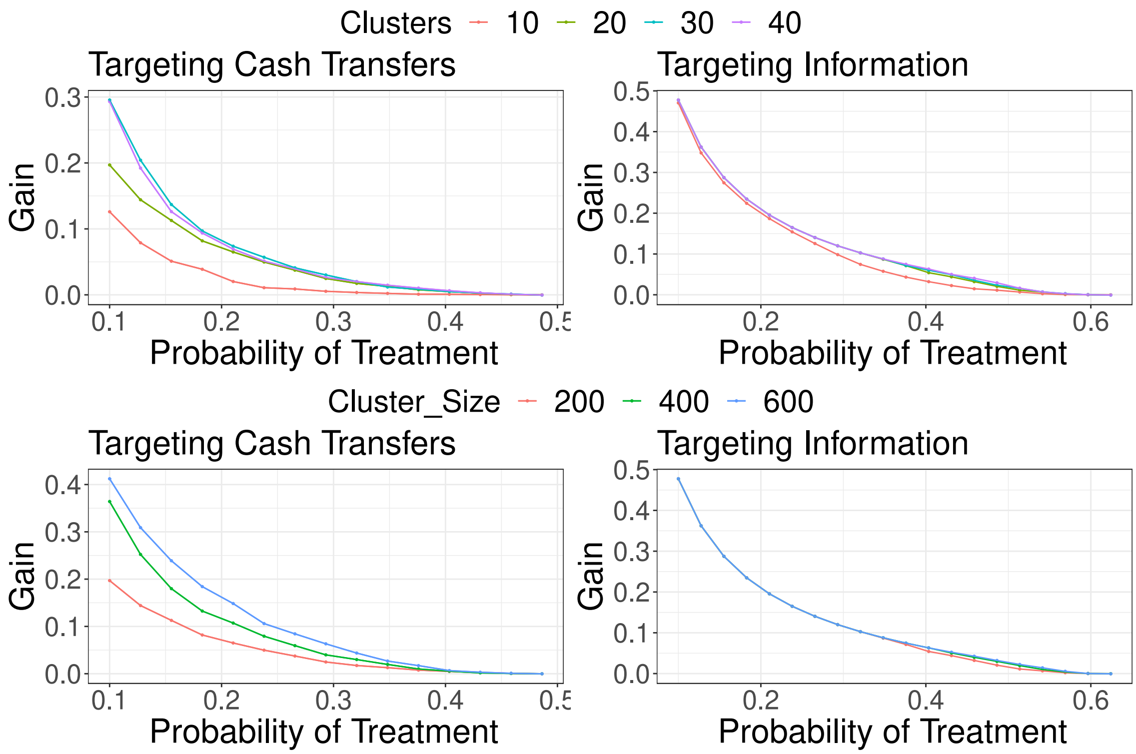


Figure 6: One-wave experiment. $\rho = 2$. Expected percentage increase in welfare from increasing the probability of treatment $\beta$ by 5% upon rejection of $H_0$. Here, the x-axis reports $\beta \in [0.1, \cdots, \beta^* - 0.05]$. The panels at the top fix $n = 400$ and vary the number of clusters. The panels at the bottom fix $K = 20$ and vary $n$.

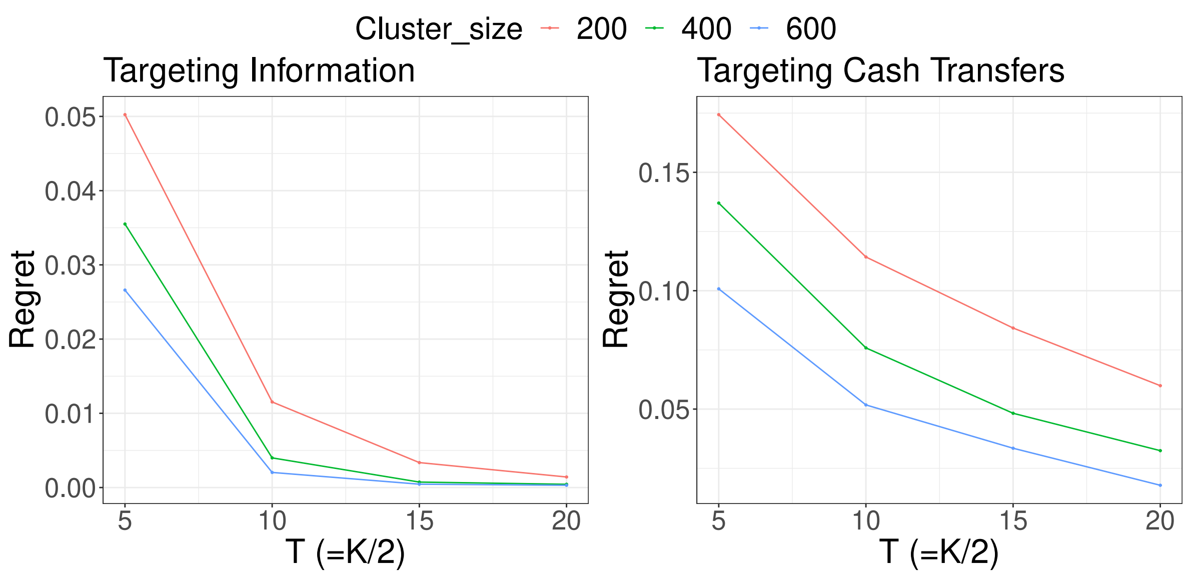


Figure 7: Adaptive experiment $\rho = 2$. 200 replications. The panel reports the out-of-sample regret of the method as a function of the number of iterations.

Table 2: Multiple-wave experiment. 200 replications. Relative improvement in welfare with respect to best competitor for $\rho = 2$. The panel at the top reports the out-of-sample regret and the one at the bottom the worst case in-sample regret across clusters.

| | Information | | | | Cash Transfer | | | |
|---|---|---|---|---|---|---|---|---|
| $T =$ | 5 | 10 | 15 | 20 | 5 | 10 | 15 | 20 |
| $n = 200$ | -0.026 | 0.014 | 0.043 | 0.033 | 0.295 | 0.390 | 0.528 | 0.322 |
| $n = 400$ | 0.0003 | 0.026 | 0.026 | 0.035 | 0.462 | 0.444 | 0.589 | 0.563 |
| $n = 600$ | 0.002 | 0.035 | 0.021 | 0.022 | 0.485 | 0.489 | 0.622 | 0.644 |
| | | | | | | | | |
| $n = 200$ | 1.103 | 1.370 | 1.451 | 1.447 | 0.254 | 0.276 | 0.305 | 0.323 |
| $n = 400$ | 1.400 | 1.616 | 1.667 | 1.626 | 0.282 | 0.329 | 0.367 | 0.379 |
| $n = 600$ | 1.546 | 1.771 | 1.828 | 1.751 | 0.279 | 0.335 | 0.364 | 0.368 |

# 6 Dynamic treatment effects: an overview

One assumption that we impose throughout the text is the absence of carry-overs. This section briefly discusses an extension with dynamic treatments. For simplicity, we omit covariates and assume that $X_i = 1$, and defer to Appendix A.3 formal details.

In the presence of carry-over effects, outcomes also depend on past treatments. For treatments assigned with exogenous parameters $(\beta_{k,1}, \cdots, \beta_{k,t})$ as in Definition 2.3, we let

$$Y_{i,t}^{(k)} = \Gamma(\beta_{k,t}, \beta_{k,t-1}) + \varepsilon_{i,t}^{(k)}, \quad \mathbb{E}_{\beta_{k,1:t}}\Big[\varepsilon_{i,t}^{(k)}\Big] = 0,$$

for some unknown $\Gamma(\cdot)$, $\varepsilon_{i,t}^{(k)}$. The components $\beta_{k,t}, \beta_{k,t-1}$ capture present and carry-over effects that result from individual and neighbors' treatments in the past two periods. We study the problem of estimating a path of treatment probabilities $(0, \beta_1, \cdots, \beta_T)$ from an experiment, where, in the first period, we assume for simplicity that none of the individuals is treated. We then implement this path on a new population without having access to the outcomes of such a new population. We maximize long-run welfare, defined as follows:

$$\mathcal{W}(\{\beta_s\}_{s=1}^{T^*}) = \sum_{t=1}^{T^*} q^t \Gamma(\beta_t, \beta_{t-1}), \quad \text{for a given horizon } T^*, \text{ and discounting factor } q < 1.$$

The long-run welfare defines the cumulative (discounted) welfare effect obtained from a certain sequence of decisions $(\beta_1, \beta_2, \cdots)$. Our goal is to find the sequence that maximizes the long-run welfare.[56] We start from the following observation: by the first-order conditions

$$\frac{\partial \Gamma(\beta_t, \beta_{t-1})}{\partial \beta_t} + q \frac{\partial \Gamma(\beta_{t+1}, \beta_t)}{\partial \beta_t} = 0, \quad \text{for all } t. \tag{20}$$

Equation (20) shows that the choice of the welfare-maximizing parameter $\beta_{t+1}$ may depend on the previous two decisions. Using ideas from reinforcement learning (Sutton and Barto, 2018), we parametrize future treatment probabilities based on past treatment probabilities as $\beta_{t+1} = h_\theta(\beta_t, \beta_{t-1}), \theta \in \Theta$, for some given function $h_\theta(\cdot)$, and find the parameter $\theta \in \Theta$, which maximizes welfare. The algorithm estimates the function $\Gamma(\cdot)$ using a single wave experiment and then maximizes

$$\widehat{\theta} \in \arg\max_{\theta \in \Theta} \sum_{t=1}^{T^*} q^t \widehat{\Gamma}(\beta_t, \beta_{t-1}), \quad \beta_t = h_\theta(\beta_{t-1}, \beta_{t-2}) \quad \forall t \geq 1, \quad \beta_0 = \beta_{-1} = 0.$$

Our main insight here is on how to design the experiment to estimate $\widehat{\Gamma}(\cdot)$. If $\widehat{\Gamma}$ is estimated with randomization based on a simple grid-search procedure, the rate of convergence of the regret would be $1/\sqrt{K}$ (see Appendix A.3). However, in Appendix A.3 we show that by using local perturbations – similarly to what was discussed in previous sections – and leveraging information from the estimated gradient, we can achieve a convergence rate of the out-of-sample regret of order $1/K$ (but not faster in $n$).[57] The single-wave procedure comes at a cost: the rate is specific to the one-dimensional setting and carry-overs over two consecutive periods; in $p$ dimensions, the rate would be of much slower order due to the curse of dimensionality. This is different from the adaptive design discussed in previous sections, where the dimension does not affect the rate in $K$, and it opens new research questions on the optimal design of experiments in the presence of carry-over effects.

# 7 Conclusions

This paper makes two main contributions. First, it introduces a single-wave experimental design to estimate the marginal effect of the policy and test for policy optimality. The experiment also allows us to identify and estimate treatment effects, which can be of independent interest. Second, it introduces an adaptive experiment to maximize welfare. We derive asymptotic properties for inference and provide a set of guarantees on the in-sample

[56] Alternatively, we may also estimate the same treatment probability each period (see Appendix B.4).

[57] The idea is as follows: we randomize probabilities $(\beta_1, \beta_2) \in [0, 1]^2$ from a coarse grid and use small groups (three) clusters to estimate the partial derivatives at each point. We extrapolate the value of $\Gamma(\cdot)$ throughout the set $[0, 1]^2$ with a first-order Taylor approximation around the closest point in the grid, using the information on the estimated marginal effect.

and out-of-sample regret. To our knowledge, this is the first paper to study inference on marginal effects and adaptive experimentation with unobserved interference.

In a single-wave experiment, we encourage researchers to identify and report estimates of the marginal effects. While the literature on welfare maximization often focuses on point estimates of the policy, we show that we can use the information on the marginal effects to conduct hypothesis testing on policy optimality and, ultimately, incorporate uncertainty in decision-making. Future research may explore notions of efficiency in this setting.

Our work opens new questions also from a theoretical perspective. The main assumption is that such clusters are observable before the experiment starts. We leave to future research to study properties when (i) such clusters are not fully disconnected, in the same spirit of Leung (2021); (ii) such clusters need to be estimated, similarly to graph-clustering procedures as in Ugander et al. (2013); (iii) clusters present different distributions, as we discuss in Appendix A.2. Similarly, it may be interesting to study the properties of our method, as the degree of interference is not bounded. This is theoretically possible, as illustrated in Theorem 3.1, and we leave its comprehensive analysis to future research. An open question is also how to combine policy learning with estimation procedures that impute the network (e.g., Alidaee et al., 2020; Breza et al., 2020).

Finally, an important assumption of this paper and, in general, of the literature on adaptive experimentations is that welfare is a function of observable characteristics. We leave to future research to study whether (and how) we may allow for unobserved utilities.

# References

Adusumilli, K., F. Geiecke, and C. Schilter (2019). Dynamically optimal treatment allocation using reinforcement learning. *arXiv preprint arXiv:1904.01047*.

Akbarpour, M., S. Malladi, and A. Saberi (2018). Just a few seeds more: value of network information for diffusion. *Available at SSRN 3062830*.

Alatas, V., A. Banerjee, A. G. Chandrasekhar, R. Hanna, and B. A. Olken (2016). Network structure and the aggregation of information: Theory and evidence from indonesia. *American Economic Review 106*(7), 1663–1704.

Alatas, V., A. Banerjee, R. Hanna, B. A. Olken, and J. Tobias (2012). Targeting the poor: evidence from a field experiment in indonesia. *American Economic Review 102*(4), 1206–40.

Alidaee, H., E. Auerbach, and M. P. Leung (2020). Recovering network structure from aggregated relational data using penalized regression. *arXiv preprint arXiv:2001.06052*.

Andrews, I., T. Kitagawa, and A. McCloskey (2019). Inference on winners. Technical report, National Bureau of Economic Research.

Armstrong, T. and S. Shen (2015). Inference on optimal treatment assignments.

Aronow, P. M. and C. Samii (2017). Estimating average causal effects under general interference, with application to a social network experiment. *The Annals of Applied Statistics 11*(4), 1912–1947.

Athey, S. and G. W. Imbens (2018). Design-based analysis in difference-in-differences settings with staggered adoption. Technical report, National Bureau of Economic Research.

Athey, S. and S. Wager (2021). Policy learning with observational data. *Econometrica 89*(1), 133–161.

Bai, Y. (2019). Optimality of matched-pair designs in randomized controlled trials. *Available at SSRN 3483834*.

Baird, S., J. A. Bohren, C. McIntosh, and B. Özler (2018). Optimal design of experiments in the presence of interference. *Review of Economics and Statistics 100*(5), 844–860.

Bakirov, N. K. and G. J. Szekely (2006). Student's t-test for gaussian scale mixtures. *Journal of Mathematical Sciences 139*(3), 6497–6505.

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

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

Basse, G. W. and E. M. Airoldi (2018b). Model-assisted design of experiments in the presence of network-correlated outcomes. *Biometrika 105*(4), 849–858.

Bhattacharya, D. and P. Dupas (2012). Inferring welfare maximizing treatment assignment under budget constraints. *Journal of Econometrics 167*(1), 168–196.

Bhattacharya, D., P. Dupas, and S. Kanaya (2013). Estimating the impact of means-tested subsidies under treatment externalities with application to anti-malarial bednets. Technical report, National Bureau of Economic Research.

Bloch, F., M. O. Jackson, and P. Tebaldi (2019). Centrality measures in networks. *Available at SSRN 2749124*.

Bottou, L., F. E. Curtis, and J. Nocedal (2018). Optimization methods for large-scale machine learning. *Siam Review 60*(2), 223–311.

Breza, E., A. G. Chandrasekhar, T. H. McCormick, and M. Pan (2020). Using aggregated relational data to feasibly identify network structure without network data. *American Economic Review 110*(8), 2454–84.

Brooks, R. L. (1941). On colouring the nodes of a network. In *Mathematical Proceedings of the Cambridge Philosophical Society*, Volume 37, pp. 194–197. Cambridge University Press.

Bubeck, S., N. Cesa-Bianchi, et al. (2012). Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends® in Machine Learning 5*(1), 1–122.

Bubeck, S., R. Munos, and G. Stoltz (2011). Pure exploration in finitely-armed and continuous-armed bandits. *Theoretical Computer Science 412*(19), 1832–1852.

Cai, J., A. De Janvry, and E. Sadoulet (2015). Social networks and the decision to insure. *American Economic Journal: Applied Economics 7*(2), 81–108.

Canay, I. A., J. P. Romano, and A. M. Shaikh (2017). Randomization tests under an approximate symmetry assumption. *Econometrica 85*(3), 1013–1030.

Carneiro, P., J. J. Heckman, and E. Vytlacil (2010). Evaluating marginal policy changes and the average effect of treatment for individuals at the margin. *Econometrica 78*(1), 377–394.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018). Double/debiased machine learning for treatment and structural parameters.

Chernozhukov, V., K. Wuthrich, and Y. Zhu (2018). Practical and robust $t$-test based inference for synthetic control and related methods. *arXiv preprint arXiv:1812.10820*.

Chetty, R. (2009). Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods. *Annu. Rev. Econ. 1*(1), 451–488.

Eckles, D., B. Karrer, and J. Ugander (2017). Design and analysis of experiments in networks: Reducing bias from interference. *Journal of Causal Inference 5*(1).

Egger, D., J. Haushofer, E. Miguel, P. Niehaus, and M. W. Walker (2019). General equilibrium effects of cash transfers: experimental evidence from kenya. Technical report, National Bureau of Economic Research.

Elliott, G. and R. P. Lieli (2013). Predicting binary outcomes. *Journal of Econometrics 174*(1), 15–26.

Flaxman, A. D., A. T. Kalai, and H. B. McMahan (2004). Online convex optimization in the bandit setting: gradient descent without a gradient. *arXiv preprint cs/0408007*.

Goldsmith-Pinkham, P. and G. W. Imbens (2013). Social networks and the identification of peer effects. *Journal of Business & Economic Statistics 31*(3), 253–264.

Graham, B. S., G. W. Imbens, and G. Ridder (2010). Measuring the effects of segregation in the presence of social spillovers: A nonparametric approach. Technical report, National Bureau of Economic Research.

Hadad, V., D. A. Hirshberg, R. Zhan, S. Wager, and S. Athey (2019). Confidence intervals for policy evaluation in adaptive experiments. *arXiv preprint arXiv:1911.02768*.

Hazan, E., K. Levy, and S. Shalev-Shwartz (2015). Beyond convexity: Stochastic quasi-convex optimization. In *Advances in Neural Information Processing Systems*, pp. 1594–1602.

Hirano, K. and J. R. Porter (2020). Asymptotic analysis of statistical decision rules in econometrics. *Handbook of Econometrics*.

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

Hu, Y., S. Li, and S. Wager (2021). Average treatment effects in the presence of interference. *arXiv preprint arXiv:2104.03802*.

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

Ibragimov, R. and U. K. Müller (2010). t-statistic based correlation and heterogeneity robust inference. *Journal of Business & Economic Statistics 28*(4), 453–468.

Imai, K., G. King, and C. Nall (2009). The essential role of pair matching in cluster-randomized experiments, with application to the mexican universal health insurance evaluation. *Statistical Science 24*(1), 29–53.

Imai, K. and M. L. Li (2019). Experimental evaluation of individualized treatment rules. *arXiv preprint arXiv:1905.05389*.

Jackson, M. O. and A. Wolinsky (1996). A strategic model of social and economic networks. *Journal of economic theory 71*(1), 44–74.

Jagadeesan, R., N. S. Pillai, A. Volfovsky, et al. (2020). Designs for estimating the treatment effect in networks with interference. *Annals of Statistics 48*(2), 679–712.

Janson, S. (2004). Large deviations for sums of partly dependent random variables. *Random Structures & Algorithms 24*(3), 234–248.

Johari, R., H. Li, I. Liskovich, and G. Weintraub (2020). Experimental design in two-sided platforms: An analysis of bias. *arXiv preprint arXiv:2002.05670*.

Karrer, B., L. Shi, M. Bhole, M. Goldman, T. Palmer, C. Gelman, M. Konutgan, and F. Sun (2021). Network experimentation at scale. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pp. 3106–3116.

Kasy, M. and A. Sautmann (2019). Adaptive treatment assignment in experiments for policy choice. *Econometrica*.

Kato, M. and Y. Kaneko (2020). Off-policy evaluation of bandit algorithm from dependent samples under batch update policy. *arXiv preprint arXiv:2010.13554*.

Kitagawa, T. and A. Tetenov (2018). Who should be treated? Empirical welfare maximization methods for treatment choice. *Econometrica 86*(2), 591–616.

Kitagawa, T. and G. Wang (2021). Who should get vaccinated? individualized allocation of vaccines over sir network. *Journal of Econometrics*.

Kleinberg, R. D. (2005). Nearly tight bounds for the continuum-armed bandit problem. In *Advances in Neural Information Processing Systems*, pp. 697–704.

Laber, E. B., D. J. Lizotte, M. Qian, W. E. Pelham, and S. A. Murphy (2014). Dynamic treatment regimes: Technical challenges and applications. *Electronic journal of statistics 8*(1), 1225.

Leung, M. P. (2019). Inference in models of discrete choice with social interactions using network data. *Available at SSRN 3446926*.

Leung, M. P. (2020). Treatment and spillover effects under network interference. *Review of Economics and Statistics 102*(2), 368–380.

Leung, M. P. (2021). Network cluster-robust inference. *arXiv preprint arXiv:2103.01470*.

Li, S. and S. Wager (2020). Random graph asymptotics for treatment effect estimation under network interference. *arXiv preprint arXiv:2007.13302*.

Lubold, S., A. G. Chandrasekhar, and T. H. McCormick (2020). Identifying the latent space geometry of network models through analysis of curvature. Technical report, National Bureau of Economic Research.

Manski (2004). Statistical treatment rules for heterogeneous populations. *Econometrica 72*(4), 1221–1246.

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

Mbakop, E. and M. Tabord-Meehan (2021). Model selection for treatment choice: Penalized welfare maximization. *Econometrica 89*(2), 825–848.

Muandet, K., K. Fukumizu, B. Sriperumbudur, and B. Schölkopf (2016). Kernel mean embedding of distributions: A review and beyond. *arXiv preprint arXiv:1605.09522*.

Muralidharan, K. and P. Niehaus (2017). Experimentation at scale. *Journal of Economic Perspectives 31*(4), 103–24.

Ogburn, E. L., O. Sofrygin, I. Diaz, and M. J. van der Laan (2017). Causal inference for social network data. *arXiv preprint arXiv:1705.08527*.

Pouget-Abadie, J. (2018). *Dealing with Interference on Experimentation Platforms.* Ph. D. thesis.

Rai, Y. (2018). Statistical inference for treatment assignment policies. *Unpublished Manuscript*.

Ross, N. et al. (2011). Fundamentals of stein's method. *Probability Surveys 8*, 210–293.

Russo, D., B. Van Roy, A. Kazerouni, I. Osband, and Z. Wen (2017). A tutorial on thompson sampling. *arXiv preprint arXiv:1707.02038*.

Saez, E. (2001). Using elasticities to derive optimal income tax rates. *The review of economic studies 68*(1), 205–229.

Sasaki, Y. and T. Ura (2020). Welfare analysis via marginal treatment effects. *arXiv preprint arXiv:2012.07624*.

Sävje, F., P. M. Aronow, and M. G. Hudgens (2021). Average treatment effects in the presence of unknown interference. *The Annals of Statistics 49*(2), 673–701.

Shamir, O. (2013). On the complexity of bandit and derivative-free stochastic convex optimization. In *Conference on Learning Theory*, pp. 3–24. PMLR.

Sriperumbudur, B. K., K. Fukumizu, A. Gretton, B. Schölkopf, G. R. Lanckriet, et al. (2012). On the empirical estimation of integral probability metrics. *Electronic Journal of Statistics 6*, 1550–1599.

Stoye, J. (2009). Minimax regret treatment choice with finite samples. *Journal of Econometrics 151*(1), 70–81.

Sutton, R. S. and A. G. Barto (2018). *Reinforcement learning: An introduction.* MIT press.

Tabord-Meehan, M. (2018). Stratification trees for adaptive randomization in randomized controlled trials. *arXiv preprint arXiv:1806.05127*.

Ugander, J., B. Karrer, L. Backstrom, and J. Kleinberg (2013). Graph cluster randomization: Network exposure to multiple universes. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 329–337. ACM.

Vazquez-Bare, G. (2017). Identification and estimation of spillover effects in randomized experiments. *arXiv preprint arXiv:1711.02745*.

Viviano, D. (2019). Policy targeting under network interference. *arXiv preprint arXiv:1906.10258*.

Viviano, D. (2020). Experimental design under network interference. *arXiv preprint arXiv:2003.08421*.

Wager, S. and K. Xu (2021). Experimenting in equilibrium. *Management Science*.

Wainwright, M. J. (2019). *High-dimensional statistics: A non-asymptotic viewpoint*, Volume 48. Cambridge University Press.

# Appendix to "Policy design in experiments with unknown interference"


**Appendix A Main extensions** **2**

A.1 Estimation with global interference . . . . . . . . . . . . . . . . . . . . . . . . . 2

A.2 Matching clusters with distributional embeddings . . . . . . . . . . . . . . . . 4

A.3 Policy choice with dynamic treatments . . . . . . . . . . . . . . . . . . . . . . 6

**Appendix B Additional extensions** **9**

B.1 Testing multiple entries of the marginal effect with one wave experiment . . 9

B.2 Regret bounds with quasi-concavity . . . . . . . . . . . . . . . . . . . . . . . 11

B.3 Non separable fixed effects . . . . . . . . . . . . . . . . . . . . . . . . . . . 12

B.4 Experimental design with dynamic treatments and stationary policies . . . . 12

B.5 Staggered adoption . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 13

**Appendix C Derivations** **14**

C.1 Notation and definitions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 14

C.2 Lemmas . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 15

C.3 Proof of the theorems in the main text . . . . . . . . . . . . . . . . . . . . . 24

C.4 Proofs for the extensions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 31

C.5 Proof of the corollaries . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 42

**Appendix D Numerical studies: additional results** **43**

D.1 One-wave experiment . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 43

D.2 Multiple-wave experiment . . . . . . . . . . . . . . . . . . . . . . . . . . . . 43

D.3 Calibrated experiment with covariates for cash transfers . . . . . . . . . . . . 43

D.4 Calibrated experiment for information diffusion to increase vaccination . . . 45

D.5 Additional figures . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 46

**Appendix E Additional Algorithms** **52**

# Appendix A Main extensions

This section contains three extensions. Additional extensions are included in Appendix B.

## A.1 Estimation with global interference

In this section, we relax the local dependency assumptions. The treatment affects each unit in a cluster $k$ through a global interference mechanism mediated by a random variable $p_t^{(k)}$. For example, we can think of $p_t^{(k)}$ as the average number of treated units in a cluster or to the adjusted price in a particular market or village due to the program. For simplicity, we consider the case where $X_i^{(k)} = 1$, i.e., the policy of interest is a global policy (e.g., the probability of treatment).

We discuss assumptions on the outcome model below.

**Assumption A.1** (Global interference)**.** Let treatments be assigned as in Definition 2.3 with exogenous vector of parameters $\beta_{k,1:t}$. Let

$$Y_{i,t}^{(k)} = \alpha_t + \tau_k + g\Big(p_t^{(k)}, \beta_{k,t}\Big) + \varepsilon_{i,t}^{(k)}, \quad \mathbb{E}_{\beta_{k,1:t}}\Big[\varepsilon_{i,t}^{(k)} | p_t^{(k)}\Big] = 0,$$

for some function $g(\cdot)$ unknown to the researcher, bounded and twice continuously differentiable with bounded derivatives, and unobservable $p_t^{(k)}$. Assume in addition that $\varepsilon_{i,t}^{(k)} \perp \varepsilon_{j \notin \mathcal{I}_i^{(k)},t}^{(k)} | \beta_{k,1:t}, p_t^{(k)}$ for some set $|\mathcal{I}_i^{(k)}| = \mathcal{O}(\gamma_N)$.

Assumption A.1 states that the outcome within each cluster is a function of a common factor, and treatment assignment rule $\beta_{k,t}$ plus unobservables centered around zero and locally dependent. The factor $p_t^{(k)}$ also depends on the treatment assignments of all individuals. This is formalized below.

**Assumption A.2** (Global interference component)**.** Let treatments be assigned as in Definition 2.3. Assume that

$$p_t^{(k)} = q(\beta_{k,t}) + o_p(\eta_n),$$

with $q(\beta)$ being unknown, bounded and twice continuously differentiable in $\beta$ with uniformly bounded derivatives.

Assumption A.2 states that the factor can be expressed as the sum of two components. The first component $q(\cdot)$ depends on the policy parameter $\beta_{k,t}$ assigned at time $t$ and on the distribution of covariates of all units in a cluster.[58] The second component is a stochastic

[58]Observe that we can equivalently relax Assumption A.2 and assume that $q(\beta, \cdot)$ depends on the empirical distribution of covariates and use basic concentration arguments (Wainwright, 2019) to show that asymptotically the two definitions are equivalent.

component that depends on the realized treatment effects. We illustrate an example below.

**Example A.1** (Within cluster average)**.** Suppose that

$$Y_{i,t}^{(k)} = t(\bar{D}_t^{(k)}, \nu_{i,t}), \quad \nu_{i,t}^{(k)} \sim_{i.i.d.} \mathcal{P}_\nu, \quad D_{i,t}^{(k)} \sim_{i.i.d.} \text{Bern}(\beta)$$

where $t(\cdot)$ is some arbitrary (smooth) function. Then $p_t^{(k)} = \bar{D}_t^{(k)}$ i.e., individuals depend on the average exposure in a cluster. We can write

$$Y_{i,t}^{(k)} = t(p_t^{(k)}, \nu_{i,t}^{(k)}) \quad \text{where} \quad p_t^{(k)} = \beta + \underbrace{(\bar{D}_t^{(k)} - \beta)}_{=O_p(n^{-1/2})},$$

which satisfies Assumption A.2 for $\eta_n = n^{-1/3}$ or larger. □

Example A.1 illustrates how we can accommodate global interference mechanism whenever individuals depend on statistics of treatment assignments, such as their average. Example A.1 does not allow for local spillovers between units but only for global interference.[59]

We are interested in the marginal effects defined below.

$$V_g(\beta) = \frac{\partial W_g(\beta)}{\partial \beta}, \quad W_g(\beta) = g(q(\beta), \beta).$$

Estimation of the marginal effect follows similarly to Equation (7). The following theorem guarantees consistency.

**Theorem A.1.** *Let Assumption A.1, A.2 hold with subgaussian $\varepsilon_{i,t}^{(k)}$, $X = 1$. Then for $\hat{V}_{(k,k+1)}$ estimated as in Algorithm 2, for $k$ being odd holds:*

$$\left|\widehat{V}_{(k,k+1)} - V_g(\beta)\right| = \mathcal{O}_p\Big(\sqrt{\frac{\gamma_N \log(\gamma_N)}{\eta_n^2 n}} + \eta_n\Big) + o_p(1).$$

The proof is in Appendix C.4.1. Theorem A.1 guarantees consistency of the estimated gradient. The experiment can be conducted similarly to what discussed in Section 4 and omitted for the sake of brevity.

[59] Assumption A.2 builds on the model of demand as a function of individual prices in Wager and Xu (2021). The difference is that we do not rely on a specific modeling assumption of market interactions. Instead, we model outcomes as functions of exposures in a certain cluster and exploit the two-clusters variation for consistent estimation, as we discuss below.

## A.2 Matching clusters with distributional embeddings

In this section, we turn to the problem of matching clusters, allowing for covariates having different distributions in different clusters. In particular, we assume that $X_i^{(k)} \sim_{i.i.d.} F_X^{(k)}$. The main distinction from previous sections is that $F_X^{(k)}$ is cluster-specific. The section works as follows: first, we characterize the bias of the difference in means estimators; second, we propose a matching algorithm that minimizes the worst-case bias.

We start from the simple setting with two clusters $k, k'$ only, and two periods $t \in \{0, 1\}$. Treatments are assigned as follows

$$\begin{aligned} t = 0: &\quad D_{i,0}^{(h)} \sim \pi(X_i^{(h)}; \beta_0), \quad h \in \{k, k'\} \\ t = 1: &\quad D_{i,1}^{(k)} \sim \pi(X_i^{(k)}; \beta), \quad D_{i,1}^{(k')} \sim \pi(X_i^{(k')}; \beta'). \end{aligned} \tag{A.1}$$

Namely, at time $t = 0$, treatments are assigned with a parameter $\beta_0$. At time $t = 1$ treatments are assigned with parameter $\beta$ in cluster $k$ and $\beta'$ in cluster $k'$.

The estimand of interest is the difference in the average effects in cluster $k$, formally

$$\omega_k = \int y(x; \beta) dF_X^{(k)}(x) - \int y(x; \beta') dF_X^{(k)}(x).$$

We study properties of the difference in differences estimator

$$\widehat{\omega}_k(k') = \Big[\bar{Y}_1^{(k)} - \bar{Y}_1^{(k')}\Big] - \Big[\bar{Y}_0^{(k)} - \bar{Y}_0^{(k')}\Big],$$

which defines a difference in differences between the two clusters over two consecutive periods.

Our focus is to control the bias of the estimator. This is defined in the following lemma.

**Lemma A.2.** *Let Assumption 2.1, 2.2, and treatments assigned as in Equation (A.1). Then*

$$\mathbb{E}[\widehat{\omega}_k(k')] - \omega_k = \int \Big(y(x; \beta') - y(x; \beta_0)\Big) d\Big(F_X^{(k)}(x) - F_X^{(k')}(x)\Big).$$

Lemma A.2 shows that the bias depends on the difference between the expectations averaged over two different distributions. Unfortunately, the bias is unknown since it depends on the function $y(\cdot)$, which is not identifiable with finitely many clusters. We therefore bound the worst-case error over a class of functions $x \mapsto [y(x; \beta') - y(x; \beta_0)] \in \mathcal{M}$, with $\mathcal{M}$ defined below. The proof follows directly from Lemma 2.1, and rearrangement.

We start by defining $\mathcal{M}$ be a reproducing kernel Hilbert space (RKHS) equipped with a norm $|| \cdot ||_{\mathcal{M}}$.[60] Without loss of generality, we study the worst-case functionals over the

[60] A RKHS is an Hilbert space of functions where all the evaluations functionals are bounded, namely,

unit-ball. Formally, we focus on bounding the worst-case error of the form[61]

$$\sup_{[y(\cdot;\beta')-y(\cdot;\beta_0)]\in\mathcal{M}:||y(\cdot;\beta')-y(\cdot;\beta_0)||_{\mathcal{M}}\leq 1}\Big|\omega_k-\mathbb{E}[\widehat{\omega}_k(k')]\Big| = \sup_{f\in\mathcal{M}:||f||_{\mathcal{M}}\leq 1}\Big\{\int f(x)d(F_X^{(k)}-F_X^{(k')})\Big\}. \tag{A.2}$$

The right-hand side is know as the maximum mean discrepancy (MMD), a measure of distances in RKHS (see Muandet et al., 2016, and references therein). It is known that the MMD can be consistently estimated using kernels. In particular, given a particular choice of a kernel $k(\cdot)$, which corresponds to a certain RKHS, we can estimate

$$\begin{aligned}\widehat{\text{MMD}}^2(k,k') &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i} h\Big(X_i^{(k)},X_i^{(k')},X_j^{(k)},X_j^{(k')}\Big),\\ h(x_i,y_i,x_j,y_j) &= k(x_i,x_j)+k(y_i,y_j)-k(x_i,y_j)-k(x_j,y_i).\end{aligned} \tag{A.3}$$

The estimator estimates the squared MMD. It only depends on the kernel function $k(\cdot)$ and hence can be easily constructed in a finite sample without requiring an explicit characterization of RKHS. Here, $\widehat{\text{MMD}}^2(k,k')\to_p \Big|\Big|\mu_{F_X^{(k)}}-\mu_{F_X^{(k')}}\Big|\Big|_{\mathcal{M}}^2$ (Sriperumbudur et al., 2012).

We now turn to the problem of matching clusters. We do so using the estimated MMD in Equation (A.3). We first note that the estimator $\widehat{\text{MMD}}^2(k,k')$ only depends on *pre-treatment* variables, and hence can be computed *before* treatments are assigned. As a result, given cluster $k$, we can match $k$ with the $k'\neq k$ having the smallest estimated MMD. Formally, the following matching algorithm is considered:

- construct
$$k'\in\arg\min_{k\neq k}\widehat{\text{MMD}}^2(k,k'). \tag{A.4}$$
based on the minimum estimated MMD in Equation (A.3).

- Randomize treatments as in Equation (A.1);

- Estimate $\widehat{\omega}_k(k')$.

When instead we want to match without replacement clusters, we can minimize some aggregate measures of error (e.g., the sum of estimated MMD across clusters).

where for each $f\in\mathcal{M}$, and $x\in\mathcal{X}$, $f(x)\leq C||f||_{\mathcal{M}}$ for a finite constant $C$. Intuitively, assuming that $[y(\cdot;\beta')-y(\cdot;\beta_0)]\in\mathcal{M}$ imposes smoothness conditions on the average effect as a function of $x$.

[61] Here Equation (A.2) follows directly from Lemma A.2 and the fact that the integral is a scalar.

## A.3 Policy choice with dynamic treatments

This section studies an experimental design when carry-overs occur. For simplicity, we omit covariates and assume that $X_i = 1$.

We start our discussion by introducing the dynamic model. For the sake of brevity, we directly impose a high-level condition on the outcome model.

**Assumption A.3** (Dynamic model)**.** For treatments assigned with exogenous parameters $(\beta_{k,1}, \cdots, \beta_{k,t})$ as in Definition 2.3, let the followig hold

$$Y_{i,t}^{(k)} = \Gamma(\beta_t, \beta_{t-1}) + \varepsilon_{i,t}^{(k)}, \quad \mathbb{E}_{\beta_{k,1:t}}\left[\varepsilon_{i,t}^{(k)}\right] = 0,$$

for some unknown $\Gamma(\cdot)$, $\varepsilon_{i,t}^{(k)}$.

The components $\beta_{k,t}, \beta_{k,t-1}$ capture present and carry-over effects that result from individual and neighbors' treatments in the past two periods. Here, we allow for both panels and repeated cross-sections. We study the problem of estimating a path of treatment probabilities $(0, \beta_1, \cdots, \beta_T)$ from an experiment, where, in the first period, we assume for simplicity that none of the individuals is treated. This path is then implemented on a new population without having access to the outcomes of such a new population.

We provide a simple example below.

**Example A.2.** Suppose that

$$Y_{i,t}^{(k)} = D_{i,t}^{(k)}\phi_1 + \frac{\sum_{j\neq i} A_{i,j}^{(k)} D_{i,t-1}}{\sum_{j\neq i} A_{i,j}^{(k)}}\phi_2 + \nu_{i,t}^{(k)}, \quad D_{i,t}^{(k)} \sim_{i.i.d.} \mathrm{Bern}(\beta_t).$$

That is, individuals depend on their present treatment assignment and on the treatment assignments of the neighbors in the previous period. Let $\nu_{i,t}$ be a zero-mean random variable. The expression simplifies to

$$Y_{i,t}^{(k)} = \beta_t\phi_1 + \beta_{t-1}\phi_2 + \varepsilon_{i,t}^{(k)}$$

where $\varepsilon_{i,t}^{(k)}$ is zero mean, and depends on neighbors' and individual assignments. □

We now define the long-run welfare.

**Definition A.1** (Long-run welfare)**.** Given an horizon $T^*$, define the long-run welfare as follows:

$$\mathcal{W}(\{\beta_s\}_{s=1}^{T^*}) = \sum_{t=1}^{T^*} q^t \Gamma(\beta_t, \beta_{t-1}),$$

for a known discounting factor $q < 1$, where $\beta_0 = 0$.

The long-run welfare deifines the cumulative (discounted) welfare obtained from a certain sequence of decisions $(\beta_1, \beta_2, \cdots)$.

Our goal is to maximize the long-run welfare. Whenever the optimal policy is stationary, i.e., every period we are interested in estimating the same treatment probability, we can apply results to previous sections to this case. This is illustrated in Appendix B.4.

If, instead, we are interested in estimating a *sequence* of decisions, as discussed in the main text, we parametrize future treatment probabilities based on past treatment probabilities as follows

$$\beta_{t+1} = h_\theta(\beta_t, \beta_{t-1}), \quad \theta \in \Theta.$$

The parametrization is imposed for computational convenience. For some arbitrary large $T^*$, the objective function takes the following form

$$\begin{aligned} \widetilde{W}(\theta) &= \sum_{t=1}^{T^*} q^t \Gamma\Big(\beta_t, \beta_{t-1}\Big), \\ \beta_t &= h_\theta(\beta_{t-1}, \beta_{t-2}) \quad \text{for all } t \geq 1, \quad \beta_0 = \beta_{-1} = 0. \end{aligned} \tag{A.5}$$

Here $\widetilde{W}(\theta)$ denotes the long-run welfare indexed by a given policy's parameter $\theta$. The objective function defines the discounted cumulative welfare induced by the policy $h_\theta$.

**Definition A.2** (Non-stationary policy decisions)**.** A non-stationary policy is defined as as a map $h_\theta : \mathcal{B} \times \mathcal{B}, \theta \in \Theta$. Define the non-stationary estimand as follows:

$$h_{\theta^*}(\cdot), \quad \theta^* \in \arg\max_{\theta \in \Theta} \widetilde{W}(\theta).$$

The algorithm estimates the function $\Gamma(\cdot)$ using a single wave experiment, i.e., we use a single period of experimentation. We then use the estimated function $\Gamma(\cdot)$ and its gradient for estimating the optimal policy.

The randomization and estimators are described in Algorithm E.2. We conduct the randomization using two periods of experimentations only. We partition the space $[0, 1]^2$ into a grid $\mathcal{G}$ of equally spaced components $(\beta_1^r, \beta_2^r)$ for each triad of clusters $r$. Within each triad, we induce small deviations to the parameters $\beta$. For each triad $r$, the algorithm returns

$$\widetilde{\Gamma}(\beta_2^r, \beta_1^r), \quad \widehat{g}_1(\beta_2^r, \beta_1^r), \quad \widehat{g}_2(\beta_2^r, \beta_1^r)$$

where the latter two components are the estimated partial derivatives of $\Gamma(\cdot)$, and $\widetilde{\Gamma}(\beta_2^r, \beta_1^r)$ is the within cluster average.

For each pair of parameters $(\beta_2, \beta_1)$, we estimate $\widehat{\Gamma}(\beta_2, \beta_1)$ as follows

$$\begin{aligned} &\widehat{\Gamma}(\beta_2, \beta_1) = \widetilde{\Gamma}(\beta_2^r, \beta_1^r) + \widehat{g}_2(\beta_2^r, \beta_1^r)(\beta_2 - \beta_2^r) + \widehat{g}_1(\beta_2^r, \beta_1^g)(\beta_1 - \beta_1^r), \\ &\text{where} \quad (\beta_1^r, \beta_2^r) = \arg \min_{(\tilde{\beta}_1, \tilde{\beta}_2) \in \mathcal{G}} \Big\{ ||\beta_1 - \tilde{\beta}_1||^2 + ||\beta_2 - \tilde{\beta}_2||^2 \Big\}. \end{aligned} \tag{A.6}$$

The idea is as follows: we estimate $\Gamma(\beta_2, \beta_1)$ at $(\beta_2, \beta_1)$ using a a first-order Taylor approximation around the closest pairs of parameters in the grid $\mathcal{G}$. Given $\widehat{\Gamma}$, we estimate the welfare-maximizing parameter[62]

$$\widehat{\theta} \in \arg\max_{\theta \in \Theta} \sum_{t=1}^{T^*} q^t \widehat{\Gamma}(\beta_t, \beta_{t-1}), \quad \beta_t = h_\theta(\beta_{t-1}, \beta_{t-2}) \quad \forall t \geq 1, \quad \beta_0 = \beta_{-1} = 0.$$

In the following theorem, we study the behaviour of $\widehat{\theta}$, in terms of out-of-sample regret.

**Theorem A.3** (Out-of-sample regret)**.** *Let Assumption A.3 hold. Let $X = 1$, and suppose that $\Gamma(\beta_2, \beta_1)$ is twice differentiable with bounded derivatives. Let treatments be assigned as in Algorithm E.2. Suppose in addition that $\varepsilon_{i,t}^{(k)} \perp \varepsilon_{j \not\in \mathcal{I}_i^{(k)}}^{(k)}$ where $|\mathcal{I}_i^{(k)}| \leq \gamma_N$, for some arbitrary $\gamma_N$ and $\varepsilon_{i,t}^{(k)}$ is sub-gaussian. Let $\gamma_N \log(\gamma_N)/(\eta_n^2 n) = o(1)$. Then*

$$\lim_{n \to \infty} P\Big( \sup_{\theta \in \Theta} \widetilde{W}(\theta) - W(\widehat{\theta}) \leq \frac{\bar{C}}{K} \Big) = 1$$

*for a constant $\bar{C}$ independent of $K$.*

The proof is in Appendix C.4.2. To our knowledge, Algorithm E.2 is novel to the literature of experimental design.[63]

Theorem A.3 shows that with probability converging to one as the size of each cluster increases, the regret scales at a rate $1/K$, linearly in the number of clusters. To gain further intuition on the derivation of the theorem, observe that we can bound

$$\sup_{\theta \in \Theta} \widetilde{W}(\theta) - \widetilde{W}(\widehat{\theta}) \leq 2 \sum_t q^t \times \underbrace{\sup_{(\beta_1, \beta_2) \in [0,1]^2} \Big| \widehat{\Gamma}(\beta_2, \beta_1) - \Gamma(\beta_2, \beta_1) \Big|}_{(A)}.$$

[62] Here, $\widehat{\theta}$ can be obtained using off-the-shelf algorithms. A simple example is running in-parallel multiple gradient descent algorithms initialized over different starting points and choosing the one which leads to the largest objective $\sum_{t=1}^{T^*} q^t \widehat{\Gamma}(\beta_t, \beta_{t-1})$.

[63] We note that optimal dynamic treatments have been studied in the literature on bio-statistics, see, e.g., Laber et al. (2014), while here we consider the different problem of the design of the experiment. Adusumilli et al. (2019) discuss off-line policy estimation in the presence of dynamic budget constraints with *i.i.d.* observations. The authors assume no carry-overs, and do not discuss the problem of experimental design.

To bound $(A)$ observe first that each element in the grid $\mathcal{G}$ has a distance of order $1/\sqrt{K}$ since the grid has two dimensions and $K/3$ components. As a result for any element $(\beta_2, \beta_1)$, we can write

$$\Gamma(\beta_2, \beta_1) = \underbrace{\Gamma(\beta_2^r, \beta_1^r)}_{(B)} + \underbrace{\frac{\partial \Gamma(\beta_2^r, \beta_1^r)}{\partial \beta_1^r}(\beta_1 - \beta_1^r) + \frac{\partial \Gamma(\beta_2^r, \beta_1^r)}{\partial \beta_2^r}(\beta_2 - \beta_2^r)}_{(C)} + \underbrace{\mathcal{O}\Big(||\beta_1 - \beta_1^r||^2 + ||\beta_2 - \beta_2^r||^2\Big)}_{(D)}$$

where $\beta^r \in \mathcal{G}$ is some value in the grid such that $(D)$ is of order $1/K$. We can then show that $(B) + (C) - \widehat{\Gamma}(\beta_2^r, \beta_1^r)$ converges in probability to zero as $n$ grows which is possible since we use the estimated gradient to construct $\widehat{\Gamma}$. If instead we had not used information on the estimated gradient, the rate would be dominated by $(C)$ which is of order $1/\sqrt{K}$.

However, we note that different from previous sections, the rate $1/K$ is specific to the one-dimensional setting and carry-overs over two consecutive periods. In $p$ dimensions, the rate would be of order $1/K^{2/(p+1)}$ due to the curse of dimensionality.

# Appendix B Additional extensions

## B.1 Testing multiple entries of the marginal effect with one wave experiment

In the following lines we extend Algorithm 2 to testing the following null

$$H_0 : V^{(j)}(\beta) = 0, \text{ for some } p \geq l \geq 1,$$

where we consider a generic number of dimensions tested $l$. We introduce the algorithmic procedure in Algorithm B.1.

We define $\mathcal{K}_j$ the set of pairs in Algorithm B.1 used to estimate the $j^{th}$ entry of $V(\beta)$. Define

$$\bar{V}_n^{(j)} = \frac{2l}{K} \sum_{k \in \mathcal{K}_j} \widehat{V}_k,$$

the average marginal effect for coordinate $j$ estimated from those clusters is used to estimate the effect of the $j^{th}$ coordinate. We construct

$$Q_{j,n} = \frac{\sqrt{K/(2l)}\bar{V}_n^{(j)}}{\sqrt{(K/(2l) - 1)^{-1} \sum_{k \in \mathcal{K}_j} (\widehat{V}_k^{(j)} - \bar{V}_n^{(j)})^2}}, \quad \mathcal{T}_n = \max_{j \in \{1, \cdots, l\}} |Q_{j,n}|, \tag{B.3}$$

where $\mathcal{T}_n$ denotes the test statistics. The choice of the $l$-infinity norm is motivated by its

**Algorithm B.1** One wave experiment for inference

**Require:** Value $\beta \in \mathbb{R}^p$, $K$ clusters, 2 periods of experimentation, number of tests $t$.

1: Match clusters into pairs $K/2$ pairs with consecutive indexes $\{k, k+1\}$;

2: $t = 0$ (*baseline*):

a: Treatments are assigned at some baseline $\beta_0$ $D_{i,0}^{(h)} \sim \pi(X_i^{(h)}, \beta_0), h \in \{1, \cdots, K\}$ (e.g., none of the individuals is treated).

b: Collect baseline values: for $n$ units in each cluster observe $Y_{i,0}^{(h)}, h \in \{1, \cdots, K\}$.

3: $t = 1$ (*experimentation-wave*)

4: Assign each pair of clusters $\{k, k+1\}$ to a coordinate $j \in \{1, \cdots, p\}$ (with the same number of pairs to each coordinate)

5: For each pair $\{k, k+1\}$, $k$ is odd, assigned to coordinate $j$

a: Randomize

$$D_{i,1}^{(h)} \sim \begin{cases} \pi(X_i^{(h)}, \beta + \eta_n \underline{e}_j) \text{ if } h = k \\ \pi(X_i^{(h)}, \beta - \eta_n \underline{e}_j) \text{ if } h = k+1 \end{cases}, \quad n^{-1/2} < \eta_n \leq n^{-1/4}$$

b: For $n$ units in each cluster $h \in \{k, k+1\}$ observe $Y_{i,1}^{(h)}$.

c: Estimate the marginal effect for coordinate $j$ as

$$\widehat{V}_k = \frac{1}{2\eta_n}\Big[\bar{Y}_1^{(k)} - \bar{Y}_0^{(k)}\Big] - \frac{1}{2\eta_n}\Big[\bar{Y}_1^{(k+1)} - \bar{Y}_0^{(k+1)}\Big] \tag{B.1}$$

**return**

$$\widetilde{V}_n = \Big[\widehat{V}_1, \widehat{V}_3, \cdots, \widehat{V}_{K-1}\Big] \tag{B.2}$$

theoretical properties: the statistics $Q_{j,n}$ follows an unknown distribution as a result of possibly heteroskedastic variances of $\widehat{V}_k$ across different clusters. However, the upper-bound on the critical quantiles of the proposed test-statistic for unknown variance attains a simple expression under the proposed test-statistics. From a conceptual stand-point, the proposed test-statistic is particularly suited when a large deviation occurs over one dimension of the vector.[64] We now introduce the following theorem.

**Theorem B.1** (Nominal coverage)**.** *Let Assumptions 2.1, 2.2, 3.2 hold. Let* $n^{1/4}\eta_n = o(1), \gamma_N^2/N^{1/4} = o(1)$, $K < \infty$. *Let* $K \geq 4l$, $H_0$ *be as defined in Equation (8). For any* $\alpha \leq 0.08$*,*

$$\lim_{n \to \infty} P\Big(\mathcal{T}_n \leq q_\alpha \Big| H_0\Big) \geq 1 - \alpha, \text{ where } q_\alpha = \mathrm{cv}_{K/(2l)-1}\Big(1 - (1-\alpha)^{1/l}\Big), \tag{B.4}$$

*with* $\mathrm{cv}_{K/(2l)-1}(h)$ *denotes the critical value of a two-sided t-test with level* $h$ *with test-statistic*

[64] Observe that alternatively, we may also consider randomization tests as discussed in Canay et al. (2017). This is omitted for the sake of brevity.

*having $K/(2l) - 1$ degrees of freedom.*

The proof is in Appendix C.4.3.

## B.2 Regret bounds with quasi-concavity

In the following lines, we provide guarantees on the regret bounds for the adaptive algorithm in Section 4 under quasi-concavity. We replace Assumption 4.2 with the following condition.

**Assumption B.1** (Local strong concavity and strict quasi-concavity)**.** Assume that the following conditions hold.

(A) For every $\beta, \beta' \in \mathcal{B}$, such that $W(\beta') - W(\beta) \geq 0$, then $V(\beta)^\top(\beta' - \beta) \geq 0$,

(B) For every $\beta \in \mathcal{B}$, $||V(\beta)||_2 \geq \mu||\beta - \beta^*||_2$, for a positive constant $\mu > 0$;

(C) $\frac{\partial^2 W(\beta)}{\partial \beta^2}\Big|_{\beta=\beta^*}$ has negative eigenvalues bounded away from zero at $\beta^*$, with $\beta^* \in \tilde{\mathcal{B}} \subset \mathcal{B}$ being in the interior of $\mathcal{B}$.

Condition (A) imposes a quasi-concavity of the objective function. The condition is *equivalent* to common definitions of quasi concavity (Hazan et al., 2015). Condition (A) holds when increasing the probability of treating more neighbors has decreasing marginal effects (see, e.g., Figure 5). Condition (B) assumes that the marginal effect only vanishes at the optimum, ruling out regions over which marginal effects remain constant at zero. A notion of strict quasi-concavity can be found in Hazan et al. (2015), where the authors assume (A) and that the gradient vanishes at the optimum only. Condition (C) also imposes that the function has negative definite Hessian at $\beta^*$ only but not necessarily globally. The above restrictions guarantee strong concavity locally at the optimum, but not necessarily globally. We now introduce out-of-sample guarantees in this setting. In such a case, the choice of the learning rate consists of a gradient norm rescaling, as discussed in Remark 7.

**Theorem B.2.** *Let Assumptions 2.1, 2.2, 4.1, B.1 hold, and choose $\alpha_{k,w}$ as in Equation (15), for arbitrary $v \in (0,1)$, $\kappa$ as defined in Equation (C.7), and $\epsilon_n$ as in Lemma C.12. Take a small $1/4 > \xi > 0$, and let $n^{1/4-\xi} \geq \bar{C}\sqrt{\log(n)p\gamma_N T^2 e^{B_p T}\log(KT)}$, $\eta_n = 1/n^{1/4+\xi}$, for finite constants $\infty > B_p, \bar{C} > 0$. Then for $T \geq \zeta^{1/v}$, for a finite constant $\zeta < \infty$, with probability at least $1 - 1/n$,*

$$W(\beta^*) - W(\hat{\beta}^*) = \mathcal{O}(\check{T}^{-1+v}).$$

The proof is in Appendix C.4.4.

## B.3 Non separable fixed effects

In the following lines, we show how we can leverage direct and marginal spillover effects to identify (and then estimate) the marginal effects when fixed effects are non-separable in time and cluster identity.

**Theorem B.3** (Marginal effects with non-separable fixed effects)**.** *Let $X = 1$, and suppose that $m(d, 1, \beta)$ is bounded and twice differentiable with bounded derivatives for $d \in \{0, 1\}$. Let Assumptions 2.1 hold. Suppose that fixed-effects are non-separable, with*

$$Y_{i,t}^{(k)} = m(D_{i,t}^{(k)}, 1, \beta) + \alpha_{k,t} + \varepsilon_{i,t}^{(k)}, \quad \mathbb{E}[\varepsilon_{i,t}^{(k)}] = 0, \quad D_{i,t}^{(k)} \sim_{i.i.d.} \mathrm{Bern}(\beta),$$

*and $m(1, 1, \beta)$ being a constant function in $\beta$. Then*

$$\mathbb{E}\Big[\hat{\Delta}_k(\beta) + \hat{S}(0, \beta)(1 - \beta) - (1 - \beta)\hat{S}(1, \beta)\Big] = V(\beta) + \mathcal{O}(\eta_n).$$

The proof is in Appendix C.4.5. Theorem B.3 shows that we can use the information on the spillover and direct treatment effects to identify the marginal effects in the presence of non-separable time and cluster fixed effects. The theorem leverages the assumption that spillovers only occur on the control individuals but not the treated. This is often the case in practice: for example, whenever an individual is vaccinated, whether her friends are vaccinated may have little effect on her outcome. Consistency follows similarly to what has been discussed in Theorem 3.1 and is omitted for the sake of brevity.

## B.4 Experimental design with dynamic treatments and stationary policies

Here, we return to the case of dynamic treatments of Section A.3 and study optimization with stationary decisions.

We note that in some applications, we may be interested in stationary estimands of the following form

$$\beta^{**} \in \arg\sup_{\beta \in \mathcal{B}} \Gamma(\beta, \beta). \tag{B.5}$$

Equation (B.5) defines the vector of parameters that maximizes welfare, *under the constraint* that the decision remains invariant over time. For this case, we propose the following algorithm.

**"Patient" gradient descent** The algorithm works as follows. We begin our iteration from the starting value $\beta_0$, we evaluate $\Gamma(\beta_0, \beta_0)$, and compute its *total* derivative $\nabla(\beta_0)$.

We then update the current policy choice in the direction of the total derivative and wait for one more iteration before making the next update. Formally, the first three iteration consists of the following updates:

$$\Gamma(\beta_0, \beta_0) \Rightarrow \Gamma(\beta_0 + \nabla(\beta_0), \beta_0) \Rightarrow \Gamma(\beta_0 + \nabla(\beta_0), \beta_0 + \nabla(\beta_0)).$$

**Estimation and updates** The estimation procedure follows similarly to Section 4, with a small modifications: for every period $t$ the policy stays enforced for one more period $t+1$, without necessitating that data are collected over the period $t+1$. At the end of all such iterations, we construct an estimator $\widehat{\beta}_T^{**}$ as in Section 4.

**Theorem B.4.** *Let Assumption 2.1, 4.1, A.3 hold. Take a small* $\xi > 0$*. Let* $n^{1/4-\xi} \geq \bar{C}\sqrt{\log(n)\gamma_N T^B e^{B\sqrt{p}T}\log(KT)}$*,* $\eta_n = 1/n^{1/4+\xi}$*, for finite constants* $\infty > B, \bar{C} > 0$*. Let* $T \geq \zeta$*, for a finite constant* $\zeta < \infty$*. Let* $\beta \mapsto \Gamma(\beta, \beta)$ *satisfying strict strong concavity. Then with probability at least* $1 - 1/n$*, for some* $J > 0$*,* $\alpha_{k,w} = J/w$

$$||\beta^{**} - \hat{\beta}_T^{**}||^2 = \mathcal{O}(T^{-1}).$$

The proof is in Appendix C.4.6.

## B.5 Staggered adoption

In this section, we sketch the experimental design in the presence of staggered adoption, i.e., when treatments are assigned only once to individuals and post-treatment outcomes are collected once. The algorithm works similarly to what was discussed in Section 4 with one small difference: every period, we only collect information from a given clusters' pair and update the policy for the subsequent pair and proceed in an iterative fashion.

The algorithm is included in Algorithm E.3. For simplicity, we only discuss the case where $\beta \in \mathbb{R}$ (single coordinate), while for the case where $\beta \in \mathbb{R}^p$ the algorithm works similarly with the only difference that every perturbation to each coordinate requires a new pair of clusters (with in total $2p$ times the number of iterations many clusters).

**Theorem B.5** (In-sample regret)**.** *Let the conditions in Theorem 4.3 hold and let* $\beta \in \mathbb{R}$*, with* $\check{\beta}^t$ *estimated as in Algorithm E.3. Then with probability at least* $1 - 1/n$*,*

$$\frac{1}{T}\sum_{t=1}^{T}\Big[W(\beta^*) - W(\check{\beta}^t)\Big] \leq \bar{C}\frac{p\log(T)}{T}$$

*for a finite constant* $\bar{C} < \infty$*.*

See Appendix C.4.7 for the proof.

# Appendix C Derivations

## C.1 Notation and definitions

First, we introduce conventions and notation. We define $x \lesssim y$ if $x \leq cy$ for a positive constant $c < \infty$. For $K$ many clusters, we say that

$$\lfloor k \rceil = \begin{cases} k \text{ if } k \leq K \\ k - K \text{ otherwise.} \end{cases}$$

We will refer to $\hat{V}_{(k,k+1)}$ as $\hat{V}_k$ for $k$ is odd for short of notation. Also, we define $\check{V}_{k,s} = \hat{V}_{\lfloor k+2 \rceil, s}$.

The following definition introduces the notion of a dependency graph (see also Janson 2004).

**Definition C.1** (Dependency graph)**.** For given random variables $R_1, \cdots, R_n$, $W_n \in \{0,1\}^{n\times n}$ is a non-random matrix defined as dependecy graph of $(R_1, \cdots, R_n)$ if, for any $i$, $R_i \perp R_{j:W_n^{(i,j)}=0}$. We denote the dependency neighbors $N_i = \{j : W_n^{(i,j)} = 1\}$. □

Intuitively, a dependency graph denotes a deterministic adjacency matrix with entry $(i,j)$ equal to one if $(i,j)$ are statistical dependent.

**Definition C.2.** (Proper Cover) Given an adjacency matrix $A_n$, with $n$ rows and columns, a family $\mathcal{C}_n = \{\mathcal{C}_n(j)\}_j$ of disjoint subsets of $\{1, \cdots, n\}$ is a proper cover of $A_n$ if $\cup_j \mathcal{C}_n(j) = \{1, \cdots, n\}$ and $\mathcal{C}_n(j)$ contains units such that for any pair of elements $\{i, k \in \mathcal{C}_n(j)\}$, then $A_n^{(i,k)} = 0$. □

Namely, a proper cover of $A_n$ defines a set of disjoint sets, where each disjoint set contains some indexes of units that are not neighbors in $A_n$. Note that a proper cover always exists, since, if $A_n$ is fully connected, then the number of disjoint sets is just $n$, one for each element.

The size of the smallest proper cover is the chromatic number, defined as $\chi(A_n)$.

**Definition C.3.** (Chromatic Number) The chromatic number $\chi(A_n)$, denotes the size of the smallest proper cover of $A_n$. □

In the following lines we define the oracle descent procedure absent of sampling error. Let $\beta \in \mathcal{B} = [\mathcal{B}_1, \mathcal{B}_2]^p$, where $\mathcal{B}_1, \mathcal{B}_2$ are finite. Also, let $P_{\mathcal{B}_1,\mathcal{B}_2}$ be the projection operator onto $\mathcal{B}$.

**Definition C.4** (Oracle gradient descent under strong concavity)**.** We define

$$\beta_w^{**} = P_{\mathcal{B}_1,\mathcal{B}_2}\Big[\beta_{w-1}^{**} + \alpha_{w-1}V(\beta_{w-1}^{**})\Big], \quad \beta_1^{**} = \beta_0, \tag{C.1}$$

with $\alpha_w = \frac{\eta}{w+1}$, equal for all clusters. $\square$

Note that in our proofs, we will refer to the general $p$-dimensional case for the multi-wave experiment, which uses $\check{T} = T/p$ waves. See Algorithm E.1.

## C.2 Lemmas

### C.2.1 Preliminary lemmas

**Lemma C.1.** *(Ross et al., 2011) Let $X_1, ..., X_n$ be random variables such that $\mathbb{E}[X_i^4] < \infty$, $\mathbb{E}[X_i] = 0$, $\sigma^2 = \mathrm{Var}(\sum_{i=1}^n X_i)$ and define $W = \sum_{i=1}^n X_i/\sigma$. Let the collection $(X_1, ..., X_n)$ have dependency neighborhoods $N_i$, $i = 1, ..., n$ and also define $D = \max_{1\leq i\leq n}|N_i|$. Then for $Z$ a standard normal random variable, we obtain*

$$d_W(W, Z) \leq \frac{D^2}{\sigma^3}\sum_{i=1}^n \mathbb{E}|X_i|^3 + \frac{\sqrt{28}D^{3/2}}{\sqrt{\pi}\sigma^2}\sqrt{\sum_{i=1}^n \mathbb{E}[X_i^4]}, \tag{C.2}$$

*where $d_W$ denotes the Wasserstein metric.*

**Lemma C.2.** *(Brook's Theorem, Brooks (1941)) For any connected undirected graph $G$ with maximum degree $\Delta$, the chromatic number of $G$ is at most $\Delta$ unless $G$ is a complete graph or an odd cycle, in which case the chromatic number is $\Delta + 1$.*

### C.2.2 Concentration for local dependency graphs

In the following lemma we study concentration of the average of locally dependent random variables (see also Janson 2004 for concentration with local dependency graphs).

**Lemma C.3** (Concentration for dependency graphs)**.** *Define $\{R_i\}_{i=1}^n$ sub-gaussian random variables, forming a dependency graph with adjacency matrix $A_n$ with maximum degree bounded by $\gamma_N$. Then with probability at least $1-\delta$, for any $\delta \in (0, 1)$,*

$$\Big|\frac{1}{n}\sum_{i=1}^n (R_i - \mathbb{E}[R_i])\Big| \leq \bar{C}\sqrt{\frac{\gamma_N \log(\gamma_N/\delta)}{n}}.$$

*for a finite constant $\bar{C} < \infty$.*

*Proof of Lemma C.3.* First, we construct a proper cover $\mathcal{C}_n$ as in Definition C.2, with chromatic number $\chi(A_n)$. We can write

$$\left|\frac{1}{n}\sum_{i=1}^{n}(R_i - \mathbb{E}[R_i])\right| \leq \sum_{\mathcal{C}_n(j)\in\mathcal{C}_n} \underbrace{\left|\frac{1}{n}\sum_{i\in\mathcal{C}_n(j)}(R_i - \mathbb{E}[R_i])\right|}_{(A)}.$$

Here, we sum over each subset of index $\mathcal{C}_n(j) \in \mathcal{C}_n$ in the proper cover, and then we sum over each element in the subset $\mathcal{C}_n(j)$. Observe now that by definition of the dependency graph, components in $(A)$ are mutually independent. Using the Chernoff's bound (Wainwright, 2019), we have that with probability at least $1-\delta$, for any $\delta \in (0,1)$

$$\left|\sum_{i\in\mathcal{C}_n(j)}(R_i - \mathbb{E}[R_i])\right| \leq \bar{C}\sqrt{|\mathcal{C}_n(j)|\log(1/\delta)},$$

for a finite constant $\bar{C} < \infty$, where $|\mathcal{C}_n(j)|$ denotes the number of elements in $\mathcal{C}_n(j)$. As a result, using the union bound, we obtain that with probability at least $1-\delta$, for any $\delta \in (0,1)$

$$\left|\frac{1}{n}\sum_{i=1}^{n}(R_i - \mathbb{E}[R_i])\right| \leq \underbrace{\frac{\bar{C}}{n}\sum_{\mathcal{C}_n(j)\in\mathcal{C}_n}\sqrt{|\mathcal{C}_n(j)|\log(\chi(A_n)/\delta)}}_{(B)}.$$

Using concavity of the square-root function, after multiplying and dividing (B) by $\chi(A_n)$, we have

$$\begin{aligned}(B) &\leq \frac{\bar{C}}{n}\chi(A_n)\sqrt{\frac{1}{\chi(A_n)}\sum_{\mathcal{C}_n(j)\in\mathcal{C}_n}|\mathcal{C}_n(j)|\log(\chi(A_n)/\delta)} \\ &= \frac{\bar{C}}{n}\sqrt{\chi(A_n)n\log(\chi(A_n)/\delta)}.\end{aligned}$$

The last equality follows by the definition of proper cover. The final result follows by Lemma C.2. □

### C.2.3 Proof of Lemma 2.1 and local dependence

Lemma 2.1 is stated as a corollary of Lemma C.4.

**Lemma C.4.** *Let Assumption 2.1, 2.2 hold. For treatment assigned as in Assumption 2.3 with exogenous parameter $\beta_{k,t}$ in cluster $k$ at time $t$, Lemma 2.1 hold. Also, $\varepsilon_{i,t}^{(k)} \perp \{\varepsilon_{j,t}^{(k)}\}_{j\notin\mathcal{I}_i^{(k)}}|\beta_{k,t}$ for a set $|\mathcal{I}_i^{(k)}| = \mathcal{O}(\gamma_N)$.*

*Proof of Lemma C.4.* Under Assumption 2.2, we can write for some function $g$,

$$r\Big(D_{i,t}^{(k)}, D_{j:A_{i,j}^{(k)}>0,t}^{(k)}, X_i^{(k)}, X_{j:A_{i,j}^{(k)}>0}^{(k)}, A_i^{(k)}, U_i^{(k)}, U_{j:A_{i,j}^{(k)}>0}^{(k)}, \nu_{i,t}^{(k)}\Big) = g(Z_{i,t}^{(k)}).$$

Here, $Z_{i,t}^{(k)}$ depends on $A_i^{(k)}$, i.e., the edges of individual $i$, and on unobservables and observables of all those individuals such that $A_{i,j}^{(k)} > 0$, namely,

$$Z_{i,t}^{(k)} = \Big[D_{i,t}^{(k)}, X_i^{(k)}, \nu_{i,t}^{(k)}, A_i^{(k)} \otimes \Big(X^{(k)}, U^{(k)}, D_t^{(k)}\Big)\Big].$$

Importantly, under Assumption 2.1, $A_i^{(k)}$ is a function of $\Big\{\Big[X_j^{(k)}, U_j^{(k)}\Big], j : 1\{i_k \leftrightarrow j_k\} = 1\Big\}$, only, and each entry depends on $(X_j, U_j, X_i, U_i)$ through the same function $f$ for each individual. What is important, is that $\sum_j 1\{i_k \leftrightarrow j_k\} = \gamma_N^{1/2}$ for each unit $i$. Therefore, for some function $\tilde{g}$ (which depends on $f$ in Assumption 2.1), we can equivalently write

$$Z_{i,t}^{(k)} = \tilde{g}(D_{i,t}^{(k)}, \nu_{i,t}^{(k)}, X_i^{(k)}, \tilde{Z}_{i,t}^{(k)}), \quad \tilde{Z}_{i,t}^{(k)} = \Big\{\Big[X_j^{(k)}, U_j^{(k)}, D_{j,t}^{(k)}\Big], j : 1\{i_k \leftrightarrow j_k\} = 1\Big\},$$

where $\tilde{Z}_{i,t}^{(k)}$ is the vector of $\Big[X_j^{(k)}, U_j^{(k)}, D_{j,t}^{(k)}\Big]$ of all individuals $j$ with $1\{i_k \leftrightarrow j_k\} = 1$.

Now, observe that since $(U_i^{(k)}, X_i^{(k)}) \sim_{i.i.d.} F_{X|U} F_U$, and $\{\nu_{i,t}\}$ are $i.i.d.$ conditionally on $U^{(k)}, X^{(k)}$ (Assumption 2.2) and treatments are randomized independently (Assumption 2.3), we have

$$\Big[X_j^{(k)}, U_j^{(k)}, \nu_{j,t}^{(k)}, D_{j,t}^{(k)}\Big]\Big|\beta_{k,t} \sim_{i.i.d.} \mathcal{D}(\beta_{k,t})$$

is $i.i.d$ with some distribution $\mathcal{D}(\beta_{k,t})$ which only depends on the exogenous coefficient $\beta_{k,t}$ governing the distribution of $D_{i,t}^{(k)}$ under Definition 2.3. As a result for $\beta_{k,t}$ being exogenous, Lemma 2.1 holds since $\sum_j 1\{i_k \leftrightarrow j_k\} = \gamma_N^{1/2}$ for all $i$, hence $\tilde{Z}_{i,t}$ are identically distributed across units $i$.

Similarly, also $\varepsilon_{i,t}^{(k)}|\beta_{k,t}$ is a measurable function of a vector $\Big[X_j^{(k)}, U_j^{(k)}, \nu_{j,t}^{(k)}, D_{j,t}^{(k)}\Big]_{j:1\{i_k \leftrightarrow j_k\}=1}$.[65] As a result since such vectors are independent conditional on $\beta_{k,t}$, $\varepsilon_{i,t}^{(k)}$ is mutally independent with $\varepsilon_{v,t}^{(k)}$ for all $v$ such that they do not share a common element $\Big[X_j^{(k)}, U_j^{(k)}, \nu_{j,t}^{(k)}, D_{j,t}^{(k)}\Big]$, that is, such that $\max_j 1\{i_k \leftrightarrow j_k\}1\{v_k \leftrightarrow j_k\} = 0$.

There are at most $\gamma_N^{1/2} + \gamma_N$ many of $\varepsilon_{v,t}^{(k)}$ which can share a common neighbor with $\varepsilon_{i,t}^{(k)}$ ($\gamma_N^{1/2}$ many neighhbors and $\gamma_N$ many neighbors of the neighbors), which concludes the proof. ∎

[65] Here for notational convenience convenience only, we are letting $1\{i_k \leftrightarrow i_k\} = 1$.

### C.2.4 Concentration of the average outcomes

In this subsection, we provide three auxiliary lemmas.

**Lemma C.5.** *Suppose that treatments are assigned as in Assumption 2.3 with*

$$
\begin{aligned}
&D_{i,0}^{(k)} \sim \pi(X_i^{(k)}, \beta_0), \quad D_{i,0}^{(k+1)} \sim \pi(X_i^{(k+1)}, \beta_0)\\
&D_{i,t}^{(k)} \sim \pi(X_i^{(k)}, \beta), \quad D_{i,t}^{(k+1)} \sim \pi(X_i^{(k+1)}, \beta')
\end{aligned}
$$

*with exogenous parameters $\beta_0, \beta, \beta'$. Let Assumption 2.1, 2.2 hold. Then with probability at least $1-\delta$, for any $\delta \in (0,1)$*

$$
\Big|\bar{Y}_t^{(k)} - \bar{Y}_t^{(k+1)} - \bar{Y}_0^{(k)} + \bar{Y}_0^{(k+1)} - \int (y(x,\beta) - y(x,\beta'))dF_X(x)\Big| = \mathcal{O}(\sqrt{\frac{\gamma_N \log(\gamma_N/\delta)}{n}}).
$$

*Proof of Lemma C.5.* First, note that by Lemma C.4, we can write

$$
\begin{aligned}
\mathbb{E}\Big[\bar{Y}_t^{(k)} - \bar{Y}_t^{(k+1)}\Big] &= \int (y(x,\beta) - y(x,\beta'))dF_X(x) + \tau_k - \tau_{k+1}\\
\mathbb{E}\Big[\bar{Y}_0^{(k)} - \bar{Y}_0^{(k+1)}\Big] &= \tau_k - \tau_{k+1}.
\end{aligned} \tag{C.3}
$$

In addition, by Lemma C.4, $\varepsilon_{i,t}^{(k)}$ (and so $Y_{i,t}^{(k)}$) form a dependency graph with maximum degree bounded by $\gamma_N$. The proof completes by invoking Lemma C.3. □

**Lemma C.6.** *Let $y(x,\beta)$ be twice differentiable with uniformly bounded derivatives for all $x \in \mathcal{X}, \beta \in \mathcal{B}$. Then for all $\beta \in \mathcal{B}$, where $\mathcal{B}$ is a compact space*

$$
\begin{aligned}
\int \Big[y(x, \beta + \eta_n \underline{e}_j) - y(x, \beta - \eta_n \underline{e}_j)\Big] dF_X(x) &= 2\eta_n \int \frac{\partial y(x,\beta)}{\partial \beta^j} dF_X(x) + \mathcal{O}(\eta_n^2)\\
&= 2\eta_n V^{(j)}(\beta) + \mathcal{O}(\eta_n^2).
\end{aligned}
$$

*Proof of Lemma C.6.* We can write

$$
\begin{aligned}
y(x, \beta + \eta_n \underline{e}_j) &= y(x,\beta) + \frac{\partial y(x,\beta)}{\partial \beta^j}\eta_n + \mathcal{O}(\eta_n^2)\\
y(x, \beta - \eta_n \underline{e}_j) &= y(x,\beta) - \frac{\partial y(x,\beta)}{\partial \beta^j}\eta_n + \mathcal{O}(\eta_n^2)
\end{aligned}
$$

from the mean-value theorem which guarantees that the first equality holds. The second equality holds by the dominated convergence theorem. □

**Lemma C.7.** *Let the conditions in Lemma C.5 hold. Let $y(x,\beta)$ be twice differentiable in $\beta$ with uniformly bounded derivatives for all $x \in \mathcal{X}, \beta \in \mathcal{B}$. Suppose that $\beta = \check{\beta} + \eta_n \underline{e}_j$ and*

$\beta' = \check{\beta} - \eta_n \underline{e}_j$. *Then with probability at least* $1-\delta$*, for any* $\delta \in (0,1)$

$$\Big|\frac{\bar{Y}_t^{(k)} - \bar{Y}_t^{(k+1)} - \bar{Y}_0^{(k)} + \bar{Y}_0^{(k+1)}}{2\eta_n} - V^{(j)}(\check{\beta})\Big| = \mathcal{O}\Big(\sqrt{\frac{\gamma_N \log(\gamma_N/\delta)}{\eta_n^2 n}} + \eta_n\Big)$$

*Proof of Lemma C.7.* Using Lemma C.5 and the triangular inequality, with probability at least $1-\delta$,

$$\begin{aligned}
&\Big|\frac{\bar{Y}_t^{(k)} - \bar{Y}_t^{(k+1)} - \bar{Y}_0^{(k)} + \bar{Y}_0^{(k+1)}}{2\eta_n} - V^{(j)}(\check{\beta})\Big| \\
&\leq \Big|\frac{1}{2\eta_n}\int (y(x, \check{\beta} + \eta_n \underline{e}_j) - y(x, \check{\beta} - \eta_n \underline{e}_j)) dF_X(x) - V^{(j)}(\check{\beta})\Big| + \mathcal{O}\Big(\sqrt{\frac{\gamma_N \log(\gamma_N/\delta)}{\eta_n^2 n}}\Big).
\end{aligned}$$

The component

$$\int (y(x, \check{\beta} + \eta_n \underline{e}_j) - y(x, \check{\beta} - \eta_n \underline{e}_j)) dF_X(x) = 2\eta_n V^{(j)}(\check{\beta}) + \mathcal{O}(\eta_n^2),$$

by Lemma C.6. □

### C.2.5 Proof of Lemma 4.1

To prove the claim it suffices to show that $\check{\beta}_k^w$ is independent of potential outcomes and covariates in cluster $k$ for all $w \in \{1, \cdots, \check{T}\}$, since $\beta_{k,t}$ is a deterministic function of some coefficient $\check{\beta}_k^w$ (see Algorithm E.1). Take $k$ to be odd. To show that the claim holds it suffices to show that $\check{\beta}_k^w$ is a function of observables and unobservables only of those units in clusters $k' \notin \{k, k+1\}$. The recursive claim that we want to prove is the following: for all $w$, $\check{\beta}_k^w$ is exogenous with respect to potential outcomes and covariates in clusters with index $\{h > \lfloor k + 2w + 1\rceil$ or $h \in \{k, k+1\}\}$. Clearly, for $\check{\beta}_k^1$ the lemma holds, since $\check{\beta}_k^1$ depends on the gradient in the pair $\{\lfloor k+2\rceil, \lfloor k+3\rceil\}$ only. Suppose that the lemma holds for all $w \leq \check{T} - 1$. Then consider $\check{\beta}_k^{\check{T}}$. Observe that $\check{\beta}_k^{\check{T}}$ is chosen based on the gradient $\hat{V}_{k+2,\check{T}-1}$ estimated in the previous wave in clusters $\{\lfloor k+2\rceil, \lfloor k+3\rceil\}$, and $\check{\beta}_k^{\check{T}-1}$. By the recursive algorithm, $\check{\beta}_k^{\check{T}-1}$ is exogenous with respect to covariates and potential outcomes in clusters with index $\{h > \lfloor k + 2\check{T} - 1\rceil$ or $h \in \{k, k+1\}\}$, which is possible since $K \geq 2\check{T}$, hence $\lfloor k + 2\check{T} - 1\rceil < k$. We only need to prove exogeneity of $\hat{V}_{k+2,\check{T}-1}$. The gradient estimated $\hat{V}_{k+2,\check{T}-1}$ is a function of the unobservables and observables at any time $t \leq T$ (where $T = \check{T}p$) in clusters $\{\lfloor k+2\rceil, \lfloor k+3\rceil\}$ and the policy $\check{\beta}_{k+2}^{\check{T}-1}$. Since $K \geq 2\check{T}$, again by the recursive algorithm $\check{\beta}_{k+2}^{\check{T}-1}$ is exogenous with respect to potential outcomes and covariates in clusters with index $\{h \geq \lfloor k + 2\check{T}\rceil$ or $h \in \{k, k+1\}\}$ which completes the proof.

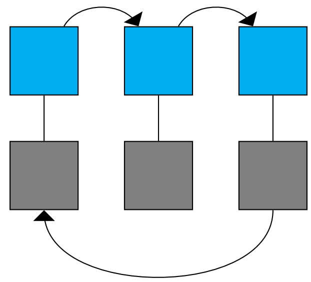

Figure C.1: Idea of the proof. Let $p = 1$. Since we have three clusters pairs (each pair of boxes), by assumption $T = 2$. Then the treatments at $T = 2$ in the first pair are assigned using information from the second pair at $T = 1$. Treatments in the second pair at $T = 1$, depend on information at $T = 0$ in the third pair. Hence, the parameter used at $T = 2$ in the first pair must be independent of covariates and potential outcomes in the first pair of clusters. The same reasoning applies to the other pairs of clusters.

### C.2.6 Lemmas for the adaptive experiment with strong concavity

In this section, we discuss theoretical guarantees of the algorithm, assuming the global strong concavity of the objective function $W(\beta)$. The following lemma follows by standard properties of the gradient descent algorithm (Bottou et al., 2018).

**Lemma C.8.** *For the learning rate as $\alpha_w = \eta/(w+1)$, and $\beta_w^{**}$ as defined in Equation (C.6), under Assumption 3.1, 4.1, 4.2, with $\sigma$-strong concavity, for $M \geq \eta \geq 1/\sigma$, for any $M \in [1/\sigma, \infty)$, and let $L = \max\{2(\mathcal{B}_2 - \mathcal{B}_1)^2, G^2 M^2\}$, with $G = \sup_\beta ||\frac{\partial W(\beta)}{\partial \beta}||_\infty$. Then the following holds:*

$$||\beta_w^{**} - \beta^*||^2 \leq \frac{Lp}{w}$$

*for a constant $L < \infty$.*

*Proof of Lemma C.8.* The proof follows standard arguments of the gradient descent method (Bottou et al., 2018), where, here, we leverage strong concavity and the assumption that the gradient is uniformly bounded. Denote $\beta^*$ the estimand of interest and recall the definition of $\beta_w^{**}$ in Equation (C.6). We define $\nabla_{w-1}$ the gradient evaluated at $\beta_{w-1}^{**}$. From strong concavity, we can write

$$W(\beta^*) - W(\beta_w^{**}) \leq \frac{\partial W(\beta_w^{**})}{\partial \beta}(\beta^* - \beta_w^{**}) - \frac{\sigma}{2}||\beta^* - \beta_w^{**}||_2^2$$
$$W(\beta_w^{**}) - W(\beta^*) \leq \frac{\partial W(\beta^*)}{\partial \beta}(\beta_w^{**} - \beta^*) - \frac{\sigma}{2}||\beta^* - \beta_w^{**}||_2^2.$$

As a result, since $\frac{\partial W(\beta^*)}{\partial \beta} = 0$, we have

$$\Big(\frac{\partial W(\beta^*)}{\partial \beta} - \frac{\partial W(\beta^{**}_w)}{\partial \beta}\Big)(\beta^* - \beta^{**}_w) = \frac{\partial W(\beta^{**}_w)}{\partial \beta}(\beta^* - \beta^{**}_w) \geq \sigma ||\beta^{**}_w - \beta^*||_2^2. \tag{C.4}$$

In addition, we can write:

$$||\beta^{**}_w - \beta^*||_2^2 = ||\beta^* - P_{\mathcal{B}_1,\mathcal{B}_2}(\beta^{**}_w + \alpha_{w-1}\nabla_{w-1})||_2^2 \leq ||\beta^* - \beta^{**}_w - \alpha_{w-1}\nabla_{w-1}||_2^2$$

where the last inequality follows from the fact that $\beta^* \in [\mathcal{B}_1, \mathcal{B}_2]^p$. Observe that we have

$$||\beta^* - \beta^{**}_w||_2^2 \leq ||\beta^* - \beta^{**}_{w-1}||_2^2 - 2\alpha_{w-1}\nabla_{w-1}(\beta^* - \beta^{**}_{w-1}) + \alpha^2_{w-1}||\nabla_{w-1}||_2^2.$$

Using Equation (C.4), we can write

$$||\beta^{**}_{w+1} - \beta^*||_2^2 \leq (1 - 2\sigma\alpha_w)||\beta^{**}_w - \beta^*||_2^2 + \alpha_w^2 G^2 p.$$

We now prove the statement by induction. Clearly, at time $w = 1$, the statement trivially holds. Consider a general time $w$. Then using the induction argument, we write

$$\begin{aligned} ||\beta^{**}_{w+1} - \beta^*||_2^2 &\leq (1 - 2\frac{1}{w+1})\frac{Lp}{w} + \frac{Lp}{(w+1)^2} \leq (1 - 2\frac{1}{w+1})\frac{Lp}{w} + \frac{Lp}{w(w+1)} \\ &= (1 - \frac{1}{w+1})\frac{Lp}{w} = \frac{Lp}{w+1}. \end{aligned}$$

□

**Lemma C.9.** *Let Assumption 2.2, 2.1, 4.1 hold. Let $\alpha_w$ be as defined in Lemma C.8. Then with probability at least $1-\delta$, for any $\delta \in (0,1)$, for all $w \geq 1$,*

$$\Big|\Big|P_{\mathcal{B}_1,\mathcal{B}_2-\eta_n}\Big[\sum_{s=1}^{w}\alpha_s \check{V}_{k,s}\Big] - P_{\mathcal{B}_1,\mathcal{B}_2}\Big[\sum_{s=1}^{w}\alpha_s V(\beta^{**}_s)\Big]\Big|\Big|_\infty = \mathcal{O}(P_{\check{T}}(\delta))$$

*where $P_1(\delta) = \alpha_1 \times \mathrm{err}(\delta)$ and $P_w(\delta) = B_p\alpha_w P_{w-1}(\delta) + P_{w-1}(\delta) + \alpha_w \mathrm{err}_w(\delta)$, for a finite constant $B_p < \infty$, and $\mathrm{err}_w(\delta) = \mathcal{O}\Big(\sqrt{\gamma_N \frac{\log(p\check{T}K/\delta)}{\eta_n^2 n}} + p\eta_n\Big)$.*

*Proof of Lemma C.9.* Recall that by Lemma C.7 we can write for every $k$ and $w \in \{1, \cdots, \check{T}\}$ (here using the union bound),

$$\check{V}^{(j)}_{k,w} = V^{(j)}(\check{\beta}^w_{k+2}) + \mathcal{O}\Big(\sqrt{\gamma_N \frac{\log(K\check{T}/\delta)}{\eta_n^2 n}} + \eta_n\Big).$$

We now proceed by induction. We first prove the statement, assuming that the constraint is always attained. We then discuss the case of the constraint not being attained. Define (where we suppress the dependence with $p$ for simplicity)

$$B = p \sup_{\beta} \Big|\Big| \frac{\partial^2 W(\beta)}{\partial \beta^2} \Big|\Big|_\infty.$$

**Unconstrained case** Consider $w = 1$. Then since we initialize parameters at $\beta_0$ (recall that $\beta_0 = \beta_1^{**}$), for all clusters, we can write with probability $1-\delta$, for any $\delta \in (0,1)$,

$$\Big|\Big| \alpha_1 \check{V}_{k,1} - \alpha_1 V(\beta_0) \Big|\Big|_\infty = \alpha_1 \mathrm{err}(\delta).$$

Consider $t = 2$, then we obtain for every $j \in \{1, \cdots, p\}$,

$$\alpha_2 \check{V}_{k,2}^{(j)} = \alpha_2 V^{(j)}(\check{\beta}_{k+2}^2) + \alpha_2 \mathrm{err}(\delta) = \alpha_2 V^{(j)}(\beta_1^{**} + \alpha_1 V(\beta_1^{**}) + \alpha_1 \check{V}_{k,w} - \alpha_1 V(\beta_1^{**})) + \alpha_2 \mathrm{err}(\delta).$$

Using the mean value theorem and Assumption 3.1, we obtain

$$\begin{aligned}
&\Big|\Big| \alpha_2 \check{V}_{k,2} - \alpha_2 V(\beta_2^{**}) \Big|\Big|_\infty \leq \alpha_2 \mathrm{err}(\delta) + B\alpha_2\alpha_1 \mathrm{err}(\delta) \\
&\qquad \Rightarrow \Big|\Big| \sum_{w=1}^2 \alpha_w \check{V}_{k,w} - \sum_{w=1}^2 \alpha_w V(\beta_w^{**}) \Big|\Big|_\infty \leq \alpha_2 \mathrm{err}(\delta) + B\alpha_2\alpha_1 \mathrm{err}(\delta) + \alpha_1 \mathrm{err}(\delta).
\end{aligned}$$

Consider now a general $w$. Then we can write with probability $1-\delta$, for any $\delta \in (0,1)$,

$$\alpha_w \check{V}_{k,w} = \alpha_w V(\check{\beta}_{k+2}^{w-1}) + \alpha_w \mathrm{err}(\delta).$$

Let $\tilde{P}_w^{(j)}(\delta) = \alpha_w \tilde{P}_{w-1}^{(j)}(\delta) + \tilde{P}_{w-1}^{(j)}(\delta) + \alpha_w \mathrm{err}(\delta)$, with $\tilde{P}_1^{(j)}(\delta) = \alpha_1 \mathrm{err}(\delta)$, the cumulative error for the $jth$ coordinate. Then, recursively, we have (here, $\tilde{P}_{w-1}(\delta)$ is the vector of cumulative errors)

$$\alpha_w \check{V}_{k,w} = \alpha_w V(\beta_w^{**} + \tilde{P}_{w-1}(\delta)) + \alpha_w \mathrm{err}(\delta).$$

Using the mean value theorem and Assumption 3.1, we obtain

$$\alpha_w \check{V}_{k,w} = \alpha_w V(\beta_w^{**}) + \alpha_w B \max_j \tilde{P}_{w-1}^{(j)}(\delta) + \alpha_w \mathrm{err}(\delta).$$

Therefore, with probability $1 - w\tilde{\delta}$ (using the union bound)

$$\Big|\Big|\sum_{s=1}^{w}\alpha_s\check{V}_{k,s} - \sum_{s=1}^{w}\alpha_s V(\beta_s^{**})\Big|\Big|_\infty \leq \Big|\Big|\alpha_w\check{V}_{k,w} - \alpha_w V(\beta_w^{**})\Big|\Big|_\infty + \Big|\Big|\sum_{s=1}^{w-1}\alpha_s\check{V}_{k,s} - \sum_{s=1}^{w-1}\alpha_s V(\beta_s^{**})\Big|\Big|_\infty$$
$$\leq \alpha_w B P_{w-1}(\tilde{\delta}) + \alpha_w \mathrm{err}(\tilde{\delta}) + P_{w-1}(\tilde{\delta}),$$

where $P_{w-1}(\tilde{\delta})$ captures the largest cumulative error up-to iteration $w-1$ as defined in the statement of the lemma (the log-term as a function of $p$ follows from the union bound). The proof completes once we write $\delta = \tilde{\delta}/w$.

**Constrained case** Since the statement is true for $w = 1$, we can assume that it is true for all $s \leq w-1$ and prove the statement by induction. Since $\mathcal{B}$ is a compact space, we can write,

$$\Big|\Big|P_{\mathcal{B}_1,\mathcal{B}_2-\eta_n}\Big[\sum_{s=1}^{w}\alpha_s\check{V}_{k,s}\Big] - P_{\mathcal{B}_1,\mathcal{B}_2}\Big[\sum_{s=1}^{w}\alpha_s V(\beta_s^{**})\Big]\Big|\Big|_\infty$$
$$\leq \Big|\Big|P_{\mathcal{B}_1,\mathcal{B}_2-\eta_n}\Big[\sum_{s=1}^{w}\alpha_s\check{V}_{k,s}\Big] - P_{\mathcal{B}_1,\mathcal{B}_2-\eta_n}\Big[\sum_{s=1}^{w}\alpha_s V(\beta_s^{**})\Big]\Big|\Big|_\infty + \mathcal{O}(p\eta_n)$$
$$\leq 2\Big|\Big|\sum_{s=1}^{w}\alpha_s\check{V}_{k,s} - \sum_{s=1}^{w}\alpha_s V(\beta_s^{**})\Big|\Big|_\infty + \mathcal{O}(p\eta_n)$$

completing the proof. □

**Lemma C.10.** *Let the conditions in Lemma C.9 hold. Then with probability at least $1-\delta$, for any $\delta \in (0,1)$, for all $w \geq 1, k \in \{1, \cdots, K\}$,*

$$||\beta^* - \check{\beta}_k^w||_2^2 \leq \frac{Lp}{w} + pw^{2B_p} \times \mathcal{O}\Big(\gamma_N \frac{\log(p\check{T}K/\delta)}{\eta_n^2 n} + p^2\eta_n^2\Big),$$

*for finite constants $B_p, L < \infty$.*

*Proof.* Using the triangular inequality, we can write

$$||\beta^* - \check{\beta}_k^w||_2^2 \leq ||\beta^* - \beta_w^{**}||_2^2 + ||\check{\beta}_k^w - \beta_w^{**}||_2^2.$$

The first component on the right-hand side is bounded by Lemma C.8. Using Lemma C.9, we bound the second component with probability at least $1-\delta$, as follows

$$||\check{\beta}_k^w - \beta_w^{**}||_2^2 \leq p||\check{\beta}_k^w - \beta_w^{**}||_\infty^2 = p \times \mathcal{O}(P_w^2(\delta)).$$

We conclude the proof by explicitely characterizing the rate of $P_w(\delta)$ as defined in Lemma

C.9. Following Lemma C.9, we can define recursively $P_w(\delta)$ for any $1 \leq w \leq \check{T}$ (recall that $\alpha_w \propto 1/w$) as

$$P_w(\delta) = (1 + \frac{B}{w})P_{w-1}(\delta) + \frac{1}{w}\mathrm{err}_n(\delta), \quad P_1(\delta) = \mathrm{err}_n(\delta).$$

where $\mathrm{err}_n = \mathcal{O}(\sqrt{\gamma_N \frac{\log(p\check{T}K/\delta)}{\eta_n^2 n}} + p\eta_n)$. Take, without loss of generality, $B \geq 1$ (if $B < 1$), we can find an upper bound with a different $B = 1$. Substituting recursively each term, we can write[66]

$$P_w(\delta) \leq \mathrm{err}_n(\delta) \sum_{s=1}^{w} \frac{1}{s} \prod_{j=s}^{w} (\frac{B}{j} + 1).$$

We now write

$$\begin{aligned} \sum_{s=1}^{w} \frac{1}{s} \prod_{j=s}^{w} (\frac{B}{j} + 1) &\leq \sum_{s=1}^{w} \frac{1}{s} \exp(\sum_{j=s}^{w} \frac{B}{j}) \leq \sum_{s=1}^{w} \frac{1}{s} \exp\Big(1 + B\log(w) - B\log(s)\Big) \\ &\lesssim \sum_{s=1}^{w} \frac{1}{s^2} e^{B\log(w)+1} \lesssim w^B, \end{aligned}$$

completing the proof.

□

## C.3 Proof of the theorems in the main text

### C.3.1 Proof of Theorem 3.1

First observe that for any $\delta \in (0,1)$,

$$\mathbb{E}\Big[\hat{V}_k(\beta)\Big] = V^{(1)}(\beta) + \mathcal{O}(\eta_n), \quad P\Big(\Big|\hat{V}_k(\beta) - V^{(1)}(\beta)\Big| > \mathcal{O}\Big(\eta_n + \sqrt{\frac{\gamma_N \log(\gamma_N/\delta)}{n\eta_n^2}}\Big)\Big) \leq \delta,$$

with the proof of the first claim follows similarly as in the proof of Lemma C.6 and the second claim being a direct corollary of Lemma C.7. Finally observe that with probability at least $1-\delta$, for any $\delta \in (0,1)$, we also have

$$\Big|\widehat{V}_k(\beta) - V^{(1)}(\beta)\Big| = \mathcal{O}(\eta_n) + \mathcal{O}\Big(\sqrt{\frac{\rho_n}{\delta n \eta_n^2}}\Big),$$

by Chebishev inequality and the triangular inequality.

[66] The expression is $\leq$ instead of $=$ since the first term in the expression $\mathrm{err}_n(\delta)$ multiplies by $(B/w+1)$ instead of just 1.

### C.3.2 Proof of Theorem 3.2

Consider Algorithm 2 for a generic coordinate $j$. Let $\beta$ be the target parameter as in Algorithm 2. By Lemma C.6, we have

$$\mathbb{E}[\widehat{V}_k^{(j)}] = V^{(j)}(\beta) + \mathcal{O}(\eta_n).$$

We have

$$\frac{\widehat{V}_k^{(j)} - \mathbb{E}[\widehat{V}_k^{(j)}]}{\sqrt{\mathrm{Var}(\widehat{V}_k^{(j)})}} = \frac{\widehat{V}_k^{(j)} - V^{(j)}(\beta)}{\sqrt{\mathrm{Var}(\widehat{V}_k^{(j)})}} + \mathcal{O}\Big(\frac{\eta_n}{\sqrt{\mathrm{Var}(\widehat{V}_k^{(j)})}}\Big).$$

Observe that under Assumption 3.2,

$$\mathcal{O}\Big(\frac{\eta_n}{\sqrt{\mathrm{Var}(\widehat{V}_k^{(j)})}}\Big) \leq \mathcal{O}(\eta_n^2 \times \sqrt{n}).$$

First, observe that by Lemma C.4, and the fact that covariates are independent, then $Y_{i,t}^{(k)} - Y_{i,0}^{(k)}$ form a locally dependent graph of maximum degree of order $\mathcal{O}(\gamma_N)$. We now invoke Lemma C.1. We can write

$$\begin{aligned}
&d_W\Big(\frac{1}{2\eta_n\sqrt{\mathrm{Var}(\widehat{V}_k^{(j)})}}\Big[\bar{Y}_t^{(k)} - \bar{Y}_0^{(k)}\Big] - \frac{1}{2\eta_n\sqrt{\mathrm{Var}(\widehat{V}_k^{(j)})}}\Big[\bar{Y}_t^{(k+1)} - \bar{Y}_0^{(k+1)}\Big], \mathcal{G}\Big) \\
&\leq \underbrace{\frac{\gamma_N^2}{\sigma^3}\sum_{h\in\{k,k+1\}}\sum_{i=1}^n\Big[\mathbb{E}\Big|\frac{Y_{i,t}^{(k)} - Y_{i,0}^{(k)}}{\eta_n n}\Big|^3\Big]}_{(A)} + \underbrace{\frac{\sqrt{28}\gamma_N^{3/2}}{\sqrt{\pi}\sigma^2}\sqrt{\sum_{i=1}^n\Big[\mathbb{E}\Big|\frac{Y_{i,t}^{(k)} - Y_{i,0}^{(k)}}{\eta_n n}\Big|^4\Big]}}_{(B)}, \\
&\mathcal{G} \sim \mathcal{N}(0,1), \quad \sigma^2 = \mathrm{Var}\Big(\frac{1}{2\eta_n}\Big[\bar{Y}_t^{(k)} - \bar{Y}_0^{(k)}\Big] - \frac{1}{2\eta_n}\Big[\bar{Y}_t^{(k+1)} - \bar{Y}_0^{(k+1)}\Big]\Big)
\end{aligned}$$

and $d_W$ denotes the Wasserstein metric. We now inspect each argument on the right hand side. Under Assumption 3.2, $\sigma^2 \geq C_k \frac{1}{n\eta_n^2}$ for a constant $C_k > 0$, and the third and fourth moment are bounded. Hence, we have for a constant $C' < \infty$,

$$(A) \leq C' \frac{\gamma_N^2}{n^3\eta_n^3} \times n^{5/2}\eta_n^3 \asymp \frac{\gamma_N^2}{n^{1/2}} \to 0.$$

Similarly, for (B), we have

$$(B) \leq c' \frac{\gamma_N^{3/2} n\eta_n^2}{\eta_n^2 n^{3/2}} \asymp \frac{\gamma_N^{3/2}}{n^{1/2}} \to 0.$$

The proof completes.

### C.3.3 Proof of Theorem 3.3

By Lemma 2.1, we can write (we omit the superscript $k$ from $X^{(k)}$ for sake of brevity)

$$
\begin{aligned}
&\mathbb{E}\Big\{\frac{1}{2n}\sum_{i=1}^{n}\Big[\frac{D_{i,1}^{(k+1)}Y_{i,1}^{(k+1)}}{\pi(X_i,\beta+\eta_n\underline{e}_1)}-\frac{(1-D_{i,1}^{(k+1)})Y_{i,1}^{(k+1)}}{1-\pi(X_i,\beta+\eta_n\underline{e}_1)}\Big]+\frac{1}{2n}\sum_{i=1}^{n}\Big[\frac{D_{i,1}Y_{i,1}^{(k)}}{\pi(X_i,\beta-\eta_n\underline{e}_1)}-\frac{(1-D_{i,1}^{(k)})Y_{i,1}^{(k)}}{1-\pi(X_i,\beta-\eta_n\underline{e}_1)}\Big]\Big\}\\
&=\frac{1}{2}\mathbb{E}\Big\{\Big[\frac{D_{i,1}^{(k+1)}Y_{i,1}^{(k+1)}}{\pi(X_i,\beta+\eta_n\underline{e}_1)}-\frac{(1-D_{i,1}^{(k+1)})Y_{i,1}^{(k+1)}}{1-\pi(X_i,\beta+\eta_n\underline{e}_1)}\Big]+\frac{1}{2n}\sum_{i=1}^{n}\Big[\frac{D_{i,1}Y_{i,1}^{(k)}}{\pi(X_i,\beta-\eta_n\underline{e}_1)}-\frac{(1-D_{i,1}^{(k)})Y_{i,1}^{(k)}}{1-\pi(X_i,\beta-\eta_n\underline{e}_1)}\Big]\Big\}\\
&=\underbrace{\frac{1}{2}\int\Big[m(1,x,\beta+\eta_n\underline{e}_1)-m(0,x,\beta+\eta_n\underline{e}_1)+m(1,x,\beta-\eta_n\underline{e}_1)-m(0,x,\beta-\eta_n\underline{e}_1)\Big]dF_X(x)}_{(i)}.
\end{aligned}
$$

The last equality follows from Lemma 2.1 and exogeneity of $\beta$. Doing a Taylor expansion to each component around $\beta$, we obtain

$$
\begin{aligned}
(i)=&\int\Big[m(1,x,\beta)-m(0,x,\beta)+\frac{\partial m(1,x,\beta)}{2\partial\beta^1}\eta_n-\frac{\partial m(0,x,\beta)}{2\partial\beta^1}\eta_n-\frac{\partial m(1,x,\beta)}{2\partial\beta^1}\eta_n+\frac{\partial m(0,x,\beta)}{2\partial\beta^1}\eta_n\Big]dF_X(x)\\
&+O(\eta_n^2)=\int\Big[m(1,x,\beta)-m(0,x,\beta)\Big]dF_X(x)+O(\eta_n^2),
\end{aligned}
$$

which completes the proof, since $\eta_n=o(n^{-1/4})$.

### C.3.4 Proof of Theorem 3.4

We are interested in studying

$$
\mathbb{E}\Big\{\frac{1}{2n}\sum_{h\in\{k,k+1\}}\frac{v_h}{\eta_n}\sum_{i=1}^{n}\Big[\frac{Y_{i,1}^{(h)}(1-D_{i,1}^{(h)})}{1-\pi(X_i^{(h)},\beta+v_h\eta_n\underline{e}_1)}-\bar{Y}_0^{(h)}\Big]\Big\},\quad v_h=\begin{cases}1 \text{ if } h=k\\ -1 \text{ otherwise.}\end{cases}
$$

Using Lemma 2.1, similarly to the derivation of Lemma C.6, we can write the above expression equal to

$$
\frac{1}{2\eta_n}\int[m(0,x,\beta+\eta_n\underline{e}_1)-m(0,x,\beta-\eta_n\underline{e}_1)]dF_X(x).
$$

Note that from the mean value theorem, and Assumption 3.1

$$
m(0,x,\beta+\eta_n\underline{e}_1)-m(0,x,\beta-\eta_n\underline{e}_1)=m(0,x,\beta)-m(0,x,\beta)+2\frac{\partial m(0,x,\beta)}{\partial\beta^1}\eta_n+O(\eta_n^2)
$$

which completes the proof.

### C.3.5 Proof of Theorem 4.2

Consider Lemma C.10 where we choose $\delta = 1/n$. We can write for each $k$

$$||\beta^* - \check{\beta}_k^{\check{T}}||_2^2 \leq \frac{pL}{\check{T}} + \mathcal{O}(1/\check{T}),$$

for a finite constant $L < \infty$, since, under the conditions for $n$ stated in the theorem, for finite $B$, the second component is $\mathcal{O}(1/\check{T})$. Note that

$$||\beta^* - \frac{1}{K}\sum_{k=1}^{K} \check{\beta}_k^{\check{T}}||_2^2 \leq \frac{1}{K}\sum_{k=1}^{K} ||\beta^* - \check{\beta}_k^{\check{T}}||_2^2$$

by Jensen's inequality, which completes the proof.

### C.3.6 Theorem 4.3

By the mean value theorem and Assumption 3.1, we have

$$\sum_{w=1}^{\check{T}} W(\beta^*) - W(\check{\beta}_k^w) \leq \bar{C} \sum_{w=1}^{\check{T}} ||\beta^* - \check{\beta}_k^w||_2^2,$$

for a finite constant $\bar{C} < \infty$, since $\frac{\partial W(\beta^*)}{\partial \beta} = 0$, and the Hessian is uniformly bounded (Assumption 3.1). By Lemma C.10, choosing $\delta = 1/n$, and for $n$ satisfying the conditions in Theorem 4.3, it follows that for all $k$,

$$\sum_{w=1}^{\check{T}} W(\beta^*) - W(\check{\beta}_k^w) \leq \sum_{w=1}^{\check{T}} \frac{p\kappa'}{w} \lesssim p\log(\check{T})$$

for $\kappa' < \infty$ being a finite constant. The proof completes.

### C.3.7 Proof of Theorem 4.4

Recall that from Assumption 2.1, the maximum degree is $\gamma_N^{1/2}$. We break the proof into several steps. We will write the model dividing by $1/\sum_{j\neq i} A_{i,j} 1\{X_j = x\}$ instead of $1/\max\{\sum_{j\neq i} A_{i,j} 1\{X_j = x\}, 1\}$ for notational convenience, but implicitely consider the expression $1/\sum_{j\neq i} A_{i,j} 1\{X_j = x\}$ equal to one if $\sum_{j\neq i} A_{i,j} 1\{X_j = x\} = 0$.

**Upper bound on $W_N^*$** We first provide an upper bound on the largest achievable welfare. Recall that $\Delta = c$. Therefore, we can write

$$W_N^* \leq \frac{1}{N}\sum_{i=1}^{N} \sup_{\mathcal{P}} \mathbb{E}\Big[\mathbb{E}_{D\sim\mathcal{P}(A,X)}\Big[s\Big(\Big[\frac{\sum_{j\neq i} A_{i,j}D_j 1\{X_j = x\}}{\sum_{j\neq i} A_{i,j} 1\{X_j = x\}}\Big]_{x\in\{1,\cdots,|\mathcal{X}|\}}\Big)\Big|A, X\Big]\Big].$$

Let

$$\beta^G = \arg\max_{\beta_1,\cdots,\beta_{|\mathcal{X}|}\in[0,1]^{|\mathcal{X}|}} s\Big(\beta_1,\cdots,\beta_{|\mathcal{X}|}\Big).$$

Note that since $D_j \in \{0,1\}$, we can write

$$\sup_{\mathcal{P}} \mathbb{E}\Big[\mathbb{E}_{D\sim\mathcal{P}(A,X)}\Big[s\Big(\Big[\frac{\sum_{j\neq i} A_{i,j}D_j 1\{X_j = x\}}{\sum_{j\neq i} A_{i,j} 1\{X_j = x\}}\Big]_{x\in\{1,\cdots,|\mathcal{X}|\}}\Big)\Big|A, X\Big]\Big] \leq s\Big(\beta_1^G,\cdots,\beta_{|\mathcal{X}|}^G\Big).$$

**Lower bound on $W(\beta^*)$** Using the fact that $\mathcal{B} = [0,1]^{|\mathcal{X}|}$, we can write[67]

$$\begin{aligned} W(\beta^*) &= \max_{\beta\in[0,1]^{|\mathcal{X}|}} \mathbb{E}_\beta\Big[s\Big(\Big[\frac{\sum_{j\neq i} A_{i,j}D_j 1\{X_j = x\}}{\sum_{j\neq i} A_{i,j} 1\{X_j = x\}}\Big]_{x\in\{1,\cdots,|\mathcal{X}|\}}\Big)\Big] \\ &\geq \mathbb{E}_{\beta^G}\Big[s\Big(\Big[\frac{\sum_{j\neq i} A_{i,j}D_j 1\{X_j = x\}}{\sum_{j\neq i} A_{i,j} 1\{X_j = x\}}\Big]_{x\in\{1,\cdots,|\mathcal{X}|\}}\Big)\Big], \end{aligned}$$

where we use the fact that $\beta^G = (\beta_1^G,\cdots,\beta_{|\mathcal{X}|}^G) \in [0,1]^{|\mathcal{X}|}$, and $\Delta(\cdot) = c(\cdot)$. From the mean value theorem

$$\begin{aligned} &\mathbb{E}_{\beta^G}\Big[s\Big(\Big[\frac{\sum_{j\neq i} A_{i,j}D_j 1\{X_j = x\}}{\sum_{j\neq i} A_{i,j} 1\{X_j = x\}}\Big]_{x\in\{1,\cdots,|\mathcal{X}|\}}\Big)\Big] \\ &= s(\beta^G) + \mathbb{E}_{\beta^G}\Big\{\frac{\partial s(\beta)}{\partial\beta}\Big|_{\beta\in\Big[\Big[\frac{\sum_{j\neq i} A_{i,j}D_j 1\{X_j=x\}}{\sum_{j\neq i} A_{i,j} 1\{X_j=x\}}\Big]_{x\in\{1,\cdots,|\mathcal{X}|\}},\beta^G\Big]} \times \\ &\times \Big(\Big[\frac{\sum_{j\neq i} A_{i,j}D_j 1\{X_j = x\}}{\sum_{j\neq i} A_{i,j} 1\{X_j = x\}}\Big]_{x\in\{1,\cdots,|\mathcal{X}|\}} - \beta^G\Big)\Big\}, \end{aligned}$$

where $\frac{\partial s(\cdot)}{\partial\beta}$ is evaluated at a point between the shared of treated neighbors and $\beta^G$.

**Bound on the difference** Combining the two bounds, we can write

$$W_N^* - W(\beta^*) \leq$$
$$\underbrace{\Big|\mathbb{E}_{\beta^G}\Big\{\frac{\partial s(\beta)}{\partial\beta}\Big|_{\beta\in\Big[\Big[\frac{\sum_{j\neq i} A_{i,j}D_j 1\{X_j=x\}}{\sum_{j\neq i} A_{i,j} 1\{X_j=x\}}\Big]_{x\in\{1,\cdots,|\mathcal{X}|\}},\beta^G\Big]} \times \Big(\Big[\frac{\sum_{j\neq i} A_{i,j}D_j 1\{X_j = x\}}{\sum_{j\neq i} A_{i,j} 1\{X_j = x\}}\Big]_{x\in\{1,\cdots,|\mathcal{X}|\}} - \beta^G\Big)\Big\}\Big|}_{(I)},$$

[67] $\mathbb{E}_\beta\Big[s\Big(\Big[\frac{\sum_{j\neq i} A_{i,j}D_j 1\{X_j=x\}}{\sum_{j\neq i} A_{i,j} 1\{X_j=x\}}\Big]_{x\in\{1,\cdots,|\mathcal{X}|\}}\Big)\Big]$ does not depend on $i$ similarly to Lemma 2.1.

where we took the absolute value in the last equation.

**Bound with Cauchy-Schwarz** We can now bound $(I)$ as follows.

$$(I) \leq \sup_{\beta} \Big|\Big|\frac{\partial s(\beta)}{\partial \beta}\Big|\Big|_2 \times |\mathcal{X}| \max_{x \in \mathcal{X}} \underbrace{\sqrt{\mathbb{E}_{\beta^G}\Big[\Big(\frac{\sum_{j\neq i} A_{i,j} D_j 1\{X_j = x\}}{\sum_{j\neq i} A_{i,j} 1\{X_j = x\}} - \beta_x^G\Big)^2\Big]}}_{(II)},$$

where we first used Cauchy-Schwarz and then bound the first component by the supremum over $\beta, x$ and the second component by the largest term over $x \in \mathcal{X}$ times the number of elements $|\mathcal{X}|$.

**Bound for** $(II)$ we now want to bound $(II)$. We do so fixing $X_j = x$ and show that for all $x \in \mathcal{X}$ (and so also for the maximum) we can obtain a useful bound. Recall that here $\mathbb{E}_{\beta^G}$ indicates that $D_{i,t}|X_i^{(k)} = x \sim_{i.i.d.} \mathrm{Bern}(\beta_x^G)$. Now, note that

$$\mathbb{E}_{\beta^G}\Big[\frac{\sum_{j\neq i} A_{i,j} D_j 1\{X_j = x\}}{\sum_{j\neq i} A_{i,j} 1\{X_j = x\}}\Big| X^{(k)}, A^{(k)}\Big] = \beta_x^G.$$

Also, note that since we take the maximum over $x \in \mathcal{X}$, here $\beta_x$ is fixed, and therefore, $\mathrm{Var}(\beta_x) = 0$. Therefore, we can write

$$\mathbb{E}_{\beta^G}\Big[\Big(\frac{\sum_{j\neq i} A_{i,j} D_j 1\{X_j = x\}}{\sum_{j\neq i} A_{i,j} 1\{X_j = x\}} - \beta_x^G\Big)^2\Big] = \mathbb{E}_{\beta^G}\Big[\mathrm{Var}\Big(\frac{\sum_{j\neq i} A_{i,j} D_j 1\{X_j = x\}}{\sum_{j\neq i} A_{i,j} 1\{X_j = x\}}\Big| X^{(k)}, A^{(k)}\Big)\Big].$$

In addition,

$$\mathrm{Var}_{\beta^G}\Big(\frac{\sum_{j\neq i} A_{i,j} D_j 1\{X_j = x\}}{\sum_{j\neq i} A_{i,j} 1\{X_j = x\}}\Big| X^{(k)}, A^{(k)}\Big) = \beta_x^G(1 - \beta_x^G)\Big/ \sum_{j\neq i} A_{i,j} 1\{X_j = x\},$$

since treatments are independent conditional on $X^{(k)}$, independent of $A^{(k)}$ conditional on $X^{(k)}$ by construction, and binary. Let $\kappa' = \kappa P(X = x)$, where, without loss of generality, $P(X = x) > 0$ since $|\mathcal{X}| < \infty$, and hence, if $P(X = x) = 0$, we can re-index $s(\cdot)$, for all types except $X = x$ and conduct the same analysis as above, without the case $X = x$. Note that

$$\mathbb{E}\Big[\beta_x^G(1-\beta_x^G)\Big/\sum_{j\neq i}A_{i,j}1\{X_j=x\}\Big] = \beta_x^G(1-\beta_x^G)\mathbb{E}\Big[1\Big/\sum_{j\neq i}A_{i,j}1\{X_j=x\}\Big]$$
$$\leq \beta_x^G(1-\beta_x^G)P\Big(\sum_{j\neq i}A_{i,j}1\{X_j=x\}<\kappa'\gamma_N^{1/4}\Big) + \beta_x^G(1-\beta_x^G)P\Big(\sum_{j\neq i}A_{i,j}1\{X_j=x\}\geq\kappa'\gamma_N^{1/4}\Big)\frac{1}{\kappa'\gamma_N^{1/4}}$$
$$\leq \beta_x^G(1-\beta_x^G)\underbrace{P\Big(\sum_{j\neq i}A_{i,j}1\{X_j=x\}<\kappa'\gamma_N^{1/4}\Big)}_{(III)}+\frac{1}{\kappa'\gamma_N^{1/4}},$$

where in the second inequality we used the fact that we are implicitely using in our notation that $1/\sum_{j\neq i}A_{i,j}1\{X_j=x\}$ equals one if $\sum_{j\neq i}A_{i,j}1\{X_j=x\}=0$ for notational convenience.

**Bound for** $(III)$ We are left to derive a bound for $(III)$, since the second term converges to zero as $\gamma_N\to\infty$. Define $h_x(X_i,U_i)=P(X=x)\int l(X_i,U_i,x,u)dF_{U|X=x}(u)$. Note that (recall that $1\{i\leftrightarrow j\}$ are fixed)

$$\mathbb{E}[A_{i,j}1\{X_j=x\}|X_i,U_i]=\mathbb{E}[l(X_i,U_i,X_j,U_j)1\{X_j=x\}1\{i\leftrightarrow j\}|X_i,U_i]=h_x(X_i,U_i)1\{i\leftrightarrow j\},$$

since, conditional on $(X_i,U_i)$,the indicator $1\{i\leftrightarrow j\}$ is fixed (exogenous), and $(X_i,U_i)\sim_{i.i.d.} F_XF_{U|X}$. Also, recall that $\sum_j 1\{i\leftrightarrow j\}=\gamma_N^{1/2}$. Hence, only $\gamma_N^{1/2}$ many edges of $i$ can at most be non-zero, while the remaining ones are zero almost surely. Therefore, using Hoeffding's inequality (Wainwright, 2019), and using independence conditional on $X_i,U_i$,

$$P\Big(\Big|\frac{1}{\gamma_N^{1/2}}\sum_{j\neq i}A_{i,j}1\{X_j=x\}-h_x(X_i,U_i)\Big|\leq\bar{C}\sqrt{\frac{\log(2\gamma_N)}{\gamma_N^{1/2}}}\Big|X_i,U_i\Big)\geq 1-1/\gamma_N, \tag{C.5}$$

for a finite constant $\bar{C}<\infty$. Observe that $h_x(X_i,U_i)\geq\kappa'>0,\kappa'=P(X=x)\kappa$ almost surely by assumption. Define the event

$$\mathcal{E}=\Big\{|\sum_{j\neq i}A_{i,j}1\{X_j=x\}-\gamma_N^{1/2}h_x(X_i,U_i)|\leq\bar{C}\sqrt{\log(2\gamma_N)\gamma_N^{1/2}}\Big\}.$$

We can write

$$\begin{aligned}
P\Big(\sum_{j\neq i} A_{i,j}1\{X_j = x\} < \kappa'\gamma_N^{1/4}\Big) &= P\Big(\sum_{j\neq i} A_{i,j}1\{X_j = x\} - \gamma_N^{1/2}h_x(X_i,U_i) + \gamma_N^{1/2}h_x(X_i,U_i) < \kappa'\gamma_N^{1/4}\Big) \\
&\leq P\Big(\gamma_N^{1/2}h_x(X_i,U_i) < \kappa'\gamma_N^{1/4} + |\sum_{j\neq i} A_{i,j}1\{X_j = x\} - \gamma_N^{1/2}h_x(X_i,U_i)|\Big) \\
&\leq P\Big(\gamma_N^{1/2}h_x(X_i,U_i) < \kappa\gamma_N^{1/4} + |\sum_{j\neq i} A_{i,j}1\{X_j = x\} - \gamma_N^{1/2}h_x(X_i,U_i)|\Big|\mathcal{E}\Big) \\
&+ P\Big(\gamma_N^{1/2}h_x(X_i,U_i) < \kappa'\gamma_N^{1/4} + |\sum_{j\neq i} A_{i,j}1\{X_j = x\} - \gamma_N^{1/2}h(X_i,U_i)|\Big|\mathcal{E}^c\Big) \times P\Big(\mathcal{E}^c\Big)
\end{aligned}$$

Note that by Equation (C.5) (which holds conditionally and so also unconditionally)

$$P\Big(\gamma_N^{1/2}h_x(X_i,U_i) < \kappa'\gamma_N^{1/4} + |\sum_{j\neq i} A_{i,j}1\{X_j = x\} - \gamma_N^{1/2}h_x(X_i,U_i)|\Big|\mathcal{E}^c\Big) \times P\Big(\mathcal{E}^c\Big) \leq \frac{1}{\gamma_N} = o(1).$$

Finally, we can write for a finite constant $\bar{C} < \infty$,

$$\begin{aligned}
&P\Big(\gamma_N^{1/2}h_x(X_i,U_i) < \kappa'\gamma_N^{1/4} + |\sum_{j\neq i} A_{i,j}1\{X_j = x\} - \gamma_N^{1/2}h_x(X_i,U_i)|\Big|\mathcal{E}\Big) \\
&\leq P\Big(\gamma_N^{1/2}h_x(X_i,U_i) < \kappa'\gamma_N^{1/4} + \bar{C}\sqrt{\log(2\gamma_N)\gamma_N^{1/2}}\Big|\mathcal{E}\Big) \\
&\leq P\Big(\inf_{x,x',u'} \gamma_N^{1/2}h_x(x',u') < \kappa'\gamma_N^{1/4} + \bar{C}\sqrt{\log(2\gamma_N)\gamma_N^{1/2}}\Big|\mathcal{E}\Big) \\
&= 1\Big\{\inf_{x,x',u'} h_x(x',u') < \kappa'\gamma_N^{-1/4} + \bar{C}\sqrt{\log(2\gamma_N)}\gamma_N^{-1/4}\Big\}
\end{aligned}$$

which equals to zero for $N, \gamma_N$ large enough, since $\inf_{x,x',u'} h_x(x',u') > 0$.

## C.4 Proofs for the extensions

### C.4.1 Proof of Theorem A.1

Observe that we can write

$$\begin{aligned}
\mathbb{E}\Big[\bar{Y}_t^{(k)}\Big|p_t^{(k)}\Big] &= \alpha_t + \tau_k + g\Big(q(\beta+\eta_n) + o_p(\eta_n), \beta+\eta_n\Big), \\
\mathbb{E}\Big[\bar{Y}_t^{(k)}\Big|p_t^{(k+1)}\Big] &= \alpha_t + \tau_k + g\Big(q(\beta-\eta_n) + o_p(\eta_n), \beta-\eta_n\Big).
\end{aligned}$$

From a Taylor expansion in its first argument around $q(\beta+\eta_n)$, we obtain

$$g\Big(q(\beta+\eta_n) + o_p(\eta_n), \beta+\eta_n\Big) = g\Big(q(\beta+\eta_n), \beta+\eta_n\Big) + o_p(\eta_n)$$

and similarly once we subctract $\eta_n$. Therefore, we obtain

$$\mathbb{E}\Big[\bar{Y}_t^{(k)}\Big|p_t^{(k)}\Big] - \mathbb{E}\Big[\bar{Y}^{(k)}\Big|p_t^{(k+1)}\Big] = \tau_k - \tau_{k+1} + g\Big(q(\beta+\eta_n), \beta+\eta_n\Big) + o_p(\eta_n) - g\Big(q(\beta-\eta_n), \beta-\eta_n\Big).$$

We can now proceed with a Taylor expansion around of the functions $g(\cdot)$ around $\beta$ to obtain (this follows similarly to Lemma C.6)

$$g\Big(q(\beta+\eta_n), \beta+\eta_n\Big) - g\Big(q(\beta-\eta_n), \beta-\eta_n\Big) = 2V_g(\beta)\eta_n + O(\eta_n^2).$$

In addition observe that since at the baseline $\beta_0$ is the same for both clusters,

$$\mathbb{E}[Y_0^{(k)} - Y_0^{(k+1)} | p_t^{(k)}, p_t^{(k+1)}] = \tau_k - \tau_{k+1} + o_p(\eta_n).$$

The proof concludes from Lemma C.3 and the local dependence assumption in Assumption A.1.

### C.4.2 Proof of Theorem A.3

First, we bound

$$\sup_{\theta\in\Theta} \widetilde{W}(\theta) - W(\widehat{\theta}) \leq 2\sum_t q^t \times \underbrace{\sup_{(\beta_1,\beta_2)\in[0,1]^2} \Big|\widehat{\Gamma}(\beta_2,\beta_1) - \Gamma(\beta_2,\beta_1)\Big|}_{(A)} \lesssim (A),$$

since $\sum_t q^t < \infty$. To bound $(A)$ observe first that each element in the grid $\mathcal{G}$ has a distance of order $1/\sqrt{K}$, since the grid has two dimensions and $K/3$ components. As a result for any element $(\beta_2, \beta_1)$, we can write

$$\Gamma(\beta_2,\beta_1) = \underbrace{\Gamma(\beta_2^r,\beta_1^r)}_{(B)} + \underbrace{\frac{\partial\Gamma(\beta_2^r,\beta_1^r)}{\partial\beta_1^r}(\beta_1-\beta_1^r) + \frac{\partial\Gamma(\beta_2^r,\beta_1^r)}{\partial\beta_2^r}(\beta_2-\beta_2^r)}_{(C)} + \underbrace{\mathcal{O}\Big(||\beta_1-\beta_1^r||^2 + ||\beta_2-\beta_2^r||^2\Big)}_{(D)}$$

where $\beta^r \in \mathcal{G}$ is some value in the grid such that $(B)$ is of order $1/K$. We can now write

$$\begin{aligned}
(A) \leq & \underbrace{\sup_{(\beta_1^r,\beta_2^r)\in\mathcal{G}, ||\beta_1^r-\beta_1||^2+||\beta_2^r-\beta_2||^2\lesssim 1/K} \Big|\widetilde{\Gamma}(\beta_2^r,\beta_1^r) - \Gamma(\beta_2^r,\beta_1^r)\Big|}_{(i)} \\
& + \underbrace{\sup_{(\beta_1^r,\beta_2^r)\in\mathcal{G}} \Big|\widehat{g}_2(\beta_2^r,\beta_1^r) - \frac{\partial\Gamma(\beta_2^r,\beta_1^r)}{\partial\beta_2^r}\Big|\Big(|\beta_2-\beta_2^r| + |\beta_1-\beta_1^r|\Big)}_{(ii)} \\
& + \underbrace{\sup_{(\beta_1^r,\beta_2^r)\in\mathcal{G}} \Big|\widehat{g}_1(\beta_2^r,\beta_1^r) - \frac{\partial\Gamma(\beta_2^r,\beta_1^r)}{\partial\beta_1^r}\Big|\Big(|\beta_2-\beta_2^r| + |\beta_1-\beta_1^r|\Big)}_{(iii)} + \mathcal{O}\Big(|\beta_2-\beta_2^r|^2 + |\beta_1-\beta_1^r|^2\Big).
\end{aligned}$$

We now study each component separately. We start from $(i)$. We observe that under Assumption A.3, by doing a Taylor expansion around $(\beta_1^r, \beta_2^r)$, it is easy to observe that we can write

$$\mathbb{E}[\bar{Y}_{t+1}^{(k)}] = \Gamma(\beta_2^r, \beta_1^r) + \mathcal{O}(\eta_n).$$

Therefore by Lemma C.3, and the union bound over $K$ many elements in $\mathcal{G}$ as $\gamma_N \log(\gamma_N K)/n \to 0$, $(i) \to 0$. Consider now $(ii)$. We observe that since $\mathcal{B}$ is compact, we have $\Big(|\beta_2 - \beta_2^r| + |\beta_1 - \beta_1^r|\Big) = \mathcal{O}(1)$. In addition, similarly to what discussed in Lemma C.7, it follows that with probability at least $1-\delta$,

$$\Big|\widehat{g}_1(\beta_2^r, \beta_1^r) - \frac{\partial \Gamma(\beta_2^r, \beta_1^r)}{\partial \beta_1^r}\Big| = \mathcal{O}\Big(\sqrt{\frac{\gamma_N \log(\gamma_N/\delta)}{\eta_n^2 n}} + \eta_n\Big).$$

Hence, by the union bound as $\frac{\gamma_N \log(\gamma_N K)}{\eta_n^2 n} = o(1)$ $(ii) = o_p(1)$ and similarly $(iii)$. The proof concludes after observing that $|\beta_1^r - \beta_1|^2 + |\beta_2^r - \beta_2|^2 \lesssim 1/K$ by construction of the grid.

### C.4.3 Proof of Theorem B.1

Let $\tilde{K} = K/2l$. Take

$$t_z^j = \frac{\frac{1}{\sqrt{z}} \sum_{i=1}^z X_i^j}{\sqrt{(z-1)^{-1} \sum_{i=1}^z (X_i^j - \bar{X}^j)^2}}, \quad X_i^j \sim \mathcal{N}(0, \sigma_i^j).$$

Recall that by Theorem 1 in Ibragimov and Müller (2010), we have that for $\alpha \leq 0.08$

$$\sup_{\sigma_1, \cdots, \sigma_q} P(|t_z| \geq \mathrm{cv}_\alpha) = P(|T_{q-1}| \geq \mathrm{cv}_\alpha),$$

where $\mathrm{cv}_\alpha$ is the critical value of a t-test with level $\alpha$, and $T_{z-1}$ is a t-student random variable with $z-1$ degrees of freedom. The equality is attained under homoskedastic variances (Ibragimov and Müller, 2010). We now write

$$P\Big(\mathcal{T}_n \geq q | H_0\Big) = P\Big(\max_{j \in \{1, \cdots, l\}} |Q_{j,n}| \geq q | H_0\Big) = 1 - P\Big(|Q_{j,n}| \leq q \forall j | H_0\Big) = 1 - \prod_{j=1}^l P\Big(|Q_{j,n}| \leq q | H_0\Big),$$

where the last equality follows by between cluster independence. Observe now that by Theorem 3.2 and the fact that the rate of convergence is the same for all clusters (Assumption

3.2) [68], for all $j$, for some $(\sigma_1, \cdots, \sigma_z)$, $z = \tilde{K}$,

$$\sup_q \left| P\Big(|Q_{j,n}| \leq q | H_0\Big) - P\Big(|t^j_{\tilde{K}}| \leq q\Big)\right| = o(1).$$

As a result, we can write

$$\sup_{\sigma_1, \cdots, \sigma_K} \lim_{n\to\infty} 1 - \prod_{j=1}^{l} P\Big(|Q_{j,n}| \leq q | H_0\Big) = 1 - \prod_{j=1}^{l} \inf_{\sigma^j_1, \cdots, \sigma^j_{\tilde{K}}} P\Big(|t^j_{\tilde{K}}| \leq q\Big),$$

where we used the fact that we use different pairs of (independent) clusters for each entry $j$. Using the result in Bakirov and Szekely (2006), we have

$$\inf_{\sigma^j_1, \cdots, \sigma^j_{\tilde{K}}} P\Big(|t^j_{\tilde{K}}| \leq q\Big) = P\Big(|T_{\tilde{K}-1}| \leq q | H_0\Big).$$

Therefore,

$$1 - \prod_{j=1}^{l} \inf_{\sigma^j_1, \cdots, \sigma^j_{\tilde{K}}} P\Big(|t^j_{\tilde{K}}| \leq q | H_0\Big) = 1 - P^l\Big(|T_{\tilde{K}-1}| \leq q\Big).$$

Setting the expression equal to $\alpha$, we obtain

$$1 - P^l\Big(|T_{\tilde{K}-1}| \leq q\Big) = \alpha \Rightarrow P\Big(|T_{\tilde{K}-1}| \geq q\Big) = 1 - (1-\alpha)^{1/l}.$$

The proof completes after solving for $q$.

### C.4.4 Proof of Theorem B.2

In this subsection, we derive the theorem for the gradient descent method under Assumption B.1, for our extension where we relax global strong concavity. The derivation is split into the following lemmas. First define the oracle descent as follows.

**Definition C.5** (Oracle gradient descent)**.** We define for positive constants $\infty > \mu, \kappa > 0$, $\kappa$ as defined in Lemma C.11, arbitrary $v \in (0,1)$

$$\beta^*_w = \begin{cases} P_{\mathcal{B}_1, \mathcal{B}_2}\Big[\beta^*_{w-1} + \alpha_{w-1} V(\beta^*_{w-1})\Big] \text{ if } ||V(\beta^*_w)||_2 \geq \frac{\kappa}{\mu \check{T}^{1/2-v/2}} \\ \beta^*_{w-1} \text{ otherwise} \end{cases} \quad \beta^*_1 = \beta_0, \tag{C.6}$$

for $\alpha_w = \frac{J}{\check{T}^{1/2-v/2}||V(\beta^*_{w-1})||}$, $J < 1$. □

[68] Here we use continuity of the Gaussian distribution, and the fact that $l$ is finite.

**Lemma C.11** (Adaptive gradient descent for quasi-concave functions and locally strong concave)**.** *Let $\mathcal{B}$ be compact. Define $G = \max\{\sup_{\beta \in \mathcal{B}} 2||\beta||^2, 1\}$. Let Assumption 3.1, 4.1, B.1 hold. Let $\kappa$ be a positive finite constant, defined as in Equation* (C.7)*. Then for any $v \in (0,1)$, for $\check{T} \geq ((G+1)/J)^{1/v}$, the following holds:*

$$||\beta^*_{\check{T}} - \beta^*||^2 \leq \kappa \check{T}^{-1+v}.$$

*Proof of Lemma C.11.* To prove the statement, we use properties of gradient descent methods with gradient norm rescaling (Hazan et al., 2015), with modifications to the original arguments to explicitely obtain an almost linear rate, and account for the formalization of local strong concavity based on the Hessian which we provide in our context.

**Preliminaries** Clearly, if the algorithm terminates at $w$, under Assumption B.1 (B), this implies that

$$||\beta^*_w - \beta^*||_2^2 \leq \kappa \check{T}^{-1+v},$$

proving the claim. Therefore, assume that the algorithm did not terminate at time $w$. This implies that for any $\check{w} \geq 1$, $||\beta^*_{\check{w}} - \beta^*||_2^2 > \kappa \check{T}^{-1+v}$. Define $\epsilon = \check{T}^{-1+v}$ and let $\nabla_w$ be the gradient evaluated at $\beta^*_w$. For every $\beta \in \mathcal{B}$, define $H(\beta)\Big|_{[\beta^*, \beta]}$ the Hessian evaluated at some point $\tilde{\beta} \in [\beta^*, \beta]$, such that

$$W(\beta) = W(\beta^*) + \frac{1}{2}(\beta - \beta^*)^\top H(\beta)\Big|_{[\beta^*, \beta]}(\beta - \beta^*),$$

which always exists by the mean-value theorem and differentiability of the objective function. Define

$$\frac{1}{2}(\beta - \beta^*)^\top H(\beta)\Big|_{[\beta^*, \beta]}(\beta - \beta^*) = f(\beta) \leq 0,$$

where the inequality follows by definition of $\beta^*$ (note that $f(\beta)$ also depends on $\tilde{\beta}$, whose dependence we implicitely suppressed).

**Claim** We claim that

$$-|\lambda_{max}| ||\beta - \beta^*||^2 \leq f(\beta) \leq -|\lambda_{\min}| ||\beta - \beta^*||^2$$

for constants $\lambda_{\max} > \lambda_{\min} > 0$. The lower bound follows directly by Assumption 3.1, while the upper bound follows directly from Assumption B.1 (C), compactness of $\mathcal{B}$, and continuity of the Hessian. We provide details for the upper bound in the following paragraph.

**Proof of the claim on the upper bound** We now use a contradiction argument. Suppose that the upper bound does not hold. Then since $\mathcal{B}$ is compact (and hence $||\beta - \beta^*||$ is bounded away from infinity for all $\beta \in \mathcal{B}$), and $\beta^*$ is unique by (A, B) in Assumption B.1, there must exist a sequence $\beta_s \in \mathcal{B}, \beta_s \to \beta^*$ such that $f(\beta_s) \geq o(||\beta_s - \beta^*||^2)$. Recall that twice continuously differentiability of $W(\beta)$, we have that $H(\beta_s) \to H(\beta^*)$. As a result, we can find, for $s \geq S$, for $S$ large enough, a point in the sequence such that (since $p$ is finite)

$$2f(\beta_s) \leq (\beta_s - \beta^*)^\top H(\beta^*)(\beta_s - \beta^*) + \delta(s)||\beta_s - \beta^*||^2,$$

for $\delta(s) = |\tilde{\lambda}_{\max}(s)|$, where $\tilde{\lambda}_{\max}(s)$ is the maximum eigenvalue of $H(\beta)\Big|_{\beta \in [\beta_s, \beta^*]} - H(\beta^*)$. Note that such decomposition holds by symmetry of $H(\beta)\Big|_{\beta \in [\beta_s, \beta^*]} - H(\beta^*)$. Since $H(\beta^*)$ is negative definite, the above expression is bounded as follows

$$2f(\beta_s) \leq -(|\tilde{\lambda}_{min}| - \delta(s))||\beta_s - \beta^*||^2,$$

where $|\tilde{\lambda}_{min}| > 0$ is the minimum eigenvalue of $H(\beta^*)$ (in absolute value) bounded away from zero by Assumption B.1 (C). By continuity of the Hessian, $\delta(s) \to 0$, and we reach a contradiction. This result implies strong concavity *locally* at the optimum. Here $\lambda_{\max} > \lambda_{\min}$ since the lower bound holds for any finite and large enough $\lambda_{\max}$.

**Cases** Define

$$\kappa = \frac{|\lambda_{max}|}{|\lambda_{min}|} \geq 1. \tag{C.7}$$

Observe now that if $||\beta_w^* - \beta^*||^2 \leq \epsilon\kappa$, the claim trivially holds. Therefore, consider the case where

$$||\beta_w^* - \beta^*||^2 > \epsilon\kappa.$$

**Comparisons within the neighborhood** Take $\tilde{\beta} = \beta^* - \sqrt{\epsilon}\frac{\nabla_w}{||\nabla_w||_2}$. Observe that

$$\begin{aligned}
W(\tilde{\beta}) - W(\beta_w^*) &= \frac{1}{2}(\tilde{\beta} - \beta^*)^\top H(\tilde{\beta})\Big|_{[\beta^*, \tilde{\beta}]}(\tilde{\beta} - \beta^*) - \frac{1}{2}(\beta_w^* - \beta^*)^\top H(\beta_w^*)\Big|_{[\beta^*, \beta_w^*]}(\beta_w^* - \beta^*) \\
&\geq -|\lambda_{\max}|\epsilon + |\lambda_{\min}|\epsilon\kappa = 0.
\end{aligned}$$

As a result, for all $\beta_w^* : ||\beta_w^* - \beta^*||^2 > \epsilon\kappa$, using quasi-concavity

$$\nabla_w^\top(\tilde{\beta} - \beta_w^*) \geq 0 \Rightarrow \nabla_w^\top(\beta^* - \beta_w^*) \geq \sqrt{\epsilon}||\nabla_w||_2 \tag{C.8}$$

**Plugging in the above expression in the definition of $\beta_w^*$** By construction of the algorithm, we write

$$||\beta^* - \beta_{w+1}^*||^2 \leq ||\beta^* - \beta_w^*||^2 - 2\alpha_w J \nabla_w^\top (\beta^* - \beta_w^*) + J^2\alpha_w^2 ||\nabla_w||^2.$$

By Equation (C.8), we can write

$$||\beta^* - \beta_{w+1}^*||^2 \leq ||\beta^* - \beta_w^*||^2 - 2J\alpha_w \sqrt{\epsilon} ||\nabla_w||_2 + J^2\alpha_w^2 ||\nabla_w||^2.$$

Plugging in the expression for $\alpha_w$, and using the fact that $J \leq 1$, we have

$$||\beta^* - \beta_{w+1}^*||^2 \leq ||\beta^* - \beta_w^*||^2 - J\epsilon.$$

**Recursive argument** Recall that since the algorithm did not terminate, $||\beta^* - \beta_{\check{w}}^*||^2 > \epsilon\kappa$, for all $\check{w} \leq w$. Using this argument recursively, we obtain

$$||\beta^* - \beta_{\check{T}}^*||^2 \leq ||\beta^* - \beta_0||^2 - J\sum_{s=1}^{\check{T}} \epsilon = 2\max_{\beta \in \mathcal{B}} ||\beta||^2 - J\check{T}^v \leq G + 1 - J\check{T}^v.$$

Whenever $\check{T} > (G/J + 1/J)^{1/v}$, we have a contradiction. The proof completes. $\square$

**Lemma C.12.** *Let Assumption 2.1, 2.2, 4.1, B.1 hold. Assume that*

$$\epsilon_n \geq \sqrt{p}\Big[\bar{C}\sqrt{\gamma_N \frac{\log(\gamma_N \check{T} K/\delta)}{\eta_n^2 n}} + \eta_n\Big], \quad \frac{1}{4\mu\check{T}^{1/2-v/2}} - \epsilon_n \geq 0$$

*for a finite constant $\bar{C} < 0$.*

*Then with probability at least $1-\delta$, for any $\delta \in (0,1)$ for any $w \leq \check{T}$,*

$$\textit{either (i)}\Big|\Big|\check{\beta}_k^w - \beta_w^*\Big|\Big|_\infty = \mathcal{O}(P_w(\delta) + p\eta_n), \ \textit{or (ii)}\Big|\Big|\check{\beta}_k^w - \beta^*\Big|\Big|_2^2 \leq \frac{p}{\check{T}^{1-v}}$$

*where $P_1(\delta) = \mathrm{err}(\delta)$ and $P_w(\delta) = \frac{2\sqrt{p}}{\nu_n} B_p \frac{1}{\check{T}^{1/2-v/2}} P_{w-1}(\delta) + P_{w-1}(\delta) + \frac{2\sqrt{p}}{\nu_n}\frac{1}{\check{T}^{1/2-v/2}}\mathrm{err}(\delta)$, for a finite constant $B_p < \infty$, and $\mathrm{err}(\delta) = \mathcal{O}\Big(\sqrt{\gamma_N \frac{\log(\gamma_N p \check{T} K/\delta)}{\eta_n^2 n}} + p\eta_n\Big)$, with $\nu_n = \frac{1}{\mu\check{T}^{1/2-v/2}} - 2\epsilon_n$.*

*Proof of Lemma C.12.* First, by Lemma 4.1, the estimated coefficients are exogenous. Hence, by invoking Lemma C.7 and the union bound, we can write for every $k$ and $t$, $\delta \in (0,1)$,

$$\check{V}_{k,w}^{(j)} = V^{(j)}(\check{\beta}_{k+2}^w) + \mathcal{O}\Big(\sqrt{\gamma_N \frac{\log(\gamma_N K \check{T}/\delta)}{\eta_n^2 n}} + \eta_n\Big).$$

We now proceed by induction. We first prove the statement, assuming that the constraint is always attained. We then discuss the case of the constrained solution. Define

$$B = p \sup_{\beta} \Big|\Big| \frac{\partial^2 W(\beta)}{\partial \beta^2} \Big|\Big|_{\infty}.$$

**Unconstrained case** Consider $w = 1$. Then since all clusters start from the same starting point $\beta_0$ recall that $(\beta_1^* = \beta_0)$, we can write with probability $1-\delta$, by the union bound over $p$ (which hence enters in the $\log(p)$ component of $\mathrm{err}_n$) and Lemma C.7

$$\Big|\Big| \check{V}_{k,1} - V(\beta_1^*) \Big|\Big|_{\infty} \leq \mathrm{err}(\delta). \tag{C.9}$$

Consider now the case where the algorithm stops. This implies that it must be that $||\check{V}_{k,1}||_2 \leq \frac{1}{\mu \check{T}^{1/2-v/2}} - \epsilon_n$. By Lemma C.7

$$||V(\beta_1^*)||_2 \leq ||\check{V}_{k,1}||_2 + \sqrt{p}\mathrm{err}(\delta) \leq \frac{1}{\mu \check{T}^{1/2-v/2}} - \epsilon_n + \sqrt{p}\mathrm{err}(\delta) \leq \frac{1}{\mu \check{T}^{1/2-v/2}}. \tag{C.10}$$

since $\epsilon_n \geq \sqrt{p}\mathrm{err}(\delta)$. As a result, also the oracle algorithm stops at $\beta_1^*$ by construction of $\epsilon_n$. Suppose the algorithm does not stop. Then it must be that $||\check{V}_{k,1}|| \geq \frac{1}{\mu \check{T}^{1/2-v/2}} - \epsilon_n$ and

$$||V_1(\beta_1^*)|| \geq \frac{1}{\mu \check{T}^{1/2-v/2}} - \epsilon_n - \sqrt{p}\mathrm{err}_1 \geq \frac{1}{\mu \check{T}^{1/2-v/2}} - 2\epsilon_n := \nu_n > 0.$$

Observe now that

$$\begin{aligned} \Big|\Big| \frac{\check{V}_{k,1}}{||\check{V}_{k,1}||_2} - \frac{V(\beta_1^*)}{||V(\beta_1^*)||_2} \Big|\Big|_{\infty} &\leq \Big|\Big| \frac{\check{V}_{k,1} - V(\beta_1^*)}{||V(\beta_1^*)||_2} \Big|\Big|_{\infty} + \Big|\Big| \frac{\check{V}_{k,1}(||\check{V}_{k,1}||_2 - ||V(\beta_1^*)||_2)}{||V(\beta_1^*)||_2 ||\check{V}_{k,1}||_2} \Big|\Big|_{\infty} \\ &\leq \Big|\Big| \frac{\check{V}_{k,1} - V(\beta_1^*)}{||V(\beta_1^*)||_2} \Big|\Big|_{\infty} + \sqrt{p} \Big|\Big| \frac{\check{V}_{k,1} - V(\beta_1^*)}{||V(\beta_1^*)||_2} \Big|\Big|_{\infty}. \end{aligned} \tag{C.11}$$

The last inequality follows from the reverse triangular inequalities and standard properties of the norms. Then with probability at least $1-\delta$, for any $\delta \in (0,1)$

$$(\mathrm{C.11}) \leq \frac{1}{\nu_n} \times 2\sqrt{p}\mathrm{err}(\delta).$$

completing the claim for $w = 1$. Consider now a general $w$. Define the error until time $w-1$ as $P_{w-1}$.Then for every $j \in \{1, \cdots, p\}$, by Assumption 3.1, we have with probability at least

$1 - w\delta$ (using the union bound),

$$\begin{aligned}
&\check{V}_{k,w}^{(j)} = V^{(j)}(\check{\beta}_{k+2}^{w}) + \text{err}(\delta) = V^{(j)}(\beta_w^* + P_w(\delta)) + \text{err}(\delta) \\
&\Rightarrow \Big|\Big|\check{V}_{k,w} - V(\beta_w^*)\Big|\Big|_\infty \leq BP_w(\delta) + \text{err}(\delta),
\end{aligned}$$

where the above inequality follows by the mean-value theorem and Assumption 3.1. Suppose now that $||\check{V}_{k,w}||_2 \leq \frac{1}{\mu \check{T}^{1/2-v/2}} - \epsilon_n$. Then for the same argument as in Equation (C.10), we have

$$||V(\check{\beta}_k^w)||_2 \leq \frac{1}{\mu \check{T}^{1/2-v/2}}.$$

Under Assumption B.1 (B) this implies that

$$||\check{\beta}_k^w - \beta^*||_2^2 \leq \frac{1}{\check{T}^{1-v}},$$

which proves the statement. Suppose instead that the algorithm does not stop. Then we can write by the induction argument

$$\Big|\Big|\check{\beta}_k^w + \frac{1}{\check{T}^{1/2-v/2}}\frac{\check{V}_{k,w}}{||\check{V}_{k,w}||_2} - \beta_w^* - \frac{1}{\check{T}^{1/2-v/2}}\frac{V(\beta_w^*)}{||V(\beta_w^*)||_2}\Big|\Big|_\infty \leq P_w(\delta) + \frac{1}{\check{T}^{1/2-v/2}}\underbrace{\Big|\Big|\frac{\check{V}_{k,w}}{||\check{V}_{k,w}||_2} - \frac{V(\beta_w^*)}{||V(\beta_w^*)||_2}\Big|\Big|_\infty}_{(B)}. \tag{C.12}$$

Using the same argument in Equation (C.11), we have with probability at least $1 - \delta$,

$$(B) \leq \frac{2\sqrt{p}}{\nu_n}\Big[\text{err}(\delta) + BP_w(\delta)\Big],$$

which completes the proof for the unconstrained case. The $\check{T}$ component in the error expression follows from the union bound across all $\check{T}$ events.

**Constrained case** Since the statement is true for $w = 1$, we can assume that it is true for all $s \leq w - 1$ and prove the statement by induction. Since $\mathcal{B}$ is a compact space, we can write

$$\begin{aligned}
&\Big|\Big|P_{\mathcal{B}_1,\mathcal{B}_2-\eta_n}\Big[\sum_{s=1}^{w}\alpha_{k,s}\check{V}_{k,s}\Big] - P_{\mathcal{B}_1,\mathcal{B}_2}\Big[\sum_{s=1}^{w}\alpha_s V(\beta_s^*)\Big]\Big|\Big|_\infty \\
&\leq \Big|\Big|P_{\mathcal{B}_1,\mathcal{B}_2-\eta_n}\Big[\sum_{s=1}^{w}\alpha_{k,s}\check{V}_{k,s}\Big] - P_{\mathcal{B}_1,\mathcal{B}_2-\eta_n}\Big[\sum_{s=1}^{w}\alpha_s V(\beta_s^*)\Big]\Big|\Big|_\infty + p\mathcal{O}(\eta_n) \\
&\leq 2\Big|\Big|\sum_{s=1}^{w}\alpha_{k,s}\check{V}_{k,s} - \sum_{s=1}^{w}\alpha_s V(\beta_s^*)\Big|\Big|_\infty + p\mathcal{O}(\eta_n).
\end{aligned}$$

For the first component in the last inequality, we follow the same argument as above. $\square$

**Lemma C.13.** *Let the conditions in Lemma C.12 hold. Then with probability at least* $1-\delta$*, for any* $k \in \{1, \cdots, K\}$*, for any* $v \in (0,1), \delta \in (0,1)$*,* $\check{T} \geq \zeta^{1/v}$*, for* $\zeta < \infty$ *being a finite constant*

$$||\beta^* - \check{\beta}_k^{\check{T}}||_2^2 \leq \frac{\kappa}{\check{T}^{1-v}} + \check{T} e^{B_p \sqrt{p} \check{T}} \times \mathcal{O}\Big(\gamma_N \frac{\log(p \gamma_N \check{T} K/\delta)}{\eta_n^2 n} + p^2 \eta_n^2\Big),$$

*with* $\kappa, B_p < \infty$ *being constants independent on* $(n, \check{T})$ *and* $\epsilon_n$ *as defined in Lemma C.12.*

*Proof.* We invoke Lemma C.12. Observe that we only have to check that the result holds for (i) in Lemma C.12, since otherwise the claim trivially holds. Using the triangular inequality, we can write

$$||\beta^* - \check{\beta}_k^{\check{T}}||_2^2 \leq ||\beta^* - \beta_{\check{T}}^*||_2^2 + ||\check{\beta}_k^{\check{T}} - \beta_{\check{T}}^*||_2^2.$$

The first component on the right-hand side is bounded by Lemma C.11, with $\check{T} \geq \zeta^{1/v}, \zeta$ being a constant defined in Lemma C.11.

Using Lemma C.12, we bound with probability at least $1-\delta$, the second component as follows

$$||\check{\beta}_k^{\check{T}} - \beta_{\check{T}}^*||_2^2 \leq p||\check{\beta}_k^{\check{T}} - \beta_{\check{T}}^*||_\infty^2 = p \times \mathcal{O}(P_{\check{T}}^2(\delta)).$$

We conclude the proof by explicitely defining recursively, for all $1 < w \leq \check{T}$,

$$P_w = (1 + \frac{2B_p\sqrt{p}}{\nu_n \check{T}^{1/2-v/2}}) P_{w-1} + \frac{1}{\check{T}^{1/2-v/2}} \mathrm{err}_n(\delta), \quad P_1 = \mathrm{err}_n(\delta).$$

where $\mathrm{err}_n(\delta) = \frac{2\sqrt{p}}{\nu_n} \mathcal{O}(\sqrt{\gamma_N \frac{\log(pTK/\delta)}{\eta_n^2 n}} + p\eta_n)$, and $B < \infty$ denotes a finite constant. Using a recursive argument, we obtain

$$P_w = \mathrm{err}_n(\delta) \sum_{s=1}^w \alpha_s \prod_{j=s}^w (\frac{2B_p\sqrt{p}}{\nu_n \check{T}^{1/2-v/2}} + 1).$$

Recall now that $\nu_n \geq \frac{1}{2\mu \check{T}^{1/2-v/2}}$ as in Lemma C.12. As a result we can bound the above expression as

$$\sum_{s=1}^w \alpha_s \prod_{j=s}^w (\frac{2B_p\sqrt{p}}{\nu_n \check{T}^{1/2-v/2}} + 1) \leq \sum_{s=1}^w \alpha_s \prod_{j=s}^w (\frac{8\mu^2 \check{T}^{1/2-v/2} B\sqrt{p}}{\check{T}^{1/2-v/2}} + 1) \leq \sum_{s=1}^w \alpha_s \exp\Big(\sum_{j=s}^w 8\mu^2 B_p \sqrt{p}\Big).$$

Now we have

$$\exp\Big(\sum_{j=s}^w 8\mu^2 B_p \sqrt{p}\Big) \leq \exp\Big(8\mu^2 (w-s) B_p \sqrt{p}\Big).$$

We now write

$$P_w(\delta) \leq \mathrm{err}_n(\delta) \sum_{s=1}^{w} \alpha_s \exp\Big(8\mu^2 (w-s) B_p \sqrt{p}\Big) \leq \mathrm{err}_n(\delta) \check{T}^{1/2+v} \exp\Big(8\mu^2 \check{T} B_p \sqrt{p}\Big),$$

where we replaced $w$ with $\check{T}$. The proof completes.

□

**Corollary 6.** *Theorem B.2 holds.*

*Proof.* Consider Lemma C.12 where we choose $\delta = 1/n$. Observe that we choose $\epsilon_n \leq \frac{1}{4\mu \check{T}^{1/2-v/2}}$, which is attained by the conditions in Lemma C.12 as long as $n$ is small enough such that

$$\sqrt{p}\Big[\bar{C}\sqrt{\log(n)\gamma_N \frac{\log(p\gamma_N \check{T} K)}{\eta_n^2 n}} + \eta_n\Big] \leq \frac{1}{4\mu \check{T}^{1/2-v/2}}$$

attained under the assumptions stated in Lemma C.12. As a result, we have $\nu_n = \frac{1}{4\mu \check{T}^{1/2-v/2}}$. By Lemma C.13 for all $k$, with probability at least $1 - 1/n$,

$$||\check{\beta}_k^{\check{T}} - \beta^*||^2 \lesssim \frac{p}{\check{T}^{1-v}}.$$

Also, we have

$$||\beta^* - \frac{1}{K}\sum_k \check{\beta}_k^{\check{T}}||_2^2 \leq \frac{1}{K}\sum_k ||\check{\beta}_k^{\check{T}} - \beta^*||^2.$$

The proof concludes by Theorem C.13 and Assumption 3.1, after observing that

$$W(\beta^*) - W(\hat{\beta}^*) \lesssim ||\beta^* - \hat{\beta}^*||_2^2.$$

□

### C.4.5 Proof of Theorem B.3

By Lemma 2.1, we can write

$$\mathbb{E}[\hat{\Delta}_k(\beta)] = m(1,1,\beta) - m(0,1,\beta) + O(\eta_n^2).$$

Following the same strategy as in the proof of Theorem 3.4, it is easy to show that

$$\mathbb{E}[\hat{S}(0,\beta)] = \frac{\partial m(0,1,\beta)}{\partial \beta} + \frac{1}{2}\Big[\alpha_{t,k} - \alpha_{t-1,k} - \alpha_{t,k+1} + \alpha_{t-1,k+1}\Big] + \mathcal{O}(\eta_n).$$

Similarly, we can write

$$\mathbb{E}[\hat{S}(1,\beta)] = \frac{\partial m(1,1,\beta)}{\partial \beta} + \frac{1}{2}\Big[\alpha_{t,k} - \alpha_{t-1,k} - \alpha_{t,k+1} + \alpha_{t-1,k+1}\Big] + \mathcal{O}(\eta_n).$$

The proof completes after noticing that $\frac{\partial m(1,1,\beta)}{\partial \beta} = 0$.

### C.4.6 Proof of Theorem B.4

The proof follows directly from Lemma C.10, after noticing that every two periods, the function is evaluated at the same vector of parameter $\Gamma(\check{\beta}^w, \check{\beta}^w)$. Therefore, we can apply all our results to the function $\beta \mapsto \Gamma(\beta, \beta)$ which satisfies the same conditions as $W(\beta)$.

### C.4.7 Proof of Theorem B.5

The proof mimics the proof of Theorem 4.3.

Consider Lemma C.10 where we choose $\delta = 1/n$. Note that we can directly apply Lemma C.10 also to the gradient estimated with Algorithm E.3, since, by the circular-cross fitting argument, each parameter $\check{\beta}_k^w$ is estimated using sequentially pairs of different clusters as in Algorithm E.3. The rest of the proof follows verbatim from the one of Theorem 4.3 and omitted for brevity.

## C.5 Proof of the corollaries

In this subsection we provide proofs to the corollary which do not directly follow from the corresponding theorem.

### C.5.1 Proof of Corollary 2

The result directly follows from Theorem B.1, here applied to $l = 1$. The reader may refer to the proof of Theorem B.1 for details.

### C.5.2 Proof of Corollary 3

The corollary follows from Lemma C.3 and the triangular inequality. Note that the rate is $Kn$ since, after pooling observations from clusters, we can equivalently interpret $\bar{\Delta}_n$ as the estimator from two clusters, each with $K/2 \times n$ observations.

### C.5.3 Proof of Corollary 4

The corollary follows from a second-order Taylor expansion, using the assumption that the Hessian is uniformly bounded.

# Appendix D Numerical studies: additional results

## D.1 One-wave experiment

In Figure D.3 we report the power plot for $\rho = 6$. In Figure D.4 we report the welfare gain from increasing $\beta$ by 5% upon rejection of $H_0$ for $\rho = 6$.

## D.2 Multiple-wave experiment

In Table D.1 the comparison with competitors for $\rho = 6$. Results are robust as in the main text.

In Figure D.5 we report a comparison among different learning rates, which are the one which rescales by $1/t$, the one that rescales by $1/\sqrt{T}$ and the one that rescales by $1/\sqrt{t}$. The best performing learning rate rescales the step size by a factor of order $1/\sqrt{t}$.

## D.3 Calibrated experiment with covariates for cash transfers

In this subsection, we turn to a calibrated experiment where we also control for covariates, as discussed in Section 2.3. We use data from Alatas et al. (2012, 2016). We estimate a function heterogenous in the distance of the household's village from the district's center. We use information from approximately four hundred observations, whose eighty percent or more neighbors are observed. We let $X_i \in \{0, 1\}, X_i = 1$ if the household is far from the district's center than the median household, and estimate

$$Y_i|X_i = x \quad = \quad \phi_0 + \tilde{X}_i\tau + D_i\phi_{1,x} + \frac{\sum_{j\neq i} A_{j,i}D_j}{\max\{\sum_{j\neq i} A_{j,i}, 1\}}\phi_{2,x} + \Big(\frac{\sum_{j\neq i} A_{j,i}D_j}{\max\{\sum_{j\neq i} A_{j,i}, 1\}}\Big)^2\phi_{3,x} + \eta_i, \tag{D.1}$$

where $\eta_i$ are unobservables centered on zero conditional on $X_i = x$, and $\tilde{X}_i$ denotes controls which also include $X_i$.[69] Using the estimated parameter, we can then calibrate the simulations as follows.

[69]We also control for the education level, village-level treatments, i.e., how individuals have been targeted in a village (i.e., via a proxy variable for income, a community-based method, or a hybrid), the size of the village, the consumption level, the ranking of the individual poverty level, the gender, marital status, household size, the quality of the roof and top (which are indicators of poverty).

We let $\eta_{i,t}, \sim \mathcal{N}(0, \sigma^2)$, where $\sigma^2$ is the residual variance from the regression. We then generate the network and the covariate as follows:

$$A_{i,j} = 1\Big\{||U_i - U_j||_1 \leq 2\rho/\sqrt{N}\Big\}, \quad U_i \sim_{i.i.d.} \mathcal{N}(0, I_2), \quad X_i = 1\{U_i^{(1)} > 0\}.$$

Here, $U_i^{(1)}$ is continuous and captures a measure of distances. Individuals are more likely to be friends if they have similar distances from the center, and $X_i$ is equal to one if an individual is far from the district's center from the median household. We fix $\rho = 1.5$ to guarantee that the objective's function optimum is approximately equal to the optimum observed from the data (in calibration, the optimum is $\beta \approx 0.26$, while $\beta^* \approx 0.29$ on the data).

We then generate data

$$Y_{i,t}|X_i = x \quad = \quad D_i\hat{\phi}_{1,x} + \frac{\sum_{j\neq i} A_{j,i}D_j}{\max\{\sum_{j\neq i} A_{j,i}, 1\}}\hat{\phi}_{2,x} + \Big(\frac{\sum_{j\neq i} A_{j,i}D_j}{\max\{\sum_{j\neq i} A_{j,i}, 1\}}\Big)^2\hat{\phi}_{3,x} + \eta_{i,t}. \quad \text{(D.2)}$$

where we removed covariates that did not interact with the treatment rule (i.e., do not affect welfare computations).

Our policy function is

$$\pi(x; \beta) = x\beta + (1-x)(1-\beta)$$

where $\beta$ is the probability of treatment for individuals farer from the center. Here, we implicitely imposed a budget constraint $\beta P(X_i = 1) + (1-\beta)P(X_i = 0) = 1/2$, where, by construction $P(X_i = 1) = 1/2$.

We collect results for the one-wave experiment in Figure D.6, D.7 (left-panel), where we report power and the relative improvement from improving by 5% the treatment probability for people in remote areas as discussed in the main text. Welfare improvements (and power) are increasing in the cluster size and the number of clusters. However, such improvements are negligible as we increase clusters from twenty to forty, suggesting that twenty clusters are sufficient to achieve the largest welfare effects.[70] In the right-hand side panel of Figure D.7 we report the out-of-sample regret. The regret is generally decreasing in the number of iterations, especially as the regret is further away from zero. As the regret gets almost zero (0.06%), the regret oscillates around zero as the number of iterations increases due to sampling variation. This behavior is suggestive that for some applications, few iterations (in this case, ten) are sufficient to reach the optimum, up to a small error. In Table D.2, we

[70]The order of magnitude of the welfare gain is smaller compared to simulations with the unconditional probability since, here, we always treat exactly half of the population. As a result, welfare oscillates between 0.24 and 0.29 only (as opposed to zero to one as in the unconditional case), as shown in Figure 2.

observe perfect coverage for $n = 600$, and under-coverage by no more than five percentage points in the remaining cases.

## D.4 Calibrated experiment for information diffusion to increase vaccination

In this subsection, we study the performance of the methods to increase vaccination against COVID-19 through information diffusion.

Specifically, at the beginning of March 2020 (before the vaccination campaign was extensively implemented), we ran an M-Turk experiment where each individual was assigned either of two arms.[71] A control arm, which consisted of a survey asking basic questions on characteristics of the participants, and a treatment arm. The treatment arm was first assigned simple survey questions. Then, individuals under treatment were assigned three questions about COVID, whose correct answer was rewarded a small economic incentive. The goal of these questions is to increase awareness of the severity of COVID.[72] Each correct answer rewarded a small bonus and was displayed right after the participant submitted her answer to the three questions (before the end of the survey). In addition, at the end of the survey, participants were asked again one of the three questions, whose correct answer rewards a bonus. Participants were made aware of the bonuses and the survey's structure. The scope of the treatment was to increase awareness of the severity of the disease by asking questions and showing the correct answers to facilitate information transmission. At the end of the survey, both controls and treated units were asked when they would have done the vaccine. Our outcome of interest is binary and equals whether individuals would have done the vaccine either as soon as possible or during the spring. We estimate the model with 1035 participants.[73] We estimate treatment effects by running a simple linear regression, where the treatment dummy interacts with the dummy, indicating whether the individual classifies herself as liberal, conservative, or "prefer not to say". We find substantial heterogeneity, with positive effects on early vaccination on liberals only. We consider a model and policy function as in Section D.3, where $X$, in this case, denotes whether an individual is liberal or conservative (drawn with the same DGP as in Section D.3 for simplicity), and $\rho = 2$ as in

[71] The experiment was certified an IRB exempt by UCSD, Human Research Protections Program.

[72] These were multiple answers questions. The first question asked (i) Which of these events caused more deaths of Covid in the US? (more answers allowed), giving four options (World War I and II, 50 times more than 9/11, US Civil war); What is the percentage of people in the US who had Covid within the last year? (approximately); The number of people infected from Covid in the last year is comparable to ... (giving three options).

[73] We collected information from 2411 participants. We removed 158 observations that had already received the vaccine in March 2020 and 203 observations that took less than thirty seconds and more than five minutes to take the survey. We also removed all those observations which were not living in the US.

the main text. We calibrate $\hat{\phi}_{1,x}$ to the estimated direct treatment effect and fix the percentage of treatment units to fifty percent. A challenge here is that we do not know spillovers $\phi_{2x}, \phi_{3x}$. Therefore, we choose $\phi_{3,x} = r\phi_{2,x}$, where $r$ is estimated from data on information diffusion from Cai et al. (2015) for simplicity.[74] We choose $\phi_{2,x} + r\phi_{2,x} = \max\{\alpha\phi_{1,x}, 0\}$, i.e., total spillovers equal direct effect $\phi_{1,x}$ times a constant $\alpha \in \{0.1, 0.2, 0.3, 0.4\}$ if these are positive, and zero otherwise.

We run our one wave experiment with these calibrations and collect results in Figure D.8. In the figure, we report the relative welfare improvement of increasing treatment probabilities for liberals of ten percent, upon rejection of the null hypothesis $H_0$ in Equation (19). We note two facts: (i) clusters equal to twenty are sufficient to have sufficient power; (ii) as the size of the spillovers increase, welfare improvements upon rejection increase and similarly power. These results illustrate the benefit of the method, which, here, can lead to up to a twelve percentage points increase in early vaccination. It would be interesting in the future to test these results on real-world experiments run through online platforms.

## D.5 Additional figures

Table D.1: Multiple-wave experiment in Section 5. Relative improvement in welfare with respect to best competitor for $\rho = 6$. The panel at the top reports the out-of-sample regret and the one at the bottom the worst case in-sample regret across clusters.

| | Information | | | | Cash Transfer | | | |
|---|---|---|---|---|---|---|---|---|
| $T =$ | 5 | 10 | 15 | 20 | 5 | 10 | 15 | 20 |
| $n = 200$ | -0.023 | 0.012 | 0.039 | 0.029 | 0.405 | 0.475 | 0.542 | 0.358 |
| $n = 400$ | 0.001 | 0.020 | 0.019 | 0.029 | 0.546 | 0.483 | 0.602 | 0.548 |
| $n = 600$ | 0.0004 | 0.030 | 0.022 | 0.020 | 0.571 | 0.481 | 0.614 | 0.643 |
| | | | | | | | | |
| $n = 200$ | 1.238 | 1.488 | 1.546 | 1.516 | 0.347 | 0.415 | 0.429 | 0.416 |
| $n = 400$ | 1.494 | 1.690 | 1.704 | 1.624 | 0.482 | 0.576 | 0.550 | 0.498 |
| $n = 600$ | 1.579 | 1.791 | 1.809 | 1.689 | 0.606 | 0.689 | 0.664 | 0.586 |

[74] Other choices are possible but omitted for brevity.

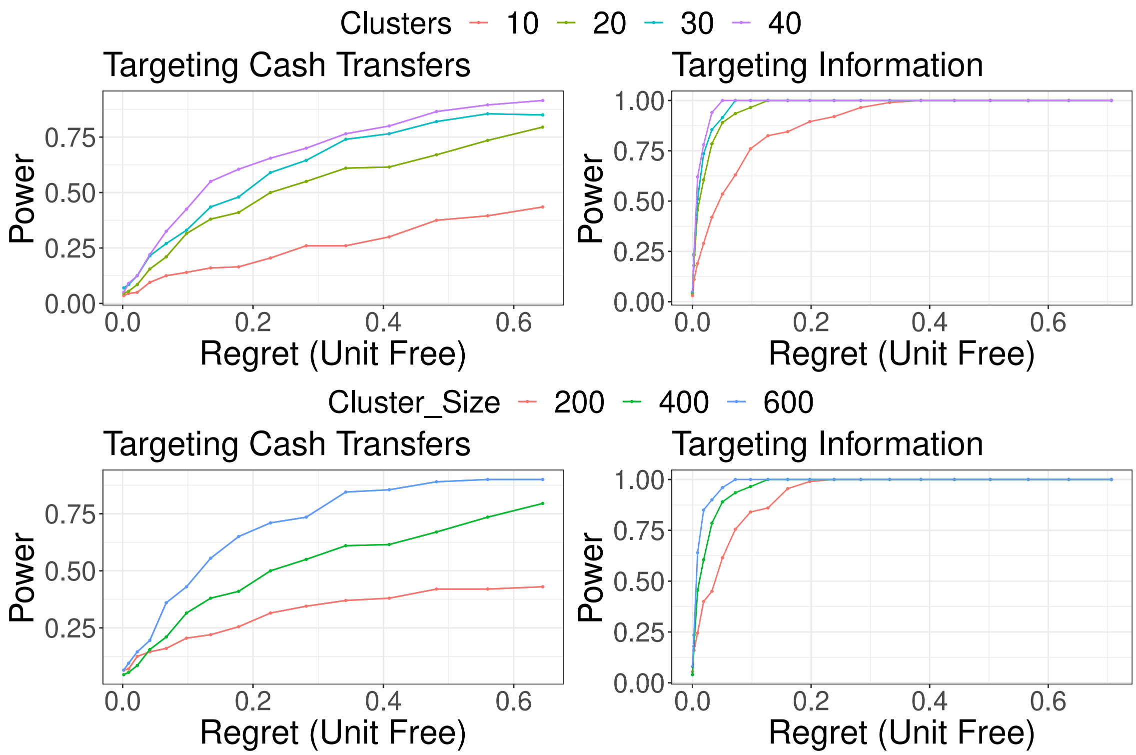


Figure D.1: One-wave experiment in Section 5. 200 replications. Power plot for $\rho = 2$. The panels at the top fix $n = 400$ and varies $K$. The panels at the bottom fix $K = 20$ and vary $n$.

Table D.2: Single-wave experiment in Section D.3, 200 replications. Coverage for tests with size 5%.

| $K =$ | 10 | 20 | 30 | 40 |
|---|---|---|---|---|
| $n = 200$ | 0.940 | 0.950 | 0.895 | 0.900 |
| $n = 400$ | 0.970 | 0.940 | 0.905 | 0.935 |
| $n = 600$ | 0.950 | 0.970 | 0.950 | 0.940 |

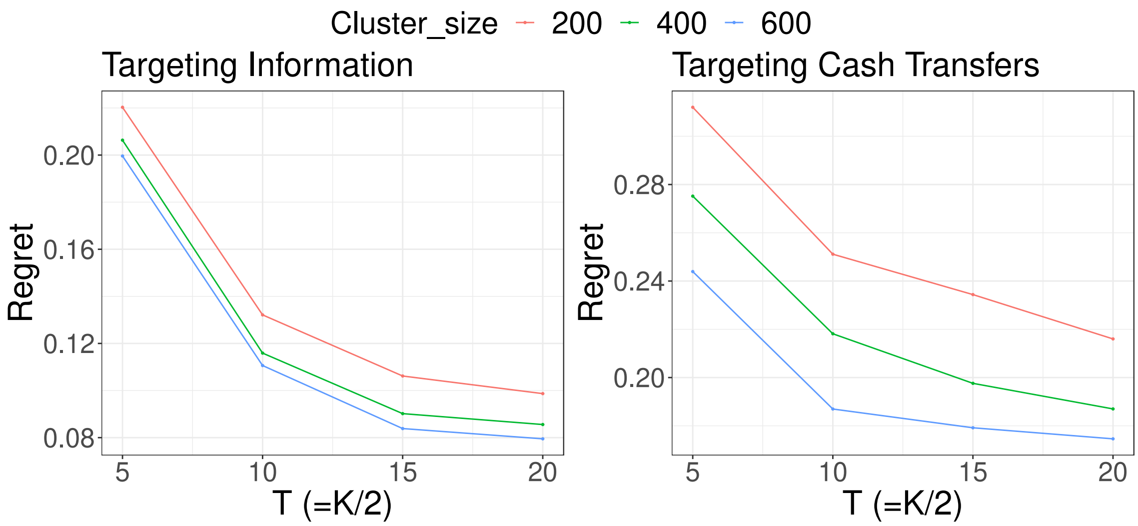


Figure D.2: Multi-wave experiment in Section 5. 200 replications. In-sample regret, average across clusters, for $\rho = 2$.

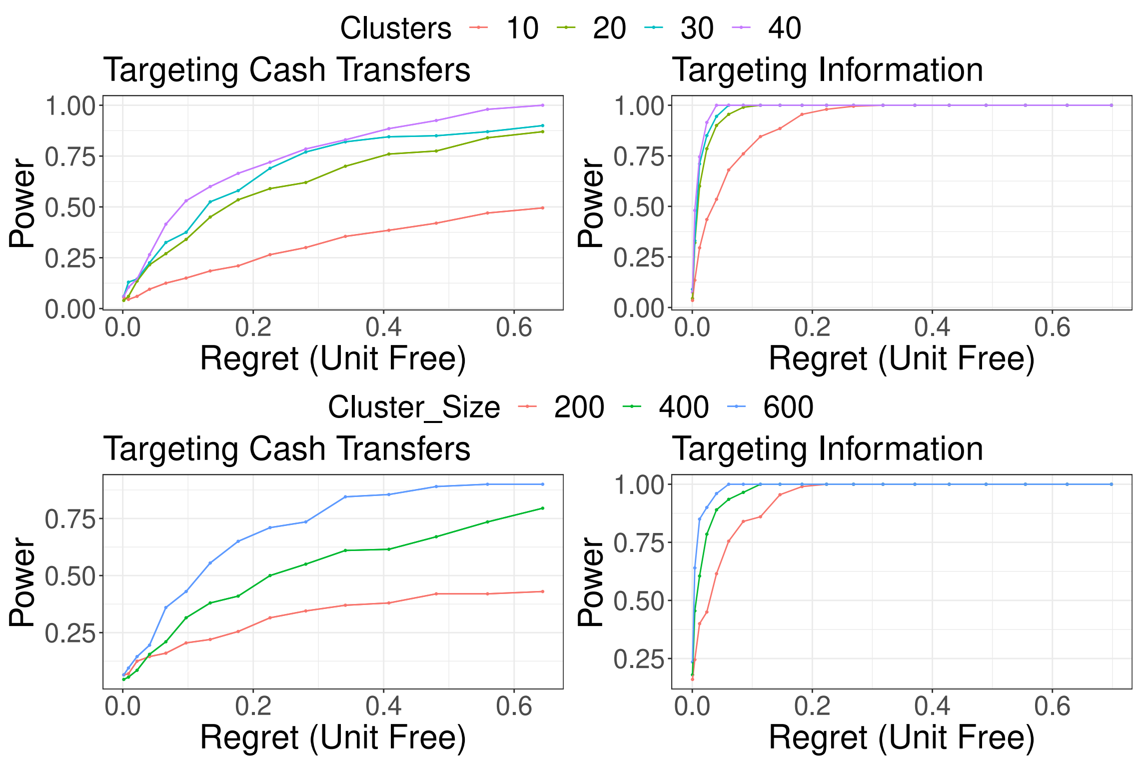


Figure D.3: One-wave experiment in Section 5. Power plot for $\rho = 6$. The panels at the top fix $n = 400$ and varies $K$. The panels at the bottom fix $K = 20$ and vary $n$.

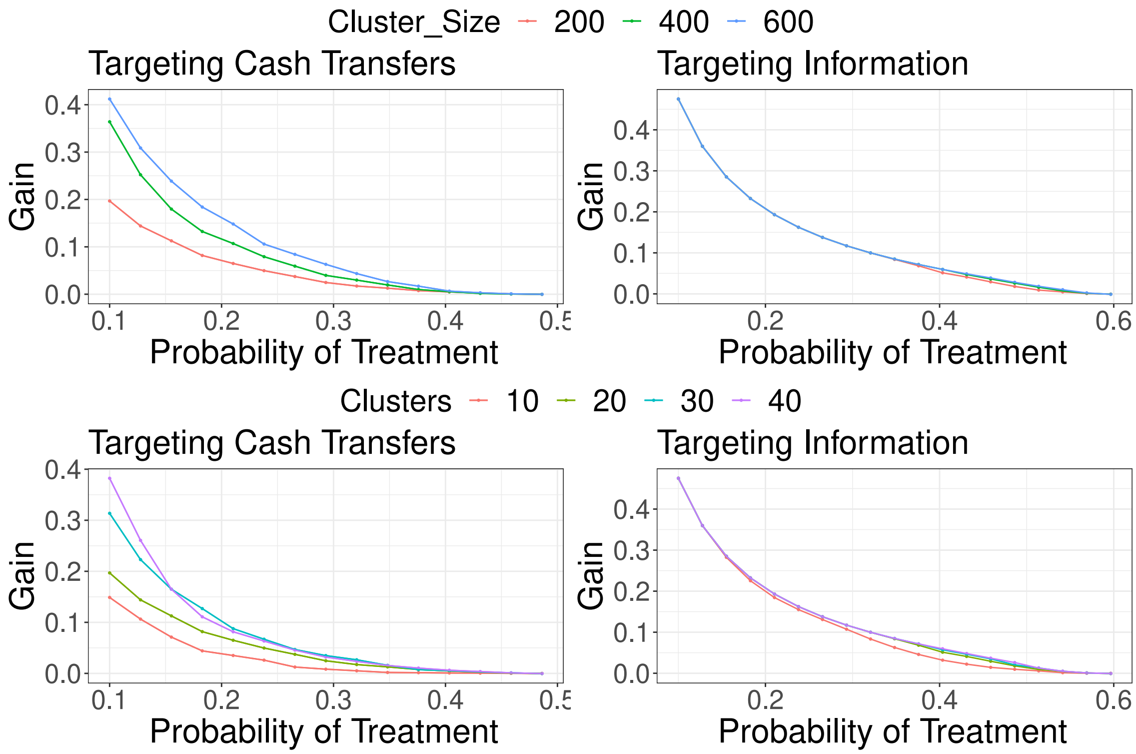


Figure D.4: One-wave experiment in Section 5. $\rho = 6$. Expected percentage increase in welfare from increasing the probability of treatment $\beta$ by 5% upon rejection of $H_0$. Here, the x-axis reports $\beta \in [0.1, \cdots, \beta^* - 0.05]$. The panels at the top fix $n = 400$ and varies the number of clusters. The panels at the bottom fix $K = 20$ and vary $n$.

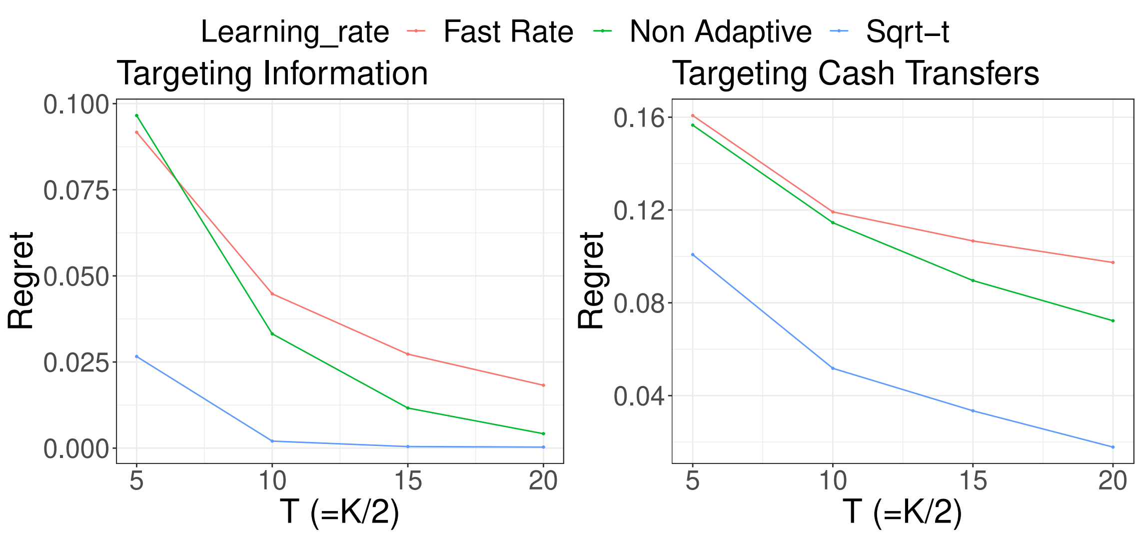


Figure D.5: Comparisons among different learning rates with experiment as in Section 5. 200 replications, $\rho = 2, n = 600, K = 2T$. Fast rate denotes a rescaling of order $1/t$; non-adaptive depends on a rescaling of order $1/\sqrt{T}$; the last one (Sqrt-t) depends on a rescaling of order $1/\sqrt{t}$.

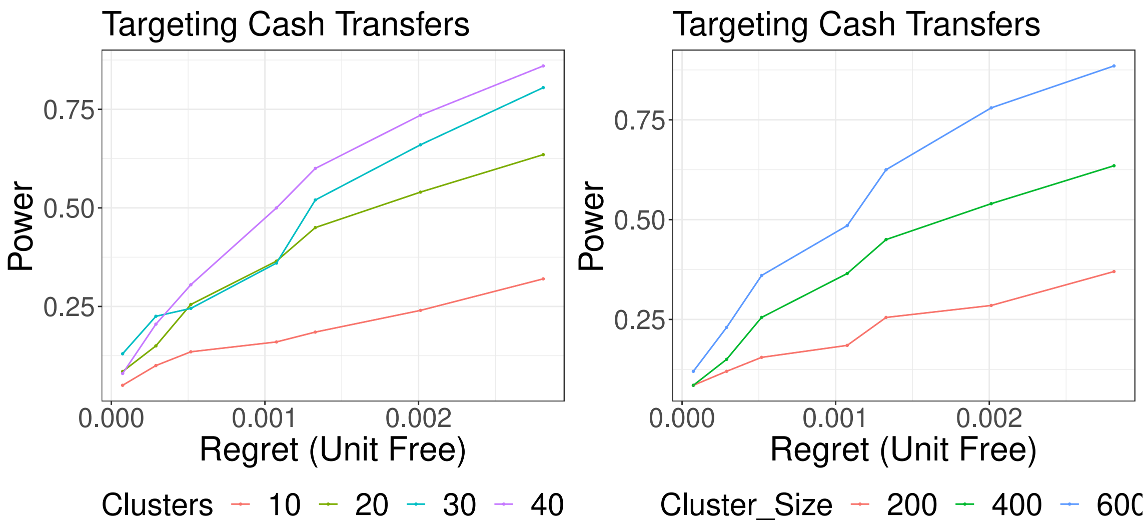


Figure D.6: Single-wave experiment in Section D.3. Power, 200 replications.

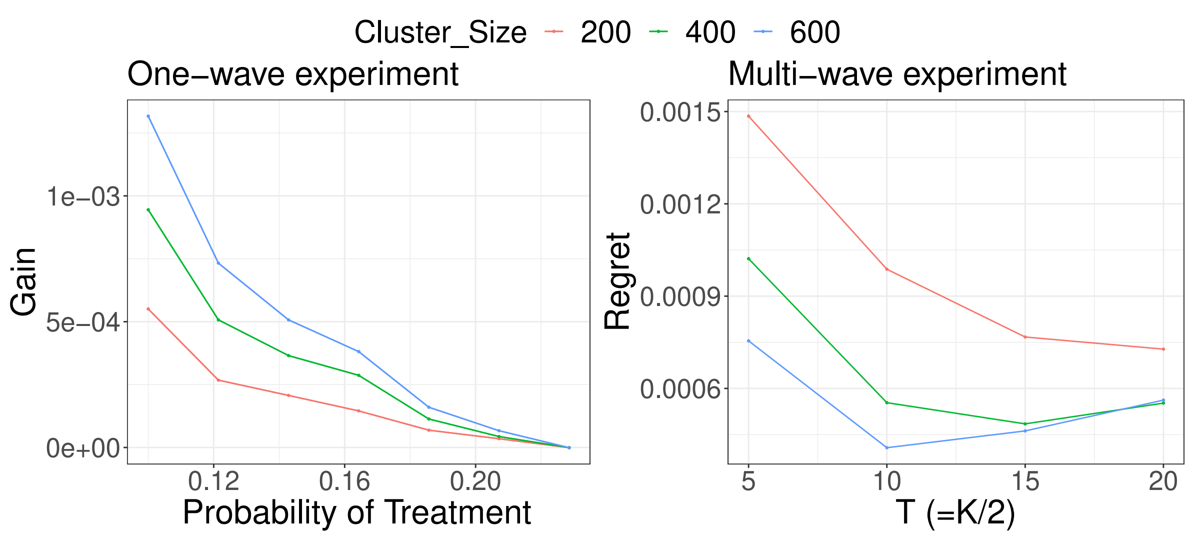


Figure D.7: Experiment in Section D.3. Left-hand side panel reports the expected percentage increase in welfare from increasing the probability of treatment $\beta$ by 5% to individuals in remote areas upon rejection of $H_0$. Here, the x-axis reports $\beta \in [0.1, \cdots, \beta^* - 0.05]$. The right-hand side panel reports the in-sample regret. 400 replications.

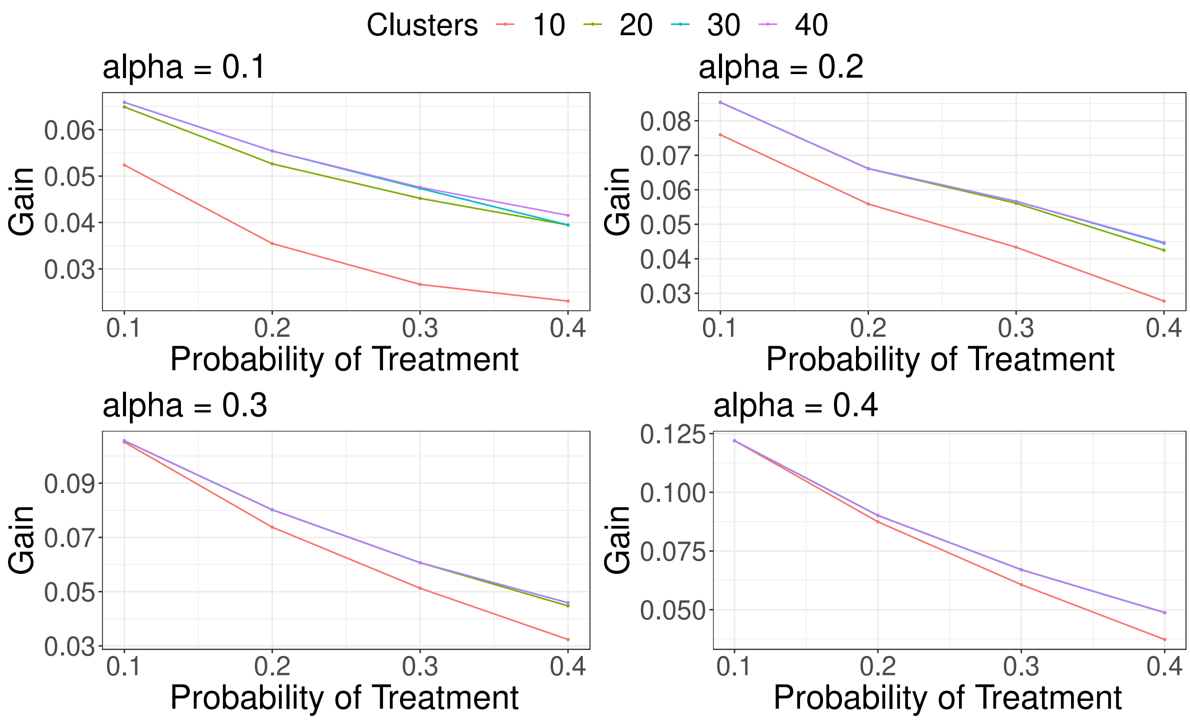


Figure D.8: One wave experiment in Section D.4. Relative welfare improvement for increasing the probability of treatment by ten percent, conditional on rejecting the null hypothesis. 200 replications. Different panels correspond to different levels of spillover effects (captured by $\alpha$).

# Appendix E Additional Algorithms

**Algorithm E.1** Adaptive Experiment with Many Coordinates

**Require:** Starting value $\beta_0 \in \mathbb{R}$, $K$ clusters, $T+1$ periods of experimentation, constant $\bar{C}$.

1: Create pairs of clusters $\{k, k+1\}, k \in \{1, 3, \cdots, K-1\}$;

2: $t = 0(baseline)$:

a: Assign treatments as $D_{i,0}^{(h)} | X_i^{(h)} = x \sim \pi(x; \beta_0)$ for all $h \in \{1, \cdots, K\}$.

b: For $n$ units in each cluster observe $Y_{i,0}^{(h)}, h \in \{1, \cdots, K\}$.

c: For cluster $k$ initalize a gradient estimate $\widehat{V}_{k,t} = 0$ and initial parameters $\check{\beta}_k^o = \beta_0$.

3: **while** $1 \leq w \leq \check{T} = \frac{T}{p}$ **do**

4: **for each** $j \in \{1, \cdots, p\}$ **do**

a: Define

$$\check{\beta}_h^w = \begin{cases} P_{\mathcal{B}_1, \mathcal{B}_2 - \eta_n}\Big[\check{\beta}_h^{w-1} + \alpha_{h+2,w-1}\widehat{V}_{h+2,w-1}\Big], & h \in \{1, \cdots, K-2\}, \\ P_{\mathcal{B}_1, \mathcal{B}_2 - \eta_n}\Big[\check{\beta}_h^{w-1} + \alpha_{h+2,w-1}\widehat{V}_{1,w-1}\Big], h \in \{K-1, K\}. \end{cases}$$

Here, $P_{\mathcal{B}_1, \mathcal{B}_2 - \eta_n}$ denotes the projection operator onto the set $[\mathcal{B}_1, \mathcal{B}_2 - \eta_n]^p$.

b: Assign treatments as (for a finite constant $\bar{C}$)

$$D_{i,t}^{(h)} | X_{i,t}^{(h)} = x \sim \pi(x, \beta_{h,w}), \quad \beta_{h,w} = \begin{cases} \check{\beta}_h^w + \eta_n \underline{e}_j \text{ if } h \text{ is odd} \\ \check{\beta}_h^t - \eta_n \underline{e}_j \text{ if } h \text{ is even} \end{cases}, \quad \bar{C} n^{-1/2} < \eta_n < \bar{C} n^{-1/4}$$

where $\underline{e}_j$ is the vector of zero, with entry $j$ equal to one (see Equation (9)).

c: For $n$ units in each cluster $h \in \{1, \cdots, K\}$ observe $Y_{i,t}^{(h)}$.

d: For each pair $\{k, k+1\}$, estimate

$$\hat{V}_{k,w}^{(j)} = \hat{V}_{k+1,w}^{(j)} = \frac{1}{2\eta_n}\Big[\bar{Y}_t^{(k)} - \bar{Y}_0^{(k)}\Big] - \frac{1}{2\eta_n}\Big[\bar{Y}_t^{(k+1)} - \bar{Y}_0^{(k+1)}\Big].$$

e: $t \leftarrow t + 1$.

5: **end for**

f: $w \leftarrow w + 1$.

6: **end while**

7: Return $\hat{\beta}^* = \frac{1}{K}\sum_{k=1}^K \check{\beta}_k^{\check{T}}$

**Algorithm E.2** Dynamic Treatment Effects with $\beta \in \mathbb{R}$

**Require:** Parameter space $\mathcal{B}$, clusters $\{1, \cdots, K\}$, two periods $\{t, t+1\}$, perturbation $\eta_n$.

1: Group clusters into groups $r \in \{1, \cdots, K/3\}$ of $\{k, k+1, k+2\}$;

2: Construct a grid of parameters $\mathcal{G} \subset [0,1]^2$ equally spaced on $[0,1]^2$;

3: Assign each parameter $(\beta_1^r, \beta_2^r) \in \mathcal{G}$ to a different triad $r$.

4: **for** each $r \in \{1, \cdots, K/3\}$ **do**

5: Randomize treatments as follows:

$$\begin{aligned} D_{i,t}^{(k)} | X_i^{(k)}, \beta_1^r, \beta_2^r &\sim \pi(X_i^{(k)}, \beta_2^r) \\ D_{i,t+1}^{(k)} | X_i^{(k)}, \beta_1^r, \beta_2^r &\sim \pi(X_i^{(k)}, \beta_1^r) \\ D_{i,t}^{(k+1)} | X_i^{(k)}, \beta_1^r, \beta_2^r &\sim \pi(X_i^{(k)}, \beta_2^r + \eta_n) \\ D_{i,t+1}^{(k+1)} | X_i^{(k)}, \beta_1^r, \beta_2^r &\sim \pi(X_i^{(k)}, \beta_1^r) \\ D_{i,t}^{(k+2)} | X_i^{(k)}, \beta_1^r, \beta_2^r &\sim \pi(X_i^{(k)}, \beta_2^r) \\ D_{i,t+1}^{(k+2)} | X_i^{(k)}, \beta_1^r, \beta_2^r &\sim \pi(X_i^{(k)}, \beta_1^r + \eta_n) \end{aligned}. \tag{E.1}$$

6: **end for**

7: For each $k \in \{1, 4, \cdots, K-2\}$ estimate

$$\widehat{g_{1,k}} = \frac{\bar{Y}_{t+1}^{(k)} - \bar{Y}_{t+1}^{(k+2)}}{\eta_n}, \quad \widehat{g_{2,k}} = \frac{\bar{Y}_{t+1}^{(k)} - \bar{Y}_{t+1}^{(k+1)}}{\eta_n}, \quad \widetilde{\Gamma}_k = \frac{1}{3} \sum_{h \in \{k,k+1,k+2\}} \bar{Y}_{t+1}^{(h)} \tag{E.2}$$

**Algorithm E.3** Adaptive Experiment with staggered adoption

**Require:** Starting value $\beta \in \mathbb{R}$, $K$ clusters, $T+1$ periods of experimentation.

1: Create pairs of clusters $\{k, k+1\}, k \in \{1, 3, \cdots, K-1\}$;

2: $t = 0$:

a: For $n$ units in each cluster observe the baseline outcome $Y_{i,0}^{(h)}, h \in \{1, \cdots, K\}, \check{\beta}^0 = \beta$.

b: Initalize a gradient estimate $\widehat{V}_t = 0$

3: **while** $1 \leq t \leq T$ **do**

a: Sample without replaceent one pair of clusters $\{k, k+1\}$ not observed in previous iterations;

b: Define

$$\check{\beta}^t = \check{\beta}^{t-1} + \alpha_t \widehat{V}_t;$$

c: Assign treatments as

$$D_{i,t}^{(h)} \sim \pi(1, \beta_t), \quad \beta_t = \begin{cases} \check{\beta}^t + \eta_n \text{ if } h \text{ is even} \\ \check{\beta}^t - \eta_n \text{ if } h \text{ is odd} \end{cases}, \quad n^{-1/2} < \eta_n \leq n^{-1/4}$$

d: For $n$ units in each cluster $h \in \{1, \cdots, K\}$ observe $Y_{i,t}^{(h)}$.

4: **end while**

5: Return $\hat{\beta}^* = \check{\beta}^T$