LEARNING UNDERSPECIFIED MODELS

IN-KOO CHO AND JONATHAN LIBGOBER

ABSTRACT. This paper examines whether one can learn to play an optimal action while only knowing part of the true specification of the environment. We choose the monopoly market as our laboratory, where the monopolist is endowed with an *underspecified* model of the market demand, facing model uncertainty. In contrast to the conventional studies on model uncertainty (e.g., Hansen and Sargent (2007)), the monopolist can observe market outcomes and learn about market demand. In contrast to existing learning models, which typically fix the specification of the monopolist's model exogenously (e.g., Evans and Honkapohja (2001)), the monopolist learns to choose a specification of the demand curve. We formulate the learning dynamics as an algorithm that forecasts the optimal price based on the data. We assume that the monopolist has a lexicographic preference over the algorithm's payoff and complexity cost, seeking an algorithm with a minimum number of parameters while achieving the maximum payoff (Rubinstein (1986)). Inspired by PAC learnability (Shalev-Shwartz and Ben-David (2014)), we develop a new notion of learnability by requiring that the algorithm must produce an accurate forecast with a reasonable amount of data *uniformly* over the set of underspecified models. We show that for the set of demand curves with strictly decreasing uniformly Lipschitz continuous marginal revenue curve, the optimal algorithm recursively estimates the slope and the intercept of the 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 underspecified demand curves.

KEYWORDS. Optimal Sales Mechanism, Underspecified model, Uniform Learnability, 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)). Yet, it is an excessively strong assumption that a decision maker is fully aware of *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 true 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 true 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 the model uncertainty.

We choose the monopolistic profit maximization problem (Myerson (1981)) as our laboratory. Our exercise has four features, which together differentiate our research from the existing studies. First, in contrast to the Bayesian models, the monopolist lacks any information to form a prior over the set of possible demand curves. The model uncertainty motivates the robust approach in Hansen and Sargent (2007) and Morris and Bergemann (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 Morris and Bergemann (2005)).[1] In our case, the seller has an opportunity to observe data to learn about the demand curve. In conventional learning models, the specification of the model is exogenously fixed, and the algorithm calculates the unknown parameters. In our exercise, the seller learns the demand curve's specification and parameters to learn from the data.

Third, the robust approach focuses on the best response against the worst possible conjecture, hoping that the strategy can perform "reasonably well" over the admissible set of demand curves. In our case, the learning algorithm must produce a good approximation to the true optimal strategy with high confidence *uniformly* over the set of admissible demand curves with a reasonable amount of data. If so, we say that the algorithm is uniformly learnable. Our notion of learnability is closely related to the PAC criterion in the machine learning literature (Shalev-Shwartz and Ben-David (2014)). Given $\epsilon > 0$ error bound and $1 - \rho > 0$ confidence requirement, we search for a learning algorithm $\mathcal{A}$ and a stopping time $T(\epsilon, \rho)$ so that after $T(\epsilon, \rho)$ time steps, algorithm $\mathcal{A}$ produces a forecast that is within ϵ neighborhood of the true optimal strategy with probability $1 - \rho$ uniformly over the set of admissible demand curves. Like the PAC criterion, uniform learnability disciplines data complexity. The number $T(\epsilon, \rho)$ of the time steps should increase at the rate of a polynomial function of $1/\epsilon$ and the logarithmic rate of $1/\rho$.[2]

Fourth, the monopolist in our exercise bears the computational cost. Absent these costs, an existing proposal from Cole and Roughgarden (2014) achieves a PAC guarantee by non-parametrically estimating the demand curve from the buyer's reported valuation in the revelation game. In this paper, we consider two sources of computational cost: the source of data and the complexity of an algorithm. In the revelation game, the buyer's private information is freely available to the algorithm. While the direct mechanism is a widely used mathematical model, an actual trading protocol is rarely the revelation game. Instead, one has to invert a buyer's strategy to infer his valuation, which is not a trivial exercise. We assume it is cheaper to obtain public data than the reported private information of a buyer. We measure the complexity of an algorithm by the number of parameters an algorithm has to estimate.[3] Following Rubinstein (1986), we assume that the monopolist has a lexicographic preference over the payoff and the complexity cost. The

---

[1] A remarkable exception is a series of papers by Hansen and Sargent (e.g., Hansen and Sargent (2020)) on robust learning models where the authors introduce the possibility of learning while maintaining the uncertainty aversion.

[2] We defer a discussion of the precise difference between uniform learnability and PAC to Section 8.

[3] 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)).

seller first seeks a uniformly learnable algorithm to maximize the expected profit and then chooses the simplest learning algorithm among those that achieve the maximum payoff.

We show that a uniformly learnable algorithm must estimate at least two parameters (Proposition 4.3). We construct an algorithm estimating two parameters that learns uniformly the optimal strategy of the monopolist over the set of demand curves with strictly decreasing uniformly Lipschitz continuous marginal revenue curve. The algorithm recursively estimates a linear demand curve (which can be parameterized by the slope and the intercept) using price and quantity, learning uniformly the optimal strategy of the monopolist (Theorem 5.1). Within a "reasonable amount" of time, the monopolistic seller behaves as if he knows the actual demand curve.

We point out the 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 paramters 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 possibly, the maximum profit, not the demand curve. The choice of the specification of the model for the demand curve should be a part of the profit maximzing decision of the monopolist. If the monopolist bears the computational cost, the chosen specification should be the simplest one among those which guarantee the profit maximizing price with a probability close to 1. The monopolist chooses a misspecified model (such as a linearity of the demand curve). 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. In this sense, a misspecified model would be a representation of the procedural rationality of the seller (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. In Section 3, we formally describe the problem and define the basic concepts, with a summary of the main result. Section 4 describe the research strategy. The construction of the algorithm involves technical steps that obscure the main feature of the algorithm. Instead, we take a little detour by constructing a "weaker" algorithm and shows that the weaker algorithm satisfies a "weaker" version of PAC guaranteeing property (called the uniform learnability). After analyzing the weaker algorithm satisfying the uniform learnability, we modify the algorithm to obtain the main result. The weaker notion of convergence is vital for extending our results. In Section 5, we construct the simplest algorithm among algorithms that uniformly learn

the optimal strategy. We state the main result, whose proof is in the appendix. Instead of a formal proof, Section 6 illustrates how the monopolist can efficiently and uniformly learn the optimal price of a non-linear demand curve through a linear demand curve. Section 7 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 8, we first modify the uniformly learnable algorithm to PAC guarantee the optimal price. We focus on the case where the monopolist is infinitely patient so that the objective function is the long run average payoff. In Section 8.2, we extend our result to the case where the monopolist discounts the future payoff. Section 9 concludes the paper.

2. Literature Review

Our paper differs from three main approaches to investigating the decision problem with underspecified models. As discussed above, the most prominent approach is that a decision-maker 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 typically differ from the optimal solution against the truth, sometimes significantly. While data could, in principle, bring the decision-maker closer to optimality, typically, these models are silent on how the decision-maker 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 decision-maker to learn from data, the specification is fixed exogenously and often excludes the true 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)).[4] 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, PAC guaranteeing the optimal price. The linear model, which is misspecified, is not imposed by the modeler but derived from optimization by the monopolist.

This paper follows a third approach initiated by machine learning models[5] (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]In the operations research, Besbes and Zeevi (2015) examined the monopoly profit maximization problem, where the monopolist is endowed with a linear demand curve, estimating the demand curve to choose the price. Besbes and Zeevi (2015) demonstrated that one could construct an algorithm to let the monopolist learn the actual optimal price for a class of demand curves satisfying a set of regularity conditions.

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

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

Following Cole and Roughgarden (2014), we freely borrow the 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 the memory, he will search for an algorithm that uses a simpler specification of the demand curve.

3. Description

3.1. **Demand.** There are N buyers, each of whom is indexed by $i \in \{1, \ldots, N\}$ and is endowed with reservation value $v_i \in [\underline{v}, \overline{v}]$. 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

$$F(p) = \frac{1}{N} \sum_{i=1}^{N} F_i(p) \quad \text{and} \quad \epsilon_2 = q - (1 - F(p)), \tag{3.1}$$

where $\mathbb{E}\epsilon_2 = 0$.

We interpret $1 - F(p)$ as the expected quantity of sales and call the (expected) demand curve. The actual amount q of sales as a random variable, whose expected value is $1 - F(p)$ if the price is p. If $p = 0$, $q = 1$ with probability 1, and $F(p) \to 1$ as $p \to \overline{v}$. We can treat F as a distribution function. Let f be the density function of F.

3.2. **What the Monopolist Knows.** In the beginning, the complete specification F is not available to the monopolist. Consequently, the specification of the expected profit

$$p(1 - F(p))$$

is not fully known to the monopolist. Since the monopolist does not have a prior over the set of feasible distributions over the valuations, the monopolist does not have sufficient information to write the maximization problem. Instead, the monopolist knows a proper subset of actual properties of F. We write the set of distributions that satisfy the said properties as $\mathcal{F}$.

Let us list the properties we are interested in and define the distributions accordingly.

IH The support of F is $[\underline{p}, \overline{p}]$ and $f(p) > 0 \ \forall p \in [\underline{p}, \overline{p}]$. $\forall F \in \mathcal{F}$, 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. 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. $\mathcal{F}^\eta$ is a compact convex subset of $\mathcal{F}^0$ and is a large subset of $\mathcal{F}^0$ if we choose large $\eta > 0$. We collect the useful properties into a lemma without proof.

Lemma 3.1.   (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 $\mathcal{F}^0$.

In the rational benchmark, the monopolist knows the distribution F of the buyer's valuations. In our case, the monopolist only knows that the actual distribution is an element of $\mathcal{F}^\eta$: $F \in \mathcal{F}^\eta$.

3.3. **Specification.** Because the notion of a specification is central to our exercise, we spell out its precise meaning and related concepts according to how we use the terms in the paper. We aim to clarify the distinction between various approaches that have been proposed in the literature.

Consider a decision problem represented by an outcome function $G : \Sigma \to \Delta(Y)$, where Y is a set of *outcomes*, Σ is the strategy space and $\Delta(Y)$ is the collection of all distributions over Y. Let $\mathcal{G}$ be the set of all outcome functions. In our problem, Σ will be a price the seller plans on charging possibly in a randomized fashion and Y will be the realization of the corresponding realized price and quantity.

In the monopolist market, the decision variable of the monopolist is to choose a price possibly in a randomized fashion, and the outcome is the probability distribution of price and quantity, which is completely determined by the aggregate demand function. Instead of G, we regard the demand function as an outcome function for the rest of the paper to simplify the exposition. Since the expected demand function and the distribution of the valuation has 1-1 correspondence, we treat the distribution of buyer's valuation as the outcome function. For the rest of the paper, we regard G as an aggregate distribution over the valuations of the buyers.

Definition 3.2. *A* specification *a pair* $(R, X(R))$ *where* R *is a formal statement and* $X(R)$ *is the collection of all feasible parameters under* R.

In principle, we can include the parametric restriction beyond R in the definition of specification. We chose not to do so for two reasons. First, the ensuing analysis does not rely on specific parametric restrictions. The Second reason is the convention. By specification, we often refer to the functional restriction on the probability distribution rather than the parametric restrictions. In practice, R describes the functional form of an outcome function and $x \in X(R)$ is the parameter to determine the distribution over outcomes.

Define $S(R, X(R))$ as the collection of all outcome functions satisfying the formal description and the parmeteric constraint, whose generic element is written as F. Since we treat the aggregate distribution of valuations as the outcome function, this paper regards F as the aggregate distribution of the valuations of buyers. For each realized value $x \in X(R)$, we call $S(R, x)$ a *model*, which describes the properties of the outcome function for a given parameter x. We say a specification is *complete* if $\left|S(x)\right| = 1$ for all $x \in X$. Otherwise, the specification is *incomplete*. A complete specification is necessary for the decision maker to write down the objective function.

We suppose that $\mathcal{F}^*$ represents the collection of all true distributions.

Definition 3.3. *A specification $(R, X(R))$ is* correct *if $\forall F \in \mathcal{F}^*$, $\exists x \in X(R)$ such that $F \in S(R, x)$. Otherwise, we say that S is* misspecified.

Our notion of specification is more general than the corresponding definition in Esponda and Pouzo (2014) who focus on the case of $|\mathcal{F}^*| = 1$. If $\mathcal{F}^*$ is assumed to be a singleton, our notions would coincide with the definition of Esponda and Pouzo (2014). A model misspecified if true $F^* = \{\mathcal{F}^*\} \neq S(R, x)$ $\forall x \in X(R)$. Our definition of specification is handy in examining the case where $|\mathcal{F}^*| > 1$. Under the robust approach, for example, the decisionmaker entertains a set of possible functions instead of a single outcome function. Note that the definition of specification does not preclude the possibility that $\exists x \in X(R)$, $S(R, x) = F \in \mathcal{F}^*$.

We can describe a correct model using different formal statements. It is natural to focus on the most "parsimonious" specification that contains no superfluous restrictions on the outcome functions.

Definition 3.4. *We call a correct specification $(R^*, X(R^*)$* minimal *if there is no $\tilde{R} \neq R^*$ such that $(R^* \vee \tilde{R}, X(R^* \vee \tilde{R})$ is a correct specification, where $\vee$ is the logical "or."*

From now on, we mean the minimal correct specification by the correct specification. We are ready to define underspecification.

Definition 3.5. *Let $(R^*, X(R^*)$ be the (minimal) correct specification, and suppose that a specification $(R, X(R))$ is given. We say that $(R, X(R))$ is* underspecified*, if $\forall x \in X(R^*)$, $S(R^*, x) \subset S(R, X(R))$. We say that $(R, X(R))$ is* strictly underspecified*, if the inclusion is strict.*

Misspecification can be an incomplete specification, although we usually choose a misspecified model to complete the specfciation to write down the optimization problem. Under-specification is an incomplete specification, but not misspecification because the decision maker knows a part of the true specification. By the definition, the underspecification is incomplete specification so that the decision maker cannot write down the optimization problem, unless one imposes a further restriction on $S(R, X(R))$.

A robust decision maker is typically endowed with $\left|S(R, x)\right| > 1$ for $x \in X(R)$. If $\forall F \in \mathcal{F}^*$, $\exists x \in X$ such that $F \in S(R, x)$, then the decision maker's model is underspecified, but not misspecified. To be underspecified, the model must be consistent with a part of the correct specification.

As we have introduced many closely related concepts, let us examine an example.

Example 3.6. *Suppose that the set of true distributions, $\mathcal{F}^*$, is a singleton that contains a truncated Gaussian distribution: $\mathcal{F}^* = \{F^*\}$. That is, a single distribution over the valuations of buyers can be a true distribution. Suppose that R^* is the statement that buyers' true distributions of valuations are truncated Gaussian distributions parameterized by three numbers: mean, variance of the untruncated Gaussian distribution and the cut-off point which determines the lower bound of the support of the distribution of the consumer valuations.[6] Thus, $X(R^*) = \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}$. The specification is minimal, because a truncated Gaussian distribution can be identified by three parameters, and cannot be identified by fewer parameters.*

Given $x^* = (x_1^*, x_2^*, x_3^*) \in X(R^*)$, we can identify a unique truncated Gaussian distribution. The specification is complete, because x^* determines a unique distribution. The specification is correct, because any distribution is a trauncated Gaussian distribution if and only if the distribution satisfies R^* parameterized by three numbers under R^*.

Suppose R^I says that the distribution satisfies the increasing hazard rate property with a finite mean and variance, whose valuation is bounded from below. Let $x = (x_1, x_2, x_3)$ be the feasible parameter of means, variances and the lower bound of the valuations. Then, $X(R^I) = \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R} = X(R^*)$.

Since $S(R^*, X(R^*)) \subset S(R, X(R))$, $S(R, X(R))$ is under-specified. $\forall x$, $S(R, x)$ admits multiple distributions that share the identical first and second moments, but may have different higher order moments. Thus, $(X, R(X))$ is incomplete specification.

Suppose that R^U says that the distribution of the valuations is a uniform distribution over $[\underline{v}, \overline{v}]$ where $0 \leq \underline{v} < \overline{v} < \infty$. Note that $S(R^U, X(R^U)) \cap S(R^*, X(R^*)) = \emptyset$. Thus, $(R^U, X(R^U))$ is a mis-specification.

3.4. **The Seller's Decision Problem.** We assume that the monopolist sellers learn 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.

Time is discrete: $t = 1, 2, 3, \ldots$. In each period, N consumers enters 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 the game Γ in each period.

The details of Γ profoundly influence the statistical procedure that the monopolist uses to learn about the optimal price. Let us illustrate two examples of per-period game Γ. Both games are based on the same monopoly market but generate substantially different data ex post.

Example 3.7. *Suppose that Γ_{CL} is the static monopoly market. At the beginning of period t, the monopolist estimates the expected demand curve $\hat{Q}_{t-1}(p)$ where p is the price, where $\hat{Q}_{t-1}(p)$ 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 choose p_t satisfying*

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

---

[6]This is not the only way to identify a truncated Gaussian.

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 valuatio 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 statistical procedure $\mathcal{A}$, for example, by minimizing the forecasting errors of the quantity.

In Example 3.7, the seller observes only the aggregate outcome from the market (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 for statistical procedure $\mathcal{A}$ uses.

Example 3.8. *Suppose that Γ_{CR} is the revelation game (Cole and Roughgarden (2014)) of the monopoly market game in Example 3.7. The strategy of each buyer in period t is to report $\hat{v}_{i,t}$ conditioned on true valuation $v_{i,t}$ in period t.*

The monopolist uses the following statistical procedure, which we call $\mathcal{A}_{CR}$. At the beginning of period t, the monopolist has the estimated aggregate distribution $\hat{F}_{t-1}$ of buyers' valuations. The monopolist randomly sample $K(\leq N)$ elements from $\{\hat{v}_{1,t}, \ldots, \hat{v}_{N,t}\}$ and construct an empirical distribution $\tilde{F}_t$ from K sample of reported types. The monopolist updates the estimated aggregate distribution according to

$$\hat{F}_t(v) = \hat{F}_{t-1}(v) + \frac{1}{t}\left(\tilde{F}_t(v) - \hat{F}_{t-1}(v)\right) \tag{3.2}$$

and choose p_t that maximizes the estimated expected profit

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

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 the 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. The monopolist's payoff in period t is $p_t q_t$.

By the nature of the revelation game, the outcome of one period game Γ in Example 3.8 in period t is

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

Statistical procedure $\mathcal{A}_{CR}$ 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, 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.

We will be intentionally vague about the details of Γ. We admit any monopoly market trading protocol but require that the monopolist can observe the delivery price and the quantity at the end of Γ.[7]

Assumption 3.9. *Let (p_t, q_t) be the profile of the pairs of price and quantity in period $t \geq 1$, where q_t is amount of goods sold at price p_t. At the end of one period game Γ, the monopolist observes (p_t, q_t) possibley along with other variables.*

---

[7]We focus on the uniform price 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 $p_{k,t}$.

The examples highlight three components necessary for the monopolist to learn the optimal price: data, forecast, and the statistical procedure that maps data to forecast. Let us explain each component one by one.

3.5. **Data.** The outcome O_t specifies what a player can observe and is exogenously determined by the one-period game Γ. 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 $t \geq 1$.

By the data, we mean a subset of the outcome that the statistical procedure uses as input of an algorithm. Let D_t be the data in period t, which is a "sub-vector" of an outcome O_t in period t. The decision-maker can choose to ignore some elements in the outcome as in Example 3.8. While outcome O_t is determined by Γ, the algorithm spells out what elements of an outcome can be a data.

Let $\dim D_t$ be the number of components in D_t, which cannot be larger than the number of the observed outcomes in O_t in period t. In Example 3.7, $\dim D_t = 2$ and in Example 3.8, $\dim D_t = K$ and. If the collection and the process of data incur the cost, the configuration of data is a part of the strategic choice of the monopolist in designing a statistical procedure. Define $\mathcal{D}_t = (D_1, \ldots, D_{t-1})$ and $\mathcal{D}$ as the collection of all feasible $\mathcal{D}_t$ $\forall t \geq 1$.

3.6. **Forecast.** The outcome of the statistical procedure is a forecast, which is a model for the distribution of the valuations of the buyers. The range of the statistical procedure is $\mathcal{F}^\eta$. If the forecast of the distribution is outside of $\mathcal{F}^\eta$, the decision maker knows the forecast is wrong and has little reason to follow the recommendation of the forecast.

Because the forecast entails a set of parameters to be estimated, the specification of forecast is often driven by the convenience of estimation. While we can be completely general about the range of an algorithm, we opt for a parametric family of specifications which are widely used in practice but also sufficiently general to approximate a large class of functions.

Let R^ℓ be the statement that

(1)

$$F(p) = 1 - \left(\sum_{i=0}^{\ell} \beta_i p^i \right) \tag{3.3}$$

    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 Liptschitz continuous with parameter $\eta > 0$.
(3) F satisfies the increasing hazard rate property.

so that the expected demand curve

$$Q(p) = \sum_{i=0}^{\ell} \beta_i p^i$$

is a strictly decreasing smooth function with strictly decreasing marginal revenue curve whose slope is uniformly bounded away from 0. Define $X(R^\ell)$ as the set of all admissible

$\ell + 1$ coefficients $(\beta_i)_{i=0}^{\ell}$. Define

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

and $X(R^f)$ as the set of all feasible coefficients. 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 (3.3) for some $\ell \geq 0$.

Under our definition of misspecification, we can say that $\mathcal{A}$ is correctly specified if $\exists \ell$ such that $\mathcal{F}^* \subset S(R^{\ell}, X(R^{\ell}))$ and otherwise, misspecified. The decision maker may want to choose $S(R^{\ell}, X(R^{\ell}))$ with a reasonable size of ℓ to simplify the estimation process. Since the decision maker does not know the true models $\mathcal{F}^*$, choosing a simpler model increases the chance of choosing a misspecified model of a demand curve. Since $\mathcal{F}^*$ is not known to the monopolist, it would be excessive to expect the monopolist to find ℓ so that $\mathcal{F}^* \subset S(R^{\ell}, X(R^{\ell}))$. A natural question is whether the monopolist has to choose a very large ℓ in order to behave as if he learns the correct demand curve. Our ensuing analysis shows otherwise.

3.7. **Algorithm.** An algorithm (a.k.a., statistical procedure) is a mapping from the history of outcomes into a forecast

$$\mathcal{A} : \mathcal{O} \rightarrow S(R^f, X(R^f))$$

where $\mathcal{O}$ is the set of all histories of outcomes. $\mathcal{A}$ can use only a subset of O_t as data D_t. Even if a history of outcomes differ, the sequence of data used by $\mathcal{A}$ can be identical, leading to the same forecast.

To maintain some algorithmic simplicity, we focus on recursive algorithms.

A1: Recursive algorithm $\mathcal{A}$ is a recursive algorithm if $\exists \Psi$ such that

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

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. We require that the internal state variable Ω_t should be updated according to the fixed function Ψ based on what the algorithm can observe.

The goal of the monopolist is to maximize expected profit by choosing an optimal price. Let

$$\varphi = (\varphi_p, \varphi_q) : \mathcal{F}^{\eta} \rightarrow \Delta(\mathbb{R}_+) \times \Delta(\mathbb{R}_+)$$

be the recommendation by the algorithm to choose a (randomized) price $\varphi_p(\mathcal{O}_t)$ conditioned on $\mathcal{O}_t$, based on the expected sales of $\varphi_q(\mathcal{O}_t)$. Thus, the recommendation predicts the expected profit would be $\mathbb{E}\varphi_p(\mathcal{O}_t)\varphi_q(\mathcal{O}_t)$. We allow a randomized price, because the algorithm may need to generate data to learn more about the demand curve. Assuming the obedient monopolist, we interpret $\varphi_p(\mathcal{A}(\mathcal{O}_t))$ as the actual price charged by the monopolist in period t, who expects the amount of sales to be $\varphi_q(\mathcal{A}(\mathcal{O}_t))$.

By an algorithm, we mean statistical procedure $\mathcal{A}$ along with the formal description of $D_t \subset O_t$. Whenever the meaning is clear from the context, we simple refers to $\mathcal{A}$ as

an algorithm along with the data struture $D_t \subset O_t$ and the recommendation function $\varphi = (\varphi_p, \varphi_1)$.

3.8. **Complexity Measure.** Since $\mathcal{A}$ is a recursive algorithm, the input 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 in period t. Since the demand curve forecast is a polynomial, $\forall t \geq 1$, $\forall \mathcal{O}_t$, $\exists \ell$ and $(\beta_i)_{i=0}^{\ell}$ such that

$$\mathcal{A}(\mathcal{O}_t) = 1 - \sum_{i=1}^{\ell} \beta_i p^i.$$

Thus, $\dim(\mathcal{A}(\mathcal{O}_t)) = \ell + 1$. Our measure of complexity focuses on the size of memory necessary for the operation of the algorithm in each period t.

Example 3.10. *In Example 3.8, 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 value of $\hat{F}_t(\hat{v})$ for each reported value $\hat{v}$ up to $t - 1$ round. In Example 3.7, $\mathcal{A}(\mathcal{O}_{t-1})$ 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 3.11. *We can implement algorithm $\mathcal{A}_{CR}$ in Example 3.8 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 $\mathcal{A}_{CR}^o$. In each period $t \leq T - 1$, the algorithm takes the K number of reported valuations of the buyers. For $t \leq T - 1$, $\dim(D_t) = K$ and $\dim(\mathcal{A}_{CR}^o(\mathcal{O}_t)) = 0$ 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 $\mathcal{A}_{CR}^o(\mathcal{O}_T)$.*

In Example 3.7, $\dim D_t = 2$ and $\dim \mathcal{A}(\mathcal{O}_t) = 2$ remain constant until the algorithm produces the final forecast in period T.

Example 3.12. *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 in put. Thus, $\dim(\mathcal{A}_{LSE}(\mathcal{O}_{t-1})) = 2$ 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, $\dim(\mathcal{A}_{LSE}(\mathcal{O}_{t-1}), D_t, \Omega_t) = 6$.

Definition 3.13. *The complexity of $\mathcal{A}$ up to time T is*

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

and if the algorithm operates indefinitely,

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

3.9. **Algorithm Game.** The algorithm game is a normal form game between the monopolist and the nature.[8] The monopolist's strategy space is A that the set of algorithms satisfying **A1** (along with recommendation function φ and the choice of $D_t \subset O_t$ as a part of the algorithm). The nature's strategy space is $\mathcal{F}^*$, which we assume is a subset of $\mathcal{F}^\eta$. After $(\mathcal{A}, F) \in \mathsf{A} \times \mathcal{F}^*$ is selected, the continuation game is played between the algorithm and the sequence of short run consumers.

Conditioned on $(\mathcal{A}, F) \in \mathsf{A} \times \mathcal{F}^*$, the algorithm generates recursively p_t and

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

amount of goods is sold, where $\epsilon_{2,t}$ is defined according to (3.1) $\forall t \geq 1$. Let (p_t, q_t) be the stochastic process of the pair of price and quantity realized according to $(\mathcal{A}, F)$. The payoff of the monopolist is the expected long run average payoff.

$$\mathcal{U}(\mathcal{A}, F) = \lim_{T \to \infty} \frac{1}{T} \mathbb{E} \sum_{t=1}^{T} \left[p_t q_t \mid F \in \mathcal{F}^* \right] \tag{3.4}$$

We need not write down the payoff function of the nature, because the only move of the nature is to choose $F \in \mathcal{F}^*$.

The natural optimization problem of the monopolist would be to choose

$$\mathcal{A} \in \arg\max_{\mathcal{A} \in \mathsf{A}} \mathcal{U}(\mathcal{A}, F)$$

for given $F \in \mathcal{F}^*$. Because the performance of an algorithm is evaluated against the actual distribution of valuation, however, the monopolist cannot observe at the time when he selects an algorithm. Moreover, the monopolist does not know $\mathcal{F}^*$, and does not have a prior over $\mathcal{F}^*$. The assessment of an algorithm is evaluated by ex-post expected payoff after the true demand curve is revealed to the monopolist. The expected profit maximization problem of the monopolist is not well defined, because the monopolist's model is underspecified, and therefore, incompletely specified.

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

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

---

[8]We suppress the decision problem of the buyer, assuming that the buyer uses a dominant strategy to purchase a good if $v_{i,t} \geq p$.

Instead, the monopolist is searching for a dominant solution of the game:

$$\mathcal{A} \in \bigcap_{\Pi \in \Delta(\mathcal{F}^\eta)} \arg\max_{\mathcal{A} \in \mathsf{A}} \int_F \mathcal{U}(\mathcal{A}, F) d\Pi(F). \tag{3.5}$$

where $\Delta(\mathcal{F}^\eta)$ is the probability distribution over $\mathcal{F}^\eta$. Since $\mathcal{F}^\eta$ is underspecified, $\mathcal{F}^* \subset \mathcal{F}^\eta$. Thus, $\mathcal{A}$ induces the best response against the true distribution in the long run.

We assume that the monopolist incurs the complexity cost. The monopolist has lexicographic preference over the long run average profit and the complexity cost.

Definition 3.14. $\mathcal{A}^* \in \mathsf{A}$ *is an optimal algorithm if*

$$\mathcal{A}^* \in \bigcap_{\Pi \in \Delta(\mathcal{F}^\eta)} \arg\max_{\mathcal{A} \in \mathsf{A}} \int_F \mathcal{U}(\mathcal{A}, F) d\Pi(F)$$

and

$$\text{comp}(A^*) \le \text{comp}(A) \qquad \forall A \in \bigcap_{\Pi \in \Delta(\mathcal{F}^\eta)} \arg\max_{\mathcal{A} \in \mathsf{A}} \int_F \mathcal{U}(\mathcal{A}, F) d\Pi(F).$$

3.10. **Useful Results.** To make the profit maximization problem meaningful, the monopolistic seller must know ex ante that an algorithm can find the optimal price of any demand curve uniformly over $\mathcal{F}^\eta$. For $F \in \mathcal{F}^\eta$, define $(b^*(F), q^*(F))$ as

$$b^*(F) = \arg\max_p p(1 - F(p))$$

Thanks to the increasing hazard rate property, the optimization problem admits a unique solution $b^*(F)$.

It is not straightforward to verify whether a given algorithm is a dominant strategy in A, because of the complexity of A. We exploit the PAC guaranteeing property developed in the computer science literature (Shalev-Shwartz and Ben-David (2014)) to find an optimal algorithm.

Definition 3.15. $\mathcal{A}$ *PAC guarantees* $\mathcal{F}^\eta$ *if* $\forall \mu > 0$, $\forall \lambda \in (0, 1)$ *and* $\overline{T}(\mu, \lambda)$ *such that*

$$\mathbb{P}\left(\left|\varphi(\mathcal{A}(\mathcal{O}_t)) - (b^*(F), 1 - F(b^*(F)))\right| \ge \mu\right) \le \lambda \qquad \forall t \ge \overline{T}(\mu, \lambda) \tag{3.6}$$

where $\overline{T}(\mu, \lambda) \sim \mathsf{O}\left(-\frac{\log \lambda}{\mu^p}\right)$ *for some* $p > 0$.

If $\mathcal{A}$ PAC guarnatees b^* over $\mathcal{F}^\eta$, $\mathcal{A}$ learns $b^*(F)$ uniformly over $\mathcal{F}^\eta$. Since $\mathcal{F}^* \subset \mathcal{F}^\eta$, $\mathcal{A}$ forecast 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. For later reference, we state the observation without a proof.

Proposition 3.16. *If* $\mathcal{A}$ *PAC guarantees* $\mathcal{F}^\eta$, *then* $\mathcal{A}$ *is a dominant strategy in* A.

Following Cole and Roughgarden (2014), one can prove that algorithm $\mathcal{A}_{CR}$ in Example 3.8 PAC guarantees $\mathcal{F}^\eta$ if Γ is the revelation game version of the monopoly market.

Theorem 3.17. *Suppose that* Γ *is the revelation game of the monopoly problem.* $\mathcal{A}_{CR}$ *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 3.7. Also, if a dominant strategy exists, we ask whether a simpler algorithm can PAC gurantee b^* over $\mathcal{F}^\eta$.

As a first step to find an optimal algorithm, we show that the forecast of $\mathcal{A}$ must be sufficiently elaborate to PAC guarantee $\mathcal{F}^\eta$. Recall that $F \in S(R^\ell, X(R^\ell))$ implies that F is ℓ-th order polynomial. If $\ell = 0$, $\forall F \in S(R^0, X(R^0))$ is a denerate distribution concentrated at $v \in [\underline{p}, \overline{p}]$, which induces the aggregate expected demand curve where $p > v$, the expected demand is 0, and $p \leq v$, the expected demand is 1. Note that $F \in F \in S(R^0, X(R^0))$ is parameterized by a single number.

Proposition 3.18. *If $\mathcal{A} : \mathcal{O} \to S(R^0, X(R^0))$, $\mathcal{A}$ is not a dominant strategy in* A*.*

Proof. See Appendix A. □

To be a dominant strategy in A, the range of a dominant strategy algorithm must be at least $S(R^1, X(R^1))$. The complexity of a forecast must be at least 2. Since the algorithm essentially forecasts the demand curve, the minimal data is the pair of price and quanity and the complexity of input data must be at least 2. If we can construct a recursive algorithm with forecast complexity of 2 and data complexity of 2 that can PAC gurantees $\mathcal{F}^\eta$, we have an optimal algorithm in A.

Let us summarize the main result of the paper.

Theorem 3.19. *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* $\mathrm{comp}(\mathcal{A}) = 4$.

4. Detour

The construction of the optimal algorithm involves technical steps, which tend to obscure the central ideas of the algorithm. Instead of directly constructing the optimal algorithm in Theorem 3.19, we take a detour by constructing an algorithm which satisfies weaker criteria of optimality and PAC learnability. The weaker notion is based on the performance of the algorithm over a finite number of periods. We need to weaken the criterion for optimality in order to accommodate the finite number of observations which are not sufficient to average out all stochastic components in the forecast. The detour has several important benefits.

First, the weaker notions of dominance and PAC learnability allow us to examine some algorithms such as $\mathcal{A}^o_{CR}$ in Example 3.11, whose complexity is ∞ if the algorithm has to run indefinitely. The weaker notion of PAC learnability is defined for an algorithm that terminates in a finite number of rounds, producing the final forecast, which the monopolist uses for the continuation game.[9] We can compare $\mathcal{A}^o_{CR}$ to a weaker version of the optimal algorithm in Theorem 3.19 to highlight the substantive difference of our algorithm over the classic benchmark of $\mathcal{A}^o_{CR}$.

Second, we need to relax the criterion of best response by allowing the response to be close to the best response, and define the weaker notion of dominance accordingly. We

---

[9]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.

show that the weaker notion of dominance is equivalent to the weaker notion of PAC guaranteeing property. The weaker notion of dominance is useful for extending to the case where the monopolist discounts the future payoff.

Third, the algorithm constructed to meet the weaker notion of PAC learnability is easier to construct than the optimal algorithm in Theorem 3.19. The algorithm generates more robust numerical results, which greatly helps to compare the numerical performance of our algorithm to our benchmark algorithm $\mathcal{A}_{CR}^o$.

To define a weaker notion of a dominant strategy in A, let us imagine a monopolist who can run algorithm $\mathcal{A}$ for a finite number of rounds, say T, to learn the optimal price and the maximized profit. At the end of period T, the algorithm produces the final forecast $(\varphi_p(\mathcal{O}_T), \varphi_q(\mathcal{O}_T))$, which the monopolist uses for the continuation game.[10] Thus, the long run average payoff is

$$\mathcal{U}(\mathcal{A}, F, T) = \mathbb{E}\varphi_p(\mathcal{O}_T)(1 - F(\varphi_p(\mathcal{O}_T))).$$

Define for a small $\epsilon > 0$, and $F \in \mathcal{F}^*$,

$$\mathsf{BR}^\epsilon(F, T) = \left\{ \mathcal{A} \mid \mathcal{U}(\mathcal{A}, F, T) \geq b^*(F)(1 - b^*(F)) \right\}$$

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 uniform over $\mathcal{F}^\eta$.

We define the uniform learnability as a weaker version of PAC guarantee. PAC guarantee requires that the recommendation of the algorithm must converge to the actual optimal price and expected quantity uniformly in probability. Instead, we only require the uniform convergence in expectation.

Definition 4.1. $\mathcal{A}$ *is uniformly learnable of* $\mathcal{F}^\eta$ *if* $\forall \lambda \in (0, 1)$, $\exists \mathcal{A}$ *and* $\overline{T}(\lambda)$ *such that*

$$\mathbb{E}\left(\left| \varphi(\mathcal{A}(D_{\overline{T}(\mu, \lambda)})) - (b^*(F), 1 - F(b^*(F))) \right| \right) \leq \lambda \tag{4.7}$$

where $\overline{T}(\lambda) \sim \mathsf{O}(-\log \lambda)$.

Given the accuracy bound μ and the confidence bound $1 - \lambda$, the uniform learnability requires us to construct an algorithm $\mathcal{A}$ and stopping time $\overline{T}(\mu, \lambda)$ such that the algorithm produces a forecast by period $\overline{T}(\mu, \lambda)$ satisfying accuracy and confidence bounds.

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 4.2. $\forall \epsilon > 0$, $\mathcal{A}$ *is an* ϵ *dominant strategy if and only if* $\forall \lambda \in (0, 1)$, $\exists \mathcal{A}$ *and* $\overline{T}(\lambda)$ *such that*

$$\mathbb{E}\left(\left| \varphi(\mathcal{A}(D_{\overline{T}(\mu, \lambda)})) - (b^*(F), 1 - F(b^*(F))) \right| \right) \leq \lambda$$

---

[10]Following the terminology of the computer science, we can call T as the training period, and $(\varphi_p(\mathcal{O}_t), \varphi_q(\mathcal{O}_t))$ as the final hypothesis.

Proof. See Appendix B   □

Note that we drop the data complexity from the weak learnability in establishing the equivalence condition with the ϵ dominance. The data complexity dictates the rate at which the expected profit recommended by the algorithm converges to the actual maximized profit. Because the monopolist is infinitely patient, he is not concerned of the rate.

Following the same logic as in Proposition 4.3, we can establish the same lower bound.

Proposition 4.3. *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 $\epsilon > 0$.*

5. Construction

We now construct a recursive algorithm

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

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

5.1. **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:

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

and estimates (β_0, β_1) according to the least square estimation over $\mathcal{O}_t$. Let $\mathcal{H}$ be the set of all linear demand functions, parameterized by (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 $F \in \mathcal{F}^\eta$,

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

or equivalently,

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

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 $\mathcal{H}_B \subset \mathcal{H}$ be the collection of the linear demand, which induces the market outcome in K. The seller knows that the market demand curve is in $\mathcal{H}_B$. Since $\mathcal{F}$ typically contains non-linear demand curves, $\mathcal{H}_B$ is a misspecified model. Nevertheless, the number of parameters the seller has to keep track of is minimal.

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 $\mathbb{E}\epsilon_{1,t}^2 = \sigma_1^2$. We interpret $\epsilon_{1,t}$ as the implementation error or the small experimentation by the monopolist seller. We choose $\epsilon_{1,t}$ from a small interval $[-\epsilon, \epsilon]$ according to a fixed distribution, say the uniform distribution. We control the size of $\epsilon > 0$ to achieve the desired level of accuracy of the algorithm.

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

Since the quantity 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, we directly update $(\beta_{0,t-1}, \beta_{1,t-1})$.

Fix $a > 0$ (which will be a parameter of the constructed algorithm.) 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
$$b_{t+1} = -\frac{\beta_{0,t}}{2\beta_{1,t}} = -\frac{\beta_{0,t-1}}{2\beta_{1,t-1}} + \frac{a}{2\beta_{1,t-1}} = b_t + \frac{a}{2\beta_{1,t-1}} < b_t,$$

where small $a > 0$ is a parameter of the algorithm that will be determined to satisfy the precision and the confidence requirements. 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

$$\begin{bmatrix} \beta_{0,t} \\ \beta_{1,t} \end{bmatrix} = \begin{bmatrix} \beta_{0,t-1} \\ \beta_{1,t-1} \end{bmatrix} + aR_{t-1}^{-1} \begin{bmatrix} 1 \\ -\frac{\beta_{0,t-1}}{2\beta_{1,t-1}} + \epsilon_{1,t} \end{bmatrix} \phi(\beta_{t-1}, \epsilon_t) \qquad (5.8)$$

where

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

and $a > 0$ is a parameter we choose to meet the accuracy and confidence requirement of the algorithm. Since the monopolist designs the size of the experiments, the variance σ_1^2 of $\epsilon_{1,t}$ is a known parameter.[11]

We need to impose a bound to $(\beta_{0,t}, \beta_{1,t})$ 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.[12]

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

$$\begin{bmatrix} \beta_{0,t} \\ \beta_{1,t} \end{bmatrix} = \begin{bmatrix} \beta_{0,t-1} \\ \beta_{1,t-1} \end{bmatrix} + aR_{t-1}^{-1} \begin{bmatrix} 1 \\ -\frac{\beta_{0,t-1}}{2\beta_{1,t-1}} + \epsilon_{1,t} \end{bmatrix} \phi(\beta_{t-1}, \epsilon_t) \qquad (5.9)$$

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

$$\varphi_{t-1} \equiv \varphi(\beta_{t-1}, \epsilon_t) = R_{t-1}^{-1} \begin{bmatrix} 1 \\ -\frac{\beta_{0,t-1}}{2\beta_{1,t-1}} + \epsilon_{1,t} \end{bmatrix} \phi(\beta_{t-1}, \epsilon_t). \qquad (5.10)$$

Treating the estimated demand curve

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

---

[11]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) are unknown to the seller and 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.
[12]This mapping is known as the projection facility in the literature of the stochastic approximation (Kushner and Yin (1997)).

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 $(\beta_{0,t}, \beta_{1,t})$. The translating function maps the estimated linear demand curve into the mean forecast price and quantity

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

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 $\mathcal{O}_t$, where

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

is the sequence of aggregate market outcomes up to period $t-1$. 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

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

so that $\mathrm{comp}_T(\mathcal{A}_a) = 4$ $\forall T \geq 1$. We will choose $a > 0$ to satisfy the accuracy and confidence requirement.

To emphasize that the optimal price b^* is a function of the underlying (aggregate) distribution F, we sometimes write $b^*(F)$ instead of b^*. Let $\beta^*(F)$ be the pair of estimators that induce the optimal price for F:

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

Theorem 5.1. $\forall \mu > 0$, $\forall \lambda > 0$, $\exists \overline{T}(\mu, \lambda)$ and $\exists a > 0$ such that

$$\mathbb{P}\left(\left| \varphi\left(\mathcal{A}_a(\mathcal{O}_{\overline{T}(\mu,\lambda)}) \right) - (b^*(F), q^*(F)) \right| > 4\mu \right) \leq \lambda \qquad \forall F \in \mathcal{F}^\eta$$

where $\overline{T}(\mu, \lambda) \sim \mathsf{O}\left(-\frac{\log \lambda}{\mu^p} \right)$ for some $p > 0$.

Proof. See Appendix C. $\square$

6. HEURISTICS

In this section, we provide a heuristic explanation about how to prove Theorem 5.1, delegating formal details to the Appendix. 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

$$\frac{1 - F(b^*(F))}{f(b^*(F))} - b^*(F) = 0. \tag{6.11}$$

For fix $\tau > 0$, define

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

Following the control theory literature (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:[13]

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

Using the recursive nature of the algorithm, we can write

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

which implies

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

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

$$= \left[\frac{1}{T(a,\tau)} \sum_{k=1}^{T(a,\tau)} \mathbb{E}_{t+k-1} \varphi_{t+k} \right] + \left[\frac{1}{T(a,\tau)} \sum_{k=1}^{T(a,\tau)} \varphi_{t+k} - \mathbb{E}_{t+k-1} \varphi_{t+k} \right] \quad (6.12)$$

where φ_{t+k} is defined in (5.10). Our proof will analyze this expression in detail. Define

$$\xi_{t+k} = \phi_{t+k} - \mathbb{E}_{t+k-1} \varphi_{t+k} \quad (6.13)$$

which is a martingale difference.

6.1. **Tracking the Mean.** Let us examine the first term in (6.12) as $a \to 0$ for a small fixed $\tau > 0$. $\beta_{t+T(a,\tau)} - \beta_t$ is an average of $\{\phi_{t+k}\}_{k=1}^{T(a,\tau)}$. We can approximate

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

or more concisely as

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

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

---

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

the estimator to minimize the forecasting error. Given (β_0, β_1), the next period's quantity q' and price p' are

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

where

$$b = -\frac{\beta_0}{2\beta_1}. \tag{6.14}$$

To fit the observed data best, the monopolist chooses the new coefficients (β'_0, β'_1) passing through $(b + \epsilon, 1 - F(b + \epsilon))$ and $(b - \epsilon, 1 - F(b - \epsilon))$. A simple calculation shows

$$\begin{aligned} \beta'_0 &= 1 - F(b) + bf(b) \\ \beta'_1 &= -f(b) \end{aligned}$$

modulo linear approximation error at the order of ϵ^2. Results from stochastic approximation theory[14] 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)

$$\begin{aligned} \dot{\beta}_0 &= \beta'_0 - \beta_0 = 1 - F(b) + bf(b) - \beta_0 \\ \dot{\beta}_1 &= \beta'_1 - \beta_1 = -f(b) - \beta_1. \end{aligned} \tag{6.15}$$

Since (6.14) 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 (6.15), we have

$$\dot{b} = -\frac{f(b)}{2\beta_1} \left[\frac{1 - F(b)}{f(b)} - b \right]. \tag{6.16}$$

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

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

With the increasing hazard rate property,

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

---

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

We can show that $\exists c > 0$ such that[15]

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

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 $1/\mu$.

6.2. **Calculating the 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 (6.13). 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 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 Chernoff's bound, 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 rate for all feasible distributions of buyer's valuation, which proves that our algorithm is efficient (Shalev-Shwartz and Ben-David (2014)).

---

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

7. 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 the numerical analysis 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.[16] 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 error as a blue bar 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 in orange color. 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 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 Cole and Roughgarden (2014) allows

---

[16]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 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 $\mathcal{A}_a$.

Figure 1 reports the distribution of the forecasting errors of $\mathcal{A}_a$ in blue and $\mathcal{A}^o_{CR}$ in orange when $\mathcal{A}^o_{CR}$ takes two valuation reports per period. Note that the distribution of forecasting errors of CR is spread out more than CL. 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.
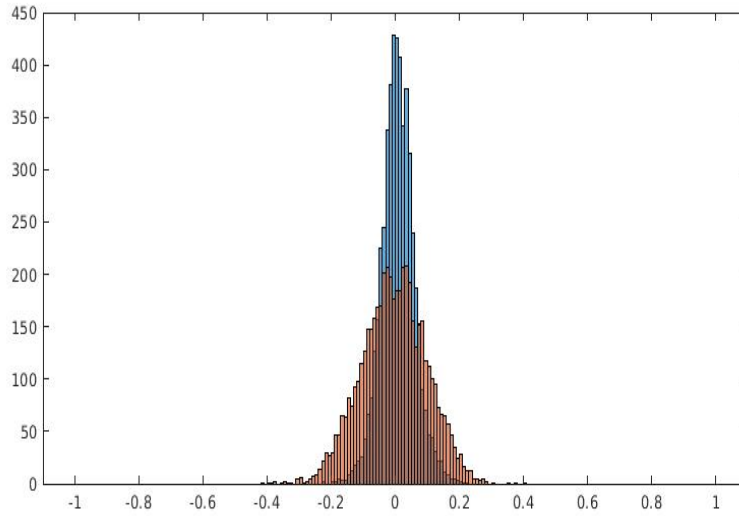

Figure 1: Blue bars represent the density of forecasting errors of our algorithm $\mathcal{A}_a$, and orange ones represent the distribution of errors of the algorithm of Cole and Roughgarden [2014]. If the algorithm of Cole and Roughgarden [2014] takes 2 reports per period, the variance is 3.5 times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

The performance of our algorithm $\mathcal{A}_a$ is comparable to that of Cole and Roughgarden (2014). We can reduce the linear approximation error by lowering the price perturbation,
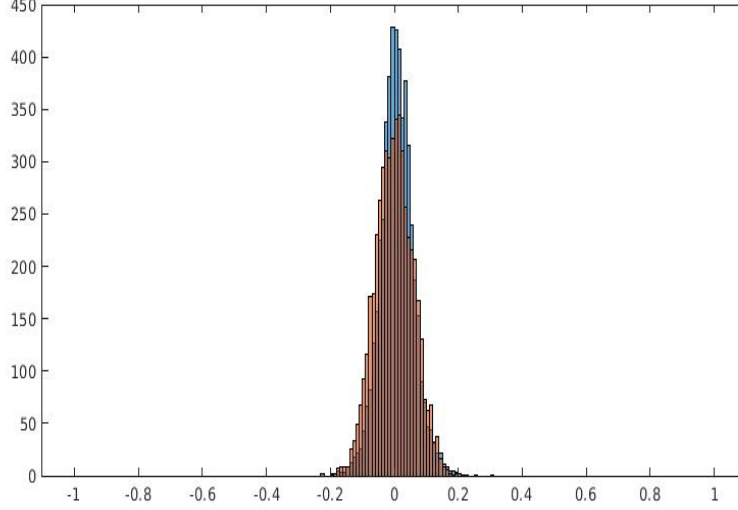
Figure 2: If the algorithm of Cole and Roughgarden [2014] takes 10 reports per period, the variance is comparable to that of $\mathcal{A}_a$. The distributions of the forecasting errors almost overlap with each other.

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.

8. EXTENSION

8.1. **PAC Guarantee.** We can modify $\mathcal{A}_a$ to satisfy the PAC guaranteeing property (3.6) to establish Theorem 3.19. We modify the algorithm so that the optimal solution calculated from $\mathcal{A}(\mathcal{O}_t)$ converges to the true optimal solution in probability. Recall (5.9) where we choose parameter $a_t = a > 0$. Instead of a positive constant gain function $a > 0$, we can choose $a = \frac{1}{t^\omega}$ where $0 < \omega < 1$ so that a_t is decreasing as t increases. We need to choose the initial value of the gain function as $\frac{1}{t_1^\omega}$. 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.

Let us state Theorem 3.19 more formally.

Theorem 8.1. $\forall \mu > 0$, $\forall \rho > 0$, we can construct algorithm $\mathcal{A}^{\frac{1}{t^\omega}}$ with initial value of the gain function set as $1/t_1^\omega$ such that $\exists T(\mu, \rho) > 0$ such that $\forall t \geq T(\mu, \rho)$,

$$\mathbb{P}\left(\left|\varphi\left(\mathcal{A}_{\frac{1}{t^\omega}}(\mathcal{O}_t)\right) - (b^*(F), 1 - F(b^*(F)))\right| > 4\mu\right) \leq \rho \qquad \forall F \in \mathcal{F}^\eta$$

and $T(\mu, \rho) = \mathsf{O}\left(\dfrac{\left(\log \frac{1}{\rho}\right)^{\frac{1}{1-\omega}}}{\mu^\kappa}\right)$ for some $\kappa \in (0, \infty)$.

Proof. See Appendix D. □

The proof of Theorem 8.1 follows the same reasoning as the proof of Theorem 5.1 despite additional complications and technical steps to deal with the decreasing gain $a_t = 1/t^\omega$.

8.2. **Discounting.** We have assume that the monopolist is infinitely patient. A direct consquence of not discounting the future payoff is that the monopolist does not consider how quickly the recommendation of the algorithm converges to the actual maximum profit. A natural question is whether the main result can be extended to the case with discounting. Instead of $\mathcal{U}(\mathcal{A}, F)$ defined in (3.4), suppose that the monopolist's payoff function is

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

for some $\delta \in (0, 1)$. We can define ϵ dominance with respect to $\mathcal{U}^\delta(\mathcal{A}, F)$.

Definition 8.2. *$\mathcal{A}$ is an ϵ dominant algorithm if $\forall \epsilon > 0$, $\exists \underline{\delta} \in (0, 1)$ such that $\forall \delta \in (\underline{\delta}, 1)$, $\forall F \in \mathcal{F}^\eta$,*

$$\mathcal{U}^\delta(\mathcal{A}, F) \geq b^*(F)(1 - F(b^*(F))) - \epsilon \tag{8.17}$$

We can characterize PAC guaranteeing property in terms of ϵ dominance in the algorithm game.

Proposition 8.3. *Suppose that the monopolist discounts the future payoff. $\mathcal{A}$ uniformly learns $\mathcal{F}^\eta$ if and only if $\mathcal{A}$ is an ϵ dominant algorithm in* A.

Proof. See Appendix E. □

$\mathcal{A}_a$ constructed in Theorem 5.1 is uniformly learnable. Thus, $\mathcal{A}_a$ is an ϵ dominant algorithm if the monopolist discounts the future payoff but is sufficiently patient.

9. 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* they may be grossly misspecified in terms of the demand curves that they consider. Thus, within a polynomial 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 4.3

To satisfy the optimality, the expected profit from the forecast must be equal to the actual optimal price and quantity: $\forall F \in \mathcal{F}^\eta$,

$$\lim_{t \to \infty} \mathbb{E}\varphi_p(\mathcal{O}_t)\varphi_q(\mathcal{O}_t) = b^*(F)(1 - b^*(F)).$$

$\forall F \in S(R^0, X(R^0))$, the valuation of buyers is concentrated at a single value $v \in [\underline{p}, \overline{p}]$. Thus, the best response must be $(v, 1)$. The quantity associated with the optimal price of $F \in \mathcal{F}^\eta$ is not 1. Consider F whose support is $[\underline{p}, \overline{p}]$. If the demand is 1, the price must be $\underline{p}$, generating the lowest possible profit for the monopolist. For such a distribution, $\mathcal{A}$ does not prescribe the best reponse.

Appendix B. Proof of Proposition 4.2

Suppose that $\mathcal{A}$ is an ϵ dominant strategy. To show that $\mathcal{A}$ is uniformly learnable, fix $\lambda > 0$. We then choose $\epsilon = \lambda$. Since $\mathcal{A}$ is an ϵ dominant strategy $\exists T$ so that

$$\mathbb{E}\varphi_p(\mathcal{A}(\mathcal{O}_T))(1 - F(\varphi_p(\mathcal{A}(\mathcal{O}_T)))) \geq b^*(F)(1 - b^*(F)) - \epsilon \qquad \forall F \in \mathcal{F}^\eta.$$

Thus,

$$\mathbb{E}\left(\left|\varphi(\mathcal{A}(D_{\overline{T}(\mu,\lambda)})) - (b^*(F), 1 - F(b^*(F)))\right|\right) \leq \epsilon = \lambda \qquad \forall F \in \mathcal{F}^\eta$$

as desired.

To show the converse, suppose that $\mathcal{A}$ is uniformly learnable. To show that $\mathcal{A}$ is an ϵ dominant strategy, fix $\epsilon > 0$. Since $\mathcal{A}$ is uniformly learnable, $\exists T$ such that

$$\mathbb{E}\left(\left|\varphi(\mathcal{A}(D_{\overline{T}(\mu,\lambda)})) - (b^*(F), 1 - F(b^*(F)))\right|\right) \leq \epsilon \qquad \forall F \in \mathcal{F}^\eta.$$

Thus,

$$\mathbb{E}\varphi_p(\mathcal{A}(\mathcal{O}_T))(1 - F(\varphi_p(\mathcal{A}(\mathcal{O}_T)))) \geq b^*(F)(1 - b^*(F)) - \epsilon \qquad \forall F \in \mathcal{F}^\eta$$

as desired.

Appendix C. Proof of Theorem 5.1

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.

C.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 $T(a)$.

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

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

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

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

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

$$|\xi_t| \le \xi.$$

Define

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

As shown in (C.21), 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

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

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 $\bar{b}$ 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

$$
\begin{aligned}
\beta_{T(a)} - \beta^*(F) &= \beta_0 + a\sum_{t=1}^{T(a)} \bar{b}(\beta_{t-1}) + a\sum_{t=1}^{T(a)} \xi_t - \beta^*(F) \\
&= \beta_0 + \int_0^\tau \bar{b}(\beta(s))ds - \beta^*(F) & \text{(C.18)} \\
&\quad + a\sum_{t=1}^{T(a)} \bar{b}(\beta_{t-1}) - \int_0^\tau \bar{b}(\beta(s))ds & \text{(C.19)} \\
&\quad + a\sum_{t=1}^{T(a)} \xi_t & \text{(C.20)}
\end{aligned}
$$

We examine (D.27), (D.28) and (D.29) one by one.

C.2. Convergence and Stability.

We can write

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

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

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

To simplify notation, we write

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

and $b(\tau)$ as the continuous process constructed from b_t via linear interpolation $\forall \tau \ge 0$. 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

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

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

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

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

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

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

$$\dot{\beta} = R^{-1} \begin{bmatrix} 1 - F(b) - \beta_0 - \beta_1 b + \mathsf{O}(\epsilon^2) \\ b\left[1 - F(b) - \beta_0 - \beta_1 b\right] - \left[f(b) + \beta_1\right]\sigma_1^2 + \mathsf{O}(\epsilon^3) \end{bmatrix}. \tag{C.21}$$

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

$$1 - F\left(b^*(F)\right) = \frac{\beta_0}{2} - \mathsf{O}(\epsilon^2)$$

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

Thus,

$$\frac{1 - F\left(b^*(F)\right)}{f\left(b^*(F)\right)} = b^*(F) + \mathsf{O}(\epsilon). \tag{C.22}$$

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

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

$|\mathsf{O}(x)| \le \eta|x|$. 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

$$\dot{b} = -\frac{f(b)}{2\beta_1}\left[\frac{1-F(b)}{f(b)} - b\right] \equiv \mathsf{R}(b) \qquad\qquad (\text{C.23})$$

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

Lemma C.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 (D.30) 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

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

$\square$

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 $\tau = \overline{\tau}(\mu)$.

C.3. **Riemann Residual.** Let us consider (D.28)

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

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

$$\overline{a} = \frac{2\mu}{\overline{\tau}(\mu)\eta'} \tag{C.24}$$

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

$$\mathbf{R}(a, F) \leq \mu. \tag{C.25}$$

Thus,

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

which implies

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

C.4. **Lower Bound of Confidence.** Next, we examine (D.29). Consider $\frac{\xi_t}{\xi}$, which is a martingale difference with

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

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

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

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 $\forall \mu' \in (0, 1/2)$,

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

where

$$\mathsf{H}(p \mid p_0) = p\log\frac{p}{p_0} + (1-p)\log\frac{1-p}{1-p_0} \tag{C.26}$$

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

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

Then,

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

Let

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

After substitution, we have

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

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

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

C.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 (C.22) of ODE is within μ neighborhood of $\beta^*(F)$ $\forall F \in \mathcal{F}$.

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

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

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

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

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 (C.22). We chose $\overline{\tau}(\mu)$ so that

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

Given $\mu > 0$, we chose $a_1 > 0$ in (C.24) so that the Riemann residual $\mathsf{R}(a, F)$ satisfies (C.25).

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 5.1. Let us summarize what we have at this point.

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

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

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

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

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

and therefore,

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

Since

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

$T(\bar{a}, \mu)$ 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 C.2 implies that $\mathcal{A}_a$ uniformly learns $\mathcal{F}^\eta$.

Appendix D. Proof of Theorem 8.1

To make the paper self-contained, we present the proof, although we are essentially replicating the proof of Theorem 5.1.

D.1. **Preliminaries.** Following Kushner and Yin (1997), we construct the (fictious) time from the gain function $a_t = \frac{1}{t^\omega}$. 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

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

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

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

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

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

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

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

As shown in (C.21), 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

$$
\begin{aligned}
\beta_{m(K+\tau)} - \beta^*(F) &= \beta^K(0) + \sum_{t=t_K}^{m(K+\tau)} a_t \bar{b}(\beta_{t-1}) + \sum_{t=t_K}^{m(K+\tau)} a_t \xi_t - \beta^*(F) \\
&= \beta^K(0) + \int_0^\tau \bar{b}(\beta(s)) ds - \beta^*(F) & \text{(D.27)} \\
&\quad + \sum_{t=t_K}^{m(K+\tau)} a_t \bar{b}(\beta_{t-1}) - \int_0^\tau \bar{b}(\beta(s)) ds & \text{(D.28)} \\
&\quad + \sum_{t=t_K}^{m(K+\tau)} a_t \xi_t & \text{(D.29)}
\end{aligned}
$$

We examine (D.27), (D.28) and (D.29) one by one.

D.2. **Convergence and Stability.** Following Kushner and Yin (1997), we can write

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

which is often written as

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

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):

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

where

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

and

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

By following the same argument as in Section D.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

$$\dot{b} = -\frac{f(b)}{2\beta_1} \left[\frac{1 - F(b)}{f(b)} - b \right] \equiv \mathsf{R}(b) \tag{D.30}$$

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 C.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,

$$\overline{\tau}(\mu) = \sup_{(\beta_0(0), \beta_1(0)) \in \mathsf{B}} \tau(\mu) < \infty \tag{D.31}$$

and

$$\overline{\tau}(\mu) \sim -\log \mu \tag{D.32}$$

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,

$$m(K + \tau) = \mathsf{O}\left(\tau^{\frac{1}{1-\omega}}, K^{\frac{1}{1-w}} \right). \tag{D.33}$$

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.

D.3. **Riemann Residual.** Fix $\mu > 0$ and $\tau = \bar{\tau}(\mu)$ defined in (D.31). Let us consider (D.28)

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

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

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

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

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

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

$$\frac{1}{t_K^\omega} = \frac{2\mu}{\bar{\tau}(\mu)\eta'} \tag{D.34}$$

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

$$\mathbf{R}(t_K, \tau, F) \leq \mu. \tag{D.35}$$

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 (D.33), we conclude that $m(K + \bar{\tau}(\mu)) - t_K$ increases at the polynomial rate of $\frac{1}{\mu}$.

D.4. **Lower Bound of Confidence.** Next, we examine (D.29). Let $a_t = \frac{1}{t^\omega}$ where $0 < \omega < 1$. Fix K so that t_K satisfies (D.34). Then, choose $\bar{\tau}(\mu)$ according to (D.31). We need to show that $\exists \mathsf{h} > 0$ such that $\forall T \geq m(K + \bar{\tau}(\mu))$

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

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

Lemma D.1.
$$\left(\frac{e^\lambda + e^{-\lambda}}{2}\right)^{a_t} \geq \frac{e^{\lambda a_t} + e^{-\lambda a_t}}{2} \tag{D.36}$$

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. $\square$

We follow the proof of Corollary 2.4.7 from (Dembo and Zeitouni 1998).[17] We have

$$\mathbb{E}[e^{\lambda \sum_{t=1}^T a_t \xi_t / \xi}] \leq \prod_{t=1}^T \frac{e^{-a_t \lambda} + e^{a_t \lambda}}{2} \leq \left(\frac{e^\lambda + e^{-\lambda}}{2}\right)^{\sum_{t=1}^T a_t} \tag{D.37}$$

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 (D.37).

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

$$\mathbb{P}\left(\frac{1}{\sum_{t=1}^T a_t}\left|\sum_{t=1}^T a_t \xi_t / \xi\right| \geq \mu\right) \leq e^{-\left(\sum_{t=1}^T a_t\right)\mathsf{H}\left(\frac{1+\mu}{2}\mid\frac{1}{2}\right)}. \tag{D.38}$$

where H is the relative entropy of the binomial distribution defined in (C.26). 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 (D.38) larger. Putting this together, we finally obtain

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

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

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

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

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

---

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

follows.

APPENDIX E. PROOF OF PROPOSITION 8.3

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 $\forall \delta \in (\underline{\delta}, 1)$, (8.17) holds. Choose $\lambda = \frac{\epsilon}{2} < \epsilon$. Since $\mathcal{A}$ is uniformly learnable, $\exists T$ such that

$$\mathbb{E} \left| \varphi_p(\mathcal{A}(\mathcal{O}_T))(1 - F(\varphi_p(\mathcal{A}(\mathcal{O}_T)))) - b^*(F)(1 - b^*(F)) \right| \leq \lambda \qquad \forall F \in \mathcal{F}^\eta. \tag{E.39}$$

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.

Suppose that $\mathcal{A}$ is an ϵ dominant strategy. Since $\mathcal{A}$ is an ϵ dominant strategy, $\forall \epsilon > 0$, $\underline{\delta}$ such that $\forall \delta \in (\underline{\delta}, 1)$,

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

Suppose that $\mathcal{A}$ is not uniformly learnable. Then, $\exists \lambda > 0$, $\forall T$, $\exists F \in \mathcal{F}^\eta$ such that

$$\mathbb{E} \left| \varphi_p(\mathcal{A}(\mathcal{O}_T))(1 - F(\varphi_p(\mathcal{A}(\mathcal{O}_T)))) - b^*(F)(1 - b^*(F)) \right| > \lambda.$$

or equivalently,

$$\mathbb{E} \varphi_p(\mathcal{A}(\mathcal{O}_T))(1 - F(\varphi_p(\mathcal{A}(\mathcal{O}_T)))) < b^*(F)(1 - b^*(F)) - \lambda. \tag{E.40}$$

Since the algorithm stops at T and the monopolists uses $\varphi_p(\mathcal{A}(\mathcal{O}_T))$ for the continuation game,

$$\mathcal{U}^\delta(\mathcal{A}, F) \leq (1 - \delta^T)\overline{\mathcal{U}} + \delta^T \left(b^*(F)(1 - b^*(F)) - \lambda \right)$$

where

$$\overline{\mathcal{U}} = \sup_{F \in \mathcal{F}^\eta} b^*(F)(1 - b^*(F)) < \infty.$$

Given $\lambda > 0$, fix T and F satisfying (E.40). We choose $\epsilon = \frac{\lambda}{2}$ and let $\delta \uparrow 1$ to satisfy

$$\mathcal{U}^\delta(\mathcal{A}, F) \leq (1 - \delta^T)\overline{\mathcal{U}} + \delta^T \left(b^*(F)(1 - b^*(F)) - \lambda \right) \leq b^*(F)(1 - b^*(F)) - \epsilon$$

which contradicts the hypothesis that $\mathcal{A}$ is an ϵ dominant strategy.

REFERENCES

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.

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.

MORRIS, S., AND D. BERGEMANN (2005): "Robust Mechanism Design," *Econometrica*, 73(6), 1771–1813.

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/