LEARNING UNDERSPECIFIED MODELS

IN-KOO CHO AND JONATHAN LIBGOBER

Abstract. This paper examines whether a decision maker can learn to play an optimal action when he possesses an *underspecified* model. As a laboratory, we choose a monopoly market where the monopolist is initially endowed with a partial specification of the true demand curve. We formulate the learning dynamics as an algorithm that chooses the specification of the demand curve in response to history of outcomes. We represent the monopolist's decision problem as a zero sum game against nature which chooses the true demand curve, while the monopolist chooses an algorithm to maximize the long run average payoff. Instead of the maxmin strategy, the monopolist seeks a *dominant algorithm* that learns the profit maximizing outcome against any true demand. We construct a dominant algorithm that is simplest among all dominant algorithms. For the set of demand curves with strictly decreasing uniformly Lipschitz continuous marginal revenue curve, our algorithm recursively estimates the slope and intercept of a linear demand curve, even if the actual demand curve is not linear. The monopolist chooses a misspecified model to save computational cost while learning the true optimal decision, *uniformly* over the set of demand curves.

Keywords. Monopoly market, Underspecified model, Algorithm, Dominant strategy, PAC guarantee, Complexity Cost

1. Introduction

A rational decision maker is endowed with a correctly specified model of the state and the relationship between his action and consequence (Simon (1987)). Nevertheless, it is an extreme assumption that a decision maker knows *all relevant states* and knows the *complete specification* of the function that maps his strategy to payoffs. This paper supposes that the decision maker possesses a partial specification of the actual environment, knowing only some but not all properties of the data generating process. In such cases, we say that the decision maker is endowed with an *underspecified* model. He faces model uncertainty (Hansen and Sargent (2007)) as multiple models, including the actual model, could be consistent with the partial specification. Our research objective is to understand how a decision maker learns to choose a specification to find the optimal decision under model uncertainty.

We choose monopolistic profit maximization (Myerson (1981)) as our laboratory. Our exercise has four features that differentiate our research from existing work. First, in

Date: December 29, 2022.

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.

contrast to the Bayesian model, the monopolist lacks the information needed to form a prior over the set of demand curves. This kind of model uncertainty motivates the robust approach in Hansen and Sargent (2007) and Bergemann and Morris (2005).

Second, our approach differs from the robust approach to the model uncertainty, which typically does not provide an opportunity for the decision maker to learn from the data (Hansen and Sargent (2007) and Bergemann and Morris (2005)).[1] In our case, the seller can observe data to learn about the demand curve. In conventional learning models, the model specification is exogenously fixed, and the decision maker calculates the unknown parameters. In our exercise, the seller can choose the demand curve's specification and parameters in response to a history of outcomes. We formulate a strategy of the monopolistic seller as an algorithm that maps a history of outcomes to a demand curve forecast.

Third, the robust approach, as well as many problems in computer science, formulates the decision maker's problem as a zero sum game between nature and the decision maker where nature chooses an actual demand curve and the monopolist chooses an algorithm. The existing literature focuses on maxmin solutions, where the decision maker chooses a best response against an adversarial nature who is trying to minimize the decision maker's payoff. Here, we are searching for a dominant strategy of the decision maker in the game against nature, seeking to achieve the best response to *any* choice of nature, ideally with a reasonable amount of data.

Fourth, the monopolist in our exercise bears a computational cost. 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.[2] Following Rubinstein (1986), we assume that the monopolist has a lexicographic preference over the payoff and the complexity cost. The seller first seeks a dominant algorithm, and if multiple dominant algorithms exist, he chooses an algorithm with the minimum complexity.

Finding a dominant algorithm of the monopolist in the game against nature is not trivial. We exploit the convergence criterion developed in the machine learning literature called the PAC (Probabilistically Almost Correct) criterion (Shalev-Shwartz and Ben-David (2014)) to find a dominant algorithm. An algorithm $\mathcal{A}$ PAC guarantees the unknown demand curve if given $\epsilon > 0$ error bound and $1 - \rho > 0$ confidence requirement, we can find a stopping time $T(\epsilon, \rho)$ so that in every period after $T(\epsilon, \rho)$, algorithm $\mathcal{A}$ produces a forecast that is within ϵ neighborhood of the actual optimal strategy with probability $1 - \rho$ uniformly over the set of admissible demand curves. We show that if an algorithm PAC guarantees the unknown demand curve, the algorithm is a dominant strategy.

We show that a dominant algorithm must estimate at least two parameters (Proposition 5.8 and Proposition 10.3). We construct a recursive algorithm estimating a linear demand curve parameterized by two numbers: slope and intercept. The algorithm uses price and quantity as input and does not need to update any internal state variables. We show that the constructed algorithm is the simplest algorithm among those which PAC guarantee the

[1]Exceptions are a series of papers following Hansen and Sargent (2020) and Epstein and Schneider (2007).
[2]This type of complexity measure is consistent with the complexity measure for a neural work (cf. Rumelhart, McClelland, and the PDP Research Group (1986), Wasserman (1989) and Weisbuch (1990)).

optimal price over the set of demand curves with strictly decreasing uniformly Lipschitz continuous marginal revenue curve (Theorem 9.1) within a "reasonable amount" of time.

Conventional learning models (e.g., Evans and Honkapohja (2001), Cho and Kasa (2015) and Cho and Kasa (2017)) provide a good understanding of the conditions under which the forecast of an algorithm converges as well as the large deviation properties of the learning dynamics for a fixed demand curve. A PAC guarantee requires that the convergence and the large deviation properties of the algorithm hold uniformly over the set of all feasible demand curves that nature can choose. The exercise is much simpler than checking the expected payoff of the monopolist against all possible demand curves.

Absent complexity costs, an existing proposal from Cole and Roughgarden (2014) identifies a PAC learnable algorithm that non-parametrically estimates the demand curve from the buyer's reported valuation in the revelation game. In the revelation game, the buyer's private information is freely available to the algorithm. But, our analysis covers a larger class of monopoly markets where the price and the quantity are observed at the end of the game.

We point out a fundamental difference between articulating (a) "demand is of the form $Q(p) = \beta_0 + \beta_1 p$" where p is the price and (β_0, β_1) are the parameters to be estimated, versus (b) "the profit-maximizing price and quantity are the same as those induced by the demand curve $Q(p) = \beta_0 + \beta_1 p$ for some (β_0, β_1)." In the conventional learning model, we impose the decision maker with misspecified models like (a), assuming cognitive bias (Fudenberg, Lanzani, and Strack (2021)) or inability to comprehend complex dynamics (Cho and Kasa (2017)). It is inaccurate to describe the monopolist who chooses (b) as a decision maker endowed with a misspecified model. Under (b), the monopolist is perfectly aware of all feasible demand curves. The monopolist is fully capable of estimating the actual demand curve through a correctly specified model, from which he can calculate the profit maximizing price.

The monopolist's goal is to learn about the profit-maximizing price and the maximum profit, not the demand curve.[3] In our exercise, the choice of the model specification for the demand curve 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 which approximately guarantee the profit maximizing price.

However, the point is not that the monopolist is unaware of other possibilities, as in most learning models with misspecification, but that he would not search for a correct specification if doing so is costly. While a monopolist could use an algorithm calculating parameters of a correctly specified model, such as a non-parametric estimation algorithm (as in Cole and Roughgarden (2014)), 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)).

The rest of the paper is organized as follows. Section 2 reviews the literature, clarifying our contribution beyond the existing literature. Section 3 formally defines and discusses underspecification in depth. In Section 4, we formally describe the problem and define

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

the basic concepts along with examples. Section 5 describes some intermediate results we use to analyze the model. The construction of the algorithm involves technical steps that could obscure the main feature of the algorithm. We first construct a "weaker" algorithm and show that the weaker algorithm satisfies a "weaker" version of PAC guaranteeing property, called uniform learnability. After analyzing the weaker algorithm satisfying uniform learnability, we modify the algorithm to obtain the main result. In Section 6, we construct the simplest algorithm among algorithms that uniformly learn the optimal strategy. We state the main result, whose full proof is in the appendix, but with the heuristic details described in Section 7. Section 8 reports numerical exercises to show that the performance of our algorithm is comparable to a more elaborate algorithm of Cole and Roughgarden (2014). In Section 9, we modify the "weaker" algorithm to construct the optimal algorithm. We discuss some extensions in Section 10. Section 11 concludes the paper.

2. Literature Review

Our paper differs from three main approaches to investigating decision problems with model uncertainty. As discussed above, the most prominent approach is that a decisionmaker chooses his actions assuming uncertainty about the environment resolves in favor of the worst-case (e.g., Hansen and Sargent (2007), Carroll (2015), Carroll (2017), Du (2018) and Libgober and Mu (2021)). The choice that maximizes the objective under the worst conjecture will sometimes differ significantly from the optimal solution against the truth. While data could bring the decisionmaker closer to optimality, these models are silent on how the decisionmaker might do this. In contrast, the monopolist in our model seeks to find a *true* optimal price for an unknown demand curve by observing data.

In the second approach, the modeler completes the specification of the decision problem by imposing a (parametric) model, where the decision maker estimates the parameters using data. While this approach does allow the decisionmaker to learn from data, the specification is fixed exogenously and often excludes the actual mapping from actions into outcomes (e.g., Cho and Kasa (2015), Cho and Kasa (2017), Heidhues, Köszegi, and Strack (2018), Esponda, Pouzo, and Yamamoto (2021), Frick, Iijima, and Ishii (2021) and Fudenberg, Lanzani, and Strack (2021)). In contrast, the monopolist in our model is not committed to a particular specification. The monopolist can use the non-parametric estimation method of the demand curve to avoid the misspecification problem. The monopolist chooses the linear model with two parameters to minimize the computational cost, achieving a PAC guarantee. The linear model is not imposed by the modeler but derived from optimization by the monopolist.

This paper follows a third approach initiated by machine learning models[4] (e.g., Huang, Mansour, and Roughgarden (2018), Cole and Roughgarden (2014), Gonczarowski and Weinberg (2021) and Goncalves and Furtado (2020)). A typical approach in this literature (Cole and Roughgarden (2014)) assumes that the seller non-parametrically estimates the demand curve. This approach avoids the underspecification issue, and the seller can

[4]Nisan, Tardos, and Vazirani (2007) report early applications of the algorithms to game theoretic models, including the monopoly problem.

achieve the true optimal outcome as the estimator converges to the actual demand curve. The analysis then focuses on the consistency of the estimator and data complexity.

Following Cole and Roughgarden (2014), we freely borrow critical concepts developed in the machine learning literature (e.g., Shalev-Shwartz and Ben-David (2014)). We depart from Cole and Roughgarden (2014), however, by considering the complexity of the algorithm and the cost of obtaining data. An essential advantage of the non-parametric estimation technique is to avoid the misspecification of the demand curve. The downside is that the 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.

3. Defining Specifications

Because the notion of a specification is central to our exercise, we spell out its precise meaning and related concepts. Consider a decision problem represented by an outcome function $G : \Sigma \to \Delta(Y)$, where Σ is the strategy space, Y is a set of *outcomes* and $\Delta(Y)$ is the collection of all distributions over Y. Let $\mathcal{G}$ be the set of all outcome functions.

Let $\mathcal{R}$ be a collection of *formal statements*. We assume that each $R \in \mathcal{R}$ is associated with a set $X(R)$. While we do not impose any further restrictions on $X(R)$, we consider a situation where R describes the functional form of an outcome function and $x \in X(R)$ is the parameter that specifies the distribution over outcomes. We interpret $X(R)$ as the set of all admissible parameters. We can also consider the "union" of two formal statements (a logical "or") of R' and R'' as $R' \vee R''$.

Definition 3.1. *A* specification *is a pair $(R, X(R))$ where R is a formal statement and $X(R)$ is the collection of all feasible parameters under R.*

Each $(R, X(R))$ is associated with a set of possible outcome functions. Define $S(R, X(R))$ as the collection of outcome functions satisfying the formal description R and the parametric constraint $X(R)$, whose generic element is written as G. We take $S(R, X(R)) \subset \mathcal{G}$.

We say a specification is *complete* if $\left|S(x)\right| = 1$ for all $x \in X(R)$. Otherwise, the specification is *incomplete*. Note that a complete specification is necessary to define the decisionmaker's objective function. If $(R, X(R))$ is complete, $\forall x \in X(R)$, $S(R, x)$ is a unique outcome function, which we call a model. We call $S(R, X(R))$ a model class induced by specification $(R, X(R))$.

Definition 3.2. *Suppose that G^* is the true distribution. A specification $(R, X(R))$ is* correct *if $\exists x \in X(R)$ such that $G^* \in S(R, x)$. Otherwise, we say that $(R, X(R))$ is* misspecified*.*

Our notion of misspecification coincides with the definition of misspecification by Esponda and Pouzo (2014).[5]

Definition 3.3. *Let $(R^*, X(R^*))$ be the correct specification, and suppose that a specification $(R, X(R))$ is given. We say that $(R, X(R))$ is underspecified, if $\forall x \in X(R^*)$,*

[5]Our formulation of specifications can be generalized to admit multiple outcome functions instead of a single outcome function. Note that we admit that the specification can be incomplete yet correct.

$S(R^*, x) \subset S(R, X(R))$. *We say that* $(R, X(R))$ *is strictly underspecified if the inclusion is strict.*

Underspecification is not a misspecification because the actual specification implies the decision maker's specification. By definition, underspecifications are incomplete specifications. The decisionmaker cannot write down the optimization problem unless one imposes a further restriction on $(R, X(R))$.

4. Description

4.1. **Demand.** There are N buyers, each of whom is indexed by $i \in \{1, \dots, N\}$ and is endowed with reservation value $v_i \in [\underline{v}, \overline{v}]$. Let $F_i(v_i)$ be the distribution of valuation of buyer i. We assume that v_i and v_j are independent $\forall i \neq j$. Given p, buyer i purchases one unit of the good if $p \leq v_i$. If the seller charges p, the (normalized) aggregate demand is

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

Define

$$F(p) = \frac{1}{N}\sum_{i=1}^{N} F_i(p) \quad \text{and} \quad \epsilon_2 = q - (1 - F(p)), \tag{4.1}$$

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

We assume that the aggregate distribution F satisfies the increasing hazard rate property and a Lipschitz continuity property.

- (Increasing Hazard Rate.) The support of F is $[\underline{p}, \overline{p}]$. $\forall F \in \mathcal{F}$, its density function f is continuous over $[\underline{p}, \overline{p}]$ and its hazard rate

  $$\frac{f(p)}{1 - F(p)}$$

  is increasing. Let $\mathcal{F}^0$ be the set of all distributions over the buyer's valuations satisfying the increasing hazard rate property. To avoid technical problems, we assume the Lipschitz continuity of the density function.
- (Lipschitz.) Define

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

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

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

$$b^*(F) = \arg\max_p p(1 - F(p)). \tag{4.2}$$

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

Lemma 4.1. (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) $\cup_{\eta>0}\mathcal{F}^\eta$ *is a dense subset of* $\mathcal{F}^0$.

4.2. **Seller's Problem.** In the rational benchmark, the monopolist knows the distribution F of the buyer's valuations. Instead, the monopolist is endowed with $\mathcal{F}^\eta$. We assume that the monopolist seller 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 consumers with IID draws of their valuations in each period.

Time is discrete: $t = 1, 2, 3, \ldots$. In each period, N consumers enter the market, each of whom is endowed with reservation value $v_{i,t}$ drawn according to distribution function $F_i(\cdot)$. The monopolist and N consumers play a normal form game Γ in each period. 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 one period game Γ. Both games are based on the same monopoly market but generate substantially different data ex post.

Example 4.2. *Suppose that* Γ_{CL} *is the "textbook" static monopoly market (also known as the post price mechanism). At the beginning of period* t*, the monopolist estimates the expected demand curve* $\hat{Q}_{t-1}(p)$ *where* p *is the price.* $\hat{Q}_{t-1}(p)$ *is a strictly decreasing* ℓ*-th order polynomial function of* p *with a strictly decreasing Lipschitz continuous marginal revenue curve whose slope is bounded away from 0. 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, receiving payoff* $v_{i,t} - p_t$*. If* $p_t < v_{i,t}$*, buyer* i *rejects this offer, receiving 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. Since the buyer's valuation is drawn each period,* q_t *is a random variable. The monopolist's payoff in period* t *is* $p_t q_t$*. 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 4.2, the seller observes only the amount of sales 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 forecasting rule $\mathcal{A}$ uses to calculate $\hat{Q}_t$.

Example 4.3. *Suppose that* Γ_{CR} *is the revelation game (Cole and Roughgarden (2014)) of the monopoly market game in Example 4.2. 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*, conditioned on his true type* $v_{i,t}$*. The monopolist observes* $(\hat{v}_{1,t}, \ldots, \hat{v}_{N,t})$ *and decides how to allocate the goods. 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 forecasting rule to update $\hat{F}_{t-1}$, which we call $\mathcal{A}_{CR}$. 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

$$\hat{F}_t(v) = \hat{F}_{t-1}(v) + \frac{1}{t}\left(\tilde{F}_t(v) - \hat{F}_{t-1}(v)\right). \tag{4.3}$$

Example 4.3 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 outcome of the one period game Γ_{CR} in Example 4.3 in period t is

$$O_t = (\hat{v}_{1,t}, \ldots, \hat{v}_{N,t}, p_t, q_t). \tag{4.4}$$

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 is a primitive of Γ_{CR}.

We admit any monopoly market trading protocol but require that the monopolist charges a single delivery price and can observe the quantity at the end of any one period game Γ.[6]

Assumption 4.4. *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 price p_t. At the end of one period game Γ, the monopolist observes at least a pair (p_t, q_t) of the price and the realized demand in period t.*

4.3. **Components of Algorithms.** The examples highlight the necessary components of an algorithm, including data, forecast, and the forecasting rule that maps data to forecast as well as the behavior rule of the monopolist responding to the forecast. Let us explain each component one by one to define the algorithm.

4.3.1. *Data.* The outcome O_t specifies what a player can observe at the end of period t and is exogenously determined by the one-period game Γ. We regard O_t as an array of elements as in (4.4). 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 $t \geq 1$.[7]

By the data, we mean a 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. While outcome O_t is determined by the one period game Γ, the algorithm spells out what elements of an outcome can be data. Let $\dim D_t$ be the number of components in D_t. Being a "sub-array" of O_t, $\dim D_t \leq \dim O_t$.

[6]We focus on uniform price auctions for simplicity. We can admit Γ where the monopolist can charge discriminatory prices. Then, we require that the monopolist can observe $(p_{1,t}, q_{1,t}, \ldots, p_{K,t}, q_{K,t})$ where $q_{k,t}$ is the amount of goods sold at price $p_{k,t}$.

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

The decisionmaker can choose to ignore some elements in the outcome as in Example 4.3. In Example 4.2, $\dim D_t = 2$ and in Example 4.3, $\dim D_t = K$ and. Define $\mathcal{D}_t = (D_1, \ldots, D_{t-1})$ and $\mathcal{D}$ as the collection of all feasible $\mathcal{D}_t$ $\forall t \geq 1$, which we call the data structure (of an algorithm).

4.3.2. *Forecast.* The central component of the algorithm is the forecasting rule that predicts the distribution of the buyers using a history of data:

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

We focus on recursive forecasting rules.

Definition 4.5. *$\mathcal{A}$ is a recursive forecasting rule if $\exists \Phi$ 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 variable in period t used by the algorithm, but not a part of the forecast.

Naturally, Ω_t is not observed by an outsider at the end of Γ in period t. We require that the internal state variable Ω_t be updated according to the fixed function Φ based on the algorithm's observation. Let $\dim \Omega_t$ be the number of elements in Ω_t, treating Ω_t as an array of variables.

We can be completely general about the range of the forecast rule within $\mathcal{F}^\eta$.[8] To simplify the exposition, however, we opt for a parametric family of specifications widely used in practice but also sufficiently general to approximate a large class of functions.

Let R^ℓ be the statement that the functional form of the aggregate distribution of valuation is an ℓ-th order polynomial. A canonical example would be the distribution F is an ℓ-th order polynomial of price p:

(1)
$$F(p) = 1 - \left(\sum_{i=0}^{\ell} \beta_i p^i\right) \tag{4.5}$$
if $0 < F(p) < 1$. The support of F is $[\underline{p}, \overline{p}]$.

(2) $F'(p) = f(p) > 0$ whenever $0 < F(p) < 1$ and Lipschitz continuous with Lipschitz constant $\eta > 0$.

(3) F satisfies the increasing hazard rate property.

Define $X(R^\ell)$ as the set of all admissible $\ell+1$ coefficients $(\beta_i)_{i=0}^{\ell}$ so that (4.5) is an element of $\mathcal{F}^\eta$. The expected demand curve

$$Q(p) = 1 - F(p) = \sum_{i=0}^{\ell} \beta_i p^i \tag{4.6}$$

is a strictly decreasing smooth function with a strictly decreasing marginal revenue curve whose slope is uniformly bounded away from 0.

The monopolist is essentially forecasting the expected demand curve. To "test" whether the forecast is sensible, the forecasting rule must impose restrictions between the price and

[8]See Section 10.2.

the quantity. The forecasting rule must take at least (p_t, q_t) as an input in each period if $\ell \geq 1$. We assume that any feasible algorithm satisfies the property.[9]

Assumption 4.6. *If $\mathcal{A}(\mathcal{O}_t) \in S(R^\ell, X(R^\ell))$ for $\ell \geq 1$, then $\dim D_t \geq 2$.*

Define

$$R^f = \bigvee_{i=0}^{\infty} R^\ell$$

and $X(R^f)$ as the set of all admissible coefficients for R^f. We assume that the set of admissible forecasts is $S(R^f, X(R^f))$ that the collection of all distribution functions that has the functional form of (4.5) for some $\ell \geq 0$. Throughout this paper, a forecasting rule is

$$\mathcal{A} : \mathcal{O} \to S(R^f, X(R^f)). \tag{4.7}$$

We measure the complexity of the data according to the size $\dim D_t$ of the array of D_t. Abusing the notation, we write $\dim \mathcal{A}(\mathcal{O}_t)$ as the number of parameters in $\mathcal{A}(\mathcal{O}_t)$. Suppose that $\mathcal{A}(\mathcal{O}_t) \in S(R^\ell, X(R^\ell))$. $\dim \mathcal{A}(\mathcal{O}_t) \geq \ell + 1$ and the equality holds if $\mathcal{A}(\mathcal{O}_t)$ contains only the price as in (4.5). If $\mathcal{A}(\mathcal{O}_t)$ contains other variables than the price, the strict inequality holds.

Definition 4.7. A *is the set of all recursive forecasting rules satisfying Assumption 4.6 and* (4.7).

4.3.3. *Behavior Rule.* The monopolist chooses $\mathcal{A} \in \mathsf{A}$ to forecast the expected demand curve. We define a behavior rule as a mapping from the forecast to a (randomized) price:

$$\varphi_p : S(R^f, X(R^f)) \to \Delta(\mathbb{R}_+).$$

A canonical example of the behavior rule would be to choose the optimal price of the forecast

$$\varphi_p(\hat{F}_t) = b^*(\hat{F}_t) = \arg\max_p p(1 - \hat{F}_t(p)).$$

where $\hat{F}_t = \mathcal{A}(\mathcal{O}_t) \in S(R^f, X(R^f))$. To allow the monopolist to explore the demand, we admit a randomized price

$$\varphi_{p,t}(\hat{F}_t) = b^*(\hat{F}_t) + \epsilon_{1,t} \tag{4.8}$$

where $\epsilon_{1,t}$ is a white noise with a small variance $\sigma^2_{1,t} \geq 0$. While we can be completely general about $\varphi_{p,t}$, this paper fixes the behavior rule as (4.8) to simplify the exposition.

4.4. **Algorithm.** We are ready to define an algorithm.

Definition 4.8. *An algorithm (or statistical procedure) is a profile $(\mathcal{A}, \mathcal{D}, (\varphi_{p,t})_{t\geq 1})$ of the forecasting rule, the data structure, and the behavior rule. An algorithm maps a history $\mathcal{O}_t$ of outcomes to a (randomized) action*

$$\varphi_p(\mathcal{A}(\mathcal{O}_t)) \in \Delta(\mathbb{R}_+) \qquad \forall \mathcal{O}_t \in \mathcal{O}$$

by combining the forecasting rule and the behavior rule. Let A be the set of all feasible algorithms of the monopolist.

[9]As we admit an arbitrary forecasting rule, the forecasting rule may use an exotic function such as the space-filling curve to forecast two dimensional objects through single dimensional data. While such a forecasting rule is hardly sensible, we choose to assume it away.

Because we fix the behavior rule of the monopolist as (4.8), the forecasting rule $\mathcal{A}$ is the component where the data is processed. Because virtually all computational exercises occur in $\mathcal{A}$, we refer to forecasting rule $\mathcal{A}$ as an algorithm whenever the meaning is clear from the context.

4.5. **Complexity Measure.** 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 memory size necessary for the algorithm's operation in each period t.

Example 4.9. *In Example 4.3, algorithm $\mathcal{A}_{CR}$ collects K reports of the reported reservation values from buyers in period t, then $\dim(D_t) = K$. $\mathcal{A}_{CR}(\mathcal{O}_t)$ is the empirical distribution $\hat{F}_t$, which can be highly non-linear. To represent $\hat{F}_t(v)$ as a collection of parameters, we need to remember virtually all realized values of $\hat{F}_t(\hat{v})$ for each reported value $\hat{v}$ up to $t-1$ round. In Example 4.2, $\mathcal{A}(\mathcal{O}_{t-1})$ is an estimated linear demand curve. Then, $\dim(\mathcal{A}(\mathcal{O}_{t-1})) = 2$. Since the seller updates the estimated demand using (p_t, q_t), $\dim(D_t) = 2$.*

In some algorithms, $\dim D_t$ or $\dim \mathcal{A}(\mathcal{O}_t)$ can change over time.

Example 4.10. *We can implement algorithm $\mathcal{A}_{CR}$ in Example 4.3 as a batch process as in Cole and Roughgarden (2014). Instead of updating the empirical distribution in each period in response to K new observations, we can calculate the empirical distribution at the end of terminal round T. Let us call the batch mode algorithm $\mathcal{A}^o_{CR}$. In each period $t \leq T-1$, the algorithm takes the K reported valuations of the buyers. For $t \leq T-1$, $\dim(D_t) = K$ and $\dim(\mathcal{A}^o_{CR}(\mathcal{O}_t)) = 0$ since the algorithm produces no forecast. In period T, the algorithm calculates the empirical distribution using KT observations of valuations. But, $\dim(D_T) = KT$ because the algorithm requires reading all existing data to produce $\mathcal{A}^o_{CR}(\mathcal{O}_T)$.*

In Example 4.2, $\dim D_t = 2$ and $\dim \mathcal{A}(\mathcal{O}_t) = 2$ remain constant until the algorithm produces the final forecast in period T.

Example 4.11. *Suppose that the monopolist estimates a linear demand*

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

through the least square estimation algorithm $\mathcal{A}_{LSE}$, which recursively estimate (β_0, β_1) according to

$$\begin{bmatrix}\beta_{0,t}\\ \beta_{1,t}\end{bmatrix} = \begin{bmatrix}\beta_{0,t-1}\\ \beta_{1,t-1}\end{bmatrix} + \frac{1}{t} C_{t-1}^{-1} \begin{bmatrix}1\\ p_t\end{bmatrix} \left(q_t - \beta_{0,t-1} - \beta_{1,t-1} p_t\right)$$

and

$$C_t = C_{t-1} + \frac{1}{t}\left(\begin{bmatrix}1 & p_t\\ p_t & p_t^2\end{bmatrix} - C_{t-1}\right).$$

The algorithm forecasts the demand curve as $q = \beta_{0,t} + \beta_{1,t} p$ which requires two parameters, using (q_t, p_t) in each period as inputs. Thus, $\dim(\mathcal{A}_{LSE}(\mathcal{O}_t)) = 2$ and $\dim D_t = 2$. But,

the algorithm has to keep track of $C_t = (c_{ij,t})_{i,j\in\{1,2\}}$*, which is a symmetric matrix where* $c_{11,t} = 1\ \forall t \geq 1$*. To keep the record of* C_t*, the algorithm has to remember* $c_{12,t}$ *and* $c_{22,t}$*. Thus,* $\dim(\Omega_t) = 2$*. Thus,* $\dim(\mathcal{A}_{LSE}(\mathcal{O}_t), D_t, \Omega_t) = 6$*.*

Definition 4.12. *The complexity of* $\mathcal{A}$ *up to time* T *is*

$$\text{comp}_T(\mathcal{A}) = \max_{1\leq t\leq T} \dim(\mathcal{A}(\mathcal{O}_{t-1}), D_t, \Omega_t)$$

and if the algorithm operates indefinitely,

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

4.6. **Algorithm Game.** The algorithm game is a normal form game between the monopolist and nature. The monopolist's strategy space is A as defined in Definition 4.8. 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 (Rubinstein (1986)). We treat forecasting rule $\mathcal{A}$ as an algorithm, as we fix the behavior rule as (4.8), and almost all computational tasks are associated with forecasting rule $\mathcal{A}$.

Conditioned on $(\mathcal{A}, F) \in \mathsf{A}\times\mathcal{F}^\eta$, the algorithm generates a forecast about the aggregate distribution of valuation $\hat{F}_t$. Based on the forecast, the monopolist chooses a (randomized) price $p_t = \varphi_{p,t}(\hat{F}_t)$ expecting that the market demand is

$$1 - \hat{F}_t(p_t)$$

on average, but the actual demand is

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

Learning the demand curve is only a tool for learning the maximum profit. Thus, the monopolist measures the performance of the forecast $\hat{F}_t = \mathcal{A}(\mathcal{O}_t)$ according to how accurately the algorithm forecasts the profit-maximizing market outcome:

$$(b^*(F), 1 - F(b^*(F))). \tag{4.9}$$

We assume that the objective function of the monopolist in the algorithm game is the (negative value of) the mean squared error of the forecast by the algorithm

$$\mathcal{U}(\mathcal{A}, F) = \lim_{T\to\infty} -\frac{1}{T}\mathbb{E}\sum_{t=1}^{T}\left|(\varphi_{p,t}(\hat{F}_t), 1 - \hat{F}_t(\varphi_{p,t}(\hat{F}_t))) - (b^*(F), 1 - F(b^*(F)))\right|^2 \tag{4.10}$$

Note that $\mathcal{U}(\mathcal{A}, F) \leq 0$ and the equality holds if and only if $\mathcal{A}$ generates a sequence $\left\{\left(\hat{F}_t, \varphi_{p,t}(\hat{F}_t)\right)\right\}_t$ where $\hat{F}_t = \mathcal{A}(\mathcal{O}_t)$ such that

$$\left(\varphi_{p,t}(\hat{F}_t), 1 - \hat{F}_t(\varphi_{p,t}(\hat{F}_t))\right) \to (b^*(F), 1 - F(b^*(F))).$$

If one follows the robust approach, we would assume the monopolist optimizes against an adversarial nature whose goal is to minimize the monopolist's payoff:

$$\max_{\mathcal{A}\in\mathsf{A}} \min_{F\in\mathcal{F}^\eta} \mathcal{U}(\mathcal{A}, F).$$

Instead, we assume that the monopolist is searching for a dominant solution of the game.

Definition 4.13. *$\mathcal{A}$ is a dominant algorithm if*

$$\mathcal{A} \in \bigcap_{F \in \mathcal{F}^\eta} \arg\max_{\mathcal{A} \in \mathsf{A}} \mathcal{U}(\mathcal{A}, F). \tag{4.11}$$

or equivalently,

$$\mathcal{U}(\mathcal{A}, F) \geq \mathcal{U}(\mathcal{A}', F) \qquad \forall \mathcal{A}' \in \mathsf{A}, \forall F \in \mathcal{F}^\eta.$$

A dominant algorithm $\mathcal{A}$ induces a best response against any actual distribution in the long run. The monopolist incurs the complexity cost as defined in Definition 4.12. To minimize the impact of the complexity cost, we assume that the monopolist has lexicographic preference over the long run average profit and the complexity cost.

We are searching for a dominant algorithm with minimal complexity in A.

Definition 4.14. *$\mathcal{A}^* \in \mathsf{A}$ is an optimal algorithm if $\mathcal{A}^* \in \bigcap_{F \in \mathcal{F}^\eta} \arg\max_{\mathcal{A} \in \mathsf{A}} \mathcal{U}(\mathcal{A}, F)$ and $\operatorname{comp}(A^*) \leq \operatorname{comp}(A)$ $\forall A \in \bigcap_{F \in \mathcal{F}^\eta} \arg\max_{\mathcal{A} \in \mathsf{A}} \mathcal{U}(\mathcal{A}, F)$.*

The main result of the paper is to construct an optimal algorithm.

Theorem 4.15. *We can construct an optimal algorithm*

$$\mathcal{A} : \mathcal{O} \to S(R^1, X(R^1))$$

where $\dim(\mathcal{A}(\mathcal{O}_t)) = \dim(D_t) = 2$ $\forall t \geq 1$ and $\operatorname{comp}(\mathcal{A}) = 4$.

The optimal algorithm we construct presumes that the demand curve is linear and estimates the two coefficients of the linear demand. While the recursive least square algorithm inspires the optimal algorithm, the optimal algorithm is simpler than the recursive least square algorithm as it does not require keeping track of the internal states to achieve $\operatorname{comp}(\mathcal{A}) = 4$. As illustrated in Example 4.11, the recursive least square algorithm requires keeping track of the covariance matrix so that $\dim \Omega_t = 2$ and the complexity of the recursive least square algorithm is 6.

5. Preliminaries

5.1. **PAC.** It is not straightforward to verify whether a given algorithm is a dominant strategy in A in the algorithm game against nature. Since $\mathcal{U}(\mathcal{A}, F) \leq 0$ $\forall (\mathcal{A}, F)$, $\mathcal{A}$ is a dominant algorithm if

$$\inf_{F \in \mathcal{F}^\eta} \mathcal{U}(\mathcal{A}, F) = 0. \tag{5.12}$$

As it turns out, it is easier to construct an algorithm $\mathcal{A}$ with (5.12). Formally, we are constructing an algorithm satisfying a notion of uniform convergence that has been developed and widely used in the computer science literature (Shalev-Shwartz and Ben-David (2014)).

Definition 5.1. *$\mathcal{A}$ PAC guarantees $\mathcal{F}^\eta$ if $\forall \mu > 0$, $\forall \lambda \in (0,1)$ and $\overline{T}(\mu, \lambda)$ such that*

$$\mathbb{P}\left(\left| \left(\varphi_{p,t}(\hat{F}_t)), 1 - \hat{F}_t(\varphi_{p,t}(\hat{F}_t))) \right) - \left(b^*(F), 1 - b^*(F) \right) \right| \geq \mu \right) \leq \lambda \qquad \forall t \geq \overline{T}(\mu, \lambda) \tag{5.13}$$

where $\hat{F}_t = \mathcal{A}(\mathcal{O}_t)$ and $\overline{T}(\mu, \lambda) \sim \mathsf{O}\left(-\frac{\log \lambda}{\mu^p} \right)$ for some $p > 0$.

If $\mathcal{A}$ PAC guarantees $\mathcal{F}^\eta$, then $\mathcal{A}$ learns $b^*(F)$ uniformly over $\mathcal{F}^\eta$. $\mathcal{A}$ forecasts the best response for every $F \in \mathcal{F}^\eta$ accurately with high confidence. Because $\mu > 0$ and $\lambda \in (0,1)$ is arbitrary, we can show that $\mathcal{A}$ is a dominant strategy of A.

Proposition 5.2. *If $\mathcal{A}$ PAC guarantees $\mathcal{F}^\eta$, then $\mathcal{A}$ is a dominant strategy in A.*

Proof. To prove that $\mathcal{A}$ is a dominant strategy, it suffices to show that $\forall \epsilon > 0$

$$\mathcal{U}(\mathcal{A}, F) \geq -\epsilon \qquad \forall F. \tag{5.14}$$

Fix $\mu > 0$ and $\lambda > 0$. If $\mathcal{A}$ PAC guarantees $\mathcal{F}^\eta$, $\exists T(\mu, \lambda) > 0$ such that $\forall t \geq \overline{T}(\mu, \lambda)$, $\forall F$,

$$\mathbb{P}\left(\left|\Big(\varphi_{p,t}(\hat{F}_t)), 1 - \hat{F}_t(\varphi_{p,t}(\hat{F}_t))\Big) - \big(b^*(F), 1 - b^*(F)\big)\right| \geq \mu\right) \leq \lambda.$$

We can choose $\mu > 0$ and $\lambda > 0$ sufficiently small so that $\epsilon > \mu(1-\lambda)$. Then, $\forall t \geq T(\mu, \lambda)$,

$$\mathbb{E}\left|\Big(\varphi_{p,t}(\hat{F}_t)), 1 - \hat{F}_t(\varphi_{p,t}(\hat{F}_t))\Big) - \big(b^*(F), 1 - b^*(F)\big)\right|^2 \leq \epsilon \qquad \forall F.$$

Thus,

$$\mathcal{U}(\mathcal{A}, F) \geq -\epsilon \qquad \forall F.$$

□

Following Cole and Roughgarden (2014), one can prove that algorithm $\mathcal{A}_{CR}$ in Example 4.3 PAC guarantees $\mathcal{F}^\eta$ if the one period game Γ is the revelation game.

Theorem 5.3. *Suppose that Γ_{CR} is the revelation game of the monopoly problem in Example 4.3. $\mathcal{A}_{CR}$ PAC guarantees $\mathcal{F}^\eta$, and therefore a dominant strategy in A.*

Proof. See Cole and Roughgarden (2014). □

We have yet to show whether a dominant strategy exists if Γ is not the revelation game as in Example 4.2. Also, if a dominant strategy exists, we ask whether a simpler algorithm than $\mathcal{A}_{CR}$ can PAC guarantee $\mathcal{F}^\eta$.

We prove Theorem 5.4 by constructing a recursive algorithm $\mathcal{A}$ with $\text{comp}(\mathcal{A}) = 4$, that PAC guarantees $\mathcal{F}^\eta$.

Theorem 5.4. *We can construct a recursive algorithm*

$$\mathcal{A} : \mathcal{O} \to S(R^1, X(R^1))$$

that PAC guarantees $\mathcal{F}^\eta$, where $\dim(\mathcal{A}(\mathcal{O}_t)) = \dim(D_t) = 2$ $\forall t \geq 1$ and $\text{comp}(\mathcal{A}) = 4$.

Proof. We state the result more formally as Theorem 9.1 whose proof can be found in Appendix C. □

Since any PAC guaranteeing algorithm is a dominant algorithm, the upper bound of the complexity of the optimal algorithm is no more than 4. From Theorem 5.4, we can derive the lower bound of the complexity of an optimal algorithm.

Proposition 5.5. *Suppose a recursive algorithm exists that PAC guarantees $\mathcal{F}^\eta$. If $\mathcal{A} : \mathcal{O} \to S(R^0, X(R^0))$, $\mathcal{A}$ is not a dominant strategy in A. Moreover, $\dim(D_t) \geq 2$ $\forall t \geq 1$ if $\mathcal{A}$ is a dominant strategy.*

Proof. See Appendix A. □

5.2. **Uniform Learnability.** Instead of directly constructing an algorithm satisfying the PAC guaranteeing property, we first construct an algorithm satisfying a weaker requirement.

Definition 5.6. *$\mathcal{A}$ uniformly learns $\mathcal{F}^\eta$ if $\forall \lambda \in (0,1)$, $\forall \mu > 0$, $\exists \overline{T}(\mu,\lambda)$ such that*

$$\mathbb{P}\left(\left|(\varphi_p(\mathcal{A}(\mathcal{O}_{\overline{T}(\mu,\lambda)})), 1 - F(\varphi_p(\mathcal{A}(\mathcal{O}_{\overline{T}(\mu,\lambda)})))) - (b^*(F), 1 - b^*(F))\right| \geq \mu\right) \leq \lambda \tag{5.15}$$

where $\overline{T}(\mu,\lambda) \sim \mathsf{O}\left(-\frac{\log \lambda}{\mu^p}\right)$ for some $p > 0$, and $\varphi_{p,t} = \varphi_p$ $\forall t \geq 1$.

If $\mathcal{A}$ PAC guarantees $\mathcal{F}^\eta$, $\mathcal{A}$ uniformly learns $\mathcal{F}^\eta$.[10] After analyzing the properties of a uniformly learnable algorithm, we modify the algorithm to satisfy the PAC guaranteeing property in Section 9. The detour has several benefits.

First, the weaker notion of PAC learnability allows us to examine algorithms like $\mathcal{A}^o_{CR}$ in Example 4.10, whose complexity is ∞ if the algorithm has to run indefinitely. We can use the uniform learnability for an algorithm that terminates in a finite number of rounds, producing the final forecast, which the monopolist uses for the continuation game.[11] We can analyze the optimality and the complexity of $\mathcal{A}^o_{CR}$, to learn the substantive difference between $\mathcal{A}^o_{CR}$ and our algorithm without being distracted by technical details.

Second, a uniformly learnable algorithm is simpler than a PAC guaranteeing algorithm because we can choose $\varphi_{p,t} = \varphi_p$ $\forall t \geq 1$. The uniformly learnable algorithm generates more robust numerical results, which greatly helps to compare the numerical performance of our algorithm to our benchmark algorithm $\mathcal{A}^o_{CR}$. The modification of the uniformly learnable algorithm to a PAC guaranteeing algorithm follows the well established technical procedure in stochastic approximation, which preserves the asymptotic properties of the algorithm (Kushner and Yin (1997)).

Third, uniform learnability implies a weaker version of dominance. The weaker notion of dominance is useful for extending the result to the case where the monopolist discounts the future payoff (Section 10.1). To define a weaker notion of dominance, imagine a monopolist who can run algorithm $\mathcal{A}$ for a finite number of rounds, say T, to learn the market outcome that maximizes the profit. At the end of period T, the algorithm produces the final forecast $\varphi_p(\mathcal{O}_T)$, which the monopolist uses for the continuation game. Thus, the long run average payoff is

$$\mathcal{U}(\mathcal{A}, F, T) = -\mathbb{E}\left|(\varphi_p(\hat{F}_T), 1 - \hat{F}_T(\varphi_p(\hat{F}_T))) - (b^*(F), 1 - b^*(F))\right|^2$$

where $\hat{F}_T = \mathcal{A}(\mathcal{O}_T)$. Note that $\mathcal{A}$ is a best response to F is $\mathcal{U}(\mathcal{A}, F, T) = 0$. Define for a small $\epsilon > 0$, and $F \in \mathcal{F}^\eta$,

$$\mathsf{BR}^\epsilon(F,T) = \left\{\mathcal{A} \mid \mathcal{U}(\mathcal{A}, F, T) \geq -\epsilon\right\}$$

[10]The PAC guaranteeing property requires that the forecast of the algorithm must converge to the actual profit maximizing outcome uniformly in probability, while uniform learnability requires uniform convergence in distribution.

[11]A typical machine learning algorithm (e.g., Schapire and Freund (2012)) goes through a finite number of rounds of training, and terminates with a final forecast.

as the set of ϵ best responses against F if the algorithm can run for T rounds. Define the set of ϵ dominant strategies as

$$\bigcup_{T\geq 1}\bigcap_{F\in\mathcal{F}^\eta}\mathsf{BR}^\epsilon(F,T)$$

and its element is called an ϵ dominant strategy. If $\mathcal{A}$ is an ϵ dominant strategy, then $\exists T$ such that $\mathcal{A}$ generates ϵ best response by T rounds for all $F\in\mathcal{F}^\eta$. The substance of the definition is that the termination time and the amount of deviation from the best response are defined uniformly over $\mathcal{F}^\eta$.

It is straightforward to show that any uniformly learnable $\mathcal{A}$ is an ϵ dominant solution. Let us state the observation without proof for later reference.

Proposition 5.7. *$\forall\epsilon>0$, $\mathcal{A}$ is an ϵ dominant strategy if $\mathcal{A}$ uniformly learns $\mathcal{F}^\eta$.*

We can establish the same lower bound by following the same logic as in Proposition 5.8.

Proposition 5.8. *Suppose that a uniformly learnable algorithm exists. If $\mathcal{A}:\mathcal{O}\to S(R^0,X(R^0))$, $\mathcal{A}$ is not an ϵ dominant strategy in A for any sufficiently small $\epsilon>0$.*

6. Construction

We now construct a recursive algorithm

$$\mathcal{A}_a:\mathcal{O}\to S(R^1,X(R^1))$$

with $\dim(\mathcal{A}(\mathcal{O}_{t-1}))=\dim(D_t)=2$ and $\dim(\Omega_t)=0\ \forall t\geq 1$, where we select $a>0$ to meet the accuracy and confidence bound (μ,λ). We show that for a sufficiently small $a>0$, the constructed algorithm $\mathcal{A}_a$ uniformly learns $\mathcal{F}^\eta$ in Theorem 6.1.

6.1. Linear Recursive Algorithm.

Our algorithm uses price and quantity as data. $D_t=(q_t,p_t)$ is the pair of quantity 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

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

$\mathbb{E}\epsilon_{2,t}=0$ and $\mathbb{E}\epsilon_{2,t}^2<\infty$ uniformly, but the actual size of the second moment can depend on p_t and $F\in\mathcal{F}$.

Let

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

be the history at the beginning of period t. The monopolist assumes that the aggregate demand is a linear function in $S(R^1,X(R^1))$:

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

and estimates (β_0,β_1).

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 $b^*(F)\in K$ $\forall F\in\mathcal{F}^\eta$. 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.

Recall that if (β_0, β_1) can support $b^*(F)$ for some $F \in \mathcal{F}^\eta$,

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

or equivalently,

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

Since $(1 - F(b^*(F)), b^*(F)) \in K$, there exists a compact set $\mathsf{B} \subset (0, \infty) \times (-\infty, 0)$ such that $(\beta_0, \beta_1) \in \mathsf{B}$ if the linear demand can support a true optimal price. 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 $\mathcal{H}_\mathsf{B} \subset S(R^1, X(R^1))$ be the collection of the linear demand curves, which induce the market outcome in K. The seller knows that the market demand curve is in $\mathcal{H}_\mathsf{B}$. Since the seller does not know (β_0, β_1), the seller recursively estimates the parameters using the least square estimation method while choosing the pricing rule based on the estimated linear demand curve. Let $(\beta_{0,t-1}, \beta_{1,t-1})$ be the least square estimator at the end of period $t-1$. Given the estimated demand curve

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

the monopolist calculates the optimal price

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

but incurs an implementation error $\epsilon_{1,t}$ 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 $\mathbb{E}\epsilon_{1,t}^2 = \sigma_1^2$.[12] We control the size of $\sigma_1^2 > 0$ to achieve the desired level of accuracy of the algorithm imposed by uniform learnability (5.15).

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 quantity in period t is

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

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 $\epsilon_t = (\epsilon_{1,t}, \epsilon_{2,t})$.

[12]We can also choose $\epsilon_{1,t}$ from a small interval $[-\bar{\epsilon}, \underline{\epsilon}]$ according to a fixed distribution, say the uniform distribution.

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

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

so that

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

Similarly, if $q_t = 1$,

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

so that

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

If $0 < q_t < 1$, then

$$\begin{bmatrix}\beta_{0,t}\\ \beta_{1,t}\end{bmatrix} = \begin{bmatrix}\beta_{0,t-1}\\ \beta_{1,t-1}\end{bmatrix} + aR_{t-1}^{-1}\begin{bmatrix}1\\ -\frac{\beta_{0,t-1}}{2\beta_{1,t-1}} + \epsilon_{1,t}\end{bmatrix}\phi(\beta_{t-1}, \epsilon_t) \tag{6.16}$$

where

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

Since the monopolist designs the size of the experiments, the variance σ_1^2 of $\epsilon_{1,t}$ is a known parameter.[13] Since $a > 0$ controls the amount of change of the estimator $\beta_t - \beta_{t-1}$, we call $a > 0$ the gain function. Since the gain function is a constant, our algorithm is known as the constant gain algorithm (Evans and Honkapohja (2001)).

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.[14]

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

$$\begin{bmatrix}\beta_{0,t}\\ \beta_{1,t}\end{bmatrix} = \begin{bmatrix}\beta_{0,t-1}\\ \beta_{1,t-1}\end{bmatrix} + aR_{t-1}^{-1}\begin{bmatrix}1\\ -\frac{\beta_{0,t-1}}{2\beta_{1,t-1}} + \epsilon_{1,t}\end{bmatrix}\phi(\beta_{t-1}, \epsilon_t) \tag{6.17}$$

[13]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 mean and the variance of the prices.

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

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

$$\psi_{t-1} \equiv \psi(\beta_{t-1}, \epsilon_t) = R_{t-1}^{-1} \begin{bmatrix} 1 \\ -\frac{\beta_{0,t-1}}{2\beta_{1,t-1}} + \epsilon_{1,t} \end{bmatrix} \phi(\beta_{t-1}, \epsilon_t). \tag{6.18}$$

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

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

where $\epsilon_{1,t}$ is an i.i.d. white noise uniformly distributed over $[-\epsilon, \epsilon]$ for a fixed $\epsilon > 0$ so that $\mathbb{E}\epsilon_{1,t} =$ and $\mathbb{E}\epsilon_{1,t}^2 = \sigma_1^2$. 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 $(\beta_{0,t}, \beta_{1,t})$.

Let $\mathcal{A}_a$ be the recursive algorithm with constant gain parameter $a > 0$ $\forall t \geq 1$. $\forall a \geq 0$, algorithm $\mathcal{A}_a$ produces β_t following history of data

$$\mathcal{D}_t = ((q_1, p_1), \ldots, (q_{t-1}, p_{t-1}))$$

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

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

and $\dim(\Omega_t) = 0$, since the algorithm does not keep track of any internal state variables. Thus, $\text{comp}_T(\mathcal{A}_a) = 4$ $\forall T \geq 1$. We choose the gain function $a > 0$ and the variance of experiments $\epsilon_t > 0$ to satisfy the accuracy μ and confidence requirement $1 - \lambda$ in (5.13).

Theorem 6.1. *$\forall \mu > 0$, $\forall \lambda > 0$, $\exists a > 0$, $\exists \epsilon > 0$ and $\exists \overline{T}(\mu, \lambda)$ such that*

$$\mathbb{P}\left(\left| \left(\varphi_p\left(\hat{F}_t\right), 1 - \hat{F}_t\left(\varphi_p\left(\hat{F}_t\right)\right)\right) - \left(b^*(F), 1 - b^*(F)\right) \right| > 4\mu \right) \leq \lambda \qquad \forall F \in \mathcal{F}^\eta$$

where $\hat{F}_t = \mathcal{A}_a(\mathcal{O}_{\overline{T}(\mu,\lambda)})$ and $\overline{T}(\mu, \lambda) \sim \mathsf{O}\left(-\frac{\log \lambda}{\mu^p}\right)$ for some $p > 0$.

Proof. See Appendix B. □

7. Heuristics

We provide a heuristic explanation about how to prove Theorem 6.1, delegating formal details to Appendix B. The argument can be strengthened to prove Theorem 5.4. But, doing so requires covering additional technical details.

An important implication of the increasing hazard rate property is that the optimal price $b^*(F)$ associated with F is unique and completely determined by the first order condition

$$\frac{1-F(b^*(F))}{f(b^*(F))} - b^*(F) = 0. \tag{7.19}$$

For fixed $\tau > 0$, define

$$T(a,\tau) = \left\lceil \frac{\tau}{a} \right\rceil.$$

Following Kushner and Yin (1997), we interpret τ as the (clock) time, and $a > 0$ as the time interval between two adjacent observations of (q_t, p_t) and (q_{t-1}, p_{t-1}). Thus, $T(a,\tau)$ is the number of time steps that can be "squeezed" into $\tau > 0$ (clock) time, where we think of a as representing the length of a period. We are interested in the dynamics of β_t over small $\tau > 0$ when the monopolist can observe data very frequently:

$$\frac{\beta_{t+T(a,\tau)} - \beta_t}{\tau}$$

with $\beta_t = \beta_0$, whose property can be inferred by taking limits:[15]

$$\lim_{\tau\to 0}\lim_{a\to 0} \frac{\beta_{t+T(a,\tau)} - \beta_t}{\tau}.$$

Using the recursive nature of the algorithm, we can write

$$\beta_{t+T(a,\tau)} - \beta_t = \frac{\tau}{T(a,\tau)} \sum_{k=1}^{T(a,\tau)} \psi_{t+k}$$

which implies

$$\begin{aligned} \frac{\beta_{t+T(a,\tau)} - \beta_t}{\tau} &= \frac{1}{T(a,\tau)} \sum_{k=1}^{T(a,\tau)} \psi_{t+k} \\ &= \left[\frac{1}{T(a,\tau)} \sum_{k=1}^{T(a,\tau)} \mathbb{E}_{t+k-1}\psi_{t+k}\right] + \left[\frac{1}{T(a,\tau)} \sum_{k=1}^{T(a,\tau)} \psi_{t+k} - \mathbb{E}_{t+k-1}\psi_{t+k}\right] \end{aligned} \tag{7.20}$$

where ψ_{t+k} is defined in (6.18). Our proof will analyze this expression in detail. Define

$$\xi_{t+k} = \psi_{t+k} - \mathbb{E}_{t+k-1}\psi_{t+k} \tag{7.21}$$

which is a martingale difference.

7.1. **Tracking the Mean.** Let us examine the first term in (7.20) as $a \to 0$ for a small fixed $\tau > 0$. $\beta_{t+T(a,\tau)} - \beta_t$ is an average of $\{\psi_{t+k}\}_{k=1}^{T(a,\tau)}$. We can approximate

$$\lim_{a\to 0} \frac{\beta_{t+T(a,\tau)} - \beta_t}{\tau} \simeq \mathbb{E}\left[\psi_t \mid \beta_t = \beta\right]$$

or more concisely as

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

[15]Under the conditions we have imposed, a limit exists (cf. Kushner and Yin (1997)).

To derive the formula of the right hand side, 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

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

where

$$b = -\frac{\beta_0}{2\beta_1}. \tag{7.22}$$

To fit the observed data best, the monopolist chooses on average the new coefficients (β_0', β_1') passing through $\big(b+\epsilon, 1-F\left(b+\epsilon\right)\big)$ and $\big(b-\epsilon, 1-F\left(b-\epsilon\right)\big)$. 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 ϵ^2. Results from stochastic approximation theory[16] 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)

$$\begin{aligned} \dot{\beta}_0 &= \beta_0' - \beta_0 = 1 - F(b) + bf(b) - \beta_0 \\ \dot{\beta}_1 &= \beta_1' - \beta_1 = -f(b) - \beta_1. \end{aligned} \tag{7.23}$$

Since (7.22) 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 (7.23), we have

$$\dot{b} = -\frac{f(b)}{2\beta_1}\left[\frac{1-F(b)}{f(b)} - b\right]. \tag{7.24}$$

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 (7.24). The stochastic approximation implies that the least square learning algorithm weakly converges to $b^*(F)$ (Kushner and Yin (1997)).

So far, we have only shown convergence "pointwise" for F, allowing the amount of data needed to achieve the desired level of accuracy to depend on F. We need to do additional work to show uniform convergence over $\mathcal{F}^\eta$. Since $\frac{1-F(b^*(F))}{f(b^*(F))} - b^*(F) = 0$,

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

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

With the increasing hazard rate property,

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

We can show that $\exists c > 0$ such that[17]

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

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

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

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

Since $b^*(F) \leq [\underline{p}, \overline{p}]$ $\forall F \in \mathcal{F}^\eta$, the initial condition of the ordinary differential equation can be selected from a compact set. Since the distance $|b - b^*(F))|$ vanishes uniformly at the order of $e^{-c\tau}$, it takes $\mathsf{O}(-\log \mu)$ amount of time for b to enter the μ neighborhood of $b^*(F)$. This observation proves that the amount of data to approximate the optimal price is uniform over $\mathcal{F}^\eta$. We can also show that the number of data to achieve μ accuracy increases at the polynomial rate of $1/\mu$.

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

$$\frac{1}{T(a,\tau)} \sum_{k=1}^{T(a,\tau)} \xi_{t+k}$$

where ξ_{t+k} is defined as (7.21). By the law of large numbers, we know the average converges to 0. To satisfy the confidence requirement, we need to find $\rho > 0$ such that

$$\mathbb{P}\left(\left|\frac{1}{T(a,\tau)} \sum_{k=1}^{T(a,\tau)} \xi_{t+k}\right| > \mu\right) \leq e^{-\rho T(a,\tau)}$$

holds uniformly for $F \in \mathcal{F}^\eta$. This part of the exercise is to calculate the tail portion of the probability distribution of the average of ξ_{t+k}. 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, which is 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 $\mathcal{F}^\eta$.

In our case, the algorithm uses (q_t, p_t) which is not IID (not even martingale), because of $p_t = b_t + \epsilon_{1,t}$ and b_t is responding to the realized sequence of data. The data generating process is endogenous, making the stochastic process (q_t, p_t) non-stationary. As a result, ξ_{t+k} is not IID but a martingale difference. We need to invoke the Azuma-Hoeffding-Bennett inequality (Dembo and Zeitouni (1998)) to calculate the uniform exponential

[17]If we assume that $\inf_p f(p) > 0$ uniformly, the proof is straightforward. Without this assumption, we need some additional work (Lemma B.1).

rate for all feasible distributions of buyer's valuation, which proves that our algorithm is efficient (Shalev-Shwartz and Ben-David (2014)).

8. Numerical Experiments

While our algorithm $\mathcal{A}_a$ has the same asymptotic properties as the algorithm $\mathcal{A}^o_{CR}$ (Cole and Roughgarden (2014)), we need to rely on numerical experiments to compare the performance of the algorithm with a finite number of data. It is possible to write the algorithm of Cole and Roughgarden (2014) in a recursive form. Yet, their algorithm is best implemented in an off-line form because a recursive formulation of the algorithm of Cole and Roughgarden (2014) requires storing the empirical distribution of the valuation after the t period, which almost always requires remembering t data points.

Instead of $\mathcal{F}^\eta$, we generate the set of feasible distributions over the buyer's valuations from a truncated Gaussian distribution at the mean, where the Gaussian distribution has a mean of 10. By changing the standard deviation of the Gaussian distribution, we can generate different distributions of the valuations, each of which satisfies the increasing hazard rate property and other regularity properties of $\mathcal{F}^\eta$. We select 5000 standard deviations, ranging from 11 to 16, with an increment of 0.001. For each distribution of valuations, we calculate the actual optimal price. As the standard deviation becomes larger, so does the actual optimal price. We assume that there are 100 buyers whose reservation values are drawn from the same distribution F. The realized amount of delivery at a price p is a random variable with a mean of $1 - F(p)$.

For each distribution, we run our algorithm $\mathcal{A}_a$ with $a = 0.0001$ and $\epsilon = 0.75$ (the size of price perturbation) for $T = 300,000$ rounds. At the end of T rounds, we calculate the forecast price and compare it to the optimal price $b^*(F)$ of the actual distribution to calculate the forecasting error for each F. We generate the distribution of forecasting error over the set of feasible distribution functions.[18] For a small $a > 0$, the distribution of the forecasting error has a mean 0 with a small variance. We found the mean is 0.0081, and the variance is 0.0027. Because of the linear approximation of a non-linear demand curve, the linear approximation error contributes to the forecasting error, which vanishes as we reduce the size of price perturbation $\epsilon > 0$. We plot the distribution of forecasting errors in a histogram in Figures 1 and 2 in blue bars. The horizontal axis shows the forecasting error at T. The heights of each part represent the number of distributions whose forecasting error is within the 0.01 neighborhood of each grid.

We calculate the optimal price forecast from Cole and Roughgarden (2014) by drawing K numbers of valuations each period for T rounds to calculate the empirical distribution (thus, KT samples of valuations) and the optimal price from the empirical distribution. We calculate the forecast error and plot the distribution in Figure 1 and 2 using orange bars. In Figure 1, $K = 2$ so that $\mathcal{A}^o_{CR}$ uses the same number of data as $\mathcal{A}_a$ in T rounds. In Figure 2, $K = 10$ so that $\mathcal{A}^o_{CR}$ uses five times as many data as $\mathcal{A}_a$.

Table 1 reports means and variances from the numerical exercises. Because the non-parametric estimation method in Cole and Roughgarden (2014) does not incur any linear approximation error, the mean of the forecasting error is closer to 0. Interestingly, the

[18]The distribution can be approximated as a Gaussian distribution which is a solution of the Ornstein-Ullenbeck equation (cf.Kushner and Yin (1997)).

Per Period	$\mathcal{A}^o_{CR}$	
	Mean	Variance
2	−0.0039	0.0102
4	−0.0002	0.0064
6	−0.0018	0.0048
8	0.0001	0.0041
10	0.0003	0.0035

TABLE 1. The first column reports the number of valuations the algorithm of Cole and Roughgarden [2014] takes per period. For $\mathcal{A}^o_{CR}$, we draw a different number of samples, which changes the mean and the variance of the forecasting errors. The true profit-maximizing price is roughly ranging from 10 to 20. The mean and the variance of the forecasting errors of $\mathcal{A}_{CL}$ are 0.0081 and 0.0027, respectively.

variance of forecasting errors is larger for the same number of data as $\mathcal{A}_a$, which takes two data points (price and quantity) in each period. If $\mathcal{A}^o_{CR}$ (Cole and Roughgarden (2014)) allows the algorithm to receive ten valuation reports (which is five times as much as data $\mathcal{A}_a$ uses), the variance becomes comparable to the variance of the forecasting error of $\mathcal{A}_a$. Note that the distribution of forecasting errors of CR is spread out more than CL in Figure 1. Figure 2 reports the distribution of forecasting errors when CR takes ten reported values in each period. As Table 1 indicates, the two distributions of the forecasting errors become closer.

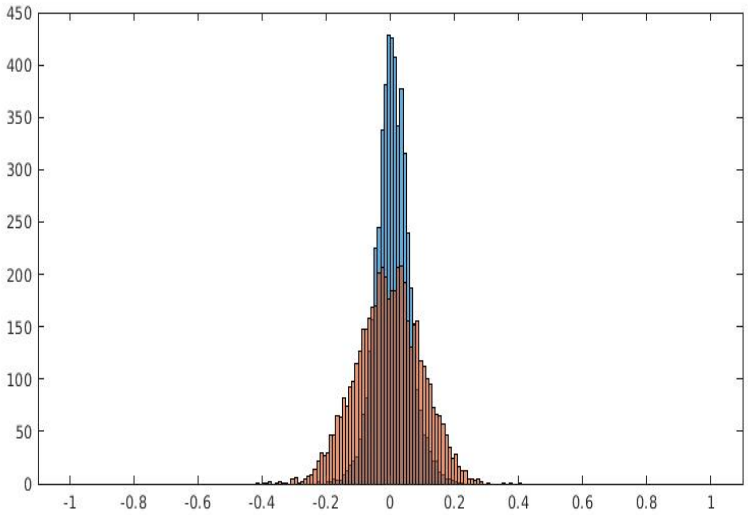

FIGURE 1. Blue bars represent the density of forecasting errors of our algorithm $\mathcal{A}_a$, and orange ones represent the distribution of errors of the algorithm of Cole and Roughgarden [2014]. If the algorithm of Cole and Roughgarden [2014] takes 2 reports per period, the variance is 3.5 times as large as that of $\mathcal{A}_a$.

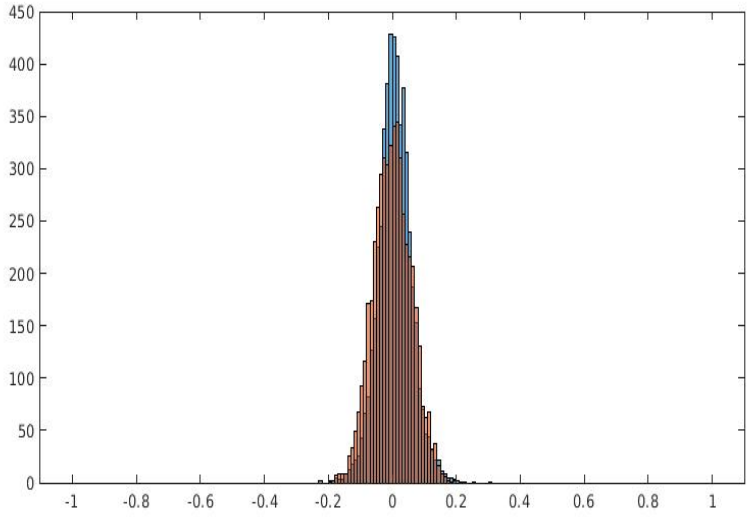

FIGURE 2. If the algorithm of Cole and Roughgarden [2014] takes 10 reports per period, the variance is comparable to that of $\mathcal{A}_a$. The distributions of the forecasting errors almost overlap with each other.

The performance of our algorithm $\mathcal{A}_a$ is comparable to that of $\mathcal{A}^o_{CR}$ (Cole and Roughgarden (2014)). We can reduce the linear approximation error by lowering the price perturbation, choosing a smaller $\epsilon > 0$. We can reduce the forecasting error variance by decreasing the gain $a > 0$. On the other hand, Cole and Roughgarden (2014) can reduce the variance of the forecasting error simply by collecting more data about the valuations.

9. PAC GUARANTEE

We can modify $\mathcal{A}_a$ to satisfy the PAC guaranteeing property (5.13) to establish Theorem 5.4. Recall (6.17) where we choose parameter $a_t = a > 0$. Instead of a positive constant $a > 0$, we can choose $a_t = \frac{1}{(t+t_1)^\omega}$ where $0 < \omega < 1$ and t_1 is a parameter to control the size of the initial value of a_t. Note that a_t decreases as t increases. We need to choose $t_1 > 0$ sufficiently large to achieve the desired level of accuracy. We choose $\omega < 1$ to satisfy the requirement for the data complexity. Let $\mathcal{A}_{\frac{1}{t^\omega}}$ be the modified algorithm often called a decreasing gain algorithm version of $\mathcal{A}_a$.

Let us state Theorem 5.4 more formally.

Theorem 9.1. *$\forall \mu > 0$, $\forall \lambda > 0$, we can construct algorithm $\mathcal{A}_{\frac{1}{t^\omega}}$ with initial value of the gain function set as $a_1 = \frac{1}{(1+t_1)^\omega}$ such that $\exists t_1$, $\exists T(\mu, \lambda) > 0$, $\forall t \geq T(\mu, \lambda)$,*

$$\mathbb{P}\left(\left|\left(\varphi_p\left(\hat{F}_t\right), 1 - \hat{F}_t\left(\varphi_p\left(\hat{F}_t\right)\right)\right) - \left(b^*(F), 1 - b^*(F)\right)\right| > 4\mu\right) \leq \lambda \qquad \forall F \in \mathcal{F}^\eta$$

and $T(\mu,\lambda)=\mathsf{O}\left(\frac{\left(\log\frac{1}{\lambda}\right)^{\frac{1}{1-\omega}}}{\mu^{\kappa}}\right)$ *for some* $\kappa\in(0,\infty)$ *where* $\hat{F}_t=\mathcal{A}_{\frac{1}{t^{\omega}}}(\mathcal{O}_t)$.

Proof. See Appendix C. □

The proof of Theorem 9.1 follows the same reasoning as the proof of Theorem 6.1 except for additional complications and technical steps to deal with the decreasing gain $a_t=1/(t+t_1)^{\omega}$.

10. Extensions

10.1. **Discounting.** We have assumed that the monopolist is infinitely patient. A natural question is whether the main result can be extended to the case with discounting. Suppose that the monopolist's payoff function is

$$\mathcal{U}^{\delta}(\mathcal{A},F)=-(1-\delta)\sum_{t=1}^{\infty}\mathbb{E}\left|(\varphi_{p,t}(\hat{F}_t),1-\hat{F}_t(\varphi_{p,t}(\hat{F}_t)))-(b^*(F),1-F(b^*(F)))\right|^2\delta^{t-1}$$

for some $\delta\in(0,1)$. It is immediate to define ϵ dominance for $\mathcal{U}^{\delta}(\mathcal{A},F)$:

Definition 10.1. $\mathcal{A}$ *is an* ϵ *dominant algorithm if* $\forall\epsilon>0$, $\exists\underline{\delta}\in(0,1)$ *such that* $\forall\delta\in(\underline{\delta},1)$, $\forall F\in\mathcal{F}^{\eta}$,

$$\mathcal{U}^{\delta}(\mathcal{A},F)\geq-\epsilon \tag{10.25}$$

Proposition 10.2. *Suppose that the monopolist discounts the future payoff.* $\mathcal{A}$ *is an* ϵ *dominant algorithm in* A *if* $\mathcal{A}$ *uniformly learns* $\mathcal{F}^{\eta}$.

Proof. See Appendix D. □

$\mathcal{A}_a$ constructed in Theorem 6.1 is uniformly learnable. Thus, the same $\mathcal{A}_a$ is an ϵ dominant algorithm if the monopolist discounts the future payoff but is sufficiently patient.

10.2. **Using General Restrictions.** The monopolist in our algorithm design problem is restricted to use polynomials to model demand—or, perhaps more precisely, to associate each optimal price and profit with a particular polynomial under which that price and profit would be optimal. One could derive analogous results with particular, tractable families for demand. But to allow *arbitrary* parameterizations, some added assumptions are needed to make sure that the parameterizations are sufficiently well-behaved so that the number of parameters has intrinsic meaning.

Consider a modification to the algorithm design problem where the monopolist is allowed to choose an algorithm with an arbitary set Θ, where each $\theta\in\Theta$. Since this need not have the structure of a particular set of demand curves, we will additionally impose that the algorithm is equipped with a *translating function* maps forecasts into predictions of optimal price and quantity (and thus profit):

$$\varphi:\Theta\to\mathbb{R}^2$$

The translation function takes each $\theta\in\Theta$ into the (human-readable) forecast about the optimal price p_t and the expected demand q_t at the optimal price (thus, the maximized

expected profit). For example, if $\mathcal{A}(\mathcal{D}_t)$ is the empirical distribution of F, then the translation function calculates the estimated optimal price and the expected quantity (thus, estimated expected profit). We are particularly interested in the case where each $\theta \in \Theta$ is associated with some $F \in \mathcal{F}^\eta$—and thus the predictions of price and quantity coincide with what would emerge under that F. This generalizes our main model, where the assumption was that parameters are associated with the coefficients of some polynomial demand curve.

We impose the following assumption on possible translation functions:

A2. Piecewise Lipschitz.
There exists a countable collection of sets $\mathcal{B}$ such that (i) $\cup_{B\in\mathcal{B}} B = \Theta$ and (ii) φ is Lipschitz continuous on each B.

Importantly, φ need not be continuous on Θ, as long as it is (Lipschitz) continuous on each B. This therefore covers the case where Θ is just the empirical observation up to some time T, as in the non-parametric algorithm of Cole and Roughgarden (2014).

Proposition 10.3. *If $\mathcal{A}$ satisfies* **A1** *and is a dominant algorithm that correctly predicts the optimal price and quantity using a specification Θ and a translating function satisfying* **A2**, *then* $\dim \mathcal{A}(\mathcal{D}_{t-1}) \geq 2$, *implying* $\text{comp}_T(\mathcal{A}) \geq 2\ \forall T \geq 1$. *Furthermore, both price and quantity data must be used for any such algorithm.*

Proof. See Appendix E. □

The idea of the proof is that, under a Lipschitz continuous mapping, the dimension of a set must be maintained. We specifically use the notion of the Hausdorff dimension, and note that since the Hausdorff dimension of a countable union of sets is simply the maximum of the Hausdorff dimension of any element, the Hausdorff dimension of Θ is equal to the Hausdorff dimension of some $B \in \mathcal{B}$. We then note that the dimension of the set of price-quantity pairs that might emerge for $F \in \mathcal{F}^\eta$ is two dimensional; thus, for Θ to be able to correctly predict any price and profit, it must be two-dimensional as well.

11. Concluding Remarks

Through a sequence of linear demand curves, the seller *can* still learn how to choose the optimal price at an exponential rate, *even though* the monopolist could make non-vanishing errors regarding the profit following other prices if the true demand curve were nonlinear. Under our algorithm, within a polynomial amount time, the monopolistic seller behaves as if he knows the actual demand curve. Even though a non-parametric estimation of the demand curve is feasible, as in Cole and Roughgarden (2014), the monopolist chooses to use a simple yet misspecified model of the demand curve to choose his price to save the computational cost.

In the monopolistic market, the buyer's decision problem is simple. The truthful revelation of his valuation is a dominant strategy, and the myopic decision to buy is an optimal choice. Even though the monopolist faces strategic buyers, the monopolist's decision problem is reduced to a single person optimization problem. An important extension of our exercise is to examine the duopoly market, where two firms complete through closely related products, say substitutes. The decision problem of each duopolist interacts with

each other. The choice of the model of the other firm's behavior is determined as an equilibrium outcome.

Appendix A. Proof of Proposition 5.8

If a PAC guaranteeing algorithm exists, a dominant algorithm must forecast the profit maximizing outcome accurately: $\forall F \in \mathcal{F}^\eta$,

$$\lim_{t\to\infty} (\varphi_p(\mathcal{O}_t), 1 - \hat{F}_t(\varphi_p(\mathcal{O}_t))) = (b^*(F), 1 - b^*(F))$$

with probability 1, where $\hat{F}_t = \varphi_p(\mathcal{O}_t)$. $\forall F \in S(R^0, X(R^0))$, the valuation of buyers is concentrated at a single value $v \in [\underline{p}, \overline{p}]$. Thus, the profit maximizing outcome must be $(v, 1)$. Choose any $F \in \mathcal{F}^\eta$ where the quantity associated with the optimal price of F is not 1. The algorithm's forecast does not converge to the true profit maximizing outcome with probability 1, contradicting the hypothesis that the algorithm is a dominant algorithm.

We have shown that for any dominant algorithm $\mathcal{A}$, $\exists \ell \geq 1$ such that $\mathcal{A}(\mathcal{O}_t) \in S(R^\ell, X(R^\ell))$. Thus, $\dim D_t \geq 2$ by Assumption 4.6.

Appendix B. Proof of Theorem 6.1

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.

B.1. **Preliminaries.** Fix $\tau > 0$ and consider an interval $[0, \tau)$ of real-time. Fix small $a > 0$, and divide the interval into subintervals of size a, with a possible exception of the last subinterval. Define

$$T(a) = \left\lceil \frac{\tau}{a} \right\rceil$$

be the number of the subintervals (treating the last subintervals as the full size subinterval) in $[0, \tau)$, where a is the gain coefficient of the updating term in the recursive formula. Recall that $\epsilon_{2,t}$ is distributed uniformly over $[-\epsilon, \epsilon]$ where $\epsilon > 0$. We will choose τ, a, ϵ to meet the algorithm's accuracy and confidence requirement, which in turn determines $T(a)$.

Recall that $b^*(F)$ is the optimal price of the true distribution F. Let $\beta^*(F)$ be the pair of estimators that induce the optimal price for F:

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

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

$$\beta_{T(a)} - \beta^*(F).$$

For $t \geq 1$, we can write the recursive formula as

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

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

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

As shown in (B.29), 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

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

Given $\beta_0, \beta_1, \ldots, \beta_{T(a)}$, define a continuous time process obtained by the linear interpolation: $\forall s \in [a(t-1), at)$,

$$\beta^a(s) = \frac{(s - a(t-1))\beta_t + (at - s)\beta_{t-1}}{a}.$$

Define

$$\beta(s) = \lim_{a \to 0} \beta^a(s)$$

pointwise. The existence of the limit point is guaranteed by the fact that β_t is contained in a compact set and $\overline{b}$ is a Lipschitz continuous function.

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

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

We can write

$$\begin{aligned}
\beta_{T(a)} - \beta^*(F) &= \beta_0 + a \sum_{t=1}^{T(a)} \overline{b}(\beta_{t-1}) + a \sum_{t=1}^{T(a)} \xi_t - \beta^*(F) \\
&= \beta_0 + \int_0^\tau \overline{b}(\beta(s))ds - \beta^*(F) && \text{(B.26)} \\
&\quad + a \sum_{t=1}^{T(a)} \overline{b}(\beta_{t-1}) - \int_0^\tau \overline{b}(\beta(s))ds && \text{(B.27)} \\
&\quad + a \sum_{t=1}^{T(a)} \xi_t && \text{(B.28)}
\end{aligned}$$

We examine (C.35), (C.36) and (C.37) one by one.

B.2. **Convergence and Stability.** We can write

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

where $\beta(0) = \beta_0$, which is often written as

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

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 meaning is clear from the context, 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):

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

where

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

and

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

Since $\epsilon_{1,t}$ has small support, and F is differentiable, it is more convenient to write

$$F\left(b_t + \epsilon_{1,t}\right) = F\left(b_t\right) + f\left(b_t\right)\epsilon_{1,t} + \mathsf{O}(\epsilon^2).$$

We are interested in the column vector

$$\mathbb{E}\begin{bmatrix}1\\ b_t+\epsilon_{1,t}\end{bmatrix}\phi(\beta_{t-1},\epsilon_t).$$

The first component is

$$1-F\left(b_t\right)-\beta_{0,t-1}-\beta_{1,t-1}b_t+\mathsf{O}(\epsilon^2).$$

The second component is

$$\left[-f\left(\frac{\beta_{0,t-1}}{2\beta_{1,t-1}}\right)-\beta_{1,t-1}\right]\sigma_1^2+\mathsf{O}(\epsilon^3).$$

We can write

$$\dot{\beta}=R^{-1}\begin{bmatrix}1-F\left(b\right)-\beta_0-\beta_1 b+\mathsf{O}(\epsilon^2)\\ b\left[1-F\left(b\right)-\beta_0-\beta_1 b\right]-\left[f\left(b\right)+\beta_1\right]\sigma_1^2+\mathsf{O}(\epsilon^3)\end{bmatrix}. \tag{B.29}$$

At the stationary point $b^*(F)$ where the right hand side of ODE vanishes,

$$\begin{aligned}1-F\left(b^*(F)\right) &= \frac{\beta_0}{2}-\mathsf{O}(\epsilon^2)\\ f\left(b^*(F)\right) &= -\beta_1+\frac{\mathsf{O}(\epsilon^3)}{\mathsf{O}(\epsilon^2)}.\end{aligned}$$

Thus,

$$\frac{1-F\left(b^*(F)\right)}{f\left(b^*(F)\right)}=b^*(F)+\mathsf{O}(\epsilon). \tag{B.30}$$

To simplify notation, let us ignore O term and treat $b^*(F)$ as the solution of

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

The uniqueness of the solution is implied by the increasing hazard rate property of F. Since $\forall F\in\mathcal{F}$, f is Lipschitz continuous

$$|f(x)-f(x')|\leq\eta|x-x'|,$$

$|\mathsf{O}(x)|\leq\eta|x|$. By reducing the size of the support of $\epsilon_{1,t}$, we can achieve the desired level of accuracy uniformly.

Let us proceed with the calculation after suppressing O terms. Note that

$$R^{-1}=\frac{1}{\sigma_1^2}\begin{bmatrix}b^2+\sigma_1^2 & -b\\ -b & 1\end{bmatrix}$$

We write ODE of $\beta=(\beta_0,\beta_1)$ without linear approximation error O as

$$\begin{bmatrix}\dot{\beta}_0\\ \dot{\beta}_1\end{bmatrix}=\begin{bmatrix}(1-F(b)-\beta_0-\beta_1 b)+b(f(b)+\beta_1)\\ -(f(b)+\beta_1)\end{bmatrix}.$$

By definition of b

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

at every moment of time. Thus,

$$\dot{\beta}_0+2\dot{\beta}_1 b+2\beta_1\dot{b}=0.$$

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

$$\dot{b}=-\frac{f(b)}{2\beta_1}\left[\frac{1-F(b)}{f(b)}-b\right]\equiv\mathsf{R}(b) \tag{B.31}$$

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

Lemma B.1. *$\exists c > 0$ such that*

$$\mathsf{R}(b) \geq -c(b - b^*(F)) \qquad b \leq b^*(F)$$

and

$$\mathsf{R}(b) \leq -c(b - b^*(F)) \qquad b \geq b^*(F).$$

Proof. We constructed the projection facility so that $\beta_1 < 0$ and $\beta_0 > 0$ and moreover,

$$\sup_{F \in \mathcal{F}^\eta} \beta_1 < 0.$$

If $\inf_{F \in \mathcal{F}^\eta} f(b) > 0$, then the proof is trivial. Since we only assume that $F \in \mathcal{F}^\eta$, we need more work.

Consider an iso-(expected) profit curve in the space of (q, p)

$$\Pi = pq.$$

Its slope is

$$-\left.\frac{dq}{dp}\right|_{\Pi} = \frac{1 - F(p)}{p}.$$

If $b = b^*(F)$, then the slope of the iso-profit curve must be equal to the slope of the demand curve $f(p)$ at $p = b^*(F)$:

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

Since $b^*(F) \in [\underline{v}, \overline{v}]$, the slope of an iso-profit curve at the optimal price must be uniformly bounded: $\exists \underline{M}, \overline{M} > 0$ such that

$$\underline{M} \leq f(b^*(F)) = \frac{1 - F(b^*(F))}{b^*(F)} \leq \overline{M}.$$

Since $F \in \mathcal{F}^\eta$ is uniformly Lipschitz continuous, for a sufficiently small $\epsilon > 0$,

$$f(b^*(F) - \epsilon) \geq \underline{M} - \eta\epsilon > \frac{\underline{M}}{2}$$

and similarly,

$$f(b^*(F) + \epsilon) \leq -\underline{M} + \eta\epsilon < -\frac{\underline{M}}{2}.$$

We prove that the right hand side of the ODE

$$\dot{b} = -\frac{f(b)}{\beta_1}\left(\frac{1 - F(b)}{f(b)} - b\right)$$

is strictly bounded away from 0 over $b < b^*(F) - \epsilon$ and $b > b^*(F) + \epsilon$.

If $f(b) = 0$, the increasing hazard rate property implies that $f(b') = 0 \ \forall b' < b$, and in particular $f(\underline{v}) = 0$. Since $f(b^*(F)) > 0$, $b < b^*(F)$. By the construction of the algorithm along the boundary, $\dot{b} = -\frac{1}{2\beta_1} > 0$ uniformly, because $\sup_{F \in \mathcal{F}^\eta} \beta_1 < 0$.

Suppose $f(b) > 0$ and $b \leq b^*(F) - \epsilon$. We know that $\mathsf{R}(b)$ defined in (C.38) is strictly decreasing because of the increasing hazard rate property. We also know that $\mathsf{R}(b^*(F) - \epsilon) \geq \frac{\underline{M}}{2}$. Thus,

$$\mathsf{R}(b) \geq \mathsf{R}(b^*(F) - \epsilon) \geq \frac{\underline{M}}{2} > 0.$$

Similarly, if $b \geq b^*(F) + \epsilon$,

$$\mathsf{R}(b) \leq \mathsf{R}(b^*(F) + \epsilon) \leq -\frac{\underline{M}}{2} < 0.$$

The increasing hazard rate property implies that if $b > b^*(F)$,

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

and if $b < b^*(F)$,

$$\frac{1 - F(b)}{f(b)} - b > -(b - b^*(F)).$$

Thus,

$$|\mathsf{R}(b) - \mathsf{R}(b^*(F))| \geq \frac{M}{2}|b - b^*(F)|$$

over $b \in [b^*(F) - \epsilon, b^*(F) + \epsilon]$. Since $b \in [\underline{v}, \overline{v}]$, $\exists c > 0$ such that

$$\mathsf{R}(b) - \mathsf{R}(b^*(F)) = R(b) \geq -c(b - b^*(F)) \qquad b \leq b^*(F)$$

and

$$\mathsf{R}(b) - \mathsf{R}(b^*(F)) = R(b) \leq -c(b - b^*(F)) \qquad b \geq b^*(F).$$

□

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

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

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

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

and

$$\overline{\tau}(\mu) \sim -\log \mu$$

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

B.3. **Riemann Residual.** Let us consider (C.36)

$$\mathbf{R}(a, F) = a\sum_{t=1}^{T(a)} \overline{b}(\beta_{t-1}) - \int_0^\tau \overline{b}(\beta(s))ds$$

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, the difference between the discrete value and the integration is at most $\frac{\eta' a^2}{2}$. Thus,

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

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

$$\overline{a} = \frac{2\mu}{\overline{\tau}(\mu)\eta'} \tag{B.32}$$

so that $\forall a \leq \overline{a}$, $\forall F \in \mathcal{F}$,

$$\mathbf{R}(a, F) \leq \mu. \tag{B.33}$$

Thus,

$$\overline{a} = \mathsf{O}\left(-\frac{\mu}{\log \mu}\right)$$

which implies

$$\frac{\tau(\mu)}{\overline{a}} = \mathsf{O}\left(\frac{(\log \mu)^2}{\mu}\right).$$

B.4. **Lower Bound of Confidence.** Next, we examine (C.37). Consider $\frac{\xi_t}{\xi}$, which is a martingale difference with

$$\left|\frac{\xi_t}{\xi}\right| \leq 1 \text{ and } \mathbb{E}\left(\frac{\xi_t}{\xi}\right)^2 \leq 1.$$

Since $\frac{\xi_t}{\xi}$ satisfies the large deviation property, $\forall \mu' > 0$, $\exists \rho(\mu', F) > 0$ (called the rate function) such that

$$\mathbb{P}\left(\left|\frac{1}{T}\sum_{t=1}^{T}\frac{\xi_t}{\xi}\right| > \mu'\right) \leq e^{-\rho(\mu',F)T}.$$

We need a uniform rate function $\rho(\mu', F) > 0$ over $\mathcal{F}$. By Azuma-Hoeffding-Bennett inequality (Corollary 2.4.7 in Dembo and Zeitouni (1998)), we have $\forall \mu' \in (0, 1/2)$,

$$\mathbb{P}\left(\left|\frac{1}{T}\sum_{t=1}^{T}\frac{\xi_t}{\xi}\right| > \mu'\right) \leq e^{-2T\mathsf{H}\left(\mu'+\frac{1}{2}\,|\,\frac{1}{2}\right)}$$

where

$$\mathsf{H}(p \mid p_0) = p\log\frac{p}{p_0} + (1-p)\log\frac{1-p}{1-p_0} \tag{B.34}$$

for $p, p_0 \in (0,1)$. Let

$$T(a,\mu) = \left\lceil \frac{\overline{\tau}(\mu)}{a} \right\rceil.$$

Then,

$$\mathbb{P}\left(\left|a\sum_{t=1}^{T(a,\mu)}\xi_t\right| < \overline{\tau}(\mu)\xi\mu'\right) \leq e^{-2T\mathsf{H}\left(\mu'+\frac{1}{2}\,|\,\frac{1}{2}\right)}.$$

Let

$$\mu' = \frac{\mu}{\overline{\tau}(\mu)\xi}.$$

After substitution, we have

$$\mathbb{P}\left(\left|\frac{1}{T(a,\mu)}\sum_{t=1}^{T(a,\mu)}\xi_t\right| > \mu\right) \leq e^{-2T(a,\mu)\mathsf{H}\left(\mu'+\frac{1}{2}\,|\,\frac{1}{2}\right)}.$$

Since $\overline{\tau}(\mu) = \mathsf{O}(-\log\mu)$,

$$\frac{\mu}{\overline{\tau}(\mu)\xi} = \mathsf{O}(\mu)$$

and therefore, $T(a,\mu)$ increases at the polynomial speed as $\mu \to 0$. Let

$$\rho = 2\mathsf{H}\left(\mu' + \frac{1}{2} \mid \frac{1}{2}\right) > 0.$$

B.5. **Combine the Pieces.** Fix $\mu > 0$. Recall that we assume that $\epsilon_{1,t}$ is distributed over $[-\epsilon, \epsilon]$. We first choose $\epsilon > 0$ so that the stationary solution (B.30) of ODE is within μ neighborhood of $\beta^*(F)$ $\forall F \in \mathcal{F}$.

Since $\forall f \in \mathcal{F}$

$$|f(p) - f(p')| \leq \eta|p - p'|.$$

The Taylor residual is bounded by $\eta\epsilon$ for small $\epsilon > 0$:

$$|\mathsf{O}(\epsilon^3)| \leq |\mathsf{O}(\epsilon^2)| \leq \eta\epsilon^2 \qquad \forall t \geq 1.$$

Note that the last term is independent of $F \in \mathcal{F}$. Choose $\epsilon > 0$ sufficiently small so that

$$\left|\beta^s - \beta^*(F)\right| < \mu \qquad \forall F \in \mathcal{F}$$

where β^s solves (B.30). We chose $\overline{\tau}(\mu)$ so that

$$|\beta(\overline{\tau}(\mu)) - \beta^s| < \mu.$$

Given $\mu > 0$, we chose $a_1 > 0$ in (B.32) so that the Riemann residual $\mathsf{R}(a, F)$ satisfies (B.33).

By the construction, $\forall a \in (0, a_1)$,

$$\mathbb{P}\left(\left|\frac{1}{T(a,\tau)}\sum_{t=1}^{T(a,\tau)}\xi_t\right| > \mu\right) < e^{-T(a,\tau)\rho}.$$

Combining these results, we have $\forall a \in (0, a_1)$,

$$\mathbb{P}\left(\left|\beta_{T(a,\tau)} - \beta^*(F)\right| \geq 4\mu\right) \leq e^{-T(a,\tau)\rho}$$

where $T(a, \tau)$ increases linearly with respect to $1/a$ and at the polynomial speed with respect to $1/\mu$.

So far, we have proved a result which is slightly weaker than Theorem 6.1. Let us summarize what we have at this point.

Proposition B.2. *$\forall \mu > 0$, $\exists \tau(\mu) > 0$, $\rho > 0$ and $\overline{a} > 0$ such that $\forall a \in (0, \overline{a})$, if $T = \lceil \frac{\tau(\mu)}{a} \rceil$ $\forall a \in (0, \overline{a})$,*

$$\mathbb{P}\left(\left|\varphi_p\left(\mathcal{A}_a(\mathcal{O}_{T(a,\mu)})(1 - F(\mathcal{A}_a(\mathcal{O}_{T(a,\mu)}))\right) - b^*(F)(1 - F(b^*(F)))\right| > 4\mu\right) \leq e^{-\rho T(a,\mu)} \qquad \forall F \in \mathcal{F}^\eta$$

and $\tau(\mu) \sim -\log \mu$ for small $\mu > 0$ and $\overline{a} = \mathsf{O}\left(-\frac{\mu}{\log \mu}\right)$.

B.6. **Final Step.** For fixed $\lambda > 0$ less than 1, we can choose $\overline{a} > 0$ sufficiently small so that

$$e^{-\frac{\rho\tau(\mu)}{\overline{a}}} = \lambda$$

and therefore,

$$\frac{\tau(\mu)}{\overline{a}} = -\frac{\log \lambda}{\rho}.$$

Since

$$T(\overline{a}, \mu) = \left\lceil \frac{\tau(\mu)}{\overline{a}} \right\rceil = \left\lceil -\frac{\log \lambda}{\rho} \right\rceil,$$

$T(\overline{a}, \mu)$ increases at the logarithmic speed with respect to $1/\lambda$. An important observation is that the estimator's accuracy depends only on $\tau(\mu)$, and the approximation error vanishes at the linear rate of $a \to 0$. Thus, the minimum number of the time steps to satisfy the accuracy requirement increases at the rate of $-\frac{\log \mu}{a}$. Proposition B.2 implies that $\mathcal{A}_a$ uniformly learns $\mathcal{F}^\eta$.

Appendix C. Proof of Theorem 9.1

To make the paper self-contained, we present the proof, although we are essentially replicating the proof of Theorem 6.1.

C.1. **Preliminaries.** Following Kushner and Yin (1997), we construct the (fictious) time from the gain function $a_t = \frac{1}{(t+t_1)^\omega}$ where t_1 is selected to satisfy the accuracy requirement. Recall that ϵ is the size of $\epsilon_{1,t}$ which the monopolist adds to b_t to experiment. We change ϵ to decreasing. Let ϵa_t be the support of $\epsilon_{1,t}$. Since $\omega \in (0, 1)$, $\sum_{t=1}^{\infty} a_t = \infty$. 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

$$m(\tau) = K$$

defined above. We refer to $\mathbb{R}_+$ as the fictitious time or simple, the time, and $\{1, 2, 3, \ldots\}$ as the number of periods or rounds. 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

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

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

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

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

The convergence result follows from the stochastic approximation results (Kushner and Yin (1997)). In addition, we need to prove that the convergence is uniform ove $\mathcal{F}^\eta$.

For $t \geq 1$, we can write the recursive formula as

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

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

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

As shown in (B.29), 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 $\bar{b}(\beta_{t-1})$ in place of $\bar{b}_{t-1}(\beta_{t-1})$, dropping the time subscript from $\bar{b}_{t-1}$. We can write the recursive formula as

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

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

We can write

$$\begin{aligned}
\beta_{m(K+\tau)} - \beta^*(F) &= \beta^K(0) + \sum_{t=t_K}^{m(K+\tau)} a_t \bar{b}(\beta_{t-1}) + \sum_{t=t_K}^{m(K+\tau)} a_t \xi_t - \beta^*(F) \\
&= \beta^K(0) + \int_0^\tau \bar{b}(\beta(s))ds - \beta^*(F) && \text{(C.35)} \\
&\quad + \sum_{t=t_K}^{m(K+\tau)} a_t \bar{b}(\beta_{t-1}) - \int_0^\tau \bar{b}(\beta(s))ds && \text{(C.36)} \\
&\quad + \sum_{t=t_K}^{m(K+\tau)} a_t \xi_t && \text{(C.37)}
\end{aligned}$$

We examine (C.35), (C.36) and (C.37) one by one.

C.2. **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 often written 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 meaning is clear from the context, 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):

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

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 following the same argument as in Section C.1,

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

at every moment of time. 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

$$\dot{b} = -\frac{f(b)}{2\beta_1}\left[\frac{1 - F(b)}{f(b)} - b\right] \equiv \mathsf{R}(b) \tag{C.38}$$

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

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

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

Invoking Lemma B.1, 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

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

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

$$\overline{\tau}(\mu) = \sup_{(\beta_0(0), \beta_1(0)) \in \mathsf{B}} \tau(\mu) < \infty \tag{C.39}$$

and

$$\overline{\tau}(\mu) \sim -\log \mu \tag{C.40}$$

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

By the definition,

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

We use $\simeq$ instead of $=$, because of the truncation error, which vanishes as $K \to \infty$.

We need to calculate how $m(K+\tau)$ changes with respect to τ for a large K. Thus, for a large K, $m(K+\tau)$ is approximated by x solving

$$\tau = \int_{t_K}^{x} \frac{1}{s^\omega} ds$$

which implies

$$x = \left[(1-\omega)\tau + t_K^{1-\omega}\right]^{\frac{1}{1-\omega}} = \mathsf{O}(\tau^{\frac{1}{1-\omega}}, t_K) = m(K+\tau).$$

Similar calculation shows that $t_K = \mathsf{O}(K^{\frac{1}{1-w}})$. Thus,

$$m(K+\tau) = \mathsf{O}\left(\tau^{\frac{1}{1-\omega}}, K^{\frac{1}{1-w}}\right). \tag{C.41}$$

As $\mu \to 0$, the amount of data to meet the accuracy requirement increases at the rate of $(-\log \mu)^{\frac{1}{1-\omega}}$ which slower than the polynomial rate of m.

C.3. **Riemann Residual.** Fix $\mu > 0$ and $\tau = \overline{\tau}(\mu)$ defined in (C.39). Let us consider (C.36)

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

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

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

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

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

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

$$\frac{1}{t_K^\omega} = \frac{2\mu}{\overline{\tau}(\mu)\eta'} \tag{C.42}$$

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

$$\mathbf{R}(t_K, \tau, F) \le \mu. \tag{C.43}$$

Thus,

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

or equivalently,

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

combined with (C.41), we conclude that $m(K + \overline{\tau}(\mu)) - t_K$ increases at the polynomial rate of $\frac{1}{\mu}$.

C.4. **Lower Bound of Confidence.** Next, we examine (C.37). Let $a_t = \frac{1}{t^\omega}$ where $0 < \omega < 1$. Fix K so that t_K satisfies (C.42). Then, choose $\overline{\tau}(\mu)$ according to (C.39). We need to show that $\exists \mathsf{h} > 0$ such that $\forall T \ge m(K + \overline{\tau}(\mu))$

$$\mathbb{P}\left(\left|\sum_{t=t_K}^{T} a_t \xi_t\right| \ge \mu\right) \le e^{-(T-t_K)\mathsf{h}}.$$

We first prove following inequality, for all $a_t \in [0,1]$ and $\lambda \ge 0$:

Lemma C.1.

$$\left(\frac{e^{\lambda}+e^{-\lambda}}{2}\right)^{a_t} \geq \frac{e^{\lambda a_t}+e^{-\lambda a_t}}{2} \tag{C.44}$$

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

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

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 is non-negative. Therefore, at any $a_t \in [0,1]$, we have the right hand side of the inequality (which is convex) is 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).[19] We have

$$\mathbb{E}[e^{\lambda \sum_{t=1}^{T} a_t \xi_t/\xi}] \leq \prod_{t=1}^{T} \frac{e^{-a_t\lambda}+e^{a_t\lambda}}{2} \leq \left(\frac{e^{\lambda}+e^{-\lambda}}{2}\right)^{\sum_{t=1}^{T} a_t} \tag{C.45}$$

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

$$\mathbb{P}\left(\frac{1}{\sum_{t=1}^{T} a_t}\left|\sum_{t=1}^{T} a_t\xi_t/\xi\right| \geq \mu\right) \leq e^{-\lambda\mu\sum_{t=1}^{T} a_t}\mathbb{E}[e^{\lambda\sum_{t=1}^{T} a_t\xi_t/\xi}] \leq e^{-\lambda\mu\sum_{t=1}^{T} a_t}\left(\frac{e^{\lambda}+e^{-\lambda}}{2}\right)^{\sum_{t=1}^{T} a_t}.$$

The first inequality follows from the exponential Chebyshev inequality. The second inequality follows from (C.45).

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

$$\mathbb{P}\left(\frac{1}{\sum_{t=1}^{T} a_t}\left|\sum_{t=1}^{T} a_t\xi_t/\xi\right| \geq \mu\right) \leq e^{-\left(\sum_{t=1}^{T} a_t\right)\mathsf{H}\left(\frac{1+\mu}{2}\,|\,\frac{1}{2}\right)}. \tag{C.46}$$

where H is the relative entropy of the binomial distribution defined in (B.34). Set $\mathsf{h} = \mathsf{H}\left(\frac{1+\mu}{2} \mid \frac{1}{2}\right)$. Recall that $a_t = \frac{1}{t^{\omega}}$, so that $\sum_{t=1}^{T} \frac{1}{t^{\omega}}$. We can bound this sum via an integral, to obtain

$$\sum_{t=t_K}^{T} \frac{1}{t^{\omega}} \geq \frac{1}{t_K^{\omega}} + \int_{t_K}^{T} \frac{1}{\tau^{\omega}} d\tau > \frac{T^{1-\omega} - t_K^{1-\omega}}{1-\omega}.$$

Therefore, replacing $\sum_{t=t_K}^{T} a_t$ with $\frac{T^{1-\omega}-t_K^{1-\omega}}{1-\omega}$ makes the right hand side of (C.46) larger. Putting this together, we finally obtain

$$\mathbb{P}\left(\frac{1}{\sum_{t=1}^{T} a_t}\left|\sum_{t=K}^{T} a_t\xi_t\right| \geq \xi\mu\right) \leq e^{-\left(\frac{T^{1-\omega}-t_K^{\omega}}{1-\omega}\right)H\left(\frac{1+\mu}{2}\,|\,\frac{1}{2}\right)}.$$

Suppose that $1-\rho$ is the lower bound of the admissible confidence level. To satify the confidence requirement, the right hand side must satisfy

$$e^{-\left(\frac{T^{1-\omega}-t_K^{\omega}}{1-\omega}\right)H\left(\frac{1+\mu}{2}\,|\,\frac{1}{2}\right)} \leq \rho.$$

In particular, the inequality must hold for $T = m(K + \overline{\tau}(\mu))$, from which

$$T = \mathsf{O}\left((-\log\rho)^{\frac{1}{1-\omega}}\right)$$

[19] Replace λ with $a_t\lambda$ to establish (2.4.8). Doing so (and setting $v = 1$), and applying the previous inequality.

follows.

Appendix D. Proof of Proposition 10.2

Suppose that $\mathcal{A}$ is uniformly learnable. To show that $\mathcal{A}$ is an ϵ dominant strategy, fix $\epsilon > 0$. We have to show $\exists \underline{\delta} \in (0,1)$ such that $\forall \delta \in (\underline{\delta}, 1)$, (10.25) holds. Choose $\lambda = \frac{\epsilon}{2} < \epsilon$. Since $\mathcal{A}$ is uniformly learnable, $\exists T$ such that

$$\mathbb{E}\left|\varphi_p(\mathcal{A}(\mathcal{O}_T))(1 - F(\varphi_p(\mathcal{A}(\mathcal{O}_T)))) - b^*(F)(1-b^*(F))\right| \leq \lambda \qquad \forall F \in \mathcal{F}^\eta. \tag{D.47}$$

Thus,

$$\mathcal{U}^\delta(\mathcal{A}, F) \geq (1-\delta^T)\cdot 0 + \delta^T\left(b^*(F)(1-b^*(F)) - \lambda\right).$$

Note that $\sup_{F\in\mathcal{F}^\eta} b^*(F)(1-b^*(F)) < \infty$. Since $\lambda = \frac{\epsilon}{2} < \epsilon$, we can choose $\underline{\delta} < 1$ sufficiently close to 1 so that

$$\underline{\delta}^T\left(b^*(F)(1-b^*(F)) - \lambda\right) \geq b^*(F)(1-b^*(F)) - \epsilon \qquad \forall F \in \mathcal{F}^\eta.$$

Thus, $\forall \delta \in (\underline{\delta}, 1)$,

$$\mathcal{U}^\delta(\mathcal{A}, F) \geq b^*(F)(1-b^*(F)) - \epsilon \qquad \forall F \in \mathcal{F}^\eta$$

as desired.

Appendix E. Proof of Proposition 10.3

We first prove that $\dim D_t \geq 2$. Since the market outcome in period t is (p_t, q_t), we show that if the algorithm uses only one of the two elements, the algorithm cannotbe dominant. We only show that this holds for $D_t = \{p_t\}$. The proof for the case where $D_t = \{q_t\}$ follows a symmetric argument.

Let F and F^α be the distribution functions of the uniform distribution over intervals $[\underline{p}, \underline{p}+1]$ and $[\underline{p}+\alpha, \underline{p}+\alpha+1]$. Note that the optimal price of F^α is

$$b^*(F^\alpha) = \frac{1+\underline{p}+\alpha}{2}.$$

Suppose that $D_t = \{p_t\}$. Then,

$$\mathcal{A}(\mathcal{D}_t : F^\alpha) = \mathcal{A}(\mathcal{D}_t : F) \qquad \forall \alpha.$$

Both $\mathcal{A}(\mathcal{D}_t : F^\alpha)$ and $\mathcal{A}(\mathcal{D}_t : F)$ have the same input and, therefore, generate the identical forecast for any F^α.

$$\varphi_p(\mathcal{A}(\mathcal{D}_t : F^\alpha)) = \varphi_p(\mathcal{A}(\mathcal{D}_t : F)) \qquad \forall \alpha, \forall t \geq 1. \tag{E.48}$$

Since $\mathcal{A}$ is a dominant algorithm $\mathcal{F}$,

$$\lim_{t\to\infty} \varphi_p(\mathcal{A}(\mathcal{D}_t : F^\alpha)) = b^*(F^\alpha) = \frac{1+\underline{p}+\alpha}{2} \qquad \forall \alpha$$

with probability 1. If so, (E.48) cannot hold.

Next, we show that $\dim(\mathcal{A}(\mathcal{D}_{t-1})) \geq 2$. Suppose that $\dim(\mathcal{A}(\mathcal{D}_{t-1})) = 1$. Let

$$\theta_t = \mathcal{A}(\mathcal{D}_t) \in \mathbb{R}$$

by the hypothesis of the proof. Since Θ is compact, $\{\theta_t\}$ has a convergent subsequence. After re-numbering the subsequence, let

$$\theta_t \to \theta^* \in \Theta.$$

Since $\mathcal{A}$ is a dominant algorithm, $\forall F \in \mathcal{F}$,

$$\theta_t = \mathcal{A}(\mathcal{D}_t : F) \to \theta^*$$

implies that

$$\varphi(\theta_t) \to \varphi(\theta^*) = (b^*(F), q^*(F)).$$

We can therefore consider the following set:

$$\{(a,b) \mid \exists \theta \in \Theta \text{ such that } a = \varphi_p(\theta), b = \varphi_q(\theta)\} \tag{E.49}$$

Note that every $\theta \in \Theta$ is an element of a set B which is single-dimensional by hypothesis (i.e., Θ is the union of all such sets B). The Lipschitz requirement on φ implies that the Hausdorff dimension (see,

e.g., Ambrosio and Tilli (2004))[20] of the image of each B under φ cannot be more than the Hausdorff dimension of B. To see this, note that if $\mathcal{H}^k(A)$ is the k-dimensional Hausdorff measure of a set A and c_B is a Lipschitz constant of φ on the set B, then Proposition 3.1.4 of Ambrosio and Tilli (2004) states that $\mathcal{H}^k(\varphi(B)) \leq c_B^k \mathcal{H}^k(B)$. In particular, since any B is (at most) one-dimensional (being a subset of a one-dimensional space), the right hand side is 0 for every B and every $k > 1$, and therefore the left hand side is also 0 for every $k > 1$. Now, if the Hausdorff measure of every set in a countable collection is 0, then the Hausdorff measure of the union of all sets is 0 as well. Furthermore, Θ is the countable union of the set of B, where each B has k-dimensional Hausdorff measure 0 for any $k > 1$. Thus, the image of Θ under φ must have k-dimensional Hausdorff measure 0 for any $k > 1$, since this holds for every $B \in \mathcal{B}$ and since the image of Θ under φ is equal to the union of the image of every $B \in \mathcal{B}$ under φ. Thus, the k-dimensional Hausdorff measure of $\varphi(\Theta)$ must be 0 for $k > 1$, and therefore the image of Θ under φ cannot have Hausdorff dimension greater than 1.

Since $\mathcal{A}$ is a dominant algorithm $\mathcal{F}$, (E.49) is exactly

$$\{(a, b) \mid \exists F \in \mathcal{F} \text{ such that } a = b^*(F), b = 1 - F(b^*(F))\} \tag{E.50}$$

To obtain the contradiction, it suffices to observe that (E.50) is in fact not one dimensional. This can be seen using linear demand curves (or the uniform distribution function F) which is a subset of $\mathcal{F}$. That is, if $q = b_0 + \beta_1 p$, then $q^* = \frac{\beta_0}{2}$ and $p^* = -\frac{\beta_0}{2\beta_1}$. Since for every fixed β_0, (a) the set of (q^*, p^*) generated by some β_1 is a one dimensional set, (b) these sets are also disjoint for distinct β_0, and (c) the set of valid β_0 is itself one dimensional, we therefore have that the set in (E.50) is two dimensional for this class of $\mathcal{F}$. This contradiction establishes that there cannot be a single-dimensional parameter which is used in the 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.

Ambrosio, L., and P. Tilli (2004): *Topics on Analysis in Metric Spaces*, vol. 25 of *Oxford Lecture Series in Mathematics and Its Applications*. Oxford University Press.

Bergemann, D., and S. Morris (2005): "Robust Mechanism Design," *Econometrica*, 73(6), 1771–1813.

Besbes, O., and A. Zeevi (2015): "On the (Surprising) Sufficiency of Linear Models for Dynamic Pricing with Demand Learning," *Management Science*, 61(4), 723–739.

Carroll, G. (2015): "Robustness and Linear Contracts," *American Economic Review*, 105(2), 536–563.

——— (2017): "Robustness and Separation in Multidimensional Screening," *Econometrica*, 85(2), 453–488.

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

——— (2017): "Gresham's Law of Model Averaging," *American Economic Review*, 107(11), 1–29.

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.

Du, S. (2018): "Robust Mechanisms Under Common Valuation," *Econometrica*, 86(5), 1569–1588.

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.

[20]Briefly, the Hausdorff dimension of a set is the supremum of the set of $d \geq 0$ such that the $d-$dimensional Hausdorff measure is infinite. Roughly speaking, the d-dimensional Hausdorff measure of a set is obtained by finding the minimal volume of an open cover of the set by balls of radius no more thanδ, treating the volume of each as proportional to δ^d (and the total volume as the sum of the volume of all of the balls), and taking $\delta \to 0$. Equivalently, it is the infimum of the set of $d \geq 0$ such that the $d-$dimensional Hausdorff measure is 0. Note that it is immediate from this latter definition that the Hausdorff dimension of a countable collection of sets cannot become larger by taking their union.

Epstein, L. G., and M. Schneider (2007): "Learning under Ambiguity," *Review of Economic Studies*, 74(4), 1275–1303.

Esponda, I., and D. Pouzo (2014): "An Equilibrium Framework for Players with Misspecified Models," University of Washington and University of California, Berkeley.

Esponda, I., D. Pouzo, and Y. Yamamoto (2021): "Asymptotic Behavior of Bayesian Learners with Misspecified Models," *Journal of Economic Theory*.

Evans, G. W., and S. Honkapohja (2001): *Learning and Expectations in Macroeconomics*. Princeton University Press.

Frick, M., R. Iijima, and Y. Ishii (2021): "Belief Convergence under Misspecified Learning: A Martingale Approach," Discussion paper, Yale University and Pennsylvania State University.

Fudenberg, D., G. Lanzani, and P. Strack (2021): "Limit Points of Endogenous Misspecified Learning," *Econometrica*.

Goncalves, D., and B. Furtado (2020): "Statistical Mechanism Design: Robust Pricing and Reliable Projections," Discussion paper, Columbia University.

Gonczarowski, Y. A., and S. M. Weinberg (2021): "The Sample Complexity of Up-to-ϵ Multi-dimensional Revenue Maximization The Sample Complexity of Up-to-ϵ Multi-dimensional Revenue Maximization," *Journal of the ACM*, 68(3).

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

Hansen, L. P., and T. J. Sargent (2020): "Structured ambiguity and model misspecification," *Journal of Economic Theory*, p. 105165.

Heidhues, P., B. Köszegi, and P. Strack (2018): "Unrealistic Expectations and Misguided Learning," *Econometrica*, 86(4), 1159–1214.

Huang, Z., Y. Mansour, and T. Roughgarden (2018): "Making the Most of Your Samples," *SIAM Journal on Computing*, 47(3), 651–674.

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

Libgober, J., and X. Mu (2021): "Informational Robustness in Intertemporal Pricing," *Review of Economic Studies*, 88(3), 1224–1252.

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

Nisan, N., E. Tardos, and V. V. Vazirani (eds.) (2007): *Algorithmic Game Theory*. Cambridge University Press.

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

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

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

Schapire, R. E., and Y. Freund (2012): *Boosting: Foundations and Algorithms*. MIT Press.

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

Simon, H. (1987): *Models of Man*, Continuity in Administrative Science. Ancestral Books in the Management of Organizations. Garland Publisher.

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/