Strategyproof Decision-Making in Panel Data Settings and Beyond

Keegan Harris [1] **Anish Agarwal** [2] **Chara Podimata** [3] **Zhiwei Steven Wu** [1]

Abstract

We consider the classical problem of decision-making using *panel data*, in which a decision-maker gets noisy, repeated measurements of multiple *units* (or *agents*). We consider a setup where there is a pre-intervention period, when the principal observes the outcomes of each unit, after which the principal uses these observations to assign a treatment to each unit. Unlike this classical setting, we permit the units generating the panel data to be strategic, i.e. units may modify their pre-intervention outcomes in order to receive a more desirable intervention. The principal's goal is to design a *strategyproof* intervention policy, i.e. a policy that assigns units to their correct interventions despite their potential strategizing. We first identify a necessary and sufficient condition under which a strategyproof intervention policy exists, and provide a strategyproof mechanism with a simple closed form when one does exist. When there are two interventions, we establish that there always exists a strategyproof mechanism, and provide an algorithm for learning such a mechanism. For three or more interventions, we provide an algorithm for learning a strategyproof mechanism if there exists a sufficiently large gap in the principal's rewards between different interventions. Finally, we empirically evaluate our model using real-world panel data collected from product sales over 18 months. We find that our methods compare favorably to baselines which do not take strategic interactions into consideration, even in the presence of model misspecification.

[*]Equal contribution [1]School of Computer Science, Carnegie Mellon University, Pittsburgh, USA [2]Department of Industrial Engineering and Operations Research, Columbia University, New York, USA [3]Sloan School of Management, Massachusetts Institute of Technology, Cambridge, USA. Correspondence to: Keegan Harris <keeganh@cs.cmu.edu>.

ICML (International Conference on Machine Learning) Workshop on The Many Facets of Preference-based Learning, Honolulu, Hawaii, USA.

1. Introduction

In *panel data* (or *longitudinal data*) settings, one observes repeated, noisy, measurements of a collection of *units* over a period of time, during which the units undergo different interventions. For example, units can be individuals, companies, or geographic locations, and interventions can represent discounts, health therapies, or tax regulations. This is a ubiquitous way to collect data, and, as a result, the analysis of panel data has a long history in econometrics and statistics. A common goal in the literature is to analyze how a *principal* (e.g., business platform, regulatory agency) can do "counterfactual inference", i.e., estimate what will happen to a unit if it undergoes a variety of possible interventions. The ultimate goal of such counterfactual inference is to enable data-driven decision-making, where one does not just estimate statistical parameters of interest, but actually uses data to make better decisions. In medical domains, for example, the goal typically is not just estimating health outcomes for patients under different health therapies, but also a policy that selects appropriate therapies for new patients. However, the leap from counterfactual inference to data-driven decision-making comes with additional challenges: namely, when units know that they will be assigned disparate interventions based on their reported data, they have incentives to strategize with their reports. Such strategic interactions in panel data settings are observed in practice. For example, Caro et al. (2010) observe that Zara store managers strategically misreported store inventory information to higher-ups in order to maximize sales at their local branch.

A running example we will use throughout this paper is that of an e-commerce platform that wishes to give one of several possible discounts to a new user to maximize some future metric of interest, say, engagement levels. Suppose the company uses historical data to build a model that estimates the "counterfactual" trajectory of engagement levels of a new user under different discount policies, based on their observed trajectory of engagement levels thus far. If a user knew this were the case, then there is a clear incentive for them to strategically modify their engagement levels to receive a larger discount. Such strategic manipulations in response to data-driven decision-making have been observed in other domains such as lending (Homonoff et al., 2021) and search engine optimization (Davis, 2006). In this paper, we focus on *strategyproof* intervention policies, i.e., policies that assign the correct treatment to the units despite them strategically altering their data. Concretely, we answer two questions:

Q1: Is it possible to design intervention policies that are robust to strategic modifications by units to receive a more favorable intervention? We call such policies *strategyproof*.

Q2: Can we leverage the structure typically present in panel data to derive computationally-efficient algorithms for learning strategyproof intervention policies?

1.1. Main contributions

The first contribution of our work is a general framework for assigning interventions in the presence of strategic units in the panel data setting, which we describe in Section 2. We build off of the *synthetic interventions* framework (Agarwal et al., 2020a), which itself is a generalization of the canonical framework of *synthetic controls* (Abadie and Gardeazabal, 2003; Abadie et al., 2010). In particular, we study the setting in which there is a "pre-intervention" period when all units are under control (i.e. no intervention), and then a "post-intervention" period when each unit undergoes *exactly one* of many possible interventions (including control). Unlike previous work, we allow each unit to strategically modify their pre-intervention outcomes within an "effort budget" of size δ in order to receive a more desirable intervention. Each unit has a "correct"/ground-truth intervention that they should be assigned to, based on their unmodified, true pre-intervention outcomes. We call this "correct" intervention the unit's *type*. The goal of the principal is to deploy a (publicly known) policy π that is *strategyproof*, i.e., assigns the "correct" intervention to all units, despite the fact that they may have strategically modified their pre-intervention outcomes.

Given that units know the principal's policy π and they are allowed to best-respond with altered pre-intervention outcomes that belong anywhere within a δ-ball of their true pre-intervention outcomes, it may seem like **Q1** has a negative answer. However, in Section 3, we derive a *necessary and sufficient condition* for a strategyproof intervention policy to exist. On a technical level, in order to obtain this full characterization, we translate our problem from the primal space (i.e., the *principal's policy space* to the dual one (i.e., the space of the units' actions), and derive properties that units of the same type must share. We show that our necessary and sufficient condition for a strategyproof intervention policy to exist is satisfied when there are two interventions (Theorem 3.4), but it is in general *not* satisfied for more than two (Theorem 3.5). Importantly, we also show that assigning three or more interventions in a panel data setting with strategic agents can be interpreted as an instance of *multiclass strategic classification*. As such, our impossibility result for strategyproof intervention policies translates to a novel impossibility result for strategyproof classification with three or more classes. To the best of our knowledge, we are the first to both draw this connection and discuss multiclass strategic classification altogether.[1]

Addressing **Q2**, we utilize the underlying low-rank structure inherent in many panel data settings to provide an algorithm for learning a strategyproof intervention policy from historical data when there is a single treatment and control (Algorithm 1). The analysis of Algorithm 1 relies on two steps: First, we show that its performance when making decisions on strategic units is equivalent to the performance of another algorithm which acts on non-strategic units. Second, we show that the loss of this equivalent algorithm is upper-bounded by the estimation error of the relationship between pre-intervention outcomes and rewards on historical data (Theorem 4.2). We also provide analogous guarantees for an extension to the setting with an arbitrary number of treatments (Algorithm 2)—under an additional assumption on the difference in rewards between the best and next-best intervention for each unit type (Corollary 4.3). We complement our theoretical results with experiments based on panel data from product sales at several stores over the course of 18 months. We find that the intervention policy of Algorithm 1 outperforms a baseline policy that does not take strategic interactions into account—even when the algorithm's estimate of the unit effort budget δ is misspecified. See Appendix A for an overview of related work.

2. Strategic interactions in panel data settings

Notation Subscripts are used to index the unit and time-step, superscripts are reserved for interventions. We use i to index units, t time-steps, and d interventions. For $x \in \mathbb{N}$, we use the shorthand $[\![x]\!] := \{1, 2, \ldots, x\}$ and $[\![x]\!]_0 := \{0, 1, \ldots, x-1\}$. Finally, all proofs can be found in the Appendix.

Decision making in panel data settings Consider a setting in which the principal observes the outcomes of m units for T time-steps, where $y_{i,t}^{(d)} \in \mathbb{R}$ is the outcome of unit i at time t under intervention d. We assume that unit outcomes are generated via a *latent factor model*, a popular assumption in the panel data setting (e.g., references in Appendix A).

Assumption 2.1 (Latent Factor Model)**.** *The outcome for unit i at time t under treatment $d \in [\![k]\!]_0$ is $y_{i,t}^{(d)} = \langle \mathbf{u}_t^{(d)}, \mathbf{v}_i \rangle$, where $\mathbf{u}_t^{(d)} \in \mathbb{R}^s$ is a latent factor which depends only on time t and intervention d, and $\mathbf{v}_i \in \mathbb{R}^s$ is a latent factor which only depends on unit i. We assume $|y_{i,t}^{(d)}| \leq 1$.*[2]

[1] Our setting and results are incomparable to (Meir et al., 2012), who consider agents which strategize with their reported *labels*, rather than their *features*.

[2] Most work in panel data with latent factor models assumes

Note that Assumption 2.1 does not require the principal to know $\mathbf{u}_t^{(d)}$ or $\mathbf{v}_i$. We assume that the latent dimension s is known to the principal for ease of analysis, although several principled heuristics exist for estimating s in practice from data (see, e.g. Section 2.2.1 in (Agarwal et al., 2020a) for details).

Consider a pre-intervention period of T_0 time-steps, for which each unit is under the same intervention, i.e., under *control*. After the pre-intervention period, the principal assigns an *intervention* $d_i \in [\![k]\!]_0$ to each unit $i \in [\![m]\!]$. W.l.o.g. we denote control by $d = 0$. Once assigned intervention d_i, unit i remains under d_i for the remaining $T - T_0$ time-steps. We use $\mathbf{y}_{i,pre} := [y_{i,1}^{(0)}, \ldots, y_{i,T_0}^{(0)}]^\top \in \mathbb{R}^{T_0}$ to refer to the set of unit i's pre-treatment *observed* outcomes under control, and $\mathbf{y}_{i,post}^{(d)} := [y_{i,T_0+1}^{(d)}, \ldots, y_{i,T}^{(d)}]^\top \in \mathbb{R}^{T-T_0}$ to refer to the set of unit i's post-intervention *potential* outcomes under intervention d. We denote the set of possible pre-treatment outcomes by $\mathcal{Y}_{pre}$.

Definition 2.2 (Intervention Policy). *An intervention policy $\pi : \mathcal{Y}_{pre} \to [\![k]\!]_0$ is a mapping from pre-treatment outcomes to interventions.*

For a given unit i, we denote the intervention assigned to them by intervention policy π as d_i^π.Given an intervention policy π, units may have an incentive to strategically modify their pre-treatment outcomes in order to receive a more desirable intervention. In our e-commerce example, this would correspond to users strategically modifying their engagement levels for the pre-intervention period (e.g., by artificially reducing their time spent on the platform), to "trick" the online marketplace into assigning them a higher discount than the one which would maximize the marketplace's revenue in the post-intervention period. In particular, we study a game between a principal and a population of units. The principal moves first by commiting to an intervention policy. Each unit then *best-responds* to the given intervention policy by strategically modifying their pre-intervention outcomes as follows:

Definition 2.3 (Strategic Responses to Intervention Policies). *Assume that interventions are ordered in increasing unit preference (i.e., units prefer d to d' for $d > d'$). Given an intervention policy $\pi : \mathcal{Y}_{pre} \to [\![k]\!]_0$, unit i best-responds to π by modifying their pre-treatment outcomes as*

$$\tilde{\mathbf{y}}_{i,pre} \in \arg \max_{\widehat{\mathbf{y}}_{i,pre} \in \mathcal{Y}_{pre}} \pi(\widehat{\mathbf{y}}_{i,pre}) \text{ s.t. } \|\widehat{\mathbf{y}}_{i,pre} - \mathbf{y}_{i,pre}\|_2 \leq \delta,$$

where $\delta \in \mathbb{R}_{>0}$ is the unit effort budget and is known to the principal. We assume that if a unit is indifferent between two modifications, they chose the one which requires the smallest effort investment.

$y_{i,t}^{(d)}$ is a *noisy* measurement of the product of latent factors. We consider such settings in Section 4, but present the simpler setup here for ease of exposition.

By Definition 2.3, the goal of each unit is to obtain the most desirable intervention possible when interventions are assigned according to π, subject to the constraint that their modification is bounded in ℓ_2 norm by δ. Such budget assumptions are common in the literature on algorithmic decision making in the presence of strategic agents (e.g., (Chen et al., 2020; Kleinberg and Raghavan, 2020; Harris et al., 2021b)), and are useful for modeling "hard constraints" in a unit's ability to manipulate. For example, in some settings the manipulation of pre-treatment outcomes may have some associated monetary cost, and units may have a fixed budget which they cannot exceed. In other settings the manipulation of pre-treatment outcomes may take time, and the δ-ball represents the set of all possible pre-treatment outcomes a unit could achieve in the amount of time in the pre-treatment period. Given Definition 2.3, the goal of the principal is to design an intervention policy to maximize their *reward* in the presence of such strategic manipulations.

Definition 2.4. *(Principal Reward) The principal's reward for unit i under intervention d is a weighted sum of unit i's outcomes in the post-treatment time period. Specifically, $r_i^{(d)} = \sum_{t=T_0+1}^{T} \omega_t \cdot y_{i,t}^{(d)}$, where $\omega_t \in \mathbb{R}$ for $t > T_0$ are known to the principal.*

We say that unit i is of *type* d if assigning them intervention d maximizes the principal's reward.

Definition 2.5 (Unit Type). *Unit i is of type d if $d \in \arg\max_{d' \in [\![k]\!]_0} r_i^{(d')}$.*

While in general the principal's reward for a unit i may be a function of *all* of unit i's outcomes (not just those in the post-intervention period), we only consider intervention policies which intervene after a fixed pre-treatment time period (for which all units are under control), in line with the synthetic interventions and synthetic controls literature. Linear rewards can capture many settings; e.g. in e-commerce, the online marketplace may wish to maximize the total amount of user engagement on the platform in the post-intervention period (this corresponds to $\omega_t = 1$ for $t > T_0$). As we show, the principal's reward for a given unit may be rewritten as a function of that unit's pre-treatment outcomes when an additional linear span assumption is satisfied.

Assumption 2.6 (Linear Span Inclusion). *We assume $\sum_{t=T_0+1}^{T} \omega_t \cdot \mathbf{u}_t^{(d)} \in \text{span}\{\mathbf{u}_1^{(0)}, \ldots, \mathbf{u}_{T_0}^{(0)}\}$.*

Assumption 2.6 can be viewed as a form of "causal transportability" *over time* which allows the principal to learn something about outcomes in the post-intervention time period from the pre-intervention time period. Such assumptions are fairly common in the literature on learning from panel data (e.g. Assumption 8 in (Agarwal et al., 2020a)).

Lemma 2.7 (Reward Reformulation). *Under Assumption 2.6, $r_i^{(d)}$ may be rewritten as $r_i^{(d)} = \langle \boldsymbol{\beta}^{(d)}, \mathbf{y}_{i,pre} \rangle$, for some $\boldsymbol{\beta}^{(d)} \in \mathbb{R}^{T_0}$.*

Observe that any strategic modification by a unit in the pre-intervention period does *not* change their latent factor $\mathbf{v}$ (or therefore, their post-intervention outcomes). Given knowledge of a unit's latent factor, it would be trivial for the principal to assign them the correct intervention. However, this knowledge is usually not available; instead it must be estimated from the unit's (strategically modified) pre-treatment behavior. Therefore, we are interested in characterizing and learning intervention policies which assign the correct intervention to each unit in the presence of strategic manipulations. Borrowing language from the game theory literature, we refer to such intervention policies as *strategyproof*.

Definition 2.8 (Strategyproof Intervention Policy). *An intervention policy π is strategyproof if $\pi(\tilde{\mathbf{y}}_{i,pre}) = \arg\max_{d \in [\![k]\!]_0} r_i^{(d)}$ for every unit i, where $\tilde{\mathbf{y}}_{i,pre} \in \mathbb{R}^{T_0}$ are unit i's strategically-modified pre-treatment outcomes according to Definition 2.3.*

In Section 4, we focus on the problem of *learning* strategyproof intervention policies from historical data which has *not* been strategically modified, as is the case when, e.g., interventions are assigned according to a *randomized control trial* (since in such settings, units do not have an incentive to strategize).

3. Characterizing strategyproof intervention policies

To define the necessary and sufficient condition under which a strategyproof intervention policy exists, we need to first introduce the notion of a *best-response ball*.

Definition 3.1 (Best-Response Ball). *The best-response ball of a set of units $\mathcal{U}$ is the set of all pre-intervention outcomes $\tilde{\mathcal{Y}}_{pre}(\mathcal{U})$ such that $\tilde{\mathbf{y}}_{pre} \in \tilde{\mathcal{Y}}_{pre}(\mathcal{U})$ if $\|\tilde{\mathbf{y}}_{pre} - \mathbf{y}_{i,pre}\|_2 \leq \delta$ for any unit $i \in \mathcal{U}$, where $\mathbf{y}_{i,pre} \in \mathcal{Y}_{pre}$ denotes the unmodified pre-intervention outcomes associated with unit i.*

The best-response ball for an *individual unit* is its set of feasible modifications according to Definition 2.3. The best-response ball for a *set of units* is the union of the balls of all units contained within the set. Equipped with this definition, we are now ready to introduce our sufficient and necessary condition, which we call *separation of types*.

Condition 3.2 (Separation of Types). *For a given problem instance, let $\mathcal{U}^{(d)}$ denote the set of all units of type d (recall Definition 2.5). Separation of types is satisfied if*

$$\forall d \in [\![k]\!]_0, \; \nexists \, i \in \mathcal{U}^{(d)} \text{ s.t. } \tilde{\mathcal{Y}}_{pre}(i) \subseteq \bigcup_{d'=0}^{d-1} \tilde{\mathcal{Y}}_{pre}(\mathcal{U}^{(d')}).$$

In other words, separation of types is satisfied if for all interventions $d \in [\![k]\!]_0$, there does not exist any unit i of type d whose best-response ball $\tilde{\mathcal{Y}}_{pre}(i)$ is a complete subset of the best-response balls of units with types less than d.

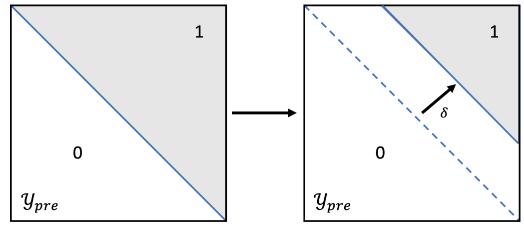

Figure 1: **Left:** The optimal policy in the non-strategic setting assigns the control (intervention 0) to the units left of the blue line and intervention 1 to the units to the right. **Right (Theorem 3.4)**: When units are strategic, the optimal decision boundary is shifted by δ in the direction of the decision boundary.

Theorem 3.3. *Separation of types (Condition 3.2) is both necessary and sufficient for a strategyproof intervention policy (as defined in Definition 2.8) to exist.*

Necessity follows from leveraging Definition 3.1 to show that if Condition 3.2 does *not* hold, there will always be at least one unit who can strategize to receive a better intervention. We show sufficiency by giving a strategyproof intervention policy if Condition 3.2 holds. We will revisit the computational complexity of evaluating this policy later in the section. The significance of Theorem 3.3 is not obvious *a priori*, as it is not immediately clear if/when Condition 3.2 holds in our panel data setting. We begin to address this question by showing that Condition 3.2 always holds in the important special case when there is only a single treatment and control.

Theorem 3.4. *If $d \in \{0, 1\}$, separation of types (Condition 3.2) always holds under Assumption 2.1 and Assumption 2.6. Moreover, the following closed-form intervention policy is strategyproof: Assign intervention d_i to unit i, where*

$$d_i = \begin{cases} 1 & \text{if } \; \langle \boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(0)}, \tilde{\mathbf{y}}_{i,pre} \rangle - \delta \|\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(0)}\|_2 > 0 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

See Figure 1 (left) for an example of such an intervention policy. Intuitively, the idea of Theorem 3.4 is to shift the true decision boundary in such a way to account for potential manipulations. While this shift prevents units from "gaming" the policy to receive the intervention when it is not in the principal's best interest, it may require some units who should receive the intervention to strategize in order to do so. Perhaps somewhat surprisingly, this line of reasoning does not carry over to the setting where there are more than two treatments.

Theorem 3.5. *There exists an instance with three interventions such that Condition 3.2 is not satisfied.*

See Figure 2 (right) for a visualization of one such

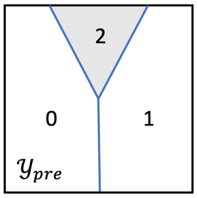

Figure 2: A setting with three interventions for which no strategyproof intervention policy exists.

Algorithm 1 Learning Strategyproof Interventions with One Treatment

Input: Trajectories $\{(\mathbf{y}_{i,pre}, \mathbf{y}_{i,post}^{(0)})\}_{i \in \mathcal{N}^{(0)}}$, $\{(\mathbf{y}_{i,pre}, \mathbf{y}_{i,post}^{(1)})\}_{i \in \mathcal{N}^{(1)}}$

Compute $r_i^{(d_i)} = \sum_{t=T_0+1}^{T} \omega_t \cdot y_{i,t}^{(d_i)}$ for $i \in [\![n]\!]$.

For $d \in \{0, 1\}$, use $\{(\mathbf{y}_{i,pre}, r_i^{(d)})\}_{i \in \mathcal{N}^{(d)}}$ to estimate $\boldsymbol{\beta}^{(d)}$ as $\widehat{\boldsymbol{\beta}}^{(d)}$.

For $i = n+1, \ldots, n+m$ **:**

Assign intervention

$$d_i^A = \begin{cases} 1 & \text{if } \langle \widehat{\boldsymbol{\beta}}^{(1)} - \widehat{\boldsymbol{\beta}}^{(0)}, \tilde{\mathbf{y}}_{i,pre} \rangle - \delta \|\widehat{\boldsymbol{\beta}}^{(1)} - \widehat{\boldsymbol{\beta}}^{(0)}\|_2 > 0 \\ 0 & \text{otherwise.} \end{cases}$$

setting. At a high level, Condition 3.2 cannot hold since the decision boundaries $\langle \boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(0)}, \tilde{\mathbf{y}}_{i,pre} \rangle = 0$ and $\langle \boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(0)}, \tilde{\mathbf{y}}_{i,pre} \rangle = 0$ must both be shifted by δ in order to prevent some units from strategizing to receive intervention 2. However, this prevents other units who should receive intervention 2 from receiving it, since the amount that they would need to modify their pre-intervention outcomes under these shifts is strictly greater than δ. We conclude this section by providing a strategyproof intervention policy for an arbitrary number of interventions whenever separation of types is satisfied.

Theorem 3.6. *When Condition 3.2 is satisfied, the following intervention policy is strategyproof and can be evaluated in time polynomial in T_0 and k under Assumption 2.1 and Assumption 2.6: Assign intervention d_i to unit i, where $d_i = \min\{d \in [\![k]\!]_0 \; : \; \tilde{\mathbf{y}}_{i,pre} \in \tilde{\mathcal{Y}}_{pre}(\mathcal{U}^{(d)})\}$.*

Proof sketch. The form of the above intervention policy follows from the proof of sufficiency in Theorem 3.3. We show that membership to the set $\tilde{\mathcal{Y}}_{pre}(\mathcal{U}^{(d)})$ can be checked by solving a (convex) quadratic program (QP) with polynomial size, which implies that d_i can be computed by solving at most k such QPs.

4. Learning strategyproof intervention policies

We now shift our focus from characterizing strategyproof intervention policies to *learning* them from historical data. While Theorem 3.6 provides a characterization of a strategyproof intervention policy when one exists, deploying such an intervention policy requires knowledge of the underlying relationships between pre-treatment outcomes and principal rewards, which may not be known *a priori*. Additionally, it may be unreasonable to assume that the latent factor model holds *exactly*, due to measurement error or randomness in the outcomes of each unit.With this in mind, we overload the notation of $y_{i,t}^{(d)}$ and consider the following relaxation of Assumption 2.1 throughout the sequel.

Assumption 4.1 (Latent Factor Model; revisited)**.** *Suppose the outcome for unit i at time t under treatment $d \in [\![k]\!]_0$ takes the following factorized form:*

$$\mathbb{E}[y_{i,t}^{(d)}] = \langle \mathbf{u}_t^{(d)}, \mathbf{v}_i \rangle \quad \textit{and} \quad y_{i,t}^{(d)} = \mathbb{E}[y_{i,t}^{(d)}] + \varepsilon_{i,t},$$

where $\mathbf{u}_t^{(d)} \in \mathbb{R}^s$ and $\mathbf{v}_i \in \mathbb{R}^s$ are defined as in Assumption 2.1, and $\varepsilon_{i,t}$ is zero-mean sub-Gaussian random noise with variance at most σ^2. We assume that $|\mathbb{E}[y_{i,t}^{(d)}]| \leq 1$.

Note that under Assumption 4.1, the reward reformulation (Lemma 2.7) now holds *in expectation*. Inspired by the linear form of the strategyproof intervention policy of Theorem 3.4 for two interventions, we begin by deriving performance guarantees for a "plug-in" version of this intervention policy. Our algorithm proceeds as follows: Given historical trajectories of the form $\{(\mathbf{y}_{i,pre}, \mathbf{y}_{i,post}^{(d)})\}_{i \in \mathcal{N}^{(d)}}$ for each $d \in \{0, 1\}$, we can calculate the principal reward for assigning intervention d_i to unit i as $r_i^{(d_i)} := \sum_{t=T_0+1}^{T} \omega_t \cdot y_{i,t}^{(d_i)}$ for $i \in [\![n]\!]$, where $\mathcal{N}^{(d)}$ denotes the set of historical (non-strategic) units who received intervention d. Given the $(\mathbf{y}_{i,pre}, r_i^{(d)})$ pairs as training data, Algorithm 1 uses an error-in-variables regression method (e.g., principal component regression (Jolliffe, 1982b; Bair et al., 2006)) to estimate $\boldsymbol{\beta}^{(0)}, \boldsymbol{\beta}^{(1)}$. The last step is to use the estimated linear coefficients to construct a "plug-in" estimator of intervention policy (1) to use when assigning interventions to the m (strategic) out-of-sample units.

Theorem 4.2. *Suppose $d \in \{0, 1\}$, d_i^A is the intervention assigned to unit i by Algorithm 1, d_i^* is the optimal intervention to assign to unit i, and $\widehat{r}_i^{(d)} := \langle \widehat{\boldsymbol{\beta}}^{(d)}, \mathbf{y}_{i,pre} \rangle$ is the estimated principal reward under intervention d. Then*

$$\begin{aligned} &\frac{1}{m} \sum_{i \in [\![m]\!]} \left(\mathbb{E}[r_{n+i}^{(d_{n+i}^A)}] - \mathbb{E}[r_{n+i}^{(d_{n+i}^*)}] \right)^2 \\ &\leq \frac{4}{m} \max_{d \in \{0,1\}} \sum_{i \in [\![m]\!]} \left(\widehat{r}_{n+i}^{(d)} - \mathbb{E}[r_{n+i}^{(d)}] \right)^2 \end{aligned} \tag{2}$$

Theorem 4.2 shows that the difference in performance of Algorithm 1 and a strategyproof intervention policy that makes no mistakes can be bounded by the difference between the actual and estimated rewards under each intervention. Therefore, if $\widehat{\boldsymbol{\beta}}^{(0)}, \widehat{\boldsymbol{\beta}}^{(1)}$ are good estimates $\boldsymbol{\beta}^{(0)}, \boldsymbol{\beta}^{(1)}$, Algorithm 1 will perform well. In Appendix E, we give end-to-end performance guarantees for Algorithm 2 when the principal uses *principal component regression* (Jolliffe, 1982a) for estimation. Since we are dealing with strategically manipulated data, we are unable to apply prior results for learning from panel data in a black-box way. Our key insight which enables us to obtain performance guarantees for Algorithm 1 is that its performance is matched by another intervention policy which makes decisions on units *which are not strategic* (intervention policy 8 in the Appendix). Given this observation, the bound follows readily from algebraic manipulation.

Next we show that analogous performance guarantees can be obtained for the extension of Algorithm 1 to the setting where there are more than two interventions, when there is a sufficiently large *gap in the principal's expected rewards* for each unit type. This property is natural in many settings of interest; in our e-commerce running example, it corresponds to the principal deriving very different rewards from offering a discount that is not optimal for each group.We now present performance guarantees for Algorithm 2, which is an extension of Algorithm 1 to settings with more than two interventions.

Corollary 4.3 (Informal; detailed version in Corollary E.4)**.** *For $\alpha > 0$, suppose the principal's expected rewards satisfy a sufficiently large reward gap $g(\alpha)$ (Assumption E.3). Then,*

$$\frac{1}{m} \sum_{i \in [\![m]\!]} \left(\mathbb{E}[r_{n+i}^{(d_{n+i}^A)}] - \mathbb{E}[r_{n+i}^{(d_{n+i}^*)}] \right)^2 \leq \frac{k^2}{m} \max_{d \in [\![k]\!]_0} \sum_{i \in [\![m]\!]} \left(\widehat{r}_{n+i}^{(d)} - \mathbb{E}[r_{n+i}^{(d)}] \right)^2$$

with probability at least $1-\alpha$, where d_{n+i}^A is the intervention assigned to unit $n+i$ by Algorithm 2, and d_{n+i}^* and $\widehat{r}_{n+i}^{(d)}$ are defined as in Theorem 4.2.

Intuitively, a gap assumption is not needed in the single treatment regime since a unit will only modify their pre-treatment behavior in order to receive the (single) treatment. This is in contrast to the multi-treatment setting, where a unit's best response may be in one of several directions depending on which treatment(s) they are capable of receiving under a particular intervention policy. Obtaining performance guarantees for learning algorithms which do not require a gap assumption appears challenging for the general case, as the unit best response is not guaranteed to converge smoothly as $\{\widehat{\boldsymbol{\beta}}^{(d)}\}_{d=0}^{k-1}$ approaches $\{\boldsymbol{\beta}^{(d)}\}_{d=0}^{k-1}$.

5. Experiments

Setup Our initial dataset consists of weekly sales data from three products at nine different stores over the course of 18 months. We consider two interventions: `discount` (the product is on sale) and `no discount` (the product is not on sale). We define a unit to be a (store, product) pair which was under `no discount` for five consecutive weeks, followed by either `discount` or `no discount` for three consecutive weeks. Using these (unit, intervention, outcome) tuples, we run a synthetic interventions (Agarwal et al., 2020a) procedure to generate counterfactual outcomes for all units under both `discount` and `no discount`, and use the resulting trajectories as the ground-truth rewards for each unit under both interventions. In order to train our model, we randomly assign interventions to 50% of the units (135 trajectories), and we use the remaining 50% to test the performance. Under such a setting, strategic behavior may arise when, for example, a local store manager wishes to maximize the number of products sold at their specific location, while the owner of the store chain ultimately wants to maximize revenue. In this case, the local store manager could conceivably have an incentive to strategically misreport their weekly revenue during the pre-treatment time period so that their products are given a discount and their sales increase.

Table 1: Normalized Δ revenue and standard deviation over 10 runs for various estimates of δ (denoted by $\widehat{\delta}$). $\widehat{\delta} = 0$ corresponds to the naive policy which does not consider strategic interactions.

$\widehat{\delta}/\delta$	Normalized Δ Revenue	Std
0 (Naive Policy)	0.237	0.110
0.2	0.527	0.126
0.5	0.831	0.033
1 (Algorithm 1)	0.989	0.011
2	0.943	0.014
5	0.846	0.024

Results See Table 1 for a summary of our results. For an intervention policy π, we are interested in the increase in revenue from assigning interventions according to π, as opposed to the alternative. We normalize w.r.t. the *optimal* improvement in revenue, i.e. the best possible improvement if the principal were able to observe both counterfactual trajectories before assigning an intervention. Denote the intervention assigned by policy π to unit $n+i$ as d_{n+i}^{π} and the intervention not assigned by π to unit $n+i$ as $\neg d_{n+i}^{\pi}$. Formally,

$$\text{Normalized } \Delta \text{ Revenue} := \frac{\sum_{i \in [\![m]\!]} \left(r_{n+i}^{(d_{n+i}^{\pi})} - r_{n+i}^{(\neg d_{n+i}^{\pi})} \right)}{\sum_{i \in [\![m]\!]} \left(r_{n+i}^{(d_{n+i}^{*})} - r_{n+i}^{(\neg d_{n+i}^{*})} \right)}.$$

Note that Normalized Δ Revenue is at most 1. We find that Algorithm 1 is able to achieve near-optimal improvement in revenue, in contrast to the relatively poor performance of the naive policy which does not consider incentives. We also examine the performance of Algorithm 1 when the principal's estimate of δ is misspecified as $\widehat{\delta}$ (i.e. the principal's estimate of the agent's effort budget is incorrect; the naive policy is denoted by $\widehat{\delta} = 0$), and find that it degrades gracefully as a function of $\widehat{\delta}/\delta$.

References

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

Alberto Abadie, Alexis Diamond, and Jens Hainmueller. Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American statistical Association*, 105(490):493–505, 2010.

Anish Agarwal, Devavrat Shah, and Dennis Shen. Synthetic interventions. *arXiv preprint arXiv:2006.07691*, 2020a.

Anish Agarwal, Devavrat Shah, and Dennis Shen. On principal component regression in a high-dimensional error-in-variables setting. *arXiv preprint arXiv:2010.14449*, 2020b.

Anish Agarwal, Munther Dahleh, Devavrat Shah, and Dennis Shen. Causal matrix completion. *arXiv preprint arXiv:2109.15154*, 2021a.

Anish Agarwal, Devavrat Shah, Dennis Shen, and Dogyoon Song. On robustness of principal component regression. *Journal of the American Statistical Association*, 116 (536):1731–1745, 2021b. doi: 10.1080/01621459.2021.1928513.

Saba Ahmadi, Hedyeh Beyhaghi, Avrim Blum, and Keziah Naggita. The strategic perceptron. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, pages 6–25, 2021.

Muhammad Jehangir Amjad, Devavrat Shah, and Dennis Shen. Robust synthetic control. *Journal of Machine Learning Research*, 19:1–51, 2018.

Muhummad Amjad, Vishal Mishra, Devavrat Shah, and Dennis Shen. mrsc: Multi-dimensional robust synthetic control. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 3(2), 2019.

Joshua D. Angrist and Jörn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, 2009. ISBN 9780691120348.

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

Dmitry Arkhangelsky, Susan Athey, David A. Hirshberg, Guido W. Imbens, and Stefan Wager. Synthetic difference in differences, 2020.

Orley C Ashenfelter and David Card. Using the longitudinal structure of earnings to estimate the effect of training programs, 1984.

Susan Athey, Mohsen Bayati, Nikolay Doudchenko, Guido Imbens, and Khashayar Khosravi. Matrix completion methods for causal panel data models. *Journal of the American Statistical Association*, 116(536):1716–1730, 2021.

Jushan Bai. Inferential theory for factor models of large dimensions. *Econometrica*, 71(1):135–171, 2003. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/3082043.

Jushan Bai. Panel data models with interactive fixed effects. *Econometrica*, 77(4):1229–1279, 2009. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/40263859.

Jushan Bai and Serena Ng. Matrix completion, counterfactuals, and factor analysis of missing data, 2020.

Eric Bair, Trevor Hastie, Debashis Paul, and Robert Tibshirani. Prediction by supervised principal components. *Journal of the American Statistical Association*, 101(473): 119–137, 2006.

Yahav Bechavod, Katrina Ligett, Steven Wu, and Juba Ziani. Gaming helps! learning from strategic interactions in natural dynamics. In *International Conference on Artificial Intelligence and Statistics*, pages 1234–1242. PMLR, 2021.

Yahav Bechavod, Chara Podimata, Steven Wu, and Juba Ziani. Information discrepancy in strategic learning. In *International Conference on Machine Learning*, pages 1691–1715. PMLR, 2022.

Eli Ben-Michael, Avi Feller, and Jesse Rothstein. The augmented synthetic control method, 2020.

Marianne Bertrand, Esther Duflo, and Sendhil Mullainathan. How much should we trust differences-in-differences estimates? *The Quarterly journal of economics*, 119(1): 249–275, 2004.

Felipe Caro, Jérémie Gallien, Miguel Díaz, Javier García, José Manuel Corredoira, Marcos Montes, José Antonio Ramos, and Juan Correa. Zara uses operations research to reengineer its global distribution process. *Interfaces*, 40(1):71–84, 2010.

Gary Chamberlain. Panel data. In Z. Griliches† and M. D. Intriligator, editors, *Handbook of Econometrics*, volume 2, chapter 22, pages 1247–1318. Elsevier, 1 edition, 1984. URL https://EconPapers.repec.org/RePEc:eee:ecochp:2-22.

Mark K. Chan and Simon Kwok. The PCDID Approach: Difference-in-Differences when Trends are Potentially Unparallel and Stochastic. Working Papers 2020-03, University of Sydney, School of Economics, March 2020. URL https://ideas.repec.org/p/syd/wpaper/2020-03.html.

Yiling Chen, Yang Liu, and Chara Podimata. Learning strategy-aware linear classifiers. *Advances in Neural Information Processing Systems*, 33:15265–15276, 2020.

Victor Chernozhukov, Kaspar Wuthrich, and Yinchu Zhu. Practical and robust t-test based inference for synthetic control and related methods, 2020.

Harold Davis. *Search engine optimization*. " O'Reilly Media, Inc.", 2006.

Jinshuo Dong, Aaron Roth, Zachary Schutzman, Bo Waggoner, and Zhiwei Steven Wu. Strategic classification from revealed preferences. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pages 55–70, 2018.

N. Doudchenko and G. Imbens. Balancing, regression, difference-in-differences and synthetic control methods: A synthesis. *NBER Working Paper No. 22791*, 2016.

Iván Fernández-Val, Hugo Freeman, and Martin Weidner. Low-rank approximations of nonseparable panel models, 2020.

Ganesh Ghalme, Vineet Nair, Itay Eilat, Inbal Talgam-Cohen, and Nir Rosenfeld. Strategic classification in the dark. In *International Conference on Machine Learning*, pages 3672–3681. PMLR, 2021.

Moritz Hardt, Nimrod Megiddo, Christos Papadimitriou, and Mary Wootters. Strategic classification. In *Proceedings of the 2016 ACM conference on innovations in theoretical computer science*, pages 111–122, 2016.

Keegan Harris, Valerie Chen, Joon Sik Kim, Ameet Talwalkar, Hoda Heidari, and Zhiwei Steven Wu. Bayesian persuasion for algorithmic recourse. *arXiv preprint arXiv:2112.06283*, 2021a.

Keegan Harris, Hoda Heidari, and Steven Z Wu. Stateful strategic regression. *Advances in Neural Information Processing Systems*, 34:28728–28741, 2021b.

Keegan Harris, Dung Daniel T Ngo, Logan Stapleton, Hoda Heidari, and Steven Wu. Strategic instrumental variable regression: Recovering causal relationships from strategic responses. In *International Conference on Machine Learning*, pages 8502–8522. PMLR, 2022.

Tatiana Homonoff, Rourke O'Brien, and Abigail B Sussman. Does knowing your fico score change financial behavior? evidence from a field experiment with student loan borrowers. *Review of Economics and Statistics*, 103 (2):236–250, 2021.

Cheng Hsiao, H. Steve Ching, and Shui Ki Wan. A panel data approach for program evaluation: Measuring the benefits of political and economic integration of hong kong with mainland china. *Journal of Applied Econometrics*, 27(5):705–740, 2012. doi: https://doi.org/10.1002/jae.1230.

Meena Jagadeesan, Celestine Mendler-Dünner, and Moritz Hardt. Alternative microfoundations for strategic classification. In *International Conference on Machine Learning*, pages 4687–4697. PMLR, 2021.

Ian T Jolliffe. A note on the use of principal components in regression. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 31(3):300–303, 1982a.

Ian T. Jolliffe. A note on the use of principal components in regression. *Journal of the Royal Statistical Society*, 31 (3):300–303, 1982b.

Jon Kleinberg and Manish Raghavan. How do classifiers induce agents to invest effort strategically? *ACM Transactions on Economics and Computation (TEAC)*, 8(4): 1–23, 2020.

Sagi Levanon and Nir Rosenfeld. Strategic classification made practical. In *International Conference on Machine Learning*, pages 6243–6253. PMLR, 2021.

Kathleen T. Li. Inference for factor model based average treatment effects. *Available at SSRN 3112775*, 2018.

Kathleen T. Li and David R. Bell. Estimation of average treatment effects with panel data: Asymptotic theory and implementation. *Journal of Econometrics*, 197(1):65 – 75, 2017. ISSN 0304-4076. doi: https://doi.org/10.1016/j.jeconom.2016.01.011.

Kung-Yee Liang and Scott L. Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22, 04 1986. ISSN 0006-3444. doi: 10.1093/biomet/73.1.13. URL https://doi.org/10.1093/biomet/73.1.13.

Reshef Meir, Ariel D Procaccia, and Jeffrey S Rosenschein. Algorithms for strategyproof classification. *Artificial Intelligence*, 186:123–156, 2012.

Hyungsik Roger Moon and Martin Weidner. Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica*, 83(4):1543–1579, 2015. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/43616977.

Hyungsik Roger Moon and Martin Weidner. Dynamic linear panel regression models with interactive fixed effects. *Econometric Theory*, 33(1):158–195, 2017. doi: 10.1017/S0266466615000328.

Evan Munro. Learning to personalize treatments when agents are strategic. *arXiv preprint arXiv:2011.06528*, 2020.

M. Hashem Pesaran. Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74(4):967–1012, 2006. ISSN 00129682, 14680262. URL `http://www.jstor.org/stable/3805914`.

Yonadav Shavit, Benjamin Edelman, and Brian Axelrod. Causal strategic linear regression. In *International Conference on Machine Learning*, pages 8676–8686. PMLR, 2020.

Yiqing Xu. Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis*, 25(1):57–76, 2017. doi: 10.1017/pan.2016.2.

A. Related work

Strategic responses to algorithmic decision making A growing line of work at the intersection of computer science and economics aims to model the effects of using algorithmic assessment tools in high-stakes decision-making settings (e.g., (Hardt et al., 2016; Dong et al., 2018; Chen et al., 2020; Kleinberg and Raghavan, 2020; Shavit et al., 2020; Munro, 2020; Ahmadi et al., 2021; Bechavod et al., 2021; 2022; Ghalme et al., 2021; Harris et al., 2021a;b; 2022; Jagadeesan et al., 2021; Levanon and Rosenfeld, 2021)). Hardt et al. (2016) introduce the problem of *strategic classification*, in which a "jury" (principal) deploys a classifier, and a "contestant" (agent), best-responds by strategically modifying their observable features. Subsequent work has studied online learning settings (Dong et al., 2018; Chen et al., 2020; Ahmadi et al., 2021), repeated interactions (Harris et al., 2021b), social learning settings (Bechavod et al., 2022), and settings in which the model being used to make decisions is (partially) unknown to the strategic agents (Ghalme et al., 2021; Harris et al., 2021a; Bechavod et al., 2022). Perhaps the line of work most relevant to ours is that of (Shavit et al., 2020; Munro, 2020; Bechavod et al., 2021; Harris et al., 2022), which aims to identify causal relationships between observable features and outcomes in the presence of strategic responses to various linear models. In contrast, we study a panel data setting in which the principal must assign one of several interventions to strategic units based on longitudinal data which may not have any underlying linear structure. In Appendix B, we discuss the connections between our panel data setting and that of *multiclass* strategic classification. In particular, intervening on strategic units which exhibit a latent factor model structure may be viewed as a particular instance of multiclass classification where agents strategically modify their observable features. We are the first to study such a multiclass strategic classification setting, to the best of our knowledge, and we find that new ideas are required to handle the multiclass nature of the decision-making problem at hand.

Panel data methods in econometrics As stated earlier, this is a setting where one gets repeated measurements of multiple heterogeneous units over time. Prominent frameworks for causal estimation in such settings include *difference-in-differences* (Ashenfelter and Card, 1984; Bertrand et al., 2004; Angrist and Pischke, 2009) and *synthetic controls* (Abadie and Gardeazabal, 2003; Abadie et al., 2010; Hsiao et al., 2012; Doudchenko and Imbens, 2016; Athey et al., 2021; Li and Bell, 2017; Xu, 2017; Amjad et al., 2018; 2019; Li, 2018; Arkhangelsky et al., 2020; Bai and Ng, 2020; Ben-Michael et al., 2020; Chan and Kwok, 2020; Chernozhukov et al., 2020; Fernández-Val et al., 2020; Agarwal et al., 2021b; 2020b). In these frameworks, there is a notion of a "pre-intervention" period where all the units are under control (i.e, no intervention), after which a subset of units receive one of many possible interventions. The goal of these works is to estimate what would have happened to a unit that undergoes an intervention (i.e., a "treated" unit) if it had remained under control (i.e., no intervention), in the potential presence of unobserved confounding. That is, they estimate the counterfactual if a treated unit remains under control for all T time-steps. A critical aspect that enables the methods above is the structure between units and time under control. One elegant encoding of this structure is through a *latent factor model* (also known as an interactive fixed effect model), (Chamberlain, 1984; Liang and Zeger, 1986; Arellano and Honore, 2000; Bai, 2003; 2009; Pesaran, 2006; Moon and Weidner, 2015; 2017). In such models, it is posited that there exist low-dimensional latent unit and time factors that capture unit- and time-specific heterogeneity, respectively, in the potential outcomes. Since the goal in these works is to estimate outcomes under control, no structure is imposed on the potential outcomes under intervention. In (Agarwal et al., 2020a; 2021a), the authors extend this latent factor model to incorporate latent factorization across interventions as well, which allows for identification and estimation of counterfactual mean outcomes under intervention rather than just under control. In essence, we extend these previous works to allow for the pre-intervention outcomes to be strategically manipulated by units to receive a more favorable intervention. What we find noteworthy is that the latent factor model typically assumed in these settings leads to strategyproof estimators that have a simple closed form.

B. Implications for multiclass strategic classification

We now highlight an impossibility result for multiclass strategic classification, which readily follows from Theorem 3.5 and may be of independent interest.

Background on strategic classification When subjugated to algorithmic decision making, decision subjects (agents) have an incentive to strategically modify their input to the algorithm in order to receive a more desirable prediction. In the context of machine learning models, such settings have been formalized in the literature under the name of *strategic classification* (see, e.g., (Hardt et al., 2016; Dong et al., 2018; Chen et al., 2020)). In the (binary) strategic classification setting, the principal commits to an *assessment rule* (usually a linear model), which maps from *observable features* to *binary predictions*. Using knowledge of the assessment rule, strategic agents may modify their observable features in order to maximize their chances of receiving a desirable classification, subject to some constraint on the amount of modification

Protocol: multiclass strategic classification

1. Using historical data collected from n non-strategic agents, the principal learns and publicly commits to an assessment policy $\pi : \mathcal{Y} \to [\![k]\!]_0$
2. m new agents arrive and strategically modify their observable features from $\mathbf{y}_i$ to $\tilde{\mathbf{y}}_i$ according to Assumption B.2, for $i \in \{n+1, \ldots, n+m\}$
3. Principal observes $\tilde{\mathbf{y}}_i$ and assigns prediction $d_i = \pi(\tilde{\mathbf{y}}_i)$ to agent i
4. Principal receives reward $r_i^{(d_i)} = \mathbb{1}\{d_i = d_i^*\}$

Figure 3: Summary of the multiclass strategic classification setting we consider.

which is possible (e.g., a best-response analogous to our Definition 2.3). Given an agent's modified features, the principal uses their assessment rule to make a prediction about the agent. After the prediction is made, the principal receives some feedback about how accurate the assessment rule's prediction was. Under such a setting, the goal of the principal is to deploy an assessment rule with high accuracy on strategic agents.

Using ideas similar to those used in Theorem 3.5, we show that an impossibility result holds for the *multiclass* generalization of the (binary) strategic classification setting, where each strategic agent now belongs to one of $k \geq 3$ classes (as opposed to the binary setting, where $k = 2$). We consider a setting in which a *principal* interacts with m strategic *agents*. Each agent i has a set of *initial* observable features[3] $\mathbf{y}_i \in \mathcal{Y}$. These features are privately observable by the agents and they are not revealed to the principal. Instead, the agents report potentially altered features $\tilde{\mathbf{y}}_i \in \mathcal{Y}$ to the principal. Given observed features $\tilde{\mathbf{y}}_i$, the principal makes a *prediction* d_i from some set of possible *classes* $[\![k]\!]_0$. We assume that each agent has some true label d_i^*, and the principal receives reward 1 if $d_i = d_i^*$ and reward 0 otherwise. In contrast to the principal's reward, we assume that each agent's reward $r_i^A(d)$ is a function of the prediction alone, i.e., $r_i^A(d) = r^A(d), \forall i \in [\![m]\!]$, and is known to the principal.

The principal's *policy* $\pi : \mathcal{Y} \to [\![k]\!]_0$ is a mapping from observable features to predictions. In particular, given a set of training data consisting of $\{(\mathbf{y}_i, d_i)\}_{i=1}^n$ pairs from n non-strategic agents, the goal of the principal is to deploy a policy which minimizes the *out-of-sample* error on m strategic units.

Definition B.1 (Out-of-sample error)**.** *The out-of-sample error of a policy π is defined as the empirical probability that π makes an incorrect prediction on the m test agents. Formally,*

$$\frac{1}{m} \sum_{i=n+1}^{n+m} \mathbb{1}\{d_i \neq d_i^*\}$$

Given a principal policy π, it is natural for an agent to modify their observable features in a way which maximizes their reward. Specifically, we assume that agent i strategically modifies their observable features based on the principal's policy, subject to a constraint on the amount of modification which is possible. In addition to being a common assumption in the literature (similar assumptions are made in, e.g., (Chen et al., 2020; Kleinberg and Raghavan, 2020; Harris et al., 2021b)), this *budget* constraint on the amount an agent can modify their features reflects the fact that agents have inherent constraints on the amount of time and resources they can spend on modification.

Assumption B.2 (Agent Best Response)**.** *We assume that agent i best-responds to the principal's policy π in order to maximize their expected reward, subject to the constraint that their modified observable features $\tilde{\mathbf{y}}_i$ are within an ℓ_2 ball of radius δ of their initial observable features $\mathbf{y}_i$. Formally, we assume that agent i solves the following optimization to determine their modified observable features:*

$$\tilde{\mathbf{y}}_i \in \arg\max_{\hat{\mathbf{y}}_i \in \mathcal{Y}} \; r^A(\pi(\hat{\mathbf{y}}_i))$$
$$\textit{s.t. } \|\hat{\mathbf{y}}_i - \mathbf{y}_i\|_2 \leq \delta$$

[3] Our notation is different than the standard one adopted in the strategic classification literature in order to match our notation from the rest of the paper.

Furthermore, we assume that if an agent is indifferent between modifying their observable features and not modifying, they choose not to modify. See Figure 3 for a summary of the setting we consider. We are now ready to present our main result for multiclass strategic classification, which follows straightforwardly from Theorem 3.5.

Corollary B.3. *Suppose $\mathcal{Y} = \mathbb{R}^2$, $k = 3$, and $r^A(2) > r^A(1) = r^A(0)$. For the following labeling over $\mathcal{Y}$, there exists a distribution over agents such that no policy can achieve perfect classification if agents strategically modify according to Assumption B.2.*

$$d_i^* = \begin{cases} 0 & \text{if } \langle \boldsymbol{\beta}_{10}, \mathbf{y}_i \rangle < 0 \text{ and } \langle \boldsymbol{\beta}_{20}, \mathbf{y}_i \rangle < 0 \\ 1 & \text{if } \langle \boldsymbol{\beta}_{21}, \mathbf{y}_i \rangle < 0 \text{ and } \langle \boldsymbol{\beta}_{10}, \mathbf{y}_i \rangle \geq 0 \\ 2 & \text{if } \langle \boldsymbol{\beta}_{21}, \mathbf{y}_i \rangle \geq 0 \text{ and } \langle \boldsymbol{\beta}_{20}, \mathbf{y}_i \rangle \geq 0 \end{cases}$$

where $\boldsymbol{\beta}_{20} = [1 \ 0.5]^\top$, $\boldsymbol{\beta}_{21} = [-1 \ 0.5]^\top$, and $\boldsymbol{\beta}_{10} = [2 \ 0]^\top$.

C. Proofs from Section 2

Lemma C.1 (Reward Reformulation)**.** *If $\sum_{t=T_0+1}^{T} \omega_t \cdot \mathbf{u}_t^{(d)} \in \text{span}(\{\mathbf{u}_1^{(0)}, \ldots, \mathbf{u}_{T_0}^{(0)}\})$, then $r_i^{(d)}$ can be written as $r_i^{(d)} = \langle \boldsymbol{\beta}^{(d)}, \mathbf{y}_{i,pre} \rangle$, for some $\boldsymbol{\beta}^{(d)} \in \mathbb{R}^{T_0}$.*

Proof. From Assumption 2.1 and Definition 2.4,

$$r_i^{(d)} = \left\langle \sum_{t=T_0+1}^{T} \omega_t \cdot \mathbf{u}_t^{(d)}, \mathbf{v}_i \right\rangle.$$

Applying $\sum_{t=T_0+1}^{T} \omega_t \cdot \mathbf{u}_t^{(d)} \in \text{span}(\{\mathbf{u}_1^{(0)}, \ldots, \mathbf{u}_{T_0}^{(0)}\})$,

$$r_i^{(d)} = \left\langle \sum_{t=1}^{T_0} \beta_t^{(d)} \cdot \mathbf{u}_t^{(0)}, \mathbf{v}_i \right\rangle$$

for some $\boldsymbol{\beta}^{(d)} = [\beta_1^{(d)}, \ldots, \beta_{T_0}^{(d)}]^\top \in \mathbb{R}^{T_0}$. □

D. Proofs from Section 3

Theorem D.1. *Separation of types (Condition 3.2) is both necessary and sufficient for a strategyproof intervention policy (as defined in Definition 2.8) to exist.*

The following two lemmas cover the necessity and sufficiency cases and immediately imply Theorem 3.3. Intuitively, separation of types is *necessary* because if it does not hold, then there are units with lower type that can always pretend to be of higher type, thus leading the principal to intervene wrongly on some subset of the population.

Lemma D.2 (Necessity)**.** *Suppose separation of types (Condition 3.2) does not hold. Then there exists no mapping $\pi : \mathcal{Y}_{pre} \to [\![k]\!]_0$ which can intervene perfectly on all unit types.*

Proof. Assume that separation of types (Condition 3.2) is violated for some unit i of type d. Since $\mathcal{Y}_{pre}(i) \subseteq \bigcup_{d'=0}^{d-1} \mathcal{Y}_{pre}(\mathcal{U}^{(d')})$, any valid modified pre-treatment behavior of unit i can also be obtained by some other unit $i' \in \bigcup_{d'=0}^{d-1} \mathcal{U}^{(d')}$ by Definition 3.1. Therefore, no policy which assigns interventions according to a unit's observed pre-treatment outcomes can perfectly intervene on both $\mathcal{U}^{(d)}$ and $\{\mathcal{U}^{(d')}\}_{d'=1}^{d-1}$. □

Next we show that separation of types is sufficient for a strategyproof intervention policy to exist, by providing a strategyproof intervention policy whenever separation of types holds. Recall that strategyproofness is defined with respect to whether the intervention assigned to a unit matches its type and *not* with respect to whether modification of the pre-treatment outcomes takes place.

Lemma D.3 (Sufficiency)**.** *Suppose separation of types (Condition 3.2) holds. Then the following intervention policy is strategyproof:*

Assign intervention d_i to unit i, where

$$d_i = \min\{d \in [\![k]\!]_0 \ : \ \tilde{\mathbf{y}}_{i,pre} \in \mathcal{Y}_{pre}(\mathcal{U}^{(d)})\} \tag{3}$$

Proof. No unit of type $d' < d$ can receive intervention d by construction, since their pre-treatment outcomes will be in $\mathcal{Y}_{pre}(\mathcal{U}^{(d')})$ by definition. Therefore, it suffices to show that any unit of type d can receive intervention d.

Consider a unit i of type d. Since Condition 3.2 holds, we know that there exists a vector of pre-treatment outcomes $\tilde{\mathbf{y}}_{pre} \in \mathcal{Y}_{pre}(i)$ such that $\tilde{\mathbf{y}}_{pre} \notin \bigcup_{d'=0}^{d-1} \mathcal{Y}_{pre}(\mathcal{U}^{(d')})$. Since $\mathcal{Y}_{pre}(i) \subseteq \mathcal{Y}_{pre}(\mathcal{U}^{(d)})$, unit i can receive intervention d by strategically modifying their pre-treatment outcomes to $\tilde{\mathbf{y}}_{pre}$. □

Theorem D.4. *If $d \in \{0, 1\}$, separation of types (Condition 3.2) always holds under the latent factor model with linear rewards. Moreover, the following closed-form intervention policy is strategyproof: Assign intervention d_i to unit i, where*

$$d_i = \begin{cases} 1 & \textit{if} \quad \langle \boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(0)}, \tilde{\mathbf{y}}_{i,pre} \rangle - \delta \|\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(0)}\|_2 > 0 \\ 0 & \textit{o.w.} \end{cases} \tag{4}$$

We call the hyperplane $\langle \boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(0)}, \tilde{\mathbf{y}}_{i,pre} \rangle - \delta \|\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(0)}\|_2 = 0$ the decision boundary *for interventions* 0 *and* 1.

Proof. If $d \in \{0, 1\}$,we can simplify Condition 3.2 to

$$\nexists \ i \in \mathcal{U}^{(1)} \text{ s.t. } \mathcal{Y}_{pre}(i) \subseteq \mathcal{Y}_{pre}(\mathcal{U}^{(0)}).$$

By the reward reformulation (Lemma 2.7), $\mathcal{U}^{(0)}$ and $\mathcal{U}^{(1)}$ are separated by a single hyperplane: $\langle \boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(0)}, \tilde{\mathbf{y}}_{pre} \rangle = 0$. Therefore, by the definition of best-response ball (Definition 3.1), this simplified version of separation of types must always hold, and a strategyproof intervention policy may be obtained by shifting the hyperplane $\langle \boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(0)}, \tilde{\mathbf{y}}_{pre} \rangle = 0$ by δ (the unit effort budget) in the direction of $\boldsymbol{\beta}^{(0)} - \boldsymbol{\beta}^{(1)}$. Note that such an intervention policy is strategyproof since there exists at least one valid modification of pre-treatment outcomes for all units of type 1 to receive treatment (namely, $\tilde{\mathbf{y}}_{pre} = \mathbf{y}_{pre} + \delta \cdot (\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(0)}) / \|\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(0)}\|_2$), and there exists no valid modification of pre-treatment outcomes for any unit of type 0 to receive treatment (due to Definition 2.3). □

Theorem D.5. *There exists an instance with three interventions such that Condition 3.2 is not satisfied.*

Proof. Suppose $d \in \{0, 1, 2\}$, $T_0 = 2$, and units prefer intervention 2 over interventions 1 and 0, of which they are indifferent between. Suppose that

$$\boldsymbol{\beta}^{(0)} = [-1 \ \ 0.5]^\top, \quad \boldsymbol{\beta}^{(1)} = [1 \ \ 0.5]^\top, \quad \boldsymbol{\beta}^{(2)} = [0 \ \ 1]^\top, \text{ and } \quad \mathbf{y}_{i,pre} = \mathbf{v}_i.$$

Consider the following set of unit types: Let

$$\mathcal{U}^{(0)} = \{\mathbf{v} \ : \ \langle \boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(0)}, \mathbf{v} \rangle = -\alpha, \ \mathbf{v}[1] < 0\}, \quad \mathcal{U}^{(1)} = \{\mathbf{v} \ : \ \langle \boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)}, \mathbf{v} \rangle = -\alpha, \ \mathbf{v}[1] > 0\},$$

and $\mathbf{v}^{(2)} = [0 \quad \zeta]^\top$, where $\alpha, \zeta > 0$. Such a setting is possible, e.g. when $\mathbf{u}_1^{(0)} = [1 \quad 0]^\top$, $\mathbf{u}_2^{(0)} = [0 \quad 1]^\top$, $\sum_{t=T_0+1}^{T} \mathbf{u}_t^{(0)} = [-1 \ \ 0.5]^\top$, $\sum_{t=T_0+1}^{T} \mathbf{u}_t^{(1)} = [1 \ \ 0.5]^\top$, $\sum_{t=T_0+1}^{T} \mathbf{u}_t^{(2)} = [0 \ \ 1]^\top$, and $\omega_{T_0+1} = \omega_{T_0+2} = \cdots = \omega_T = 1$. Observe that a necessary condition for correctly intervening on units in $\mathcal{U}^{(0)}$ is that the intervention policy should not assign intervention $d = 2$ to any units with pre-treatment outcomes $\mathbf{y}_{pre}$ such that $\|\mathbf{y}_{pre} - \mathbf{v}\|_2 \leq \delta$, where $\mathbf{v} \in \mathcal{U}^{(0)}$. This is because such $\mathbf{v}$'s could best respond and get intervention 2 instead of their type, which is 0. An analogous necessary condition holds for units in $\mathcal{U}^{(1)}$. By Definition 3.1, any intervention policy which correctly intervenes on unit $\mathbf{v}_i$ if $\mathbf{v}_i \in \mathcal{U}^{(0)}$ or $\mathbf{v}_i \in \mathcal{U}^{(1)}$ must not assign intervention $d = 2$ if $\tilde{\mathbf{y}}_{i,pre}$ is such that

$$\begin{aligned} &\langle \boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)}, \tilde{\mathbf{y}}_{i,pre} \rangle \in [-\alpha - \delta \|\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)}\|_2, -\alpha + \delta \|\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)}\|_2] \\ \text{or } &\langle \boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(0)}, \tilde{\mathbf{y}}_{i,pre} \rangle \in [-\alpha - \delta \|\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(0)}\|_2, -\alpha + \delta \|\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(0)}\|_2]. \end{aligned}$$

Algorithm 2 Learning Strategyproof Interventions under the Expected Reward Gap Assumption

Input: Trajectories $\{\{(\mathbf{y}_{i,pre}, \mathbf{y}_{i,post}^{(d)})\}_{i \in \mathcal{N}^{(d)}}\}_{d=0}^{k-1}$
Compute $r_i^{(d_i)} = \sum_{t=T_0+1}^{T} \omega_t \cdot y_{i,t}^{(d_i)}$ for $i \in [\![n]\!]$.
For $d \in [\![k]\!]$, use $\{(\mathbf{y}_{i,pre}, r_i^{(d)})\}_{i \in \mathcal{N}^{(d)}}$ to estimate $\boldsymbol{\beta}^{(d)}$ as $\widehat{\boldsymbol{\beta}}^{(d)}$.
For $i = n+1, \ldots, n+m$ **:**
Assign intervention $d_i^B = d$ to unit i if

$$\langle \widehat{\boldsymbol{\beta}}^{(d)} - \widehat{\boldsymbol{\beta}}^{(d')}, \tilde{\mathbf{y}}_{i,pre} \rangle - \delta \|\widehat{\boldsymbol{\beta}}^{(d)} - \widehat{\boldsymbol{\beta}}^{(d')}\|_2 > 0 \ \forall d' < d$$
$$\text{and } \langle \widehat{\boldsymbol{\beta}}^{(d)} - \widehat{\boldsymbol{\beta}}^{(d')}, \tilde{\mathbf{y}}_{i,pre} \rangle + \delta \|\widehat{\boldsymbol{\beta}}^{(d)} - \widehat{\boldsymbol{\beta}}^{(d')}\|_2 \geq 0 \ \forall d' > d$$

However, if this condition is satisfied, it will be impossible to correctly intervene on unit $\mathbf{v}_i$ if $\mathbf{v}_i = \mathbf{v}^{(2)}$ and α, ζ are small enough. To see this, note that in order for both

$$\langle \boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)}, \tilde{\mathbf{y}}_{i,pre} \rangle > \delta \|\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)}\|_2 - \alpha$$
$$\text{and } \langle \boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(0)}, \tilde{\mathbf{y}}_{i,pre} \rangle > \delta \|\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(0)}\|_2 - \alpha$$

to hold,

$$\delta \|\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)}\|_2 - \alpha < (\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)})[2](\zeta + \delta)$$
$$\text{and } \delta \|\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(0)}\|_2 - \alpha < (\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(0)})[2](\zeta + \delta).$$

This implies that intervening perfectly on all units is not possible unless $\frac{1}{2}\zeta + \alpha > \delta(\sqrt{1.25} - 0.5)$, which does not hold for sufficiently small α, ζ. In other words, the condition on α, ζ implies that if the pre-intervention outcomes of units of different types are sufficiently close, intervening perfectly on these units is generally not possible. □

Theorem D.6. *When Condition 3.2 is satisfied, the following intervention policy is strategyproof and can be evaluated in polynomial time: Assign intervention d_i to unit i, where*

$$d_i = \min\{d \in [\![k]\!]_0 \ : \ \tilde{\mathbf{y}}_{i,pre} \in \tilde{\mathcal{Y}}_{pre}(\mathcal{U}^{(d)})\} \tag{5}$$

It suffices to show that $\tilde{\mathbf{y}}_{i,pre} \in \tilde{\mathcal{Y}}_{pre}(\mathcal{U}^{(d)})$ may be checked in polynomial time.

Proposition D.7. *In the latent factor model with linear rewards for the principal (Assumptions 2.1 and 2.4), $\tilde{\mathbf{y}}_{pre} \in \mathcal{Y}_{pre}(\mathcal{U}^{(d)})$ if and only if $OPT^{(d)} \leq \delta$, where $OPT^{(d)}$ is the solution to the following optimization:*

$$\begin{aligned} OPT^{(d)} := \min_{\widehat{\mathbf{y}}_{pre}} \ & \|\widehat{\mathbf{y}}_{pre} - \tilde{\mathbf{y}}_{pre}\|_2 \\ \textit{s.t.} \ & \langle \boldsymbol{\beta}^{(d)} - \boldsymbol{\beta}^{(d')}, \widehat{\mathbf{y}}_{pre} \rangle \geq 0 \ \textit{for all} \ d' \in [\![k]\!]_0. \end{aligned} \tag{6}$$

Proof. Observe that by using the reward reformulation (Lemma 2.7), the definition of a best-response ball (Definition 3.1) may be rewritten as

$$\tilde{\mathbf{y}}_{pre} \in \mathcal{Y}_{pre}(\mathcal{U}) \ \text{ if } \ \|\tilde{\mathbf{y}}_{pre} - \mathbf{y}_{pre}\|_2 \leq \delta$$
$$\text{for any } \mathbf{y}_{pre} \in \mathcal{Y}_{pre} \text{ such that } \langle \boldsymbol{\beta}^{(d)} - \boldsymbol{\beta}^{(d')}, \mathbf{y}_{pre} \rangle \geq 0 \ \text{ for all } \ d' \in [\![k]\!]_0.$$

Therefore, $\tilde{\mathbf{y}}_{pre} \in \mathcal{Y}_{pre}(\mathcal{U})$ if and only if $OPT^{(d)}$ is at most δ. □

E. Proofs from Section 4

Theorem E.1. *Suppose $d \in \{0, 1\}$. Algorithm 1 achieves out-of-sample performance*

$$\frac{1}{m} \sum_{i=n+1}^{n+m} \left(\mathbb{E}[r_i^{(d_i^A)}] - \mathbb{E}[r_i^{(d_i^*)}] \right)^2 \leq \frac{4}{m} \max_{d \in \{0,1\}} \sum_{i=n+1}^{n+m} \left(\widehat{r}_i^{(d)} - \mathbb{E}[r_i^{(d)}] \right)^2 \tag{7}$$

where d_i^A is the intervention assigned to unit i by Algorithm 1, d_i^* is the optimal intervention to assign to unit i, and $\widehat{r}_i^{(d)} := \langle \widehat{\boldsymbol{\beta}}^{(d)}, \mathbf{y}_{i,pre} \rangle$ is the estimated principal reward under intervention d.

The proof of Theorem 4.2 relies on the following proposition, which shows that the interventions assigned by the intervention policy of Algorithm 1 on strategic units match the interventions assigned according to the following intervention policy on units which are always truthful. We say that a unit is truthful if they do not modify their pre-intervention outcomes.

Lemma E.2. *Consider the following intervention policy:*

$$d_i^B = \begin{cases} 1 & \textit{if} \;\; \widehat{r}_i^{(1)} - \widehat{r}_i^{(0)} > 0 \\ 0 & \textit{otherwise}, \end{cases} \tag{8}$$

where $\widehat{\boldsymbol{\beta}}^{(0)}, \widehat{\boldsymbol{\beta}}^{(1)}$ are defined as in Algorithm 1. (Recall that $\widehat{r}_i^{(d)} := \langle \widehat{\boldsymbol{\beta}}^{(d)}, \mathbf{y}_{i,pre} \rangle$.) The intervention policy of Algorithm 1 assigns the same interventions to strategic units that intervention policy (8) assigns to truthful units.

Proof. The proof proceeds on a case-by-case basis. Fix a (strategic) unit $i \in \{n+1, \ldots, n+m\}$.

Case 1: Suppose intervention policy (8) assigns intervention $d_i^B = 1$ to unit i. Since $d_i^B = 1$, $\langle \widehat{\boldsymbol{\beta}}^{(1)} - \widehat{\boldsymbol{\beta}}^{(0)}, \mathbf{y}_{i,pre} \rangle > 0$. If Algorithm 1 assigns intervention 1 to unit i without any modification to their pre-treatment outcome, then the claim holds trivially. One valid modification is:

$$\tilde{\mathbf{y}}_{i,pre} = \mathbf{y}_{i,pre} + \delta \frac{\widehat{\boldsymbol{\beta}}^{(1)} - \widehat{\boldsymbol{\beta}}^{(0)}}{\left\| \widehat{\boldsymbol{\beta}}^{(1)} - \widehat{\boldsymbol{\beta}}^{(0)} \right\|_2}.$$

Supposing unit i modifies to $\tilde{\mathbf{y}}_{i,pre}$,

$$\langle \widehat{\boldsymbol{\beta}}^{(1)} - \widehat{\boldsymbol{\beta}}^{(0)}, \tilde{\mathbf{y}}_{i,pre} \rangle = \langle \widehat{\boldsymbol{\beta}}^{(1)} - \widehat{\boldsymbol{\beta}}^{(0)}, \mathbf{y}_{i,pre} \rangle + \delta \left\langle \widehat{\boldsymbol{\beta}}^{(1)} - \widehat{\boldsymbol{\beta}}^{(0)}, \frac{\widehat{\boldsymbol{\beta}}^{(1)} - \widehat{\boldsymbol{\beta}}^{(0)}}{\left\| \widehat{\boldsymbol{\beta}}^{(1)} - \widehat{\boldsymbol{\beta}}^{(0)} \right\|_2} \right\rangle > \delta \| \widehat{\boldsymbol{\beta}}^{(1)} - \widehat{\boldsymbol{\beta}}^{(0)} \|_2.$$

Therefore, unit i can receive intervention $d_i^A = 1$ under the intervention policy of Algorithm 1.

Case 2: Suppose intervention policy (8) assigns intervention $d_i^B = 0$ to unit i, i.e., $\langle \widehat{\boldsymbol{\beta}}^{(1)} - \widehat{\boldsymbol{\beta}}^{(0)}, \mathbf{y}_{pre} \rangle \leq 0$. To receive intervention $d_i^A = 1$ by the policy of Algorithm 1, it needs to be the case that $\langle \widehat{\boldsymbol{\beta}}^{(1)} - \widehat{\boldsymbol{\beta}}^{(0)}, \tilde{\mathbf{y}}_{i,pre} \rangle - \delta \| \widehat{\boldsymbol{\beta}}^{(1)} - \widehat{\boldsymbol{\beta}}^{(0)} \|_2 > 0$. However according to Definition 2.3, the most a unit can manipulate their pre-treatment outcomes by is δ, so

$$\langle \widehat{\boldsymbol{\beta}}^{(1)} - \widehat{\boldsymbol{\beta}}^{(0)}, \tilde{\mathbf{y}}_{i,pre} \rangle \leq \left\langle \widehat{\boldsymbol{\beta}}^{(1)} - \widehat{\boldsymbol{\beta}}^{(0)}, \mathbf{y}_{i,pre} + \frac{\delta(\widehat{\boldsymbol{\beta}}^{(1)} - \widehat{\boldsymbol{\beta}}^{(0)})}{\| \widehat{\boldsymbol{\beta}}^{(1)} - \widehat{\boldsymbol{\beta}}^{(0)} \|_2} \right\rangle \leq \delta \| \widehat{\boldsymbol{\beta}}^{(1)} - \widehat{\boldsymbol{\beta}}^{(0)} \|_2,$$

where for the last inequality we have used the fact that $\langle \widehat{\boldsymbol{\beta}}^{(1)} - \widehat{\boldsymbol{\beta}}^{(0)}, \mathbf{y}_{pre} \rangle \leq 0$. Therefore, no valid strategic modification to unit i's pre-treatment outcomes exists for which the intervention policy of Algorithm 1 assigns intervention $d_i^A = 1$ to unit i. □

Since the performance of the intervention policy of Algorithm 1 on strategic units matches that of intervention policy (8) on truthful units, we can analyze the performance of intervention policy (8) on truthful units without any loss of generality. The analysis of the performance of intervention policy (8) on truthful units completes the proof of Theorem 4.2. If $d_i^A = d_i^*$, then $\mathbb{E}[r_i^{(d_i)}] - \mathbb{E}[r_i^{(d_i^*)}] = 0$. If $d_i^A \neq d_i^*$, we know that

$$\mathbb{E}[r_i^{(d_i^*)}] - |\widehat{r}_i^{(d_i^*)} - \mathbb{E}[r_i^{(d_i^*)}]| \leq \widehat{r}_i^{(d_i^*)} \leq \widehat{r}_i^{(d_i^A)} \leq \mathbb{E}[r_i^{(d_i^A)}] + |\widehat{r}_i^{(d_i^A)} - \mathbb{E}[r_i^{(d_i^A)}]|.$$

Therefore,

$$
\begin{aligned}
\frac{1}{m}\sum_{i=n+1}^{n+m}\left(\mathbb{E}[r_i^{(d_i^A)}]-\mathbb{E}[r_i^{(d_i^*)}]\right)^2 &\leq \frac{1}{m}\sum_{i=n+1}^{n+m}\left(|\widehat{r}_i^{(d_i^A)}-\mathbb{E}[r_i^{(d_i^A)}]|+|\widehat{r}_i^{(d_i^*)}-\mathbb{E}[r_i^{(d_i^*)}]|\right)^2\\
&\leq \frac{1}{m}\sum_{i=n+1}^{n+m}\left(\sum_{d=0}^{1}|\widehat{r}_i^{(d)}-\mathbb{E}[r_i^{(d)}]|\right)^2\\
&= \frac{1}{m}\sum_{i=n+1}^{n+m}\left(\sum_{d=0}^{1}(\widehat{r}_i^{(d)}-\mathbb{E}[r_i^{(d)}])^2\right.\\
&\left.+\sum_{d=0}^{k-1}\sum_{d'=0,d'\neq d}^{1}|\widehat{r}_i^{(d)}-\mathbb{E}[r_i^{(d)}]||\widehat{r}_i^{(d')}-\mathbb{E}[r_i^{(d')}]|\right)\\
&\leq \frac{4}{m}\max_{d\in[\![k]\!]_0}\sum_{i=n+1}^{n+m}\left(\widehat{r}_i^{(d)}-\mathbb{E}[r_i^{(d)}]\right)^2
\end{aligned}
$$

Assumption E.3 (Expected Reward Gap)**.** *Suppose that for each unit type d, $\boldsymbol{\beta}^{(d)}, \widehat{\boldsymbol{\beta}}^{(d)} \in [-\bar{\beta}, \bar{\beta}]^{T_0}$ for $\bar{\beta} \in \mathbb{R}_+$ and there is a gap in the principal's expected reward between assigning units their type and assigning them any other intervention. Formally, for some $\alpha > 0$ (specified in Corollary E.4) for each unit type $d \in [\![k]\!]_0$: $\forall v \in \mathcal{V}^{(d)}, \quad \mathbb{E}[r_v^{(d)}] - \mathbb{E}[r_v^{(d')}] > \gamma^{(d,d')}$ for all $d' < d$, where*

$$\gamma^{(d,d')} := (\sqrt{T_0}+\delta)(\|\boldsymbol{\beta}^{(d)}-\widehat{\boldsymbol{\beta}}^{(d)}\|_2+\|\boldsymbol{\beta}^{(d')}-\widehat{\boldsymbol{\beta}}^{(d')}\|_2)+\delta\|\boldsymbol{\beta}^{(d)}-\boldsymbol{\beta}^{(d')}\|_2+6\sigma\bar{\beta}\sqrt{2T_0\log(1/\alpha)},$$

and δ, σ are defined as in Definition 2.3 and Assumption 4.1 respectively.

The gap in Assumption E.3 depends on three terms: one which goes to zero as $\{\widehat{\boldsymbol{\beta}}^{(d)}\}_{d=0}^{k-1} \to \{\boldsymbol{\beta}^{(d)}\}_{d=0}^{k-1}$, one which is proportional to the maximum amount of modification possible in the pre-treatment period, and one which is proportional to the amount of measurement noise. Note that under Assumption E.3, separation of types (Condition 3.2) holds by design (in expectation). Intuitively, such a gap between unit rewards allows the principal to incentivize *truthful* unit behavior, as it is possible to design an intervention policy such that no unit has an incentive to modify their pre-intervention outcomes. When units are truthful, linear intervention policies are optimal due to Lemma 2.7.

Corollary E.4. *Suppose the principal's expected rewards satisfy the gap assumption (Assumption E.3). Then, for any $\alpha > 0$, with probability at least $1-\alpha$, Algorithm 2 achieves out-of-sample performance*

$$\frac{1}{m}\sum_{i=n+1}^{n+m}\left(\mathbb{E}[r_i^{(d_i^A)}]-\mathbb{E}[r_i^{(d_i^*)}]\right)^2 \leq \frac{k^2}{m}\max_{d\in[\![k]\!]_0}\sum_{i=n+1}^{n+m}\left(\widehat{r}_i^{(d)}-\mathbb{E}[r_i^{(d)}]\right)^2,$$

where d_i^A is the intervention assigned to unit i by Algorithm 2, d_i^* is the optimal intervention to assign to (strategic) unit i, and $\widehat{r}_i^{(d)} := \langle \widehat{\boldsymbol{\beta}}^{(d)}, \mathbf{y}_{i,pre}\rangle$ is the estimated principal reward under intervention d.

The proof of Corollary E.4 proceeds analogously to that of Theorem 4.2. We begin by showing that the interventions assigned by the intervention policy of Algorithm 2 on strategic units match the interventions assigned according to the intervention policy in the following lemma. However, unlike in the proof of Theorem 4.2, we also show that behaving truthfully in the pre-intervention period is a (weakly) dominant strategy for each unit under Assumption E.3.

Lemma E.5. *Consider the setting of Corollary E.4 and the following intervention policy:*

Assign intervention $d_i^B = d$ to unit i if

$$\widehat{r}_i^{(d)} - \widehat{r}_i^{(d')} > 0 \text{ for all } d' < d \quad \text{and} \quad \widehat{r}_i^{(d)} - \widehat{r}_i^{(d')} \geq 0 \text{ for all } d' > d, \tag{9}$$

where $\{\widehat{\boldsymbol{\beta}}^{(d)}\}_{d\in[\![k]\!]_0}$ are defined as in Algorithm 2. For any $\alpha > 0$, the intervention policy of Algorithm 2 assigns the same interventions to strategic units that intervention policy (9) assigns to truthful units with probability at least $1-\alpha$.

Proof. Suppose that the intervention policy of Algorithm 2 would assign intervention $d_i^A = d$ to unit i if they were truthful. Unit i cannot obtain any intervention $d' > d$ under the intervention policy of Algorithm 2, due to an argument analogous to Case 2 in the proof of Lemma E.2.

Next we show that if $\tilde{\mathbf{y}}_{i,pre} = \mathbf{y}_{i,pre}$, the intervention policy of Algorithm 2 also assigns intervention d to unit i. Consider $d' < d$.

$$\begin{aligned}\langle \widehat{\boldsymbol{\beta}}^{(d)} - \widehat{\boldsymbol{\beta}}^{(d')}, \tilde{\mathbf{y}}_{i,pre}\rangle - \delta\|\widehat{\boldsymbol{\beta}}^{(d)} - \widehat{\boldsymbol{\beta}}^{(d')}\|_2 \geq \mathbb{E}[r_i^{(d)}] - \mathbb{E}[r_i^{(d')}] - 6\sigma\bar{\beta}\sqrt{2T_0\log(1/\alpha)} \\ - (\sqrt{T_0} + \delta)(\|\boldsymbol{\beta}^{(d)} - \widehat{\boldsymbol{\beta}}^{(d)}\|_2 + \|\boldsymbol{\beta}^{(d')} - \widehat{\boldsymbol{\beta}}^{(d')}\|_2) \\ - \delta\|\boldsymbol{\beta}^{(d)} - \boldsymbol{\beta}^{(d')}\|_2,\end{aligned}$$

with probability at least $1-\alpha$, which follows from algebraic manipulation and a Hoeffding bound. Since the expected reward gap is sufficiently large, $\langle \widehat{\boldsymbol{\beta}}^{(d)} - \widehat{\boldsymbol{\beta}}^{(d')}, \tilde{\mathbf{y}}_{i,pre}\rangle - \delta\|\widehat{\boldsymbol{\beta}}^{(d)} - \widehat{\boldsymbol{\beta}}^{(d')}\|_2 > 0$ with probability at least $1-\alpha$ if $\tilde{\mathbf{y}}_{i,pre} = \mathbf{y}_{i,pre}$. Therefore, unit i can receive intervention d under the intervention policy of Algorithm 2 by behaving truthfully with probability at least $1-\alpha$. □

If $d_i^A = d_i^*$, then $\mathbb{E}[r_i^{(d_i)}] - \mathbb{E}[r_i^{(d_i^*)}] = 0$. If $d_i^A \neq d_i^*$, we know that

$$\mathbb{E}[r_i^{(d_i^*)}] - |\widehat{r}_i^{(d_i^*)} - \mathbb{E}[r_i^{(d_i^*)}]| \leq \widehat{r}_i^{(d_i^*)} \leq \widehat{r}_i^{(d_i^A)} \leq \mathbb{E}[r_i^{(d_i^A)}] + |\widehat{r}_i^{(d_i^A)} - \mathbb{E}[r_i^{(d_i^A)}]|.$$

Therefore,

$$\begin{aligned}\frac{1}{m}\sum_{i=n+1}^{n+m}\left(\mathbb{E}[r_i^{(d_i^A)}] - \mathbb{E}[r_i^{(d_i^*)}]\right)^2 &\leq \frac{1}{m}\sum_{i=n+1}^{n+m}\left(|\widehat{r}_i^{(d_i^A)} - \mathbb{E}[r_i^{(d_i^A)}]| + |\widehat{r}_i^{(d_i^*)} - \mathbb{E}[r_i^{(d_i^*)}]|\right)^2 \\ &\leq \frac{1}{m}\sum_{i=n+1}^{n+m}\left(\sum_{d=0}^{k-1}|\widehat{r}_i^{(d)} - \mathbb{E}[r_i^{(d)}]|\right)^2 \\ &= \frac{1}{m}\sum_{i=n+1}^{n+m}\left(\sum_{d=0}^{k-1}(\widehat{r}_i^{(d)} - \mathbb{E}[r_i^{(d)}])^2\right. \\ &\quad \left. + \sum_{d=0}^{k-1}\sum_{d'=0, d'\neq d}^{k-1}|\widehat{r}_i^{(d)} - \mathbb{E}[r_i^{(d)}]||\widehat{r}_i^{(d')} - \mathbb{E}[r_i^{(d')}]|\right) \\ &\leq \frac{k^2}{m}\max_{d\in[\![k]\!]_0}\sum_{i=n+1}^{n+m}\left(\widehat{r}_i^{(d)} - \mathbb{E}[r_i^{(d)}]\right)^2\end{aligned}$$

E.1. Application of PCR to obtain end-to-end guarantees

In order to leverage the out-of-sample guarantees for PCR, we make the following two assumptions on $\{\mathbb{E}[Y_{pre}^{(d)}]\}_{d=0}^{k-1}$ and $\mathbb{E}[Y'_{pre}]$ (as defined in Appendix F).

Assumption E.6 (Subspace Inclusion)**.** *The rowspace of* $\mathbb{E}[Y'_{pre}]$ *is contained within that of* $\mathbb{E}[Y_{pre}^{(d)}]$*, for all* $d \in [\![k]\!]_0$.

Assumption E.6 can be thought of as a sufficient condition for "causal transportability" from the n training units to the m out-of-sample units. Next we impose a condition on the singular values of $\mathbb{E}[Y_{pre}^{(0)}]$ and $\mathbb{E}[Y_{pre}^{(1)}]$, although this assumption is not strictly necessary for the PCR results of Agarwal et al. (2021b) to apply, as more general results may be obtained in terms of a *signal to noise* ratio, albeit at the cost of generally worse rates.

Assumption E.7 (Balanced Spectra)**.** *For all* $d \in [\![k]\!]_0$*, the* s *non-zero singular values* $\{s_l^{(d)}\}_{l=1}^r$ *of* $\mathbb{E}[Y_{pre}^{(d)}]$ *satisfy* $s_l^{(d)} = \Theta(\sqrt{n^{(d)}T_0/s})$ *for* $l \in [\![s]\!]$*. The* s *non-zero singular values* $\{s'_l\}_{l=1}^s$ *of* $\mathbb{E}[Y'_{pre}]$ *satisfy* $s'_l = \Theta(\sqrt{mT_0/s})$ *for* $l \in [\![s]\!]$.

We are now ready to state our formal result for the convergence rates of Algorithm 2 using PCR. An analogous bound may be obtained for Algorithm 1.

Corollary E.8. *Consider $\boldsymbol{\beta}^{(d)} \in \text{rowspan}(\mathbb{E}[Y_{pre}^{(d)}])$ for $d \in [\![k]\!]_0$ and the procedure of Algorithm 2, where $\{\widehat{\boldsymbol{\beta}}^{(d)}\}_{d=0}^{k-1}$ are given by PCR with $p = s$. Let $\|\boldsymbol{\beta}^{(d)}\|_2 = \Omega(1)$ and $\|\boldsymbol{\beta}^{(d)}\|_1 = \mathcal{O}(\sqrt{T_0})$ for all $d \in [\![k]\!]_0$. If Assumptions 4.1 (latent factor model), E.3 (expected reward gap), E.6 (subspace inclusion), E.7 (balanced spectra) hold, then with probability at least $1 - \mathcal{O}(\alpha + \sum_{d=0}^{k-1}((n^{(d)} \wedge m)T_0)^{-10})$,*

$$\frac{1}{m}\sum_{i=n+1}^{n+m}\left(\mathbb{E}[r_i^{(d_i^A)}] - \mathbb{E}[r_i^{(d_i^*)}]\right)^2 \leq k^2 \max_{d \in [\![k]\!]_0} C_{noise} s^3 \log((n^{(d)} \wedge m)T_0)$$
$$\cdot \left(\left(\frac{1 \vee \frac{T_0}{m}}{n^{(d)} \wedge T_0} + \frac{n^{(d)} \vee T_0}{(n^{(d)} \wedge T_0)^2} + \frac{1}{m}\right)\|\boldsymbol{\beta}^{(d)}\|_1^2 + \left(\frac{\sqrt{n^{(d)}}}{n^{(d)} \wedge T_0}\right)\|\boldsymbol{\beta}^{(d)}\|_1\right)$$

The proof of Corollary E.8 follows from applying Theorem F.1 to Corollary E.4.

F. Further Background on Principal Component Regression

Principal component regression We first describe the basics of PCR, using the notation of our setting. Let

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

and

$$Y'_{pre} := [\mathbf{y}_{n+i,pre}^\top : i \in [\![m]\!]] \in \mathbb{R}^{m \times T_0}.$$

Denote the singular value decomposition of $Y_{pre}^{(d)}$ as

$$Y_{pre}^{(d)} = \sum_{l=1}^{n^{(d)} \wedge T_0} s_l^{(d)} \widehat{\mathbf{u}}_l^{(d)} (\widehat{\mathbf{v}}_l^{(d)})^\top,$$

where $s_l^{(d)} \in \mathbb{R}$ is the l-th singular value, $\widehat{\mathbf{u}}_l^{(d)} \in \mathbb{R}^{n^{(d)}}$ is the l-th left singular vector, and $\widehat{\mathbf{v}}_l^{(d)} \in \mathbb{R}^{T_0}$ is the l-th right singular vector. Denote the vector of observed principal rewards under intervention d as

$$\mathbf{r}^{(d)} := [r_i^{(d)} : i \in \mathcal{N}^{(d)}]^\top \in \mathbb{R}^{n^{(d)}}.$$

For a given hyperparameter $p \leq n^{(d)} \wedge T_0$, we can use PCR to estimate $\boldsymbol{\beta}^{(d)}$ as

$$\widehat{\boldsymbol{\beta}}^{(d)} := \left(\sum_{l=1}^{p} \frac{1}{s_l^{(d)}} \widehat{\mathbf{v}}_l^{(d)} (\widehat{\mathbf{u}}_l^{(d)})^\top\right) \mathbf{r}^{(d)}.$$

In order to perform out-of-sample prediction, PCR first de-noises Y'_{pre} by computing

$$\widehat{Y}'_{pre} = \sum_{l=1}^{p} s_l \widehat{\mathbf{u}}_l \widehat{\mathbf{v}}_l^\top, \text{ where } Y'_{pre} = \sum_{l=1}^{m \wedge T_0} s_l \widehat{\mathbf{u}}_l \widehat{\mathbf{v}}_l^\top.$$

Principal rewards are then estimated as $\widehat{\mathbf{r}}^{(d)} := \widehat{Y}'_{pre} \widehat{\boldsymbol{\beta}}^{(d)}$.

We now restate the out-of-sample prediction results for PCR of Agarwal et al. (2021b), using the notation of our setting. See Table 2 for a summary of the key similarities/differences in our notation. Let $\mathsf{snr}^{(d)}$ denote the "signal-to-noise" ratio of $\mathbb{E}[Y_{pre}^{(d)}]$, defined as

$$\mathsf{snr}^{(d)} := \frac{s_s^{(d)}}{\sqrt{n^{(d)}} + \sqrt{T_0}},$$

where $s_s^{(d)}$ is the smallest non-zero singular value of $Y_{pre}^{(d)}$. Similarly, we can define the signal-to-noise ratio of $\mathbb{E}[Y'_{pre}]$ as

$$\mathsf{snr}_{test} := \frac{s_s}{\sqrt{m} + \sqrt{T_0}},$$

where s_s is the smallest non-zero singular value of $\mathbb{E}[Y'_{pre}]$.

Table 2: A summary of the main notational differences between our setting and that of Agarwal et al. (2021b).

Notation of Agarwal et al. (2021b)	Our Notation
C_{noise}	C_{noise}
n	$n^{(d)}$
m	m
r	s
r'	s
ρ	1
p	T_0
$\boldsymbol{\beta}^*$	$\boldsymbol{\beta}^{(d)}$
b	$T - T_0$
X	$\mathbb{E}[Y_{pre}^{(d)}]$
X'	$\mathbb{E}[Y'_{pre}]$
Z	$Y_{pre}d$
Z'	Y'_{pre}
$\mathbf{y}$	$\mathbb{E}[\mathbf{r}^{(d)}]$

Theorem F.1 (Agarwal et al. (2021b))**.** *Let Assumption 2.1 and Assumption E.6 hold. Consider* $\boldsymbol{\beta}^{(d)} \in \text{rowspan}(\mathbb{E}[Y_{pre}^{(d)}])$ *and PCR with* $p = r$*. Let (i)* $\mathsf{snr}^{(d)} \geq C_{noise}$*, (ii)* $\|\boldsymbol{\beta}^{(d)}\|_2 = \Omega(1)$*, (iii)* $\|\boldsymbol{\beta}^{(d)}\|_1 = \mathcal{O}(\sqrt{T_0})$*. Then with probability at least* $1 - \mathcal{O}(1/((n^{(d)} \wedge m)T_0)^{10})$*,*

$$MSE_{test} := \frac{1}{m} \sum_{i=n+1}^{n+m} \left(\widehat{r}_i^{(d)} - \mathbb{E}[r_i^{(d)}]\right)^2 \leq C_{noise} \log((n^{(d)} \wedge m)T_0)$$
$$\cdot \left(\frac{s(1 \vee \frac{T_0}{m})\|\boldsymbol{\beta}^{(d)}\|_1^2}{(\mathsf{snr}^{(d)})^2} + \frac{s(n^{(d)} \vee T_0)\|\boldsymbol{\beta}^{(d)}\|_1^2}{(\mathsf{snr}^{(d)})^4} + \frac{s\|\boldsymbol{\beta}^{(d)}\|_1^2}{\mathsf{snr}_{test}^2 \wedge m} + \frac{\sqrt{n^{(d)}}\|\boldsymbol{\beta}^{(d)}\|_1}{(\mathsf{snr}^{(d)})^2} \right),$$

where $\widehat{r}_i^{(d)} := \langle \widehat{\boldsymbol{\beta}}^{(d)}, \mathbf{y}_{i,pre} \rangle$ *and* C_{noise} *depends only on constants related to noise terms. Furthermore,*

$$\mathbb{E}[MSE_{test}] \leq$$
$$C_{noise} s \log((n^{(d)} \wedge m)T_0) \left(\frac{1 \vee \frac{T_0}{m}}{(\mathsf{snr}^{(d)})^2} + \frac{n^{(d)} \vee T_0}{(\mathsf{snr}^{(d)})^4} + \frac{1}{\mathsf{snr}_{test}^2 \wedge m} \right) \|\boldsymbol{\beta}^{(d)}\|_1^2 + \frac{C(T-T_0)^2}{((n^{(d)} \wedge m)T_0)^{10}}$$

for some $C > 0$*.*

Corollary F.2 (Agarwal et al. (2021b))**.** *Let the setup of Theorem F.1 hold. Further, let Assumption E.7 hold. Then with probability at least* $1 - \mathcal{O}(1/((n^{(d)} \wedge m)T_0)^{10})$*,*

$$MSE_{test} \leq C_{noise} s^3 \log((n^{(d)} \wedge m)T_0)$$
$$\cdot \left(\left(\frac{1 \vee \frac{T_0}{m}}{n^{(d)} \wedge T_0} + \frac{n^{(d)} \vee T_0}{(n^{(d)} \wedge T_0)^2} + \frac{1}{m} \right) \|\boldsymbol{\beta}^{(d)}\|_1^2 + \left(\frac{\sqrt{n^{(d)}}}{n^{(d)} \wedge T_0} \right) \|\boldsymbol{\beta}^{(d)}\|_1 \right).$$

Further,

$$\mathbb{E}[MSE_{test}] \leq C_{noise} s^3 \log((n^{(d)} \wedge m)T_0)$$
$$\cdot \left(\frac{1 \vee \frac{T_0}{m}}{n^{(d)} \wedge T_0} + \frac{n^{(d)} \vee T_0}{(n^{(d)} \wedge T_0)^2} + \frac{1}{m} \right) \|\boldsymbol{\beta}^{(d)}\|_1^2 + \frac{C(T-T_0)^2}{((n^{(d)} \wedge m)T_0)^{10}}$$

G. Experiments

The dataset we use can be found at `https://raw.githubusercontent.com/susanli2016/Machine-Learning-with-Python/master/data/Sales_Product_Price_by_Store.csv`.

We ran experiments on a 2020 MacBook Air with 16 GB of RAM.

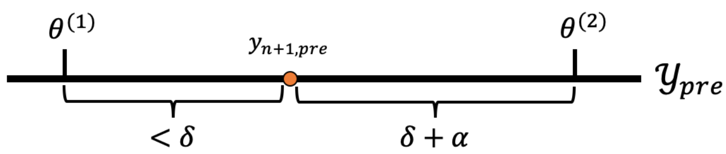

Figure 4: Visualization of the setting in Example H.1. If $\theta^{(1)}$ and $\theta^{(2)}$ are perfectly known to the principal, unit $n+1$ will modify their pre-intervention outcome to receive intervention 1 (left of $\theta^{(1)}$). However if the principal's estimate of $\theta^{(2)}$ is sufficiently inaccurate, unit $n+1$ may be able to modify their pre-intervention outcome to receive intervention 2 (right of $\theta^{(2)}$).

H. Example from Section 4

To build intuition as to why such a gap may be necessary, consider the following example.

Example H.1. *Consider the one-dimensional setting in Figure 4, where the unit can manipulate by δ in either direction. Specifically, let $\mathcal{Y}_{pre} \in \mathbb{R}$, $d \in \{0, 1, 2\}$, and let the optimal intervention policy*[4] *be:*

$$d_i^* = \begin{cases} 1 & \text{if } \tilde{y}_{i,pre} \leq \theta^{(1)} \\ 2 & \text{if } \tilde{y}_{i,pre} \geq \theta^{(2)} \\ 0 & \text{o.w.} \end{cases} \tag{10}$$

for some $\theta^{(1)} \in \mathbb{R}$, $\theta^{(2)} \in \mathbb{R}$, and $\theta^{(1)} < \theta^{(2)}$. Suppose the principal deploys the following "plug-in" estimate of intervention policy (10):

$$d_i^P = \begin{cases} 1 & \text{if } \tilde{y}_{i,pre} \leq \widehat{\theta}^{(1)} \\ 2 & \text{if } \tilde{y}_{i,pre} \geq \widehat{\theta}^{(2)} \\ 0 & \text{o.w.} \end{cases}$$

for $\widehat{\theta}^{(1)} \in \mathbb{R}$ and $\widehat{\theta}^{(2)} \in \mathbb{R}$. Moreover, suppose that the principal has perfect knowledge of $\theta^{(1)}$ (i.e., $\widehat{\theta}^{(1)} = \theta^{(1)}$) and after observing data from n non-strategic units, $|\widehat{\theta}^{(2)}(n) - \theta^{(2)}| = c/n$ for some $c > 0$. Now consider a strategic unit $n+1$ with $y_{n+1,pre} = y$ such that $y - \theta^{(1)} < \delta$ and $\theta^{(2)} - y = \delta + \alpha$, for some $\alpha > 0$. If the principal had perfect knowledge of $\theta^{(2)}$, unit $n+1$'s best-response would be to modify their pre-treatment outcome to $\theta^{(1)}$ and receive intervention 1. However, given the expression for $\widehat{\theta}^{(2)}$, we can write the best-response of unit $n+1$ as

$$\tilde{y}_{n+1,pre} = \begin{cases} \theta^{(1)} & \text{if } \widehat{\theta}^{(2)} = \theta^{(2)} + \frac{c}{n} \\ \widehat{\theta}^{(2)} = y + \delta + \alpha - \frac{c}{n} & \text{if } \widehat{\theta}^{(2)} = \theta^{(2)} - \frac{c}{n} \end{cases}$$

as long as the number of non-strategic units $n < \frac{c}{\alpha}$. Therefore under this setting, if the $(n+1)$-st unit has pre-intervention outcome y, then their best-response may be highly discontinuous as long as $n < \frac{c}{\alpha}$ (recall that α can be chosen to be arbitrarily small), despite the fact that $\theta^{(1)}$ is perfectly known to the principal and $\widehat{\theta}^{(2)}$ converges to $\theta^{(2)}$ at the "fast" rate of $\mathcal{O}(1/n)$.

[4]We do not specify the principal's reward function so as to simplify the exposition.