Learning the Distribution Map in Reverse Causal Performative Prediction

Daniele Bracale[*]
dbracale@umich.edu
Department of Statistics
University of Michigan

Subha Maity[*]
smaity@uwaterloo.ca
Department of Statistics
& Actuarial Science
University of Waterloo

Yuekai Sun
yuekai@umich.edu
Department of Statistics
University of Michigan

Moulinath Banerjee
moulib@umich.edu
Department of Statistics
University of Michigan

Abstract

In numerous predictive scenarios, the predictive model affects the sampling distribution; for example, job applicants often meticulously craft their resumes to navigate through a screening system. Such shifts in distribution are particularly prevalent in social computing, yet, the strategies to learn these shifts from data remain remarkably limited. Inspired by a microeconomic model that adeptly characterizes agents' behavior within labor markets, we introduce a novel approach to learning the distribution shift. Our method is predicated on a *reverse causal model*, wherein the predictive model instigates a distribution shift exclusively through a finite set of agents' actions. Within this framework, we employ a microfoundation model for the agents' actions and develop a statistically justified methodology to learn the distribution shift map, which we demonstrate to effectively minimize the performative prediction risk.

1 Introduction

Predictive models frequently guide decisions in strategic settings, but their environmental impact is often ignored. For example, a lender assessing loan default risk to set interest rates might inadvertently increase default risk for riskier applicants by charging higher rates, creating a self-fulfilling prophecy. In addition, borrowers might manipulate their profiles to appear

less risky and secure lower rates. These influences are widespread: crime predictions shape police resource allocation, affecting crime rates, while recommendations alter user preferences and consumption. Goodhart's law (Strathern, 1997), summarized as "when a measure becomes a target, it ceases to be a good measure," also hints how predictive models affect their environments through performative *agents*.

From a probabilistic perspective, performativity manifests itself as a shift in distribution; agents adapt to prediction models by altering their attributes over time. Returning to the default risk assessment example, the self-fulfilling prophecy leads to label shift (Lipton et al., 2018), while borrowers manipulating their profiles cause a class conditional shift (Maity et al., 2022). To mitigate these changes, while practitioners often retrain models (Cormier et al., 2016), a more proactive measure is to incorporate environmental impacts during model training, known as *strategic/performative prediction* (Hardt et al., 2015; Perdomo et al., 2020), which adjusts risk minimization to account for distribution shifts. This method, while effective, requires understanding these shifts – a topic not well-covered in the literature. *Our paper proposes a framework to estimate these distribution shifts*, which makes training of prediction tools for performative environments more accessible to practitioners.
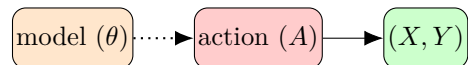

Figure 1: Reverse causal performative prediction

In this paper, we reformulate the setup of *reverse causal performative prediction* as introduced in Somerstep et al. (2024), which can be roughly stated as "the agents doctor their attributes by taking appropriate actions, and the distributions of modified attributes depend on the predictive model only through these actions". To formalize the assumption, we denote the *ex-post* distribution of attribute $(X, Y, A) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{A}$ after deploying

---
* Equal contribution.

a model $\theta \in \Theta$ as $\mathcal{D}(\theta)$, where X denotes the covariates, Y is the response and A is the action. The parameter space Θ is of appropriate dimension, depending on the model used by the learner (*e.g.* $\dim(\Theta) = \dim(\mathcal{X}) + 1$ for linear models and $\dim(\Theta) \gg \dim(\mathcal{X})$ for neural networks). Our reformulation states that $\mathcal{D}(\theta)$ depends only through the distribution of agents' action, *i.e.* having specified an action $a \in \mathcal{A}$ the $\mathcal{D}(X, Y \mid A = a; \theta)$ is invariant of θ and therefore, denoting the invariant distribution as $\mathcal{D}(X, Y \mid A)$ the $\mathcal{D}(\theta)$ is decomposed as

$$\mathcal{D}(\theta) = \mathcal{D}_A(\theta) \times \mathcal{D}(X, Y \mid A). \tag{1.1}$$

In the second paragraph of our conclusion Section 6 we discuss this reverse causal model into the broader context of performative prediction, but for now we move on toward our goal of learning $\mathcal{D}(\theta)$ within this reverse causal model. To learn $\mathcal{D}(\theta)$, in our paper, we assume that the learner has access to adequate i.i.d. samples from each of the $\mathcal{D}(X, Y \mid A)$, or better yet, we assume that $\mathcal{D}(X, Y \mid A)$ is known. Several instances where $\mathcal{D}(\cdot \mid A)$ is known can be found in Fang and Moro (2011, §§ 2.1, 3.1, 3.2, 4.1, 4.3, 6.6.1, 6.6.2) where they analyze models in which Z represents a score and A denotes group membership (e.g., *qualified* vs. *unqualified*) (Fang and Moro, 2011, §§ 3.2, 4.1, 4.3). Then, the only part that remains to be learned is $\mathcal{D}_A(\theta)$, *i.e.* the distribution map for the actions of the agents. A proposal for learning this $\mathcal{D}_A(\theta)$ follows.

To estimate $\mathcal{D}_A(\theta)$, we consider a microfoundation framework whose motivation comes from the Caote-Loury labor market model (Coate and Loury, 1993) (see Theorem 2.2). The best course of action of an agent is formalized by the *cost-adjusted utility maximization problem*:

$$A \triangleq \arg\max_{a \in \mathcal{A}} \mathrm{B}_a(\theta) - C_a, \tag{1.2}$$

where $\mathrm{B}_a(\theta)$ is the benefit of taking an action a under the model $\theta \in \Theta$, and C_a is the cost of the same a. Typically, the $\mathrm{B}_a(\theta)$ is known to the learner. Furthermore, we frame C_a as a random quantity, which encodes an agent's own preference for the action a. For example, while preparing for a test, a candidate from a wealthier background would be more willing to enroll in a test-prep course. Given that $\mathrm{B}_a(\theta)$ is known, in this microfoundation setting, we learn $\mathcal{D}_A(\theta)$ through learning the distribution of $\{C_a; a \in \mathcal{A}\}$. To keep things simple, in this paper we assume that *agent's action space $\mathcal{A}$ is finite*. While it may be possible to learn this distribution for an arbitrary $\mathcal{A}$, this will involve considerably more technical challenges and beyond the scope of this paper; further related discussion is deferred to the concluding Section 6.

At this point, it may seem that we have restricted our reverse causal setting to merely a microfoundation setting described in (1.2), where the agents are necessarily strategic. But we argue in Section 2.2 (*cf.* Theorem 2.5) that by incorporating a random cost into our microfoundation we make it completely general in the reverse causal setting, as, for a suitable choice of benefit $\mathbf{B}(\theta) \triangleq \{\mathrm{B}_a(\theta) : a \in \mathcal{A}\}$ and random cost $\mathbf{C} \triangleq \{C_a : a \in \mathcal{A}\}$, any $\mathcal{D}_A(\theta)$ can be expressed by our microfoundation setup. This is not possible in a usual strategic prediction setting (Hardt et al., 2015), where the deterministic cost can only allow a strategic agent. In contrast, our setting also allows for *non-strategic agents*. By *non-strategic* we mean that two agents with the same *ex-ante z* do not necessarily have the same *ex-post z'* if the cost is random. We defer further related discussion to Section 2.2.

We end the introductory section with a description of our contribution, followed by a brief survey of the related literature.

1. In Section 2 we introduce the microfoundations in reverse causal models for learning the distribution map, and we show that a microfoundation framework can recover any distribution shift in a reverse causal model by randomizing the cost.

2. In Section 3 we provide a method to estimate the distribution map non-parametrically, for a general number of actions.

3. In Section 4 we provide a sequential *optimal design* algorithm based on the doubling trick (*i.e.*, every episode is twice the previous one) with 2 actions. We prove a convergence result of the *relative efficiency* of the estimated design density.

4. In Section 5 we propose an algorithm based on the doubling trick to estimate the distribution map $\widetilde{\mathcal{D}}(\theta)$ with a regret bound of $\widetilde{\mathcal{O}}(M^{\frac{1}{1+\eta}})$, where η is the uniform convergence rate of the distribution map in the exploration phase and $\widetilde{\mathcal{O}}$ excludes factors $\log(M)$.

5. We consider the Coate-Loury labor market model Coate and Loury (1993) as a baseline, for which we show convergence results of the relative efficiency in both theoretical and numerical experiments.

1.1 Related work

The performative prediction, initially introduced by Perdomo et al. (2020), encompasses various aspects, with a notable instance being strategic classification Hardt et al. (2016); Shavit et al. (2020); Levanon and Rosenfeld (2022). In the context of strategic prediction, agents are presumed to be utility-maximizing, leading to an implicit determination of the distribution map by the agents' utilities. A growing body of literature

focuses on developing algorithms for computing performative optimal policies Izzo et al. (2021); Levanon and Rosenfeld (2021); Miller et al. (2021). In a related line of research, performativity is addressed in online settings with the aim of devising algorithms that effectively minimize Stackelberg regret Dong et al. (2018); Chen et al. (2020); Jagadeesan et al. (2022).

A key challenge that impedes the broader adoption of performative prediction lies in practitioners often not knowing the distribution map. In general, there are two ways to bypass this problem. The first treats the performative prediction problem (2.4) as a zeroth-order (*i.e.* derivative free) optimization problem, relying on samples from $\mathcal{D}(\theta_t)$ (θ_t being optimization algorithm iterates) to solve it without explicit knowledge of the distribution map $\mathcal{D}(\theta)$ Izzo et al. (2021); Miller et al. (2021). However, as the authors Miller et al. (2021) state, this method often leads to slow convergence and curse of dimensionality. The second approach deals with unknown agent responses by imposing microfoundation models Jang et al. (2022). This approach benefits from fast convergence rates, leveraging white-box access to the distribution map Hardt et al. (2016); Levanon and Rosenfeld (2021); Cutler et al. (2021), but the distribution map is often unknown in practice, and a misspecification in this map can hinder the finding of the correct solutions Lin and Zrnic (2023). Our proposed approach complements the second approach: by learning in a non-parametric way the microfoundations model from agent response data, we address issues related to misspecified distribution map, while benefiting from faster optimization algorithms by utilizing the white-box access to the distribution map. To the best of our knowledge, current results for estimating the distribution map assume a parametric model on $\mathcal{D}(\theta)$, for example, Miller et al. (2021) assumes a location-scale distribution map generated by the linear transformation $z_\theta = z_0 + \mu\theta$.

2 Reverse causal performative prediction

In a standard strategic prediction setting, agents update their covariates X, but do not update their responses Y. Horowitz and Rosenfeld (2023) considered a more general setting in which agents modify both X and Y using a causal/structural model. Here, agents still change their covariates, but such changes can propagate through the causal model and (indirectly) lead to changes in their responses. In their reverse causal model, Somerstep et al. (2024) assumes that the agents update their $Y \to Y'$, which leads to a change in $X \sim \Phi(\cdot \mid Y')$. We further generalize this by assuming that the agents can modify both $Z = (X, Y)$ as a re-

sponse to θ, but only through their action A, leading to the structural model in Figure 1. Thus, the *ex-post* distribution $\mathcal{D}(\theta)$ depends on θ only through the marginal distribution of A, *i.e.* through $\mathcal{D}_A(\theta)$. Since $\mathcal{D}(\theta)$ can be decomposed as $\mathcal{D}(\theta) = \mathcal{D}_A(\theta) \times \mathcal{D}(Z \mid A; \theta)$, the reverse causal setting is equivalently stated in the following assumption.

Assumption 2.1 (Reverse causal model)**.** *The $\mathcal{D}(Z \mid A; \theta)$ are not affected by the prediction model θ. Moving forward, we drop this dependence on θ and simply denote this by $\mathcal{D}(Z \mid A)$.*

Our reverse causal model simplifies to the one in Somerstep et al. (2024, Section 2) if we let $A = Y$.

A prominent example of reverse causal model is the Coate-Loury (CL) labor market model (Coate and Loury, 1993). We describe a variation of this model below, which will serve as a running example throughout our paper.

Example 2.2 (Coate-Loury labor market model (Coate and Loury, 1993))**.** *Employers seek to hire skilled workers, but while making hiring decisions, are uncertain whether the workers are skilled ($Y = 1$) or unskilled ($Y = 0$). Instead, they observe a score $X \in \mathcal{X} = [0,1]$, which is an (imperfect) skill assessment of workers and hire those whose scores exceed a threshold $\theta \in [0,1]$, i.e. hires if $X > \theta$. Workers respond strategically to the hiring policy by making skill investments:*

$$A \triangleq \arg\max_{a \in \{0,1\}} \omega \bar{F}(\theta \mid a) - C \cdot a, \quad (2.1)$$

where $\omega > 0$ is the wage paid to hired workers and $\bar{F}(\theta \mid a) = \mathbb{P}(X > \theta | Y = a)$ is the survival function of X, interpreted as the probability of hiring a (skilled and unskilled) worker, and finally, $C > 0$ is a random cost of worker-specific investment for acquiring skill $Y = 1$. Here, the action A is identical to the response Y: whether to be skilled ($A = Y = 1$) by making an investment or not ($A = Y = 0$), and thus we absorb Y within A while describing the distribution shift. As the distribution of $X \mid A$ remains unaffected by the threshold θ, the only component of the distribution of (X, A) that changes is the marginal distribution of A, leading to the distribution shift map

$$\mathcal{D}(X, A; \theta) = \mathcal{D}_A(\theta) \times \mathcal{D}(X \mid A), \quad \mathcal{D}_A(\theta) \overset{d}{=} \text{Ber}(\pi(\theta))$$

where $\overset{d}{=}$ denotes 'equally distributed' and $\pi(\theta)$ is the proportion of skilled workers, determined by the workers whose (expected) wage incentive $I(\theta) \in [0, \omega]$ as below:

$$\pi(\theta) \triangleq \mathbb{P}_\theta\{A = 1\} = \mathbb{P}\{C \le I(\theta)\}, \quad (2.2)$$

$$I(\theta) \triangleq \omega\{\bar{F}(\theta \mid 1) - \bar{F}(\theta \mid 0)\}. \quad (2.3)$$

2.1 Learning prediction model in reverse causal model

As a proactive approach to learning a model in a performative prediction setting, Hardt et al. (2015); Perdomo et al. (2020) propose to minimize *performative risk* by modifying the standard risk minimization problem to account for the shift in the underlying distribution:

$$\min_{\theta \in \Theta} \mathrm{PR}(\theta), \quad \mathrm{PR}(\theta) \triangleq \mathbb{E}_{\mathcal{D}(\theta)}[\ell(\theta; Z)], \quad (2.4)$$

where ℓ is an appropriate loss function. For our reverse causal model (Theorem 2.1), the performative risk simplifies to

$$\mathrm{PR}(\theta) = \sum_{a \in \mathcal{A}} \mathcal{D}_{A=a}(\theta)\mathbb{E}\big[\ell(\theta; Z) \mid A = a\big], \quad (2.5)$$

where $\mathbb{E}[\ell(\theta; Z) \mid A]$ is efficiently estimated given adequate samples from the distribution of $Z \mid A = a$. To keep things simple, we assume that $\mathcal{D}(Z \mid A = a)$ is known for the moment.

Example 2.3 (Coate-Loury model)**.** *Returning to Theorem 2.2, assume that the firm with hiring policy $\widehat{Y}_\theta(X) = \mathbb{I}\{X > \theta\}$ incurs a loss of $\delta_0 > 0$ (resp. gain of $\delta_1 > 0$) for hiring an unskilled (resp. skilled) worker. Expressing the gain as a negative loss, we can express the loss function as $\ell(\theta; Z) \triangleq \widehat{Y}_\theta(X)\{-\delta_1 Y + \delta_0(1 - Y)\}$. In this case the performative risk $\mathrm{PR}(\theta)$ is:*

$$\sum_{a \in \{0,1\}} \mathcal{D}_{A=a}(\theta)\mathbb{E}\big[\ell(\theta; Z) \mid A = a\big] \quad (2.6)$$
$$= -\pi(\theta)\delta_1 \bar{F}(\theta \mid 1) + (1 - \pi(\theta))\delta_0 \bar{F}(\theta \mid 0).$$

2.2 Microfoundation framework for learning the distribution shift

In order to minimize the performative risk (2.5), we require $\mathcal{D}(\theta)$. Since we assume that we know $\mathbb{E}[\ell(\theta; Z) \mid A]$, it remains to learn $\mathcal{D}_A(\theta)$. For this purpose, we propose a microfoundation framework: an agent pays a random cost C_a for their action a and strategically chooses their best course of action by maximizing their cost-adjusted utility, as described in (1.2)

$$A \triangleq \arg\max_{a \in \mathcal{A}} \mathrm{B}_a(\theta) - C_a$$

where, for an action a, $\mathrm{B}_a(\theta)$ is the benefit realized under the model θ, and C_a is the random cost that encodes the agent's own preference for the action. We eliminate the edge cases where the above arg max has multiple solutions and an agent is in conflict between multiple actions. Such conflictions can be easily eliminated by assuming that the distributions of C_a's are continuous, for which the arg max is unique with probability one.

Example 2.4 (Coate-Loury model)**.** *In our Theorem 2.2, one may set $\mathrm{B}_a(\theta) = \omega \bar{F}(\theta \mid A = a)$, $C_1 = C$ and $C_0 \equiv 0$.*

Under this microfoundation framework (1.2) the $\mathcal{D}_A(\theta)$ is expressed as:

$$\mathcal{D}_{A=a}(\theta) = \mathbb{P}\big\{ \mathrm{B}_a(\theta) - C_a \geq \mathrm{B}_{a'}(\theta) - C_{a'}, \forall a' \neq a \big\}. \quad (2.7)$$

Typically, the $\mathrm{B}_a(\theta)$ is known to the learner, *e.g.* in Coate Loury Theorem 2.2, the firm typically has access to $\mathcal{D}(X \mid A)$ from a historical dataset, and can easily calculate $\mathrm{B}_a(\theta) = \mathbb{P}(X > \theta \mid A = a)$. It only remains to learn the distribution of $\mathbf{C} \triangleq \{C_a; a \in \mathcal{A}\}$ and, moving ahead, in Section 3 we propose an algorithm for learning this. But, before moving ahead, we make the following remark about the generality of the microfoundation framework in the broader context of learning $\mathcal{D}(\theta)$ in a reverse causal model.

Generality of microfoundation framework: How restrictive is our framework (1.2), given that agents are strategic? As we show here, by incorporating a random cost in our microfoundation, we made it completely general in the reverse causal model that may include *non-strategic* agents as well. By *non-strategic* we mean that two agents with the same *ex-ante z* do not necessarily have the same *ex-post z'* if the cost is random. Moreover, in the following lemma, we show that, for a suitably chosen benefit, any $\mathcal{D}_A(\theta)$ can be expressed using the microfoundation framework.

Lemma 2.5. *For any $\mathcal{D}_A(\theta)$ and a benefit function $\mathbf{B}(\theta) \triangleq \{\mathrm{B}_a(\theta); a \in \mathcal{A}\}$ satisfying the property that for every $a, a' \in \mathcal{A}$ with $a \neq a'$, $\mathcal{D}_{A=a}(\theta)$ can be written as a fixed increasing function of $\mathrm{B}_a(\theta) - \mathrm{B}_{a'}(\theta)$, there exists a random cost vector $\mathbf{C}$, that induces $\mathcal{D}_A(\theta)$ through the microfoundation equation (1.2). Furthermore, if $\{\mathcal{D}_A(\theta) : \theta \in \Theta\} = \Delta^{\mathcal{A}}$ then for a given $\mathbf{B}(\theta)$ the distribution of such a $\mathbf{C}$ that satisfies $C_0 = 0$ is unique.*

The complete proof is in Appendix D, but here we explain the key idea in the simple case of $|\mathcal{A}| = 2$, *i.e.* $\mathcal{A} = \{0, 1\}$. We set $C_0 \equiv 0$. Since $\mathcal{D}_{A=1}(\theta)$ increases with $\mathrm{B}_1(\theta) - \mathrm{B}_0(\theta)$, *i.e.* $\mathcal{D}_{A=1}(\theta) = G(\mathrm{B}_1(\theta) - \mathrm{B}_0(\theta))$ for an increasing function $G : \mathbb{R} \to [0, 1]$, we extend G to an appropriate c.d.f. $\tilde{G}$ over $\mathbb{R}$, for which a random cost $C_1 \sim \tilde{G}$ leads to microfoundation (1.2), because

$$\mathbb{P}\big\{C_1 \leq \mathrm{B}_1(\theta) - \mathrm{B}_0(\theta)\big\} = \tilde{G}(\mathrm{B}_1(\theta) - \mathrm{B}_0(\theta))$$
$$= G(\mathrm{B}_1(\theta) - \mathrm{B}_0(\theta)) = \mathcal{D}_{A=1}(\theta). \quad (2.8)$$

If $\{\mathcal{D}_1(\theta) : \theta \in \Theta\} = [0, 1]$ then the uniqueness of the distribution of C_1 is easily realized by noticing that the collection $\{\mathrm{B}_1(\theta) - \mathrm{B}_0(\theta) : \theta \in \Theta\}$ specifies all the quantiles of the distribution of C_1.

Misspecification of benefit function: Although the learner must specify a $\mathbf{B}(\theta)$ for the microfoundation, Theorem 2.5 indicates that for any choice of $\mathbf{B}(\theta)$ that meets the the lemma's condition, the microfoundation

can accurately express $\mathcal{D}_A(\theta)$. This allows a flexible choice of $\mathbf{B}(\theta)$, as any misspecifications in the benefit are corrected within the estimated cost distribution, thereby ensuring the proper specification of $\mathcal{D}_A(\theta)$.

3 Learning the distribution of agents' action

In this section, we propose a simple algorithm for estimating the $\mathcal{D}_A(\theta)$, which is the remaining unknown part of the distribution shift. Recall the expression of $\mathcal{D}_a(\theta) \triangleq \mathcal{D}_{A=a}(\theta)$ in (2.7), which we first re-express in a compact vector inequality form with a matrix $\mathbf{L}_a = \{\mathbf{L}_a(u, v); u \in \mathcal{A} - \{a\}, v \in \mathcal{A}\} \in \mathbb{R}^{(\mathcal{A}-\{a\}) \times \mathcal{A}}$ defined as

$$\mathbf{L}_a(u, v) = \begin{cases} 1 & \text{if } v = a, u \in \mathcal{A} - \{a\} \\ -1 & \text{if } u = v \neq a \\ 0 & \text{otherwise} \end{cases}$$

With the $\mathbf{L}_a$ the $\mathcal{D}_a(\theta)$ can be compactly expressed as

$$\begin{aligned} \mathcal{D}_a(\theta) &= \mathbb{P}\{C_a - C_{a'} \leq \mathrm{B}_a(\theta) - \mathrm{B}_{a'}(\theta), \forall a' \neq a\} \\ &= \mathbb{P}\{\mathbf{L}_a \mathbf{C} \preceq \mathbf{L}_a \mathbf{B}(\theta)\} = \mathrm{F}_a\left(\mathbf{L}_a \mathbf{B}(\theta)\right) \end{aligned}$$
(3.1)

where $\preceq$ is the element-wise inequality operator, $\mathbf{C} \triangleq \{C_a; a \in \mathcal{A}\}$, $\mathbf{B}(\theta) \triangleq \{\mathrm{B}_a(\theta); a \in \mathcal{A}\}$, F_a is the multivariate cumulative distribution function (cdf) of $\mathbf{L}_a \mathbf{C}$. Learning $\mathcal{D}_A(\cdot)$ reduces to learning $\{\mathrm{F}_a(\cdot); a \in \mathcal{A}\}$. For this purpose, we assume that the learner deploys predictive models $\theta_1, \ldots, \theta_M$ and observes the estimated proportions of the actions of agents $\widehat{\pi}(\theta_1), \ldots, \widehat{\pi}(\theta_M)$, i.e., for each m the $\widehat{\pi}(\theta_m) = \{\widehat{\pi}_a(\theta_m), a \in \mathcal{A}\}$ consists of the estimated proportion of agents $\widehat{\pi}_a(\theta_m)$ who chose to take the action a as a response to the model θ_m. In the following remark, we explain how a learner may estimate these proportions.

Remark 3.1. *We explain two settings for estimating $\widehat{\pi}(\theta_m)$: with or without direct access to agents' actions. For the model θ_m*

(1) If the learner has access to the actions $\{A_{m,i}; i \in [n_m]\}$ from n_m i.i.d. draws from $\mathcal{D}(\theta_m)$ then they can directly estimate the proportions as $\widehat{\pi}_a(\theta_m) = \frac{1}{n_m} \sum_{i=1}^{n_m} \mathbb{I}\{A_{m,i} = a\}$.

(2) If the actions are not directly accessible to the learner, who can only access i.i.d. $\{Z_{m,i}; i \in [n_m]\}$ from $\mathcal{D}(\theta_m)$, the moment matching approach from Lipton et al. (2018) for estimating label shift can be easily adapted here. Since we assume that we know $\mathcal{D}(Z \mid A; \theta_m)$, for a 'suitable' choice of $h_m : \mathcal{Z} \to \mathbb{R}^{\mathcal{A}}$, that has a good discriminating power over $\mathcal{D}(Z \mid A; \theta_m)$, i.e. smallest singular

value of the following matrix is sufficiently larger than zero

$$\Delta_m \triangleq \left[\mathbb{E}_{\mathcal{D}(\theta_m)}[h_m(Z) \mid A = a]; a \in \mathcal{A}\right] \in \mathbb{R}^{\mathcal{A} \times \mathcal{A}},$$

the $\widehat{\pi}(\theta_m)$ can then be simply estimated as $\widehat{\pi}(\theta_m) = \Delta_m^{-1}\{\frac{1}{n_m} \sum_{i=1}^{n_m} h_m(Z_{m,i})\}$. As explained in Lipton et al. (2018, Theorem 3), the error in this estimator is inversely proportional to the smallest singular value of Δ_m.

Having estimated the $\widehat{\pi}(\theta_m)$'s, the $\{\mathrm{F}_a; a \in \mathcal{A}\}$ can be learned by solving a least squares (LS) problem with an appropriate monotone shape constraint. Recall that F_a is the cumulative distribution function (cdf) of $\mathbf{L}_a \mathbf{C}$ and thus a coordinate-wise monotone function. We incorporate this shape constraint in our LS problem:

$$\begin{aligned} &\sum_{a \in \mathcal{A}, m \in [M]} n_m \{\mathrm{F}_a\left(\mathbf{L}_a \mathbf{B}(\theta_m)\right) - \widehat{\pi}_a(\theta_m)\}^2, \\ &\text{subject to } \sum_{a \in \mathcal{A}} \mathrm{F}_a\left(\mathbf{L}_a \mathbf{B}(\theta_m)\right) = 1, \ \forall \ m \in [M], \end{aligned}$$
(3.2)

where F_a are multivariate cdfs. The loss corresponding to the estimate $\widehat{\pi}(\theta_m)$ is weighted by n_m based on its estimation error: a $\widehat{\pi}(\theta_m)$ calculated from a larger sample of size n_m is statistically more accurate and therefore is weighted with a weight of higher importance n_m. The LS problem in (3.2) is a coordinate-wise monotone regression problem (Chatterjee et al., 2018; Han et al., 2019); it is a convex quadratic program that can be efficiently solved. Armed with these $\widehat{\mathrm{F}}_a$, the learner plugs them in to obtain (an estimate of) the distribution map $\widehat{\mathcal{D}}_A(\theta)$, where for an $a \in \mathcal{A}$ the $\widehat{\mathcal{D}}_a(\theta) = \widehat{\mathrm{F}}_a(\mathbf{L}_a \mathbf{B}(\theta))$ and any θ. We summarize our approach in Algorithm 1.

Algorithm 1 Learning discrete agent responses

1: Observe agent responses $\widehat{\pi}(\theta_1), \ldots, \widehat{\pi}(\theta_M) \in \Delta^{|\mathcal{A}|-1}$ to predictive models $\theta_1, \ldots, \theta_M$
2: Estimate $\{\mathrm{F}_a, \ a \in \mathcal{A}\}$ as $\{\widehat{\mathrm{F}}_a; a \in \mathcal{A}\}$ with (coordinate-wise) monotone regression (3.2)
3: Estimate distribution map $\mathcal{D}_A(\cdot)$ as $\widehat{\mathcal{D}}_a(\cdot) \leftarrow \widehat{\mathrm{F}}_a(\mathbf{L}_a \mathbf{B}(\cdot)), a \in \mathcal{A}$

3.1 Example. Univariate cost: CL model.

Consider the CL model in Theorem 2.2, where let $\{\widehat{\pi}(\theta_m) : m \in [M]\} \subset [0, 1]$ be the estimated proportions of qualified workers (i.e. $\pi(\theta_m) \triangleq \mathcal{D}_{A=1}(\theta_m)$) after they have been exposed to the thresholds $\{\theta_m : m \in [M]\} \in \Theta = [0, 1]$; see Theorem 3.1 for their estimation. At this point, we set the cost of $A = 0$ (the worker remains unskilled) to zero and only estimate the random cost of becoming skilled ($A = 1$). Denoting

this cost by C, the optimization in (3.2) simplifies to just estimating F_C (c.d.f. of C) as

$$\widehat{F}_C \leftarrow \arg\min_{F \in \mathcal{F}} \sum_{m=1}^M n_m \{F(b_m) - \widehat{\pi}(\theta_m)\}^2, \ (3.3)$$

where $b_m = I(\theta_m)$ and $\mathcal{F}$ is a collection of non-negative and monotonically increasing functions.

Next, we provide a uniform convergence result for the estimated c.d.f. in (3.3) assuming that the learner observes the $A_{m,i}$'s. While it is possible to establish a similar result for the second estimator proposed in Theorem 3.1 which does not require these $A_{m,i}$'s, we leave it for future research interest as we believe this will be technically more challenging and verify the quality of estimated c.d.f. obtained from such $\widehat{\pi}(\theta_m)$ in a synthetic experiment. To establish uniform convergence, we assume that the true F_C is smooth and the design points b_m in (3.3) are "regular", which are formalized below.

Assumption 3.2. *(A1) For an $0 \leq \alpha < 1$ the true F_C is α-Hölder smooth, i.e. there exists a $K > 0$ such that $|F_C(b) - F_C(b')| \leq K|b - b'|^\alpha$ for any $b, b' \in [0, 1]$.*

(A2) The design points b_m are either equidistant over the range of $\{I(\theta) : \theta \in \Theta\} = [0, \omega]$ (ω is defined in Theorem 2.2) or drawn as i.i.d. *from a density that is bounded away from zero.*

Now we establish a uniform convergence result in the following proposition, which follows directly from (Mösching and Dümbgen, 2020, Theorem 3.3).

Proposition 3.3. *Suppose that $\widehat{\pi}(\theta_m) = {}^1/n_m \sum_{i=1}^{n_m} A_{m,i}$ and define $n \triangleq \sum_m n_m$. Under assumptions (A1) and (A2) the $\widehat{F}_C$ defined in (3.3) has the following rate of convergence:*

$$\sup_{b \in [\delta_n, w - \delta_n]} |\widehat{F}_C(b) - F_C(b)| = \mathcal{O}_P\left\{\left(\log n/n\right)^{\alpha/(2\alpha+1)}\right\}$$

where, the $\delta_n = \kappa\left(\log n/n\right)^{1/(2\alpha+1)}$ with an universal constant κ (see Mösching and Dümbgen (2020, Remark 3.1 and 3.2)).
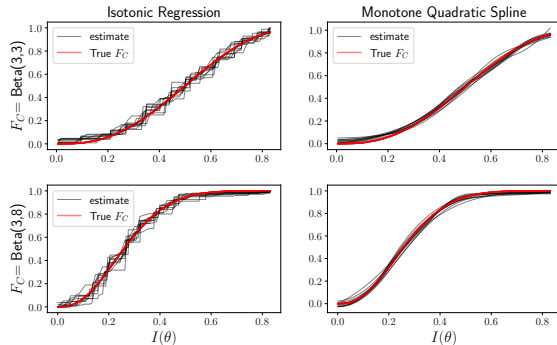

Figure 2: Estimation of F_C

Estimating proportions with unknown actions: Here we employ the second method proposed in Theorem 3.1, where the discriminating function h_m is chosen simply as the classifier for action A using X. This classifier can be fitted on a previously collected dataset, where both A and X are observed. Having obtained the $\widehat{\pi}(\theta_m)$ for thresholds θ_m, we use (3.3) to estimate $\widehat{F}_C$. In Figure 2 we compare the $\widehat{F}_C$ with the true F_C, where we utilized two techniques for estimating a monotonically increasing functions: isotonic regression (Robertson et al., 1988) and monotone quadratic spline (Meyer, 2008). As the latter approach yields a smooth estimate, this is a better estimator, because the true F_C is smooth.

While we move on to design of model θ_m that is to be deployed, in the appendix, we demonstrate how F_C can be estimated within an appropriate parametric model (see Appendix A), and results for estimating distribution map with multiple actions $|\mathcal{A}| \geq 3$ (see Appendix B).

4 Optimal design for deploying models under binary actions

In this section, we assume that $\mathcal{A} = \{0, 1\}$, the deployed models θ_m belong to a known bonded convex set Θ, and without loss of generality, $B_0(\theta) - B_1(\theta) \in [0, 1]$ for every $\theta \in \Theta$, so that $b_m = b(\theta_m) \triangleq \mathbf{L}_0\mathbf{B}(\theta_m) = B_0(\theta_m) - B_1(\theta_m) \in [0, 1]$. For a function $\theta \mapsto f(\theta)$, we denote $\|f\|_{L^\infty(\Theta)} = \sup_{\theta \in \Theta} |f(\theta)|$. We propose an optimal design algorithm that aims to find the distribution d of the i.i.d. values $b_m \in [0, 1]$ to minimize the mean integrated square error

$$\text{MISE}\{\widehat{F} \mid d\} \triangleq \mathbb{E}_d \int \{\widehat{F}(b) - F(b)\}^2 db$$

between the true cdf $F \triangleq F_1$, $(F_0 = 1 - F)$ and the estimated $\widehat{F} \triangleq \widehat{F}_1$ as defined in Section 3. Whenever we say that we draw a $b_m \in [0, 1]$, we also mean dawning the corresponding $\theta_m \in \Theta$ that leads to $b_m = b(\theta_m)$.

Example 4.1. *To build an intuition, assume that for each $m \in [M]$*

$$\widehat{\pi}(\theta_m) = F(b_m) + \epsilon(b_m), \quad \mathbb{E}[\epsilon] = 0, \text{var}[\epsilon] = \sigma^2(b_m).$$

Since $\mathcal{A} = \{0, 1\}$, the estimator in (3.2) reduces to $\widehat{F} = \arg\min_{F \in \mathcal{F}} \sum_m n_m(\widehat{\pi}(\theta_m) - F(b_m))^2$.

The design density d that minimizes the mean integrated squared error $\text{MISE}\{\widehat{F} \mid d\}$ is given by

$$d^\star(b) = {}^{\sigma(b)}/\int \sigma(u)du, \quad (4.1)$$

where $\sigma^2(b) = \text{var}[\widehat{F}(b) - F(b)]$ (see Müller (1984); Zhao and Yao (2012)) which, intuitively affirms that more

design points b_m are needed where the noise level for $\widehat{F}(b_m) - F(b_m)$ is high. In particular, if $\sigma(\cdot)$ is constant, then $d^\star(\cdot)$ is the uniform density. Although the design density takes the form of (4.1), we're yet to learn $\sigma(\cdot)$, which we learn through a sequential method.

4.1 Sequential method with doubling trick

Different approaches have been developed to estimate the optimal design; most related is *episode-sequential* design by Zhao and Yao (2012). Further developing on this design algorithm, we propose our new algorithm that uses a doubling trick to minimize the *relative efficiency loss* (Zhao and Yao, 2012) with respect to the optimal design $d^\star$ (as described in (4.1)):

$$\mathrm{REL}(d) \triangleq 1 - \frac{\mathrm{MISE}\{\widehat{F}|d^\star\}}{\mathrm{MISE}\{\widehat{F}|d\}}.$$

We divide the model choices $[M]$ into K episodes that increase exponentially as $L_k = \tau_o 2^{k-1}$, where τ_o is the (pre-fixed) length of the first episode. With this choice of L_k, we have $K = \lfloor \log_2(1 + M/\tau_o) \rfloor$. We state our algorithm in 2 and establish an upper bound for relative efficiency in Theorem 4.2.

Algorithm 2 Sequential Design Algorithm with doubling trick

1: Fix $\tau_o \in \mathbb{N}$, $K = \log_2(1 + M/\tau_o)$, $L_k = \tau_o 2^{k-1}$. Define the episodes $\mathcal{I}_1 = \{1, \ldots, L_1\}$ and $\mathcal{I}_k = \{1 + \sum_{j=1}^{k-1} L_j, \ldots, \sum_{j=1}^{k} L_j\}$, for $k = 2, \ldots, K$.
2: Let the initial density $\widehat{d}^{(1)}$ be the uniform in $[0,1]$.

3: **for** k in $[1, K]$ **do**
4:   For every $m \in \mathcal{I}_k$ sample $b_m \overset{\text{i.i.d.}}{\sim} \widehat{d}^{(k)}$ and record $\widehat{\pi}(\theta_m)$ (see Remark 3.1).
5:   Based on $\{\widehat{\pi}(\theta_m)\}_{m \in \mathcal{I}_k}$ update $\widehat{F}^{(k)}(b)$, $\widehat{\sigma}^{(k)}(b)$ and the design density $\widehat{d}^{(k+1)}(b) \propto \widehat{\sigma}^{(k)}(b)$.
6: **end for**

Theorem 4.2. *Assume that there exists a K_0 and a $\delta \in (0,1)$ such that $\|\widehat{d}^{(k)} - d^\star\|_{L^\infty(0,1)} = \mathcal{O}_P((\log(L_k)/L_k)^\delta)$ for all $k \geq K_0$ (where $\widehat{d}^{(k)}$ is defined in Algorithm 2). Then, as $M \to \infty$*

$$\mathrm{REL}(\widehat{d}^{(K)}) = \mathcal{O}_P((\log(M)/M)^\delta). \qquad (4.2)$$

In Theorem 4.2 $\mathrm{REL}(\widehat{d}^{(K)})$, has the same rate of convergence as the method in (Zhao and Yao, 2012, Theorem 2). The advantage of Theorem 4.2 is that, at every episode k being $\widehat{d}^{(k)}$ a functional of i.i.d. data, then the unknown δ can be determined using classical uniform convergence results (*e.g.* Mösching and Dümbgen (2020); Yang and Barber (2019)), which was not possible in Zhao and Yao (2012), where the

data are not i.i.d. When $\sigma(\cdot)$ is a continuous function of F (see *e.g.* Example 4.3, where $\sigma^2 \propto F\bar{F}$) the value δ coincides with the value of η that satisfies $\|\widehat{F}^{(k)} - F\|_{L^\infty(0,1)} = \mathcal{O}_P((\log(L_k)/L_k)^\eta)$. In such cases, for non-parametric estimates such as kernel and spline (e.g. (Zhao and Yao, 2012, §3)), we have $\delta = \eta = \beta/2\beta+1$ where β is the degree of smoothness of F.
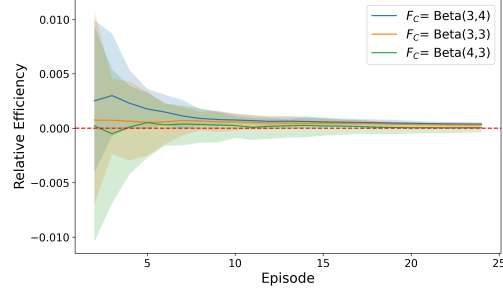

Figure 3: Relative efficiencies and their 95% confidence intervals for three different F_C's.

Example 4.3 (CL model). *Going back to Theorem 2.2, continued in Section 3.1, suppose that $w = 1$, so that $b_m \in [0,1]$. Suppose that for every b_m we observe $\widehat{\pi}(\theta_m) = 1/n_m \sum_{i=1}^{n_m} A_{m,i}$, where n_m is the number of observations at point b_m. We have $\widehat{\pi}(\theta_m) = F_C(b_m) + \epsilon(b_m)$ where $\epsilon(b_m) \triangleq 1/n_m \sum_{i=1}^{n_m} (A_{m,i} - F_C(b_m))$, $\mathbb{E}[\epsilon(b)] = 0$ and $\sigma^2(b) = \mathbb{V}[\epsilon(b)] \propto F_C(b)\bar{F}_C(b)$. Thus, the optimal density design would be $d^\star(b) \propto (F_C(b)\bar{F}_C(b))^{1/2}$. Since F_C is not known, we use Algorithm 2 to estimate it. In particular, at episode k, we estimate F_C using isotonic regression using data $\{b_m, \widehat{\sigma}^2(b_m) \propto \widehat{\pi}(\theta_m)(1 - \widehat{\pi}(\theta_m))\}_{m \in \mathcal{I}_k}$. In Figure 3 we see that the relative efficiency converges to zero as K increases.*

5 Regret analysis on performative risk

In this section, we demonstrate how performative risk can be easily minimized by "plugging-in" the estimated distribution map. As in Section 4, assume that $\mathcal{A} = \{0,1\}$ and without loos of generality $b_m = b_m(\theta_m) \in [0,1]$. In order to find the *performative optimum* the learner would minimize the performative risk (2.5)

$$\theta^\star \in \arg\min_{\theta \in \Theta} \left\{ \mathrm{PR}(\theta) \triangleq \sum_{a \in \mathcal{A}} \mathcal{D}_a(\theta) \mathbb{E}[\ell(\theta; Z) \mid A = a] \right\}, \qquad (5.1)$$

where $\mathcal{D}_a(\theta)$ remains unknown. In a plug-in approach, the learner estimates $\theta^\star \in \Theta$ by plugging-in an estimate $\widehat{\mathcal{D}}_a(\theta)$ as in Algorithm 1, and then minimizes the performative prediction risk. Toward this goal of minimizing $\mathrm{PR}(\theta)$ what $\{\theta_m; m \in [M]\} \subset \Theta$ should be deployed? In this section, we explore this question: an optimal design choice of $\{\theta_m; m \in [M]\}$ that minimizes the total (random) regret

$$\mathrm{Reg}(M) \triangleq \sum_{m \in [M]} \mathrm{PR}(\theta_m) - \mathrm{PR}(\theta^\star), \qquad (5.2)$$

where the randomness comes from θ_m. To this end, we propose a design algorithm similar to Algorithm 2 that relies on a doubling trick and alternates between exploration and exploitation. We divide drawing M models into $K = \log_2(1 + M/\tau_0)$ episodes denoted as $\{\mathcal{J}_k; k \in [K]\}$, where the lengths of the episodes are $|\mathcal{J}_k| = \tau_0 2^{k-1}$ and τ_0 is the number of models drawn in the first episode (initial phase) and does not affect the rate of convergence. We then divide each episode $\mathcal{J}_k$ into two sub-phases: the *exploration* phase $(\mathcal{I}_k)$, where we use Algorithm 2 to estimate $\mathcal{D}_a(\theta)$ and then *exploitation* phase $(\mathcal{I}'_k)$, where we constantly deploy

$$\widehat{\theta}^{(k)} \in \arg\min_{\theta \in \Theta}\Big\{\widehat{\mathrm{PR}}^{(k)}(\theta) \triangleq \sum_{a \in \mathcal{A}} \widehat{\mathcal{D}}_a^{(k)}(\theta)\mathbb{E}[\ell(\theta; Z)|A = a]\Big\}.$$
$$(5.3)$$

We summarize our procedure in Algorithm 3. Next

Algorithm 3 Sequential Design with doubling trick

1: Let $\mathcal{J}_1, \ldots, \mathcal{J}_K$ sequential episodes with $\mathcal{J}_1 = \{1, \ldots, \tau_o\}$, and $\mathcal{J}_k = \mathcal{I}_k \cup \mathcal{I}'_k$ for $k = 2, \ldots, K$, where $\mathcal{I}_k = \{\tau_o(2^{k-1} - 1) + 1, \ldots, a_k\}$ is the exploration index phase, where $a_k = (\tau_o 2^{k-1})^\alpha$ for some $\alpha \in (0, 1)$, and $\mathcal{I}'_k = \{a_k + 1, \ldots, \tau_o(2^k - 1)\}$ is the exploitation index phase.
2: For $k = 1$ let $\widehat{d}^{(1)}$ be the uniform density in $[0, 1]$.
3: **for** $k = 1, \ldots K$ **do**
4:     For every $m \in \mathcal{I}_k$ sample $b_m \overset{\text{i.i.d.}}{\sim} \widehat{d}^{(k)}$ and record $\widehat{\pi}(\theta_m)$ (see Remark 3.1).
5:     Based on $\{\widehat{\pi}(\theta_m)\}_{m \in \mathcal{I}_k}$ update $\{\widehat{\mathcal{D}}_a^{(k)}(\cdot); a \in \mathcal{A}\}$, $\widehat{\sigma}^{(k)}$ and the design density $\widehat{d}^{(k+1)} \propto \widehat{\sigma}^{(k)}$.
6:     For every $m \in \mathcal{I}'_k$ deploy the same $\widehat{\theta}^{(k)} = \arg\max_\theta \widehat{\mathrm{PR}}^{(k)}(\theta)$, where $\widehat{\mathrm{PR}}^{(k)}$ is defined in (5.3).
7: **end for**

---

we evaluate the goodness of Algorithm 3 in terms of the total regret (5.2). Notice that there is a trade-off between the exploration and exploitation phases: a larger exploration phase produces a more accurate $\widehat{\mathrm{PR}}^{(k)}(\theta)$, but does not decrease overall regret. For this reason, in Algorithm 3 we balance between these two phases by letting the length of exploration phase $a_k = |\mathcal{I}_k| = |\mathcal{J}_k|^\alpha = (\tau_o 2^{k-1})^\alpha$ where $\alpha \in (0, 1)$ is chosen optimally by looking at the rate of convergence in regret. This choice of α, along with the rate of convergence for both regret and relative efficiencies, is provided in the following theorem.

Theorem 5.1. *Assume that* $\mathbb{E}[\ell(\theta, X)|A = a]$ *is uniformly bounded in* θ *and* a. *Consider the procedure of Algorithm 3. Suppose that there exist* $\eta, \delta \in (0, 1)$ *and* $K_0 > 0$ *such that* $\|\mathcal{D}(1) - \widehat{\mathcal{D}}^{(k)}(1)\|_{L^\infty(\Theta)} = \mathcal{O}_P((\log(|\mathcal{I}_k|)/|\mathcal{I}_k|)^\eta)$ *and* $\|\widehat{d}^{(k)} - $

$d^\star\|_{L^\infty(0,1)} = \mathcal{O}_P\big((\log(|\mathcal{I}_k|)/|\mathcal{I}_k|)^\delta\big)$ *for* $k \geq K_0$. *Then the total regret is minimized at* $\alpha = 1/1+\eta$ *for with the regret and relative efficiencies are*

$$\mathrm{Reg}(M) = \mathcal{O}_P\big(M^{1/1+\eta}\log^\eta(M)\big),$$
$$\mathrm{REL}(\widehat{d}^{(K)}) = \mathcal{O}_P\big(M^{-\delta/1+\eta}\log^\delta(M)\big) \text{ as } M \to \infty.$$

Remark 5.2. *As we mention in Section 4.1, when* $\sigma(\cdot)$ *is a continuous function of* F *(see e.g. Theorem 4.3, where* $\sigma^2 \propto F\bar{F}$), *then* $\delta \equiv \eta$.

Remark 5.3. *For one-dimensional isotonic regression we have* $\eta = 1/3$ *(see Theorem 3.3) that leads to regret* $\mathcal{O}_P(M^{3/4})$. *For other standard non-parametric estimates, such as kernel and spline, see e.g. (Zhao and Yao, 2012, Section 3),* $\eta = \beta/2\beta+1$ *where* β *is the smoothness parameter for* F, *leading to regret* $\mathcal{O}(M^{2\beta+1/3\beta+1})$.

Comparing our Theorem 5.1 with Jagadeesan et al. (2022, Theorem 3), the key difference lies in our use of a plug-in estimator for the $\mathcal{D}(\theta)$ shift versus their need for i.i.d. samples from $\mathcal{D}(\theta)$ in each optimization step of PR. Their regret bound is $\mathcal{O}(M^{d+1/d+2})$, where d is the zooming dimension of θ, which in the worst case is identical to $\dim(\theta)$. In contrast, our bound is independent of $\dim(\theta)$, because the distribution map itself depends on θ only through the unidimensional $b(\theta) = \mathbf{L}_0 \mathbf{B}(\theta)$. Consequently, we achieve $\mathrm{Reg}(M) = \widetilde{\mathcal{O}}_P(M^{2\beta+1/3\beta+1})$ (Theorem 5.3), which is of lower order than $\mathcal{O}(M^{d+1/d+2})$ whenever $\beta \geq 1$ and $d \geq 2$.

6 Conclusions

We propose a microfoundation framework with random cost to learn the distribution map in reverse causal performative prediction settings. We also provide a design choice of predictive models for the learner to efficiently estimate the distribution map and minimize the performative risk. The performative prediction settings are extremely relevant in ML research, as they are routinely realized in many strategic environments. We believe that our proposal of learning the distribution map will make the problem of performative risk minimization more accessible to practitioners.

Where does our reverse causal framework fit into learning the distribution map in broader performative prediction problems? Our focus is primarily on finite action spaces, but the reverse causal model with an arbitrary A offers significant potential in performative prediction. Specifically, any $\mathcal{D}(X, Y; \theta)$ can be fit into reverse causal model by setting $A = (X, Y)$, making $\mathcal{D}(X, Y \mid A; \theta)$ trivially independent of θ. A balance can be struck between this general A and finite action spaces by choosing an action variable A

that captures all relevant information about θ, ensuring $\mathcal{D}(X, Y \mid A; \theta)$ remains independent of θ. Extending our approach to arbitrary action spaces may involve using generative models, such as a generative model h with a latent variable U (with a known distribution), to express $\mathcal{D}_A(\theta)$ via $h(U, \theta)$. For instance, in the CL model 2.2, setting $U \sim \text{unif}(0, 1)$ and $h(U, \theta) = \mathbb{I}\{U \leq \pi(\theta)\}$ allows A to be modeled as $h(U, \theta)$, where $\mathcal{D}_{A=1}(\theta) = P(h(U, \theta) = 1) = \pi(\theta)$, making A and $h(U, \theta)$ identically distributed. Learning the generative map h could benefit from the techniques in Goodfellow et al. (2014); Arjovsky et al. (2017). While exploring this further could broaden learning of distribution shifts, it is outside this paper's scope.

7 Code Availability

The codes are available at `https://github.com/dbracale/LearningDistributionPP-Code`.

References

Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *International conference on machine learning*, pages 214–223. PMLR, 2017.

Sabyasachi Chatterjee, Adityanand Guntuboyina, and Bodhisattva Sen. On matrix estimation under monotonicity constraints. *Bernoulli*, 24(2):1072–1100, May 2018. ISSN 1350-7265. doi: 10.3150/16-BEJ865.

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

Stephen Coate and Glenn C. Loury. Will Affirmative-Action Policies Eliminate Negative Stereotypes? *The American Economic Review*, 83(5):1220–1240, 1993. ISSN 0002-8282.

Q. Cormier, M. Milani Fard, K. Canini, and M. R. Gupta. Launch and iterate: Reducing prediction churn. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16, pages 3179–3187, Red Hook, NY, USA, December 2016. Curran Associates Inc. ISBN 978-1-5108-3881-9.

Joshua Cutler, Dmitriy Drusvyatskiy, and Zaid Harchaoui. Stochastic optimization under time drift: iterate averaging, step decay, and high probability guarantees. *arXiv preprint arXiv:2108.07356*, 2021.

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*, EC '18, pages 55–70, New York, NY, USA, June 2018. Association for Computing Machinery. ISBN 978-1-4503-5829-3. doi: 10.1145/3219166.3219193.

Hanming Fang and Andrea Moro. Theories of statistical discrimination and affirmative action: A survey. *Handbook of social economics*, 1:133–200, 2011.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

Qiyang Han, Tengyao Wang, Sabyasachi Chatterjee, and Richard J. Samworth. Isotonic regression in general dimensions. *The Annals of Statistics*, 47(5): 2440–2471, October 2019. ISSN 0090-5364, 2168-8966. doi: 10.1214/18-AOS1753.

Moritz Hardt, Nimrod Megiddo, Christos Papadimitriou, and Mary Wootters. Strategic Classification. *arXiv:1506.06980 [cs]*, November 2015.

Moritz Hardt, Nimrod Megiddo, Christos Papadimitriou, and Mary Wootters. Strategic Classification. In *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science*, ITCS '16, pages 111–122, New York, NY, USA, January 2016. Association for Computing Machinery. ISBN 978-1-4503-4057-1. doi: 10.1145/2840728.2840730.

Guy Horowitz and Nir Rosenfeld. Causal strategic classification: A tale of two shifts. In *International Conference on Machine Learning*, pages 13233–13253. PMLR, 2023.

Zachary Izzo, Lexing Ying, and James Zou. How to Learn when Data Reacts to Your Model: Performative Gradient Descent. In *Proceedings of the 38th International Conference on Machine Learning*, pages 4641–4650. PMLR, July 2021.

Meena Jagadeesan, Tijana Zrnic, and Celestine Mendler-Dünner. Regret Minimization with Performative Feedback. In *Proceedings of the 39th International Conference on Machine Learning*, pages 9760–9785. PMLR, June 2022.

Sooyong Jang, Sangdon Park, Insup Lee, and Osbert Bastani. Sequential covariate shift detection using classifier two-sample tests. In *International Conference on Machine Learning*, pages 9845–9880. PMLR, 2022.

Sagi Levanon and Nir Rosenfeld. Strategic Classification Made Practical, June 2021.

Sagi Levanon and Nir Rosenfeld. Generalized strategic classification and the case of aligned incentives. In *International Conference on Machine Learning*, pages 12593–12618. PMLR, 2022.

Licong Lin and Tijana Zrnic. Plug-in performative optimization. *arXiv preprint arXiv:2305.18728*, 2023.

Zachary C. Lipton, Yu-Xiang Wang, and Alex Smola. Detecting and Correcting for Label Shift with Black Box Predictors. *arXiv:1802.03916 [cs, stat]*, July 2018.

Subha Maity, Mikhail Yurochkin, Moulinath Banerjee, and Yuekai Sun. Understanding new tasks through the lens of training data via exponential tilting. In *The Eleventh International Conference on Learning Representations*, September 2022.

Mary C Meyer. Inference using shape-restricted regression splines. 2008.

John P. Miller, Juan C. Perdomo, and Tijana Zrnic. Outside the Echo Chamber: Optimizing the Performative Risk. In *Proceedings of the 38th International Conference on Machine Learning*, pages 7710–7720. PMLR, July 2021.

Alexandre Mösching and Lutz Dümbgen. Monotone least squares and isotonic quantiles. 2020.

Hans-Georg Müller. Optimal designs for nonparametric kernel regression. *Statistics & Probability Letters*, 2 (5):285–290, 1984.

Juan C. Perdomo, Tijana Zrnic, Celestine Mendler-Dünner, and Moritz Hardt. Performative Prediction. *arXiv:2002.06673 [cs, stat]*, June 2020.

Tim Robertson, Richard Dykstra, and FT Wright. Order restricted statistical inference. *(No Title)*, 1988.

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

Seamus Somerstep, Yuekai Sun, and Yaacov Ritov. Learning in reverse causal strategic environments with ramifications on two sided markets. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=vEfmVS5ywF.

Marilyn Strathern. 'improving ratings': audit in the british university system. *European review*, 5(3): 305–321, 1997.

Fan Yang and Rina Foygel Barber. Contraction and uniform convergence of isotonic regression. 2019.

Zhibiao Zhao and Weixin Yao. Sequential design for nonparametric inference. *Canadian Journal of Statistics*, 40(2):362–377, 2012.

Checklist

1. For all models and algorithms presented, check if you include:

   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes]

   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Yes]

   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Yes]

2. For any theoretical claim, check if you include:

   (a) Statements of the full set of assumptions of all theoretical results. [Yes]

   (b) Complete proofs of all theoretical results. [Yes]

   (c) Clear explanations of any assumptions. [Yes]

3. For all figures and tables that present empirical results, check if you include:

   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Yes]

   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Yes]

   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Yes]

   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Not Applicable]

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:

   (a) Citations of the creator If your work uses existing assets. [Not Applicable]

   (b) The license information of the assets, if applicable. [Not Applicable]

   (c) New assets either in the supplemental material or as a URL, if applicable. [Not Applicable]

   (d) Information about consent from data providers/curators. [Not Applicable]

   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]

5. If you used crowdsourcing or conducted research with human subjects, check if you include:

   (a) The full text of instructions given to participants and screenshots. [Not Applicable]

   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]

   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

Learning the Distribution Map in Reverse Causal Performative Prediction: Supplementary Materials

A Parametric model the distribution of the cost in the CL model

Suppose that F_C is the cdf of a Gaussian $\mathcal{N}(\mu, \sigma)$ that is $g(b) = \Phi(\frac{b-\mu}{\sigma})$, where Φ is the cdf of a $\mathcal{N}(0,1)$. We have $F_C(b(\theta)) = \Phi(\frac{b-\mu}{\sigma})$. If we observe models $\{\theta_m\}_{m \in [M]}$ and corresponding proportions $\{\widehat{\pi}_m\}_{m \in [M]}$ and benefits $\{b_m\}_{m \in [M]}$, to estimate F_C we can solve the linear model

$$\phi_m \equiv \Phi^{-1}(\widehat{\pi}_m) = \frac{b_m}{\sigma} - \frac{\mu}{\sigma} + \epsilon_m$$

with the constraint that $\sigma > 0$. The solution is

$$\widehat{\sigma}_M = \begin{cases} \frac{\sum_{m=1}^{M} \Delta b_m \Delta \phi_m}{\sum_{m=1}^{M} \Delta b_m^2} & \text{when } \sum_{m=1}^{M} \Delta b_m \Delta \phi_m > 0 \\ 0 & \text{otherwise} \end{cases}$$

and $\widehat{\mu}_M = \bar{b} - \widehat{\sigma}_M \bar{\phi}$, where $\Delta b_m = b_m - \bar{b}$, $\Delta \phi_m = \phi_m - \bar{\phi}$, and $\bar{b}$ and $\bar{\phi}$ are the average of the b_m and ϕ_m, respectively. Similarly, we can choose a logistic model by replacing the standard Gaussian cdf with $\Phi(b) = 1/(1 + e^{-b})$, the standard logistic cdf. The advantages of parametric models are simplicity of interpretation, computational efficiency, and convergence rate $\mathcal{O}_P(\sqrt{M})$.
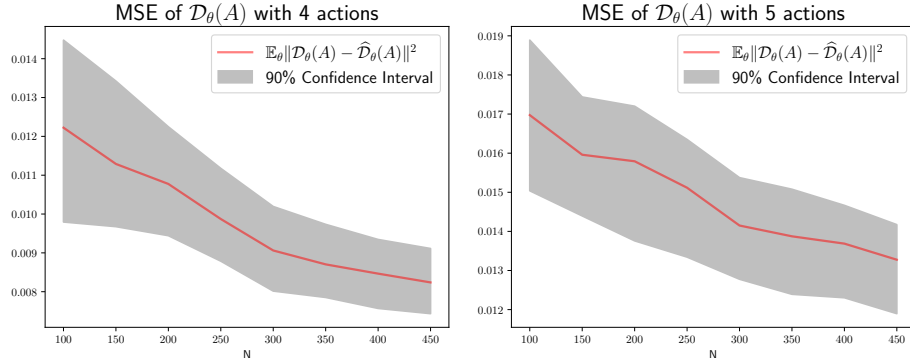
B Simulation: multivariate cost estimation

Figure 4: Multivariate distribution map estimation.

In this section, we provide a simulation setup for a general number of actions $|\mathcal{A}|$. We set $\mathbf{C} \sim \mathcal{N}_{|\mathcal{A}|}(\boldsymbol{\mu}, \Sigma)$, with $\boldsymbol{\mu} = \mathbf{0}$ and $\Sigma = \mathbf{I}$. We select random benefits $\{\mathbf{B}(\theta_m), m \in [M]\}$ with $M = 500$. For every $a \in \mathcal{A}$, we define $\widehat{\pi}_a(\theta_m) = \pi_a(\theta_m) + \epsilon_a(\theta_m)$, where $\pi_a(\theta_m) = \mathrm{F}_a(L_a \mathbf{B}(\theta_m))$, where F_a is the cdf of $\mathbf{L}_a \mathbf{C} \sim \mathcal{N}_{|\mathcal{A}|-1}(\mathbf{L}_a \boldsymbol{\mu}, \mathbf{L}_a \Sigma \mathbf{L}_a^\top)$ and i.i.d. $\epsilon_m(a) \sim \mathcal{N}(0, \tau)$, with $\tau = 0.1$. We redefined $\widehat{\pi}_a(\theta_m) \leftarrow \max\{0, \min\{1, \widehat{\pi}_a(\theta_m)\}\}$ since they represents probabilities. For $a \in \mathcal{A} \setminus \{0\}$, we get $\widehat{F}_a$ as multivariate isotonic regression with data $\{(\widehat{\pi}_a(\theta_m), \mathbf{L}_a \mathbf{B}(\theta_m)), m \in [M]\}$, and $\widehat{\mathrm{F}}_0 = 1 - \sum_{a \in \mathcal{A} \setminus \{0\}} \widehat{\mathrm{F}}_a$. Then we have $\widehat{\mathcal{D}}_A(\theta_m) = \{\widehat{\mathrm{F}}_a(\mathbf{L}_a \mathbf{B}(\theta_m)); a \in \mathcal{A}\}$ and $\mathcal{D}_A(\theta_m) = \{\mathrm{F}_a(\mathbf{L}_a \mathbf{B}(\theta_m)); a \in \mathcal{A}\}$ and we plot the quantity $\sum_{m \in [N]} \|\mathcal{D}_A(\theta_m) - \widehat{\mathcal{D}}_A(\theta_m)\|_2^2$, for $N \in \{50, 100, 150, 200, \dots, 500\}$. We repeat this process for 10 times, with the same $\mathbf{C}$ and $\{\mathbf{B}(\theta_m), m \in [M]\}$ and with different seed to simulate $\{\pi_a(\theta_m), a \in \mathcal{A}, m \in [M]\}$ and we show the result in Figure 4.

C   Regret Analysis: comparison with zero order methods

We compare our result stated in Theorem 5.1, with the result provided by Jagadeesan et al. (2022). They define

$$\text{Reg}_{mean}(M) = \sum_{m=1}^{M} \left[\mathbb{E}[\text{PR}(\theta_m)] - \min_\theta \text{PR}(\theta) \right]$$

where the expectation is taken over the possible randomness in the choice of $\{\theta_m\}_{m\in[M]}$, a sequence of deployed models. The main difference between the two algorithms is that, while we estimate PR by plugging in an estimate of the distribution shift, they sample from the distribution shift to estimate PR. In Theorem 4.2 they prove that

$$\text{Reg}_{mean}(M) = \mathcal{O}(M^{\frac{d+1}{d+2}} \epsilon^{\frac{d}{d+2}} + \sqrt{M})$$

where d is the zooming dimensions (which is equal to the dimension of θ in the worst-case scenario) and ϵ is the sensitivity of the distribution map, which is a value proportional to the amount of distribution shift. They assume that:

  i) the loss ℓ is bounded in $[0,1]$ (it is the same as our Assumption in Theorem 5.1).

  ii) ℓ is Lipschitz in the two variables (which we don't require);

When $\epsilon = 0$, meaning that the performative effect vanishes, $\text{Reg}_{mean}(M)$ achieves the standard bound $\mathcal{O}(\sqrt{M})$ in an essentially dimension-independent manner. In our work, we assume that the distribution shift is non-trivial (*i.e.* $\epsilon > 0$). Indeed, when $\epsilon \approx 0$, the agents don't move, which means that the cost c is always null, and then, there is no need to estimate it. For this reason, our regret bound can be compared to Jagadeesan et al. (2022) when $\epsilon > 0$ and is independent of M, which leads to $\text{Reg}_{mean}(M) = \mathcal{O}(M^{\frac{d+1}{d+2}})$.. However, our result is independent of the dimension of θ, because the distribution map only depends on θ throw the unidimensional variable $b(\theta) = \mathbf{L}_0 \mathbf{B}(\theta)$. As mentioned in Remark 5.3, we can achieve $\text{Reg}(M) = \widetilde{\mathcal{O}}_P(M^{2\beta+1/3\beta+1})$, which upper-bound is lower than $\mathcal{O}(M^{d+1/d+2})$ (excluding $\log(M)$ factors) whenever $\beta \geq 1/(d-1)$, *i.e.* $\beta \geq 1$ and $d \geq 2$.

D   Proofs

Proof of Lemma 2.5. The necessary direction is trivially realized once $\mathcal{D}_{A=a}(\theta)$ is expressed as

$$\mathcal{D}_{A=a}(\theta) = \mathbb{P}\big\{ \mathrm{B}_a(\theta) - \mathrm{B}_{a'}(\theta) \geq C_a - C_{a'}, \text{ for all } a' \neq a \big\}. \tag{D.1}$$

because this necessarily implies that $\mathcal{D}_{A=a}(\theta)$ increases with $\mathrm{B}_a(\theta) - \mathrm{B}_{a'}(\theta)$ for any $a' \neq a$. Establishing the sufficient direction is more non-trivial, which we establish using an induction argument. We start with the case of $|\mathcal{A}| = 2$, where, without loss of generality, we let $\mathcal{A} = \{0, 1\}$. Furthermore, we set $C_0 \equiv 0$. In this case, since $\mathcal{D}_{A=1}(\theta)$ increases with $\mathrm{B}(1, \theta) - \mathrm{B}(0, \theta)$ there exists a unique random variable C_1 such that

$$\mathcal{D}_{A=1}(\theta) = \mathbb{P}\big\{ C_1 \leq \mathrm{B}(1, \theta) - \mathrm{B}(0, \theta) \big\}.$$

For such $\{C_0, C_1\}$ we have the desired microfoundation for $\mathcal{A} = \{0, 1\}$.

For a more general $|\mathcal{A}| = K + 1 > 2$ let $\mathcal{A} = \{0, 1, \ldots, K\}$. To establish our induction step we assume that there exists a joint distribution for the $\mathbf{C}_{-K} = \{C_a; a = 0, \ldots, K - 1\}$ such that the microfoundation holds on $\mathcal{A} - \{K\}$, *i.e.* for any $a < K$ we have

$$\mathcal{D}_{A=a|A\neq K}(\theta) = \frac{\mathcal{D}_{A=a}(\theta)}{1 - \mathcal{D}_{A=K}(\theta)} = \mathbb{P}\big\{ \mathrm{B}_a(\theta) - \mathrm{B}_{a'}(\theta) \geq C_a - C_{a'}, \text{ for all } a' < K, a' \neq a \big\}. \tag{D.2}$$

Now, we show that there exists a random variable C_K such that

$$\mathcal{D}_{A=K}(\theta) = \mathbb{P}\big\{ \mathrm{B}_K(\theta) - \mathrm{B}_{a'}(\theta) \geq C_K - C_{a'}, \text{ for all } a' < K \big\}. \tag{D.3}$$

For the process, we further simplify the above expression in the right hand side as:

$$\begin{aligned}
&\mathbb{P}\big\{ \mathrm{B}_K(\theta) - \mathrm{B}_{a'}(\theta) \geq C_K - C_{a'}, \text{ for all } a' < K \big\} \\
&= \mathbb{P}\big\{ C_K \leq C_{a'} + \mathrm{B}_K(\theta) - \mathrm{B}_{a'}(\theta), \text{ for all } a' < K \big\} \\
&= \mathbb{P}\big\{ C_K \leq \min_{a'<K} \left(C_{a'} + \mathrm{B}_K(\theta) - \mathrm{B}_{a'}(\theta) \right) \big\} \\
&= \int \mathbb{P}\big\{ C_K \leq \min_{a'<K} \left(C_{a'} + \mathrm{B}_K(\theta) - \mathrm{B}_{a'}(\theta) \right) \mid \mathbf{C}_{-K} \big\} d\mathbb{P}(\mathbf{C}_{-K})
\end{aligned} \tag{D.4}$$

At this point, for any fixed value for $\mathbf{C}_{-K}$ we can find a distribution for $C_K \mid \mathbf{C}_{-K}$ such that

$$\mathcal{D}_{A=K}(\theta) = \mathbb{P}\big\{C_K \leq \min_{a' < K} \big(C_{a'} + \mathrm{B}_K(\theta) - \mathrm{B}_{a'}(\theta)\big) \mid \mathbf{C}_{-K}\big\}$$

This is possible because $\mathcal{D}_{A=K}(\theta)$ increases with each $\mathrm{B}_K(\theta) - \mathrm{B}_{a'}(\theta)$ and thus with $\min_{a' < K} \big(C_{a'} + \mathrm{B}_K(\theta) - \mathrm{B}_{a'}(\theta)\big)$. Denote the c.d.f. of the conditional distribution of $C_K \mid \mathbf{C}_{-K}$ by $\mathrm{F}_k(\cdot \mid \mathbf{C}_{-K})$. We combine this conditional distribution with the joint distribution of $\mathbf{C}_{-K}$ to obtain our desired joint distribution of $\mathbf{C}$.

To complete the proof, we need to show that such a choice of $\mathbf{C}$ satisfies (D.1). We have already established it for $a = K$, *i.e.* (D.3) because according to (D.4)

$$\begin{aligned}
&\mathbb{P}\big\{\, \mathrm{B}_K(\theta) - \mathrm{B}_{a'}(\theta) \geq C_K - C_{a'}, \text{ for all } a' < K \big\} \\
&= \int \mathbb{P}\big\{C_K \leq \min_{a' < K} \big(C_{a'} + \mathrm{B}_K(\theta) - \mathrm{B}_{a'}(\theta)\big) \mid \mathbf{C}_{-K}\big\} d\mathbb{P}(\mathbf{C}_{-K}) \\
&= \int \mathcal{D}_{A=K}(\theta) d\mathbb{P}(\mathbf{C}_{-K}) = \mathcal{D}_{A=K}(\theta).
\end{aligned}$$

To show this for $a \neq K$ we follow the identities:

$$\begin{aligned}
&\mathbb{P}\big[\, \mathrm{B}_a(\theta) - \mathrm{B}_{a'}(\theta) \geq C_a - C_{a'}, \text{ for all } a' \neq a \big] \\
&= \mathbb{P}\big[\big\{\, \mathrm{B}_a(\theta) - \mathrm{B}_{a'}(\theta) \geq C_a - C_{a'}, \text{ for all } a' \neq a, a' < K \big\} \\
&\qquad \cap \big\{\, \mathrm{B}_a(\theta) - \mathrm{B}_K(\theta) \geq C_a - C_K \big\}\big] \\
&= \mathbb{P}\big[\big\{\, \mathrm{B}_a(\theta) - \mathrm{B}_{a'}(\theta) \geq C_a - C_{a'}, \text{ for all } a' \neq a, a' < K \big\} \\
&\qquad \cap \big\{\, \mathrm{B}_{a'}(\theta) - \mathrm{B}_K(\theta) > C_{a'} - C_K \text{ for some } a' < K \big\}\big] \quad \text{(since dist. of } C_a \text{ are continuous)} \\
&= \int d\mathbb{P}(\mathbf{C}_{-K}) \mathbb{I}\big\{\, \mathrm{B}_a(\theta) - \mathrm{B}_{a'}(\theta) \geq C_a - C_{a'}, \text{ for all } a' \neq a, a' < K \big\} \\
&\qquad \times \mathbb{P}\big\{\, \mathrm{B}_{a'}(\theta) - \mathrm{B}_K(\theta) > C_{a'} - C_K, \text{ for some } a' < K \mid \mathbf{C}_{-K}\big\} \\
&= \int d\mathbb{P}(\mathbf{C}_{-K}) \mathbb{I}\big\{\, \mathrm{B}_a(\theta) - \mathrm{B}_{a'}(\theta) \geq C_a - C_{a'}, \text{ for all } a' \neq a, a' < K \big\} \mathcal{D}_{A \neq K}(\theta) \quad \text{(by (D.3))} \\
&= \mathcal{D}_{A \neq K}(\theta) \int d\mathbb{P}(\mathbf{C}_{-K}) \mathbb{I}\big\{\, \mathrm{B}_a(\theta) - \mathrm{B}_{a'}(\theta) \geq C_a - C_{a'}, \text{ for all } a' \neq a, a' < K \big\} \\
&= \mathcal{D}_{A \neq K}(\theta) \mathcal{D}_{A=a \mid A \neq K}(\theta) \quad \text{(by (D.2))} \\
&= \mathcal{D}_{A=a}(\theta).
\end{aligned}$$

This completes the proof of existance.

Uniqueness: If $\{\mathcal{D}_A(\theta) : \theta \in \Theta\} = \Delta^{\mathcal{A}}$ the uniqueness of the random variable is easily realized by noticing that the collection $\{(\mathcal{D}_A(\theta), \{\mathrm{B}(a, \theta) - \mathrm{B}(0, \theta)\}_{a \neq 0}) : \theta \in \Theta\}$ specifies all probability-quantile pairs. $\qquad\square$

Proof of Proposition 4.2

Proof. Recall the following definitions: $L_k = \tau_o 2^{k-1}$ is the length of the k-th episode for some integer τ_o that defined the length of the first episode; $K = K(M) = \log_2(1 + M/\tau_o)$ is the total number of episodes.

The assumption $\|\widehat{d}^{(k)} - d^\star\|_{L^\infty(0,1)} = \mathcal{O}_P\big(\log(L_k)^\delta L_k^{-\delta}\big)$ for $k \geq K_0$, implies in particular for $k = K$ that for every $\varepsilon > 0$ there exists $C_1 > 0$ and $M_0 > 0$ such that

$$\|\widehat{d}^{(K)} - d^\star\|_{L^\infty(0,1)} \leq C_1 \log(L_K)^\delta L_K^{-\delta} = C_1 \log^\delta(2^K) 2^{-K\delta} = C_1 K^\delta 2^{-K\delta} = C_1 \log_2^\delta(M) M^{-\delta}, \qquad (\text{D.5})$$

with probability at least $1 - \varepsilon$, for all $M \geq M_0$. From now on we consider all the statements in our proof happening in this high-probability event. We will also use the notation $a_n \asymp (\pm 1)\epsilon_n + a^\star$ to denote $|a_n - a^\star| \lesssim \epsilon_n$. For example, by assumption we have

$$\widehat{d}^{(K)} \asymp (\pm 1)\xi + d^\star, \quad \text{uniformly in } (0,1), \text{ where } \quad \xi \triangleq \log_2^\delta(M) M^{-\delta}.$$

Step 1.

From the proof of Theorem 2-3 of Zhao and Yao (2012) we have that $\mathrm{REL}(\widehat{d}_K) = \frac{4}{3}\rho_M + o(\rho_M^2)$ as $M \to \infty$

$$\rho_M \triangleq (U_M - c^2)/U_M$$

with

$$U_M \triangleq \int \frac{\sigma^2(b)}{\widehat{d}^{(K)}(b)} \mathrm{d}b \quad \text{and} \quad c \triangleq \int \sigma(b)\mathrm{d}b,$$

so that by Equation (4.1) $d^\star = \sigma/c$. Without loss of generality, we assume that $b \in (0, 1)$.

Step 2.

From $\widehat{d}^{(K)} \asymp (\pm 1)\xi + d^\star$ (uniformly in $b \in (0, 1)$) we have

$$U_M = \int \frac{\sigma^2(b)}{\widehat{d}^{(K)}} \mathrm{d}b \asymp \int \frac{\sigma^2(b)}{(\pm 1)\xi + d^\star(b)} \mathrm{d}b$$

Then

$$\frac{U_M - c^2}{c} \asymp \int \frac{\sigma^2(b)}{c((\pm 1)\xi + d^\star(b))} \mathrm{d}b - \int \sigma(b)\mathrm{d}b = \int \frac{\sigma^2(b) - c\sigma(b)((\pm 1)\xi + d^\star(b))}{c((\pm 1)\xi + d^\star(b))} \mathrm{d}b$$

Using that $d^\star(b) = \sigma(b)/c$ we have

$$\frac{U_M - c^2}{c} \asymp -(\pm 1) \int \frac{\xi \sigma(b)}{\xi + d^\star(b)} \mathrm{d}b$$

By the Taylor expansion $\frac{\xi \sigma}{\xi + d^\star} = \frac{\sigma}{d^\star}\xi + o(\xi^2)$ as $\xi \to 0$, since $\xi \to 0$ as $M \to \infty$, we have

$$\frac{U_M - c^2}{c} \asymp -(\pm 1)\xi \int \frac{\sigma(b)}{d^\star(b)} \mathrm{d}b + o(\xi^2) = -(\pm 1)\xi c + o(\xi^2), \quad M \to \infty$$

Then using that $U_M = \int \frac{\sigma^2(b)}{\xi + d^\star(b)} \mathrm{d}b \to \int \frac{\sigma^2(b)}{d^\star(b)} \mathrm{d}b = c \int \sigma(b)\mathrm{d}b = c^2$ as $M \to \infty$ we have

$$\rho_M = \frac{U_M - c^2}{U_M} = \frac{U_M - c^2}{c} \frac{c}{U_M} \asymp -(\pm 1)\xi + o(\xi^2)$$

then as $M \to \infty$, $\rho_M \asymp (\pm 1)\xi = (\pm 1)\log_2^\delta(M)M^{-\delta}$ that is $\rho_M = \mathcal{O}_P\left(\log_2^\delta(M)M^{-\delta}\right)$. □

Proof of Theorem 5.1

Proof. From definition (5.2)

$$\mathrm{Reg}(M) = \sum_{k=1}^{K} \sum_{m \in \mathcal{J}_k} \mathrm{PR}(\theta_m^{(k)}) - \mathrm{PR}(\theta^\star)$$

$$= \sum_{k=1}^{K} \left[\sum_{m \in \mathcal{I}_k} \mathrm{PR}(\theta_m^{(k)}) - \mathrm{PR}(\theta^\star) + \sum_{m \in \mathcal{I}_k'} \mathrm{PR}(\theta_m^{(k)}) - \mathrm{PR}(\theta^\star) \right] \tag{D.6}$$

For $m \in \mathcal{I}_k$, by the assumption of boundedness of the loss function, we have that $\mathrm{PR}(\theta_m^{(k)}) - \mathrm{PR}(\theta^\star)$ is bounded, then

$$\sum_{m \in \mathcal{I}_k} \mathrm{PR}(\theta_m^{(k)}) - \mathrm{PR}(\theta^\star) \lesssim |\mathcal{I}_k| \tag{D.7}$$

For $m \in \mathcal{I}_k'$, where $\mathcal{I}_k'$ is the time-index of the exploitation phase, we split (for simplicity of notation we remove the dependence on m in $\theta_m^{(k)}$)

$$0 \leq \mathrm{PR}(\theta^{(k)}) - \mathrm{PR}(\theta^{\star}) = \mathrm{PR}(\theta^{(k)}) - \widehat{\mathrm{PR}}^{(k)}(\theta^{(k)}) + \widehat{\mathrm{PR}}^{(k)}(\theta^{(k)}) - \mathrm{PR}(\theta^{\star})$$

$$\leq \underbrace{|\mathrm{PR}(\theta^{(k)}) - \widehat{\mathrm{PR}}^{(k)}(\theta^{(k)})|}_{=:\mathcal{V}_1} + \underbrace{|\mathrm{PR}(\theta^{\star}) - \widehat{\mathrm{PR}}^{(k)}(\theta^{(k)})|}_{=:\mathcal{V}_2} \qquad (\mathrm{D.8})$$

Bound $\mathcal{V}_1$.

$$\mathcal{V}_1 = \left| \mathrm{PR}(\theta^{(k)}) - \widehat{\mathrm{PR}}^{(k)}(\theta^{(k)}) \right|$$

$$= \sum_{a \in \mathcal{A}} \left| \mathbb{E}_Z[\ell(\theta^{(k)}; Z)|A = a]\mathcal{D}_a(\theta^{(k)}) - \mathbb{E}_Z[\ell(\theta^{(k)}; Z)|A = a]\widehat{\mathcal{D}}_a^{(k)}(\theta^{(k)}) \right|$$

$$= \sum_{a \in \mathcal{A}} \mathbb{E}_Z[\ell(\theta^{(k)}; Z)|A = a] \left| \mathcal{D}_a(\theta^{(k)}) - \widehat{\mathcal{D}}_a^{(k)}(\theta^{(k)}) \right|$$

$$\lesssim \sum_{a \in \mathcal{A}} \left| \mathcal{D}_a(\theta^{(k)}) - \widehat{\mathcal{D}}_a^{(k)}(\theta^{(k)}) \right|$$

$$= \left| \mathcal{D}_1(\theta^{(k)}) - \widehat{\mathcal{D}}_1^{(k)}(\theta^{(k)}) \right| + \left| 1 - \mathcal{D}_1(\theta^{(k)}) - (1 - \widehat{\mathcal{D}}_1^{(k)}(\theta^{(k)})) \right|$$

$$\lesssim \left| \mathcal{D}_1(\theta^{(k)}) - \widehat{\mathcal{D}}_1^{(k)}(\theta^{(k)}) \right|$$

$$\lesssim \|\mathcal{D}_1 - \widehat{\mathcal{D}}_1^{(k)}\|_{L^\infty(\Theta)} \qquad (\mathrm{D.9})$$

Bound $\mathcal{V}_2$.

From definition (5.1) and (5.3) we have

$$\theta^{\star} \in \arg\min_\theta \left\{ \mathrm{PR}(\theta) = \sum_{a \in \mathcal{A}} \mathbb{E}_Z[\ell(\theta; Z)|A = a]\mathcal{D}_a(\theta) \right\}$$

and

$$\theta^{(k)} \in \arg\min_\theta \left\{ \widehat{\mathrm{PR}}^{(k)}(\theta) = \sum_{a \in \mathcal{A}} \mathbb{E}_Z[\ell(\theta; Z)|A = a]\widehat{\mathcal{D}}_a^{(k)}(\theta) \right\}.$$

Then

$$\mathrm{PR}(\theta^{\star}) - \widehat{\mathrm{PR}}^{(k)}(\theta^{(k)}) = \mathrm{PR}(\theta^{\star}) - \mathrm{PR}(\theta^{(k)}) + \mathrm{PR}(\theta^{(k)}) - \widehat{\mathrm{PR}}^{(k)}(\theta^{(k)}) \leq \mathrm{PR}(\theta^{(k)}) - \widehat{\mathrm{PR}}^{(k)}(\theta^{(k)}),$$

where we used that $\mathrm{PR}(\theta^{\star}) \leq \mathrm{PR}(\theta^{(k)})$. Then

$$\mathrm{PR}(\theta^{\star}) - \widehat{\mathrm{PR}}^{(k)}(\theta^{(k)}) \leq \mathrm{PR}(\theta^{(k)}) - \widehat{\mathrm{PR}}^{(k)}(\theta^{(k)})$$

$$\leq \left| \mathrm{PR}(\theta^{(k)}) - \widehat{\mathrm{PR}}^{(k)}(\theta^{(k)}) \right|$$

$$\leq \sum_{a \in \mathcal{A}} \left| \mathbb{E}_Z[\ell(\theta^{(k)}; Z)|A = a]\mathcal{D}_a(\theta^{(k)}) - \mathbb{E}_Z[\ell(\theta^{(k)}; Z)|A = a]\widehat{\mathcal{D}}_a^{(k)}(\theta^{(k)}) \right|$$

$$= \sum_{a \in \mathcal{A}} \mathbb{E}_Z[\ell(\theta^{(k)}; Z)|A = a] \left| \mathcal{D}_a(\theta^{(k)}) - \widehat{\mathcal{D}}_a^{(k)}(\theta^{(k)}) \right|$$

$$\lesssim \sum_{a \in \mathcal{A}} \left| \mathcal{D}_a(\theta^{(k)}) - \widehat{\mathcal{D}}_a^{(k)}(\theta^{(k)}) \right|$$

$$= \left| \mathcal{D}_1(\theta^{(k)}) - \widehat{\mathcal{D}}_1^{(k)}(\theta^{(k)}) \right| + \left| 1 - \mathcal{D}_1(\theta^{(k)}) - (1 - \widehat{\mathcal{D}}_1^{(k)}(\theta^{(k)})) \right|$$

$$\lesssim \left| \mathcal{D}_1(\theta^{(k)}) - \widehat{\mathcal{D}}_1^{(k)}(\theta^{(k)}) \right|$$

$$\lesssim \|\mathcal{D}_1 - \widehat{\mathcal{D}}_1^{(k)}\|_{L^\infty(\Theta)} \qquad (\mathrm{D.10})$$

Similarly,

$$-[\mathrm{PR}(\theta^\star) - \widehat{\mathrm{PR}}^{(k)}(\theta^{(k)})] = -\mathrm{PR}(\theta^\star) + \widehat{\mathrm{PR}}^{(k)}(\theta^\star) - \widehat{\mathrm{PR}}^{(k)}(\theta^\star) + \widehat{\mathrm{PR}}^{(k)}(\theta^{(k)}) \le \widehat{\mathrm{PR}}^{(k)}(\theta^\star) - \mathrm{PR}(\theta^\star),$$

where we used that $\widehat{\mathrm{PR}}^{(k)}(\theta^{(k)}) \le \widehat{\mathrm{PR}}^{(k)}(\theta^\star)$. Then

$$
\begin{aligned}
-[\mathrm{PR}(\theta^\star) - \widehat{\mathrm{PR}}^{(k)}(\theta^{(k)})] &\le \mathrm{PR}(\theta^\star) - \widehat{\mathrm{PR}}^{(k)}(\theta^\star) \\
&\le \left| \mathrm{PR}(\theta^\star) - \widehat{\mathrm{PR}}^{(k)}(\theta^\star) \right| \\
&\le \sum_{a \in \mathcal{A}} \left| \mathbb{E}_Z[\ell(\theta^\star; Z)|A = a]\mathcal{D}_a(\theta^\star) - \mathbb{E}_Z[\ell(\theta^\star; Z)|A = a]\widehat{\mathcal{D}}_a^{(k)}(\theta^\star) \right| \\
&= \sum_{a \in \mathcal{A}} \mathbb{E}_Z[\ell(\theta^\star; Z)|A = a]\left| \mathcal{D}_a(\theta^\star) - \widehat{\mathcal{D}}_a^{(k)}(\theta^\star) \right| \\
&\lesssim \sum_{a \in \mathcal{A}} \left| \mathcal{D}_a(\theta^\star) - \widehat{\mathcal{D}}_a^{(k)}(\theta^\star) \right| \\
&= \left| \mathcal{D}_1(\theta^\star) - \widehat{\mathcal{D}}_1^{(k)}(\theta^\star) \right| + \left| 1 - \mathcal{D}_1(\theta^\star) - (1 - \widehat{\mathcal{D}}_1^{(k)}(\theta^\star)) \right| \\
&\lesssim \left| \mathcal{D}_1(\theta^\star) - \widehat{\mathcal{D}}_1^{(k)}(\theta^\star) \right| \\
&\lesssim \|\mathcal{D}_1 - \widehat{\mathcal{D}}_1^{(k)}\|_{L^\infty(\Theta)}
\end{aligned}
\tag{D.11}
$$

Thus, by (D.10) and (D.11), we have $\mathcal{V}_2 = |\mathrm{PR}(\theta^\star) - \widehat{\mathrm{PR}}^{(k)}(\theta^{(k)})| \lesssim \|\mathcal{D}_1 - \widehat{\mathcal{D}}_1^{(k)}\|_{L^\infty(\Theta)} = \mathcal{O}_P\left([\log(|\mathcal{I}_k|)/|\mathcal{I}_k|]^\eta\right)$.

Putting together bound $\mathcal{V}_1$ and $\mathcal{V}_2$.

Form (D.8) and the two bounds we have that, for $m \in \mathcal{I}'_k$ (reintroducing the dependence on m in $\theta_m^{(k)}$)

$$\mathrm{PR}(\theta_m^{(k)}) - \mathrm{PR}(\theta^\star) = \mathcal{O}_P\left([\log(|\mathcal{I}_k|)/|\mathcal{I}_k|]^\eta\right).$$

This means that for every $\varepsilon > 0$ there exists a $C_{1,m}^{(k)} > 0$ and a $K_{0,m}^{(k)} > 0$ such that

$$\mathbb{P}\left(\mathrm{PR}(\theta_m^{(k)}) - \mathrm{PR}(\theta^\star) \le C_{1,m}^{(k)}[\log(|\mathcal{I}_k|)/|\mathcal{I}_k|]^\eta\right) \ge 1 - \frac{\varepsilon}{|\mathcal{J}_k|K}, \quad \text{for all } k \ge K_{0,m}^{(k)}.$$

Now, in the intersection of these $|\mathcal{J}_k|$ events (which intersection has probability at least $1 - \frac{\varepsilon}{K}$), using (D.7) we have

$$
\begin{aligned}
\sum_{m \in \mathcal{J}_k} \mathrm{PR}(\theta_m^{(k)}) - \mathrm{PR}(\theta^\star) &= \sum_{m \in \mathcal{I}_k} \mathrm{PR}(\theta_m^{(k)}) - \mathrm{PR}(\theta^\star) + \sum_{m \in \mathcal{I}'_k} \mathrm{PR}(\theta_m^{(k)}) - \mathrm{PR}(\theta^\star) \\
&\lesssim |\mathcal{I}_k| + |\mathcal{I}'_k||\mathcal{I}_k|^{-\eta} \log^\eta(|\mathcal{I}_k|) \\
&\lesssim 2^{\alpha k} + 2^k 2^{-\alpha k \eta} \log^\eta(2^{\alpha k}) \\
&\lesssim k^\eta(2^{\alpha k} + 2^k 2^{-\alpha k \eta}).
\end{aligned}
$$

The optimal α is achieved when the two exponents are equal that is $\alpha = 1 - \alpha\eta$, that implies $\alpha = 1/(1+\eta)$. With this choice we get that for every $\varepsilon > 0$ there exists a $C_1^{(k)} > 0$ and a $K_0^{(k)} > 0$ such that

$$\mathbb{P}\left(\sum_{m \in \mathcal{J}_k} \mathrm{PR}(\theta_m^{(k)}) - \mathrm{PR}(\theta^\star) \le C_1^{(k)} k^\eta 2^{\alpha k}\right) \ge 1 - \frac{\varepsilon}{K}, \quad \text{for all } k \ge K_0^{(k)}$$

Now, consider the intersection of these events over the K episodes (which intersection has a probability of at least $1 - \varepsilon$). In this intersection, we have that, from (D.6)

$$\mathrm{Reg}(M) \lesssim \sum_{k=1}^K k^\eta 2^{\alpha k} \lesssim K^\eta \sum_{k=1}^K 2^{\alpha k} \lesssim K^\eta 2^{\alpha K} \lesssim \log^\eta(M) M^\alpha = \log^\eta(M) M^{1/(1+\eta)},$$

where we used that $K = \log_2(1 + M/\tau_o)$. This proves that

$$\text{Reg}(M) = \mathcal{O}_P(\log^\eta(M)M^{1/(1+\eta)}).$$

To prove the convergence of the relative efficiency, we apply Theorem 4.2. $\qquad\square$