LEARNING UNDERSPECIFIED MODELS

IN-KOO CHO AND JONATHAN LIBGOBER

ABSTRACT. This paper examines learning dynamics under non-parametric model uncertainty. We choose the monopolistic profit maximization problem (Myerson (1981)) as our laboratory. We consider a monopolist who chooses a learning algorithm to select a price following a history, facing non-parametric model uncertainty about the probability distribution of the buyer's valuation and bearing the computational cost. We posit that the monopolist has a lexicographic preference over profit and computational complexity while seeking an ϵ *dominant* algorithm that prescribes an ϵ best response against any cumulative distribution function of the buyer's valuation for any small $\epsilon > 0$. We construct a simplest ϵ dominant algorithm among all dominant algorithms when the distribution of the buyer's valuation satisfies the increasing hazard rate property. Our algorithm recursively estimates two parameters of the distribution, even if the actual distribution is parameterized by many more variables. The monopolist chooses a misspecified model to save computational cost while learning the true optimal decision uniformly over the set of feasible distributions.

KEYWORDS. Learning, Non-parametric model uncertainty, Parametric forecast, Underspecification, Algorithm, Dominant strategy, Uniform Learnability, Complexity Cost

1. INTRODUCTION

This paper considers a monopolist's choice of a pricing algorithm, unsure of the functional form of the actual distribution of the buyer's valuations and seeking to minimize the computational cost to find the optimal price. Our exercise involves an examination of learning dynamics under non-parametric model uncertainty.[1] We ask whether the monopolist can learn to charge an optimal price, and if so, what would be the "simplest" algorithm to do so. We show that a learning algorithm that estimates a linear function parameterized by two numbers can achieve the goal, even though the actual distribution function of the buyer's valuations could be highly non-linear.

Date: April 19, 2025.

We thank Francisco Castro, Drew Fudenberg, Nima Haghpanah, Kevin He, Johannes Hörner, Yuhta Ishii, Stephen Morris, David Parkes, Tim Roughgarden, and Lones Smith for helpful comments. The first author acknowledges financial support from the National Science Foundation (SES-1952882) and the Korea Research Foundation (2018S1A5A2A01029483). The second author gratefully acknowledges the hospitality of Yale University and the Cowles Foundation, which hosted him during parts of this research.

[1]Throughout this paper, we use parametric uncertainty for the case where the model is determined by a finite number of parameters, and the decision maker does not have objective knowledge about the parameters as in Battigalli, Cerreia-Vioglio, Maccheroni, Marinacci, and Sargent (2022). By non-parametric uncertainty, we mean the uncertainty about an infinite number of parameters, such as the functional form of a highly non-linear function. The formal definition is in Section 2.5.

The literature has developed diverse approaches to formulating decision problems under model uncertainty (e.g., Gilboa and Schmeidler (1989), Hansen and Sargent (2007), Maccheroni, Marinacci, and Rustichini (2006), Klibanoff, Marinacci, and Mukerji (2005) and Cerreia-Vioglio, Maccheroni, Marinacci, and Montrucchio (2013)).[2] We follow the basic framework of Battigalli, Francetich, Lanzani, and Marinacci (2019) and Cerreia-Vioglio, Maccheroni, Marinacci, and Montrucchio (2013), where the formal foundations for learning under model uncertainty are developed. Our paper focuses on model uncertainty caused by *underspecification*. By underspecification, we mean that the objective restrictions on the state space distribution are insufficient to pinpoint a specific functional form of the distribution.

We consider a market where a monopolist faces a sequence of multiple short-run buyers whose valuations are drawn from a cumulative distribution function F independently over time. We normalize the production cost to be 0. A standard approach treats F as common knowledge among all players. Instead, we assume that F belongs to a set $\mathcal{F}^\eta$ that is a collection of cumulative distribution functions with a Lipschitz continuous density function $f = F'$ with Lipschitz constant $\eta > 0$, satisfying the increasing hazard rate property.[3] The monopolist is endowed with objective knowledge that the actual distribution of the buyer's valuation belongs to $\mathcal{F}^\eta$, but is *not* endowed with a prior probability distribution over $\mathcal{F}^\eta$.[4]

Since the monopolist does not know the actual cumulative distribution function of the buyer's valuations, he has to rely on the "estimated cumulative distribution function" of the buyer's valuation to find the optimal price. Suppose that the monopolist's model of the distribution function is an ℓ-th order polynomial function of valuations such as $\sum_{l=0}^{\ell} \beta_l v^l$ with a suitable restriction on the coefficients so that the polynomial is consistent with the objective knowledge about $\mathcal{F}^\eta$. For example, the polynomial must increase in v and satisfy the increasing hazard rate property.

The monopolist estimates $\ell + 1$ coefficients of an ℓ-th order polynomial to construct an estimated cumulative distribution function $\hat{F}_t$ of the buyer's valuations, from which the monopolist calculates the price to be charged in period t. Given the price, a short-run buyer decides to buy at the posted price if the price is lower than his valuation. The monopolist uses the price and the realized demand to update the estimated distribution function of the valuations. We refer to a learning algorithm as the process that both determines the functional form and estimates the coefficients. We ask whether the monopolist can learn to charge the optimal price associated with the actual distribution function F.

The cumulative distribution function determining the buyer's values may not be a polynomial. Consequently, the monopolist can misspecify the functional form of the actual distribution function. If the learning dynamics converge, the monopolist's belief and the market's outcome must constitute a self-confirming equilibrium (Battigalli, Francetich, Lanzani, and Marinacci (2019)). We aim to find a learning algorithm that converges to a self-confirming equilibrium, whose outcome coincides with the actual profit-maximizing price and quantity under non-parametric model uncertainty. In particular, we explore how

[2]One can view the bandit problem (e.g., Rothschild (1974)) as a decision problem under uncertainty.

[3]The assumptions imply that, for all F in $\mathcal{F}^\eta$, a unique profit-maximizing price exists (Cole and Roughgarden (2014)) and the slope of the marginal revenue curve of the monopolist is negative.

[4]By objective knowledge, we mean that the actual distribution must be drawn from $\mathcal{F}^\eta$ (Cerreia-Vioglio, Maccheroni, Marinacci, and Montrucchio (2013)).

the decision maker chooses an algorithm if the monopolist bears the complexity cost of the learning algorithm (Rubinstein (1986)). A higher-order polynomial is more flexible in fitting the unknown distribution function but requires estimating more coefficients. The choice of the learning algorithm is determined by balancing its accuracy and complexity.[5]

Inspired by Hansen and Sargent (2007) and Rubinstein (1986), we consider an *algorithm game*: the zero-sum normal form game between the decision maker and the nature, where the monopolist selects an algorithm. The objective function of the monopolist is the expected discounted average profit. Facing an adversarial nature under model uncertainty, the monopolist could opt for a robust solution by choosing a maxmin algorithm that maximizes his objective function under the most pessimistic conjecture.

Instead, we search for an ϵ dominant algorithm that induces an ϵ best response uniformly against any cumulative distribution function in $\mathcal{F}^\eta$ for a small $\epsilon > 0$. Because the monopolist must be committed to an algorithm before receiving any information about nature's move, it is sensible to search for an ϵ dominant algorithm that leads to an ϵ best response, independently of new information the monopolist might receive in the future.[6]

Hansen and Sargent (2007) noted that when the decision maker searches for a robust decision rule that works well against a worst-case, the decision maker intends "to construct a rule that will work well across a set of models."[7] However, the gap between the robust solution and the optimal solution depends on details of the problems such as the size of the set of the feasible models satisfying the objective constraint. Unless we impose a tight bound on the set of feasible models, the gap can be large, and the robust solution could become degenerate and uninteresting.[8] In contrast, ϵ dominance imposes an upper bound for the gap between the optimal solution and the algorithm's forecast uniformly over the set of feasible distributions of the buyer. Our solution is robust in the sense that the algorithm prescribes a solution uniformly within a small neighborhood of an optimal solution of the actual distribution.[9]

Absent complexity costs, an existing proposal from Cole and Roughgarden (2014) identifies an ϵ dominant algorithm that non-parametrically estimates the actual distribution of the buyer's valuation. We search for an ϵ dominant algorithm while minimizing complexity costs, which we show can be achieved by an algorithm that estimates two parameters of a linear function.

---

[5]As the monopolist is exposed to non-parametric model misspecification, our exercise is related to Esponda and Pouzo (2016), which introduces a formal framework for studying learning under misspecified models. Nyarko (1991) is the first paper on the learning dynamics with misspecified models. Lanzani (2023) examines the learning dynamics under model uncertainty where the decision maker is aware of possible model misspecification. While misspecification is assumed in these papers, the misspecified model is derived as a consequence of optimization in our paper.

[6]If the monopolist does not discount the future payoff, we can set $\epsilon = 0$. The monopolist needs to be more patient for a small $\epsilon > 0$.

[7]https://julia.quantecon.org/dynamic_programming/robustness.html

[8]For example, the robust solution in the monopolist's optimization problem under model uncertainty is to charge a price of zero unless the set of feasible distributions is tightly restricted.

[9]By the same token, the solution is robust against small changes in the attitude toward uncertainty (Battigalli, Francetich, Lanzani, and Marinacci (2019)). However, we focus on the case where the monopolist is uncertainty-neutral.

An extensive set of feasible algorithms combined with the lexicographic preference over an algorithm's accuracy and complexity cost makes it challenging to find a dominant algorithm of the monopolist in the game against nature. Instead of searching directly for a dominant algorithm, we translate the profit maximization problem into the problem of forecasting the optimal price and quantity to invoke analytic tools developed in the machine learning literature.

We develop a uniform learnability condition inspired by the PAC (Probabilistically Almost Correct) criterion (Shalev-Shwartz and Ben-David (2014)). An algorithm $\mathcal{A}$ uniformly learns the optimal price for the unknown distribution of the buyer's valuation if, given $\mu > 0$ error bound and $1 - \lambda > 0$ confidence requirement, there is a positive integer $T(\mu, \lambda)$ so that after $T(\mu, \lambda)$ period, algorithm $\mathcal{A}$ produces a forecast that is within μ neighborhood of the actual optimal strategy with probability $1 - \lambda$ uniformly over the set of admissible distribution functions.[10] Since we have a well-developed set of tools to investigate the learnability of an algorithm (e.g., Evans and Honkapohja (2001)), it is significantly easier to check uniform learnability than the dominance of an algorithm.

We show that if an algorithm uniformly learns $\mathcal{F}^\eta$, the algorithm is an ϵ dominant strategy as long as the discount factor of the monopolist is close to 1. We construct a recursive algorithm that forecasts the optimal price based on the estimated linear demand curve and then show that the algorithm uniformly learns $\mathcal{F}^\eta$. We show that any algorithm simpler than the constructed algorithm cannot be an ϵ dominant algorithm.

Given a distribution of the buyer's valuation, we can construct the expected market demand for each price the monopolist charges. If the actual cumulative distribution function of the buyer's valuation satisfies a mild regularity condition, the seller's expected revenue is "locally quadratic" around the optimal price. Suppose a pair of price and quantity is the market outcome maximizing the expected profit for a given demand curve. We can find a linear demand curve where the same pair of price and quantity is the profit-maximizing outcome. Even though a linear demand curve might be a misspecified model of an actual demand curve, a linear demand is sufficiently flexible to "explain" why a given pair of price and quantity is an optimal outcome. The algorithm recursively estimates the linear demand curve that best explains the market outcome. Our solution concept requires us to show "uniform" convergence over $\mathcal{F}^\eta$, strengthening the conventional learnability result (e.g., Evans and Honkapohja (2001)).

The monopolist's goal is to learn about the profit-maximizing price and the maximum profit, not the demand curve itself.[11] In our exercise, the choice of the model specification for the distribution function is a part of the profit-maximizing decision of the monopolist. If the monopolist bears the computational cost, the chosen specification should be the simplest among those that approximately guarantee the profit-maximizing price. The monopolist chooses a simpler model of demand to save computational costs. A misspecified model would represent the seller's procedural rationality (Osborne and Rubinstein (1998)).

[10] Our convergence criterion does not require the data-generating process to be i.i.d. and allows the data-generating process to be endogenous.

[11] We are aiming for adequate learning rather than complete learning (Aghion, Bolton, Harris, and Jullien (1991)). Nevertheless, we require that the algorithm achieve adequate learning *uniformly* over the set of distributions of the valuations.

We depart from Cole and Roughgarden (2014) by considering the algorithm's complexity. An essential advantage of the non-parametric estimation technique is to avoid misspecification. The downside is that the non-parametric estimation algorithm is complex because the estimator requires a possibly unbounded number of parameters to represent a highly non-linear demand curve. If the monopolistic seller incurs the cost of storing the estimator in memory, he will search for an algorithm that uses a simpler specification of the demand curve.

A two-parameter specification is generally a misspecification of the actual cumulative distribution function of buyer valuations. The convergence result of Easley and Kiefer (1988) does not apply, as Easley and Kiefer (1988) required the true parameter to lie within the support of the prior belief over the parameter set. Battigalli, Francetich, Lanzani, and Marinacci (2019) obtained convergence to self-confirming equilibrium while allowing the decision maker to entertain a misspecified model under a strong form of consistency between the conjectured models and the true data generating process, which we do not impose. Under the additional assumption of own action independence, they also obtain convergence to the objectively optimal behavior.[12]

While the details of our complexity measure differ, our work is in the spirit of Olea, Ortoleva, Pai, and Prat (2022) in measuring the complexity of the forecasting rule. Nevertheless, the nature of data-generating processes in our paper differs fundamentally from Olea, Ortoleva, Pai, and Prat (2022), where the data-generating process is exogenous. By contrast, the data-generating process in our model is endogenous, as the monopolist sets the price, which determines the quantity of goods demanded. Endogeneity of the data-generating process requires an approach that differs substantially from Olea, Ortoleva, Pai, and Prat (2022) to investigate the learning dynamics.[13] In Olea, Ortoleva, Pai, and Prat (2022), an elaborate model edges out a simpler model as the decision maker observes more data. In contrast, our monopolistic seller chooses a simple model even in the long run.

The rest of the paper is organized as follows. In Section 2, we formally describe the problem and state the main result. Section 3 constructs our algorithm, which is simplest among all algorithms that uniformly learn the optimal strategy. Section 4 provides an informal description of the proof of the main result while placing the formal proof in the appendix. Section 5 describes applying the same methodology to a general optimization problem.

2. Description

2.1. **Demand.** There are N buyers, each of whom is indexed by $i \in \{1, \ldots, N\}$ and is endowed with reservation value $v_i \in [\underline{v}, \overline{v}]$ with $\overline{v} > \underline{v} \geq 0$. Let $F(v_i)$ be the cumulative distribution function of valuation of buyer i.[14] We assume that v_i and v_j are independent for all $i \neq j$. Given a price p, buyer i purchases one unit of the good if $p \leq v_i$. If the seller

[12] We are grateful for an anonymous referee for the observation.

[13] Cho and Kasa (2015) allows the decision maker to select a particular model specification in response to the validation process, subject to the endogenous data-generating process. While Cho and Kasa (2015) imposes a finite set of feasible specifications, this paper admits a significantly more general set of specifications.

[14] Symmetry is assumed only to simplify notation.

charges p, the (normalized) aggregate demand is

$$q = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}(v_i \geq p) \quad \text{and} \quad \mathbb{E}q = 1 - F(p).$$

Define

$$\epsilon_2 = q - (1 - F(p)), \tag{2.1}$$

where $\mathbb{E}\epsilon_2 = 0$. Note that the aggregate demand q is a random variable. We interpret $1-F(p)$ as the expected quantity of sales at a price p, which we call the (expected) demand curve.

We assume the cumulative distribution function F satisfies the increasing hazard rate and Lipschitz continuity property:

- (Increasing Hazard Rate.) Define

$$\mathcal{F}^0 = \left\{ F \mid f = F' \text{ is continuous and } \frac{f(v)}{1-F(v)} \text{ is increasing in } v \in (\underline{v}, \overline{v}) \right\}.$$

- (Lipschitz.) For all $\eta > 0$, define

$$\mathcal{F}^\eta = \left\{ F \in \mathcal{F}^0 \mid \ \forall v, v' \in [\underline{v}, \overline{v}], \ \left|f(v) - f(v')\right| \leq \eta |v - v'| \right\}$$

  as the collection of all feasible distributions over the buyer's valuations.

The objective knowledge of the monopolist is that the actual cumulative distribution function F is an element of $\mathcal{F}^\eta$.

For $F \in \mathcal{F}^\eta$, define $b^*(F)$ as

$$b^*(F) = \arg\max_p p(1 - F(p)). \tag{2.2}$$

The optimization problem has a unique solution $b^*(F)$ thanks to the increasing hazard rate property. We collect useful properties into the following lemma, whose proof is omitted.

Lemma 2.1. (1) *$\forall \eta > 0$, $\mathcal{F}^\eta$ is (sequentially) compact.*

(2) *$\forall \eta > 0$, there exists a compact set K in the interior of $\mathbb{R}^2_+$ so that $\forall F \in \mathcal{F}^\eta$, $(b^*(F), 1 - F(b^*(F))) \in K$.*

(3) *If $\eta < \eta'$, then $\mathcal{F}^\eta \subset \mathcal{F}^{\eta'}$. $\cup_{\eta>0}\mathcal{F}^\eta$ is a dense subset of $\mathcal{F}^0$.*

2.2. **Underspecification.** Departing from the classic framework of Myerson (1981), we endow the monopolist with objective knowledge that the actual distribution F is an element of $\mathcal{F}^\eta$ instead of a subjective belief over the buyer's value implied by some fixed distribution. In particular, the monopolist is *not* endowed with knowledge about the actual distribution F itself, or a prior probability distribution over $\mathcal{F}^\eta$. Since $\mathcal{F}^\eta$ is not a singleton, the seller's model about the distribution of buyer's valuations is underspecified.

2.3. **Seller's Problem.** The monopolist learns the expected profit-maximizing price of the actual distribution F from the data. To incorporate the learning process by the monopolist, we consider a dynamic version of the static model, where the long-run monopolist seller faces a sequence of short-run buyers with IID draws of their valuations in each period.

Time is discrete: $t = 1, 2, 3, \ldots$. In each period, N buyers enter the market, each endowed with reservation value $v_{i,t}$ drawn independently over time according to cumulative

distribution function F. Let $v_t = (v_{1,t}, \ldots, v_{N,t})$ be a profile of valuations of buyers in period t.

Let Γ be the game that the monopolist and N consumers play in each period. We admit any monopoly market trading protocol where the monopolist charges a single delivery price and can observe the quantity at the end of the one-period game Γ.

Assumption 2.2. *Let (p_t, q_t) be the pair of price and quantity in period $t \geq 1$, where q_t is the number of goods sold at the price p_t. At the end of the one-period game Γ, the monopolist observes a pair (p_t, q_t) of the price and the realized demand in period t.*

In period t, a buyer with valuation $v_{i,t}$ obtains payoff $v_{i,t} - p_t$ if he accepts p_t, and otherwise, 0. At the end of period t, the buyer leaves the market and never returns. Thus, it is optimal for a $v_{i,t}$ buyer to purchase the good at p_t if $v_{i,t} - p_t \geq 0$. We assume that every buyer follows the same decision rule.

2.4. **Data.** From F, we can construct the expected aggregate demand curve as $1 - F(p)$ for $p \in [\underline{v}, \overline{v}]$. The monopolist lacks a full description of the demand curve and has to estimate the aggregate demand based on the history of outcomes. The details of the one-period game Γ profoundly influence the statistical procedure that the monopolist uses to learn about the optimal price.

Let us illustrate two examples of possible one-period game Γ. Both games are based on the same monopoly market but generate substantially different outcomes ex-post.

Example 2.3. *Suppose that Γ_{CL} is the "textbook" static monopoly market (also known as the posted price mechanism). At the beginning of period t, the monopolist estimates the expected demand curve by $\hat{Q}_{t-1}(p)$ where p is the price. The monopolist chooses p_t, satisfying*

$$p_t \hat{Q}_{t-1}(p_t) \geq p \hat{Q}_{t-1}(p) \qquad \forall p \geq 0.$$

If $p_t \leq v_{i,t}$, buyer i purchases one unit of goods. If $p_t > v_{i,t}$, buyer i rejects this offer. Let q_t be the actual number of buyers who purchase the good in period t. All buyers in period t leave the game and never return. Since the buyer's valuation is drawn each period, q_t is a random variable. Using (p_t, q_t), the monopolist updates $\hat{Q}_{t-1}$ to $\hat{Q}_t$ according to a forecasting rule $\mathcal{A}$, for example, by minimizing the mean squared forecasting errors of the quantity.

In Example 2.3, the seller observes only the actual sales amount in period t (q_t) and remembers the price p_t he charges. Thus, the outcome at the end of period t is (p_t, q_t), which is also the data that forecasting rule $\mathcal{A}$ uses to calculate $\hat{Q}_t$.

Example 2.4. *Suppose that Γ_{CR} is the revelation game (Cole and Roughgarden (2014)) of the monopoly market game in Example 2.3. At the beginning of period t, the monopolist has the estimated aggregate distribution $\hat{F}_{t-1}$ of buyers' valuations based on data available at the end of period $t-1$. The monopolist chooses p_t that maximizes the estimated expected profit:*

$$p_t(1 - \hat{F}_{t-1}(p_t)) \geq p(1 - \hat{F}_{t-1}(p)) \qquad \forall p \geq 0.$$

The monopolist asks each buyer to report his type. Let $\hat{v}_{i,t}$ be the reported valuation of buyer i in period t. The monopolist observes $(\hat{v}_{1,t}, \ldots, \hat{v}_{N,t})$. If $p_t \leq \hat{v}_{i,t}$, buyer i receives

one unit of goods, paying p_t. Otherwise, buyer i does not receive the goods, paying 0. Let q_t be the actual number of buyers who purchase the good in period t. All buyers in period t leave the game and never return. The monopolist's payoff in period t is $p_t q_t$.

The monopolist uses the following algorithm to update $\hat{F}_{t-1}$, which we call $\mathcal{A}_{CR}$.[15] *The monopolist randomly samples $K(\leq N)$ elements from $\{\hat{v}_{1,t}, \ldots, \hat{v}_{N,t}\}$ and constructs an empirical distribution $\tilde{F}_t$ from K samples of reported types. The monopolist updates the estimated aggregate distribution according to*

$$\hat{F}_t(v) = \hat{F}_{t-1}(v) + \frac{1}{t}\left(\tilde{F}_t(v) - \hat{F}_{t-1}(v)\right) \qquad \forall v. \tag{2.3}$$

Example 2.4 highlights the difference between an outcome of the one-period game Γ_{CR} and the data for an algorithm $\mathcal{A}_{CR}$. By the nature of the revelation game, the monopolist observes

$$O_t = (\hat{v}_{1,t}, \ldots, \hat{v}_{N,t}, p_t, q_t) \tag{2.4}$$

in period t. Forecasting rule $\mathcal{A}_{CR}$ uses only $K(\leq N)$ elements from $(\hat{v}_{1,t}, \ldots, \hat{v}_{N,t})$ to update the estimated aggregate distribution $\hat{F}_{t-1}$ to $\hat{F}_t$, while ignoring (p_t, q_t). Thus, the data for $\mathcal{A}_{CR}$ is

$$D_t = (\hat{v}_{i_1,t}, \ldots, \hat{v}_{i_K,t})$$

which is a part of the statistical procedure, while the outcome (2.4) is a primitive of Γ_{CR}.

2.5. **Algorithms.** The outcome O_t in period t is what a player can observe at the end of period t. We regard O_t as an array of elements as in (2.4). Define $\dim O_t$ as the number of elements in O_t. Let $\mathcal{O}_t = (O_1, \ldots, O_{t-1})$ be the history at the beginning of period t. Define $\mathcal{O}$ as the collection of all histories, whose generic element is $\mathcal{O}_t$ for some $t \geq 1$.[16]

An algorithm is a mapping from a history of outcomes to the forecast:

$$\mathcal{A} : \mathcal{O} \to \mathcal{F}^\eta.$$

An algorithm is referred to as a statistical procedure or a forecasting rule.

By data, we mean the subset of the outcome that the statistical procedure uses as inputs of an algorithm. Let D_t be the data in period t, which is a "sub-array" of an outcome O_t in period t. Let $\dim D_t$ be the number of components in D_t. Being a "sub-array" of O_t, $\dim D_t \leq \dim O_t$.

The decision-maker can ignore some elements in the outcome, as in Example 2.4. In Example 2.3, $\dim D_t = 2$ and in Example 2.4, $\dim D_t = K$. Define $\mathcal{D}_t = (D_1, \ldots, D_{t-1})$ and $\mathcal{D}$ as the collection of all feasible $\mathcal{D}_t$ for all $t \geq 1$, which we call the data structure of an algorithm.

Definition 2.5. *$\mathcal{A}$ is a recursive algorithm if there exists a function Φ and a sequence of states $\{\Omega_t\}_{t\geq 1}$ such that*

$$\begin{bmatrix} \mathcal{A}(\mathcal{O}_{t+1}) \\ \Omega_{t+1} \end{bmatrix} = \Phi\left(\mathcal{A}(\mathcal{O}_t), D_t, \Omega_t\right)$$

where $\mathcal{O}_{t+1}$ is obtained by concatenating $\mathcal{O}_t$ to O_t and Ω_t is any state in period t used by the algorithm, but not a part of the forecast, which can be a function of $\mathcal{O}_t$.

[15] $\mathcal{A}_{CR}$ is the recursive version of the algorithm used in Cole and Roughgarden (2014) designed for the revelation game.

[16] It is a private history of the monopolist, as O_t can include variables only the monopolist can observe.

We require updating the internal state variable Ω_t according to the fixed function Φ. Let $\dim \Omega_t$ be the number of elements in Ω_t, treating Ω_t as an array of variables.

By setting the image of $\mathcal{A}(\mathcal{O}_t)$ to be any probability distribution of the buyer's valuation and $\Omega_t = \mathcal{O}_t$, we can represent any forecasting rule of the monopolist as a recursive algorithm. To have a "workable" recursive algorithm, we need to restrict $\mathcal{A}(\mathcal{O}_t)$ to a low dimensional family of probability distributions over buyer's valuation while restricting Ω_t to a lower dimensional state. This restriction captures the bounded rationality of the monopolist, who economizes the computational resources by simplifying the forecasting procedure while trying to maximize the expected profit.[17]

To easily quantify the complexity of an algorithm, we opt for a parametric family of specifications widely used in practice but also sufficiently general to approximate a large class of functions. By a polynomial, we mean a real-valued function F from $[\underline{v}, \overline{v}]$ to $[0, 1]$ where there exists $\{\beta_i\}_{i=0}^{\infty}$ such that

$$F(v) = \sum_{i=0}^{\infty} \beta_i v^i. \tag{2.5}$$

By an ℓ-th order polynomial, we mean (2.5) satisfying that for all $i \geq \ell + 1$, $\beta_i = 0$ so that (2.5) can be written as

$$F(v) = \sum_{i=0}^{\ell} \beta_i v^i. \tag{2.6}$$

Let X^ℓ be the collection of all ℓ-th order polynomials. Under our definition, if $\ell < \ell'$, then $X^\ell \subset X^{\ell'}$. Define

$$S^{\ell,\eta} = \mathcal{F}^\eta \cap X^\ell.$$

as the collection of cumulative distribution functions in $\mathcal{F}^\eta$ that can be represented as an ℓ-th order polynomial. Clearly, if $\ell < \ell'$, $S^\ell \subset S^{\ell'}$. Define

$$S^{\infty,\eta} = \mathcal{F}^\eta \setminus \cup_{\ell \geq 0} X^\ell$$

as the collection of elements in $\mathcal{F}^\eta$ that cannot be represented by an ℓ-th order polynomial for any finite ℓ, requiring infinitely many coefficients to represent the function.[18]

A forecasting rule is

$$\mathcal{A} : \mathcal{O} \to \mathcal{F}^\eta. \tag{2.7}$$

Define $\dim \mathcal{A}(\mathcal{O}_t)$ as the number of coefficients that have to be estimated, and $\mathcal{A}(\mathcal{O}_t)(v)$ as the value of the estimated cumulative distribution function at $v \in [\underline{v}, \overline{v}]$. Recall that $S^{\ell,\eta} \subset \mathcal{F}^\eta$ for any non-negative integer ℓ. If $\mathcal{A} : \mathcal{O} \to S^{\ell,\eta}$, $\mathcal{A}$ can be considered a

[17] Sargent (1993) treats such a restriction as an exogenous constraint of the computational capability of a decision maker. Rubinstein (1986) and Osborne and Rubinstein (1998) treat the same restriction as a choice of the monopolist who bears the computational cost. As a part of optimization, the monopolist tries to reduce the computational cost, even though he is capable of much more complex calculations. This paper generally follows the view of Osborne and Rubinstein (1998).

[18] The Stone Weierstrass Theorem (Theorem 7.26, Rudin (1976)) implies that any $F \in \mathcal{F}^\eta$ can be approximated by a uniformly convergent sequence of polynomials. Even though $F \in \mathcal{F}^\eta$ can be uniformly approximated by an ℓ-th order polynomial, we do not assume that $f = F'$ is also uniformly approximated by the derivative of the ℓ-th order polynomial. Since the optimal solution depends on $f = F'$, the uniform convergence to $F \in \mathcal{F}^\eta$ does not guarantee convergence of the optimal solution.

parametric forecasting rule, and if $\mathcal{A} : \mathcal{O} \to S^{\infty,\eta}$, it can be called a non-parametric forecasting rule.

Definition 2.6. A *is the set of all recursive forecasting rules.*

Based on the estimated distribution function $\mathcal{A}(\mathcal{O}_t)$ following history $\mathcal{O}_t$, the estimated expected demand at $p \in [\underline{v}, \overline{v}]$ is

$$1 - \mathcal{A}(\mathcal{O}_t)(p). \tag{2.8}$$

Define

$$\varphi_p(\mathcal{A}(\mathcal{O}_t)) = \arg\max_p p\left(1 - \mathcal{A}(\mathcal{O}_t)(p)\right) \tag{2.9}$$

as the optimal price for the estimated demand curve (2.8) and

$$\varphi_q(\mathcal{A}(\mathcal{O}_t)) = 1 - \mathcal{A}(\mathcal{O}_t)(\varphi_p(\mathcal{A}(\mathcal{O}_t)))$$

as the expected quantity at the optimal price for the estimated demand curve $\mathcal{A}(\mathcal{O}_t)$. Let

$$\varphi(\mathcal{A}(\mathcal{O}_t)) = (\varphi_p(\mathcal{A}(\mathcal{O}_t)), \varphi_q(\mathcal{A}(\mathcal{O}_t))) \tag{2.10}$$

be the pair of the point estimate of the optimal price and the expected demand at the estimated optimal price generated by $\mathcal{A}$ conditioned on history $\mathcal{O}_t$.

2.6. **Behavior Rule.** We focus on a simple behavior rule of the monopolist in response to forecast $\mathcal{A}(\mathcal{O}_t)$. The actual price p_t in period t is

$$p_t = \varphi_p(\mathcal{A}(\mathcal{O}_t)) + \epsilon_{1,t} \tag{2.11}$$

where $\epsilon_{1,t}$ is a white noise with a small variance $\sigma_1^2 \geq 0$. Quantity q_t in period t is

$$q_t = 1 - F(p_t) + \epsilon_{2,t}. \tag{2.12}$$

Since $\epsilon_{2,t} = q_t - \mathbb{E}q_t$, $\mathbb{E}\epsilon_{2,t} = 0$ but $\epsilon_{2,t}$ is not i.i.d. random variable since its distribution is affected by (q_t, p_t).

For example, if the algorithm forecasts the distribution function of the buyer's valuations is a uniform distribution, the resulting demand curve is a linear function with coefficient $(\beta_{0,t}, \beta_{1,t})$ so that for all $p \in [\underline{v}, \overline{v}]$,

$$1 - \mathcal{A}(\mathcal{O}_t)(p) = \beta_{0,t} + \beta_{1,t}p,\ \varphi_p(\mathcal{A}(\mathcal{O}_t)) = -\frac{\beta_{0,t}}{2\beta_{1,t}} \text{ and } \varphi_q(\mathcal{A}(\mathcal{O}_t)) = \frac{\beta_{0,t}}{2}.$$

The actual distribution function of the buyer's valuations is unknown but fixed, and so is the actual expected demand curve. A natural way to measure the forecasting error would be to compare (2.10) to the optimal outcome $(b^*(F), 1 - F(b^*(F)))$ associated with the actual expected demand curve $1 - F(p)$, where $b^*(F)$ is the (static) profit-maximizing price against the (unknown) actual expected demand curve $1 - F(p)$.

The monopolist mostly follows the recommendation $\varphi_p(\mathcal{A}(\mathcal{O}_t))$ of the algorithm but experiments a little to learn more about the demand curve. Combining the algorithm and the behavior rule, we have a mapping from a history $\mathcal{O}_t$ at period t to an action (i.e., price) in period t, following any history. As we fix the behavior rule as (2.11), we regard $\mathcal{A}$ as a strategy of the long-run monopolist.

2.7. **Complexity Measure.** We measure the complexity of an algorithm by considering three components: the number of parameters in the forecast of the algorithm, the number of components of data used, and the number of state variables the algorithm has to update internally to produce a forecast.[19] Since $\mathcal{A}$ is a recursive algorithm, the input of $\mathcal{A}$ in period t is $(\mathcal{A}(\mathcal{O}_t), D_t, \Omega_t)$. Define

$$\dim(\mathcal{A}(\mathcal{O}_t), D_t, \Omega_t) = \dim(\mathcal{A}(\mathcal{O}_t)) + \dim(D_t) + \dim(\Omega_t)$$

as the number of variables the algorithm needs to process in period t. Our complexity measure focuses on the state space size necessary for the algorithm's operation in each period t.

Definition 2.7. *The complexity of $\mathcal{A}$ is*

$$\text{comp}(\mathcal{A}) = \sup_{t\geq 1} \dim(\mathcal{A}(\mathcal{O}_t), D_t, \Omega_t).$$

2.8. **Algorithm Game.** To formalize how the monopolist chooses an algorithm, we formulate the monopolist's decision problem as a zero-sum game between the monopolist and nature, called the algorithm game. The monopolist's strategy space is A as defined in Definition 2.6. Nature's strategy space is $\mathcal{F}^\eta$. After $(\mathcal{A}, F) \in \mathsf{A} \times \mathcal{F}^\eta$ is selected, the continuation game is played between the algorithm and the sequence of short-run consumers as in the machine game in Rubinstein (1986).[20]

Conditioned on $(\mathcal{A}, F) \in \mathsf{A} \times \mathcal{F}^\eta$, the algorithm generates an estimated cumulative distribution function $\mathcal{A}(\mathcal{O}_t)$ for all $\mathcal{O}_t$. The monopolist calculates $\varphi_p(\mathcal{A}(\mathcal{O}_t))$ according to (2.9) and charges $p_t = \varphi_p(\mathcal{A}(\mathcal{O}_t)) + \epsilon_{1,t}$. The actual demand is $q_t = 1 - F(p_t) + \epsilon_{2,t}$. Let (p_t, q_t) be the realized outcome at the end of period t, and write

$$\mathcal{U}(\mathcal{A}(\mathcal{O}_t), F) = p_t q_t$$

as the realized payoff of the monopolist in period t.

Define the expected discounted average payoff from $(\mathcal{A}, F)$ as[21]

$$\mathcal{V}(\mathcal{A}, F) = \mathbb{E}(1-\delta) \sum_{t=1}^{\infty} \mathcal{U}(\mathcal{A}(\mathcal{O}_t), F)\delta^{t-1}.$$

The monopolist's preference over $(\mathsf{A}, \mathcal{F}^\eta)$ is the lexicographic ordering over the robustness and the complexity cost. To highlight the impact of the complexity cost, we choose the lexicographic preference over the average payoff and the complexity of an algorithm so that the importance of the complexity cost is secondary to the long-run average discounted payoff.

Given a subjective probability measure μ over $\mathcal{F}^\eta$, define the set of best responses as

$$B(\mu) = \arg\max_{\mathcal{A}\in\mathsf{A}} \int_{F\in\mathcal{F}^\eta} \mathcal{V}(\mathcal{A}, F) d\mu(F).$$

[19]This type of complexity measure is consistent with the complexity measure for a neural network (cf. Rumelhart, McClelland, and the PDP Research Group (1986), Wasserman (1989) and Weisbuch (1990)).

[20]If nature's strategy space consists of a single distribution $F \in \mathcal{F}^\eta$, our model is reduced to Myerson (1981).

[21]The expected utility representation of the preference is based on Cerreia-Vioglio, Maccheroni, Marinacci, and Montrucchio (2013). The players are assumed to be uncertainty-neutral.

Define

$$\overline{B} = \bigcap_{\mu} B(\mu)$$

as the set of dominant algorithms. The monopolist selects the simplest dominant algorithm

$$\overline{\mathcal{A}} \in \arg\min_{\mathcal{A} \in \overline{B}} \text{comp}\mathcal{A}.$$

For a small $\epsilon > 0$, define the set of ϵ best responses as

$$B^{\epsilon}(\mu) = \left\{ \mathcal{A} \in \mathsf{A} \mid \int_{F \in \mathcal{F}^{\eta}} \mathcal{V}(\mathcal{A}, F) d\mu(F) \geq \int_{F \in \mathcal{F}^{\eta}} \mathcal{V}(\mathcal{A}', F) d\mu(F) - \epsilon \ \ \forall \mathcal{A}' \in \mathsf{A} \right\}.$$

Define the set of ϵ dominant algorithms as

$$\overline{B}^{\epsilon} = \bigcap_{\mu} B^{\epsilon}(\mu)$$

and a simplest ϵ dominant algorithm as

$$\overline{\mathcal{A}}^{\epsilon} \in \arg\min_{\mathcal{A} \in \overline{B}^{\epsilon}} \text{comp}\mathcal{A}.$$

We state the main result of the paper. Recall that if $\mathcal{A}_d : \mathcal{O} \to \mathcal{F}^{\eta}$ is an algorithm, $\varphi(\mathcal{A}_d(\mathcal{O}_t)) = (\varphi_p(\mathcal{A}_d(\mathcal{O}_t)), \varphi_q(\mathcal{A}_d(\mathcal{O}_t))) \in \mathbb{R}^2$ is the forecast of the algorithm conditioned on history $\mathcal{O}_t$, where the first component $\varphi_p(\mathcal{A}_d(\mathcal{O}_t))$ is the forecast price, and the second component $\varphi_p(\mathcal{A}_d(\mathcal{O}_t))$ is the forecast expected quantity.

Theorem 2.8. (1) *$\forall \mu > 0$, $\forall \lambda > 0$, we can construct*

$$\mathcal{A}_d : \mathcal{O} \to S^{1,\eta}$$

where $\dim(\mathcal{A}_d(\mathcal{O}_t)) = \dim(D_t) = 2$ *$\forall t \geq 1$ so that* $\text{comp}\mathcal{A}_d = 4$ *and $\exists T(\mu, \lambda)$ such that*

$$\mathbb{P}\left(\exists t \geq T(\mu, \lambda), \ \left|\varphi(\mathcal{A}_d(\mathcal{O}_t)) - (b^*(F), 1 - F(b^*(F)))\right| > 4\mu\right) \leq \lambda \qquad \forall F \in \mathcal{F}^{\eta}$$

where

$$T(\mu, \lambda) = \mathsf{O}\left(\frac{1}{\mu^{\omega}} \left(\log \frac{1}{\mu}\right)^{\frac{1}{\omega(1-\omega)}} \log \frac{1}{\lambda}\right) \tag{2.13}$$

for $\omega \in (0, 1)$.

(2) *If the monopolist discounts future payoffs, $\exists \underline{\delta} \in (0, 1)$ such that $\forall \delta \in (\underline{\delta}, 1)$, $\mathcal{A}_d$ is an ϵ dominant algorithm so that*

$$\sup_{F \in \mathcal{F}^{\eta}} \left|b^*(F)(1 - F(b^*(F))) - \mathcal{V}(\mathcal{A}, F)\right| \leq \epsilon.$$

(3) *If $\mathcal{A} : \mathcal{O} \to S^{\ell,\eta}$ is a simplest algorithm, then $\ell \geq 1$.*

Proof. See Appendix A. □

The uniformity of $T(\mu,\lambda)$ over $\mathcal{F}^\eta$ is the substance of the first part of the theorem, which proves the existence of an algorithm with complexity measure 4 that can forecast the optimal price and expected demand accurately with high confidence by $T(\mu,\lambda)$ periods uniformly over $\mathcal{F}^\eta$.

The second part of the theorem shows that the constructed algorithm $\mathcal{A}_d$ is an ϵ dominant algorithm. Since $\mathcal{A}_d$ accurately forecasts the optimal price and quantity in the limit, for all $\mu > 0$,

$$\lim_{t\to\infty} \mathbb{P}\left(\left|\mathbb{E}\mathcal{U}(\mathcal{A}_d(\mathcal{O}_t),F) - b^*(F)(1-b^*(F))\right| < \mu\right) = 1. \tag{2.14}$$

Suppose the monopolist discounts the future payoff, but the discount factor $\delta \in (0,1)$ is sufficiently close to 1. The expected discounted average payoff is close to the maximum profit for an actual cumulative distribution function $F \in \mathcal{F}^\eta$. The first part of the theorem says that by $T(\mu,\lambda)$ period, the algorithm's forecast is within a small neighborhood of the optimal price and quantity with a probability close to 1 uniformly over $\mathcal{F}^\eta$. Thus, we can choose $\underline{\delta}$ uniformly over $\mathcal{F}^\eta$ so that for all $\delta > \underline{\delta}$, the discounted average expected payoff is within ϵ neighborhood of the maximum profit for any $F \in \mathcal{F}^\eta$.

The last part of Theorem 2.8 says that the algorithm $\mathcal{A}_d$ is a simplest ϵ dominant algorithm. If $\ell = 1$, then $F \in S^{1,\eta}$ is a uniform distribution that can be parameterized by two numbers. Since the expected demand generated by the uniform distribution function is a linear function, we can estimate the demand function by the slope and the intercept of the linear function. We need at least 2 observations in each period to estimate the two parameters, such as the pair (p_t, q_t) of price and quantity. Thus, the algorithm's complexity is at least 4 if $\ell = 1$. $\mathcal{A}_d$ has a complexity of 4, implying that it is one of the simplest ϵ dominant algorithms.

3. Construction

We now construct a recursive algorithm

$$\mathcal{A}_d : \mathcal{O} \to S^{1,\eta}$$

with $\dim(\mathcal{A}_d(\mathcal{O}_{t-1})) = \dim(D_t) = 2$ and $\dim(\Omega_t) = 0$ for all $t \geq 1$ that satisfies Theorem 2.8.

$\mathcal{A}_d$ uses price and quantity as data. $D_t = (q_t, p_t)$ is the pair of actual sales amount q_t and the price p_t charged in period t. Recall that q_t is a random variable with $\mathbb{E}q_t = 1 - F(p_t)$. We can write

$$q_t = 1 - F(p_t) + \epsilon_{2,t}. \tag{3.15}$$

$\mathbb{E}\epsilon_{2,t} = 0$ and $\mathbb{E}\epsilon_{2,t}^2 < \infty$ uniformly. Let

$$\mathcal{O}_t = (O_1, \ldots, O_{t-1})$$

be the history at the beginning of period t. The monopolist assumes that the valuation is drawn from a uniform distribution so that the aggregate demand is a linear function in $S^{1,\eta}$:

$$q = \beta_0 + \beta_1 p$$

and estimates (β_0, β_1). We write the estimator of (β_0, β_1) at the end of period t as $(\beta_{0,t}, \beta_{1,t})$.

Since $F \in \mathcal{F}^\eta$, the optimal solution $b^*(F)$ must generate a positive profit. Thanks to the uniform bound η, there exists a compact set K in the interior of $\mathbb{R}^2_+$ such that $b^*(F) \in K$ for all $F \in \mathcal{F}^\eta$. Thus, (β_0, β_1) must be such that the optimal price and the expected quantity under the linear demand curve parameterized by (β_0, β_1) must be contained in K.

Note that for all $F \in \mathcal{F}^\eta$, there exists $\beta^* = (\beta_0^*, \beta_1^*)$ such that

$$1 - F(b^*(F)) = \frac{\beta_0^*}{2} \quad \text{and} \quad b^*(F) = -\frac{\beta_0^*}{2\beta_1^*}$$

or equivalently,

$$\beta_0^* = 2(1 - F(b^*(F)) \quad \text{and} \quad \beta_1^* = -\frac{1 - F(b^*(F))}{b^*(F)}. \tag{3.16}$$

To emphasize that $\beta^* = (\beta_0^*, \beta_1^*)$ is conditioned on F, we often write $\beta^*(F) = (\beta_0^*(F), \beta_1^*(F))$ instead of β^*.

Since the pair $(b^*(F), 1 - F(b^*(F)))$ of the optimal price and the expected demand at the optimal price is in K, there exists a compact set $\mathsf{B} \subset (0, \infty) \times (-\infty, 0) \subset \mathbb{R}^2$ such that $(\beta_0, \beta_1) \in \mathsf{B}$ if the linear demand can support an actual optimal outcome. If $(\beta_0, \beta_1) \notin \mathsf{B}$, then the monopolist can conclude that the estimated demand is wrong, based on what the monopolist knows.

Let $\mathcal{H}_\mathsf{B} \subset S^{1,\eta}$ be the collection of the uniform distributions generating the linear expected demand curves, supporting the market outcomes in K: for all $F \in \mathcal{F}^\eta$, there exists (β_0, β_1) satisfying (3.16). The monopolistic seller recursively estimates (β_0, β_1) using (a sort of) the least square estimation method while choosing the pricing rule based on the estimated linear demand curve.[22] Let $(\beta_{0,t-1}, \beta_{1,t-1})$ be the estimator at the end of period $t-1$. Given the estimated demand curve

$$q = \beta_{0,t-1} + \beta_{1,t-1} p, \tag{3.17}$$

the monopolist calculates the optimal price

$$b_t = -\frac{\beta_{0,t-1}}{2\beta_{1,t-1}} \equiv \varphi_p\left(\mathcal{A}_d(\mathcal{O}_t)\right)$$

but experiments with the market price by adding $\epsilon_{1,t}$ so that the actual price in period t is

$$p_t = -\frac{\beta_{0,t-1}}{2\beta_{1,t-1}} + \epsilon_{1,t}$$

where $\epsilon_{1,t}$ is i.i.d. with $\mathbb{E}\epsilon_{1,t} = 0$ and $\mathbb{E}\epsilon_{1,t}^2 = \sigma_1^2$. To be concrete, we choose $\epsilon_{1,t}$ from a uniform distribution over $[-\bar{\epsilon}, \bar{\epsilon}]$.[23] We regard $\bar{\epsilon}$ as the size of exploration by the decision maker.

The algorithm forecasts the *expected* demand

$$\varphi_q\left(\mathcal{A}_d(\mathcal{O}_t)\right) = \frac{\beta_{0,t}}{2}.$$

[22] The difference from the "orthodox" least square is R_{t-1} in (3.19).

[23] The details of the distribution of $\epsilon_{1,t}$ do not affect the paper's main conclusion, as long as $\epsilon_{1,t}$ has the second moment $\sigma_1^2 < \infty$.

Based on the estimated linear demand (3.17), the monopolist forecasts that the sales quantity will be

$$\beta_{0,t-1} + \beta_{1,t-1}\left[-\frac{\beta_{0,t-1}}{2\beta_{1,t-1}} + \epsilon_{1,t}\right] = \frac{\beta_{0,t-1}}{2} + \beta_{1,t-1}\epsilon_{1,t}$$

but the actual demand in period t is

$$q_t = 1 - F\left(-\frac{\beta_{0,t-1}}{2\beta_{1,t-1}} + \epsilon_{1,t}\right) + \epsilon_{2,t}.$$

Thus, the forecasting error is

$$\phi(\beta_{t-1}, \epsilon_t) = 1 - F\left(-\frac{\beta_{0,t-1}}{2\beta_{1,t-1}} + \epsilon_{1,t}\right) + \epsilon_{2,t} - \frac{\beta_{0,t-1}}{2} - \beta_{1,t-1}\epsilon_{1,t}$$

where $\beta_{t-1} = (\beta_{0,t-1}, \beta_{1,t-1})$ and $\epsilon_t = (\epsilon_{1,t}, \epsilon_{2,t})$.

Since the actual demand must be in the closed interval $[0, 1]$, we first take care of the cases of "corner solution" before moving to "interior solution." If actual quantity q_t is at the boundary of $[0, 1]$, we directly update $(\beta_{0,t-1}, \beta_{1,t-1})$.

Fix $a_t > 0$. If $q_t = 0$, then the algorithm concludes that the forecast price was too high and adjusts accordingly:

$$\beta_{0,t} = \beta_{0,t-1} - a_t \quad \text{and} \quad \beta_{1,t} = \beta_{1,t-1} \leq 0$$

so that

$$b_{t+1} = -\frac{\beta_{0,t}}{2\beta_{1,t}} = -\frac{\beta_{0,t-1}}{2\beta_{1,t-1}} + \frac{a_t}{2\beta_{1,t-1}} = b_t + \frac{a_t}{2\beta_{1,t-1}} < b_t,$$

Similarly, if $q_t = 1$,

$$\beta_{0,t} = \beta_{0,t-1} + a_t \quad \text{and} \quad \beta_{1,t} = \beta_{1,t-1} \leq 0$$

so that

$$b_{t+1} = b_t - \frac{a_t}{2\beta_{1,t-1}} > b_t.$$

If $0 < q_t < 1$, then

$$\begin{bmatrix}\beta_{0,t}\\ \beta_{1,t}\end{bmatrix} = \begin{bmatrix}\beta_{0,t-1}\\ \beta_{1,t-1}\end{bmatrix} + a_t R_{t-1}^{-1}\begin{bmatrix}1\\ -\frac{\beta_{0,t-1}}{2\beta_{1,t-1}} + \epsilon_{1,t}\end{bmatrix}\phi(\beta_{t-1}, \epsilon_t) \tag{3.18}$$

where

$$R_{t-1} = \begin{bmatrix}1 & -\frac{\beta_{0,t-1}}{2\beta_{1,t-1}}\\ -\frac{\beta_{0,t-1}}{2\beta_{1,t-1}} & \left(-\frac{\beta_{0,t-1}}{2\beta_{1,t-1}}\right)^2 + \sigma_1^2\end{bmatrix}. \tag{3.19}$$

Since the monopolist designs the size of the experiments, the variance σ_1^2 of $\epsilon_{1,t}$ is a known parameter.[24] Since $a_t > 0$ controls the amount of change of the estimator $\beta_t - \beta_{t-1}$, we call $a_t > 0$ the gain function.

[24]The covariance matrix R_{t-1} is known to the seller because the seller knows the mean and the variance of the price in period t. Therefore, the algorithm has to keep track of 2 estimators: $\beta_{0,t}, \beta_{1,t}$. The algorithm differs from the recursive least square estimation algorithm, where the independent variable's mean and variance (i.e., price) must be estimated. The recursive least square estimation algorithm has to keep track of 4 estimators: $\beta_{0,t}, \beta_{1,t}$ along with the prices' mean and variance.

We need to impose a bound on $(\beta_{0,t}, \beta_{1,t})$ to keep the estimator within a compact set. Let $\mathcal{B}$ be a compact convex set that contains B in the interior of $\mathcal{B}$ so that the Hausdorff distance between $\mathcal{B}$ and B is positive. If $(\beta_{0,t}, \beta_{1,t}) \notin \mathcal{B}$, then the seller can conclude that the estimator is out of the line and needs to adjust the estimator by pushing it back to B.[25]

We modify the baseline updating scheme to construct the formal updating scheme for $(\beta_{0,t}, \beta_{1,t})$:

$$\begin{bmatrix}\beta_{0,t}\\ \beta_{1,t}\end{bmatrix} = \begin{bmatrix}\beta_{0,t-1}\\ \beta_{1,t-1}\end{bmatrix} + a_t R_{t-1}^{-1} \begin{bmatrix}1\\ -\frac{\beta_{0,t-1}}{2\beta_{1,t-1}} + \epsilon_{1,t}\end{bmatrix} \phi(\beta_{t-1}, \epsilon_t) \tag{3.20}$$

if the right-hand side is in $\mathcal{B}$. Otherwise, $(\beta_{0,t}, \beta_{1,t}) = (\beta_0, \beta_1) \in \mathsf{B}$ for some fixed (β_0, β_1) in the interior of B. To simplify notation, we write

$$\psi_{t-1} \equiv \psi(\beta_{t-1}, \epsilon_t) = R_{t-1}^{-1} \begin{bmatrix}1\\ -\frac{\beta_{0,t-1}}{2\beta_{1,t-1}} + \epsilon_{1,t}\end{bmatrix} \phi(\beta_{t-1}, \epsilon_t). \tag{3.21}$$

Treating the estimated demand curve

$$q = \beta_{0,t-1} + \beta_{1,t-1} p$$

as the actual demand curve, the seller sets the price

$$p_t = -\frac{\beta_{0,t-1}}{2\beta_{1,t-1}} + \epsilon_{1,t}$$

Given p_t, the quantity in period t

$$q_t = 1 - F(p_t) + \epsilon_{2,t}$$

is realized. Using (q_t, p_t), the seller updates $(\beta_{0,t-1}, \beta_{1,t-1})$ to $(\beta_{0,t}, \beta_{1,t})$. We choose

$$a_t = \frac{1}{(t + t_0)^\omega} \tag{3.22}$$

for some $t_0 > 0$ and $\omega \in (0, 1)$. We choose t_0 to meet the accuracy and confidence requirement. We must choose $\omega < 1$ to satisfy the data complexity requirement on $T(\mu, \lambda)$.

Let $\mathcal{A}_d$ be the recursive algorithm with decreasing gain function $a_t > 0$ for all $t \geq 1$. The output of $\mathcal{A}_d$ is the estimated expected demand:

$$1 - \mathcal{A}_d(\mathcal{O}_t)(p) = \beta_{0,t} + \beta_{1,t} p \qquad \forall p \in [\underline{v}, \overline{v}].$$

From $\mathcal{A}_d$, we can derive forecasts about the optimal price and the expected demand

$$\big(\varphi_p(\mathcal{A}_d(\mathcal{O}_t)), \varphi_q(\mathcal{A}_d(\mathcal{O}_t))\big) = \left(-\frac{\beta_{0,t-1}}{2\beta_{1,t-1}}, \frac{\beta_{0,t-1}}{2}\right) \tag{3.23}$$

Let $D_t = (q_t, p_t)$ be the data in period t. The constructed algorithm is recursive: $\mathcal{A}_d(\mathcal{O}_t) = \beta_t$ is the output of the algorithm based on D_t and $\mathcal{A}_d(\mathcal{O}_{t-1})$. The input complexity

$$\dim(\mathcal{A}_d(\mathcal{O}_{t-1})) = 2 \text{ and } \dim(D_{t-1}) = 2 \qquad \forall t \geq 1,$$

[25]This mapping is known as the projection facility in the literature of the stochastic approximation (Kushner and Yin (1997)).

and $\dim(\Omega_t) = 0$, since the algorithm does not track any internal state variables. Thus, $\text{comp}(\mathcal{A}_d) = 4$.

4. Heuristics

We are interested in the asymptotic properties of the right tail $\{\beta_t\}_{t=t_K}^{\infty}$ of the sample path of β_t from t_K where $t_K \to \infty$ as $K \to \infty$. Using the recursive nature of the algorithm, we can write

$$\beta_{t+t_K} - \beta_{t_K} = \sum_{s=t_K}^{t+t_K} a_s \psi_s$$

where ψ_s is defined in (3.21), which implies

$$\beta_{t+t_K} - \beta_{t_K} = \sum_{s=t_K}^{t+t_K} a_s \psi_s = \sum_{s=t_K}^{t+t_K} a_s \mathbb{E}_{s-1}\psi_s + \sum_{s=t_K}^{t+t_K} a_s \left(\psi_s - \mathbb{E}_{s-1}\psi_s\right) \tag{4.24}$$

where ψ_s is defined in (3.21). Our proof will analyze this expression in detail. Define

$$\xi_s = \psi_s - \mathbb{E}_{s-1}\psi_s \tag{4.25}$$

which is a martingale difference. Under our assumptions, ξ_s is bounded.

4.1. Tracking the Mean.

Following the same analytic method as in conventional learning models (Evans and Honkapohja (2001)), we can show convergence "pointwise" for F, allowing the amount of data needed to achieve the desired level of accuracy to depend on F. We examine the first term in (4.24) in the limit:

$$\lim_{K\to\infty} \lim_{t\to\infty} \frac{1}{\sum_{s=t_K}^{t} a_s} \sum_{s=t_K}^{t+t_K} a_s \mathbb{E}_{s-1}\psi_s \tag{4.26}$$

with $\beta_{t_K} = \beta$. Following the stochastic approximation technique (Kushner and Yin (1997)),[26] the evolution of $\beta_{t+t_K} - \beta_t$ can be approximated by the trajectory of

$$\dot{\beta} = \mathbb{E}\left[\psi_t \mid \beta_t = \beta\right].$$

To derive the formula for the right-hand side, note that our algorithm is a simplified version of the least square estimation algorithm. Let us fix $(\beta_{0,t}, \beta_{1,t}) = (\beta_0, \beta_1) \equiv \beta$ and calculate the expected value of the estimator in the "next period" if the monopolist chooses the estimator to minimize the forecasting error. Given (β_0, β_1), the next period's quantity q' and price p' are

$$(p', q') = \begin{cases} \left(b + \epsilon, 1 - F\left(b + \epsilon\right)\right) & \text{with probability } 0.5 \\ \left(b - \epsilon, 1 - F\left(b - \epsilon\right)\right) & \text{with probability } 0.5. \end{cases}$$

where

$$b = -\frac{\beta_0}{2\beta_1} \tag{4.27}$$

[26] The same method is widely used in the conventional learning literature (e.g., Marcet and Sargent (1989), Evans and Honkapohja (2001)).

is the best response recommended by the algorithm. To fit the observed data best, the monopolist chooses, on average, the new coefficients (β_0', β_1') passing through $\left(b+\epsilon, 1-F\left(b+\epsilon\right)\right)$ and $\left(b-\epsilon, 1-F\left(b-\epsilon\right)\right)$. A simple calculation shows

$$\begin{aligned}\beta_0' &= 1-F\left(b\right)+bf(b)\\ \beta_1' &= -f(b)\end{aligned}$$

modulo linear approximation error at the order of σ_1^2. We choose $\bar{\epsilon}$ (or the size of σ_1^2) to meet the accuracy requirement. Results from stochastic approximation theory[27] shows that the asymptotic properties of the mean of $(\beta_{0,t}, \beta_{1,t})$ are dictated by the dynamic properties of the associated ordinary differential equation (ODE)

$$\begin{aligned}\dot{\beta}_0 &= \beta_0' - \beta_0 = 1-F(b)+bf(b)-\beta_0\\ \dot{\beta}_1 &= \beta_1' - \beta_1 = -f(b)-\beta_1.\end{aligned} \tag{4.28}$$

Since (4.27) holds in every period, we take the time derivative on both sides of the equality to have

$$\dot{b} = -\frac{1}{2\beta_1}\left(\dot{\beta}_1 + 2b\dot{\beta}_0\right).$$

After substituting $\dot{\beta}_1$ and $\dot{\beta}_0$ by (4.28), we have

$$\dot{b} = -\frac{f(b)}{2\beta_1}\left[\frac{1-F(b)}{f(b)} - b\right]. \tag{4.29}$$

Since the demand curve $1-F(p)$ is strictly decreasing, $\beta_1 < 0$. Thus, $-\frac{f(b)}{2\beta_1} > 0$. The term inside the bracket has a unique solution $b^*(F)$, the profit-maximizing price for (actual) distribution F. By the increasing hazard rate property, the term in the bracket is strictly decreasing for b, which makes $b^*(F)$ a stable stationary solution of (4.29). The stochastic approximation implies that the least square learning algorithm converges to $b^*(F)$ (Kushner and Yin (1997)).

4.2. **Uniform Convergence.** We must do additional work to prove uniform convergence over $\mathcal{F}^\eta$. We must verify that we can choose a uniform bound for the number of observations to achieve the desired level of accuracy and that the required number of observations increases at a polynomial rate as the accuracy requirement tightens. Since $\frac{1-F(b^*(F))}{f(b^*(F))} - b^*(F) = 0$,

$$\frac{1-F(b)}{f(b)} - b = \frac{1-F(b)}{f(b)} - b - \left(\frac{1-F(b^*(F))}{f(b^*(F))} - b^*(F)\right).$$

By the increasing hazard rate property,

$$\frac{1-F(b)}{f(b)} - b - \left(\frac{1-F(b^*(F))}{f(b^*(F))} - b^*(F)\right) \leq -(b-b^*(F))$$

if $b > b^*(F)$, in which case we can show that there exists $c > 0$ such that[28]

$$\dot{b} \leq -c(b-b^*(F)) < 0.$$

[27] See Kushner and Yin (1997) for details.

[28] If we make a stronger assumption that $\inf_p f(p) > 0$ uniformly, the proof is straightforward.

Similarly, if $b < b^*(F)$, then

$$\dot{b} \geq -c(b - b^*(F)) > 0$$

for any $F \in \mathcal{F}^\eta$.

Since $b^*(F) \leq [\underline{v}, \overline{v}]$ for all $F \in \mathcal{F}^\eta$, the initial condition of the ordinary differential equation can be selected from a compact set. The distance $|b - b^*(F))|$ vanishes uniformly at the order of $e^{-c\tau}$. Note that the right hand side of (4.26) determines how "quickly" $|b - b^*(F))|$ can vanish. If we choose smaller $a_s > 0$, it takes more observations to move b to $b^*(F)$. In particular, the number of data points increases exponentially if $a_s \sim \frac{1}{s}$, as in most conventional least square learning algorithms. We need a_s to be larger than $1/s$ so that the required amount of data increases at a polynomial rate while ensuring convergence with probability 1. To this end, we choose $a_s \sim 1/s^\omega$ where $0 < \omega < 1$ (Dupuis and Kushner (1989)).

It might not be obvious where we use the assumption that the density function $f = F'$ is uniformly Lipschitz continuous. We can prove "pointwise" convergence to $b^*(F)$ using only the continuity of density function $f = F'$ combined with the projection facility. However, the number of time steps necessary for (4.29) to converge to a small neighborhood of $b^*(F)$ depends upon the speed of b converging to $b^*(F)$, which in turn depends on $f = F'$. As the rate of changes of f, $\sup_{p \neq p'} \left| \frac{f(p) - f(p')}{p - p'} \right|$, becomes larger, we need more data to approximate the sample path of b_t by the trajectory of the ordinary differential equation (4.29). In order to prove uniform convergence, we must ensure that (4.29) approximates the sample path accurately uniformly over $\mathcal{F}^\eta$. If the rate of change of the density function $f = F'$ is uniformly bounded, we can find a uniform bound on the amount of data that guarantees the desired level of accuracy of the approximation.[29]

4.3. **Confidence Bound.** To calculate the uniform confidence bound, we need to examine the distribution of

$$\sum_{s=t_K}^{t+t_K} a_s \xi_s \tag{4.30}$$

where ξ_s is defined as (4.25). To simplify notation, let us assume for a moment $t_K = 0$. To satisfy the uniform confidence requirement, we need to find $\rho > 0$ such that

$$\mathbb{P}\left(\exists t \geq T_K, \left| \frac{1}{\sum_{s=t_K}^{t+t_K} a_s} \sum_{s=t_K}^{t+t_K} a_s \xi_s \right| > \mu \right) \leq e^{-\rho T_K}$$

holds uniformly for $F \in \mathcal{F}^\eta$. This part of the exercise calculates the tail portion of the probability distribution of (4.30). For a fixed F, the existence of $\rho > 0$ can be proved by the large deviation properties (Dembo and Zeitouni (1998)) of a recursive algorithm (Dupuis and Kushner (1989)). Our exercise is more challenging because we are searching for $\rho > 0$ *uniformly* over the set of feasible distributions.

The algorithm of Cole and Roughgarden (2014) uses the buyers' valuation, drawn independently from the same distribution. In that case, we could invoke the large deviation property of the IID sample average, such as Hoeffding's inequality, to prove that the tail probability vanishes at the exponential rate uniformly over $\mathcal{F}^\eta$.

[29] See Section A.4.

Our algorithm uses (q_t, p_t), which is not IID (or even a martingale). The data-generating process is endogenous, making the stochastic process (q_t, p_t) non-stationary. As a result, ξ_t is not IID but a bounded martingale difference. We need the full power of the concentration theorem to invoke the Azuma-Hoeffding-Bennett inequality (Dembo and Zeitouni (1998)) to calculate the uniform exponential rate for all feasible distributions of buyer's valuation,[30] which proves that our algorithm is efficient (Shalev-Shwartz and Ben-David (2014)).

5. Concluding Remarks

The basic idea of our analysis is to approximate locally a general concave objective function by a quadratic function. The linear quadratic approximation has been widely used in macroeconomics literature (e.g., Benigno and Woodford (2012)) to solve a broad class of optimization problems. Our innovation over the existing literature is to find a mild condition under which the decision maker can find the optimal solution even if he has only imprecise objective knowledge about the objective function. If so, we construct a simplest algorithm to find the optimal solution with a reasonable amount of data. We can identify the key restrictions to show how to apply our analytic method to a general problem.

Let us illustrate a simple generic optimization problem where the decision maker has to choose an optimal solution p^e from a convex compact subset of $\mathcal{R}$, possibly in a randomized fashion. Define

$$\psi : K \times X \to \mathbb{R}$$

as the outcome function of interest where X is the state space endowed with a probability distribution. The decision maker can observe the realized value of $\psi(p, x)$ ex-post but does not know the functional form of ψ. In the monopoly problem, ψ is determined by the demand curve, and W is the profit function. In a dynamic macroeconomic model, ψ can be a state transition rule in equilibrium (Marimon and Scott (1999)).

The goal of the decision-maker is to find an optimal solution p^e that solves

$$\max_{p \in K} \mathbb{E} W(\psi(p), p).$$

We assume that W and ψ are smooth functions to invoke the first-order analysis. We often impose additional conditions to ensure the existence of a unique interior solution.

(1) $\exists! b^e \in K$ such that the first-order condition holds.

$$\mathbb{E} W_1 \psi'(b^e) + W_2 = 0 \tag{5.31}$$

(2) At b^e, the second order condition holds with a strict inequality.

$$\mathbb{E} W_{11}(\psi')^2 + W_1 \psi'' + W_{12}(\psi' + 1) < 0 \tag{5.32}$$

We use the virtual utility function (Myerson (1981)) to write down the first-order condition in a static monopoly problem. In a dynamic problem, (5.31) appears as the Euler equation (Marimon and Scott (1999)). If (5.31) and (5.32) hold, we say that the first-order approach is valid.

[30] Fudenberg, Lanzani, and Strack (2023) obtained the uniform upper bound of the confidence in the Bayesian learning dynamics where the data generating process is exogenous. In our case, the data generating process is endogenous.

Let us write the optimal solution

$$b(\psi) = \arg\max_{p\in K} \mathbb{E}W(\psi(p), p).$$

obtained by solving (5.31). The increasing hazard rate property of the distribution of buyer's valuation is a classic condition to satisfy (5.31) and (5.32). Suppose W is the value function in a dynamic optimization problem. In that case, we have conditions under which the optimal solution can be derived from the Euler equation corresponding to (5.31). However, the details of the conditions are specific to the problems at hand.

The decision maker knows W, but a complete specification of outcome function ψ is unknown to the decision maker. Because ψ is generally non-linear, its calculation is challenging. Because the data-generating process is endogenous, we cannot use the law of large numbers to infer ψ.

The monopolist in our exercise behaves like a computational economist. We often replace ψ with more computationally manageable functions (called base functions) such as polynomials, although the selection of function classes is tailored to the specific task. The fundamental difference is that the choice of base functions in our exercise is potentially endogenously determined by the sequence of outcomes rather than exogenously by the modeler. The remaining question concerns the conditions under which the decision maker can learn (or calculate) ψ to learn the equilibrium outcome. For illustration, let S^ℓ be the collection of ℓ-st order polynomial. Let $\beta = (\beta_0, \ldots, \beta_\ell)$ be the coefficients of a function in S^ℓ.

Let Ψ be the set of all outcome functions that satisfy (5.31) and (5.32) that validate the first-order approach. Similarly, for all $\alpha > 0$, define Ψ^α as a subset of Ψ where for all $\psi \in \Psi^\alpha$ satisfies a stronger form of (5.32)

$$\mathbb{E}W_{11}(\psi')^2 + W_1\psi'' + W_{12}(\psi' + 1) < -\alpha \tag{5.33}$$

called α regularity following Cole and Roughgarden (2014).

The decision maker chooses a manageable class of functions for ψ to calculate the optimal choice. Since the decision maker knows W, he can verify whether $S^\ell \cap \Psi^\alpha \neq \emptyset$. The choice of Ψ^α must satisfy two conditions. Let $f_\beta(p)$ be a generic element of $S^\ell \cap \Psi^\alpha$ and $\beta = (\beta_0, \ldots, \beta_\ell)$ is a profile of coefficients.

Definition 5.1. *If $S^\ell \cap \Psi^\alpha \neq \emptyset$, we say that S^ℓ validates the first-order approach. Let*

$$b(f_\beta) = \arg\max_{p\in K} \mathbb{E}W(f_\beta(p, x), p)$$

be the optimal decision induced by f_β. If for all $b(\psi)$, $\exists f_\beta \in S^\ell \cap \Psi^\alpha$ such that $b(\psi) = b(f_\beta)$, we say that S^ℓ emulates the true outcome function.

Define ℓ^* as the smallest integer where S^{ℓ^*} validates the first-order approach and emulates the true outcome function. Since the decision maker only knows $\psi \in \Psi$, he has to find the "correct" β satisfying $b(\psi) = b(f_\beta)$. We are searching for a condition on S^{ℓ^*} and an algorithm that leads to the "correct" β.

Suppose that f_β satisfies the following condition:

$$\forall \psi \in \Psi^\alpha, \exists f_\beta \in S^{\ell^*} \cap \Psi,\ \psi(p) = f_\beta(p) \text{ and } \frac{\partial \psi(p)}{\partial p} = \frac{\partial f_\beta(p)}{\partial p}. \tag{5.34}$$

That is, f_β is sufficiently flexible to match the 0-th and 1-st derivatives of $\psi(p)$. If ψ is a demand curve, a linear demand curve can match the actual demand and the slope at p. If ψ is a deterministic function, having a smooth function with two parameters to match the first two derivatives of $\psi(p)$ suffices.

Let β_{t-1} be the least square estimator at the beginning of period t. The decision-maker solves

$$b_t(f_{\beta_{t-1}}) = \arg\max_{p\in K} \mathbb{E}W(f_{\beta_{t-1}}(p), p)$$

and chooses

$$p_t = b_t(f_{\beta_{t-1}}) + \epsilon_t$$

where ϵ_t is an i.i.d. white noise with small variance, added to allow the decision maker to experiment. The outcome at the end of period t includes $(p_t, \psi(p_t))$, which is used to update β_{t-1} to β_t solving

$$\min_\beta \frac{1}{t}\sum_{k=1}^{t}(\psi(p_k) - f_\beta(p_k))^2$$

recursively.

Because the data-generating process is endogenous, we need conditions to ensure that β_t converges. The conditions for the convergence of the estimation process tend to be specific to the problems (cf. Marcet and Sargent (1989)). Still, we can state a general result in the limit.

Proposition 5.2. *Suppose that* (5.34) *condition holds for* $\Psi^\alpha \cap S^{\ell^*}$ *for some* $\alpha > 0$. *If* $\beta_t \to \beta$, $b(f_{\beta_t}) \to b(\psi)$ *uniformly over* $\Psi^\alpha \cap S^{\ell^*}$.

Proof. The proof is essentially identical with the proof of Theorem 2.8. We sketch how each condition is used to derive the result. The estimator for β is obtained by solving

$$\min_\beta \mathbb{E}(\psi(p) - f_\beta(p))^2$$

recursively. The first-order condition is

$$\mathbb{E}(\psi(p) - f_\beta(p))\frac{\partial f_\beta}{\partial \beta} = 0.$$

Since $p_t = b(f_{\beta_{t-1}}) + \epsilon_t$,

$$\psi(p_t) - f_{\beta_{t-1}}(p_t) = \psi(b_t) + \psi'(b_t)\epsilon_t + \frac{1}{2}\psi''(b_t)\epsilon_t^2 - \left(f_{\beta_{t-1}}(b_t) + f'_{\beta_{t-1}}(b_t)\epsilon_t + \frac{1}{2}f''_{\beta_{t-1}}(b_t)\epsilon_t^2\right) + \mathsf{O}(\epsilon_t^3).$$

Since $f \in S^{\ell^*} \cap \Psi^\alpha$ can emulate ψ, the optimal solution β_t must satisfy

$$\mathbb{E}[\psi(b_t) - f_{\beta_t}(b_t)] = 0 \text{ and } \mathbb{E}[\psi'(b_t) - f'_{\beta_t}(b_t)] = 0.$$

Since the first-order approach is valid, $b^e = b(\psi)$ is a unique solution of

$$\mathbb{E}W_1(\psi(b^e), b^e)\psi'(b^e) + W_2(\psi(b^e), b^e) = 0.$$

Similarly, $f_{\beta_t} \in S^{\ell^*} \cap \Psi^\alpha$ validates the first-order approach so that $b_t^* = b(f_{\beta_t})$ is the unique solution of

$$\mathbb{E}W_1(f_{\beta_t}(b_t^*), b_t^*)f'(b_t^*) + W_2(f_{\beta_t}(b_t^*), b_t^*) = 0.$$

Since f_{β_t} matches the 0-th and 1-st derivatives of ψ, $\beta_t \to \beta$ implies $b(f_\beta) = b(\psi)$. The convergence can be made uniform over $S^{\ell^*} \cap \Psi^\alpha$ for a fixed $\alpha > 0$. □

Appendix A. Proof of Theorem 2.8

A.1. **Outline of Proof.** We first translate the problem to the forecasting problem. Instead of calculating the long-run discounted average payoff, let us consider a forecasting problem where the monopolist's algorithm forecasts the pair $(b^*(F), 1 - F(b^*(F)))$ of the profit-maximizing price and the expected demand at the profit-maximizing price at the unknown F. Inspired by PAC learnability in computer science, let us consider a uniform learnability condition.

Definition A.1. *$\mathcal{A}$ uniformly learns $\mathcal{F}^\eta$ if $\forall \lambda \in (0,1)$, $\forall \mu > 0$, $\exists \overline{T}(\mu, \lambda)$ such that*

$$\mathbb{P}\left(\exists t \geq \overline{T}(\mu,\lambda),\ \left|(\varphi_p(\mathcal{A}(\mathcal{O}_t)), \varphi_q(\mathcal{A}(\mathcal{O}_t))) - (b^*(F), 1 - b^*(F))\right| \geq \mu\right) \leq \lambda \tag{A.35}$$

where

$$\overline{T}(\mu,\lambda) \sim \mathsf{O}\left(\frac{1}{\mu^p} \log \frac{1}{\lambda}\right) \tag{A.36}$$

for some $p > 0$.

The substance of the definition is the uniformity of $\overline{T}(\mu,\lambda)$ over $\mathcal{F}^\eta$. That is, by $\overline{T}(\mu,\lambda)$ time steps, the forecast of $\mathcal{A}$ must be within μ neighborhood of the actual profit maximizing outcome uniformly for any $F \in \mathcal{F}^\eta$.

If we show that $\mathcal{A}$ uniformly learns $\mathcal{F}^\eta$, the ϵ dominance of $\mathcal{A}$ can be established easily, proving the second part of the theorem.

Proposition A.2. *Suppose that $\mathcal{A}$ uniformly learns $\mathcal{F}^\eta$ and that the monopolist discounts future payoffs. Then, $\forall \epsilon > 0$, $\exists \underline{\delta}$ such that $\forall \delta \in (\underline{\delta}, 1)$, $\mathcal{A}$ is an ϵ dominant algorithm.*

Proof. For fixed $\lambda > 0$ less than 1 and small $\mu > 0$, we know the existence of $\rho > 0$. Thus, we can choose $\overline{T}$ so that

$$e^{-\rho \overline{T}} = \lambda$$

and therefore,

$$\overline{T} = -\frac{\log \lambda}{\rho}.$$

$\overline{T}$ increases at the logarithmic speed with respect to $1/\lambda$, which is independent of $F \in \mathcal{F}^\eta$.

It remains to show that the expected discounted average payoff converges uniformly to $b^*(F)(1 - F(b^*(F)))$ as the discount factor converges to 1. Fix $\epsilon > 0$. We can then choose $\mu > 0$ and $\sigma_1^2 > 0$ such that

$$\mathbb{E} p_t q_t \geq b^*(F)(1 - F(b^*(F))) - \frac{\epsilon}{2} \qquad \forall t \geq T.$$

For a given $\mu > 0$, choose $\overline{T}$ accordingly. We have

$$\begin{aligned}
& \left|\mathbb{E}(1-\delta)\sum_{t=1}^{\infty} p_t q_t \delta^{t-1} - b^*(F)(1 - F(b^*(F)))\right| \\
= & \left|\mathbb{E}(1-\delta)\left[\sum_{t=1}^{T}(p_t q_t - b^*(F)(1 - F(b^*(F))))\delta^{t-1} - \sum_{t=T+1}^{\infty}(p_t q_t - b^*(F)(1 - F(b^*(F))))\delta^{t-1}\right]\right| \\
\leq & (1-\delta^T) b^*(F)(1 - F(b^*(F))) + \delta^T \frac{\epsilon}{2} \\
\leq & \epsilon
\end{aligned}$$

if $\delta \to 1$, since $\overline{T}$ is independent of δ. □

We prove that algorithm $\mathcal{A}_d$ constructed in Section 3 uniformly learns $\mathcal{F}^\eta$ from section A.2. Then, section A.6 proves that any simpler algorithm cannot be an ϵ dominant algorithm.

A.2. **Preliminaries.** The projection facility is only used to ensure the tightness of the set of the sample paths. It does not alter the asymptotic properties, such as the stability and the large deviation properties of the algorithm (Dupuis and Kushner (1989)). Thanks to the projection facility, we can assume that $(\beta_{0,t}, \beta_{1,t})$ is contained in a compact convex set. For the remainder of the paper, we suppress the projection facility to simplify the exposition when we examine the asymptotic properties of the algorithm.

Following Kushner and Yin (1997), we construct the (fictitious) time from the gain function $a_t = \frac{1}{(t+t_0)^\omega}$ where t_0 is selected to satisfy the accuracy requirement. Since $0 < \omega < 1$, $\sum_{t=1}^{\infty} a_t = \infty$. Thus, $\forall \tau > 0$, there is a unique K such that

$$K = \inf\{T \mid \sum_{t=1}^{T} a_t \geq \tau\}.$$

Define a mapping

$$m : \mathbb{R}_+ \to \{1, 2, 3, \ldots\}$$

where

$$m(\tau) = K$$

as defined above. We refer to $\mathbb{R}_+$ as the fictitious time, the clock time, or simply, the time. We call $\{1, 2, 3, \ldots\}$ as the number of periods, rounds or time steps. Given a discrete process $\{\beta_t\}$, define a continuous time process $\beta(\tau)$ for $\tau \geq 0$ through the linear interpolation of the sample path of the discrete time process $\{\beta_t\}$. The next step is to construct the left shift process, obtained from $\beta(t)$ by re-setting the time clock to 0 at each integer time K: $\forall K \in \{1, 2, \ldots\}$ and $\tau > 0$,

$$\beta^K(\tau) = \beta(K + \tau).$$

We have a sequence of $\{\beta^K\}$ continuous sample paths. Define

$$\bar{\beta}(\tau) = \lim_{K \to \infty} \beta^K(\tau)$$

pointwise by taking a convergent subsequence of $\{\beta^K\}$. The existence of a convergent subsequence is implied by the assumptions we imposed on β_t as in Kushner and Yin (1997).

For a fixed $F \in \mathcal{F}^\eta$, we are interested in

$$\lim_{K \to \infty} \beta_{m(K+\tau)} - \beta^*(F) \qquad \forall \tau \geq 0$$

where $\beta^*(F)$ is defined in (3.16). For $t \geq 1$, we can write the recursive formula as

$$\beta_t = \beta_{t-1} + a_t \psi(\beta_{t-1}, p_t, \epsilon_t)$$

where ψ is defined by (3.21). Since the updating term is determined by the old estimate β_{t-1}, the price in period t and the realized quantity, where the last two variables are subject to two shocks $(\epsilon_{1,t}, \epsilon_{2,t})$. Let

$$\psi(\beta_{t-1}, p_t, \epsilon_t) = \mathbb{E}_{t-1} \psi(\beta_{t-1}, p_t, \epsilon_t) + \xi_t$$

where ξ_t is the martingale difference. Since $\beta_t \in \mathsf{B}$ which is compact, ξ_t is uniformly bounded: $\exists \xi > 0$ such that

$$|\xi_t| \leq \xi.$$

Define

$$\overline{b}_{t-1}(\beta_{t-1}) = \mathbb{E}_{t-1} \psi(\beta_{t-1}, p_t, \epsilon_t).$$

The functional form of b_{t-1} is not affected by $t-1$ and is a Lipschitz continuous function of β_{t-1}. To simplify notation, we write $\overline{b}(\beta_{t-1})$ in place of $\overline{b}_{t-1}(\beta_{t-1})$, dropping the time subscript from $\overline{b}_{t-1}$. We can write the recursive formula as

$$\beta_t = \beta_{t-1} + a_t \left[\overline{b}(\beta_{t-1}) + \xi_t \right].$$

Define $\beta^*(F) = (\beta_0^*(F), \beta_1^*(F))$ as the intercept and the slope of a linear demand curve that generates the optimal price $b^*(F)$ and the expected quantity $1 - F(b^*(F))$, that solves

$$1 - F(b^*(F)) = \frac{\beta_0^*(F)}{2} \quad \text{and} \quad f(b^*(F)) = -\beta_1^*(F).$$

The Lipschitz continuity and the increasing hazard rate property guarantee the unique existence of $\beta^*(F)$.

We can write

$$\begin{aligned}\beta_{m(K+\tau)} - \beta^*(F) &= \beta^K(0) + \sum_{t=t_K}^{m(K+\tau)} a_t \bar{b}(\beta_{t-1}) + \sum_{t=t_K}^{m(K+\tau)} a_t \xi_t - \beta^*(F) \\ &= \beta^K(0) + \int_0^\tau \bar{b}(\beta(s))ds - \beta^*(F) \qquad \text{(A.37)} \\ &\quad + \sum_{t=t_K}^{m(K+\tau)} a_t \bar{b}(\beta_{t-1}) - \int_0^\tau \bar{b}(\beta(s))ds \qquad \text{(A.38)} \\ &\quad + \sum_{t=t_K}^{m(K+\tau)} a_t \xi_t \qquad \text{(A.39)}\end{aligned}$$

We examine (A.37), (A.38), and (A.39) one by one.

A.3. **Convergence and Stability.** Following Kushner and Yin (1997), we can write

$$\bar{\beta}(\tau) = \bar{\beta}(0) + \int_0^\tau \bar{b}(\bar{\beta}(s))ds$$

which is written concisely as

$$\dot{\bar{\beta}} = \bar{b}(\beta).$$

To simplify notation, we write

$$b_t = -\frac{\beta_{0,t-1}}{2\beta_{1,t-1}} \qquad \forall t \geq 1$$

and $b(\tau)$ as the continuous process constructed from b_t via linear interpolation $\forall \tau \geq 0$. If the context makes the meaning clear, we drop τ to write b instead of $b(\tau)$. The same convention applies to all other variables, such as $\beta_{0,t}$ and $\beta_{1,t}$.

We examine the properties of the ordinary differential equation (ODE):

$$\dot{\bar{\beta}} = \mathbb{E}\psi = R^{-1}\mathbb{E}\begin{bmatrix} 1 \\ -b_t + \epsilon_{1,t} \end{bmatrix} \phi(\beta_{t-1}, \epsilon_{1,t})$$

where

$$R = \begin{bmatrix} 1 & b_t \\ b_t & b_t^2 + \sigma_1^2 \end{bmatrix}$$

and

$$\phi(\beta_{t-1}, \epsilon_t) = 1 - F\left(b_t + \epsilon_{1,t}\right) + \epsilon_{2,t} - \beta_{0,t-1} - \beta_{1,t-1} b_t - \beta_{1,t-1}\epsilon_{1,t}.$$

By the definition of b,

$$\beta_0 + 2\beta_1 b = 0$$

at every moment. Thus,

$$\dot{\bar{\beta}}_0 + 2\dot{\bar{\beta}}_1 b + 2\bar{\beta}_1 \dot{b} = 0.$$

After substituting $\dot{\bar{\beta}}_0$ and $\dot{\bar{\beta}}_1$, we have

$$\dot{b} = -\frac{f(b)}{2\beta_1}\left[\frac{1 - F(b)}{f(b)} - b\right] \equiv \mathsf{R}(b) \qquad \text{(A.40)}$$

modulo linear approximation errors $\mathsf{O}(\epsilon)$.

Following the convergence theorem in Kushner and Yin (1997), we conclude that b_t converges to $b^*(F)$ in probability: $\forall \mu > 0$, $\forall \lambda > 0$, $\exists T(\mu, \lambda, F)$ such that

$$\mathbb{P}\left(\exists t \geq T(\mu, \lambda, F),\ |b_t - b^*(F)| \geq \mu\right) \leq \lambda.$$

To show the uniform convergence over $F \in \mathcal{F}^\eta$, we have to show that the number of periods to achieve the desired level of accuracy is uniform over $F \in \mathcal{F}^\eta$ and that the desired confidence level can be achieved at the exponential speed uniformly over F.

We can show that for any initial value $b(0) \in [\underline{v}, \overline{v}]$,

$$\left|b(\tau) - b^*(F)\right| \leq e^{-c\tau} \left|b(0) - b^*(F)\right| \leq e^{-c\tau}(\overline{v} - \underline{v}).$$

Let $\tau(\mu)$ be the first time when the trajectory of (A.40) enters the μ neighborhood of $b^*(F)$,

$$\left|b(\tau) - b^*(F)\right| \leq \mu.$$

Recall that $(1 - F(b^*(F)), b^*(F)) \in K \ \forall F \in \mathcal{F}$ and K is a compact subset in the interior of $\mathbb{R}^2_+$. Thus,

$$\overline{\tau}(\mu) = \sup_{(\beta_0(0), \beta_1(0)) \in \mathsf{B}} \tau(\mu) < \infty \tag{A.41}$$

and

$$\overline{\tau}(\mu) \sim -\log \mu \tag{A.42}$$

as $\mu \to 0$. Let us choose $\tau = \overline{\tau}(\mu)$.

By definition,

$$\sum_{t=t_K}^{m(K+\tau)} \frac{1}{t^\omega} \simeq \overline{\tau}(\mu).$$

where the left-hand side is approximated by

$$\int_{t_K}^{m(K+\overline{\tau}(\mu))} \frac{1}{s^\omega} ds$$

A simple calculation shows that

$$m(K + \overline{\tau}(\mu)) \sim \mathsf{O}\left(\left(\log \frac{1}{\mu}\right)^{\frac{1}{1-\omega}}\right) \tag{A.43}$$

which implies that the number of time steps to make real time of $\overline{\tau}(\mu)$ increase at the rate of $\left(\log \frac{1}{\mu}\right)^{\frac{1}{1-\omega}}$.

A.4. **Riemann Residual.** Fix $\mu > 0$ and $\tau = \bar{\tau}(\mu)$ defined in (A.41). Let us consider (A.38)

$$\mathbf{R}(t_K, \tau, F) = \sum_{t=t_K}^{m(K+\tau)} a_t \overline{b}(\beta_{t-1}) - \int_0^\tau \overline{b}(\beta(s)) ds$$

which is the Riemann residual. Since f is uniformly Lipschitz over $\mathcal{F}$, $\overline{b}(\beta)$ is uniformly Lipschitz: $\exists \eta' > 0$ such that

$$\left|\overline{b}(\beta) - \overline{b}(\beta')\right| \leq \eta' |\beta - \beta'| \qquad \forall F \in \mathcal{F}.$$

For each subinterval of size $a_t \leq 1/t_K^\omega$, the difference between the discrete value and the integration is at most $\frac{\eta'}{2t_K^2}$. Thus,

$$\mathbf{R}(t_K, \tau, F) \leq \frac{\eta'}{2t_K^{2\omega}} \overline{\tau}(\mu) t_K^\omega = \frac{\eta' \overline{\tau}(\mu)}{2t_K^\omega}.$$

Note that the right-hand side is independent of F. Thus, $\forall \mu > 0$, define

$$\frac{1}{t_K^\omega} = \frac{2\mu}{\overline{\tau}(\mu)\eta'} \tag{A.44}$$

so that $\forall F \in \mathcal{F}$,

$$\mathbf{R}(t_K, \tau, F) \leq \mu. \tag{A.45}$$

Thus,

$$\frac{1}{t_K} = \mathsf{O}\left(\left[\frac{\mu}{\log \frac{1}{\mu}}\right]^{\frac{1}{\omega}}\right)$$

or equivalently,

$$t_K = \mathsf{O}\left(\left[\frac{1}{\mu}\log\frac{1}{\mu}\right]^{\frac{1}{\omega}}\right). \tag{A.46}$$

We choose $t_0 = t_K$ defined by (A.46) to satisfy the accuracy requirement.

A.5. **Lower Bound of Confidence.** Next, we examine (A.39). Let $a_t = \frac{1}{t^\omega}$ where $0 < \omega < 1$. Fix K so that t_K satisfies (A.44). Following Dupuis and Kushner (1989), define the H-functional $\forall(\beta,\alpha) \in \mathbb{R}^2 \times \mathbb{R}^2$, $\forall\omega \in (0,1)$,

$$H^\omega(\beta,\alpha) = \lim_{\tau\to 0}\lim_{K\to\infty}\frac{a_{t_K}}{\tau}\log\mathbb{E}\left(e^{\frac{\alpha\cdot(\beta^K(\tau)-\beta^K(0))}{a_{t_K}}} \mid \beta^K(0) = \beta\right).$$

If the limit is not well defined, we can replace lim by lim sup, as in Dupuis and Kushner (1989). Let $H^1(\beta,\alpha)$ be the H-functional corresponding to $a_t = 1/t$. A simple calculation shows that

$$H^1(\beta,\alpha) = H^\omega(\beta,\alpha) \qquad \forall\omega \in (0,1).$$

Dupuis and Kushner (1989) shows that if the two stochastic processes have the same H functional, the two stochastic processes have the same rate function h. Instead of calculating the H functional for β_t associated with $a_t = 1/(t+t_0)^\omega$, we calculate the H function for β_t associated with $a_t = 1/t$. If $a_t = 1/t$, then β_t is a simple average of the past forecasting errors, which makes it straightforward to apply the concentration inequalities in Chapter 2.4 in Dembo and Zeitouni (1998).

To make the paper self-contained, we write the proof closely following the notation of Dembo and Zeitouni (1998), assuming $a_t = 1/t$. Define

$$Y_t = \frac{\xi_t}{\xi}.$$

By the definition, $|Y_t| \le 1$ and $\mathbb{E}[Y_t \mid I_{t-1}] = 0$. Thus, $\mathbb{E}[Y_t^2 \mid I_{t-1}] \le 1$, where I_{t-1} is the information available at the end of period $t-1$. Define $S_t = \sum_{s=1}^t Y_s$.

Following Dembo and Zeitouni (1998), we first prove following inequality for all $a_t \in [0,1]$ and $\lambda \ge 0$:

Lemma A.3.

$$\left(\frac{e^\lambda + e^{-\lambda}}{2}\right)^{a_t} \ge \frac{e^{\lambda a_t} + e^{-\lambda a_t}}{2} \tag{A.47}$$

Proof. Taking log of both sides, we have it suffices to show:

$$a_t \log\left(\frac{e^\lambda + e^{-\lambda}}{2}\right) \ge \log(e^{\lambda a_t} + e^{-\lambda a_t}) - \log(2).$$

We note that the inequality holds with equality at $a_t = 0$ and $a_t = 1$. The left-hand side is linear in a_t, whereas the second derivative of the right-hand side is non-negative. Therefore, at any $a_t \in [0,1]$, we have the right-hand side of the inequality (which is convex) lower than the left-hand side of the inequality (which is linear), proving the lemma. □

We follow the proof of Corollary 2.4.7 from (Dembo and Zeitouni 1998).[31] We have

$$\mathbb{E}[e^{\lambda S_t}] \le \left(\frac{e^\lambda + e^{-\lambda}}{2}\right)^t. \tag{A.48}$$

Following the proof of Corollary 2.4.7 from from (Dembo and Zeitouni 1998), we have

$$\mathbb{P}\left(\frac{1}{t}|S_t| \ge \mu\right) \le e^{-\lambda\mu t}\mathbb{E}[e^{\lambda S_t}] \le e^{-\lambda\mu t}\left(\frac{e^\lambda + e^{-\lambda}}{2}\right)^t.$$

The first inequality follows from the exponential Chebyshev inequality. The second inequality follows from (A.48).

[31] Set $v = 1$, and applying the previous inequality.

Again following (Dembo and Zeitouni 1998), choose $\mu < 1$ and set $\lambda = \frac{1}{2}\log\left(\frac{1+\mu}{1-\mu}\right)$. After a tedious calculation,

$$\mathbb{P}\left(\frac{1}{t}|S_t| \geq \mu\right) \leq e^{-t\mathsf{H}(\frac{1+\mu}{2}|\frac{1}{2})}. \tag{A.49}$$

where H is the relative entropy of the binomial distribution:

$$\mathsf{H}(p \mid p_0) = p\log\frac{p}{p_0} + (1-p)\log\frac{1-p}{1-p_0} \tag{A.50}$$

for $p, p_0 \in (0,1)$. Set $\mathsf{h} = \mathsf{H}\left(\frac{1+\mu}{2} \mid \frac{1}{2}\right)$. The conclusion follows from the observation that the right-hand side of (A.49) holds uniformly over $F \in \mathcal{F}^\eta$.

We choose the gain function $a_t > 0$ to satisfy the accuracy μ and confidence requirement $1-\lambda$ in (A.35). The data complexity required to achieve the accuracy requirement μ and the confidence bound $1-\lambda$ reveals how the accuracy bound μ and the confidence bound $1-\lambda$ determines the requirement number of observations. We approximate the evolution of the optimal price forecast b_t by a trajectory of an ordinary differential equation. We show that the optimal price forecast converges to the actual optimal price at the rate of $e^{-c\tau}$ for some constant $c > 0$ and $\tau > 0$ where τ is the clock time.

As we decrease $\mu > 0$, the neighborhood of the optimal price $b^*(F)$ decreases. A higher accuracy bound affects the required number of data in two different channels. First, it takes more time for b_t to enter the smaller neighborhood from an initial condition. The algorithm learns efficiently so that the distance between the forecast and the actual optimal price vanishes exponentially. The real time to enter the neighborhood increases at the logarithmic rate of $\log\frac{1}{\mu}$.

The data complexity is measured in terms of the number of time steps instead of the real time length. If $a_t = \frac{1}{(t+t_0)^\omega}$ is larger (i.e., smaller $\omega \in (0,1)$), β_t response more quickly to the forecasting error and b_t evolves faster. The amount of real time τ over t time steps can be approximated by

$$\tau \sim \log\frac{1}{\mu} \sim \sum_{s=t_K}^{t_K+t} a_s \sim \int_{t_K}^{t} \frac{1}{s^\omega} ds \sim \mathsf{O}\left(t^{1-\omega}\right). \tag{A.51}$$

Since τ increases at the logarithmic rate of $\log\frac{1}{\mu}$, the number of time steps increases at the rate of $(\log\frac{1}{\mu})^{\frac{1}{1-\omega}}$.

We incur the approximation error as we approximate the discrete process $\{b_t\}$ by a continuous time process. If ω is small, a_t becomes larger. The sample path of the discrete time process becomes more "jagged." To smooth out the process, we need to make a_t smaller by choosing t_0 larger. As we tighten the accuracy bound, we choose a larger t_0 to reduce the approximation error. A tedious calculation shows that

$$t_0 \sim \mathsf{O}\left(\left(\frac{1}{\mu}\log\frac{1}{\mu}\right)^{\frac{1}{\omega}}\right). \tag{A.52}$$

By combining (A.51) and (A.52), we obtain the terms involving μ. If we choose $\omega = 1$ as in the conventional least square learning algorithm (e.g.,Marcet and Sargent (1989)), the number of time steps required to achieve the desired accuracy and confidence increases exponentially. To keep the data complexity increasing at a polynomial rate, we must choose $\omega < 1$. To ensure convergence with probability 1, we should keep $\omega > 0$ to invoke the convergence result of Dupuis and Kushner (1989).

The distribution of the price and quantity forecast converges to the actual optimal outcome, and the probability of the tail portion of the distribution vanishes at an exponential rate (Dupuis and Kushner (1989)). As $1-\rho$ increases, the demand for the algorithm to reach the higher confidence level increases at $\log\frac{1}{\lambda}$. Following the same logic to translate the real-time to the number of time steps, we conclude that the time steps increase at the rate of $(\log\frac{1}{\lambda})^{\frac{1}{1-\omega}}$.

A.6. **Last Part.** It remains to show that $S^{1,\eta}$ is the simplest class of forecasts that can support an ϵ dominant algorithm for δ sufficiently close to 1. Suppose that $\exists \mathcal{A} : \mathcal{O} \to S^{0,\eta}$ which is an ϵ dominant algorithm. Since an ϵ dominant algorithm can achieve $b^*(F)(1-F(b^*(F))) - \epsilon$, its forecast of the pair of price and quantity must be close to $(b^*(F), 1-F(b^*(F)))$ uniformly over $F \in \mathcal{F}^\eta$.

Thus, $\forall\epsilon$ an ϵ dominant algorithm must forecast the profit maximizing outcome accurately within ϵ bound uniformly: $\forall\epsilon$, $\exists T, \eta, \mu > 0$ such that $\forall t \geq T$, $\forall F \in \mathcal{F}^\eta$

$$\left|(\varphi_p(\mathcal{A}(\mathcal{O}_t)), \varphi_q(\mathcal{A}(\mathcal{O}_t))) - (b^*(F), 1 - F(b^*(F)))\right| < \eta$$

with probability $1 - \mu$ such that

$$\left|\varphi_p(\mathcal{A}(\mathcal{O}_t))\varphi_q(\mathcal{A}(\mathcal{O}_t)) - b^*(F)(1 - F(b^*(F)))\right| < \epsilon \qquad \forall t \geq T.$$

Since $\mathcal{A}(\mathcal{O}_t) \in S^{0,\eta}$, the forecast distribution is degenerate, concentrated at $v^* \in [\underline{v}, \overline{v}]$. Thus,

$$\varphi_p(\mathcal{A}(\mathcal{O}_t) = v^* \text{ and } \varphi_q(\mathcal{A}(\mathcal{O}_t) = 1.$$

Choose any $F \in \mathcal{F}^\eta$ where $1-F(b^*(F)) < 1-\eta$ such as the uniform distribution over $[\underline{v}, \overline{v}]$. The algorithm's forecast does not converge to the true profit maximizing outcome with probability 1, contradicting the hypothesis that the algorithm is an ϵ dominant algorithm.

References

Aghion, P., P. Bolton, C. Harris, and B. Jullien (1991): "Optimal Learning by Experimentation," *The Review of Economic Studies*, 58(4), 621–654.

Battigalli, P., S. Cerreia-Vioglio, F. Maccheroni, M. Marinacci, and T. Sargent (2022): "A Framework for the Analysis of Self-Confirming Policies," *Theory and Decision*, 92, 455–512.

Battigalli, P., A. Francetich, G. Lanzani, and M. Marinacci (2019): "Learning and Self-Confirming Long-Run Biases," *Journal of Economic Theory*, 183, 740–785.

Benigno, P., and M. Woodford (2012): "Linear Quadratic Approximation of Optiaml Policy Problem," *Journal of Economic Theory*, 147, 1–42.

Cerreia-Vioglio, S., F. Maccheroni, M. Marinacci, and L. Montrucchio (2013): "Classical subjective expected utility," *Proceedings of the National Academy of Sciences of the United States of America*, 110(17), 6754–6759.

Cho, I.-K., and K. Kasa (2015): "Learning and Model Validation," *Review of Economic Studies*, 82, 45–82.

Cole, R., and T. Roughgarden (2014): "The Sample Complexity of Revenue Maximization," in *Proceedings of the forty-sixth annual acm symposium on theory of computing*, STOC '14, pp. 243–252, New York, New York. ACM.

Dembo, A., and O. Zeitouni (1998): *Large Deviations Techniques and Applications*. Springer-Verlag, New York, 2nd edn.

Dupuis, P., and H. J. Kushner (1989): "Stochastic Approximation and Large Deviations: Upper Bounds and w.p.1 Convergence," *SIAM Journal of Control and Optimization*, 27, 1108–1135.

Easley, D., and N. M. Kiefer (1988): "Controlling a Stochastic Process with Unknown Parameters," *Econometrica*, 56(5), 1045–1064.

Esponda, I., and D. Pouzo (2016): "Berk–Nash Equilibrium: A Framework for Modeling Agents With Misspecified Models," *Econometrica*, 84(3), 1093–1130.

Evans, G. W., and S. Honkapohja (2001): "Expectations and the Stability Problem for Optimal Monetary Policies," University of Oregon.

Fudenberg, D., G. Lanzani, and P. Strack (2023): "Pathwise Concentration Bounds for Bayesian Beliefs," *Theoretical Economics*, 18, 1585–1622.

Gilboa, I., and D. Schmeidler (1989): "Maxmin Expected Utility with Non-Unique Prior," *Journal of Mathematical Economics*, 18, 141–153.

Hansen, L. P., and T. J. Sargent (2007): *Robustness*. Princeton University Press.

Klibanoff, P., M. Marinacci, and S. Mukerji (2005): "A Smooth Model of Decision Making under Ambiguity," *Econometrica*, 73(6), 1849–1892.

Kushner, H. J., and G. G. Yin (1997): *Stochastic Approximation Algorithms and Applications*. Springer-Verlag.

Lanzani, G. (2023): "Dynamic Concern for Misspecification," MIT.

Maccheroni, F., M. Marinacci, and A. Rustichini (2006): "Ambiguity Aversion, Robustness, and the Variational Representation of Preferences," *Econometrica*, 74(6), 1447–1498.

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.

Marimon, R., and A. Scott (eds.) (1999): *Computational Methods for the Study of Dynamic Economies*. Oxford University Press.

Myerson, R. (1981): "Optimal Auction Design," *Mathematics of Operations Research*, 6, 58–73.

Nyarko, Y. (1991): "Learning in mis-specified models and the possibility of cycles," *Journal of Economic Theory*, 55(2), 416–427.

Olea, J. L. M., P. Ortoleva, M. M. Pai, and A. Prat (2022): "Competing Models," *Quarterly Journal of Economics*, 137(4), 2419–2457.

Osborne, M. J., and A. Rubinstein (1998): "Games with Procedurally Rational Players," *American Economic Review*, 88(4), 834–847.

Rothschild, M. (1974): "A two-armed bandit theory of market pricing," *Journal of Economic Theory*, 9(2), 185–202.

Rubinstein, A. (1986): "Finite Automata Play Repeated Prisoners Dilemma," *Journal of Economic Theory*, 39(1), 83–96.

Rudin, W. (1976): *Principles of Mathematical Analysis*. McGraw Hill.

Rumelhart, D. E., J. L. McClelland, and the PDP Research Group (1986): *Parallel Distributed Processing, I & II* . MIT Press.

Sargent, T. J. (1993): *Bounded Rationality in Macroeconomics*. Clarendon Press.

Shalev-Shwartz, S., and S. Ben-David (2014): *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press.

Wasserman, P. D. (1989): *Neural Computing: Theory and Practice*. Van Nostrand Reinhold, New York.

Weisbuch, G. (1990): *Complex Systems Dynamics: An Introduction to Automata Networks*, Lecture Note Volume II, Santa Fe Institute Studies in the Sciences of Complexity. Addison-Wesley.

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

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