MACHINE LEARNING FOR STRATEGIC INFERENCE IN A SIMPLE DYNAMIC GAME

IN-KOO CHO AND JONATHAN LIBGOBER

ABSTRACT. We consider a simple buyer-seller game, with a buyer whose strategy is determined via access to data and some statistical algorithm. Our model builds off Rubinstein (1993), who showed, for this environment, that the seller can exploit the limited ability of simple classifiers to implement the ex-post optimal decision rule. Taking either the set of baseline classifiers as given or dropping the assumption that the seller is profit maximizing, we argue that no statistical algorithm is capable of approximating the rational benchmark. However, allowing for algorithms to "combine" classifiers and using the seller's incentive to maximize expected profit, we show the existence of an algorithm which induces (approximately) rational behavior from the buyer. Our construction uses *boosting*, a common technique from machine learning. This algorithm shows that it is unnecessary for the buyer to be able to *fit* sophisticated classifiers, provided they can combine rudimentary classifiers in a particular way.

1. INTRODUCTION

Consumers often make purchasing decisions based on recommendations of a platform, which are in turn based on aggregated data. In these situations, what determines a seller's optimal strategy is not the resulting inference of a rational buyer, but how this strategy interacts with a statistical algorithm. This raises the question of whether statistical algorithms can induce behavior that is *as-if* rational, and if so, how.

A rational decisionmaker in an economic model updates beliefs in response to strategic choices in order to determine his or her optimal decision. When the decisionmaker's rule is the outcome of a statistical algorithm, one instead hopes to use past data in order to inform future choices. This process involves several distinct steps:

- Processing the data,
- Fitting a model, and
- Making predictions with the resulting fit.

In this paper, our interest is in cases where the algorithm is limited in the kinds of models that can be fit to data. On the other hand, it is important to distinguish the *fitting* of a model to the data, versus the *construction* of a model. Even if only a limited number of models can be fit to data, this still leaves substantial flexibility regarding how to construct classifiers. For instance, in principle one could seek to fit one classifier on part of the data

---

Date: March 20, 2020

Preliminary and incomplete: comments welcome.

This project started when the first author was visiting the University of Southern California. We are grateful for hopitality and support from USC. Financial support from the National Science Foundation is gratefully acknowledged.

and another classifier on a separate part. In this case, the algorithm could specify using each classifier, depending on whether a new observation were closer to the first part or the second part. Or, one could process observations in a particular way, and then fit a classifier on the modified dataset. As we will show in this paper, there may indeed be scope for an algorithm to be designed which can outperforms the best that can be done simply by fitting a model alone.

To study whether and how algorithms can approximate rationality, this paper builds off of the model of Rubinstein (1993) to study the question of whether an algorithm can mimic the predictions a rational actor would make in a strategic setting. Rubinstein (1993) showed that if a *rational* decisionmaker is *restricted* to use a binary threshold classifier—i.e., one that makes the same decision on a given side of a fixed threshold—then the seller can price discriminate by utilizing a particular form of randomization which "fools" these buyers into making a decision which is suboptimal, given the realized price. A similar result obtains (with an appropriate translation given the different context) if we assume that the buyer uses a statistical algorithm.

The intuition behind this result is simple—first, the optimally chosen classifier chosen can do strictly better than simply randomizing the guess, implying that the seller can exploit the incentives of the buyer in order to manipulate the decision rule. On the other hand, it is impossible for the thresholds to take the optimal decision with probability 1 when this decision is non-monotone in the price. The first point implies the buyer trades off against errors, and the second point implies that the tradeoff falls short of the fully rational response. As a result, the seller can force a different decision than would be rationally optimal for these buyers (with arbitrarily high probability).

Our reason for using Rubinstein (1993), then, is to focus attention on an environment where it is already known that there is an incentive to exploit the form of bounded rationality of the buyer. That is, while we no longer assume the buyer is rational, we drop the restriction that all they can do is use the best-fitting single threshold classifier. There are a few features of our environment that are worth emphasizing. First, one desiderata is to perform well against a sufficiently broad class of environments with access only to past data. In particular, we seek an algorithm that is independent of other parameters of the problem, for instance because there is sufficient uncertainty over them. This is in line with our motivation of having the buyer's behavior being *as driven by the data as possible.* As a result, we seek a guarantee of approximate optimality that does well across a wide variety of environments. This implies that it is difficult to know a priori which types of environments the algorithm should be concerned with. This adds a layer of non-triviality to our exercise; certain naive strategies which seek to "force" behavior upon the seller may do well in certain environments, but not well in other environments. On the other hand, our algorithm has the desireable feature that it is "parameter-free", drawing attention to the underlying model classes and the methods used to combine them, as opposed to anticipating particular behavior which the buyer may seek to optimize against.

Second, our exercise features the following tension: On the one hand, a rich set of classifiers would be necessary to have any hope of giving accurate predictions with a sufficiently high degree of confidence. On the other hand, considering classifiers with this richness would make our exercise hopeless, preventing any good guarantee for achieving a performance that is close to optimal. The non-trivial aspect of our exercise is in determining how to design the algorithm in a way that controls the complexity while allowing simultaneously to do well given any possible realized seller strategy.

Our contribution is to illustrate the following: Rather than start with a rich set of baseline classifiers, we start with a minimal set of classifiers, which are precisely those considered in Rubinstein (1993). Instead, we construct *new classifiers* online, by combining binary threshold classifiers with specified weights. This allows the complexity of the resulting classifier to respond endogenously to the data generating process. More precisely, our approach is to use the Adaptive Boosting algorithm (Schapire and Freund (2012)), which specifies exactly how to construct the weights to minimize error. The algorithm requires us to be able to (repeatedly) fit a classifier to some distribution over prices and outcomes, from some set of baseline classifers. Each classifier is weighted according to its performance on past data, and the prediction made following any price is the one which achieves the highest score.

Returning to the particular case at hand, single threshold classifiers turn out to be the smallest set of classifiers which would provide any hope of achieving a rational response, even if we augment the ability to fit classifiers with the ability to combine them. Rather, we see that the issue with single threshold classifiers is that they are not *strong learners* (i.e., they cannot ensure the optimal decision is taken with probability 1 following any price), even though they are *weak learners* (i.e., they can outperform random guesses when chosen optimally). The remarkable property of the Adaptive Boosting algorithm is that it shows that the requirement of weak learnability is actually equivalent to strong learnability. In other words, the first part of the intuition for the main result in Rubinstein (1993), outlined above, exactly tells us how to overcome the issues with the second part, once we have the algorithm in hand.

Our contribution is to show how this algorithm can be applied to the setting studied in Rubinstein (1993), providing a counterpoint to the observation that the bounded capability of statistical classifiers make them exploitable. Our main result is to exhibit a version of AdaBoost which ensures that the seller (when rational) uses a strategy that leads the buyer to behave close to rational with high probability—put differently, the rational benchmark is PAC learnable.[1] Putting this together, our message is that while it is not possible to guarantee that rationality emerges for arbitrarily seller actions, it *is* possible if the data generating process is endogenous to the statistical algorithm. This argument requires some additional steps using incentives of the seller to demonstrates that the resulting output does in fact correspond to what is traditionally thought of as subgame perfection. While our algorithm is off-the-shelf to a certain extent, some additional steps are needed in order to demonstrate convergence to the rational benchmark. This should not be surprising, since the endogeneity issue makes the problem no longer a pure statistical exercise. These modifications extend beyond the initial need to show that it is possible to do better than random guessing in this environment. As our analysis elucidates, AdaBoost is capable of handling a *particular* kind of unboundedness in the cardinality of the action space. It is thus necessary to discipline the environment further in order to achieve our results.

[1] "PAC" refers to Probably Approximately Correct; see Section 4.6.1.

While this result may seem simple once the algorithm is presented, before doing so we outline a conceptual issue which make the exercise non-standard. Without using the rationality of the seller, the learning problem is hopeless—there are just too many possible seller strategies to worry about, and a rational decision maker would need to adapt the decision one to each one individually. As mentioned, our algorithm takes some additional steps in describing how we discipline the seller's incentives sufficiently in order to maintain the good performance of the algorithm. In other words, there is no guarantee that our algorithm does well in the absence of seller incentives, and one should not expect such results to be maintained. This is largely what distinguishes our exercise from a standard statistical exercise–the incentives of the seller matter.

Therefore, the contribution of this paper is two-fold: First, we seek to further scrutinize the sense in which algorithms reflect a boundedly rational decisionmaker, as the details of this assertion turn out to matter substantially. Second, with regards to the machine learning literature, we show how by treating the endogeneity of the data generating process seriously, we can ensure that algorithms perform well even when the environment would be otherwise too complicated. This suggests the intriguing possibility for improvements to algorithms arising due to incentive considerations. Going forward, we hope our work illuminates the importance of taking seriously how algorithms interact with the data generating process. We believe these issues are of increasing importance, as the interaction between algorithms and strategic choices will only become more ubiquitous in the future.

We first review the relevant literature, and proceed to describe the baseline model, essentially reviewing Rubinstein (1993). We then describe the "supergame" during which the algorithm is determined in the follow section. In Section 5, we describe a few benchmarks which help motivate and appreciate our problem. Our proposed algorithm can be found in Section 4.6, which subsequent sections being devoted to results demonstrating the appealing features of this algorithm. Proofs are in the Appendix.

2. Literature

This paper is most closely related to the literature on learning in games when players' behavior depends on a statistical method. The single-agent problem is a particular special case. Single agent versions of this problem are the focus of Al-Najjar (2009) and Al-Najjar and Pai (2014). As the buyer essentially faces a single-agent problem given the seller's strategy, these results are particularly relevant to the analysis of Section 5.3, where we contrast our results with theirs. However, it is worth emphasizing that the data buyers receive is *endogenous* in our setting because of the strategic interactions. In contrast, their benchmarks correspond to the case of exogenous data. This problem is also studied in Spiegler (2016), who focuses on causality and defines a solution concept for behavior that arises from individuals fitting a directed acyclic graph to past observations.

Taking these approaches to games, the literature has still for the most part focused on settings where the interactions between players is *static*, typically imposing finiteness to a degree that rules out the game of Rubinstein (1993). In contrast, our setting is a simple, two-player (and two-move) sequential game. Cherry and Salant (2019) discuss a procedure whereby players' behavior arises from a statistical rule estimated by sampling past actions. This leads to an endogeneity issue similar to the one present in our environment, i.e., an interaction between the data generating process and the statistical method used to evaluate it. Eliaz and Spiegler (2018) study the problem of a statistician estimating a model in order to help an agent take an action, motivated (like us) by issues involved with the interaction between rational plays and statistical algorithms. Liang (2018), like us, is focused on games of incomplete information, asking when a class of learning rules leads to rationalizable behavior. Focusing on the application of model selection in econometrics, Olea, Ortoleva, Pai, and Prat (2019) study an auction model and ask which statistical models achieve the highest confidence in results as a function of a particular dataset.

On the other side, the literature on learning in extensive form games has typically assumed that agents experiment optimally, and hence embeds notion of rationality on the part of agents which we dispense with in this paper. Classic contributions include Fudenberg and Kreps (1995), Fudenberg and Levine (1993) and Fudenberg and Levine (2006). Most of this literature has focused on cases where there is no exogenous uncertainty regarding a player's type, and asking whether self-confirming behavior emerges as the outcome. An important exception is Fudenberg and He (2018), who study the steady-state outcomes from experimentation in a signalling game. While a rational agent in our game would need to form an expectation over an exogenous random variable, signalling issues do not arise because our seller has commitment.

Less related—although similar in spirit—is a small but growing literature on the use of machine learning algorithms by *sellers*, particularly in competitive environments. Calvano, Calzolari, Denicolò, and Pastorello (2019), Brown and MacKay (2019), and Hansen, Misra, and Pai (2020) study the question of the use of algorithms by sellers can provide a channel through which collusive behavior can be sustained. While we are focused on a very different question—namely, whether an algorithm determining the *buyer*'s strategy could yield a *rational* reply—we are similarly interested in studying the implications of constraints related to taking the strategy space to be an algorithm (and not necessarily chosen by a rational player). Our interest realtes more to implementable behavior, as opposed to positing a particular game. We anticipate a growing interest in studying how computational considerations interact with other strategic variables.

Our companion paper Cho and Libgober (2020) studies a more general version of the model presented in this paper, but discussed the same algorithm as capable of delivering an approximate rational response. That paper requires a generalization of the algorithm in this paper in order to allow for richer possible actions, in which case the set of baseline classifiers may be even more severely limited relative to the rational benchmark. On the other hand, we are able to obtain some additional results by focusing on this particular environment. For instance, we are able to argue directly that the seller has no incentive to slow down the algorithm, and are able to explicitly use our convergence rates to calculate payoff bounds. So whereas our other paper seeks to speak to a wider range of environments, this paper explores in greater depth how the algorithm performs in a particular context.

3. Baseline Model

The baseline model builds off Rubinstein (1993), which introduces a buyer-seller game where buyers make purchasing decisions using binary classifiers. In that paper, each buyer's classifier is chosen optimally, although potentially from a set that is unable to replicate a sequentially optimal response. This section reviews his model, with following sections nesting this within a *machine game* to be described.

3.1. **Strategies and Payoffs.** The seller sells a product of quality $\theta \in \{L, H\}$, yielding a buyer willingness-to-pay of v_θ with $v_H > v_L$. π_θ is the (ex-ante) probability quality is equal to θ. Before observing θ, the seller commits to a strategy

$$\sigma : \{L, H\} \to \Delta(P),$$

where $P \subset \mathbb{R}_+$ is a set of admissible prices. For technical reasons, we assume that the support of σ can have at most countably many prices.

Denote by Σ the set of possible strategies the seller can commit to, and let $\Delta(P)$ be the set of all probability distributions over P. Throughout this paper, we assume that

$$P = [v_L, v_H].$$

The restriction will turn out to be without loss, as we will see that the only valid response of a buyer to a price of $p < v_L$ would be to purchase, and the only valid response of a buyer to a price of $p > v_H$ would be to not purchase.

A buyer infers the underlying state through the offered price. The optimal decision of a buyer is to buy at p if

$$\mathbf{E}(v|p) - p = \big(\mathbf{P}(L|p)v_L + \mathbf{P}(H|p)v_H\big) - p > 0.$$

3.1.1. *The Lemons Condition.* Rubinstein (1993) focused on the case where the seller has an incentive to separate the buyers which depends on θ. If the state is L, then the production cost is $c_L = 0$. If the state is H, then the production cost depends upon the type of buyers. There are N_i buyers of type $i \in \{1, 2\}$ in every period, and $N = N_1 + N_2$. If the good is delivered to type i buyer under state H, it costs c_i to the seller. We assume that

$$c_1 > v_H > c_2 > v_L > c_L = 0 \tag{3.1}$$

and

$$v_H < \frac{N_1}{N}c_1 + \frac{N_2}{N}c_2 = \mathbf{E}[c] \tag{3.2}$$

so that the seller cannot make a positive profit in state H by selling the good to every buyer. To generate positive profit, the seller has to screen out type 1 buyer in state H. Parameters that satisfy these conditions will be said to satisfy the *lemons condition.*

In state L, $v_L N$ is the largest profit the monopolist can generate. In state H, the monopolist can make the largest profit by selling only to type 2 buyers at the highest possible price, say v_H. Thus, the upper bound of the expected profit the monopolist can ever generate is

$$\Pi^* = \pi_L v_L N + \pi_H (v_H - c_2) N_2.$$

It is optimal to accept any $p < v_L$. Therefore, s lower bound of the profit of the monopolist is

$$\Pi_* = \pi_L v_L N$$

by offering $p = v_L$ if $\theta = L$.

3.2. **Rational Buyers.** A rational buyer takes σ as given, computes $\mathbf{E}[v \mid \sigma, p] - p$, and purchases if this is strictly greater than 0 and does not purchase if this is strictly less than 0. In our context, the rational strategy of the buyer assumes optimal behavior, given the seller's pricing rule σ along with the system of beliefs conditioned on $\forall p \in P$.

For analytic convenience, let us label each price in the support of σ according to the behavior of a rational buyer. Define

$$y(\sigma, p) = \begin{cases} 1 & \text{if} \quad \big(\mathbf{P}(L|p)v_L + \mathbf{P}(H|p)v_H\big) - p > 0 \\ 1 \text{ or } -1 & \text{if} \quad \big(\mathbf{P}(L|p)v_L + \mathbf{P}(H|p)v_H\big) - p = 0 \\ -1 & \text{if} \quad \big(\mathbf{P}(L|p)v_L + \mathbf{P}(H|p)v_H\big) - p < 0. \end{cases} \tag{3.3}$$

as the function that represents the optimal behavior of a rational buyer, as we interpret $y(\sigma, p) = 1$ as "buy at p" and $y(\sigma, p) = -1$ as "do not buy at p." We call y the rational label.

Under the lemons condition, the seller cannot make more than Π_*. Because the seller has no instrument to screen type 2 buyer from type 1 buyer, the optimal strategy of the seller is to trade only under state L.

Proposition 3.1 (Rubinstein (1993))**.** *The unique equilibrium payoff of the monopolistic seller is $\Pi_* = \pi_L v_L N$. The equilibrium strategy of the seller is to charge v_L with probability 1 if the state is L, and v_H if the state is H: $\sigma(v_L|L) = 1$ and $\sigma(v_H|H) = 1$. Conditioned on v_L, all buyer accepts the offer. Conditioned on v_H or any $p \neq v_L$, no buyer purchases the good. If $p = v_L$ or v_H, the belief is computed by Bayes rule. If $v_L < p < v_H$, then the belief is concentrated at L.*

For the rest of the paper, we refer to this equilibrium as the equilibrium in the baseline model.

4. Defining Machine Games

Having outlined the basic interaction in the previous section, this section describes our formulation of the algorithm choice problem. In this setting, we assume (as in Rubinstein (1993)) that type 2 buyers are rational, but that the type 1 buyers have a strategy that is the outcome of the statistical algorithm. In contrast, the benchmark of Rubinstein (1993) will emerge when buyers are restricted to choosing best-fitting classifier from some fixed set of classifiers.

4.1. **Overview.** The basic problem of interest is one where the buyer arrives at a strategy using a statistical algorithm. Our assumption is that, for some set fixed set of classifiers $\mathcal{H}$ and a distribution over prices and decisions, and any assignment of labels $y(p) \in \{-1, 1\}$, it is possible to solve the following problem:

$$\max_{h \in \mathcal{H}} \sum_p h(p) y(p) f(p).$$

We interpret this as saying that there is some code which can find the *best fitting hypothesis* from some class $\mathcal{H}$.

We treat this step as a black box; however, our interest is in building an algorithm around this capacity. Nevertheless, we emphasize that our perspective is that the algorithm is *constrained* in the hypotheses that can be fit to data. In contrast, we imagine that an algorithm can augment the data, or specify price distributions to fit. The difficulty is in specifying these modifications, and in particular how to do this in such a way that yields desireable buyer behavior. Later, we elaborate on desireable properties of an algorithm, and discuss various reasons why one algorithm may be better than another. For now, we proceed with describing the basic interaction.

Our assumption will be that the type 1 buyer has only coarse information about the underlying game, and cannot condition the algorithm sensitively on underlying parameters. Let

$$\kappa = (c_1, c_2, v_H, v_L, c_L, \pi_H, N_1, N_2)$$

satisfying (3.2) be the parameter of the underlying model which defines the lemon's problem. Let $\mathcal{K}$ be a compact set of κ satisfying (3.2). Boundedness implies that type 1 buyer has some knowledge about the underlying game, but does not know precisely what the underlying game is because $\mathcal{K}$ typically contains a continuum of elements. $\mathcal{K}$ and a prior distribution over $\mathcal{K}$ are common knowledge among (rational) players.

4.2. **Single-Threshold Classifiers.** Our main case of interest is when the set $\mathcal{H}$ consists of single-threshold classifiers. Let

$$h : \mathbb{R} \to \{1, -1\}$$

be a singe threshold classifier parameterized by its threshold θ, which partition the real line into two parts

$$p \geq \theta \qquad \text{or} \qquad p < \theta$$

and assigns the same value for each p in each partition.[2] We interpret $h(p) = 1$ (resp. $h(p) = -1$) as the decision to buy (resp. not to buy) the product at price p. His decision rule is constrained to a threshold rule, choosing the same action whenever the price is on a given side of the threshold θ.

A sample is a pair (p, y) where σ is a probability distribution over P and $p \in P$ is the price charged by the monopolist with a positive probability, and $y \in \{1, -1\}$ which indicates the "real" optimal decision of a consumer. A sample is always associated with an underlying randomization rule σ. Whenever the meaning is clear from the context, we suppress σ.

Definition 4.1. *We say that h correctly classifies sample (p, y) if $h(p)y(\sigma, p) = 1$, and that $h(p)$ incorrectly classifies (p, y) if $h(p)y(\sigma, p) = -1$. We say the buyer emulates a rational buyer, if $h(p)y(\sigma, p) = 1$ $\forall p$ in the support of σ.*

4.3. **Algorithms.** In this section, we formally define the object that determines the buyer's strategy, namely the algorithm. Let

$$D_t = (s_1, \ldots, s_{t-1})$$

[2]The inequality can be replaced by the weak inequality and vice versa.

be the history at the beginning of period t. We maintain the assumption that all prices in the support of σ are realized in each period. Let $\mathcal{D} = \cup_{t\geq 1} D_t$ be the set of possible datasets a buyer could observe, which is the set of histories.

Definition 4.2. *A label is*

$$\gamma : \mathcal{D} \times P \to \{-1, 1\}.$$

Let Γ *be the set of all labels.*

$\gamma(D,p)$ is type 1 buyer's action following (D,p). If $\gamma(D,p) = 1$, the buyer buys at price p. If $\gamma(D,p) = -1$, then the buyer does not accept p.

If

$$\gamma(D,p) = y(\sigma, p) \qquad \forall D \in \mathcal{D} \tag{4.4}$$

holds for every $\sigma \in \Sigma$, then type 1 buyer behaves as if he is rational. If so, we say that type 1 buyer emulates a rational player.

Definition 4.3. *Let* $\tilde{\Gamma} \subset \Gamma$ *denote a subset of labels. A statistical procedure is a function*

$$\tau : \mathcal{D} \to \tilde{\Gamma}.$$

The specification of $\tilde{\Gamma}$ captures the specific properties of $\tau(D)$. Let

$$\tau(D)(p) \in \{1, -1\}$$

be the value of $\tau(D)$ assigned to p. Let $\mathcal{T}$ be the set of all feasible algorithms. Our main interest is in understanding which kinds of $\mathcal{T}$ allow for the buyer to approximate rational behavior.

4.4. **Timing of a Machine game.** To analyze the equilibrium in a model where the strategy of type 1 buyers is restricted to the single threshold rules, we examine the machine game (Rubinstein (1986)) where each player choose a strategy, and then delegates the decisions in the ensuing game to the selected strategy.

(1) In period -1, the type 1 buyer chooses an algorithm from some set of possible algorithms, i.e., $\tau \in \mathcal{T}$.
(2) In period 0, the parameters of the underlying game are realized, and a rational seller chooses a *distribution* over $\sigma(\theta)$, a probability distribution over prices conditioned on $\theta \in \{H, L\}$. That is, we take $\sigma \in \Sigma = \Delta(P) \times \Delta(P)$, and $\mu \in \Delta(\Sigma)$.
(3) In period $t \geq 1$, nature chooses state $\theta \in \{H, L\}$ with probability $\mathbf{P}(\theta) = \pi_\theta$. A strategy $\sigma \sim \mu$ is drawn and a price is realized according to θ and $\sigma(\cdot|\theta)$.
(4) Conditioned on a realized price, a buyer of type 1 or type 2 decides whether to purchase or not. Recall the type 2 buyer is rational, and therefore chooses a strategy which is a complete specification of strategic move under all possible contigencies.
(5) Payoff in period t is realized.

By considering different sets of possible algorithms (i.e., $\mathcal{T}$), one arrives at different machine games. Our interest is in understanding which properties of possible algorithms yields behavior that looks as-if rational. As a result, we will discuss various possibilities regarding the first part of this interaction.

The ensuing game is repeated infinitely many times. To simplify the analysis, we assume as in Rubinstein (1993) that a buyer observes all prices in the support of σ in each period. Let $\mathbf{P}(\sigma)$ be the set of prices in the support of σ. The outcome in period t is then

$$s_t = (p, h(p), y(\sigma, p))_{p\in\mathbf{P}(\sigma)}$$

and

$$D_{t+1} = (D_t, S_t) \qquad \forall t \geq 1.$$

Note that $h(p) = 1$ is to purchase the good at p, while $h(p) = -1$ is not to buy. We can regard

$$\frac{h(p)+1}{2}$$

as the probability of buying the good at p.

The payoff of type 1 buyer in period t following D_t in state $\theta \in \{H, L\}$ is

$$u_{b,1}(s_t, \theta) = \sum_p \left[(v_\theta - p)\frac{h(p)+1}{2}\right]\sigma(p|\theta). \tag{4.5}$$

Similarly, the payoff of type 2 (rational) buyer is

$$u_{b,1}(s_t, \theta) = \sum_p \left[(v_\theta - p)\frac{y(\sigma,p)+1}{2}\right]\sigma(p|\theta). \tag{4.6}$$

The payoff of the seller in period t following D_t is

$$u_s(s_t, L) = \sum_p \left[\left(\frac{h(p)+1}{2}N_1 + \frac{y(\sigma,p)+1}{2}N_2\right)p\right]\sigma(p|L) \tag{4.7}$$

$$u_s(s_t, H) = \sum_p \left[\frac{h(p)+1}{2}N_1(p-c_1) + \frac{y(\sigma,p)+1}{2}N_2(p-c_2)\right]\sigma(p|H). \tag{4.8}$$

Let $\delta \in (0, 1)$ be the discounting factor. In a machine game with discounting, the objective function of the seller is

$$\mathcal{U}_s(\sigma, \tau) = (1-\delta)\mathbf{E}\sum_{t=1}^{\infty}\delta^{t-1}u_s(s_t, \theta) \tag{4.9}$$

Similarly, the objective function of a type i buyer is

$$\mathcal{U}_{b,i} = (1-\delta)\mathbf{E}\sum_{t=1}^{\infty}\delta^{t-1}u_{b,i}(s_t, \theta). \tag{4.10}$$

We define Nash equilibrium for the machine game.

Definition 4.4. *$(\overline{\sigma}, \tau)$ is a Nash equilibrium if*

$$\mathcal{U}_s(\overline{\sigma}, \tau) \geq \mathcal{U}_s(\overline{\sigma}', \tau) \quad \textit{and} \quad \mathcal{U}_{b,1}(\overline{\sigma}, \tau) \geq \mathcal{U}_{b,1}(\overline{\sigma}, \tau')$$

while type 2 buyers decide according to rational label y, $\forall\overline{\sigma}', \forall\tau'$.

4.5. **Discussion of the model.** Several of our modelling choices are described below:

4.5.1. *Computational cost.* Our assumption is that the algorithm must be designed *before* observing the underlying parameters of the game. As illustrated in Lemma 5.1 below, the seller will typically faces an incentive to add prices to the support of σ in this game. And if the support of σ has many prices, finding an optimal threshold is a complex task. Because σ is endogenous, the optimization problem is even more complicated.

One might wonder why the algorithm is not reoptimized every time σ is chosen. This modelling choice can be justified by the introduction of small costs to writing an algorithm. If σ were fixed, then a type 1 buyer should be able to identify the best response even if they were required to pay a small computational cost. However, the equilibrium value of the buyer against σ is endogenous. And in the equilibrium of the game, we presume that a type 1 buyer calculates the best response for all possible σ. So unless we impose a restriction on the set of feasible pricing rules of the seller, the computational cost of calculating a best response for every σ overwhelms any potential gain from playing the game. At first glance, the prediction of a subgame perfect equilibrium does not appear to be robust against a small computational cost (Rubinstein (1986)).

We have in mind a situation where the type 1 buyer has to pay a small fixed cost for a computational *code*. This implies that it is prohibitively costly for them to seek to reoptimize against each individual seller. We then search for an algorithm which can calculate a best response on behalf of type 1 buyer. If such an algorithm exists, and if the algorithm is sufficiently simple, then a type 1 buyer can behave "as if" he is rational so that the best response of the monopolistic seller is the subgame perfect equilibrium strategy generating expected payoff Π_*. By writing a flexible algorithm, the type 1 buyer is able to respond to more strategies without incurring the additional costs.

4.5.2. *Coarse information.* The input of the algorithm are the data of the outcome, not the parameters of the game—namely, the price and the ex-post optimal decision given the price. The data from the outcome is coarse in the sense that the algorithm can use only the ordinal information of the outcome. For example, if the consumer surplus is positive, the algorithm can use the information that the surplus is positive (or negative), but not the information about the size of surplus. Relying on coarse information, the algorithm can operate over a broad class of games and its performance is not affected by the details of the games. Such robustness is particularly sought after, if the buyer has to design the algorithm based upon coarse information about the underlying game.

4.5.3. *Endogenously misspecified models.* An interesting question in our context is whether the algorithmic buyer in our model is correctly specified or not. A conventional learning algorithm aims to find the best fit model in a fixed class of models. The learning algorithm searches for a threshold rule in the class of single threshold decision rules which maximizes his expected return. If there were mostly type 2 buyers, then the equilibrium strategy of type 1 buyer in a baseline model is a single threshold rule and therefore, $\mathcal{H}$ is correctly specified in the sense of Esponda and Pouzo (2014). If the model is correctly specified, we obtain the convergence to the rational behavior under general conditions (e.g., Marcet and Sargent (1989)).

In our case, however, the seller strategically chooses her strategy to render the model of a type 1 buyer misspecified. Given the set of single threshold decision rules, the seller uses a strategy which requires two thresholds to correctly label the decision. Misspecification of type 1's model is endogenously generated by the strategic choice of the seller, which prevents type 1 buyers from learning to respond rationally to the strategy of the seller. Type 1 buyers need an algorithm that can identify the best fit model efficiently, while expanding the model class from the decision rules with a single threshold to those with multiple thresholds.

Despite a large literature on learning in economics, there has been little progress of the investigation on the evolution of model classes.[3] Exploiting the recent development in machine learning literature, we construct a learning algorithm over the model class, which allows type 1 buyer to respond rationally to a broad class of strategies of the seller. In the end, the seller finds it optimal to play the equilibrium strategy against the algorithm of a type 1 buyer, who behave as if he is rational.

4.6. **Desired properties of an algorithm.** Before proceeding with the analysis, we introduce a few other concepts from the machine learning literature which will guide are analysis.

4.6.1. *PAC.* We are interested in whether a statistical procedure can learn the strategies of the seller from the data. A fundamental criterion in the computer science literature to evaluate whether this can be done well is *probably approximately correct (PAC) learnability.*

Let s_t be an outcome in period t, which is observed by a decision maker. Suppose that

$$D_t = (s_1, \ldots, s_{t-1})$$

is a sequence of $t-1$ independently identically distributed (IID) samples. Let $\mathcal{D}_t$ be the set of all D_t. Let

$$\tilde{\Gamma}_t = \{h \mid D_t \in \mathcal{D}_t\}$$

be the set of all decision rules generated by τ. Recall that $y(\sigma, p) \in \{1, -1\}$ is the label that is the decision a rational player would make in response to (σ, p).

The class of buyer's strategies $\tilde{\Gamma}_t$ is PAC learnable if, given a sufficiently large number of IID samples, an algorithm produces a strategy with the property that the event "the wrong action is taken against an independently chosen sample with probability more than ε" occurs with probability no more than δ. Since $\tilde{\Gamma}_t$ is induced by statistical procedure τ, we call τ is PAC learnable if $\tilde{\Gamma}_t$ is PAC learnable.

Definition 4.5. *Fix $\tilde{\Sigma} \subset \Sigma$. Statistical procedure τ is PAC learnable of $\tilde{\Sigma}$, if $\forall \varepsilon, \delta > 0$, $\forall \sigma \in \tilde{\Sigma}$, $\exists T$ such that $\forall t \geq T$*

$$\mathbf{P}\left(\mathbf{P}\left(\tau(D_t)(p)y(\sigma, p) = -1\right) < \varepsilon \mid \sigma\right) > 1 - \delta.$$

$\tilde{\Gamma}_t$ is uniformly PAC learnable if T is selected uniformly over $\sigma \in \tilde{\Sigma}$.

We emphasize the order of quantifers in this definition; while δ and ε can be taken arbitrarily small, the data requirement will typically increase as they approach 0.

[3]Cho and Kasa (2015) considered a learning model with multiple but fixed model classes, which makes it difficult to examine a long run evolution of model classes.

4.6.2. *Ensemble algorithm.* Recall that $\mathcal{H}$ is the collection of all single threshold decision rules.

Definition 4.6. *Classifier H is an ensemble of $\mathcal{H}$ if $\exists h_1, \ldots, h_K \in \mathcal{H}$ and $\alpha_1, \ldots, \alpha_K \geq 0$ such that*

$$H(p) = \begin{cases} 1 & \text{if } \sum_{k=1}^{K} \alpha_k h_k(p) \geq 0 \\ -1 & \text{if } \sum_{k=1}^{K} \alpha_k h_k(p) < 0. \end{cases}$$

Note that $\alpha_k \geq 0$ is without loss of generality, since

$$\alpha_k h_k(p) = (-\alpha_k)(-h_k(p)).$$

Without loss of generality, we can assume that $\sum_{k=1}^{K} \alpha_k = 1$. We can interpret H as a weighted majority vote of $h_1, \ldots, h_K$, because the value of h_k is 1 or -1.

An ensemble algorithm constructs a classifier through a linear combination of single threshold classifiers. Since the final classifier is constructed through a basic arithmetic operation, one can easily construct an elaborate classifier from rudimentary classifiers. Ensemble algorithm has been remarkably successful in real world applications (Dietterich (2000)).

4.6.3. *Recursive algorithm.* Recall that $D_{t+1} = (D_t, S_t)$ is the concatenation of D_t and the latest observation S_t. Let X_t be any state variable that may not be observed.

Definition 4.7. *τ is recursive if $\exists \tilde{\tau}$ such that*

$$\tau(D_{t+1}) = \tilde{\tau}(\tau(D_t), S_t, X_t) \qquad \forall D_t, S_t, X_t.$$

A recursive algorithm summarizes the past history through $\tau(D_t)$, thus requiring a minimal amount of memory to implement in computers. It would be natural for an agent with limited computational capability to look for a tractable model which requires a small amount of memory to solve.

4.6.4. *Efficiency.* The data complexity has been a major interest in computer science (Shalev-Shwartz and Ben-David (2014)), as we want the algorithm to produce a good approximation with a reasonable size of data. We require that the algorithm should have the large deviation property, in the sense that the probability of mis-classification vanishes at an exponential rate, or equivalently, the required amount of data increases in logarithmic scale with respect to the probability of mis-classification.

Definition 4.8. *Statistical procedure τ satisfies the large deviation property if $\forall \epsilon > 0$, $\exists \rho > 0$ such that*

$$\limsup_{t \to \infty} \frac{1}{t} \log \mathbf{P} \left(\mathbf{P} \left(\tau(D_t)(p) y(\sigma, p) = -1 \right) > \epsilon \mid \sigma \right) < -\rho.$$

The large deviation property ensures that the standard deviation of the forecasting error vanishes at the linear rate as the number of observation increases, as the sample average of i.i.d. random variables. The finite sample property of an estimator violating the large deviation property can be extremely errative, or even misleading (Meyn (2007)).

This requirement is slightly weaker than the efficiency criterion (Shalev-Shwartz and Ben-David (2014)), which requires that the number of data needed to reduce the forecasting error below a certain bound increases at a logarithmic rate.

Definition 4.9. *Statistical procedure τ is efficient if $\forall \varepsilon, \delta > 0$, $\exists T$, $\exists \rho > 0$ such that $\forall t \geq T$*

$$\mathbf{P}_\sigma \left(\mathbf{P}_{\mathcal{D}_t} \left(\tau(D_t)(p) y(\sigma, p) = -1 \right) > \epsilon \right) < e^{-\rho t}.$$

The efficiency requires the existence of the rate function $\rho > 0$ dictates the convergence rate $\forall t \geq T$, while the large deviation property requires the rate holds only in the limit.

4.6.5. *Summary.* Putting these together, we assume that $\tau(D_t) = H_t$ should be

$$H_t(p) = \begin{cases} 1 & \text{if } \sum_{k=1}^t \alpha_k h_k(p) \geq 0 \\ -1 & \text{if } \sum_{k=1}^t \alpha_k h_k(p) < 0. \end{cases}$$

Following the observation of D_t, the selection of (α_t, h_t) is completely determined by the state in period t. To choose (α_t, h_t) $\forall t \geq 1$, we require that the algorithm relies on ordinal information of the outcome, while the nature of the ordinal information can be specific to the algorithm. We also require that τ is efficient, in the sense that the classification error of H_t vanishes at an exponential rate asymptotically.

Let $\tilde{\Gamma}$ be the set of classifiers generated by recursive ensemble algorithms, and $\mathcal{T}$ be the set of statistical procedures that satisfy our requirement.

5. Preliminary Benchmarks

In this section, we discuss a few simple instances of our model which highlight the aspects which make it non-trivial. Along the way, we justify our focus on single-threshold classifiers, describing the sense in which they are a minimal set which one might hope could achieve rationality. We also highlight the importance of considering the seller's incentives, as learnability guarantees fail rather dramatically without restrictions that the seller acts optimally.

5.1. **Single threshold classifiers.** Our first benchmark considers the case where the algorithm does not seek to do anything more sophisticated than utilizing the best fitting best single-threshold classifier. In this case, the seller can exploit the buyer under the lemons condition, which was pointed out by Rubinstein (1993). He showed that the monopolist can obtain profit arbitrarily close to the upper bound Π^*:

Lemma 5.1 (Rubinstein 1993)**.** *Suppose the type 1 buyer were restricted to choosing a single-threshold classifier, but instead chooses the optimal one in response to σ. Then $\forall \epsilon > 0$, the seller has a strategy which generates expected payoff larger than $\Pi^* - \epsilon$.*

Proof. For later reference, we sketch the proof. Since $c_1 > v_H > c_2$, the monopolistic seller needs to screen out type 1 buyers who use the single threshold decision rules, while selling to type 2 buyers. For a fixed $\epsilon > 0$, consider the following randomized pricing rule σ of the seller: in state H, she charges $v_H - \epsilon_H$ with probability 1, and in state L, she charges $v_L - \epsilon_L$ with probability ϵ and $(v_H + v_L)/2$ with probability ϵ.

We can choose $\epsilon, \epsilon_H, \epsilon_L > 0$ to satisfy

$$\pi_H \epsilon_H < \pi_L \epsilon_L$$

and

$$\frac{\epsilon_L}{\epsilon_L + 0.5(v_H - v_L)} < \epsilon < \frac{\pi_L \epsilon_L - \pi_H \epsilon_H}{\pi_L \epsilon_L}.$$

A type 2 buyer rationally accepts $v_L - \epsilon_L$ and $v_H - \epsilon_H$, and rejects $0.5(v_H + v_L)$, since

$$\mathbf{E}(v|v_L - \epsilon_L) - (v_L - \epsilon_L) = \epsilon_L, \quad \mathbf{E}(v|v_H - \epsilon_H) - (v_H - \epsilon_H) = \epsilon_H$$

but

$$\mathbf{E}(v|0.5(v_H + v_L)) = v_L < 0.5(v_H + v_L).$$

The seller randomizes over 3 prices. As a result, a type 1 buyer has 8 possible strategies, since they can specify a different decision after each price. Note that accepting $(v_H + v_c)/2$ in state L generates loss. We chose $\epsilon > 0$ so that by accepting $0.5(v_H + v_L)$, type 2 buyer suffers sufficiently large loss. A simple calculation shows that the only two candidates for the optimal threshold rule is to accept any price below v_L or accept any price above $v_H - \epsilon_H$. Since $\pi_H \epsilon_H < \pi_L \epsilon_L$, the optimal threshold rule is to accept a price below v_L.

In state L, both types of the buyer accept $v_L - \epsilon_L$ but in state H, only type 2 buyer accepts. The profit of the seller is

$$N \pi_L (v_L - \epsilon_L) + N_2 \pi_H (v_H - c_2) \simeq \Pi^*,$$

as claimed. □

Given this benchmark, we now turn to our setting, where the buyer instead must resort to an algorithm instead of calculating the optimal response. We show how this results translates over when thebuyer *is restricted to choosing a classifier that lies within the set of single-threshold classifiers*. This corresponds to a restriction on $\mathcal{T}$. Again, we do not require the code will lead to the optimal response immediately. Instead, we require that the algorithm to produce an approximately optimal response in the long run as the buyer interacts with σ.

Note that a single threshold decision rule can be parameterized by its threshold and the action assigned to each partition induced by the threshold. Since the optimal decision for $p \geq v_H$ is to reject the offer, and that for $p \leq v_L$ is to accept the offer, we can restrict to the class of single threshold decision rule parameterized by ϕ so that

$$h(p) = \begin{cases} 1 & \text{if } p \geq \phi \\ -1 & \text{if } p < \phi. \end{cases}$$

For $t \geq 1$, let ϕ_t be the estimated threshold. Consider a recursive learning algorithm

$$\phi_{t+1} = \psi(\phi_t, s_t) \tag{5.11}$$

where

$$s_t = (\sigma(\theta_t), h_t(\sigma(\theta_t)), y(\sigma, \sigma(\theta_t)))$$

is the outcome realized in period t, where h_t is the threshold rule defined by threshold ϕ_t. In a recursive algorithm, a new threshold ϕ_{t+1} is calculated by comparing the rational label $y(\sigma, p)$ against the actual decision rule in period t.

Suppose that the estimated threshold rule converges to an optimal threshold rule against σ.

$$\sup_h \mathcal{U}_{b,1}(\sigma, h, y) = \lim_{t \to \infty} \mathcal{U}_{b,1}(\sigma, h_t, y). \tag{5.12}$$

Let Ψ be the set of all recursive learning algorithms that satisfy (5.12)

The seller has a strategy which generates a long run average profit close to Π^*.

Proposition 5.2. *Suppose the type 2 buyer's algorithm is restricted to choosing outputs which are single-threshold classifiers. Then* $\forall \epsilon > 0$*, the seller has a strategy which generates expected payoff larger than* $\Pi^* - \epsilon$.

Proof. Suppose that the seller chooses the same randomized pricing rule σ as in Lemma 5.1. The best response of type 1 buyer against σ is to set the threshold between v_L and $(v_H + v_L)/2$ so that he accept $v_L - \epsilon_L$ while rejecting $(v_H + v_L)/2$ and $v_H - \epsilon_H$.

Against σ, any recursive learning algorithm generates $\{\phi_t\}$ which converges to $\phi \in \left(v_L - \epsilon_L, \frac{v_H + v_L}{2}\right)$ to emulate the best response of type 1 buyer against σ. Thus, the long run average payoff against such algorithm should be bounded from below by $\Pi^* - \epsilon$. □

To summarize, Proposition 5.2 shows that an algorithm that approximates optimal responses will need to expand the model classes used. The next few sections articulate why this is not a straightforward task.

5.2. **PAC-Learning with Observable θ.** Next, we justify our focus on single-threshold classifiers as those which *are* able to achieve rational responses in a very simple benchmark. Specificially, we first consider the case where θ is observable to the buyer, and hence there is no need to in update beliefs depending on price. In this case, the statistical exercise is straightforward. Consider the following very simple algorithm, which produces a single threshold classifier, given a set of observed prices $p_1, \ldots, p_N$:

- Set $\underline{p} = \max\{p_i \mid y(\sigma, p_i) = 1\}$ and $\overline{p} = \min\{p_i \mid y(\sigma, p_i) = 1\}$.
- Choose some threshold classifier $h(p)$ that set $h(p) = 1$ if $p \leq \underline{p}$ and $h(p) = -1$ if $p \geq \overline{p}$.

Notice that the only prices that could be mislabeled are in the interval $[\underline{p}, \overline{p}]$. But given any arbitrary distribution over prices, the probability that a subsequent price will be drawn in this interval approaches 0 as the dataset gets large.

Proposition 5.3. *Given any* σ*, the rational decision rule is PAC learnable via the above algorithm.*

This algorithm mentioned is an example of an Empirical Risk Minimization (ERM) algorithm, since it achieves perfect empirical error at every step. The fundamental theorem of statistical learning states that learnability by such an algorithm is equivalent to having finite VC dimension. In this case, the space of prices has VC dimension 1, and hence the above algorithm works.

The reason this fails in general is that when the type is unobserved, then the rational rule must condition on σ as well. We now turn to the difficulties with this step.

5.3. **Impossibility Results for Incomplete Information.** Now we show that when it is necessary to update beliefs, buyers cannot be guaranteed to do well relative to the rational benchmark. Formally, suppose that rather than a rational seller, there is an exogenous (full support) probability distribution μ over $\tilde{\Sigma} \subset \Sigma$ which determines the strategy of the seller. Otherwise, the model is unchanged, with type 1 buyers using a statistical procedure (as defined above) to arrive at their strategy.

This section is devoted to the following result:

Theorem 5.4. *Suppose $\tilde{\Sigma}$ contains any randomization over at least two prices. Then Γ is not (non-uniformly) learnable.*

The following result helps contrast when we will subsequently be interested in the case of a strategic seller:

Theorem 5.5. *No classifier that is independent of σ can produce y with probability $1-\varepsilon$, for every μ.*

The latter theorem is trivial, since it is possible to find distributions which have the same support and yet different optimal strategies with probability bounded away from 0. The former is essentially a counting exercise, using the notion of VC-dimension. This concept is also discussed in Al-Najjar (2009), Kalai (2003), Salant (2007), and Basu and Echenique (2019).

Definition 5.6. *A set of seller outcomes (that is, points in $\tilde{\Sigma}\times P$) is **shattered** by a class $\tilde{\Gamma}$ if, no matter how buyer decisions are assigned, these labels coincide with the prediction of some element of $\tilde{\Gamma}$. The **Vapnik-Chervonenkis Dimension** of a class $\tilde{\Gamma}$ is the largest number of points in $\tilde{\Sigma} \times P$ that can be shatterred by $\tilde{\Gamma}$.*

A series of well-known results in machine learning relate the VC dimension of a class to the learnability of decision rules. Whether a class is learnable depends on two aspects of the environment:

- How large is $\tilde{\Sigma} \times P$?
- How large is $\tilde{\Gamma}$?

The following proposition argues that size, in this case, reflects VC-dimension:

Proposition 5.7. *If $\tilde{\Gamma}$ has infinite VC dimension, then it is not uniformly PAC learnable. If $\tilde{\Gamma}$ cannot be written as the countable union of classes with finite VC dimension, then it is not non-uniformly PAC learnable.*

A (fully) rational buyer is a special case where $\tilde{\Gamma}$ is a singleton on the classifier predicting $y = 1$ whenever $\mathbf{E}[v \mid \sigma, p] > p$ and $y = 0$ when this inequality is flipped. This trivializes the learning problem. This is also true if the buyer is not fully rational but purchases according to some given rule (as in Rubinstein (1993)), or if the set of possible $\tilde{\Gamma}$ is itself finite. As our interest is primarily in cases where the buyer's problem is non-trivial, and hence wishes to use their model to *extrapolate* to observations they have not seen, these restrictions are unsatisfying.

Proposition 5.8. *Suppose $\tilde{\Gamma}$ contains all single threshold classifiers. Then $\tilde{\Gamma}$ is not learnable, even non-uniformly.*

The proof of Proposition 5.8 simply relies upon the fact that (1) seller randomizations are not restricted, and hence uncountable, and (2) that the buyer's statistical procedure attempts to make a different optimal response at every seller strategy. Hence the result holds even if the buyer were restricted to taking the same action following any fixed σ. The reason this proposition holds is that, without a restriction regarding optimality of the seller's behavior, the classifier must result in the buyer making the optimal decision *for every seller strategy*, including those that are suboptimal. Of course, the set of strategies that are suboptimal is an endogenous object, since it depends on how the buyer responds. Hence without positing some strategic foundations, there does not seem to be a way to strengthen this result.

Al-Najjar (2009) considers the related problem of a decisionmaker seeking to learn an underlying probability distribution over a set of outcomes. He obtains the striking result that, if the distribution over outcomes is countably additive, then there exists a hypotheses class with VC-dimension 1 which allow the decisionmaker to learn (uniformly) the probability of any event. This result holds due to the fact that any Borel measurable space is equivalent to the unit interval, and all events on the unit interval are learnable via half-spaces. Intuitively, this result holds due to the continuity of countably additive measures. When this continuity fails (for instance, with finite additive as focused on by Al-Najjar (2009)), learnability fails as well. While our problem is very different—and in particular, without an immediate way of mapping to the problem of learning a finitely-additive distribution—the lack of a restriction on how the classifiers react to different randomizations is why learning cannot be done.

For us, since we are concerned with learning the correct *labels* on each possible strategy-price pair, our agents requires much finer observations compared whether an event occurred. In this sense, our problem is more similar to learning the Borel sets; Al-Najjar (2009) notes this has infinite VC dimension and thus learnability fails. A difference in our environment is that we will, from now on, primarily be interested in the case where the set of hypothesis classes are *exogenously given*. When this is large, this can make the learning problem more difficult.[4]

More subtle is that in fact non-uniform learning is significantly weaker than the condition for uniform learning, the latter of which is used in the aforementioned papers which invoke VC dimension. In our setting, since our concern is explicitly the rational benchmark, it is less interesting if learning is hard relative to *any* possible hypothesis, as opposed to simply the "correct" one (i.e., the most rational). In other settings, having infinite VC dimension is *not* the condition which states it is impossible to learn a *particular* hypothesis with a given amount of data. Nevertheless, the strategy space (and set of possible classifiers) is still too large in order to guarantee that the true hypothesis is learned.

6. Statement of the Main Result

At this point, we are now prepared to state our main result. We have shown that the algorithm must extend beyond the model classes it is able to fit, but that taking a model class that is "too rich" will make the process of finding a best response more inefficient. Our result is to show that one can find a recursive ensemble algorithm for which the outcome of the game approximates rationality. More precisely, we construct algorithm $\tau^{\lambda}_{\hat{A}}$, which is parameterized by $\lambda > 0$. We show that for a large discount factor, the seller

[4]Notice that in Al-Najjar (2009), the set of hypotheses classes requires finding a measurable bijection from the observation space to a Borel set on the unit interval. This function need not be simple to construct.

follows the equilibrium strategy even though type 1 buyer is boundedly rational. For an outsider, the seller appears to treat type 1 buyer as if he is fully rational.

Proposition 6.1. $\exists \overline{\lambda} > 0$ *such that* $\forall \lambda \in (0, \overline{\lambda})$*, there exists* $\tau^{\lambda}_{\hat{A}} \in \mathcal{T}$ *such that the best response of the seller is the baseline equilibrium strategy* $\forall \kappa \in \mathcal{K}$*, for any* $\delta < 1$ *close to 1.*

The proposition is trivial, if the algorithm can use the parameter κ of the underlying game. The type 1 buyer can choose the equilibrium strategy of the baseline game, which is a single threshold decision rule with threshold is equal to v_L if the true state is L and v_H if the true state is H. Against the equilibrium strategy of the buyer, the seller has to play the equilibrium strategy of the baseline game. Because the algorithm cannot use the parametric information of the underlying game such as v_L, it is not feasible to implement the equilibrium strategy of the baseline game.

As we show in the next section, a uniformly PAC learnable algorithm of Σ generally fails to exist, unless we impose a restriction on Σ. Instead of imposing exogenous restriction, we exploit the rational behavior of the seller to restrict the set of strategies to those which is a best response to the algorithm. Instead of Σ, the algorithm appears as if the buyer learns the best response of the seller. Note however that the algorithm cannot observe $\kappa \in \mathcal{K}$ and therefore, cannot calculate the best response of the seller.

The rest of the paper proves Proposition 6.1 by constructing the algorithm in a number of steps, implementing the desired properties into the algorithm. Before the proof, we show that the existence of a uniformly PAC learnable algorithm is not guaranteed unless we impose a certain restriction on Σ. We then construct $\tau^{\lambda}_{\hat{A}}$ in a few steps. First, we describe the Adaptive boosting algorithm, which we denote as τ_A. Second, we constructed an intermediate algorithm $\tau_{\hat{A}}$ inspired by τ_A, which uses the ordinal data to infer the rational label $y(\sigma, p)$. Finally, we modify $\tau_{\hat{A}}$ to construct $\tau^{\lambda}_{\hat{A}}$ by treating "close" prices as equal. We prove Proposition 6.1 by $\tau^{\lambda}_{\hat{A}}$ which has the desired properties which we spelled out in Section 4.6.

7. Adaptive Boosting

We describe the candidate equilibrium algorithm, Adaptive Boosting algorithm (Schapire and Freund (2012)). To illustrate the basic structure of AdaBoost, let us assume for a moment that type 1 buyers observe σ as in Rubinstein (1993) so that $\forall p$ in the support of σ, type 1 buyer knows the value of the rational label $y(\sigma, p)$. We then relax a restrictive assumption while adding a new feature to construct the algorithm for Proposition 6.1.

7.1. Description.

Parameters and Initialization: To highlight the link between P and σ, let us write $P(\sigma) = \{p_1, \ldots, p_G\}$ as the support of σ with $G < \infty$. Recall that $y(\sigma, p)$ is the decision a rational agent would make in response to (σ, p). The algorithm has two key components: an artificial probability distribution $d_t(p)$ over $\mathbf{P}(\sigma)$, and threshold rule h_t in period t.

Define $d_1(p)$ as the uniform distribution over $P(\sigma)$, which is well defined as we assumed $G < \infty$.

Iteration: Suppose that $d_t(p)$ is defined $\forall p \in P(\sigma)$. Let $\mathbf{P}_{d_t}$ be the probability distribution over $P(\sigma)$ determined by d_t, not by σ.

Choose h_t by solving

$$\max_{h \in \mathcal{H}} \sum_{p \in P(\sigma)} h(p) y(p) d_t(p) \tag{7.13}$$

Define

$$\epsilon_t = \mathbf{P}_{d_t}\left(h_t(p) y(\sigma, p) = -1\right) \tag{7.14}$$

as the probability that the optimal classifier h_t at t misclassifies p under d_t. If $\epsilon_t = 0$, then we stop the training and use h as the forecasting rule, which perfectly forecast $y(\sigma, p)$.

Suppose that $\epsilon_t > 0$. Define

$$\alpha_t = \frac{1}{2} \log \frac{1-\epsilon_t}{\epsilon_t} = \log \sqrt{\frac{1-\epsilon_t}{\epsilon_t}}. \tag{7.15}$$

Define for each p in the support of σ, and each pair $(p, y(p))$,

$$d_{t+1}(p) = \frac{d_t(p) \exp(-\alpha_t y(\sigma, p) h_t(p))}{Z_t}$$

where

$$Z_t = \sum_{i=1}^{m} d_t(p) \exp(-\alpha_t y(\sigma, p) h_t(p)).$$

Given d_{t+1}, we can recursively define h_{t+1} and ϵ_{t+1}.

Final Output: The output is the *final hypothesis*

$$H_t(p) = \mathbf{sgn}\left(\sum_{k=1}^{t} \alpha_k h_k(p)\right)$$

which the decision maker will use to classify (σ, p), instead of h_t. Our objective of interest is

$$\mathbf{P}_{D_1}\left(H_t(p) y(\sigma, p) = -1\right)$$

and in particular, how quickly the probability of misclassification vanishes.

Let τ_A be the statistical procedure thus defined, which maps the observed data d_t at the beginning of period t into a classifier $\tau_A(d_t)$. In this case, $\tau_A(d_t) \notin \mathcal{H}$ as $\tau_A(d_t)$ typically entails multiple thresholds. τ_A is the Adaptive Boosting Algorithm (AdaBoost) invented by Schapire and Freund (2012).

7.2. **Weak Learnability.** The description of the algorithm is not quite complete: it is necessary that $\epsilon_t < \frac{1}{2}$ uniformly to ensure $\alpha_t > 0$. We must proved that the selected threshold rule in period t is better than a random guess, so that the classification error is strictly less than $1/2$.

The usefulness of this algorithm in the machine learning literature is largely due to the observation that one *only needs to start with a classifier that can outperform random guessing in order to be able to classify arbitrarily well* (Schapire and Freund (2012)). Our main result is that the above algorithm leads the buyer to making approximately optimal responses with high probability. In showing that this also works in a strategic setting, we proceed as follows:

In our model, h_t is not an accurate classifier, in the sense that the probability ϵ_t of misclassification can be arbitrary. Following the language of computer science, let us call h_t a *weak hypothesis*. The output of a statistical procedure (in this case, AdaBoost algorithm) is also a classifier, which we call the final hypothesis.

An important step is to show that an optimally selected h_t can do strictly better than a random guess that assigns 1 with probability 1/2, and −1 with probability 1/2. This property is referred to as *weak learnability* by Schapire and Freund (2012).

Definition 7.1. *h is weakly learnable if $\exists \rho > 0$ such that $\forall d$,*

$$\sum_{p \in P(\sigma)} d(p) y(\sigma, p) h(p) \geq \rho.$$

The substance of weak learnability is the fact that ρ is a *uniform* lower bound of the maximized objective function.

Definition 7.2. *If $\overline{h}$ solves*

$$\sum_{p \in P(\sigma)} d(p) y(\sigma, p) \overline{h}(p) \geq \sum_{p \in P(\sigma)} d(p) y(\sigma, p) h(p) \qquad \forall h,$$

$\overline{h}$ is an optimal weak hypothesis.

We show that an optimal weak hypothesis must be weakly learnable.

Lemma 7.3. *$\forall \sigma$ whose support has $G < \infty$ elements, $\exists \rho > 0$ such that $\forall d$*

$$\max_h \sum_{p \in P(\sigma)} d(p) y(\sigma, p) h(p) \geq \rho.$$

Proof. See Appendix A □

7.3. **Convergence.** We show that the probability that the final hypothesis inaccurately classifies σ vanishes at the exponential speed, replicating the proof by Schapire and Freund (2012).

Proposition 7.4. *Fix a positive integer $G < \infty$. $\exists \rho > 0$ such that $\forall \sigma \in \Sigma^G$ whose support is in P, $\mathbf{P}_{d_1}\big(H_t(p) y(\sigma, p) = -1\big) < e^{-\rho t}$.*

Proof. See Appendix B. □

The proof reveals that the rate at which the probability of misclassfication vanishes is determined entirely by the number of prices in the support of σ. Thus, the algorithm is efficient.

8. Algorithm with unobservable randomization scheme

8.1. **Description.** So far, we have assumed that the buyer can observe σ. We drop the assumption to construct another "intermediate" algorithm $\tau_{\hat{A}}$ before constructing the algorithm for Proposition 6.1. Now, type 1 buyer cannot observe σ, but observes the realized sign $\hat{y}_t(p)$ of

$$\sum_{t'=1}^{t-1} (v_{t'} - p) \tag{8.16}$$

where $v_{t'} \in \{v_H, v_L\}$ is the realized valuation in period t', for each p in the support of σ. That is,

$$\hat{y}_t(p) = \begin{cases} 1 & \text{if } \sum_{t'=1}^{t-1}(v_{t'} - p) \geq 0 \\ -1 & \text{if } \sum_{t'=1}^{t-1}(v_{t'} - p) < 0. \end{cases}$$

Let $f_t^y(p)$ be the empirical probability that $\hat{y}_t(p) = 1$ at the beginning of period t. Thus, $\hat{y}_t(p) = -1$ with probability $1 - f_t^y(p)$. Given $\{d_t(p), \hat{y}_t(p)\}_p$, h_t solves

$$\max_{h \in \mathcal{H}} \sum_p h(p) d_t(p)[1 \cdot f_t^y(p) - 1 \cdot (1 - f_t^y(p))]$$

and

$$\hat{\epsilon}_t = \sum_p d_t(p) \left[f_t^y(p) \mathbb{I}(h(p) = 1) + (1 - f_t^y(p)) \mathbb{I}(h(p) = -1) \right].$$

Following the same proof as in the proof of Lemma 7.3, we can show that $\exists \rho > 0$ such that

$$\hat{\epsilon}_t \leq \frac{1}{2} - \rho.$$

Since $\hat{y}_t(p)$ has the full support over $\{-1, 1\}$ $\forall t \geq 1$,

$$\hat{\epsilon}_t > 0.$$

Define

$$\alpha_t = \frac{1}{2} \log \frac{1 - \hat{\epsilon}_t}{\hat{\epsilon}_t}$$

and

$$d_{t+1}(p) = \frac{d_t(p) \exp(-\hat{y}_t(p) \alpha_t h_t(p))}{Z_t}$$

where $Z_t = \sum_p d_t(p) \exp(-\hat{y}_t(p) \alpha_t h_t(p))$. The final hypothesis is

$$\hat{H}_t(p) = \mathbf{sgn} \left(\sum_{k=1}^{t} \alpha_k h_k(p) \right)$$

Let $\tau_{\hat{A}}$ be the statistical procedure obtained by replacing y in AdaBoost algorithm τ_A in period t by $\hat{y}_t$ $\forall t \geq 1$.

8.2. **Remarks.** The ordinal information (8.16) about the average quality is necessary. Without access to (8.16), the algorithm cannot estimate $y(\sigma, p)$, which is critical for emulating the rational behavior.

The information contained in (8.16) is coarse, because the algorithm does not take any cardinal information about the parameters of the underlying game such as v_H and v_L. Without the cardinal information, the buyer cannot implement the equilibrium strategy of the baseline game, which is a single threshold rule but the threshold must be v_L. Because the algorithm does not rely on parameter values of the underlying game, the algorithm is robust against specific details of the game, if the algorithm can function as intended by the decision maker.

9. The algorithm

While $\tau_{\hat{A}}$ is designed to be robust against parametric details of the underlying problems, the algorithm is vulnerable to strategic manipulation by the rational seller. The proof of Proposition 7.4 reveals that the rate of convergence is decreasing as the number of prices in the support of σ increases. The seller can randomize over a countably infinitely many number of prices to slow down the convergence rate, and take advantage of the slow rate, if the discount factor is less than 1. By the same token, $\tau_{\hat{A}}$ may not PAC learn uniformly the strategies of the seller.

We need to revise $\tau_{\hat{A}}$ accordingly. Instead of processing individual prices, we let $\tau_{\hat{A}}$ process a group of prices at a time, treating "close" prices as the same group. In principle, we want to partition $(v_L, v_H]$ into a set of half-open intervals with size λ. Define

$$K^\lambda = \min\{k \mid v_H \in (v_{k-1}, v_k]\}.$$

Each interval is

$$(v_{k-1}, v_k]$$

where $v_0 = v_L$, and $v_k - v_{k-1} = \lambda$ with a possible exception of $k = K^\lambda$. Let us refer to each element of the partition as P_k with

$$P_0 = (v_L, v_L + \lambda], \cdots, P_{K^\lambda} = (v_{K^\lambda - 1}, v_H].$$

This construction relies on the precise information about v_L, and therefore, is not feasible. Since $\mathcal{K}$ is a compact set, we can choose

$$\underline{v}_L = \inf \mathrm{proj}_{v_L} \mathcal{K}$$

where

$$\mathrm{proj}_{v_L} \mathcal{K}$$

is the projection of $\mathcal{K}$ to the space of v_L. Similarly define

$$\overline{v}_H = \sup \mathrm{proj}_{v_H} \mathcal{K}$$

where

$$\mathrm{proj}_{v_L} \mathcal{K}$$

is the projection of $\mathcal{K}$ to the space of v_L. We partition $[\underline{v}_L, \overline{v}_H]$ into the collection of half open intervals of size $\lambda > 0$ with a possible exception of the last interval:

$$P_0 = [\underline{v}_L, \underline{v}_L + \lambda), \ldots, P_{K^\lambda} = [\underline{v}_L + (K^\lambda - 1)\lambda), \overline{v}_H]$$

where K^λ is the number of elements in the partition.

For each $k \in \{0, 1, \ldots, K^\lambda\}$, the algorithm receives an ordinal information about the average outcome from the decision, if it contains a price in the support of σ:

$$\hat{y}_t^\lambda(k) = \begin{cases} 1 & \text{if } \sum_{p \in P_k} \sum_{t'=1}^{t-1} [v_{t'} - p] \geq 0 \\ -1 & \text{if } \sum_{p \in P_k} \sum_{t'=1}^{t-1} [v_{t'} - p] < 0 \end{cases}$$

where p in the support of σ. Let $\tau_{\hat{A}}^\lambda$ be the algorithm obtained by replacing $\hat{y}_t(p)$ in $\tau_{\hat{A}}$ by $\hat{y}_t^\lambda(k)$. Note that as $\lambda \to 0$, the size of the individual elements in the partition shrinks and $\tau_{\hat{A}}^\lambda$ converges to $\tau_{\hat{A}}$ for a fixed σ.

Compared to τ_A and $\tau_{\hat{A}}$, $\tau_{\hat{A}}^{\lambda}$ takes only coarse information for two important reasons. First, the algorithm cannot differentiate two prices which are very close. This features makes the algorithm robust against strategic manipulation of the seller to slow down the speed of learning. Second, the algorithm cannot detect the precise consequence of its decision, but only the ordinal information of the past decision, aggregated over time. The second feature allows the alogorithm to operate with very little information about the details of the parameters of the underlying game.

We are ready to prove Proposition 6.1, which we state for the reference.

Proposition 9.1. $\exists \overline{\lambda} > 0$ *such that* $\forall \lambda \in (0, \overline{\lambda})$, *the seller's best response to* $\tau_{\hat{A}}^{\lambda}$ *is the equilibrium strategy of the baseline model.*

Proof. See Appendix D. □

10. Conclusion

In this paper, we have demonstrated how buyers may eventually play "as-if rationally" when they have access to single-threshold classifiers and their behavior can be determined by the outcome of a recursive ensemble algorithm (i.e., AdaBoost). As Rubinstein (1993) showed, this need not be the case when their behavior follows from the *optimally* chosen single-threshold classifier, and despite the complexity involved with determining this strategy based on data alone in a non-strategic setting. Using Rubinstein (1993) as a laboratory, this paper has articulated the following tradeoff in the design of statistical algorithms to mimic rationality: on the one hand, simply fitting a single-threshold classifier to data will fall short of rational play and be exploited by a seller. On the other hand, it may not be clear why this is the end of the story. By adding the ability to fit classifiers repeatedly and combining them in particular ways, we show how the rational benchmark can be restored. In this paper, we have taken as a black box the ability to fit these classifiers. But given this, our algorithm articulates exactly how to put these fitted classifiers together in order to construct one which can mimic rationality arbitrarily well. Going forward, given how productive the machine learning literature has been in terms of designing algorithms for the purposes of classification, we hope that our work will inspire further analysis of how these algorithms behave in strategic settings. Along these lines, we suspect that Rubinstein (1993) (or similar models) may be a useful laboratory for furthering this agenda beyond the issues we have looked at here.

Appendix A. Proof of Lemma 7.3

We need a preliminary result. Let $\overline{\theta}$ be the threshold of $\overline{h}$, and n', n'' be indexes for the prices adjacent to the threshold:

$$p_{n''} < \overline{\theta} < p_{n'}$$

but there is no other p_n satisfying

$$p_{n''} < p_n < p_{n'}.$$

Lemma A.1. *Fix an optimal weak hypothesis $\overline{h}$. If $d(p_{n'}) > 0$,*

$$\overline{h}(p_{n'})y(p_{n'}) = 1. \tag{A.17}$$

Similarly, if $d(p_{n''}) > 0$, $\overline{h}(p_{n''})y(\sigma, p_{n''}) = 1$.

Proof. Suppose that

$$\overline{h}(p_{n'})y(p_{n'}) = -1.$$

We can increase the threshold from $\overline{\theta}$ slightly so that the sign assigned by the classifier to $p_{n'}$ changes from $\overline{h}(p_{n'})$ to $-\overline{h}(p_{n'})$. That is, we shift the threshold of $\overline{h}$ by one "notch" so that the new weak hypothesis classifies one more price $p_{n'}$ correctly. Let $\tilde{h}$ be the classifier built around the new threshold.

$$\sum_n d(p_n)y(\sigma, p_n)\tilde{h}(p_n) - \sum_n d(p_n)y(\sigma, p_n)\overline{h}(p_n)$$
$$= d(p_{n'})y(\sigma, p_{n'})\tilde{h}(p_{n'}) - d(p_{n'})y(\sigma, p_{n'})\overline{h}(p_{n'})) = 2d(p_{n'}) > 0$$

which contradict the hypothesis that $\overline{h}$ is optimal with respect to $\overline{D}$. The analysis of the remaining case follows the same logic. □

To prove the lemma by way of contradiction, suppose that there exist a sequence d^k and $\overline{d}$ such that $d^k \to \overline{d}$ and

$$0 \geq \sum_p \overline{d}(p)y(\sigma, p)\overline{h}(p) \geq \sum_p \overline{d}(p)y(\sigma, p)h(p)$$

for all single threshold classifier h, where $\overline{h}$ is the optimal classifier under $\overline{d}$.

We first claim that

$$\sum_p \overline{d}(p)y(\sigma, p)\overline{h}(p) \geq 0.$$

To see this, suppose that

$$\sum_p \overline{d}(p)y(\sigma, p)\overline{h}(p) < 0.$$

Define

$$\tilde{h}(p) = -\overline{h}(p) \qquad \forall p$$

which is a feasible classifier. Then,

$$\sum_p \overline{d}(p)y(\sigma, p)\overline{h}(p) < 0 < \sum_p \overline{d}(p)y(\sigma, p)\tilde{h}(p)$$

which contradicts the hypothesis that $\overline{h}$ is an optimal choice with respect to $\overline{d}$.

Next, we claim that

$$\sum_p \overline{d}(p)y(\sigma, p)\overline{h}(p) > 0. \tag{A.18}$$

Suppose that

$$\sum_p \overline{d}(p)y(\sigma, p)\overline{h}(p) = 0. \tag{A.19}$$

We only examine the case where $\overline{d}(p) > 0 \ \forall p$. The general case follows from the same logic.

Let us consider a classifier $\tilde{h}$ defined as

$$\tilde{h}(p) = -\overline{h}(p).$$

Clearly,

$$\sum_{n=1}^{G} \overline{d}(p_n) y(\sigma, p_n) \tilde{h}(p_n) = -\sum_{n=1}^{G} \overline{d}(p_n) y(\sigma, p_n) \overline{h}(p_n) = 0.$$

Although $\tilde{h}$ has the same threshold as $\overline{h}$, $\tilde{h}$ incorrectly classifiers the adjacent elements around the threshold:

$$\tilde{h}(p_{n'}) y(\sigma, p_{n'}) = \tilde{h}(p_{n''}) y(\sigma, p_{n''}) = -1.$$

We follow the same logic as the proof of Lemma A.1 to construct a new classifier $\hat{h}$ which classifies $p_{n'}$ accurately by increasing the threshold slightly. Then,

$$\begin{aligned}\sum_{n=1}^{G} \overline{d}(p_n) y(\sigma, p_n) \hat{h}(p_n) &= \sum_{n=1}^{G} \overline{d}(p_n) y(\sigma, p_n) \hat{h}(p_n) - \sum_{n=1}^{G} \overline{d}(p_n) y(\sigma, p_n) \tilde{h}(p_n) \\ &= \overline{d}(p_{n'})(y(\sigma, p_{n'}) \hat{h}(p_{n'}) - y(\sigma, p_{n'}) \tilde{h}(p_{n'})) = 2\overline{d}(p_{n'}) > 0\end{aligned}$$

which contradicts the hypothesis that $\overline{h}$ is optimal with respect to $\overline{d}$.

Appendix B. Proof of Proposition 7.4

We replicate the proof in Schapire and Freund (2012) for later reference. Define

$$F_t(p) = \sum_{k=1}^{t} \alpha_k h_k(p).$$

Following the same recursive process described in Schapire and Freund (2012), we have

$$d_{t+1}(p) = \frac{d_1(p) \exp\left(-y(\sigma, p) \sum_{k=1}^{t} \alpha_k h_k(p)\right)}{\prod_{k=1}^{t} Z_k} = \frac{d_1(p) \exp(-y(\sigma, p) F_t(p))}{\prod_{k=1}^{t} Z_k}. \tag{B.20}$$

Following Schapire and Freund (2012), we can show that

$$\mathbf{P}\left(H_t(p) \neq y(\sigma, p)\right) = \mathbf{E} \sum_p d_1(p) \mathbb{I}(H_t(p) \neq y(\sigma, p)) \leq \mathbf{E} \sum_p d_1(p) \exp(-y(\sigma, p) F_t(p)),$$

and

$$\mathbf{P}(H_t(p) \neq y(\sigma, p)) = \mathbf{E} \prod_{k=1}^{t} Z_k.$$

Note

$$Z_k = \sum_p d_k(p) \exp\left(-y(\sigma, p) \alpha_k h_k(p)\right).$$

The rest of the proof follows from Schapire and Freund (2012), which we copy here for later reference.

$$\begin{aligned} Z_t &= \sum_p d_t(p) \exp\left(-y(\sigma, p) \alpha_t h_t(p)\right) \\ &= \sum_{y(\sigma,p) h_t(p) = 1} d_t(p) \exp(-\alpha_t) + \sum_{y(\sigma,p) h_t(p) = -1} d_t(p) \exp(-\alpha_t) \\ &= e^{-\alpha_t}(1 - \epsilon_t) + e^{\alpha_t} \epsilon_t \\ &= e^{-\alpha_t}\left(\frac{1}{2} + \gamma_t\right) + e^{\alpha_t}\left(\frac{1}{2} - \gamma_t\right) \\ &= \sqrt{1 - 4\gamma_t^2} \end{aligned}$$

where

$$\gamma_t = \frac{1}{2} - \epsilon_t.$$

By weak learnability, we know that γ_t is uniformly bounded away from 0: $\exists \gamma > 0$ such that

$$\gamma_t \geq \gamma \qquad \forall t \geq 1.$$

Recall that the maximum number of the elements in the support of σ is N. Thus,

$$d_{t+1}(p) = d_1(p) \prod_{k=1}^{t} \sqrt{1 - 4\gamma_t^2} \leq \frac{1}{N}\left(1 - 4\gamma^2\right)^{\frac{t}{2}} \leq \frac{1}{N} e^{-2\gamma^2 t}$$

where the right hand side converges to 0 at the exponential rate uniformly over p.

Appendix C. Proof of Proposition 5.8

For exposition, we first describe the argument for the case of uniform learnability, for which it suffices to show that $\tilde{\Gamma}$ has infinite VC-dimension. Consider an arbitrary σ which randomizes between at least two prices, p_L^σ, p_H^σ. Then the class $\tilde{\Gamma}$ is shatters the points (σ, p_L^σ) and (σ, p_H^σ); indeed, letting θ be an arbitrary point between p_L^σ and p_H^σ, simply consider the classifers $h_{\theta,1}, h_{\theta,2}$, where $h_{\theta,1}(p) = 1$ iff $p \geq \theta$ and $h_{\theta,2}(p) = 1$ iff $p \leq \theta$. Since these classifiers can rationalize an arbitrary assignment of labels to observations, we have that the VC dimension of any set containing σ is at least two.[5]

Now, given any finite number of distinction randomizations $S = \{\sigma_1, \ldots, \sigma_n\}$, note that the VC-dimension of the class Γ, when the sample space is S, is $2 \cdot |S|$; indeed, fixing two distinct price realizations that emerge under σ_i, the previous argument shows an arbitrary label can be assigned to either price by some classifier h_θ. Given an arbitrary assignment of labels, we then consider the product classifier $(h_\theta^i)_{i=1}^n$, where h_θ^i is the classifier that rationalizes the labels of the price observations when the randomization is σ_i. It follows that $\tilde{\Gamma}$ can shatter these $2|S|$ points, demonstrating that the VC-dimension of $\tilde{\Gamma}$ is at least $2|S|$, when there are S randomizations possible. Hence $\tilde{\Gamma}$ has infinite VC-dimension, whenever it contains an infinite number of distinct randomizations.

Now we extend the argument to show that that $\tilde{\Gamma}$ is not non-uniformly learnable—that is, it is not learnable given a true hypothesis γ^*. This holds provided we show that $\tilde{\Gamma}$ cannot be written as the countable union of hypothesis classes with finite dimension. We focus on the case where the seller can only randomize between two prices p_L and p_H. Fix any countable partition of $\tilde{\Gamma} = \cup_{n \in \mathbf{N}} \tilde{\Gamma}_n$. Note that the cardinality of the randomizations on p_L and p_H is equal to that of the unit interval. Hence any classifier can be associated with a function $h : (0,1) \times \{L, H\} \to \{-1, 1\}$, where $h(\sigma, q)$ denotes the decision when the probability of p_L is σ and p_q is realized. Hence the cardinality of the set of *all* classifiers is $2^{2^{\aleph_0}}$. On the other hand, given any countably infinite subset A of (σ, q), the cardinality of the set of classifiers that give some constant prediction to all (σ, q) outside of A is $2^{\aleph_0}$. It follows that some $\tilde{\Gamma_n}$ must contain a set of classifiers which can provide distinct predictions for an infinite number of (σ, q), and can hence shatter an infinite number of points. The result follows.

Appendix D. Proof of Proposition 6.1

We prove that the equilibrium strategy of the baseline model is the best response against $\tau_{\hat{A}}^\lambda$ if $\delta < 1$ is sufficiently close to 1. The proof is somewhat involved, because $\tau_{\hat{A}}^\lambda$ cannot use any preference parameter such as v_L to classify the price.

Lemma D.1. *Suppose that σ^e assigns$\sigma^e(v_L|L) = 1$ and $\sigma^e(v_H|H) = 1$. The long run average payoff of the seller against $\tau_{\hat{A}}^\lambda$ is $\pi_L v_L N$.*

Proof. Fix $\tau_{\hat{A}}^\lambda$ for some $\lambda > 0$. Since the support of σ^e is $\{v_L, v_H\}$, we only consider two partition 0 and K^λ which include v_L and v_H, respectively. Conditioned σ^e, $\hat{y}_t^\lambda(0) = 1$ and $\hat{y}_t^\lambda(K^\lambda) = -1$ $\forall t \geq 1$. $\tau_{\hat{A}}^\lambda$ immediately classifies $\hat{y}_t^\lambda(0) = 1$ and $\hat{y}_t^\lambda(K^\lambda) = -1$, since the classification can be implemented by a single threshold rule. The response of $\tau_{\hat{A}}^\lambda$ generates payoff stream $\pi_L v_L N$ in each period, from which the conclusion of the lemma follows. □

---

[5]In fact, the VC dimension of the set of prices with support in σ is exactly two, since adding any third point would imply that the label could only switch at most once.

We have to show that if σ assigns a positive probability to any other price than v_L, then the expected long run discounted average price is strictly less than $\pi_L v_L N$.

The next lemma is independent of an algorithm by the type 1 buyer, but uses the conditions for the lemon's problem.

Lemma D.2. (1) *If $p > v_L$ and $\mathbf{E}(v|p) - p \geq 0$, then the expected profit from p is strictly less than $\pi_L v_L N$.*

(2) *$\exists \bar{\epsilon} > 0$ such that $\forall \epsilon \in (0, \bar{\epsilon})$, $\exists \delta_\epsilon < 1$ so that $\forall \delta \in (\delta_\epsilon, 1)$, if $\sigma \in \Sigma$ assigns a positive weight to price $p > v_L + \sqrt{\epsilon}$ with*

$$\mathbf{E}(v|p) - p \geq -\epsilon,$$

then the expected payoff from σ is strictly less than the equilibrium strategy of the seller in the baseline game.

Proof. We write the proof in Rubinstein (1993) for the later reference. For any price p satisfying

$$\mathbf{P}(H|p)v_H + \mathbf{P}(L|p)v_L \geq p,$$

the revenue cannot exceed

$$\mathbf{P}(H|p)Nv_H + \mathbf{P}(L|p)Nv_L$$

but the cost is

$$\mathbf{P}(H|p)N_2c_2 + \mathbf{P}(H|p)N_1c_1.$$

Thus, the seller's expected profit is at most

$$\mathbf{P}(L|p)Nv_L + \mathbf{P}(H|p)(N_2(v_H - c_2) + N_1(v_H - c_1))$$

Because of the lemon's problem,

$$N_2(v_H - c_2) + N_1(v_H - c_1) < 0$$

and

$$\mathbf{P}(H|p) > 0$$

to satisfy

$$\mathbf{P}(H|p)v_H + \mathbf{P}(L|p)v_L \geq p > v_L.$$

Integrating over p, we conclude that the ex ante profit is strictly less than

$$\pi_L v_L N.$$

We prove the second part of the lemma. Fix $p > v_L + \sqrt{\epsilon}$ satisfying

$$\mathbf{E}(v|p) - p \geq -\epsilon \tag{D.21}$$

which implies that $\sigma(p|H) > 0$ and

$$\sigma(p|H)\pi_H(v_H - p) \geq \sigma(p|L)\pi_L(p - v_L) - \epsilon(\sigma(p|H)\pi_H + \sigma(p|L)\pi_L)$$

and therefore

$$\sigma(p|H)\pi_H(\mathbf{E}[c] - p) > \sigma(p|L)\pi_L(p - v_L) + \sigma(p|H)\pi_H(\mathbf{E}[c] - v_H) - \epsilon(\sigma(p|H)\pi_H + \sigma(p|L)\pi_L).$$

If

$$\sigma(p|H)\pi_H(\mathbf{E}[c] - v_H) - \epsilon(\sigma(p|H)\pi_H + \sigma(p|L)\pi_L) > 0, \tag{D.22}$$

then we have

$$\sigma(p|H)\pi_H(\mathbf{E}[c] - p) > \sigma(p|L)\pi_L(p - v_L)$$

from which the desired conclusion follows.

We can write (D.22) as

$$\frac{\sigma(p|H)\pi_H}{\sigma(p|L)\pi_L} > \frac{\epsilon}{\mathbf{E}[c] - v_H - \epsilon}. \tag{D.23}$$

From (D.21),

$$\frac{\sigma(p|L)\pi_L v_L + \sigma(p|H)\pi_H v_H}{\sigma(p|L)\pi_L + \sigma(p|H)\pi_H} \geq p - \epsilon$$

for $p \geq v_L + \sqrt{\epsilon}$. Thus,

$$\frac{\sigma(p|L)\pi_L v_L + \sigma(p|H)\pi_H v_H}{\sigma(p|L)\pi_L + \sigma(p|H)\pi_H} \geq v_L + \sqrt{\epsilon} - \epsilon$$

must hold $\forall p \geq v_L + \sqrt{\epsilon}$. Note that the left hand side is the convex combination of v_L and v_H, and the right hand side is increasing with respect to ϵ at the rate of $\sqrt{\epsilon}$ if $\epsilon > 0$ is sufficiently small.

$$\frac{\sigma(p|L)\pi_L v_L + \sigma(p|H)\pi_H v_H}{\sigma(p|L)\pi_L + \sigma(p|H)\pi_H} = O(\sqrt{\epsilon}).$$

Thus, $\exists \overline{\epsilon} > 0$ such that $\forall \epsilon \in (0, \overline{\epsilon})$,

$$\frac{\sigma(p|L)\pi_L v_L + \sigma(p|H)\pi_H v_H}{\sigma(p|L)\pi_L + \sigma(p|H)\pi_H} > \frac{\epsilon}{\mathbf{E}[c] - v_H - \epsilon}.$$

□

Remark D.3. *In calculating $\overline{\epsilon} > 0$, we use the assumption that the buyer knows the range of parameter values of the underlying game, if not the precise value of the game.*

We conclude that if σ is a best response of the seller, then $\forall p > v_L$,

$$\mathbf{E}(v|p) - p < 0.$$

The next lemma uses a property of $\tau^\lambda_{\hat{A}}$ that takes an average of the past observation.

Lemma D.4. *Fix $\tau^\lambda_{\hat{A}}$ and $p > v_L$ so that $\exists i \in \{0, 1, \ldots, K^\lambda\}$ so that $p \in P^i$. $\exists T^*$ such that $\forall t \geq T^*$,*

$$\mathbf{P}\left(\sum_{k=1}^{t} \alpha_k h_k(i) \geq 0\right) \leq \frac{1}{2}.$$

Proof. Recall that

$$\hat{y}_t(i) = \mathbf{sgn}\left(\sum_{p\in P^i}\sum_{k=1}^{t-1}[v_k - p]\right) = \mathbf{sgn}\left(\sum_{p\in P^i}\left[\frac{1}{t-1}\sum_{k=1}^{t-1}[v_k - p]\right]\right)$$

and

$$d_{t+1}(i) = \frac{d_1(i)\exp\left(-\sum_{k=1}^{T}\hat{y}_k(i)\alpha_k h_k(i)\right)}{\prod_{k=1}^{t} Z_k}.$$

We can show the weak learnability holds so that $\exists \rho > 0$ such that

$$\sum_p d_1(i)\exp\left(-\sum_{k=1}^{T}\hat{y}_k(i)\alpha_k h_k(i)\right) = \prod_{k=1}^{t} Z_k \leq e^{-\rho t} \qquad \forall\{\hat{y}_k, \alpha_k h_k\}_{k=1}^{t}.$$

Taking expectation over all sample paths $\{\hat{y}_k, \alpha_k h_k\}_{k=1}^{t}$,

$$\mathbf{E}\sum_{i=1}^{K^\lambda} d_1(i)\mathbf{E}\left[\exp\left(-\sum_{k=1}^{T}\hat{y}_k(i)\alpha_k h_k(i)\right) \mid \{\alpha_k h_k\}_{k=1}^{t}\right] = \prod_{k=1}^{t} Z_k \leq e^{-\rho t}.$$

Since the exponential function is convex,

$$\begin{aligned}\mathbf{E}\left[\exp\left(-\sum_{k=1}^{T}\hat{y}_k(i)\alpha_k h_k(i)\right) \mid \{\alpha_k h_k\}_{k=1}^{t}\right] &\geq \exp\left(-\sum_{k=1}^{T}\mathbf{E}\left[\hat{y}_k(i)\alpha_k h_k(i) \mid \{\alpha_k h_k\}_{k=1}^{t}\right]\right)\\ &= \exp\left(-\sum_{k=1}^{T}\mathbf{E}\left[\hat{y}_k(i)\right]\alpha_k h_k(i) \mid \{\alpha_k h_k\}_{k=1}^{t}\right).\end{aligned}$$

By the law of large numbers, $\exists \lambda_k$ such that

$$\hat{y}_k(i) = \begin{cases} -1 & \text{with probability at least } 1-\lambda_k \\ 1 & \text{with probability at most } \lambda_k \end{cases}$$

where

$$\lim_{k\to\infty} \lambda_k = 0.$$

Thus, $\mathbf{E}\hat{y}_k(i) = -1 + 2\lambda_k$. After collecting the terms, we have

$$\mathbf{E}\sum_{k=1}^{t}(1-2\lambda_k)\alpha_k h_k(i) \le -\rho t + \log N$$

which implies

$$\mathbf{E}\sum_{k=1}^{t}\alpha_k h_k(i) \le -\rho t + 2\mathbf{E}\sum_{k=1}^{t}\lambda_k\alpha_k h_k(i) + \log N.$$

Dividing both sides by t, we have

$$\mathbf{E}\frac{1}{t}\sum_{k=1}^{t}\alpha_k h_k(i) \le -\rho + 2\mathbf{E}\frac{1}{t}\sum_{k=1}^{t}\lambda_k\alpha_k h_k(i) + \frac{\log N}{t}.$$

Since $\lambda_k \to 0$ as $k \to \infty$,

$$\lim_{t\to\infty}\frac{1}{t}\sum_{k=1}^{t}\lambda_k\alpha_k h_k(i) = 0.$$

Thus, $\exists T^*$ such that $\forall t \ge T^*$,

$$\frac{1}{t}\mathbf{E}\sum_{k=1}^{t}\alpha_k h_k(i) \le -\frac{\rho}{2}$$

which implies

$$\mathbf{E}\sum_{k=1}^{t}\alpha_k h_k(i) \le -\frac{\rho t}{2}.$$

Thus, if σ is a best response to $\tau_{\hat{A}}^{\lambda}$,

$$\mathbf{P}\left(\mathbf{sgn}\left[\sum_{k=1}^{t}\alpha_k h_k(i)\right] = 1\right) \le \frac{1}{2}.$$

□

Lemma D.5. *$\exists \bar{\epsilon} > 0$, $\forall \epsilon \in (0, \bar{\epsilon})$, $\exists \delta_\epsilon < 1$ such that $\forall \delta \in (\delta_\epsilon, 1)$, and $\lambda > 0$ such that the equilibrium strategy of the baseline game is the best response of the seller against $\tau_{\hat{A}}^{\lambda}$.*

Proof. We show that $\forall p > v_L$, the expected profit cannot exceed $\pi_L v_L N$. Choose a small positive $\lambda < \epsilon$.

Suppose that $v_L < p < v_L + \sqrt{\epsilon}$. If σ is a best response, then $\forall p > v_L$, $\mathbf{E}(v|p) - p < 0$. Thus, only type 1 buyers buy, if at all. By Lemma D.4, the probability of accepting $p \in (v_L, v_L + \sqrt{\epsilon})$ is no more than 1/2. Thus, the expected payoff from such p is at most

$$\frac{1}{2}N_1\pi_L(v_L + \sqrt{\epsilon}) < N\pi_L v_L$$

for a sufficiently small $\epsilon > 0$. For a sufficinetly large $\delta < 1$, the expected discounted average payoff from $p \in (v_L, v_L + \sqrt{\epsilon})$ is strictly smaller than $N\pi_L v_L$.

Next, suppose that $p > v_L + \sqrt{\epsilon}$. We prove that $\exists \bar{\rho} > 0$ such that

$$\limsup_{t\to\infty}\frac{1}{t}\log \mathbf{P}\left(\hat{H}_t(p) \ne \hat{y}_t(p)\right) \le -\bar{\rho}.$$

We know that if $p \in \mathsf{supp}(\sigma)$ with $p > v_L$, then the previous reasoning implies that $p > v_L + \sqrt{\epsilon}$. If so,

$$\mathbf{E}(v|p) - p < -\epsilon.$$

Let P^i be the partion whre $p \in P^i$. Since any $p > v_L + \sqrt{\epsilon}$, implies

$$\mathbf{E}(v|p) - p < -\epsilon,$$

$$\mathbf{E}\left[\mathbf{E}(v|p) - p \mid P^i\right] < -\epsilon,$$

We can invoke Cramer's theorem to have $\rho_1 > 0$ such that

$$\limsup_{t\to\infty} \frac{1}{t} \log \mathbf{P}\left(\hat{y}_t(i) = 1\right) \leq -\rho_1.$$

That means, $\forall \epsilon \in (0, \overline{\rho})$, $\exists T_\epsilon$ such that $\forall t \geq T_\epsilon$,

$$\mathbf{P}\left(\hat{y}_t(i) = 1\right) \leq e^{-(\rho_1 - \epsilon)t}.$$

Let us consider the event where the strong law of large number holds

$$\mathcal{L} = \left\{\forall t \geq T_\epsilon, \forall p, \hat{y}_t(i) = 1\right\}.$$

Let $\mathcal{L}^c$ be the complement of $\mathcal{L}$. We can write $\forall t \geq T_\epsilon$,

$$\begin{aligned}
& \mathbf{P}\left(\hat{H}_t(i) \neq \hat{y}_t(i)\right) \\
= & \mathbf{P}\left(\hat{H}_t(i) \neq \hat{y}_t(i) \mid \mathcal{L}\right)\mathbf{P}(\mathcal{L}) + \mathbf{P}\left(\hat{H}_t(i) \neq \hat{y}_t(i) \mid \mathcal{L}^c\right)\mathbf{P}(\mathcal{L}^c) \\
\leq & \mathbf{P}\left(\hat{H}_t(i) \neq \hat{y}_t(i) \mid \mathcal{L}\right) + \mathbf{P}(\mathcal{L}^c) \\
\leq & \mathbf{P}\left(\hat{H}_t(i) \neq \hat{y}_t(i) \mid \mathcal{L}\right) + e^{-t(\rho_1 - \epsilon)}.
\end{aligned}$$

We calculate the upper bound of

$$\mathbf{P}\left(\hat{H}_t(i) \neq \hat{y}_t(i) \mid \mathcal{L}\right).$$

Following the same reasoning as in the proor of Lemma D.4, we know that $\exists \rho > 0$ such that

$$\sum_{i=0}^{K^\lambda} \exp\left(-\sum_{k=1}^{t} \hat{y}_k(i)\alpha_k h_k(i)\right) \leq K^\lambda e^{-\rho t}.$$

Thus,

$$\exp\left(-\sum_{k=1}^{t} \hat{y}_k(i)\alpha_k h_k(i)\right) \leq (K^\lambda)^2 e^{-\rho t}.$$

Following the same line of reasoning as in the proof of Proof of Proposition 7.4, we have

$$\begin{aligned}
& \mathbf{P}\left(\hat{H}_t(i) \neq \hat{y}_t(i) \mid \mathcal{L}\right) \\
\leq & \exp\left(-\hat{y}_t(i)\sum_{k=1}^{t} \alpha_k h_k(i)\right) \\
= & \exp\left(-\hat{y}_t(i)\sum_{k=1}^{t} \alpha_k h_k(i)\right) \exp\left(\sum_{k=1}^{t} \hat{y}_k(i)\alpha_k h_k(i)\right) \exp\left(-\sum_{k=1}^{t} \hat{y}_k(i)\alpha_k h_k(i)\right) \\
= & \exp\left(\sum_{k=1}^{t} (\hat{y}_k(i) - \hat{y}_t(i))\alpha_k h_k(i)\right) \exp\left(-\sum_{k=1}^{t} \hat{y}_k(i)\alpha_k h_k(i)\right) \\
\leq & \exp\left(\sum_{k=1}^{t} (\hat{y}_k(i) - \hat{y}_t(i))\alpha_k h_k(i)\right) (K^\lambda)^2 e^{-\rho t}.
\end{aligned}$$

Note that

$$\begin{aligned} & \exp\left(\sum_{k=1}^{t}(\hat{y}_k(i)-\hat{y}_t(i))\alpha_k h_k(i)\right) \\ = & \exp\left(\sum_{k=1}^{T_\epsilon-1}(\hat{y}_k(i)-\hat{y}_t(i))\alpha_k h_k(i)\right)\exp\left(\sum_{T_\epsilon}^{t}(\hat{y}_k(i)-\hat{y}_t(i))\alpha_k h_k(i)\right) \\ = & \exp\left(\sum_{k=1}^{T_\epsilon-1}(\hat{y}_k(i)-\hat{y}_t(i))\alpha_k h_k(i)\right) \end{aligned}$$

since over $\mathcal{L}$,

$$\hat{y}_k(i)=1 \qquad \forall t\ge T_\epsilon.$$

Since $\alpha_k<\infty$ $\forall k\in\{1,\ldots,T_\epsilon-1\}$,

$$\exp\left(\sum_{k=1}^{t}(\hat{y}_k(i)-\hat{y}_t(i))\alpha_k h_k(i)\right)\le \exp\left(2\sum_{k=1}^{T_\epsilon-1}\alpha_k(i)\right)<\infty$$

and is independent of t.

After the collecting the term, we have

$$\mathbf{P}\left(\hat{H}_t(i)\ne\hat{y}_t(i)\mid\mathcal{L}\right)\le (K^\lambda)^2\exp\left(2\sum_{k=1}^{T_\epsilon-1}\alpha_k\right)e^{-\rho t}$$

and therefore,

$$\begin{aligned} & \mathbf{P}\left(\hat{H}_t(i)\ne\hat{y}_t(i)\right) \\ \le & (K^\lambda)^2\exp\left(2\sum_{k=1}^{T_\epsilon-1}\alpha_k\right)e^{-\rho t}+e^{-(\rho_1-\epsilon)t}. \end{aligned}$$

For fixed $\epsilon>0$, let $t\to\infty$. We have

$$\limsup_{t\to\infty}\frac{1}{t}\log\mathbf{P}\left(\hat{H}_t(i)\ne\hat{y}_t(i)\right)\le-\min(\rho,\rho_1-\epsilon).$$

Because $\epsilon>0$ is arbitrary, $\overline{\rho}=\min(\rho,\rho_1)$ and

$$\limsup_{t\to\infty}\frac{1}{t}\log\mathbf{P}\left(\hat{H}_t(i)\ne\hat{y}_t(i)\right)\le-\overline{\rho}$$

as desired.

Now, we derive the conclusion. $\forall p\in(\mathsf{supp})\sigma$, if $p>v_L$, the probability of accepting such p does not exceed $\epsilon>$ and the maximum expected payoff from such p does not exceed ϵv_H if $\delta<1$ is sufficiently close to 1. Thus, for a sufficiently large $\delta<1$, the only best response against $\tau_{\hat{A}}^{\lambda}$ is the equilibrium strategy of the baseline game. □

References

Al-Najjar, N. I. (2009): "Decision Makers as Statisticians: Diversity, Ambiguity and Learning," *Econometrica*, 77(5), 1371–1401.

Al-Najjar, N. I., and M. M. Pai (2014): "Coarse decision making and overfitting," *J. Economic Theory*, 150, 467–486.

Basu, P., and F. Echenique (2019): "Learnability and Models of Decision Making under Uncertainty," forthcoming in Theoretical Economics.

Brown, Z., and A. MacKay (2019): "Competition in Pricing Algorithms," .

Calvano, E., G. Calzolari, V. Denicolò, and S. Pastorello (2019): "Artificial Intelligence, Algorithmic Pricing and Collusion," .

Cherry, J., and Y. Salant (2019): "Statistical Inference in Games," Northwestern University.
Cho, I.-K., and K. Kasa (2015): "Learning and Model Validation," *Review of Economic Studies*, 82, 45–82.
Cho, I.-K., and J. Libgober (2020): "Machine Learning for Strategic Inference in Principal-Agent Interactions," Emory University and University of Southern California.
Dietterich, T. G. (2000): "Ensemble Methods in Machine Learning," in *Multiple Classifier Systems*, pp. 1–15, Berlin, Heidelberg. Springer Berlin Heidelberg.
Eliaz, K., and R. Spiegler (2018): "A Model of Competing Narratives," Brown University and Tel Aviv University.
Esponda, I., and D. Pouzo (2014): "An Equilibrium Framework for Players with Misspecified Models," University of Washington and University of California, Berkeley.
Fudenberg, D., and K. He (2018): "Learning and Type Compatibility in Signaling Games," *Econometrica*, 86(4), 1215–1255.
Fudenberg, D., and D. M. Kreps (1995): "Learning in Extensive Form Games I: Self-confirming Equilibria," *Journal of Economic Theory*, 8(1), 20–55.
Fudenberg, D., and D. K. Levine (1993): "Steady State Learning and Nash Equilibrium," *Econometrica*, 61(3), 547–573.
——— (2006): "Superstition and Rational Learning," *American Economic Reivew*, 96, 630–651.
Hansen, K., K. Misra, and M. M. Pai (2020): "Algorithmic Collusion: Supra-competitive prices via independent algorithms," .
Kalai, G. (2003): "Learnability and rationality of choice," *Journal of Economic Theory*, 113(1), 104–117.
Liang, A. (2018): "Games of Incomplete Information Played by Statisticians," Discussion paper, University of Pennsylvania.
Marcet, A., and T. J. Sargent (1989): "Convergence of Least Squares Learning Mechanisms in Self Referential Linear Stochastic Models," *Journal of Economic Theory*, 48, 337–368.
Meyn, S. P. (2007): *Control Techniques for Complex Networks*. Cambridge University Press.
Olea, J. L. M., P. Ortoleva, M. M. Pai, and A. Prat (2019): "Competing Models," Columbia University, Princeton University and Rice University.
Rubinstein, A. (1986): "Finite Automata Play Repeated Prisoners Dilemma," *Journal of Economic Theory*, 39(1), 83–96.
——— (1993): "On Price Recognition and Computational Complexity in a Monopolistic Model," *Journal of Political Economy*, 101(3), 473–484.
Salant, Y. (2007): "On the Learnability of Majority Rule," *Journal of Economic Theory*, 135(1), 196–213.
Schapire, R. E., and Y. Freund (2012): *Boosting: Foundations and Algorithms*. MIT Press.
Shalev-Shwartz, S., and S. Ben-David (2014): *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press.
Spiegler, R. (2016): " Bayesian Networks and Boundedly Rational Expectations *," *The Quarterly Journal of Economics*, 131(3), 1243–1290.

Department of Economics, Emory University, Atlanta, GA 30322 USA
E-mail address: icho30@emory.edu
URL: https://sites.google.com/site/inkoocho

Department of Economics, University of Southern California, Los Angeles, CA 90089 USA
E-mail address: libgober@usc.edu
URL: http://www.jonlib.com/