Dynamic Pricing with an Unknown Demand Model: Asymptotically Optimal Semi-myopic Policies

N. Bora Keskin*
University of Chicago

Assaf Zeevi†
Columbia University

This version: January 31, 2014

Abstract

We consider a monopolist who sells a set of products over a time horizon of T periods. The seller initially does not know the parameters of the products' linear demand curve, but can estimate them based on demand observations. We first assume that the seller knows nothing about the parameters of the demand curve, and then consider the case where the seller knows the expected demand under an incumbent price. It is shown that the smallest achievable revenue loss in T periods, relative to a clairvoyant who knows the underlying demand model, is of order $\sqrt{T}$ in the former case and of order $\log T$ in the latter case. To derive pricing policies that are practically implementable, we take as our point of departure the widely used policy called greedy iterated least squares (ILS), which combines sequential estimation and myopic price optimization. It is known that the greedy ILS policy itself suffers from incomplete learning, but we show that certain variants of greedy ILS achieve the minimum asymptotic loss rate. To highlight the essential features of well-performing pricing policies, we derive sufficient conditions for asymptotic optimality.

Keywords: Revenue management, pricing, sequential estimation, exploration-exploitation.

1 Introduction

Pricing decisions often involve a trade-off between *learning* about customer behavior to increase long-term revenues, and *earning* short-term revenues. In this paper we examine that trade-off.

Overview of the problem. In our learning-and-earning problem there is a seller that can adjust the price of its product over time. Customer demand for the seller's product is determined by the price, according to an underlying parametric *demand model*. The seller is initially uncertain about the parameters of the demand model but can use price as a learning tool, estimating the

*Booth School of Business, e-mail: `bora.keskin@chicagobooth.edu`
†Graduate School of Business, e-mail: `assaf@gsb.columbia.edu`

model parameters based on the empirical market response to successive prices. This is a dynamic pricing problem with *demand model uncertainty*.

A particular example that motivates our study is the pricing of financial services, such as consumer and auto loans. As noted by Phillips (2005, pp. 266-267), sellers of consumer lending products are typically capable of quoting a different interest rate to each customer who requests a loan. This transaction structure, which Phillips calls *customized pricing*, creates a price experimentation opportunity for the seller. At the same time, price experimentation does not seem to be very common in practice (Phillips 2010); for many consumer lending products, a bank will very often keep the interest rate that it charges fixed, over extended periods of time. It then knows with high confidence the average market response to that one price, referred to hereafter as the *incumbent price*, but may be uncertain about market response to alternative prices. In this paper we study several closely related models of dynamic pricing with demand model uncertainty, both with and without prior knowledge of market response to an incumbent price. The models employed are somewhat stylized, intended to provide fundamental insight while lacking the fine structure needed in real-world applications.

Another feature of consumer lending that influences our problem formulation is high transaction volume: a commercial bank typically encounters hundreds of lending opportunities per day. In our model the horizon length T can be thought of as a proxy for the number of potential sales, and "demand" can be thought of as the portion of these that are successfully captured. To assess the quality of any pricing policy, we focus on the asymptotic growth of *regret*: the expected loss of revenue relative to a clairvoyant who knows the underlying demand model, as the horizon length T becomes large. When we speak later of asymptotically optimal policies, this will mean policies that achieve a minimal growth rate of regret.

Focus on semi-myopic policies. According to a report by the Boston Consulting Group, myopic pricing policies are common practice in many industries, and pricing executives usually do not engage in price experimentation or other forms of active learning (Morel, Stalk, Stanger and Wetenhall 2003). A conventional procedure is to simply combine greedy price optimization with sequential estimation as follows: whenever a seller needs to adjust its price, the first step is to estimate the parameters of a demand model, and then the seller charges whatever price maximizes expected profit, assuming that the unknown model parameters are equal to the most recent estimates. As will be explained later, this approach leads to poor profit performance due to *incomplete learning*: because there is no explicit effort to create price dispersion, there is no guarantee that parameter estimates will converge to the true values. To provide practical insight, we use the myopic approach as a starting point, modifying it to guard against poor performance. It will be shown that the resulting *semi-myopic policies* are asymptotically optimal in a well defined mathematical sense.

Related literature. The trade-off between exploration (learning) and exploitation (earning) has long been a focus of attention in statistics, engineering, and economics. There has been recent interest in this topic in the operations research and management science (OR/MS) realm, especially in the field of dynamic pricing and revenue management. Economists were the first to formulate a learn-and-earn problem in a pricing context, but recent OR/MS work has a somewhat different flavor, emphasizing development of practical policies whose performance is provably good although not necessarily optimal.

Harrison, Keskin and Zeevi (2012) analyze the performance of myopic pricing and some of its variants in a stylized Bayesian setting where one of two possible demand models is known to be true. Under mild assumptions that ensure the existence of just one "uninformative" price (that is, just one price that fails to distinguish between the two demand hypotheses), they show that the myopic price sequence converges either to the uninformative price or else to the true optimal price, each with positive probability. Having thus established that myopic pricing may lead to incomplete learning, they go on to prove positive results about variants of myopic pricing that avoid use of the uninformative price. Harrison et al. (2012) also discuss in detail how recent OR/MS papers on learning-and-earning relate to the antecedent economics literature, and we refer to their paper for a more complete list of references. Two other closely related papers on dynamic pricing with demand model uncertainty are due to Broder and Rusmevichientong (2012) and den Boer and Zwart (2013). Broder and Rusmevichientong (2012) consider a general parametric demand model to evaluate the performance of maximum-likelihood-based policies, and concentrate on the idea of combining the myopic approach with explicit price experimentation. Dividing the sales horizon into cycles of increasing length, they design a policy that applies fixed experimental prices at the beginning of each cycle, and charges the myopic maximum-likelihood-based price during the rest of the same cycle. They show that their policy is asymptotically optimal. The framework of den Boer and Zwart (2013) involves a demand model with two unknown parameters. Using a myopic maximum-likelihood-based policy as a starting point, den Boer and Zwart design a variant of the myopic policy, namely the controlled variance pricing policy, by creating taboo intervals around uninformative prices, and derive a theoretical bound on the performance of this variant. In a more recent study that has been conducted in parallel to our work, den Boer (2013) extends the controlled variance pricing policy of den Boer and Zwart (2013) to the case of multiple products with a generalized linear demand curve. A common theme in these papers is that they all consider some particular parametric family of pricing policies to find a good balance between learning and earning. In contrast, we provide in this paper *general* sufficient conditions that achieve the same goal, which sheds further light on the role and impact of price experimentation.

There are two other noteworthy differences between our work and that of Broder and Rusmevichientong (2012). First, like den Boer and Zwart (2013), Broder and Rusmevichientong study the pricing of a single product, whereas we extend our analysis to the case of many products with substitutable demand. As will be explained in detail later, learning and earning in higher dimensions requires extra care. Broder and Rusmevichientong (2012) also derive logarithmically growing regret bounds in a particular variant of their original problem, which they call the "well-separated" case. At least superficially, this might seem similar to our analysis of the incumbent-price problem, but that appearance is somewhat deceptive: at least in our view, the incumbent-price formulation is better motivated from a practical standpoint than the well-separated problem, and while the performance bounds alluded to above are of similar nature, the analysis is distinctly different in the two settings.

There are also OR/MS studies that take on different approaches to dynamic pricing to learn and earn. Kleinberg and Leighton (2003) employ a multi-armed bandit approach in a dynamic pricing context by allowing the seller to use only a discrete grid of prices within the continuum of feasible prices. They show that certain methods in multi-armed bandit literature work well for a carefully chosen price grid. A key feature of this study is that the seller needs to know the time horizon before the problem starts, so that he will be able to choose a sufficiently fine grid of prices. Carvalho and Puterman (2005) consider dynamic pricing problem with a binomial-logistic demand curve, and analyze the performance of several pricing policies. They conduct a simulation study to demonstrate that one-step lookahead policies could perform significantly better than myopic policies.

The particular myopic policy we examine in this paper is *greedy iterative least squares* (ILS), which combines the myopic approach mentioned above with sequential estimation based on ordinary least squares regression. Greedy ILS was first formulated by Anderson and Taylor (1976) in a problem motivated by medical applications. Based on simulation results, Anderson and Taylor argued that the plain version of greedy ILS could be used in regulation problems, such as stabilizing the response of patients to the dosage level of a given drug (see Lai and Robbins 1979, Section 1 for further details about this medical motivation). There are a few studies that demonstrate that the parameter estimates of greedy ILS are consistent when either charging any feasible price provides information about the unknown response curve, or greedy ILS charges only the prices that provide information (cf., e.g., Taylor 1974, Broder and Rusmevichientong 2012). However, as shown by Lai and Robbins (1982), the parameter estimates computed under greedy ILS may not be consistent, and the resulting controls may be sub-optimal. Because of this negative result, there is a clear need to "tweak" the plain version of greedy ILS in order to generate well-performing policies.

Main contributions and organization of the paper. Our study makes four main contributions to the literature on learning-and-earning. First, our formulation of the incumbent-price problem, where the seller starts out knowing one point on its demand curve, is both novel and widely applicable, at least as an approximation. As will be shown later, our conclusions for the incumbent-price problem differ sharply from those for the conventional problem. Second, unlike previous studies on this subject, our paper provides *general* sufficient conditions for asymptotic optimality of semi-myopic policies, providing general guidelines for implementing price experimentation strategies. Said optimality conditions are obtained using analytical tools that apply broadly to problems of dynamic pricing with model uncertainty. Our third contribution is the development of these tools by (i) using the concept of *Fisher information* to find natural and intuitive lower bounds on regret, and (ii) using martingale theory to control pricing errors, and construct semi-myopic policies whose order of regret matches these lower bounds. Finally, our analysis covers the pricing of multiple products with substitutable demand, which, as will be seen in what follows, is not a straightforward extension of single-product pricing. In essence, learning in a high dimensional price space requires sufficient "variation" in the directions of successive price vectors, and the necessity of spanning the price space brings forth the idea of *orthogonal pricing*.

Our paper is organized as follows. In Section 2 we formulate our problem. In Section 3 we analyze single-product pricing by first constructing a lower bound on the revenue loss of any given policy, and then providing a set of sufficient conditions for asymptotic optimality. To facilitate practical implementation of our results, we also give some examples of policies that meet these asymptotic optimality conditions. Toward the end of Section 3, we analyze the incumbent-price problem explained above, and finish that section by comparing the results obtained in the conventional setting with the ones obtained in the incumbent-price setting. In Section 4 the single-product results are generalized to a multi-product setting. Section 5 summarizes the main contributions of this paper. The proofs of formal results are postponed to the appendices.

2 Problem Formulation

Basic model elements. Consider a firm, hereafter called *the seller*, that sells n distinct products over a time horizon of T periods. In each period $t = 1, 2, \ldots$ the seller must choose an $n \times 1$ price vector $p_t = (p_{1t}, \ldots, p_{nt})$ from a given feasible set $[\ell, u]^n \subseteq \mathbb{R}^n$, where $0 \leq \ell < u < \infty$, after which the seller observes the demand D_{it} for each product i in period t. (In practice, different products may have different price intervals, and it is straightforward to generalize the following analysis to the case where the feasible set of prices is an arbitrary but fixed rectangle in $\mathbb{R}^n$.) In this paper we focus on linear demand models, which elucidate the key features of the problem by making our analysis more transparent (see Section 5 for further discussion of the linear demand assumption). Suppose that D_{it} is given by

$$D_{it} = \alpha_i + \beta_i \cdot p_t + \epsilon_{it} \qquad \text{for } i = 1, \ldots, n \text{ and } t = 1, 2, \ldots$$

where $\alpha_i \in \mathbb{R}$, $\beta_i = (\beta_{i1}, \ldots, \beta_{in}) \in \mathbb{R}^n$ are the demand model parameters, which are initially unknown to the seller, ϵ_{it} are unobservable demand shocks, and $x \cdot y = \sum_{i=1}^{n} x_i y_i$ denotes the inner product of vectors x and y. We assume that $\{\epsilon_{it}\}_{i=1,\ldots,n,\, t=1,\ldots,T}$ are independent and identically distributed random variables with mean zero and variance σ^2, and that there exists a positive constant z_0 such that $\mathbb{E}[e^{z\epsilon_{it}}] < \infty$ for all $|z| \leq z_0$. (An important example is where $\epsilon_{it} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$; in treating this case, it is more realistic to consider small values of σ so that the probability of observing a negative demand is negligible.) To simplify notation, we let $\theta := (\alpha_1, \beta_1, \ldots, \alpha_n, \beta_n) \in \mathbb{R}^{n^2+n}$ be the vector of demand model parameters for all products, and express the demand vector $D_t = (D_{1t}, \ldots, D_{nt}) \in \mathbb{R}^n$ in terms of θ as follows:

$$D_t = a + Bp_t + \epsilon_t \qquad \text{for } t = 1, 2, \ldots \tag{2.1}$$

where $\epsilon_t := (\epsilon_{1t}, \ldots, \epsilon_{nt}) \in \mathbb{R}^n$, $a := (\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^n$, and B is an $n \times n$ matrix whose i^{th} row is β_i^{T}, that is,

$$B = \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1n} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{n1} & \beta_{n2} & \cdots & \beta_{nn} \end{bmatrix}.$$

Because prices can always be expressed as increments above marginal production cost, we assume without loss of generality that the marginal cost of production is zero, and use the terms "profit" and "revenue" interchangeably. The seller's expected single-period revenue function is

$$r_\theta(p) := p \cdot (a + Bp) \qquad \text{for } \theta \in \mathbb{R}^{n^2+n} \text{ and } p \in [\ell, u]^n. \tag{2.2}$$

Let $\varphi(\theta)$ be the price that maximizes the expected single-period revenue function $r_\theta(\cdot)$, i.e.,

$$\varphi(\theta) := \arg\max\{r_\theta(p) : p \in [\ell, u]^n\}. \tag{2.3}$$

Before selling starts, nature draws the value of θ from a compact rectangle $\Theta \subseteq \mathbb{R}^{n^2+n}$. We assume that the optimal price corresponding to each $\theta \in \Theta$ is an interior point of the feasible set $[\ell, u]^n$, implying that B is negative definite. (If B were not negative definite, then there would exist a non-zero vector z such that $z^{\mathsf{T}} B z \geq 0$, and we would get a corner solution by changing the price in the direction of z.) Therefore, the first-order condition for optimality is

$$\eta(\theta, p) \;:=\; a + \big(B + B^{\mathsf{T}}\big)p \;=\; 0, \tag{2.4}$$

from which we deduce that $\varphi(\theta) = -\big(B + B^{\mathsf{T}}\big)^{-1} a \;\in\; (\ell, u)^n$. If the seller has only one product (that is, $n = 1$), then the revenue-maximizing price simply becomes $\varphi\big((\alpha, \beta)\big) = -\alpha/(2\beta)$.

Because Θ is a compact set and B is negative definite, we have the following: for all $\theta \in \Theta$ the eigenvalues of the matrix B are in the interval $[b_{\min}, b_{\max}]$, where $-\infty < b_{\min} < b_{\max} < 0$. In the single-product case, this fact implies that the slope parameter β is contained in $[b_{\min}, b_{\max}]$.

Pricing policies, induced probabilities, and performance metric. Denote by H_t the vectorized history of demands and prices observed through the end of period t. That is, $H_t :=$

$(D_1, p_1, \ldots, D_t, p_t)$. We define a *policy* as a sequence of functions $\pi = (\pi_1, \pi_2, \ldots)$, where π_{t+1} is a measurable mapping from $\mathbb{R}^{2nt}$ into $[\ell, u]^n$ for all $t = 1, 2, \ldots,$ and π_1 is a constant function. The argument of the pricing function π_{t+1} is H_t. The seller exercises a policy π to construct a nonanticipating price sequence $p = (p_1, p_2, \ldots)$, charging price vector p_t in period t. This definition of a pricing policy implies that each p_t is adapted to $H_{t-1} = (D_1, p_1, \ldots, D_{t-1}, p_{t-1})$.

Any pricing policy induces a family of probability measures on the sample space of demand sequences $D = (D_1, D_2, \ldots)$. Given the parameter vector $\theta = (\alpha_1, \beta_1, \ldots, \alpha_n, \beta_n)$ and a policy π, let $\mathbb{P}_\theta^\pi$ be a probability measure satisfying

$$\mathbb{P}_\theta^\pi(D_1 \in d\xi_1, \ldots, D_T \in d\xi_T) \;=\; \prod_{t=1}^{T} \prod_{i=1}^{n} \mathbb{P}_\epsilon(\alpha_i + \beta_i \cdot p_t + \epsilon_{it} \in d\xi_{it}) \qquad \text{for } \xi_1, \ldots, \xi_T \in \mathbb{R}^n, \quad (2.5)$$

where $\mathbb{P}_\epsilon(\cdot)$ is the probability measure of random variables ϵ_{it}, and $p = (p_1, p_2, \ldots)$ is the price sequence formed under policy π and demand realizations $D_1, D_2, \ldots$ In our context, $\mathbb{P}_\theta^\pi(A)$ is the probability of an event A given that the parameter vector is θ and the seller uses policy π. Thus, $p_t = \pi_t(H_{t-1})$ almost surely under $\mathbb{P}_\theta^\pi$, which implies that p_t is completely characterized by π, θ, and H_{t-1}. Accordingly the expected T-period revenue of the seller is

$$R_\theta^\pi(T) = \mathbb{E}_\theta^\pi\Big\{ \sum_{t=1}^{T} r_\theta(p_t) \Big\}, \quad (2.6)$$

where $\mathbb{E}_\theta^\pi$ is the expectation operator associated with the probability measure $\mathbb{P}_\theta^\pi$. The performance metric we will use throughout this study is the T-period *regret*, defined as

$$\Delta_\theta^\pi(T) = T r_\theta^* - R_\theta^\pi(T) \qquad \text{for } \theta \in \Theta \text{ and } T = 1, 2, \ldots \quad (2.7)$$

where $r_\theta^* := r_\theta\big(\varphi(\theta)\big)$ is the optimal expected single-period revenue. While deriving our lower bounds on best achievable performance, we will also make use of the worst-case regret, which is given by

$$\Delta^\pi(T) = \sup\big\{\Delta_\theta^\pi(T) : \theta \in \Theta\big\} \qquad \text{for } T = 1, 2, \ldots \quad (2.8)$$

The regret of a policy can be interpreted as the expected revenue loss relative to a clairvoyant policy that knows the value of θ in period 0; smaller values of regret are more desirable for the seller.

Greedy iterated least squares (ILS). Given the history of demands and prices through the end of period t, the least squares estimator of θ is given by

$$\widehat{\theta}_t = \arg\min_\theta \{SSE_t(\theta)\}, \quad (2.9)$$

where $SSE_t(\theta) = \sum_{s=1}^{t} \sum_{i=1}^{n} (D_{is} - \alpha_i - \beta_i \cdot p_s)^2$ for $\theta = (\alpha_1, \beta_1, \ldots, \alpha_n, \beta_n)$. Using the first-order optimality conditions of the least squares problem (2.9), the estimator $\widehat{\theta}_t$ can be expressed explicitly as $\widehat{\theta}_t = (\widehat{\alpha}_{1t}, \widehat{\beta}_{1t}, \ldots, \widehat{\alpha}_{nt}, \widehat{\beta}_{nt})$ where

$$\begin{bmatrix} \widehat{\alpha}_{it} \\ \widehat{\beta}_{it} \end{bmatrix} = \begin{bmatrix} t & \sum_{s=1}^t p_s^\mathsf{T} \\ \sum_{s=1}^t p_s & \sum_{s=1}^t p_s p_s^\mathsf{T} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{s=1}^t D_{is} \\ \sum_{s=1}^t D_{is} p_s \end{bmatrix}$$

for $i = 1, \ldots, n$. Combining the preceding least squares formula with the demand equation (2.1), one has the following:

$$\begin{bmatrix} \widehat{\alpha}_{it} - \alpha_i \\ \widehat{\beta}_{it} - \beta_i \end{bmatrix} = \mathcal{J}_t^{-1} \mathcal{M}_{it} \quad \text{for all } i = 1, \ldots, n \text{ and } t = n+1, n+2, \ldots, \tag{2.10}$$

where $\mathcal{J}_t$ is the empirical Fisher information given by

$$\begin{aligned} \mathcal{J}_t &= \begin{bmatrix} t & \sum_{s=1}^t p_s^\mathsf{T} \\ \sum_{s=1}^t p_s & \sum_{s=1}^t p_s p_s^\mathsf{T} \end{bmatrix} \\ &= \sum_{s=1}^t \left(\left[\begin{smallmatrix} 1 \\ p_s \end{smallmatrix}\right] \cdot \left[\begin{smallmatrix} 1 \\ p_s \end{smallmatrix}\right]^\mathsf{T} \right), \end{aligned} \tag{2.11}$$

and $\mathcal{M}_{it}$ is the $(n+1)\times 1$ vector, $\mathcal{M}_{it} := \sum_{s=1}^t \left(\epsilon_{is} \left[\begin{smallmatrix} 1 \\ p_s \end{smallmatrix}\right] \right)$. With a single product, the estimation error expressed in (2.10) reduces to $\widehat{\theta}_t - \theta = \mathcal{J}_t^{-1} \mathcal{M}_t$, where $\mathcal{J}_t = \sum_{s=1}^t \left[\begin{smallmatrix} 1 & p_s \\ p_s & p_s^2 \end{smallmatrix}\right]$ and $\mathcal{M}_t = \sum_{s=1}^t \left[\begin{smallmatrix} \epsilon_s \\ \epsilon_s p_s \end{smallmatrix}\right]$.

Because θ lies in the compact rectangle Θ, the accuracy of the unconstrained least squares estimate $\widehat{\theta}_t$ can be improved by projecting it into the rectangle Θ. We denote by ϑ_t the truncated estimate that satisfies $\vartheta_t := \arg\min_{\vartheta \in \Theta} \{\|\vartheta - \widehat{\theta}_t\|\}$, where by assumption the corresponding price vector $\varphi(\vartheta_t)$ is an interior point of the feasible set $[\ell, u]^n$. We call the policy that charges price $p_t = \varphi(\vartheta_{t-1})$ in period $t = 1, 2, \ldots$ the *greedy iterated least squares* (ILS) policy. Because greedy ILS cannot estimate anything without data, we fix $n+1$ non-random and linearly independent vectors for $p_1, p_2, \ldots, p_{n+1}$, and (whenever necessary) $\vartheta_0, \vartheta_1, \ldots, \vartheta_n$ can be found accordingly.

In the following section we will focus on the case $n = 1$, and explain in Section 4 how the single-product analysis extends to the case of multiple products with substitutable demand.

3 Analysis of the Single-Product Setting

To introduce our main ideas in a simple setting we assume throughout this section that the seller has a single product to sell (that is, $n = 1$), and consequently drop the product index i from the demand parameters α_i and β_i.

3.1 A Lower Bound on Regret

We use the case where $\epsilon_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$ to derive a lower bound on regret. In fact, our analysis for all lower bounds on regret is valid for a broader exponential family of distributions whose densities have the following parametric form: $f_\epsilon(\xi \,|\, \sigma) = c(\sigma)\exp\big(h(\xi) + \sum_{j=1}^k w_j(\sigma) t_j(\xi)\big)$ for $\xi \in \mathbb{R}$, where $h(\cdot)$ and $t_j(\cdot)$ are differentiable functions. This family includes a wide range of distributions such as Gaussian, log-normal, exponential, and Laplace distributions. The last paragraph of Section A.1 explains how our proof of the lower bound with Gaussian noise can be extended to the more general setting.

All the information that a pricing policy can use is that contained in the history vector H_t, which can be quantified as follows. Given $\theta = (\alpha, \beta)$, the density of H_t is

$$g_t(H_t, \theta) = \prod_{s=1}^{t} \frac{1}{\sigma}\,\phi\Big(\frac{D_s - \alpha - \beta p_s}{\sigma}\Big), \tag{3.1}$$

where $\phi(\cdot)$ denotes the standard Gaussian density. Thus, H_t has the following Fisher information matrix:

$$\mathcal{I}_t^\pi(\theta) \;:=\; \mathbb{E}_\theta^\pi\bigg\{\Big[\frac{\partial \log g_t(H_t,\theta)}{\partial\theta}\Big]^{\mathsf{T}}\frac{\partial \log g_t(H_t,\theta)}{\partial\theta}\bigg\} \;=\; \frac{1}{\sigma^2}\,\mathbb{E}_\theta^\pi\left[\mathcal{J}_t\right]. \tag{3.2}$$

To obtain a lower bound on regret, the following lemma is key.

Lemma 1 (lower bound on cumulative pricing error) *There exist finite positive constants c_0 and c_1 such that*

$$\sup_{\theta\in\Theta}\Big\{\textstyle\sum_{t=2}^T \mathbb{E}_\theta^\pi\big(p_t - \varphi(\theta)\big)^2\Big\} \;\geq\; \displaystyle\sum_{t=2}^{T} \frac{c_0}{c_1 + \sup_{\theta\in\Theta}\big\{C(\theta)\,\mathbb{E}_\theta^\pi\left[\mathcal{J}_{t-1}\right]\,C(\theta)^{\mathsf{T}}\big\}}\,, \tag{3.3}$$

where $C(\cdot)$ is a 1×2 matrix function on Θ such that $C(\theta) = [-\varphi(\theta) \;\; 1]$.

Lemma 1, whose proof relies on a multivariate version of the van Trees inequality (cf. Gill and Levit 1995, p. 64), expresses the manner in which the growth rate of regret depends on accumulated information. If the denominator on the right side of inequality (3.3) converges to a constant, then information acquisition stops prematurely (incomplete learning), and hence regret would grow linearly over time. Thus, one would expect a "good" policy to make $\sup_{\theta\in\Theta}\big\{C(\theta)\,\mathbb{E}_\theta^\pi\left[\mathcal{J}_{t-1}\right]\,C(\theta)^{\mathsf{T}}\big\}$ diverge to infinity for any given non-zero 1×2 matrix function $C(\cdot)$. If this quantity grows linearly, then we would arrive at a logarithmically growing lower bound on regret. However, if we specify the function $C(\cdot)$ as in Lemma 1, we realize that a logarithmic lower bound is in fact loose. For $C(\theta) = [-\varphi(\theta) \;\; 1]$, the quantity in the supremum on the right hand side of (3.3) becomes $\sum_{s=1}^t \mathbb{E}_\theta^\pi\big(p_s - \varphi(\theta)\big)^2$, which is almost identical to the quantity in the supremum on the left hand side. Viewing these quantities as a time-indexed sequence of real numbers, we then deduce that each term $\sup_{\theta\in\Theta}\big\{\sum_{t=1}^T \mathbb{E}_\theta^\pi\big(p_t - \varphi(\theta)\big)^2\big\}$ is at least of order $\sqrt{T}$, which leads to the following conclusion.

Theorem 1 (lower bound on regret) *There exists a finite positive constant c such that $\Delta^\pi(T) \geq c\sqrt{T}$ for any pricing policy π and time horizon $T \geq 3$.*

Remark The constant c in the above theorem, and all the constants that will appear in the following theorems are spelled out explicitly in the appendices. See Section 3.5 for a discussion of how these constants depend on characteristics of the demand shocks.

According to Theorem 1, the T-period regret of any given policy must be at least of order $\sqrt{T}$. A policy π for which the growth rate of regret does not exceed order $\sqrt{T}$ (up to logarithmic terms) will be called *asymptotically optimal*. In the next subsection we explain how greedy ILS fails to achieve this benchmark while certain variants of greedy ILS do achieve it. With slight abuse of terminology, the minimum asymptotic growth rate for regret will also be referred to as the *minimum asymptotic loss rate*.

Besbes and Zeevi (2011) and Broder and Rusmevichientong (2012) provide lower bounds on regret that grow proportional to $\sqrt{T}$ in similar dynamic pricing contexts with demand model uncertainty. The argument used in those papers relies on the Kullback-Leibler divergence, and deals with uncertainty in a single model parameter. Our proof of Theorem 1 employs an entirely different argument that uses the van Trees inequality and the concept of Fisher information, and we derive lower bounds for parameter uncertainty in two and higher dimensions.

3.2 Sufficient Conditions for Asymptotic Optimality

We now derive a set of conditions which assure that the regret of a policy is asymptotically optimal. Our starting point is the greedy ILS policy, which is commonly used in practice. As explained in Section 2, the operating principles of greedy ILS are (i) estimating at the beginning of each time period the unknown model parameters via ordinary least squares regression, and (ii) charging in that period the "greedy" or "myopic" price that would be appropriate if the most recent least squares estimates were precisely correct. Although the iterative use of this estimate-and-then-optimize method seems reasonable, it does not take into account the fact that the myopic price optimization routine will be repeated in the future. There is a threat that the greedy price will not generate enough information, causing the seller to stop learning prematurely. Thus, greedy ILS estimates may get stuck at parameter values that are not the true ones. This phenomenon, which is called *incomplete learning*, was first analyzed by Lai and Robbins (1982) in a sequential decision problem motivated by medical treatments. Recently, den Boer and Zwart (2013) showed that the same result holds in a dynamic pricing context like ours. An important consequence of incomplete learning is that the seller forfeits a constant fraction of expected revenues in each period, causing regret to grow linearly over time.

In modifying greedy ILS, our main concern is to make sure that information is gathered at an adequate rate. One measure of information is the minimum eigenvalue of the empirical information matrix $\mathcal{J}_t$, denoted by $\mu_{\min}(t)$. Although it is possible to use $\mu_{\min}(t)$ as the sole information metric in the remainder of our analysis, we choose to employ a more practical information metric, which is closely related to $\mu_{\min}(t)$. The information metric we will use is the sum of squared price deviations, which is expressed as

$$J_t \;=\; \sum_{s=1}^{t}(p_s - \overline{p}_t)^2 \,,$$

where $\overline{p}_t = t^{-1}\sum_{s=1}^{t} p_s$. The following lemma relates $\mu_{\min}(t)$ to J_t. (For this lemma, readers are reminded that ℓ and u are a given lower bound and a given upper bound, respectively, on prices.)

Lemma 2 (minimum eigenvalue of Fisher information) *Let $\mu_{\min}(t)$ be the smallest eigenvalue of $\mathcal{J}_t$. Then $\mu_{\min}(t) \geq \gamma J_t$, where $\gamma = 2/(1+2u-\ell)^2$.*

An immediate implication of the above result is that one can control the rate of information acquisition by controlling the process $\{J_t\}$. We use this observation to show that the least squares estimation errors decrease exponentially as a function of J_t.

Lemma 3 (exponential decay of estimation errors) *There exist finite positive constants ρ and k such that, under any pricing policy π,*

$$\mathbb{P}_\theta^\pi\big\{\|\widehat{\theta}_t - \theta\| > \delta,\; J_t \geq m\big\} \;\leq\; kt\exp\big(-\rho(\delta \wedge \delta^2)m\big)\,, \tag{3.4}$$

for all $\delta, m > 0$ and all $t \geq 2$.

Lemma 3 shows that, by imposing carefully designed time-dependent constraints on the growth of the process J_t, the seller can directly control how fast estimation errors converge to zero. This leads to the following recipe for asymptotic optimality.

Theorem 2 (sufficient conditions for asymptotic optimality) *Assume that $\theta \in \Theta$. Let κ_0, κ_1 be finite positive constants, and let π be a pricing policy that satisfies*

(i) $J_t \geq \kappa_0\sqrt{t}$

(ii) $\sum_{s=0}^{t}\big(\varphi(\vartheta_s) - p_{s+1}\big)^2 \leq \kappa_1\sqrt{t}$

almost surely for all t. Then there exist a finite positive constant C such that $\Delta_\theta^\pi(T) \leq C\sqrt{T}\log T$ for all $T \geq 3$.

A simple verbal paraphrase of the preceding theorem is the following: a policy will perform well if it (i) accumulates information at an adequate rate, and (ii) does not deviate "too much" from greedy pricing. The conditions laid out in this theorem describe the (asymptotically) optimal balance between learning and earning. Condition (i) takes care of learning by forcing the information metric J_t to grow at an adequate rate, whereas condition (ii) limits the deviations from the greedy ILS price, making sure that enough emphasis is given to earning. Failure to satisfy either one of these conditions may lead to poor revenue performance. One extreme example is the greedy ILS policy, which puts all emphasis on earning. Greedy ILS satisfies condition (ii) trivially, but violates condition (i) because greedy ILS prices can get stuck at an incorrect value due to the incomplete learning phenomenon, leading to a regret of order T. Another extreme example is conducting price experiments with a constant frequency. If a policy charges a test price for a constant fraction of periods, the number of price experiments in t periods would grow linearly in t, making J_t satisfy condition (i). However, such a policy cannot satisfy condition (ii) because the fixed frequency of price experiments forces the quantity on the left side of (ii) to increase linearly in t. This second extreme policy also has a regret of order T. See Section 3.5 for a discussion of the performance of policies that lie between these two extremes but still violate conditions of Theorem 2.

Unlike previous results on learning and earning, Theorem 2 covers a wide range of policies, showing that any ILS variant satisfying its conditions will perform well. The "MLE-cycle" policy of Broder and Rusmevichientong (2012) and "controlled-variance" policy of den Boer and Zwart (2013) are two examples. To satisfy the conditions of Theorem 2, the former policy charges a set of pre-determined test prices in the beginning of pre-determined cycles of increasing length, while the latter enforces deviations from the greedy ILS price with a decaying magnitude.

3.3 Asymptotically Optimal ILS Variants

To demonstrate the practical use of Theorem 2, we now construct two families of policies, similar in structure to those described immediately above, that satisfy its sufficient conditions. Because these policy families are obtained by modifying the greedy ILS policy in different ways, we call them *ILS variants*.

Example 1: Constrained variant of ILS. The *constrained iterated least squares* (CILS) policy adjusts the greedy ILS price whenever its deviation from the historical average price is not large enough. Let δ_s denote the difference between the greedy ILS price in period s and the average price in the first $s-1$ periods, that is $\delta_s := \varphi(\vartheta_{s-1}) - \overline{p}_{s-1}$. In each period t, the CILS policy with threshold parameter κ, hereafter abbreviated CILS(κ), charges the price

$$p_t = \begin{cases} \overline{p}_{t-1} + \mathrm{sgn}(\delta_t)\kappa t^{-1/4} & \text{if } |\delta_t| < \kappa t^{-1/4} \\ \varphi(\vartheta_{t-1}) & \text{otherwise.} \end{cases} \tag{3.5}$$

To show that any policy π in the CILS family satisfies the conditions of Theorem 2, note that the information metric J_t can be expressed as

$$J_t = \sum_{s=2}^{t} (1 - s^{-1})(p_s - \overline{p}_{s-1})^2.$$

By construction, the CILS policy with parameter κ has $(p_s - \overline{p}_{s-1})^2 \geq \kappa^2 s^{-1/2}$ for all s, implying that $J_t \geq \frac{1}{4}\kappa^2 t^{1/2}$. Hence, condition (i) of Theorem 2 is satisfied for $\kappa_0 = \frac{1}{4}\kappa^2$. Moreover, deviations from the greedy ILS price have the following deterministic upper bound: $|\varphi(\vartheta_{s-1}) - p_s| \le \kappa s^{-1/4}$. Thus, condition (ii) is satisfied with $\kappa_1 = 2\kappa^2$. As a result, any policy in the CILS family achieves the performance guarantee in Theorem 2. Of course, one can tune the policy parameter κ to further improve the performance. As explained in the proof of Theorem 2, the constant in front of the $\sqrt{T}\log T$ term in the upper bound is $C_1 = 4|b_{\min}|(8\widetilde{K}_0\rho^{-1}\kappa^{-2} + \kappa^2)$, where $\widetilde{K}_0 = (1 + 5k)\max_{j=1,2}\big\{\max_\theta\{(\partial\varphi(\theta)/\partial\theta_j)^2\}\big\}$, so the policy parameter that minimizes the upper bound is $\kappa^* = (8\widetilde{K}_0/\rho)^{1/4}$.

The controlled-variance policy of den Boer and Zwart (2013) is an example that belongs to the CILS family, and the performance guarantee that den Boer and Zwart prove for regret under their policy is $\mathcal{O}(\sqrt{T}\log T)$.

Example 2: ILS with deterministic testing. Our second modification of greedy ILS, called *ILS with deterministic testing* (ILS-d), conducts price experiments to gather information. To specify the experimental prices and the periods at which experiments will take place, we let $\widetilde{p}_1$, $\widetilde{p}_2$ be two distinct prices in $[\ell, u]$, and $\mathcal{T}_{1,t}$, $\mathcal{T}_{2,t}$ be two sequences of sets satisfying the following conditions: for each i and t, suppose that $\mathcal{T}_{i,t} \subseteq \mathcal{T}_{i,t+1}$, and $\mathcal{T}_{1,t}$, $\mathcal{T}_{2,t}$ are disjoint subsets of $\{1, 2, \dots, t\}$, each containing $\lfloor \kappa\sqrt{t} \rfloor$ distinct elements. In period t, an ILS-d policy with threshold parameter κ and experimental prices $\widetilde{p}_1$ and $\widetilde{p}_2$, abbreviated ILS-d$(\kappa, \widetilde{p}_1, \widetilde{p}_2)$, charges the price

$$p_t = \begin{cases} \widetilde{p}_1 & \text{if } t \in \mathcal{T}_{1,t} \\ \widetilde{p}_2 & \text{if } t \in \mathcal{T}_{2,t} \\ \varphi(\vartheta_{t-1}) & \text{otherwise.} \end{cases} \tag{3.6}$$

An ILS-d policy with parameters κ, $\widetilde{p}_1$, $\widetilde{p}_2$ conducts $\lfloor \kappa\sqrt{t} \rfloor$ experiments in the first t periods, which implies that $J_t = \sum_{s=1}^{t}(p_s - \overline{p}_t)^2 \ge \sum_{s\in\mathcal{T}_{1,t}\cup\mathcal{T}_{2,t}}(p_s - \overline{p}_t)^2 \ge \frac{1}{4}(\widetilde{p}_1 - \widetilde{p}_2)^2\lfloor \kappa\sqrt{t} \rfloor$ for all t. Moreover, because ILS-d$(\kappa, \widetilde{p}_1, \widetilde{p}_2)$ deviates from the greedy ILS price in at most $\kappa\sqrt{t}$ periods, we have $\sum_{s=0}^{t}\big(\varphi(\vartheta_s) - p_{s+1}\big)^2 \le \kappa(u-\ell)^2\sqrt{t}$ for all t. Thus, ILS-d$(\kappa, \widetilde{p}_1, \widetilde{p}_2)$ satisfies the conditions of Theorem 2 with $\kappa_0 = \frac{1}{4}(\kappa - 1)(\widetilde{p}_1 - \widetilde{p}_2)^2$ and $\kappa_1 = \kappa(u-\ell)^2$. As argued in our previous example, one can tune ILS-d by minimizing the upper bound in Theorem 2 with respect to the policy parameters.

The MLE-cycle policy of Broder and Rusmevichientong (2012) is closely related to the ILS-d family of policies. However the plain version of MLE-cycle does not update estimates at every period, meaning that strictly speaking it is not an ILS-d policy. In their numerical experiments, Broder and Rusmevichientong define and simulate refined versions of MLE-cycle that are indeed members of the ILS-d family.

Remark Because the greedy ILS policy projects unconstrained least squares estimates $\widehat{\theta}_t$ onto Θ, the ILS variants described above use the knowledge of Θ. Replacing the projected least squares estimates ϑ_t with $\widehat{\theta}_t$ in the definition of greedy ILS, we can employ the exact arguments in the proof of Theorem 2 to obtain similar asymptotically optimal ILS variants that do not use the knowledge of Θ.

3.4 An Alternative Formulation: Incumbent-Price Problem

In this subsection we consider a modification of the learn-and-earn problem where the seller is effectively uncertain about just one demand parameter. Specifically, we assume that the seller knows the expected demand under a particular price $\widehat{p}$ to be a particular quantity $\widehat{D}$. Thus the demand parameters α and β are known to be related by the equation

$$\widehat{D} = \alpha + \beta \widehat{p}. \tag{3.7}$$

where $\widehat{D}$ and $\widehat{p}$ are given constants. Because the market response to price $\widehat{p}$ is already known, one has to choose a price different from $\widehat{p}$ to acquire further information about the demand model. In that sense, $\widehat{p}$ is the unique *uninformative price* within the feasible price domain $[\ell, u]$.

This model variant is of obvious theoretical interest, and it also approximates a situation that is common in practice, where a seller has been charging a single price $\widehat{p}$ for a long time and has observed the corresponding average demand to be $\widehat{D}$. With that situation in mind, we call $\widehat{p}$ the *incumbent price*, and $\widehat{D}$ the *incumbent demand*. Of course, a seller will never know the expected demand under an incumbent price *exactly*, but if the incumbent price has been in effect long enough, residual uncertainty about the corresponding expected demand will be negligible compared with uncertainty about the sensitivity of demand to deviations from the incumbent price.

Both Taylor (1974) and Broder and Rusmevichientong (2012) have studied dynamic pricing models with linear demand and just one unknown parameter, but those studies both considered the artificial situation where the intercept α is known exactly. That is, earlier studies with a single unknown demand parameter have assumed that the seller knows exactly the market response to a price of zero. In addition to its unreality, that scenario does not confront the seller with an interesting or difficult trade-off, because the only uninformative price is zero, and a price of zero is not attractive to a profit-maximizing seller in any case. For example, when α is known exactly there is no threat of incomplete learning under greedy ILS, because greedy prices are always strictly positive. The incumbent-price model that we consider is more realistic and much more interesting mathematically.

To avoid a borderline case where the seller must stop learning in order to price optimally, we assume that the incumbent price does not exactly equal the optimal price induced by nature's choice of θ; that is, $\widehat{p} \neq \varphi(\theta)$, although $\widehat{p}$ and $\varphi(\theta)$ can be arbitrarily close to each other. For notational simplicity, in this subsection we re-express the seller's decision in period t as a deviation from the incumbent price (rather than the new price itself), and denote it by $x_t = p_t - \widehat{p}$. As a result, the demand equation (2.1) becomes

$$D_t \; = \; \widehat{D} + \beta x_t + \epsilon_t \qquad \text{for } t = 1, 2, \ldots \tag{3.8}$$

One can take the view that the only uncertain demand parameter is β, since (3.7) defines α in terms of β. In the remainder of this subsection we will assume without loss of generality that the parameter vector θ consists only of the slope parameter β, whose value is drawn in period 0 from the compact interval $\mathfrak{B} := [b_{\min}, b_{\max}]$. Therefore we can redefine the single-period revenue function as follows:

$$r_\beta(x) := (\widehat{p} + x)(\widehat{D} + \beta x) \qquad \text{for } \beta \in \mathbb{R} \text{ and } \widehat{p} + x \in [\ell, u]. \tag{3.9}$$

The price deviation that maximizes $r_\beta(\cdot)$ can then be calculated as

$$\psi(\beta) := \arg\max\{r_\beta(x) : \widehat{p} + x \in [\ell, u]\}, \tag{3.10}$$

implying by (3.7) that $\widehat{p} + \psi(\beta) = \varphi\big((\alpha, \beta)\big)$. With these new constructs, we can update the definitions of expected T-period revenue $R^\pi_\beta(T)$ and T-period regret $\Delta^\pi_\beta(T)$ by just replacing θ, p_t, and $\varphi(\cdot)$ with β, x_t, and $\psi(\cdot)$ in equations (2.6) and (2.7). Because we assume that the optimal price and the incumbent price do not coincide, the worst-case regret becomes

$$\Delta^\pi(T) = \sup\big\{\Delta^\pi_\beta(T) : b_{\min} \le \beta \le b_{\max},\ \psi(\beta) \neq 0\big\} \quad \text{for } T = 1, 2, \ldots \tag{3.11}$$

To construct a lower bound on regret, we consider the case where $\epsilon_t \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$, and quantify the information contained in the sales data observed through the end of period t. The density of the history vector H_t in the incumbent-price setting is

$$\widehat{g}_t(H_t, \beta) = \prod_{s=1}^{t} \frac{1}{\sigma}\,\phi\left(\frac{D_s - \widehat{D} - \beta x_s}{\sigma}\right). \tag{3.12}$$

Therefore the Fisher information matrix of H_t is

$$\widehat{\mathcal{I}}^\pi_t(\beta) \;:=\; \mathbb{E}^\pi_\beta\left\{\left[\frac{\partial \log \widehat{g}_t(H_t, \beta)}{\partial \beta}\right]^2\right\} \;=\; \frac{1}{\sigma^2}\,\mathbb{E}^\pi_\beta[\widehat{J}_t], \tag{3.13}$$

where $\widehat{J}_t := \sum_{s=1}^{t} x_s^2$. Because Fisher information is a scalar rather than a matrix in this setting, we will use $\widehat{J}_t$ as the only information metric throughout this subsection. The needed analog of Lemma 1 in our incumbent-price setting is the following result.

Lemma 4 (lower bound on cumulative pricing error) *Assume that the incumbent-price relation (3.7) holds. Then, there exist finite positive constants c_0 and c_1 such that*

$$\sup_{\beta \in \mathfrak{B},\, \psi(\beta) \neq 0}\big\{\textstyle\sum_{t=2}^{T} \mathbb{E}^\pi_\beta\big(x_t - \psi(\beta)\big)^2\big\} \;\ge\; \displaystyle\sum_{t=2}^{T} \frac{c_0}{c_1 + \sup_{\beta \in \mathfrak{B}}\{\mathbb{E}^\pi_\beta[\widehat{J}_{t-1}]\}}. \tag{3.14}$$

Like Lemma 1, the preceding lower bound indicates that the growth rate of regret is determined by accumulated information. However, in contrast to the original problem, the incumbent-price problem allows the seller to accumulate information "for free" as long as it does not charge the incumbent price, which is the unique uninformative action in this setting. As a result, our information metric can increase linearly without forcing regret to increase linearly. Using this observation and plugging into (3.14), we have a logarithmically growing lower bound on regret. The following corollary of Lemma 4 spells this out mathematically.

Theorem 3 (lower bound on regret) *Assume that the incumbent-price relation (3.7) holds. Then there exists a finite positive constant c such that $\Delta^{\pi}(T) \geq c \log T$ for any pricing policy π and time horizon $T \geq 3$.*

Remark A similar logarithmic lower bound is obtained by Broder and Rusmevichientong (2012, Theorem 4.5 on p. 21) in a dynamic pricing problem with a binary-valued market response. However, their model is one in which no uninformative action exists, and hence incomplete learning is impossible. Both our result and theirs are proved using the van Trees inequality, following Goldenshluger and Zeevi (2009). The essential added feature of our problem formulation is the existence of a strictly positive uninformative price (the incumbent price), which creates a risk of incomplete learning, making the seller's problem strictly more difficult.

Theorem 3 implies that the minimum asymptotic loss rate for the incumbent-price problem is $\log T$. Accordingly, any policy π that satisfies $\Delta^{\pi}(T) = \mathcal{O}(\log T)$ is considered to be asymptotical optimal in the incumbent-price setting.

To show that the logarithmic lower bound in Theorem 3 is in fact achievable, we start as before with the greedy ILS policy. The basic principles of greedy ILS stay essentially the same under assumption (3.7). At the end of each period t, greedy ILS calculates the least squares estimator $\widehat{\beta}_t$, which can be explicitly expressed as

$$\widehat{\beta}_t = \frac{\sum_{s=1}^{t} (D_s - \widehat{D}) x_s}{\sum_{s=1}^{t} x_s^2}. \tag{3.15}$$

By the demand equation (3.8), this implies that

$$\widehat{\beta}_t - \beta = \frac{M_t}{\widehat{J}_t} \tag{3.16}$$

for all $t = 1, 2, \ldots$, where $M_t := \sum_{s=1}^{t} \epsilon_s x_s$. It is known that β lies in the interval $[b_{\min}, b_{\max}]$, so we assume that the estimates are projected into this interval. That is, the greedy ILS policy uses the truncated estimate $b_t = b_{\min} \vee (\widehat{\beta}_t \wedge b_{\max})$, where $\vee$ and $\wedge$ denote the maximum and minimum respectively. The corresponding price $\widehat{p} + \psi(b_t) = \widehat{p}/2 - \widehat{D}/(2b_t)$ is by assumption an interior point of the feasible price interval $[\ell, u]$. Thus the greedy ILS policy in the incumbent-price setting charges price $p_t = \widehat{p} + \psi\big(b_{t-1}\big)$ in period $t = 1, 2, \ldots$ To generate an initial data point, we fix the first greedy ILS price deterministically and b_0 can be found accordingly.

Having explained how greedy ILS operates in the incumbent-price setting, we now determine sufficient conditions for asymptotic optimality. The next lemma, which is an analog of Lemma 3, characterizes the relationship between the estimation errors and accumulated information.

Lemma 5 (exponential decay of estimation errors) *Assume that the incumbent-price relation (3.7) holds. Then there exists a finite positive constant ρ such that, under any pricing policy π,*

$$\mathbb{P}_{\beta}^{\pi}\big\{|\widehat{\beta}_t - \beta| > \delta,\ \widehat{J}_t \geq m\big\} \leq 2 \exp\big(- \rho(\delta \wedge \delta^2) m\big) \tag{3.17}$$

for all $\delta, m > 0$ and all $t \geq 2$.

The probability bounds in the preceding lemma and its earlier counterpart, Lemma 3, are slightly different. The bound in Lemma 3, which is derived for the two-dimensional estimator $\widehat{\theta}_t$, has a polynomial term kt in front of the exponential term, whereas in Lemma 5 the corresponding polynomial term, which simply equals 2, has a lower degree because the latter lemma is derived in a one-dimensional parameter space (see the proofs in Appendix A for further details). Apart from polynomial terms, Lemma 5 appears to be identical to Lemma 3, but there is actually a fundamental difference between them, arising from the definitions of the information metrics J_t and $\widehat{J}_t$. Once the seller's prices get close to the optimal price in the incumbent-price setting, we know that the price deviations x_t will approach the optimal price deviation $\psi(\beta)$, making the information metric $\widehat{J}_t$ grow linearly. However, when we relax the incumbent-price assumption there is no such guarantee. In our original setting, a linearly growing J_t means that the seller's price p_t eventually deviates from the average price $\overline{p}_{t-1}$ by a constant, making the seller's regret grow linearly even though the true demand parameters are estimated efficiently (i.e., with exponential bounds on probability of error). Because of this distinction, Lemma 5 leads to the following asymptotic optimality conditions.

Theorem 4 (sufficient conditions for asymptotic optimality) *Assume that the incumbent-price relation (3.7) holds, that $\beta \in [b_{\min}, b_{\max}]$, and that $\psi(\beta) \neq 0$. There exist finite positive constants κ_0, κ_1, and C such that if π is a pricing policy that satisfies*

(i) $\widehat{J}_t \geq \kappa_0 \log t$

(ii) $\sum_{s=0}^{t} \big(\psi(b_s) - x_{s+1}\big)^2 \leq \kappa_1 \log t$

almost surely for all t, then $\Delta_\beta^\pi(T) \leq C \log T$ for all $T \geq 3$.

By modifying in the obvious way the definitions of CILS and ILS-d policies (Section 3.3) readers can easily construct concrete examples of families of asymptotically optimal policies for the incumbent-price problem. For example, let us re-define CILS(κ) as the policy that charges the price $p_t = \widehat{p} + \lambda_t \delta_t + \mathrm{sgn}(\delta_t)\kappa t^{-1/2}$ in period t, where $\delta_t = \psi(b_{t-1})$ is the price deviation of greedy ILS, and $\lambda_t = \max\{0, 1 - \kappa t^{-1/2}/|\psi(b_{t-1})|\} \in [0,1]$. To demonstrate the benefit of applying conditions (i) and (ii) of Theorem 4, we display in Figure 1 the simulated performance of greedy ILS and CILS in an incumbent-price problem. As we can clearly observe, the regret of greedy ILS grows linearly over time, whereas the regret of CILS grows logarithmically, in accordance with the performance guarantee in Theorem 4.

3.5 Qualitative Insights

Information envelopes. Compared to the sufficient conditions enunciated in Theorem 2 for our original problem, those given in Theorem 4 put less emphasis on learning and more emphasis on

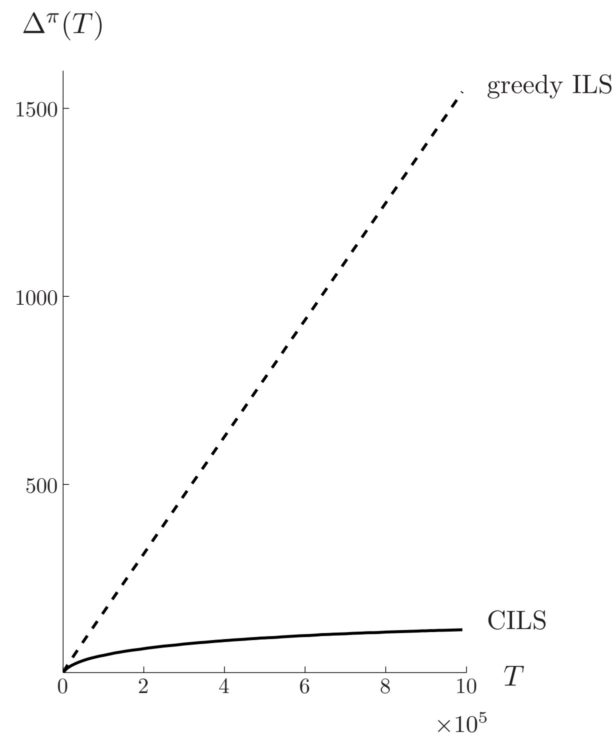

Figure 1: **Performance of greedy ILS and CILS in the incumbent-price setting.** The upper and lower curves display the T-period regret $\Delta^\pi(T)$ of greedy ILS and CILS($\kappa = 0.1$) respectively, where the problem parameters are $\alpha = 1.1$, $\beta = -0.5$, $\widehat{p} = 1$, $\widehat{D} = 0.6$, $\sigma = 0.1$, and $[\ell, u] = [0.75, 2]$. For each policy the initial price is $\widehat{p} + x_1 = u = 2$. Linearly growing regret of greedy ILS suggests that this policy does not satisfy the sufficient conditions for asymptotic optimality given in Theorem 4.

earning, and that re-balancing assures that the $\mathcal{O}(\log T)$ bound on asymptotic regret (Theorem 3) is achieved. The contrast between this $\mathcal{O}(\log T)$ bound and the $\mathcal{O}(\sqrt{T} \log T)$ bound in Theorem 2 helps us quantify the value of knowing expected demand at the incumbent price (that is, the value of knowing one point on the linear demand curve).

The first sufficient condition in both Theorem 2 and Theorem 4 describes what might be called an *information envelope*. If accumulated information, as measured by the metric J_t or $\widehat{J}_t$, is kept above the envelope, learning is "fast enough" to ensure asymptotic optimality. However, that must be accomplished without too much deviation from greedy pricing, a requirement that is expressed by the second sufficient condition in Theorems 2 and 4.

To demonstrate the practical necessity of our sufficient conditions, we simulate the performance of several pricing policies that violate those conditions. We have shown in Section 3.3 that the CILS policy with threshold parameter $\kappa = 2\lambda^{1/2}$ guarantees that J_t stays above the envelope $\lambda t^{1/2}$.

Viewing this original version of CILS as *the CILS with information envelope* $\Lambda(t) = \lambda t^{1/2}$, we can extend the definition of the CILS family to accommodate different information envelopes simply by adjusting equation (3.5). Given $0 < \gamma < 1$, if we replace the $t^{-1/4}$ terms with $t^{-\gamma/2}$ in (3.5), we obtain the CILS with information envelope $\Lambda(t) = \lambda t^{1-\gamma}$. Figure 2 shows that policies from the broader CILS family perform poorly, i.e., with loss rate exceeding $\mathcal{O}(\sqrt{T})$, if their information envelopes grow more quickly or more slowly than the asymptotically optimal information envelope $\Lambda(t) = \lambda t^{1/2}$.

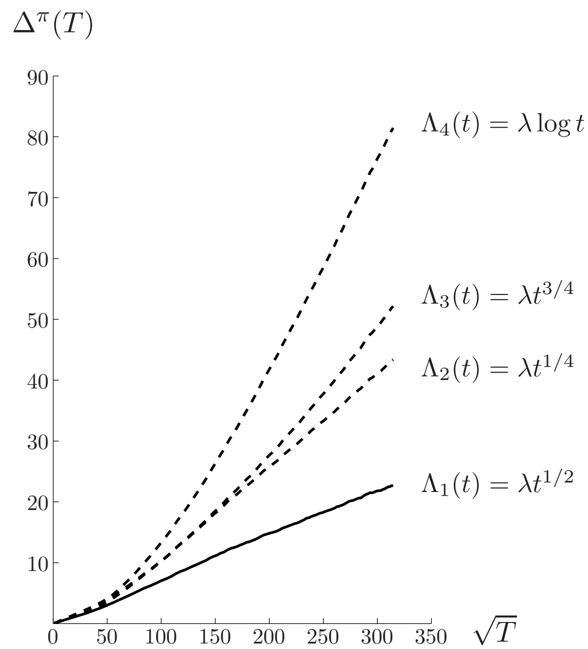

Figure 2: **Performance of CILS with different information envelopes.** Each curve displays the T-period regret $\Delta^\pi(T)$ of a CILS policy π_i with a distinct information envelope $\Lambda_i(t)$, where the problem parameters are $\alpha = 1.1$, $\beta = -0.5$, $\sigma = 0.05$, $[\ell, u] = [0.75, 2]$, and $\lambda = 4.55 \times 10^{-5}$. For each policy the first two prices are $p_1 = u = 2$, and $p_2 = \ell = 0.75$. The regret under $\Lambda_1(t) = \lambda t^{1/2}$ grows at rate $\sqrt{T}$, whereas the regret under other information envelopes grow at significantly faster rates. An ordinary least squares regression reveals that the T-period regret under $\Lambda_2(\cdot)$, $\Lambda_3(\cdot)$, and $\Lambda_4(\cdot)$ grow at rates $T^{0.64}$, $T^{0.70}$, and $T^{0.80}$, respectively.

Sensitivity to system noise. The behavior of the constants c (lower bounds) and C (upper bounds) depends on the extent of "noise" in the demand signals. To be precise, in the original setting (Theorems 1 and 2) these constants are proportional to σ, the standard deviation of the demand shocks, and in the incumbent-price setting (Theorems 3 and 4) they are both proportional to σ^2. To better understand this behavior, start off by observing that the Fisher information in both settings is inversely proportional to σ^2. At first glance, this suggests that the magnitude of the regret would be proportional to σ^2 in both cases. However, this intuition is only correct in the incumbent-price setting, because the information metric $\widehat{J}_t = \sum_{s=1}^{t}(p_s - \widehat{p})^2$ will grow linearly over time (with high probability), implying that the logarithmic information envelope constraint in Theorem 4(i) would be eventually non-binding. As a consequence, with high probability $\widehat{J}_t$ will be eventually of order $\sigma^2 t$. In contrast, in the original setting the information metric $J_t = \sum_{s=1}^{t}(p_s - \overline{p}_t)^2$ *cannot* increase linearly if the price process p_t is convergent. Consequently, if σ increases, say, the information envelope should scale up in proportion to σ so that sample paths of J_t would increase at least at the same rate as σ. In that case, the cumulative squared deviations from the greedy price would not increase faster than σ, implying by Theorem 2 that the seller's regret would scale up by order σ. (If the information envelope is inflated at a higher rate, e.g., proportional to σ^2, then the cumulative squared deviations from the greedy price increase at the same rate as the information envelope, leading to a regret of order σ^2, which is strictly worse if $\sigma > 1$.) Due to the above logic, the seller's revenue performance is less sensitive to noise in the original setting than in the incumbent-price setting.

4 Extension to Multiple Products with Substitutable Demand

In this section we extend our single-product results to the case of multiple products with substitutable demand. With a single product, the only critical issue was the *rate* of information acquisition. However, when there are two or more products the *direction* of information acquisition also becomes relevant. If all of the price deviation vectors $p_t - \overline{p}_{t-1}$ point in a particular direction, then the seller would be able to learn only a relation among the demand model parameters, rather than learning the value of every parameter. Thus a successful multi-product pricing policy needs to make sure that information is collected not only at the right rate but also evenly in all directions of the price deviation space. To collect information in such a way, we will force each set of n consecutive price deviation vectors to be an orthogonal basis for $\mathbb{R}^n$. This idea of orthogonal pricing, which will be explained in further detail below, is the most prominent feature of our approach to multi-product pricing.

4.1 Generalization of Main Ideas

To extend the analysis in Section 3.1 to multiple dimensions, we begin by deriving a lower bound on regret. Consider the case where $\epsilon_{it} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$. Given the parameter vector $\theta = (\alpha_1, \beta_1, \ldots, \alpha_n, \beta_n)$, the history H_t has the following density:

$$g_t(H_t, \theta) = \prod_{s=1}^{t} \prod_{i=1}^{n} \frac{1}{\sigma}\, \phi\Big(\frac{D_{is} - \alpha_i - \beta_i \cdot p_s}{\sigma}\Big). \tag{4.1}$$

Therefore, letting $\otimes$ denote the Kronecker product of matrices, we get by elementary algebra that the Fisher information contained in H_t is

$$\mathcal{I}_t^\pi(\theta) := \mathbb{E}_\theta^\pi\Big\{\Big[\frac{\partial \log g_t(H_t,\theta)}{\partial\theta}\Big]^\mathsf{T} \frac{\partial \log g_t(H_t,\theta)}{\partial\theta}\Big\}$$

$$= \frac{1}{\sigma^4}\,\mathbb{E}_\theta^\pi\Big\{\Big[\sum_{s=1}^t \epsilon_s \otimes \big[\begin{smallmatrix}1\\ p_s\end{smallmatrix}\big]\Big]^\mathsf{T} \sum_{s=1}^t \epsilon_s \otimes \big[\begin{smallmatrix}1\\ p_s\end{smallmatrix}\big]\Big\}$$

$$= \frac{1}{\sigma^2}\,\mathbb{E}_\theta^\pi[\mathbf{I}_n \otimes \mathcal{J}_t], \tag{4.2}$$

where the last equation follows because the ϵ_{it} are independent and $\mathbb{E}_\theta^\pi[\epsilon_{it}^2] = \sigma^2$. The following result is the analog of Lemma 1 in the multi-product setting.

Lemma 6 (lower bound on cumulative pricing error) *There exist finite positive constants c_0 and c_1 such that*

$$\sup_{\theta\in\Theta}\Big\{\sum_{t=2}^T \mathbb{E}_\theta^\pi\big\|p_t - \varphi(\theta)\big\|^2\Big\} \;\geq\; \sum_{t=2}^T \frac{c_0}{c_1 + \sup_{\theta\in\Theta}\big\{\mathsf{tr}\big(C(\theta)\,\mathbb{E}_\theta^\pi[\mathcal{J}_{t-1}]\,C(\theta)^\mathsf{T}\big)\big\}}\,, \tag{4.3}$$

where $C(\cdot)$ is an $n\times(n+1)$ matrix function on Θ such that $C(\theta) = [-\varphi(\theta)\;\; \mathbf{I}_n]$, and $\mathsf{tr}(\cdot)$ is the matrix trace operator.

Using Lemma 6, we can generalize Theorem 1 to the case of multiple products.

Theorem 5 (lower bound on regret) *There exists a finite positive constant c such that $\Delta^\pi(T) \geq c\sqrt{T}$ for any pricing policy π and time horizon $T \geq 3$.*

Remark The constant c in the preceding theorem is directly proportional to $|b_{\max}|$, where $b_{\max} < 0$ is the upper bound on the eigenvalues of the demand parameter matrix B. Note that, in the single-product case, $b_{\max}$ is simply the largest value that the slope parameter β can take. In general, the constants in the lower bounds (cf. Theorems 1, 3, 5, and 7) are proportional to $|b_{\max}|$, and the constants in the upper bounds (cf. Theorems 2, 4, 6, and 8) are proportional to $|b_{\min}|$, where $b_{\min} < 0$ is the lower bound on the eigenvalues of B. This relationship demonstrates that the seller's regret would decrease as the expected demand function becomes more "unresponsive" to changes in price.

To derive sufficient conditions of asymptotic optimality in the multivariate setting, we will first focus on policies that satisfy an orthogonal pricing condition, and then generalize the definition of our information metric. A pricing policy π is said to be an *orthogonal pricing policy* if for each $\tau = 1, 2, \ldots$ there exist $X_\tau \in \mathbb{R}^n$ and $v_\tau > 0$ such that

$$V_\tau := \big\{v_\tau^{-1}\big(p_s - \overline{p}_{s-1} - X_\tau\big) \;:\; s = n(\tau-1)+1, \ldots, n(\tau-1)+n\big\} \tag{4.4}$$

forms an orthonormal basis of $\mathbb{R}^n$, and $e \cdot X_\tau = n^{-1/2} v_\tau \|X_\tau\|$ for all $e \in V_\tau$. That is, for any given block of n periods, there is an orthonormal basis embedded within the price deviation vectors $p_s - \overline{p}_{s-1}$. Here, X_τ can be interpreted as the "nominal" price deviation in periods $n(\tau-1)+1, \ldots, n(\tau-1)+n$. In most instances, it would be sufficient to simply take X_τ equal to an n-vector of zeros, and let each price deviation vector $p_s - \overline{p}_{s-1}$ be orthogonal to the previous $n-1$ price deviation vectors, but we adopt the slightly more general definition in (4.4) to allow for a wider family of policies (see, e.g., the example at the end of this subsection). When $n = 1$, we let $X_\tau = 0$, and note that all pricing policies are orthogonal.

To characterize the rate of information acquisition for orthogonal pricing policies, we generalize the mathematical expression for sum of squared price deviations J_t, which was frequently used in Section 3. In the single-product setting where prices are scalars, the information metric we used was

$$J_t \;=\; \sum_{s=1}^{t} (p_s - \overline{p}_t)^2 \;=\; \sum_{s=2}^{t} (1 - s^{-1})(p_s - \overline{p}_{s-1})^2 \qquad \text{for } t = 1, 2, \ldots$$

Our information metric for the multi-product setting adds squared lengths of price deviation vectors over blocks of n periods. To be precise, we extend the above definition by letting

$$\begin{aligned} J_t \;&=\; n^{-1} \sum_{s=2}^{n\lfloor t/n \rfloor} (1 - s^{-1}) \big\| p_s - \overline{p}_{s-1} - X_{\lceil s/n \rceil} \big\|^2 \\ &=\; n^{-1} \sum_{\tau=1}^{\lfloor t/n \rfloor} \sum_{i=1}^{n} \Big(1 - \tfrac{1}{n(\tau-1)+i}\Big) v_\tau^2 \qquad \text{for } t = 1, 2, \ldots \end{aligned}$$

Note that the former definition of J_t is a special case of the latter when $n = 1$, in which case we assume $X_\tau = 0$ for all τ.

Using the definition of orthogonal pricing policy and our new information metric, we derive the following multivariate counterpart of Lemma 2.

Lemma 7 (minimum eigenvalue of Fisher information) *Let $\mu_{\min}(t)$ be the smallest eigenvalue of $\mathbf{I}_n \otimes \mathcal{J}_t$. Under any orthogonal pricing policy one has $\mu_{\min}(t) \geq \gamma J_t$, where $\gamma = 1/(1 + 2u - \ell)^2$.*

In the following result, we extend Lemma 3 to include orthogonal pricing policies in the multi-product setting.

Lemma 8 (exponential decay of estimation errors) *There exist finite positive constants ρ and k such that, under any orthogonal pricing policy π,*

$$\mathbb{P}_\theta^\pi \big\{ \|\widehat{\theta}_t - \theta\| > \delta,\ J_t \geq m \big\} \;\leq\; k t^{n^2+n-1} \exp\big(-\rho(\delta \wedge \delta^2) m \big) \tag{4.5}$$

for all $\delta, m > 0$ and all $t \geq 2$.

The polynomial term kt^{n^2+n-1} in the preceding probability bound reflects the effect of having a higher dimensional parameter space, thereby generalizing the polynomial term kt in Lemma 3, which is derived assuming $n = 1$. In the last result of this subsection, we generalize Theorem 2 to the case of multiple products.

Theorem 6 (sufficient conditions for asymptotic optimality) *Assume that $\theta \in \Theta$. Let κ_0, κ_1 be finite positive constants, and π be an orthogonal pricing policy satisfying*

(i) $J_t \geq \kappa_0 \sqrt{t}$

(ii) $\sum_{s=0}^{t} \left\| \varphi\big(\vartheta_{\lfloor s/n \rfloor}\big) - p_{s+1} \right\|^2 \leq \kappa_1 \sqrt{t}$

almost surely for all t. Then there exists a finite positive constant C such that $\Delta_\theta^\pi(T) \leq C\sqrt{T}\log T$ for all $T \geq 3$.

It is possible to express condition (i) of Theorem 6 in terms of the minimum eigenvalue of the empirical Fisher information matrix, $\mu_{\min}(t)$, simply by replacing J_t with $\mu_{\min}(t)$. The definition of orthogonal pricing policy and the generalization of our information metric J_t offers a practical and fairly general way of controlling $\mu_{\min}(t)$ to accumulate information at the desired rate, which is of order $\sqrt{t}$ in our setting. Naturally, one might use alternative rates for different purposes. For instance, Lai and Wei (1982) derive conditions on $\mu_{\min}(t)$ to obtain almost sure bounds on least squares errors in various settings, and in a recent study, den Boer (2013) extends some of these ideas to a dynamic pricing setting, arriving at an $\mathcal{O}(\sqrt{T}\log T)$ upper bound on the regret of a policy based on maximum quasi-likelihood estimation.

Example: Multivariate CILS. To generalize the CILS policy family to the multi-product case, let us first re-express in a more compact form the pricing rule of CILS in the single-product case, which is given in (3.5): in period t, CILS(κ) charges the price $p_t = \overline{p}_{t-1} + \lambda_t \delta_t + \operatorname{sgn}(\delta_t)\kappa t^{-1/4}$, where $\delta_t = \varphi(\vartheta_{t-1}) - \overline{p}_{t-1}$ and $\lambda_t = \max\{0, 1 - \kappa t^{-1/4}/|\delta_t|\} \in [0,1]$. The general version of CILS(κ) has a similar functional form. For each $\tau = 1, 2, \ldots$ and $i = 1, \ldots, n$, the *multivariate constrained iterated least squares* (MCILS) policy with threshold parameter κ, abbreviated MCILS$_n(\kappa)$, uses the following price vector in period $t = n(\tau - 1) + i$:

$$p_{n(\tau-1)+i} \;=\; \overline{p}_{n(\tau-1)+i-1} \;+\; \lambda_\tau \delta_\tau \;+\; \kappa \tau^{-1/4} e_{i\tau} \tag{4.6}$$

where $\delta_\tau = \varphi\big(\vartheta_{n(\tau-1)}\big) - \overline{p}_{n(\tau-1)}$, $\lambda_\tau = \max\{0, 1 - \kappa\tau^{-1/4}/\|\delta_\tau\|\} \in [0,1]$, and $\{e_{i\tau} : i = 1, \ldots, n\}$ is the orthonormal basis of $\mathbb{R}^n$ satisfying $n^{-1/2}\sum_{i=1}^{n} e_{i\tau} = \delta_\tau/\|\delta_\tau\|$. Note that, if $n = 1$, the only possible choice for e_{1t} is $\operatorname{sgn}(\delta_t)$, implying that MCILS$_1(\kappa)$ coincides with CILS(κ).

To see that any MCILS policy satisfies the conditions of Theorem 6, we first note that MCILS$_n(\kappa)$ is an orthogonal policy, because if we let $X_\tau = \lambda_\tau \delta_\tau$ and $v_\tau = \kappa\tau^{-1/4}$ for all τ the set of vectors in (4.4) satisfies the orthogonal pricing condition. Moreover, under MCILS$_n(\kappa)$, we have $J_t \geq$

$\frac{1}{2}n^{-1}\kappa^2\sum_{\tau=1}^{\lfloor t/n\rfloor}\tau^{-1/2} \geq \frac{1}{4}n^{-1}\kappa^2 t^{1/2}$, which satisfies condition (i) with $\kappa_0 = \frac{1}{4}n^{-1}\kappa^2$. Finally, for all τ and i, we have $\big\|\varphi\big(\vartheta_{n(\tau-1)}\big)-p_{n(\tau-1)+i}\big\| \leq \kappa\tau^{-1/4}+\max_i\{\overline{p}_{n(\tau-1)+i}-\overline{p}_{n(\tau-1)}\} \leq (\kappa+u-\ell)\tau^{-1/4}$; so condition (ii) is satisfied with $\kappa_1 = 2(\kappa+u-\ell)^2$. As a result, MCILS$_n(\kappa)$ achieves the performance guarantee in Theorem 6. To illustrate this fact, we display in Figure 3 the simulated performance of an MCILS policy in a particular numerical example.

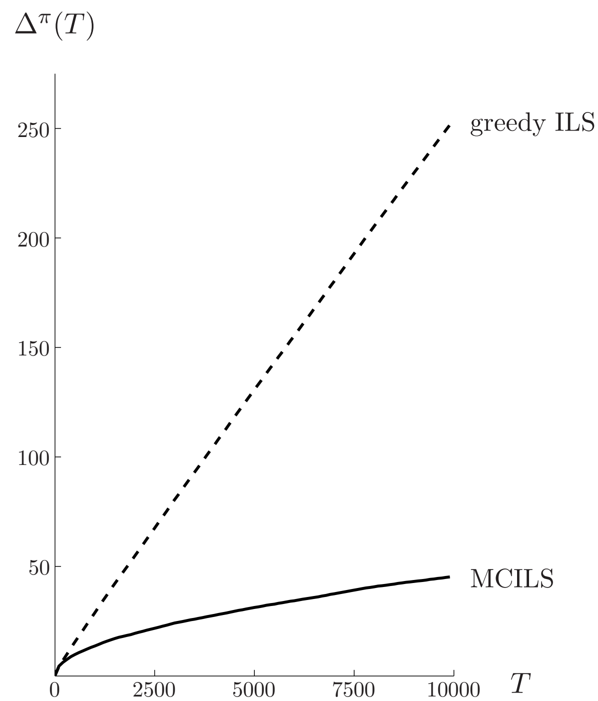

Figure 3: **Performance of greedy ILS and multivariate CILS.** The upper and lower curves depict the T-period regret $\Delta^\pi(T)$ of greedy ILS and MCILS$_2(\kappa = 0.25)$ respectively, where the problem parameters are $\alpha_1 = 1.1$, $\alpha_2 = 0.7$, $\beta_1 = (-0.5, 0.05)$, $\beta_2 = (0.05, -0.3)$, $\sigma = 0.1$, and $[\ell, u] = [0.75, 2]$. For each policy the first three price vectors are $p_1 = (\ell, \ell) = (0.75, 0.75)$, $p_2 = (u, u) = (2, 2)$, and $p_3 = (u, \ell) = (2, 0.75)$. The regret of greedy ILS grows linearly over time, whereas the regret of MCILS$_2(0.25)$ grows at rate $\sqrt{T}$.

In summary, MCILS policy family combines the orthogonal pricing condition with the rate at which CILS accumulates information. Using the same reasoning, it is straightforward to generalize the ILS-d policy family to the multi-product setting.

4.2 Incumbent-Price Problem with Multiple Products

To provide a complete picture, we now extend the incumbent-price analysis in Section 3.4 to the case of two or more products. Our formulation of the incumbent-price problem remains essentially the same, except prices are $n \times 1$ vectors rather than scalars. So assumption (3.7) is replaced with

$$\widehat{D} = a + B\widehat{p}\,, \tag{4.7}$$

where a and B are the parameters of the demand model (2.1). Assuming the seller knows that (4.7) holds with certainty, the seller's demand model becomes

$$D_t \;=\; \widehat{D} + Bx_t + \epsilon_t \qquad \text{for } t = 1, 2, \ldots \tag{4.8}$$

where $x_t = p_t - \widehat{p}$. Because the above demand model can be fully specified by learning the value of B, we treat B as the matrix containing all demand parameters. To be precise, we assume that the demand parameter vector is $b := \mathrm{vec}(B^\mathsf{T}) = (\beta_1, \ldots, \beta_n) \in \mathbb{R}^{n^2}$, and that nature selects the value of b from a compact rectangle $\mathfrak{B} \subseteq \mathbb{R}^{n^2}$ in period 0. Using the multivariate analogs of the single-period revenue function $r_\beta(\cdot)$ and the optimal price deviation function $\psi(\cdot)$, which were originally expressed in equations (3.9) and (3.10), we let $\Delta_b^\pi(T)$ denote the T-period regret under policy π and parameter vector b. Consequently the worst-case regret is $\Delta^\pi(T) := \sup\{\Delta_b^\pi(T) : \, b \in \mathfrak{B},\, \psi(b) \neq \mathbf{0}_n\}$.

As before, we consider the case where $\epsilon_{it} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$. Conditional on $b = (\beta_1, \ldots, \beta_n)$, the density of H_t in the multivariate incumbent-price setting is

$$\widehat{g}_t(H_t, b) = \prod_{s=1}^{t} \prod_{i=1}^{n} \frac{1}{\sigma}\, \phi\!\left(\frac{D_{is} - \widehat{D}_i - \beta_i \cdot x_s}{\sigma}\right). \tag{4.9}$$

Therefore, the Fisher information for H_t is

$$\widehat{\mathcal{I}}_t^\pi(b) \;:=\; \mathbb{E}_b^\pi \left\{ \left[\frac{\partial \log \widehat{g}_t(H_t, b)}{\partial b} \right]^\mathsf{T} \frac{\partial \log \widehat{g}_t(H_t, b)}{\partial b} \right\} \;=\; \frac{1}{\sigma^2}\, \mathbb{E}_b^\pi [\mathbf{I}_n \otimes \widehat{\mathcal{J}}_t]\,, \tag{4.10}$$

where $\widehat{\mathcal{J}}_t := \sum_{s=1}^t x_s x_s^\mathsf{T}$. As in the case of our previous lower bounds on pricing errors, we use the multivariate version of the van Trees inequality to get the following lemma.

Lemma 9 (lower bound on cumulative pricing error) *Assume that the incumbent-price relation (4.7) holds. Then there exist finite positive constants c_0 and c_1 such that*

$$\sup\nolimits_{b \in \mathfrak{B},\, \psi(b) \neq \mathbf{0}_n} \left\{ \sum\nolimits_{t=2}^{T} \mathbb{E}_b^\pi \big\| x_t - \psi(b) \big\|^2 \right\} \;\geq\; \sum_{t=2}^{T} \frac{c_0}{c_1 + \sup_{b \in \mathfrak{B}} \left\{ \mathbf{1}_n^\mathsf{T}\ \mathbb{E}_b^\pi[\widehat{\mathcal{J}}_{t-1}]\ \mathbf{1}_n \right\}}\,, \tag{4.11}$$

where $\mathbf{1}_n$ is an $n \times 1$ vector of ones.

By virtue of Lemma 9, we arrive at our final lower bound on regret.

Theorem 7 (lower bound on regret) *Assume that the incumbent-price relation (4.7) holds. Then there exists a finite positive constant c such that $\Delta^{\pi}(T) \geq c \log T$ for any pricing policy π and time horizon $T \geq 3$.*

To achieve the minimum asymptotic loss rate given above, we first state an orthogonal pricing condition for the multivariate incumbent-price problem: a policy π is an orthogonal pricing policy in the incumbent-price setting if for each $\tau = 1, 2, \ldots$ there exist $X_\tau \in \mathbb{R}^n$ and $v_\tau > 0$ such that $\widehat{V}_\tau := \big\{ v_\tau^{-1}\big(x_s - X_\tau\big) \,:\, s = n(\tau-1)+1, \ldots, n(\tau-1)+n \big\}$ forms an orthonormal basis of $\mathbb{R}^n$, and $e \cdot X_\tau = n^{-1/2} v_\tau \|X_\tau\|$ for all $e \in V_\tau$. We also generalize the definition of the information metric $\widehat{J}_t$ as

$$\widehat{J}_t \;=\; n^{-1} \sum_{s=1}^{n\lfloor t/n \rfloor} \big\| x_s - X_{\lceil s/n \rceil} \big\|^2 \;=\; \sum_{\tau=1}^{\lfloor t/n \rfloor} v_\tau^2 \qquad \text{for } t = 1, 2, \ldots$$

Having updated our orthogonality condition and information metric, we first show that one can control the minimum eigenvalue of the empirical Fisher information matrix.

Lemma 10 (minimum eigenvalue of Fisher information) *Assume that the incumbent-price relation (4.7) holds, and let $\widehat{\mu}_{\min}(t)$ be the smallest eigenvalue of $\mathbf{I}_n \otimes \widehat{\mathcal{J}}_t$. Then under any orthogonal pricing policy one has $\widehat{\mu}_{\min}(t) \geq \widehat{J}_t$.*

Remark In the single-product case, $\widehat{\mu}_{\min}(t)$ simply equals $\widehat{J}_t$, implying that the bound stated in Lemma 10 is tight.

Letting $\widehat{b}_t$ be the least squares estimate of b in period t, we generalize Lemma 5 as follows.

Lemma 11 (exponential decay of estimation errors) *Assume that the incumbent-price relation (4.7) holds. There exist finite positive constants ρ and k such that, under any orthogonal pricing policy π,*

$$\mathbb{P}_b^{\pi}\big\{ \|\widehat{b}_t - b\| > \delta,\; \widehat{J}_t \geq m \big\} \;\leq\; k t^{n^2-1} \exp\big(-\rho(\delta \wedge \delta^2) m \big) \tag{4.12}$$

for all $\delta, m > 0$ and all $t \geq 2$.

Finally, denoting by b_t the projection of $\widehat{b}_t$ onto $\mathfrak{B}$, we use the preceding lemma to generate the last set of sufficient conditions for asymptotic optimality.

Theorem 8 (sufficient conditions for asymptotic optimality) *Assume that the incumbent-price relation (4.7) holds, that $b \in \mathfrak{B}$, and that $\psi(b) \neq \mathbf{0}_n$. There exist finite positive constants κ_0, κ_1, and C such that if π is an orthogonal pricing policy satisfying*

(i) $\widehat{J}_t \geq \kappa_0 \log t$

(ii) $\sum_{s=0}^{t} \big\| \psi\big(b_{\lfloor s/n \rfloor}\big) - x_{s+1} \big\|^2 \leq \kappa_1 \log t$

almost surely for all t, then $\Delta_b^{\pi}(T) \leq C \log T$ for all $T \geq 3$.

5 Concluding Remarks

On the linear demand assumption. Linear demand models are commonly used in economic theory, and also in revenue management practice, but they are obviously restrictive. To justify our assumption of linear demand in the learning-and-earning problem, one possible argument is that (a) any demand function can be closely approximated by a linear function if the range of permissible prices is sufficiently narrow, and (b) narrow price ranges are often enforced in dynamic pricing contexts, because the seller, not wishing to provoke a competitive response, considers only "tactical" deviations from an incumbent price.

Alternatively, one may argue that essentially the same analysis applies with a generalized linear model (GLM) $q = F(L(p))$, where q is the vector of expected demands in response to price vector p, L is a linear function, and F is a general "link function." If the link function is known but the coefficients of L are not known, then the problem of dynamic pricing under the GLM essentially maps to the problem treated here: one can make use of generalized least squares estimation in place of the simple least squares estimates used in the present paper. den Boer and Zwart (2013) adopt this framework in their study of constrained variance pricing. The GLM framework adds a substantial layer of technical detail, so we have focused on the simpler class of linear models, striving to achieve a more transparent analysis. One can consider extensions to more general parametric families, using the method of maximum likelihood to make inferences about model parameters. Because maximum likelihood estimators ultimately are expressed as averages of suitable score functions, one can use martingale methods and associated large deviation inequalities in that context, as we have done in the linear context, but the analysis is more technical and cumbersome; see Besbes and Zeevi (2009) for an indication of the tools involved. Finally, recent work of Besbes and Zeevi (2013) indicates that in certain circumstances a misspecification that stems from assuming linear demand, while the true demand model has different functional form, may not be as detrimental as one would expect.

Discussion of main contributions. The main contributions of this paper are (a) the formulation of the incumbent-price problem, which applies to a substantial number of cases in pricing practice, (b) the generality of the price experimentation solutions offered, (c) the development of a unifying theme to analyze dynamic pricing problems with demand model uncertainty, and (d) the extension of single-product pricing analysis to the case of multiple products with substitutable demand.

With regard to (a), the main insight we derive is the sharp contrast between the results obtained in the incumbent-price setting, and the original setting. All of the regret bounds we derive in the incumbent-price setting are of order $\log T$, whereas the regret bounds in the original setting are of order $\sqrt{T}$ or higher. This difference shows that knowing average demand under the incumbent price has significant value, and our asymptotic optimality conditions in Theorems 4 and 8 provide practical guidelines on how best to use this incumbent-price information.

With regard to (b), a prominent feature of our analysis is that we do not focus on a particular parametric family of pricing policies. Instead, we derive general sufficient conditions for asymptotic optimality (cf. Theorems 2, 4, 6, and 8), and show by example that these conditions are necessary in practice. In Section 3.3, we provide a few ideas about how to construct pricing policies that satisfy the asymptotic optimality conditions, but the general nature of our results allows practitioners to design price experimentation policies that best fit their particular business setting.

With regard to (c), we employ a systematic mode of analysis for eliciting lower and upper bounds on the performance of pricing policies. The lower bounds are obtained via the van Trees inequality, which is a generalized version of the Cramér-Rao bound, and our upper bounds are obtained via an exponential martingale argument. We predict that this unifying approach will find more use in the broader context of revenue management and pricing, particularly in the analysis of problems that involve model uncertainty.

The results in this paper introduce asymptotic optimality to multi-product pricing with demand model uncertainty. Our analysis shows that learning and earning in higher dimensions is not a straightforward repetition of learning and earning in one dimension. To be precise, when there is more than one pricing decision, information should be collected by spanning the entire action space so that the seller would learn the values of all demand parameters, instead of the value of a linear combination of them. The orthogonal pricing condition, which is described in detail in Section 4, is an intuitive and easily implementable way to ensure that information is accumulated evenly in all directions. We believe that the extension to multiple products with substitutable demand is an important feature that distinguishes our work from other recent studies in the dynamic pricing literature.

Appendix A: Proofs of the Results in Section 3

A.1 A Lower Bound on Regret

Proof of Lemma 1. Let λ be an absolutely continuous density on Θ, taking positive values on the interior of Θ and zero on its boundary (cf. Goldenshluger and Zeevi 2009, p. 1632 for a standard choice of λ). Then, the multivariate van Trees inequality (cf. Gill and Levit 1995) implies that

$$\mathbb{E}_\lambda\big\{\mathbb{E}_\theta^\pi\big(p_t - \varphi(\theta)\big)^2\big\} \geq \frac{\Big(\mathbb{E}_\lambda\big[\,C(\theta)\,(\partial\varphi/\partial\theta)^\mathsf{T}\,\big]\Big)^2}{\mathbb{E}_\lambda\big[\,C(\theta)\,\mathcal{I}_{t-1}^\pi(\theta)\,C(\theta)^\mathsf{T}\,\big] + \widetilde{\mathcal{I}}(\lambda)}, \tag{A.1}$$

where $\widetilde{\mathcal{I}}(\lambda)$ is the Fisher information for the density λ, and $\mathbb{E}_\lambda(\cdot)$ is the expectation operator with respect to λ. Note that for all $\theta = (\alpha, \beta) \in \Theta$ we have $C(\theta)\,(\partial\varphi/\partial\theta)^\mathsf{T} = -\varphi(\theta)/(2\beta)$, and $\mathcal{I}_{t-1}^\pi(\theta) = \sigma^{-2}\,\mathbb{E}_\theta^\pi\,[\mathcal{J}_{t-1}]$. Using these identities and adding up inequality (A.1) over $t = 2, \ldots, T$, we obtain

$$\sum_{t=2}^{T} \mathbb{E}_\lambda\big\{\mathbb{E}_\theta^\pi\big(p_t - \varphi(\theta)\big)^2\big\} \geq \sum_{t=2}^{T} \frac{\big(\mathbb{E}_\lambda[\varphi(\theta)/(2\beta)]\big)^2}{\sigma^{-2}\mathbb{E}_\lambda\big[\, C(\theta)\, \mathbb{E}_\theta^\pi\, [\mathcal{J}_{t-1}]\ C(\theta)^\mathsf{T}\,\big] + \widetilde{\mathcal{I}}(\lambda)}.$$

Because $\mathbb{E}_\lambda(\cdot)$ is a monotone operator,

$$\sup_{\theta\in\Theta}\big\{\textstyle\sum_{t=2}^{T} \mathbb{E}_\theta^\pi\big(p_t - \varphi(\theta)\big)^2\big\} \geq \displaystyle\sum_{t=2}^{T} \frac{\inf_{\theta\in\Theta}\big\{\big(\varphi(\theta)/(2\beta)\big)^2\big\}}{\sigma^{-2}\sup_{\theta\in\Theta}\big\{C(\theta)\,\mathbb{E}_\theta^\pi\,[\mathcal{J}_{t-1}]\ C(\theta)^\mathsf{T}\big\} + \widetilde{\mathcal{I}}(\lambda)}.$$

In the preceding inequality, note that $\inf_{\theta\in\Theta}\big\{\big(\varphi(\theta)/(2\beta)\big)^2\big\} \geq \ell^2/(4b_{\min}^2)$ because $0 < \ell \leq \varphi(\theta)$, and $b_{\min} \leq \beta < 0$. So we let $c_0 = \sigma^2\ell^2/(4b_{\min}^2)$, $c_1 = \sigma^2\widetilde{\mathcal{I}}(\lambda)$, and arrive at the desired result. ■

Proof of Theorem 1. Due to the fact that $C(\theta) = [-\varphi(\theta)\ 1]$, we have $C(\theta)\mathbb{E}_\theta^\pi\,[\mathcal{J}_{t-1}]\, C(\theta)^\mathsf{T} = \sum_{s=1}^{t-1} \mathbb{E}_\theta^\pi\big(p_s - \varphi(\theta)\big)^2$. Thus, inequality (3.3) of Lemma 1 is equivalent to the following:

$$\sup_{\theta\in\Theta}\big\{\textstyle\sum_{t=2}^{T} \mathbb{E}_\theta^\pi\big(p_t - \varphi(\theta)\big)^2\big\}\ \geq\ \displaystyle\sum_{t=2}^{T} \frac{c_0}{c_1 + \sup_{\theta\in\Theta}\big\{\sum_{s=1}^{t-1} \mathbb{E}_\theta^\pi\big(p_s - \varphi(\theta)\big)^2\big\}}\,. \tag{A.2}$$

Recalling the definition of the worst-case regret (2.8), we also get

$$\begin{aligned}
\Delta^\pi(T) &= \sup_{\theta\in\Theta}\Big\{\sum_{t=1}^{T} \mathbb{E}_\theta^\pi\big[r_\theta^* - r_\theta(p_t)\big]\Big\} \\
&= \sup_{\theta\in\Theta}\Big\{\sum_{t=1}^{T} \mathbb{E}_\theta^\pi\big[\varphi(\theta)\big(\alpha + \beta\varphi(\theta)\big) - p_t(\alpha + \beta p_t)\big]\Big\} \\
&= \sup_{\theta\in\Theta}\Big\{-\beta\sum_{t=1}^{T} \mathbb{E}_\theta^\pi\big(p_t - \varphi(\theta)\big)^2\Big\} \\
&\geq |b_{\max}|\,\sup_{\theta\in\Theta}\Big\{\sum_{t=2}^{T} \mathbb{E}_\theta^\pi\big(p_t - \varphi(\theta)\big)^2\Big\},
\end{aligned}$$

because β is contained in $[b_{\min},\, b_{\max}]$. Therefore, inequality (A.2) implies that

$$\Delta^\pi(T)\ \geq\ b_{\max}^2 \sum_{t=2}^{T} \frac{c_0}{|b_{\max}|c_1 + \Delta^\pi(t-1)}. \tag{A.3}$$

Let $K_1 = b_{\max}^2 c_0$, and $K_2 = |b_{\max}|c_1$. Because $\Delta^\pi(t)$ is non-decreasing in t, the preceding inequality implies that

$$\Delta^\pi(T)\ \geq\ \frac{K_1(T-1)}{K_2 + \Delta^\pi(T)}\ \geq\ \frac{K_1 T}{2K_3\Delta^\pi(T)}\,,$$

for all $T \geq 2$, where $K_3 := K_2/\Delta^\pi(1) + 1 \leq 4K_2/(u-\ell)^2 + 1$. Thus, $\Delta^\pi(T) \geq c\sqrt{T}$, where $c = \sqrt{K_1/(2K_3)}$. ■

As mentioned earlier, the above proof holds for a broader exponential family of distributions whose densities have the form $f_\epsilon(\xi\,|\,\sigma) = c(\sigma)\exp\big(h(\xi) + \sum_{j=1}^{k} w_j(\sigma)t_j(\xi)\big)$ for $\xi \in \mathbb{R}$, where $h(\cdot)$

and $t_j(\cdot)$ are differentiable functions. To see this generalization, first note that the proofs of Lemma 1 and its subsequent counterparts (Lemmas 4, 6, and 9) rely on Fisher information matrices that are computed by using the density of Gaussian distribution with zero mean and σ^2 variance. As a matter of fact, we can obtain an almost identical Fisher information matrix as long as noise terms have a density of "similar shape." To be precise, if we assume that the distribution of the ϵ_t belongs to the exponential family mentioned above, the density of history vector H_t would be

$$g_t(H_t, \theta) = \prod_{s=1}^{t} f_\epsilon(D_s - \alpha - \beta p_s \,|\, \sigma),$$

which implies that the Fisher information matrix of H_t is

$$\mathcal{I}_t^\pi(\theta) \;:=\; \mathbb{E}_\theta^\pi \left\{ \left[\frac{\partial \log g_t(H_t, \theta)}{\partial \theta} \right]^\mathsf{T} \frac{\partial \log g_t(H_t, \theta)}{\partial \theta} \right\} \;=\; k(\sigma)\, \mathbb{E}_\theta^\pi \left[\mathcal{J}_t \right],$$

where $k(\sigma) = \mathbb{E}\big[(h'(\epsilon_t) + \sum_{j=1}^k w_j(\sigma) t_j'(\epsilon_t))^2\big]$. Replacing equation (3.2) with the preceding identity, we carry out the rest of our argument exactly the same way.

A.2 Sufficient Conditions for Asymptotic Optimality

Proof of Lemma 2. Take an arbitrary unit vector $y = (y_1, y_2) \in \mathbb{R}^2$. Note that

$$y^\mathsf{T} \mathcal{J}_t y \;=\; \sum_{s=1}^{t} (y_1 + y_2 p_s)^2 \;=\; \sum_{s=1}^{t} (y_1 + y_2 \overline{p}_t)^2 + y_2^2 \sum_{s=1}^{t} (p_s - \overline{p}_t)^2.$$

The first term on the right side above equals $(y_1 + y_2 \overline{p}_t)^2 t$. Moreover, the second term can be expressed as $y_2^2 \sum_{s=1}^{t} (p_s - \overline{p}_t)^2 = y_2^2 J_t$. Because $J_t \le (u - \ell)^2 t$, we get by elementary algebra that $y^\mathsf{T} \mathcal{J}_t y \ge \gamma J_t$, where $\gamma = 2/(1 + 2u - \ell)^2$. Recalling that the unit vector y was selected arbitrarily, we conclude by the Rayleigh-Ritz theorem that $\mu_{\min}(t) \ge \gamma J_t$. ■

Proof of Lemma 3. We will complete the proof in four steps.

Step 1: Derive an upper bound on the exponential moments of the error terms. Because the ϵ_t have a light-tailed distribution with mean zero and variance σ^2, there exists $z_0 > 0$ such that

$$\mathbb{E}_\theta^\pi[\exp(z\epsilon_t)] \;=\; 1 + z\, \mathbb{E}_\theta^\pi[\epsilon_t] + \sum_{k=2}^{\infty} \frac{z^k \mathbb{E}_\theta^\pi[\epsilon_t^k]}{k!} \;=\; 1 + \frac{1}{2}\nu(z)\sigma^2 z^2 \quad \text{for all } z \text{ satisfying } |z| \le z_0, \tag{A.4}$$

where $\nu(z) := \sum_{k=2}^{\infty} \big(2 z^{k-2} \mathbb{E}_\theta^\pi[\epsilon_t^k]\big)/(k!\sigma^2) < \infty$. Thus,

$$\begin{aligned} \mathbb{E}_\theta^\pi[\exp(z\epsilon_t)] \;&\le\; \exp\left(\frac{1}{2}\nu(z)\sigma^2 z^2\right) \\ &\le\; \exp\left(\frac{1}{2}\nu^* \sigma^2 z^2\right) \quad \text{for all } z \text{ satisfying } |z| \le z_0\,, \end{aligned} \tag{A.5}$$

where $\nu^* = \max_{|z|\le z_0}\{\nu(z)\}$. Note that, if the ϵ_t have a Gaussian distribution, then the preceding inequalities hold with equality and $\nu(z) = \nu^* = 1$ for all $z \in \mathbb{R}$.

Step 2: Define a family of supermartingales $\{Z^y_s\}$. We know by the least squares equation (2.10) that $\widehat{\theta}_t - \theta = \mathcal{J}_t^{-1}\mathcal{M}_t$. Let $y = (y_1, y_2)$ be a 2×1 vector with $\|y\| = \delta$, and define the process $\{Z^y_s,\ s = 1, 2, \ldots\}$ such that $Z^y_0 = 1$ and

$$Z^y_s := \exp\Big\{\frac{1}{\zeta^*\sigma^2}\Big(y\cdot\mathcal{M}_s - \frac{1}{2}y^\mathsf{T}\mathcal{J}_s y\Big)\Big\} \quad \text{for all } s = 1, 2, \ldots, \tag{A.6}$$

where $\zeta^* = (1\vee\delta)\big(\nu^*\vee(z^*/z_0)\big)$, and $z^* = \max_{\|y\|\le 1,\, p\in[\ell,u]}\{|y_1+y_2 p|/\sigma^2\}$. Note that Z^y_s is integrable for all s. Letting $\mathcal{F}_s := \sigma(D_1,\ldots,D_s)$, we get by the tower property that

$$\mathbb{E}^\pi_\theta[Z^y_s|\mathcal{F}_{s-1}] \;=\; \exp\Big\{\frac{1}{\zeta^*\sigma^2}\Big(y\cdot\mathcal{M}_{s-1} - \frac{1}{2}y^\mathsf{T}\mathcal{J}_s y\Big)\Big\}\; \mathbb{E}^\pi_\theta\Big[\exp\Big\{\frac{1}{\zeta^*\sigma^2}\,y\cdot(\mathcal{M}_s - \mathcal{M}_{s-1})\Big\}\Big|\mathcal{F}_{s-1}\Big].$$

Because $\mathcal{M}_s - \mathcal{M}_{s-1} = (\epsilon_s, \epsilon_s p_s)$ and $|y_1 + y_2 p_s|/(\zeta^*\sigma^2) \le z_0|y_1 + y_2 p_s|/(\delta z^*\sigma^2) \le z_0$ for all $p_s \in [\ell, u]$, inequality (A.5) implies that the conditional expectation on the right side immediately above satisfies

$$\mathbb{E}^\pi_\theta\Big[\exp\Big\{\frac{1}{\zeta^*\sigma^2}\,y\cdot(\mathcal{M}_s - \mathcal{M}_{s-1})\Big\}\;\Big|\;\mathcal{F}_{s-1}\Big] \;\le\; \exp\Big\{\frac{1}{2\zeta^*\sigma^2}\,y^\mathsf{T}\Big[\begin{smallmatrix}1 & p_s\\ p_s & p_s^2\end{smallmatrix}\Big]y\Big\}.$$

Combining the preceding identity and inequality, we deduce that

$$\mathbb{E}^\pi_\theta[Z^y_s|\mathcal{F}_{s-1}] \;\le\; \exp\Big\{\frac{1}{\zeta^*\sigma^2}\Big(y\cdot\mathcal{M}_{s-1} - \frac{1}{2}y^\mathsf{T}\mathcal{J}_{s-1}y\Big)\Big\} \;=\; Z^y_{s-1}.$$

So $(Z^y_s, \mathcal{F}_s)$ is a supermartingale for any given $y \in \mathbb{R}^2$ with $\|y\| = \delta$.

Step 3: Construct a stochastic process $\{\widetilde{Z}_s\}$ that characterizes the probability bound. Let $w_s = \delta\mathcal{J}_s^{-1}\mathcal{M}_s/\|\mathcal{J}_s^{-1}\mathcal{M}_s\|$, and define the stochastic process $\{\widetilde{Z}_s, s = 1, 2, \ldots\}$ such that $\widetilde{Z}_s = Z^{w_s}_s$ for all s. Fix $m > 0$, let ξ be an arbitrary positive real number, and $A := \{\|\mathcal{M}_t\| \le \xi t\} \in \mathcal{F}_t$. Then, we deduce the following by the union bound:

$$\begin{aligned}\mathbb{P}^\pi_\theta\big(\|\widehat{\theta}_t - \theta\| > \delta,\ J_t \ge m\big) \;&=\; \mathbb{P}^\pi_\theta\big(\|\mathcal{J}_t^{-1}\mathcal{M}_t\| > \delta,\ J_t \ge m\big)\\ &\le\; \mathbb{P}^\pi_\theta\big(\|\mathcal{J}_t^{-1}\mathcal{M}_t\| > \delta,\ J_t \ge m,\ A\big) \;+\; \mathbb{P}^\pi_\theta\big(J_t \ge m,\ A^c\big).\end{aligned} \tag{A.7}$$

To derive an upper bound the first probability on the right hand side above, note that

$$\mathbb{P}^\pi_\theta\big(\|\mathcal{J}_t^{-1}\mathcal{M}_t\| > \delta,\ J_t \ge m,\ A\big) \;\le\; \mathbb{P}^\pi_\theta\big(w_t\cdot\mathcal{M}_t \ge w_t^\mathsf{T}\mathcal{J}_t w_t,\ J_t \ge m,\ A\big) \tag{A.8}$$

because $\|\mathcal{J}_t^{-1}\mathcal{M}_t\| > \delta$ implies $w_t\cdot\mathcal{M}_t \ge w_t^\mathsf{T}\mathcal{J}_t w_t$ for the above choice of w_t. Combined with the definition of the process $\{\widetilde{Z}_s\}$, this implies that

$$\mathbb{P}^\pi_\theta\big(\|\mathcal{J}_t^{-1}\mathcal{M}_t\| > \delta,\ J_t \ge m,\ A\big) \;\le\; \mathbb{P}^\pi_\theta\Big(\widetilde{Z}_t \ge \exp\Big\{\frac{w_t^\mathsf{T}\mathcal{J}_t w_t}{2\zeta^*\sigma^2}\Big\},\ J_t \ge m,\ A\Big).$$

Because $\mu_{\min}(t)$ is the smallest eigenvalue of $\mathcal{J}_t$, and $\|w_t\| = \delta$, we deduce by the Rayleigh-Ritz theorem that $w_t^{\mathsf{T}} \mathcal{J}_t w_t \geq \delta^2 \mu_{\min}(t)$, and in conjunction with Lemma 2 we get

$$\mathbb{P}_\theta^\pi\big(\|\mathcal{J}_t^{-1}\mathcal{M}_t\| > \delta,\ J_t \geq m,\ A\big) \ \leq\ \mathbb{P}_\theta^\pi\Big(\widetilde{Z}_t \geq \exp\Big\{\frac{\gamma\,\delta^2 m}{2\zeta^*\sigma^2}\Big\},\ A\Big).$$

Step 4: Prove $\widetilde{Z}_t$ is bounded above by a supermartingale with very high probability. Noting that the function $L(y) := y\cdot\mathcal{M}_t - \frac{1}{2}y^{\mathsf{T}}\mathcal{J}_t y$ is concave, we deduce that $L(w) - L(y) \leq (w-y)\cdot\nabla L(y) = (w-y)\cdot(\mathcal{M}_t - \mathcal{J}_t y) \leq \|w-y\|\,\|\mathcal{M}_t\| + (1+u^2)\|w-y\|^2 t$. Thus, on the event $A = \{\|\mathcal{M}_t\| \leq \xi t\}$, we have $Z_t^y \geq e^{-(\xi+1+u^2)/(\zeta^*\sigma^2)}\widetilde{Z}_t$ for all y within the $(1/t)$-neighborhood of w_t. Let $\mathcal{S} = \{\widetilde{y}_1, \widetilde{y}_2, \ldots, \widetilde{y}_S\}$ be a set of $S = \lceil \pi\delta t\rceil$ distinct points that are evenly spaced on the circle $\{y \in \mathbb{R}^2 : \|y\| = \delta\}$. Therefore, at time $s = t$, w_s has to be in the $(1/t)$-neighborhood of some $y \in \mathcal{S}$, implying $Z_t^y \geq e^{-(\xi+1+u^2)/(\zeta^*\sigma^2)}\widetilde{Z}_t$ for some $y \in \mathcal{S}$. Thus, we have

$$\begin{aligned}\mathbb{P}_\theta^\pi\big(\|\mathcal{J}_t^{-1}\mathcal{M}_t\| > \delta,\ J_t \geq m,\ A\big) &\leq \mathbb{P}_\theta^\pi\Big(\widetilde{Z}_t \geq \exp\Big\{\frac{\gamma\,\delta^2 m}{2\zeta^*\sigma^2}\Big\},\ A\Big)\\ &\leq \mathbb{P}_\theta^\pi\big(Z_t^y \geq K_1\exp(\rho_0\delta^2 m) \text{ for some } y \in \mathcal{S},\ A\big)\end{aligned}$$

where $K_1 = e^{-(\xi+1+u^2)/(\zeta^*\sigma^2)}$ and $\rho_0 = \gamma/(2\zeta^*\sigma^2)$. Then, by the union bound, we have

$$\mathbb{P}_\theta^\pi\big(\|\mathcal{J}_t^{-1}\mathcal{M}_t\| > \delta,\ J_t \geq m,\ A\big) \ \leq\ \sum_{\widetilde{y}_j\in\mathcal{S}} \mathbb{P}_\theta^\pi\big(Z_t^{\widetilde{y}_j} \geq K_1\exp(\rho_0\delta^2 m),\ A\big).$$

By Markov's inequality and the fact that $Z_t^{\widetilde{y}_j}$ is a supermartingale with $Z_0^{\widetilde{y}_j} = 1$ for all $\widetilde{y}_j \in \mathcal{S}$, each term inside the preceding sum satisfies $\mathbb{P}_\theta^\pi\big(Z_t^{\widetilde{y}_j} \geq K_1\exp(\rho_0\delta^2 m),\ A\big) \leq \frac{1}{K_1}\exp(-\rho_0\delta^2 m)$. Because there are $S = \lceil\pi\delta t\rceil$ points in $\mathcal{S}$, the preceding inequality implies the following upper bound on the first term on the right hand side of (A.7):

$$\mathbb{P}_\theta^\pi\big(\|\mathcal{J}_t^{-1}\mathcal{M}_t\| > \delta,\ J_t \geq m,\ A\big) \ \leq\ \tfrac{1+K_2 t}{K_1}\ \exp(-\rho_0\delta^2 m)\,,$$

where $K_2 = (1\vee\delta)\pi$. Using a simplified version of the above exponential supermartingale argument (see, for instance, the proof of Lemma 5) and the fact that $J_t \leq (u-\ell)^2 t$ for all t, it is straightforward to show that

$$\begin{aligned}\mathbb{P}_\theta^\pi\big(J_t \geq m,\ A^c\big) &= \mathbb{P}_\theta^\pi\big(J_t \geq m,\ \|\mathcal{M}_t\| > \xi t\big)\\ &\leq \mathbb{P}_\theta^\pi\Bigg(J_t \geq m,\ \Big|\sum_{s=1}^t \epsilon_s\Big| > \frac{\xi t}{\sqrt{2}}\Bigg) \ +\ \mathbb{P}_\theta^\pi\Bigg(J_t \geq m,\ \Big|\sum_{s=1}^t \epsilon_s p_s\Big| > \frac{\xi t}{\sqrt{2}}\Bigg)\\ &\leq 2\exp\Bigg(-\frac{\xi^2 m}{4\zeta^*\sigma^2(u-\ell)^2}\Bigg) \ +\ 2\exp\Bigg(-\frac{\xi^2 m}{4\zeta^*\sigma^2(u-\ell)^2u^2}\Bigg).\end{aligned} \tag{A.9}$$

Choosing $\xi = 2(1\vee\delta)\sigma(u-\ell)\big(\zeta^*(1+u^2)\rho_0\big)^{1/2} \geq 2\delta\sigma(u-\ell)\big(\zeta^*(1+u^2)\rho_0\big)^{1/2}$, we have $\mathbb{P}_\theta^\pi\big(J_t \geq m,\ A^c\big) \leq 4\exp(-\rho_0\delta^2 m)$. Having found upper bounds on both terms on the right hand side of

(A.7), we let $\rho = (1 \vee \delta)\rho_0 = \frac{1}{2}\gamma\sigma^{-2}\big(\nu^* \vee (z^*/z_0)\big)^{-1}$ and $k = 4 + (1+K_2)/K_1$ to conclude the proof. ■

Proof of Theorem 2. First, note that the revenue loss of a policy comes from either least squares estimation errors, or perturbations from the greedy ILS price. More formally, conditioned on the parameter vector θ and policy π, we have the following upper bound on the expected value of squared pricing error in period $t+1$:

$$\mathbb{E}^\pi_\theta \big(\varphi(\theta) - p_{t+1}\big)^2 \le 2\,\mathbb{E}^\pi_\theta \big(\varphi(\theta) - \varphi(\vartheta_t)\big)^2 + 2\,\mathbb{E}^\pi_\theta \big(\varphi(\vartheta_t) - p_{t+1}\big)^2\,, \tag{A.10}$$

where ϑ_t is the truncated least squares estimate calculated at the end of period t. Consider the first term on the right side of inequality (A.10). By the mean value theorem, we have $|\varphi(\theta) - \varphi(\vartheta_t)| \le \sqrt{2K_0}\,\|\theta - \vartheta_t\|$ where $K_0 = \max_{j=1,2}\big\{\max_\theta\{(\partial\varphi(\theta)/\partial\theta_j)^2\}\big\}$. Therefore, by monotonicity of expectation we have

$$\begin{aligned}
\mathbb{E}^\pi_\theta \big(\varphi(\theta) - \varphi(\vartheta_t)\big)^2 &\le 2K_0\,\mathbb{E}^\pi_\theta \big\|\theta - \vartheta_t\big\|^2 \\
&\overset{(a)}{\le} 2K_0\,\mathbb{E}^\pi_\theta \big\|\theta - \widehat{\theta}_t\big\|^2 \\
&\overset{(b)}{=} 2K_0 \int_0^\infty \mathbb{P}^\pi_\theta \Big(\big\|\theta - \widehat{\theta}_t\big\|^2 > x,\ J_t \ge \kappa_0\sqrt{t}\Big)\,dx \\
&\le \frac{4K_0 \log t}{\rho\kappa_0\sqrt{t}} + 2K_0 \int_{\frac{2\log t}{\rho\kappa_0\sqrt{t}}}^\infty \mathbb{P}^\pi_\theta \Big(\big\|\theta - \widehat{\theta}_t\big\|^2 > x,\ J_t \ge \kappa_0\sqrt{t}\Big)\,dx \\
&\overset{(c)}{\le} \frac{4K_0 \log t}{\rho\kappa_0\sqrt{t}} + 2K_0 \int_{\frac{2\log t}{\rho\kappa_0\sqrt{t}}}^\infty kt\exp\big(-\rho(\sqrt{x}\wedge x)\kappa_0\sqrt{t}\big)\,dx \\
&= \frac{4K_0 \log t}{\rho\kappa_0\sqrt{t}} + 2K_0 \int_{\frac{2\log t}{\rho\kappa_0\sqrt{t}}}^1 k\exp\Big(-\frac{1}{2}\rho x\kappa_0\sqrt{t}\Big)\,dx + 2K_0 \int_1^\infty k\exp\Big(-\frac{1}{2}\rho\sqrt{x}\kappa_0\sqrt{t}\Big)\,dx \\
&\le \frac{4K_0}{\rho\kappa_0\sqrt{t}}\Big(\log t + \frac{k}{t} + \frac{4k}{t}\Big) \le \frac{4K_0(1+5k)\log t}{\rho\kappa_0\sqrt{t}}
\end{aligned}$$

for all $t \ge N$, where: (a) follows since $\|\theta - \vartheta_t\| \le \|\theta - \widehat{\theta}_t\|$, (b) follows by condition (i), (c) follows by Lemma 3, and N is the smallest natural number such that $kN\exp(-\frac{1}{2}\rho\kappa_0\sqrt{N}) \le 1$. Thus, inequality (A.10) becomes

$$\mathbb{E}^\pi_\theta \big(\varphi(\theta) - p_{t+1}\big)^2 \le 4\widetilde{K}_0 \log t/(\rho\kappa_0\sqrt{t}) + 2\,\mathbb{E}^\pi_\theta \big(\varphi(\vartheta_t) - p_{t+1}\big)^2 \quad \text{for all } t \ge N, \tag{A.11}$$

where $\widetilde{K}_0 = (1+5k)K_0$. Summing over $t = N, \ldots, T-1$ we obtain

$$\sum_{t=N}^{T-1} \mathbb{E}^\pi_\theta \big(\varphi(\theta) - p_{t+1}\big)^2 \le \frac{8\widetilde{K}_0}{\rho\kappa_0}\sqrt{T}\log T + 2\sum_{t=N}^{T-1} \mathbb{E}^\pi_\theta \big(\varphi(\vartheta_t) - p_{t+1}\big)^2\,.$$

By condition (ii), we have an upper bound also on the second term on the right side, implying that

$$\sum_{t=N}^{T-1} \mathbb{E}_\theta^\pi \big(\varphi(\theta) - p_{t+1}\big)^2 \;\leq\; \frac{8\widetilde{K}_0}{\rho\kappa_0}\sqrt{T}\log T + 2\kappa_1\sqrt{T} \;\leq\; \left(\frac{8\widetilde{K}_0}{\rho\kappa_0} + 2\kappa_1\right)\sqrt{T}\log T.$$

Because β is contained in the interval $[b_{\min}, b_{\max}]$, we have

$$\Delta^\pi(T) = \sup_{\theta\in\Theta}\Big\{ -\beta \sum_{t=0}^{T-1} \mathbb{E}_\theta^\pi\big(\varphi(\theta) - p_{t+1}\big)^2\Big\}$$

$$\leq |b_{\min}| \sup_{\theta\in\Theta} \sum_{t=0}^{T-1} \mathbb{E}_\theta^\pi\big(\varphi(\theta) - p_{t+1}\big)^2 .$$

Thus we deduce that $\Delta^\pi(T) \leq C_0 + C_1\sqrt{T}\log T$, where $C_0 = |b_{\min}|(u-\ell)^2 N$, and $C_1 = |b_{\min}|(8\widetilde{K}_0\rho^{-1}\kappa_0^{-1} + 2\kappa_1)$. Choosing $C = \max\{C_0, C_1\}$ concludes the proof. ■

Remark Assuming that we put equal emphasis on learning and earning (i.e., $\kappa_0 = \kappa_1$), we can minimize the constant C_1 by choosing $\kappa_0 = \kappa_1 = (4\widetilde{K}_0/\rho)^{1/2}$. With that choice, we have $C_1 = 8|b_{\min}|(\widetilde{K}_0/\rho)^{1/2}$.

A.3 An Alternative Formulation: Incumbent-Price Problem

Proof of Lemma 4. Let λ be an absolutely continuous density on $\mathfrak{B} = [b_{\min}, b_{\max}]$, taking positive values on $(b_{\min}, b_{\max})$ and zero on the extreme points $\{b_{\min}, b_{\max}\}$. By the van Trees inequality (Gill and Levit 1995), we have

$$\mathbb{E}_\lambda\big\{\mathbb{E}_\beta^\pi\big(x_t - \psi(\beta)\big)^2\big\} \geq \frac{\big(\mathbb{E}_\lambda[\psi'(\beta)]\big)^2}{\mathbb{E}_\lambda[\widehat{\mathcal{I}}_{t-1}^\pi(\beta)] + \widetilde{\mathcal{I}}(\lambda)}, \tag{A.12}$$

where $\widetilde{\mathcal{I}}(\lambda)$ is the Fisher information for the density λ, and $\mathbb{E}_\lambda(\cdot)$ is the expectation operator with respect to λ. Using the fact that $\widehat{\mathcal{I}}_{t-1}^\pi(\beta) = \sigma^{-2}\,\mathbb{E}_\beta^\pi[\widehat{J}_{t-1}]$, and adding up inequality (A.12) over $t = 2, \ldots, T$, we get

$$\sum_{t=2}^{T} \mathbb{E}_\lambda\big\{\mathbb{E}_\beta^\pi\big(x_t - \psi(\beta)\big)^2\big\} \geq \sum_{t=2}^{T} \frac{\big(\mathbb{E}_\lambda[\psi'(\beta)]\big)^2}{\sigma^{-2}\mathbb{E}_\lambda\big[\mathbb{E}_\beta^\pi[\widehat{J}_{t-1}]\big] + \widetilde{\mathcal{I}}(\lambda)}.$$

Monotonicity of $\mathbb{E}_\lambda(\cdot)$ further implies that

$$\sup_{\beta\in\mathfrak{B},\,\psi(\beta)\neq 0}\big\{\textstyle\sum_{t=2}^{T}\mathbb{E}_\beta^\pi\big(x_t - \psi(\beta)\big)^2\big\} \geq \displaystyle\sum_{t=2}^{T} \frac{\inf_{\beta\in\mathfrak{B}}\big\{\big(\psi'(\beta)\big)^2\big\}}{\sigma^{-2}\sup_{\beta\in\mathfrak{B}}\big\{\mathbb{E}_\beta^\pi[\widehat{J}_{t-1}]\big\} + \widetilde{\mathcal{I}}(\lambda)}.$$

Letting $c_0 = \sigma^2 \inf_{\beta\in\mathfrak{B}}\big\{\big(\psi'(\beta)\big)^2\big\} = \sigma^2\widehat{D}^2/(4b_{\min}^4)$, $c_1 = \sigma^2\,\widetilde{\mathcal{I}}(\lambda)$, we obtain inequality (3.14). ■

Proof of Theorem 3. Note that $\widehat{J}_s \leq (u-\ell)^2 s$ for all s. Thus, Lemma 4 implies that

$$\sup_{\beta\in\mathfrak{B},\,\psi(\beta)\neq 0}\big\{\textstyle\sum_{t=2}^{T}\mathbb{E}_\beta^\pi\big(x_t - \psi(\beta)\big)^2\big\} \;\geq\; \displaystyle\sum_{t=2}^{T} \frac{c_0}{c_1 + (u-\ell)^2(t-1)}.$$

Because β is contained in $[b_{\min}, b_{\max}]$, the preceding inequality further implies that

$$\Delta^\pi(T) = \sup_{\beta\in\mathfrak{B},\,\psi(\beta)\neq 0} \Big\{ -\beta \sum_{t=1}^{T} \mathbb{E}^\pi_\beta\big(x_t - \psi(\beta)\big)^2 \Big\}$$

$$\geq |b_{\max}| \sup_{\beta\in\mathfrak{B},\,\psi(\beta)\neq 0} \sum_{t=2}^{T} \frac{c_0}{c_1 + (u-\ell)^2(t-1)}\,.$$

To simplify notation, we let $K_1 = |b_{\max}|c_0/(u-\ell)^2$, $K_2 = c_1/(u-\ell)^2$, and deduce that

$$\Delta^\pi(T) \;\geq\; K_1 \sum_{t=1}^{T-1} \frac{1}{K_2 + t} \;\geq\; c \sum_{t=1}^{T-1} \frac{1}{t}\,,$$

where $c = K_1/(1+K_2)$. Thus, $\Delta^\pi(T) \geq c \log T$ for all $T \geq 2$. ■

Proof of Lemma 5. As in the proof of Lemma 3, we will first find an exponential supermartingale, and then apply Markov's inequality to get the desired bound. Recall that we have by (A.5) $\mathbb{E}^\pi_\beta[\exp(z\epsilon_t)] \leq \exp(\frac{1}{2}\nu^*\sigma^2 z^2)$ for all z satisfying $|z| \leq z_0$, where $\nu^* = \max_{|z|\leq z_0}\{\nu(z)\}$, and $\nu(z) = \sum_{k=2}^{\infty}\big(2z^{k-2}\mathbb{E}^\pi_\beta[\epsilon_t^k]\big)/(k!\sigma^2)$. By the least squares equation (3.15) we also have $\widehat{\beta}_t - \beta = M_t/\widehat{J}_t$. Let $y \in \mathbb{R}$, and define the process $\{\widehat{Z}^y_s,\ s = 1, 2, \ldots\}$ such that $\widehat{Z}^y_0 = 1$ and

$$\widehat{Z}^y_s := \exp\Big\{\frac{1}{\zeta^*\sigma^2}\Big(yM_s - \frac{1}{2}y^2\widehat{J}_s\Big)\Big\} \qquad \text{for all } s = 1, 2, \ldots\,, \tag{A.13}$$

where $\zeta^* = (1\vee\delta)\big(\nu^*\vee(z^*/z_0)\big)$, and $z^* = \max_{|y|\leq 1,\,\widehat{p}+x\in[\ell,u]}\{|yx|/\sigma^2\}$. The tower property implies that

$$\mathbb{E}^\pi_\beta[\widehat{Z}^y_s|\mathcal{F}_{s-1}] = \exp\Big\{\frac{1}{\zeta^*\sigma^2}\Big(yM_{s-1} - \frac{1}{2}y^2\widehat{J}_s\Big)\Big\}\,\mathbb{E}^\pi_\beta\Big[\exp\Big(\frac{yx_s\epsilon_s}{\zeta^*\sigma^2}\Big)\big|\mathcal{F}_{s-1}\Big]\,,$$

where $\mathcal{F}_s = \sigma(D_1, \ldots, D_s)$. Using inequality (A.5) and noting that $|yx_s|/(\zeta^*\sigma^2) \leq z_0|yx_s|/(\delta z^*\sigma^2) \leq z_0$ for all $\widehat{p} + x_s \in [\ell, u]$, we compute the following upper bound on the conditional expectation on the right hand side above: $\mathbb{E}^\pi_\beta[\exp\big(yx_s\epsilon_s/(\zeta^*\sigma^2)\big)|\mathcal{F}_{s-1}] \leq \exp\big(y^2x_s^2/(2\zeta^*\sigma^2)\big)$, so $(\widehat{Z}^y_s, \mathcal{F}_s)$ is a supermartingale for all $y \in \mathbb{R}$. Now choose $y = \delta$, fix $m > 0$, and note that

$$\mathbb{P}^\pi_\beta\{\widehat{\beta}_t - \beta > \delta,\ \widehat{J}_t \geq m\} = \mathbb{P}^\pi_\beta\{M_t > \delta\widehat{J}_t,\ \widehat{J}_t \geq m\}$$

$$= \mathbb{P}^\pi_\beta\{yM_t > y^2\widehat{J}_t,\ \widehat{J}_t \geq m\}\,.$$

By definition of the process $\widehat{Z}^y_t$, this implies that

$$\mathbb{P}^\pi_\beta\{\widehat{\beta}_t - \beta > \delta,\ \widehat{J}_t \geq m\} = \mathbb{P}^\pi_\beta\Big\{\widehat{Z}^y_t \geq \exp\Big(\frac{\delta^2\widehat{J}_t}{2\zeta^*\sigma^2}\Big),\ \widehat{J}_t \geq m\Big\}$$

$$\leq \mathbb{P}^\pi_\beta\Big\{\widehat{Z}^y_t \geq \exp\Big(\frac{\delta^2 m}{2\zeta^*\sigma^2}\Big)\Big\}.$$

Using Markov's inequality, and the fact that $\widehat{Z}_t^y$ is a martingale with $\widehat{Z}_0^y = 1$, we obtain the following exponential bound:

$$\mathbb{P}_\beta^\pi\{\widehat{\beta}_t - \beta > \delta,\ \widehat{J}_t \geq m\} \leq \exp\Big(-\frac{\delta^2 m}{2\zeta^*\sigma^2}\Big).$$

Repeating the above argument for $y = -\delta$, we derive the same exponential bound on $\mathbb{P}_\beta^\pi\{\widehat{\beta}_t - \beta < -\delta,\ \widehat{J}_t \geq m\}$. Finally, we let $\rho = (1 \vee \delta)/(2\zeta^*\sigma^2) = \frac{1}{2}\gamma\sigma^{-2}\big(\nu^* \vee (z^*/z_0)\big)^{-1}$, and arrive at the desired result. ■

Proof of Theorem 4. We begin with decomposing the revenue loss in period $t+1$. Note that

$$\mathbb{E}_\beta^\pi\big(\psi(\beta) - x_{t+1}\big)^2 \leq 2\mathbb{E}_\beta^\pi\big(\psi(\beta) - \psi(b_t)\big)^2 + 2\mathbb{E}_\beta^\pi\big(\psi(b_t) - x_{t+1}\big)^2, \tag{A.14}$$

where b_t is the truncated least squares estimate calculated at the end of period t. To find an upper bound on the first term on the right side of the preceding inequality, we define the sequence of events $\{A_1, A_2, \ldots\}$ such that

$$A_t \ = \bigcap_{s=\lfloor\frac{t-1}{2}\rfloor}^{t-1} \left\{\big(\psi(\beta) - \psi(b_s)\big)^2 \leq \frac{\big(\psi(\beta)\big)^2}{12}\right\} \quad \text{for all } t = 1, 2, \ldots \tag{A.15}$$

By the triangle inequality, we have

$$\big(\psi(\beta)\big)^2 \ \leq\ 3\big(\psi(\beta) - \psi(b_s)\big)^2 \ +\ 3\big(\psi(b_s) - x_{s+1}\big)^2 \ +\ 3x_{s+1}^2$$

for all s between $\lfloor(t-1)/2\rfloor$ and $t-1$. Thus one has (almost surely) the following on A_t:

$$\begin{aligned}\widehat{J}_t &\geq \sum_{s=\lfloor\frac{t-1}{2}\rfloor}^{t-1} x_{s+1}^2 \\ &\geq \sum_{s=\lfloor\frac{t-1}{2}\rfloor}^{t-1} \frac{\big(\psi(\beta)\big)^2}{3} \ - \sum_{s=\lfloor\frac{t-1}{2}\rfloor}^{t-1} \big(\psi(\beta) - \psi(b_s)\big)^2 \ - \sum_{s=\lfloor\frac{t-1}{2}\rfloor}^{t-1} \big(\psi(b_s) - x_{s+1}\big)^2.\end{aligned} \tag{A.16}$$

Note that the first sum on the right side of inequality (A.16) is greater than $\big(\psi(\beta)\big)^2(t-1)/6$, while the second sum is less than $\big(\psi(\beta)\big)^2(t-1)/24$ by definition of the event A_t. Moreover, the third sum is less than $\kappa_1 \log(t-1)$ by condition (ii) in the hypothesis. Combining all these facts, we have

$$\widehat{J}_t \geq \left(\frac{\big(\psi(\beta)\big)^2}{8} - \frac{\kappa_1 \log(t-1)}{t-1}\right)(t-1) \quad \text{on } A_t.$$

Since the sequence $s^{-1}\log s$ converges to zero as s tends to ∞, there exists a natural number $N_0(\beta)$ such that for all $t \geq N_0(\beta)$ we have $\widehat{J}_t \geq \big(\psi(\beta)\big)^2(t-1)/16$ on the event A_t.

Having shown that $\widehat{J}_t$ grows linearly on A_t, we now turn our attention to prove that each A_t is a very likely event. The union bound implies

$$\mathbb{P}_\beta^\pi(A_t^c) \ \leq \sum_{s=\lfloor(t-1)/2\rfloor}^{t-1} \mathbb{P}_\beta^\pi\left\{\big(\psi(\beta) - \psi(b_s)\big)^2 > \frac{\big(\psi(\beta)\big)^2}{12}\right\}.$$

By the mean value theorem, we deduce that $|\psi(\beta) - \psi(b_s)| \le \sqrt{K_0}\,|\beta - b_s|$, where $K_0 = \max_{\beta\in[b_{\min},b_{\max}]} \big(\psi'(\beta)\big)^2 = \widehat{D}^2/(4b_{\max}^4)$. Thus, the preceding inequality leads to

$$\mathbb{P}^\pi_\beta(A_t^c) \le \sum_{s=\lfloor (t-1)/2\rfloor}^{t-1} \mathbb{P}^\pi_\beta\left\{\sqrt{K_0}\,|\beta - b_s| > \frac{|\psi(\beta)|}{\sqrt{12}}\right\}.$$

Invoking condition (i) in the hypothesis, we further get

$$\begin{aligned}
\mathbb{P}^\pi_\beta(A_t^c) &\le \sum_{s=\lfloor (t-1)/2\rfloor}^{t-1} \mathbb{P}^\pi_\beta\left\{|\beta - b_s| > \frac{|\psi(\beta)|}{\sqrt{12K_0}},\ \widehat{J}_s \ge \kappa_0\log s\right\}\\
&\overset{\text{(a)}}{\le} \sum_{s=\lfloor (t-1)/2\rfloor}^{t-1} \mathbb{P}^\pi_\beta\left\{|\beta - \widehat{\beta}_s| > \frac{|\psi(\beta)|}{\sqrt{12K_0}},\ \widehat{J}_s \ge \kappa_0\log s\right\}\\
&\overset{\text{(b)}}{\le} \sum_{s=\lfloor (t-1)/2\rfloor}^{t-1} 2\exp\left(-\rho\kappa_0\left(\frac{|\psi(\beta)|}{\sqrt{12K_0}} \wedge \frac{\big(\psi(\beta)\big)^2}{12K_0}\right)\log s\right) \;\le\; \sum_{s=\lfloor (t-1)/2\rfloor}^{t-1} s^{-q}
\end{aligned}$$

where: (a) follows since $|\beta - \widehat{\beta}_s| \ge |\beta - b_s|$; (b) follows by Lemma 5; and $q := \frac{1}{12}\rho\kappa_0\{\frac{|\psi(\beta)|}{\sqrt{K_0}} \wedge \frac{(\psi(\beta))^2}{K_0}\}$. Choosing κ_0 greater than $36\rho^{-1}\{\frac{|\psi(\beta)|}{\sqrt{K_0}} \wedge \frac{(\psi(\beta))^2}{K_0}\}^{-1}$, we ensure that $q \ge 3$. Thus, for all $t \ge 5$, we have

$$\mathbb{P}^\pi_\beta(A_t^c) \;\le\; \sum_{s=\lfloor \frac{t-1}{2}\rfloor}^{\infty} s^{-3} \;\le\; \int_{\lfloor \frac{t-1}{2}\rfloor}^{\infty} (s-1)^{-3}ds \;\le\; 2(t-4)^{-2},$$

hence $\sum_{t=5}^{\infty} \mathbb{P}^\pi_\beta(A_t^c) \le \pi^2/3$. Recalling inequality (A.14), we can now bound the first term on the right side as follows. Let $N(\beta) = \max\{N_0(\beta), 5\}$. Then for any time period $t \ge N(\beta)$, we have by virtue of the mean value theorem that

$$\begin{aligned}
\mathbb{E}^\pi_\beta\big(\psi(\beta) - \psi(b_t)\big)^2 &= \mathbb{E}^\pi_\beta\big[\big(\psi(\beta) - \psi(b_t)\big)^2;\, A_t\big] \;+\; \mathbb{E}^\pi_\beta\big[\big(\psi(\beta) - \psi(b_t)\big)^2;\, A_t^c\big]\\
&\le K_0\,\mathbb{E}^\pi_\beta[(\beta - b_t)^2; A_t] \;+\; \mathbb{E}^\pi_\beta\big[\big(\psi(\beta) - \psi(b_t)\big)^2;\, A_t^c\big]. \qquad \text{(A.17)}
\end{aligned}$$

Since $|\beta - \widehat{\beta}_t| \ge |\beta - b_t|$, the first term on the right side of inequality (A.17) is bounded above by $K_0\,\mathbb{E}^\pi_\beta[(\beta - \widehat{\beta}_t)^2; A_t]$. By the least squares equation (3.16), this further implies that

$$\begin{aligned}
K_0\,\mathbb{E}^\pi_\beta[(\beta - b_t)^2; A_t] &\le K_0\,\mathbb{E}^\pi_\beta\left[\frac{M_t^2}{\widehat{J}_t^2};\, A_t\right]\\
&\le \frac{256K_0}{\big(\psi(\beta)\big)^4(t-1)^2}\,\mathbb{E}^\pi_\beta[M_t^2;\, A_t] \qquad \text{(A.18)}
\end{aligned}$$

since for all $t \ge N(\beta)$ we have $\widehat{J}_t \ge \big(\psi(\beta)\big)^2(t-1)/16$ on the event A_t. Noting that $\mathbb{E}^\pi_\beta[M_t^2;\, A_t] \le \mathbb{E}^\pi_\beta[M_t^2] = \sigma^2\,\mathbb{E}^\pi_\beta[\widehat{J}_t] \le \sigma^2(u-\ell)^2 t$, we get

$$K_0\,\mathbb{E}^\pi_\beta[(\beta - b_t)^2; A_t] \le \frac{K_1}{t},$$

for all $t \geq N(\beta)$, where $K_1 = 512 K_0 \sigma^2 (u-\ell)^2 \big(\psi(\beta)\big)^{-4}$. Having found an upper bound on the first term on the right side of inequality (A.17), we also notice that the second term is smaller than $4(u-\ell)^2\, \mathbb{P}^{\pi}_{\beta}(A_t^c)$, because both the optimal price $\widehat{p} + \psi(\beta)$ and the greedy ILS price $\widehat{p} + \psi(b_t)$ are contained in $[\ell, u]$. So, letting $K_2 = 4(u-\ell)^2$, we deduce by inequality (A.17) that

$$\mathbb{E}^{\pi}_{\beta}\big(\psi(\beta) - \psi(b_t)\big)^2 \;\leq\; K_1\, t^{-1} \;+\; K_2\, \mathbb{P}^{\pi}_{\beta}(A_t^c). \tag{A.19}$$

Thus, inequality (A.14) becomes

$$\mathbb{E}^{\pi}_{\beta}\big(\psi(\beta) - x_{t+1}\big)^2 \;\leq\; 2K_1\, t^{-1} \;+\; 2K_2\, \mathbb{P}^{\pi}_{\beta}(A_t^c) \;+\; 2\mathbb{E}^{\pi}_{\beta}\big(\psi(b_t) - x_{t+1}\big)^2.$$

Letting $T \geq N(\beta)$, and adding the preceding inequality over $t = N(\beta), \ldots, T$, one obtains

$$\sum_{t=N(\beta)}^{T} \mathbb{E}^{\pi}_{\beta}\big(\psi(\beta) - x_{t+1}\big)^2 \;\leq\; 2K_1 \log T \;+\; 2K_2 \pi^2/3 \;+\; 2 \sum_{t=N(\beta)}^{T} \mathbb{E}^{\pi}_{\beta}\big(\psi(b_t) - x_{t+1}\big)^2.$$

By condition (ii) in the hypothesis, the third term on the right hand side is bounded from above by $\kappa_1 \log T$, so we deduce that $\Delta^{\pi}_{\beta}(T) \leq C_0 + C_1 \log T$, with $C_0 = 4|b_{\min}|\big((u-\ell)^2 N(\beta) \;+\; K_2\pi^2/6\big)$ and $C_1 = |b_{\min}|(2K_1 + \kappa_1)$. To complete the proof, let $C = \max\{C_0, C_1\}$. ■

Appendix B: Proofs of the Results in Section 4

B.1 Generalization of Main Ideas

Proof of Lemma 6. In what follows, we will extend the proof of Lemma 1 to higher dimensions. As before, we choose λ to be an absolutely continuous density on Θ, taking positive values on the interior of Θ and zero on its boundary. Let $\mathcal{C}(\cdot)$ be $n \times (n^2+n)$ matrix function on Θ such that $\mathcal{C}(\theta) = \mathbf{1}_n^{\mathsf{T}} \otimes C(\theta)$, where $\mathbf{1}_n$ is an $n \times 1$ vector of ones. That is, $\mathcal{C}(\theta) = [\mathcal{C}_1(\theta) \cdots \mathcal{C}_n(\theta)]$, where

$$\mathcal{C}_i(\theta) \;:=\; C(\theta) \;=\; \begin{bmatrix} -\varphi_1(\theta) & 1 & 0 & \cdots & 0 \\ -\varphi_2(\theta) & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\varphi_n(\theta) & 0 & 0 & \cdots & 1 \end{bmatrix} \qquad \text{for } i = 1, \ldots, n \text{ and all } \theta \in \Theta.$$

Invoking the multivariate van Trees inequality (Gill and Levit 1995), we obtain

$$\mathbb{E}_{\lambda}\big\{\mathbb{E}^{\pi}_{\theta}\big\|p_t - \varphi(\theta)\big\|^2\big\} \geq \frac{\Big(\mathbb{E}_{\lambda}\big[\,\mathsf{tr}\big\{\mathcal{C}(\theta)\,(\partial\varphi/\partial\theta)^{\mathsf{T}}\big\}\,\big]\Big)^2}{\mathbb{E}_{\lambda}\big[\,\mathsf{tr}\big\{\mathcal{C}(\theta)\,\mathcal{I}^{\pi}_{t-1}(\theta)\,\mathcal{C}(\theta)^{\mathsf{T}}\big\}\,\big] + \widetilde{\mathcal{I}}(\lambda)}, \tag{B.1}$$

where $\widetilde{\mathcal{I}}(\lambda)$ is the Fisher information for the density λ, and $\mathbb{E}_{\lambda}(\cdot)$ is the expectation operator with respect to λ. We begin by computing $\mathsf{tr}\big\{\mathcal{C}(\theta)\,(\partial\varphi/\partial\theta)^{\mathsf{T}}\big\}$, which appears in the numerator on the right side. Applying the implicit function theorem on the first-order condition of optimality expressed in (2.4), we deduce that

$$\frac{\partial \varphi}{\partial \theta} \;=\; -\left(\frac{\partial \eta}{\partial p}\right)^{-1}\frac{\partial \eta}{\partial \theta} \;=\; -\left(B + B^{\mathsf{T}}\right)^{-1}\frac{\partial \eta}{\partial \theta}.$$

Because $B + B^{\mathsf{T}}$ is symmetric and negative definite, it has a symmetric inverse. So, we get

$$\mathsf{tr}\Big\{\mathcal{C}(\theta)\left(\frac{\partial \varphi}{\partial \theta}\right)^{\mathsf{T}}\Big\} = -\mathsf{tr}\Big\{\mathcal{C}(\theta)\left(\frac{\partial \eta}{\partial \theta}\right)^{\mathsf{T}}\left(B + B^{\mathsf{T}}\right)^{-1}\Big\}.$$

Note that for matrices A, B, C, D of such size that one can form the matrix products AC and BD, one has $(A \otimes B)(C \otimes D) = AC \otimes BD$. Using this result, which will hereafter be called *the mixed product property*, and the fact that

$$\left(\frac{\partial \eta}{\partial \theta}\right)^{\mathsf{T}} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ \varphi_1(\theta) & 0 & \cdots & 0 \\ 0 & \varphi_1(\theta) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \varphi_1(\theta) \\ \\ \vdots & \vdots & \vdots & \vdots \\ \\ 0 & 0 & \cdots & 0 \\ \varphi_n(\theta) & 0 & \cdots & 0 \\ 0 & \varphi_n(\theta) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \varphi_n(\theta) \end{bmatrix} + \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \varphi_1(\theta) & 0 & \cdots & 0 \\ \varphi_2(\theta) & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_n(\theta) & 0 & \cdots & 0 \\ \\ \vdots & \vdots & \vdots & \vdots \\ \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & \varphi_1(\theta) \\ 0 & 0 & \cdots & \varphi_2(\theta) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \varphi_n(\theta) \end{bmatrix}$$

$$= \varphi(\theta) \otimes \begin{bmatrix} \mathbf{0}_n \\ \mathbf{I}_n \end{bmatrix} + \mathbf{I}_n \otimes \begin{bmatrix} 1 \\ \varphi(\theta) \end{bmatrix},$$

we further deduce that $\mathcal{C}(\theta)\,(\partial\eta/\partial\theta)^{\mathsf{T}} = \big(\mathbf{1}_n^{\mathsf{T}}\,\varphi(\theta)\big) \otimes \mathbf{I}_n + \mathbf{1}_n^{\mathsf{T}} \otimes \mathbf{0}_n = \mathbf{I}_n \sum_{i=1}^{n} \varphi_i(\theta)$, which in turn implies that

$$\mathsf{tr}\Big\{\mathcal{C}(\theta)\left(\frac{\partial \varphi}{\partial \theta}\right)^{\mathsf{T}}\Big\} = \mathsf{tr}\Big\{\left(-B - B^{\mathsf{T}}\right)^{-1}\Big\}\sum_{i=1}^{n}\varphi_i(\theta). \tag{B.2}$$

Because the eigenvalues of the matrix B are in the interval $[b_{\min}, b_{\max}]$, we obtain the following by elementary algebra:

$$\mathsf{tr}\Big\{\left(-B - B^{\mathsf{T}}\right)^{-1}\Big\} \;\geq\; \frac{n^2}{\mathsf{tr}\big\{-B - B^{\mathsf{T}}\big\}} \;\geq\; \frac{n}{2|b_{\min}|}.$$

Combined with equation (B.2) and the fact that $\varphi_i(\theta) \geq \ell > 0$ for all i, the preceding inequality implies that $\mathsf{tr}\big\{\mathcal{C}(\theta)\,(\partial\varphi/\partial\theta)^{\mathsf{T}}\big\} \geq n^2\ell/(2|b_{\min}|)$. Now, we turn our attention to computing

$\mathsf{tr}\{\mathcal{C}(\theta)\,\mathcal{I}^{\pi}_{t-1}(\theta)\,\mathcal{C}(\theta)^{\mathsf{T}}\}$, which is in the denominator on the right side of (B.1). By the mixed product property, we have

$$\begin{aligned}\mathcal{C}(\theta)\,\mathcal{I}^{\pi}_{t-1}(\theta)\,\mathcal{C}(\theta)^{\mathsf{T}} &= \frac{1}{\sigma^2}\,\mathbb{E}^{\pi}_{\theta}\{\mathbf{1}_n^{\mathsf{T}}\mathbf{I}_n\mathbf{1}_n \otimes [-\varphi(\theta)\ \ \mathbf{I}_n]\ \mathcal{J}_{t-1}\ [-\varphi(\theta)\ \ \mathbf{I}_n]^{\mathsf{T}}\} \\ &= \frac{n}{\sigma^2}\sum_{s=1}^{t-1}\mathbb{E}^{\pi}_{\theta}\big(p_s-\varphi(\theta)\big)\big(p_s-\varphi(\theta)\big)^{\mathsf{T}}.\end{aligned}$$

Thus, $\mathsf{tr}\{\mathcal{C}(\theta)\,\mathcal{I}^{\pi}_{t-1}(\theta)\,\mathcal{C}(\theta)^{\mathsf{T}}\} = n\sigma^{-2}\sum_{s=1}^{t-1}\mathbb{E}^{\pi}_{\theta}\big\|p_s-\varphi(\theta)\big\|^2 = n\sigma^{-2}\,\mathsf{tr}\{\,C(\theta)\,\mathbb{E}^{\pi}_{\theta}[\mathcal{J}_{t-1}]\,C(\theta)^{\mathsf{T}}\,\}$. By this identity and the fact that $\mathsf{tr}\{\mathcal{C}(\theta)\,(\partial\varphi/\partial\theta)^{\mathsf{T}}\} \ge n^2\ell/(2|b_{\min}|)$, inequality (B.1) becomes

$$\mathbb{E}_{\lambda}\{\mathbb{E}^{\pi}_{\theta}\big\|p_t-\varphi(\theta)\big\|^2\} \ge \frac{c_0}{c_1+\mathbb{E}_{\lambda}\big[\,\mathsf{tr}\{\,C(\theta)\,\mathbb{E}^{\pi}_{\theta}[\mathcal{J}_{t-1}]\,C(\theta)^{\mathsf{T}}\,\}\,\big]},$$

where $c_0 = n^3\sigma^2\ell^2/(4b_{\min}^2)$ and $c_1 = \sigma^2\widetilde{\mathcal{I}}(\lambda)/n$. Summing over $2,\ldots,T$, and using the monotonicity of $\mathbb{E}_{\lambda}(\cdot)$ completes the proof. ■

Proof of Theorem 5. Note that $\mathsf{tr}\{\,C(\theta)\,\mathbb{E}^{\pi}_{\theta}[\mathcal{J}_{t-1}]\,C(\theta)^{\mathsf{T}}\,\} = \sum_{s=1}^{t-1}\mathbb{E}^{\pi}_{\theta}\big\|p_s-\varphi(\theta)\big\|^2$, implying that inequality (4.3) of Lemma 6 can also be expressed as

$$\sup\nolimits_{\theta\in\Theta}\big\{\textstyle\sum_{t=2}^{T}\mathbb{E}^{\pi}_{\theta}\big\|p_t-\varphi(\theta)\big\|^2\big\}\ \ge\ \displaystyle\sum_{t=2}^{T}\frac{c_0}{c_1+\sup_{\theta\in\Theta}\big\{\sum_{s=1}^{t-1}\mathbb{E}^{\pi}_{\theta}\big\|p_s-\varphi(\theta)\big\|^2\big\}}\,. \tag{B.3}$$

Now, because the eigenvalues of the B are contained in $[b_{\min},\, b_{\max}]$, we have by the definition of the worst-case regret (2.8),

$$\begin{aligned}\Delta^{\pi}(T) &= \sup_{\theta\in\Theta}\Big\{\sum_{t=1}^{T}\mathbb{E}^{\pi}_{\theta}\big[r^*_{\theta}-r_{\theta}(p_t)\big]\Big\} \\ &= \sup_{\theta\in\Theta}\Big\{\sum_{t=1}^{T}\mathbb{E}^{\pi}_{\theta}\big[\varphi(\theta)\cdot\big(a+B\varphi(\theta)\big)\ -\ p_t\cdot(a+Bp_t)\big]\Big\} \\ &= \sup_{\theta\in\Theta}\Big\{-\sum_{t=1}^{T}\mathbb{E}^{\pi}_{\theta}\big(p_t-\varphi(\theta)\big)^{\mathsf{T}}B\big(p_t-\varphi(\theta)\big)\Big\} \\ &\ge |b_{\max}|\,\sup_{\theta\in\Theta}\Big\{\sum_{t=2}^{T}\mathbb{E}^{\pi}_{\theta}\big\|p_t-\varphi(\theta)\big\|^2\Big\}.\end{aligned}$$

Combining the preceding inequality with (B.3), we obtain

$$\Delta^{\pi}(T)\ \ge\ b_{\max}^2\sum_{t=2}^{T}\frac{c_0}{\Delta^{\pi}(t-1)+|b_{\max}|c_1}. \tag{B.4}$$

Note that the inequality immediately above is identical to inequality (A.3) in the proof of Theorem 1. To complete the proof, we apply exactly the same argument that follows (A.3). ■

Proof of Lemma 7. First, note that $\mu_{\min}(t)$ is also the smallest eigenvalue of $\mathcal{J}_t$ because the unique eigenvalue of $\mathbf{I}_n$ is 1. Let y be an arbitrary unit vector in $\mathbb{R}^{n+1}$, and y_* be the $n \times 1$ vector consisting of the last n components of y, that is, $y_* = (y_2, \ldots, y_{n+1})$. Then,

$$y^\top \mathcal{J}_t y \;=\; \sum_{s=1}^{t}(y_1 + y_* \cdot p_s)^2 \;=\; \sum_{s=1}^{t}(y_1 + y_* \cdot \overline{p}_t)^2 + \sum_{s=1}^{t}\big(y_* \cdot (p_s - \overline{p}_t)\big)^2.$$

As in the proof of Lemma 2, we would like to find a lower bound on $y^\top \mathcal{J}_t y$ to invoke the Rayleigh-Ritz theorem. Note that the first term on the right side above is $(y_1 + y_* \cdot \overline{p}_t)^2 t$, whereas the second term equals $\sum_{s=1}^{t}\big(y_* \cdot (p_s - \overline{p}_t)\big)^2 = y_*^\top \widetilde{\mathcal{J}}_t y_*$, where $\widetilde{\mathcal{J}}_t := \sum_{s=1}^{t}(p_s - \overline{p}_t)(p_s - \overline{p}_t)^\top = \sum_{s=2}^{t}(1 - s^{-1})(p_s - \overline{p}_{s-1})(p_s - \overline{p}_{s-1})^\top$. Therefore, we have

$$y^\top \mathcal{J}_t y \;=\; (y_1 + y_* \cdot \overline{p}_t)^2 t \;+\; \sum_{s=2}^{t}(1 - s^{-1})(y_* \cdot \widetilde{x}_s)^2$$

where $\widetilde{x}_s = p_s - \overline{p}_{s-1}$ for all s. By the orthogonal pricing condition, we know that for each $\tau = 1, 2, \ldots$ there exists $X_\tau \in \mathbb{R}^n$ and $v_\tau > 0$ such that $V_\tau = \big\{v_\tau^{-1}\big(\delta_s - X_\tau\big) \;:\; s = n(\tau-1) + 1, \ldots, n(\tau-1) + n\big\}$ is an orthonormal basis of $\mathbb{R}^n$. Thus the sum on the right hand side above can also be expressed as

$$\begin{aligned}\sum_{s=2}^{t}(1 - s^{-1})(y_* \cdot \widetilde{x}_s)^2 \;&=\; \sum_{s=2}^{t}(1 - s^{-1})\big(y_* \cdot X_{\lceil s/n \rceil} \;+\; y_* \cdot (\widetilde{x}_s - X_{\lceil s/n \rceil})\big)^2 \\ &\geq\; \frac{1}{2}\sum_{\tau=1}^{\lfloor t/n \rfloor}\sum_{i=1}^{n}\big(y_* \cdot X_\tau \;+\; y_* \cdot (\widetilde{x}_{n(\tau-1)+i} - X_\tau)\big)^2.\end{aligned}$$

Letting $e_{i\tau} = v_\tau^{-1}\big(\widetilde{x}_{n(\tau-1)+i} - X_\tau\big)$, $\widetilde{y}_{i\tau} = y_* \cdot e_{i\tau}$, and recalling that $e_{i\tau} \cdot X_\tau = n^{-1/2} v_\tau \|X_\tau\|$ for all $i = 1, \ldots, n$, we deduce by the preceding inequality that

$$\begin{aligned}\sum_{s=2}^{t}(1 - s^{-1})(y_* \cdot \widetilde{x}_s)^2 \;&\geq\; \frac{1}{2}\sum_{\tau=1}^{\lfloor t/n \rfloor}\sum_{i=1}^{n}\left(n^{-1/2}\|X_\tau\|\sum_{j=1}^{n}\widetilde{y}_{j\tau} \;+\; v_\tau \widetilde{y}_{i\tau}\right)^2 \\ &=\; \frac{1}{2}\sum_{\tau=1}^{\lfloor t/n \rfloor}\sum_{i=1}^{n}\left(n^{-1}\|X_\tau\|^2\Big(\sum_{j=1}^{n}\widetilde{y}_{j\tau}\Big)^2 \;+\; 2n^{-1/2}\|X_\tau\|\Big(\sum_{j=1}^{n}\widetilde{y}_{j\tau}\Big)v_\tau \widetilde{y}_{i\tau} \;+\; v_\tau^2 \widetilde{y}_{i\tau}^2\right) \\ &\geq\; \frac{1}{2}\sum_{\tau=1}^{\lfloor t/n \rfloor} v_\tau^2 \sum_{i=1}^{n}\widetilde{y}_{i\tau}^2\,.\end{aligned}$$

Because $\|y_*\|^2 = \sum_{i=1}^{n}\widetilde{y}_{i\tau}^2$, and $\|\widetilde{x}_{n(\tau-1)+i} - X_\tau\| = v_\tau$ for all $i = 1, \ldots, n$, we deduce that

$$\sum_{s=2}^{t}(1 - s^{-1})(y_* \cdot \widetilde{x}_s)^2 \;\geq\; \frac{1}{2}\|y_*\|^2 \sum_{\tau=1}^{\lfloor t/n \rfloor} v_\tau^2\,.$$

Using the definition of the information metric J_t, we have $y^\mathsf{T}\mathcal{J}_t y \geq (y_1 + y_* \cdot \overline{p}_t)^2 t + \frac{1}{2}\|y_*\|^2 J_t$. Finally, we note that $J_t \leq (u-\ell)^2 t$, and deduce by elementary algebra that $y^\mathsf{T}\mathcal{J}_t y \geq \gamma J_t$, where $\gamma = 1/(1+2u-\ell)^2$. Because y was selected arbitrarily, the Rayleigh-Ritz theorem implies that $\mu_{\min}(t) \geq \gamma J_t$. ■

Proof of Lemma 8. To get the desired result, we will extend the proof of Lemma 3 to higher dimensions. By (A.5), we have $\mathbb{E}_\theta^\pi[\exp(z\epsilon_{it})] \leq \exp(\frac{1}{2}\nu^*\sigma^2 z^2)$ for all z satisfying $|z| \leq z_0$, where $\nu^* = \max_{|z|\leq z_0}\{\nu(z)\}$, and $\nu(z) = \sum_{k=2}^\infty \big(2z^{k-2}\mathbb{E}_\theta^\pi[\epsilon_{it}^k]\big)/(k!\sigma^2)$. Note that the least squares equation (2.10) implies that $\widehat{\theta}_t - \theta = (\mathbf{I}_n \otimes \mathcal{J}_t^{-1})\,\mathcal{M}_t$, where $\mathcal{M}_t$ is the $(n^2+n)\times 1$ vector $\mathcal{M}_t = (\mathcal{M}_{1t}, \ldots, \mathcal{M}_{nt})$. Select y to be an $(n^2+n)\times 1$ vector with $\|y\| = \delta$. Define the process $\{\mathcal{Z}_s^y,\ s = 1, 2, \ldots\}$ such that $\mathcal{Z}_0^y = 1$ and

$$\mathcal{Z}_s^y := \exp\Big\{\frac{1}{\zeta^*\sigma^2}\Big(y\cdot\mathcal{M}_s - \frac{1}{2}y^\mathsf{T}(\mathbf{I}_n\otimes\mathcal{J}_s)y\Big)\Big\} \quad \text{for all } s = 1, 2, \ldots, \tag{B.5}$$

where $\zeta^* = (1\vee\delta)\big(\nu^*\vee(z^*/z_0)\big)$, and $z^* = \max_{y\in\mathbb{R}^{n+1},\,\|y\|\leq 1,\,p\in[\ell,u]^n}\big\{|y\cdot\big[\begin{smallmatrix}1\\ p\end{smallmatrix}\big]|/\sigma^2\big\}$. To show that $\mathcal{Z}_s^y$ is a supermartingale, we let $\mathcal{F}_s := \sigma(D_1, \ldots, D_s)$, and note that

$$\mathbb{E}_\theta^\pi[\mathcal{Z}_s^y|\mathcal{F}_{s-1}] \;=\; \exp\Big\{\frac{1}{\zeta^*\sigma^2}\Big(y\cdot\mathcal{M}_{s-1} - \frac{1}{2}y^\mathsf{T}(\mathbf{I}_n\otimes\mathcal{J}_s)y\Big)\Big\}\ \mathbb{E}_\theta^\pi\Big[\exp\Big\{\frac{1}{\zeta^*\sigma^2}\,y\cdot(\mathcal{M}_s - \mathcal{M}_{s-1})\Big\}\Big|\mathcal{F}_{s-1}\Big].$$

Letting $y^i = \big(y_{(n+1)(i-1)+1}, \ldots, y_{(n+1)(i-1)+n+1}\big) \in \mathbb{R}^{n+1}$, the conditional expectation on the right hand side of the preceding inequality can be bounded as

$$\begin{aligned}
\mathbb{E}_\theta^\pi\Big[\exp\Big\{\frac{1}{\zeta^*\sigma^2}\,y\cdot(\mathcal{M}_s - \mathcal{M}_{s-1})\Big\}\Big|\mathcal{F}_{s-1}\Big] &\overset{\text{(a)}}{\leq} \prod_{i=1}^n \mathbb{E}_\theta^\pi\Big[\exp\Big\{\frac{1}{\zeta^*\sigma^2}\,\big(y^i\cdot\big[\begin{smallmatrix}1\\ p_s\end{smallmatrix}\big]\big)\epsilon_{is}\Big\}\Big|\mathcal{F}_{s-1}\Big] \\
&\overset{\text{(b)}}{\leq} \prod_{i=1}^n \exp\Big\{\frac{1}{2\zeta^*\sigma^2}\big(y^i\cdot\big[\begin{smallmatrix}1\\ p_s\end{smallmatrix}\big]\big)^2\Big\} \\
&\overset{\text{(c)}}{\leq} \exp\Big\{\frac{1}{2\zeta^*\sigma^2}\,y^\mathsf{T}\big(\mathbf{I}_n\otimes\big[\begin{smallmatrix}1\\ p_s\end{smallmatrix}\big]\big)\big(\mathbf{I}_n\otimes\big[\begin{smallmatrix}1\\ p_s\end{smallmatrix}\big]^\mathsf{T}\big)y\Big\},
\end{aligned} \tag{B.6}$$

where: (a) follows by the independence of the ϵ_{it}, and the fact that $y\cdot(\mathcal{M}_s - \mathcal{M}_{s-1}) = y\cdot\big(\epsilon_s\otimes\big[\begin{smallmatrix}1\\ p_s\end{smallmatrix}\big]\big) = \sum_{i=1}^n\big(y^i\cdot\big[\begin{smallmatrix}1\\ p_s\end{smallmatrix}\big]\big)\epsilon_{is}$; (b) follows by inequality (A.5), and the fact that $\big|y^i\cdot\big[\begin{smallmatrix}1\\ p_s\end{smallmatrix}\big]\big|/(\zeta^*\sigma^2) \leq z_0\big|y^i\cdot\big[\begin{smallmatrix}1\\ p_s\end{smallmatrix}\big]\big|/(\delta z^*\sigma^2) \leq z_0$ for all $i = 1, \ldots, n$, and $p_s \in [\ell, u]^n$; and (c) follows because $\sum_{i=1}^n\big(y^i\cdot\big[\begin{smallmatrix}1\\ p_s\end{smallmatrix}\big]\big)^2 = y^\mathsf{T}\big(\mathbf{I}_n\otimes\big[\begin{smallmatrix}1\\ p_s\end{smallmatrix}\big]\big)\big(\mathbf{I}_n\otimes\big[\begin{smallmatrix}1\\ p_s\end{smallmatrix}\big]^\mathsf{T}\big)y$. Using the mixed product property, we also get

$$\big(\mathbf{I}_n\otimes\big[\begin{smallmatrix}1\\ p_s\end{smallmatrix}\big]\big)\big(\mathbf{I}_n\otimes\big[\begin{smallmatrix}1\\ p_s\end{smallmatrix}\big]^\mathsf{T}\big) \;=\; \mathbf{I}_n\otimes\big(\big[\begin{smallmatrix}1\\ p_s\end{smallmatrix}\big]\cdot\big[\begin{smallmatrix}1\\ p_s\end{smallmatrix}\big]^\mathsf{T}\big).$$

As a result, $\mathbb{E}_\theta^\pi[\mathcal{Z}_s^y|\mathcal{F}_{s-1}] \leq \mathcal{Z}_{s-1}^y$, and hence $\mathcal{Z}_s^y$ is a supermartingale. For the rest of this proof, we choose $w_s = \delta(\mathbf{I}_n\otimes\mathcal{J}_s^{-1})\,\mathcal{M}_s/\|(\mathbf{I}_n\otimes\mathcal{J}_s^{-1})\,\mathcal{M}_s\|$, and define the stochastic process $\{\widetilde{\mathcal{Z}}_s\}$ such that $\widetilde{\mathcal{Z}}_s = \mathcal{Z}_s^{w_s}$ for all s. Fixing $m > 0$, and defining the event $A := \{\|\mathcal{M}_t\| \leq \xi t\} \in \mathcal{F}_t$ for $\xi > 0$, we obtain the following by the argument we used to derive inequality (A.7):

$$\mathbb{P}_\theta^\pi\big(\|\widehat{\theta}_t - \theta\| > \delta,\ J_t \geq m\big) \;\leq\; \mathbb{P}_\theta^\pi\big(\|(\mathbf{I}_n\otimes\mathcal{J}_t^{-1})\,\mathcal{M}_t\| > \delta,\ J_t \geq m,\ A\big) \;+\; \mathbb{P}_\theta^\pi\big(J_t \geq m,\ A^c\big). \tag{B.7}$$

As argued in the proof of Lemma 3, we use the definition of $\widetilde{\mathcal{Z}}_t$ and the Rayleigh-Ritz theorem to get the following upper bound on the first probability on the right hand side of (B.7):

$$\mathbb{P}_\theta^\pi\big(\|(\mathbf{I}_n \otimes \mathcal{J}_t^{-1})\,\mathcal{M}_t\| > \delta,\ J_t \geq m,\ A\big) \ \leq\ \mathbb{P}_\theta^\pi\Big(\widetilde{\mathcal{Z}}_t \geq \exp\Big\{\frac{\mu_{\min}(t)\delta^2}{2\zeta^*\sigma^2}\Big\},\ J_t \geq m,\ A\Big),$$

where $\mu_{\min}(t)$ is the smallest eigenvalue of empirical Fisher information matrix $\mathbf{I}_n \otimes \mathcal{J}_t$. Recalling that the pricing policy π is orthogonal, we apply Lemma 7 in the preceding inequality, and obtain

$$\mathbb{P}_\theta^\pi\big(\|(\mathbf{I}_n \otimes \mathcal{J}_t^{-1})\,\mathcal{M}_t\| > \delta,\ J_t \geq m,\ A\big) \ \leq\ \mathbb{P}_\theta^\pi\Big(\widetilde{\mathcal{Z}}_t \geq \exp\Big\{\frac{\gamma\,\delta^2 m}{2\zeta^*\sigma^2}\Big\},\ J_t \geq m,\ A\Big).$$

Using the concavity of the function $L(y) := y \cdot \mathcal{M}_t - \frac{1}{2}y^{\mathsf{T}}(\mathbf{I}_n \otimes \mathcal{J}_t)y$ as in the proof of Lemma 3, we further deduce that the following is true on the event A: $\mathcal{Z}_t^y \geq e^{-(\xi+n+n^2u^2)/(\zeta^*\sigma^2)}\widetilde{\mathcal{Z}}_t$ for all y in the $(1/t)$-neighborhood of w_t. Fixing a set $\mathcal{S}$ of $\lceil \pi\delta t\rceil^{n^2+n-1}$ distinct points on the sphere $\{y \in \mathbb{R}^{n^2+n} : \|y\| = \delta\}$, we make sure that $\mathcal{Z}_t^y \geq e^{-(\xi+n+n^2u^2)/(\zeta^*\sigma^2)}\widetilde{\mathcal{Z}}_t$ for some $y \in \mathcal{S}$. Thus we have

$$\mathbb{P}_\theta^\pi\big(\|(\mathbf{I}_n \otimes \mathcal{J}_t^{-1})\,\mathcal{M}_t\| > \delta,\ J_t \geq m,\ A\big) \ \leq\ \mathbb{P}_\theta^\pi\big(\mathcal{Z}_t^y \geq K_1 \exp(\rho_0\,\delta^2 m) \text{ for some } y \in \mathcal{S}\big),$$

where $K_1 = e^{-(\xi+n+n^2u^2)/(\zeta^*\sigma^2)}$ and $\rho_0 = \gamma/(2\zeta^*\sigma^2)$. Using the union bound on the right hand side of preceding inequality, we deduce that

$$\mathbb{P}_\theta^\pi\big(\|(\mathbf{I}_n \otimes \mathcal{J}_t^{-1})\,\mathcal{M}_t\| > \delta,\ J_t \geq m,\ A\big) \ \leq \sum_{\widetilde{y}_j \in \mathcal{S}} \mathbb{P}_\theta^\pi\big(\mathcal{Z}_t^{\widetilde{y}_j} \geq K_1 \exp(\rho\,\delta^2 m)\big).$$

Applying Markov's inequality on each term in the preceding sum, and noting that $\mathbb{E}_\theta^\pi[\mathcal{Z}_t^{\widetilde{y}_j}] \leq \mathcal{Z}_0^{\widetilde{y}_j} = 1$ for all $\widetilde{y}_j \in \mathcal{S}$, we get

$$\mathbb{P}_\theta^\pi\big(\|(\mathbf{I}_n \otimes \mathcal{J}_t^{-1})\,\mathcal{M}_t\| > \delta,\ J_t \geq m,\ A\big) \leq \tfrac{1+K_2 t^{n^2+n-1}}{K_1}\ \exp(-\rho_0\,\delta^2 m),$$

where $K_2 = \big(\pi(1 \vee \delta)\big)^{n^2+n-1}$. Moreover, by the argument used to obtain inequality (A.9), we further get

$$\begin{aligned}
\mathbb{P}_\theta^\pi\big(J_t \geq m,\ A^c\big) &= \mathbb{P}_\theta^\pi\big(J_t \geq m,\ \|\mathcal{M}_t\| > \xi t\big) \\
&\leq \sum_{i=1}^n \mathbb{P}_\theta^\pi\bigg(J_t \geq m,\ \Big|\sum_{s=1}^t \epsilon_{is}\Big| > \frac{\xi t}{\sqrt{n^2+n}}\bigg) \ +\ \sum_{i,j=1}^n \mathbb{P}_\theta^\pi\bigg(J_t \geq m,\ \Big|\sum_{s=1}^t \epsilon_{is}p_{js}\Big| > \frac{\xi t}{\sqrt{n^2+n}}\bigg) \\
&\leq 2n \exp\bigg(-\frac{\xi^2 m}{2(n^2+n)\zeta^*\sigma^2(u-\ell)^2}\bigg) \ +\ 2n^2 \exp\bigg(-\frac{\xi^2 m}{2(n^2+n)\zeta^*\sigma^2(u-\ell)^2u^2}\bigg).
\end{aligned}$$

Finally, we choose $\xi = (1\vee\delta)\sigma(u-\ell)\big(2\zeta^*(n^2+n)(1+u^2)\rho_0\big)^{1/2} \geq \delta\sigma(u-\ell)\big(2\zeta^*(n^2+n)(1+u^2)\rho_0\big)^{1/2}$ to obtain $\mathbb{P}_\theta^\pi\big(J_t \geq m,\ A^c\big) \leq 2(n^2+n)\exp(-\rho_0\,\delta^2 m)$. Now, we can construct the desired upper bound on the right hand side of (B.7) by letting $\rho = (1 \vee \delta)\rho_0 = \frac{1}{2}\gamma\sigma^{-2}\big(\nu^* \vee (z^*/z_0)\big)^{-1}$ and $k = 2(n^2+n) + (1+K_2)/K_1$. ■

Proof of Theorem 6. This proof is almost identical to the proof of Theorem 2. Replacing the absolute value in that proof with Euclidean norm, and using exactly the same arguments, we arrive at the following analog of inequality (A.11):

$$\mathbb{E}_\theta^\pi \big\|\varphi(\theta) - p_{t+1}\big\|^2 \leq \frac{2\widetilde{K}_0(n^3+n^2)(n^2-n+1)\log t}{\rho\kappa_0\sqrt{t}} + 2\,\mathbb{E}_\theta^\pi \big\|\varphi(\vartheta_{\lfloor t/n\rfloor}) - p_{t+1}\big\|^2 \quad \text{for all } t \geq N_0\,, \tag{B.8}$$

where N_0 is the smallest natural number satisfying $kN_0^{n^2-n+1}\exp(-\frac{1}{2}\rho\kappa_0\sqrt{N_0}) \leq 1$, and $\widetilde{K}_0 = (1+5k)\max_{i,j,\theta}\big\{(\partial\varphi_i(\theta)/\partial\theta_j)^2\big\}$. We let $N = \max\{N_0, n+1\}$, sum the preceding inequality over $t = N, \ldots, T-1$, and invoke condition (ii) to get

$$\sum_{t=N}^{T-1} \mathbb{E}_\theta^\pi \big\|\varphi(\theta) - p_{t+1}\big\|^2 \;\leq\; \frac{8\widetilde{K}_0 n^5}{\rho\kappa_0}\sqrt{T}\log T + 2\kappa_1\sqrt{T} \;\leq\; \left(\frac{8\widetilde{K}_0 n^5}{\rho\kappa_0} + 2\kappa_1\right)\sqrt{T}\log T.$$

Recalling that the eigenvalues of the matrix B are in the interval $[b_{\min}, b_{\max}]$, we get

$$\begin{aligned}\Delta^\pi(T) &= \sup_{\theta\in\Theta}\Big\{-\sum_{t=0}^{T-1}\mathbb{E}_\theta^\pi\big(\varphi(\theta) - p_{t+1}\big)^{\mathsf{T}} B\big(\varphi(\theta) - p_{t+1}\big)\Big\} \\ &\leq |b_{\min}| \sup_{\theta\in\Theta}\sum_{t=0}^{T-1}\mathbb{E}_\theta^\pi\big\|\varphi(\theta) - p_{t+1}\big\|^2.\end{aligned}$$

Thus we have $\Delta^\pi(T) \leq C_0 + C_1\sqrt{T}\log T$, where $C_0 = n\,|b_{\min}|\,(u-\ell)^2 N$ and $C_1 = |b_{\min}|\big(8\widetilde{K}_0 n^5\rho^{-1}\kappa_0^{-1} + 2\kappa_1\big)$. We conclude the proof by choosing $C = \max\{C_0, C_1\}$. ■

Remark As argued in the remark following the proof of Theorem 2, we can minimize C_1 with respect to κ_0 and κ_1 under the assumption that equal emphasis will be given to learning and earning. The optimal choice is $\kappa_0 = \kappa_1 = (4\widetilde{K}_0 n^5/\rho)^{1/2}$, and the corresponding minimal value is $C_1 = 8|b_{\min}|\big(\widetilde{K}_0 n^5/\rho\big)^{1/2}$.

B.2 Incumbent-Price Problem with Multiple Products

Proof of Lemma 9. As in the proofs of our previous information-based inequalities, we will first invoke the van Trees inequality (Gill and Levit 1995). Choose λ to be an absolutely continuous density on $\mathfrak{B}$, taking positive values on the interior of $\mathfrak{B}$ and zero on its boundary. Let $\mathcal{C}(b) = (B + B^{\mathsf{T}})e_1 b^{\mathsf{T}}$ for all $b \in \mathfrak{B}$, where $b = \mathrm{vec}(B^{\mathsf{T}}) = (\beta_1, \ldots, \beta_n)$, and e_1 is the first unit vector in the standard basis of $\mathbb{R}^n$. Then, the multivariate van Trees inequality (Gill and Levit 1995) implies that

$$\mathbb{E}_\lambda\big\{\mathbb{E}_b^\pi\big\|x_t - \psi(b)\big\|^2\big\} \geq \frac{\Big(\mathbb{E}_\lambda\big[\,\mathsf{tr}\big\{\mathcal{C}(b)\,(\partial\psi/\partial b)^{\mathsf{T}}\big\}\,\big]\Big)^2}{\mathbb{E}_\lambda\big[\,\mathsf{tr}\big\{\mathcal{C}(b)\,\widehat{\mathcal{I}}_{t-1}^\pi(b)\,\mathcal{C}(b)^{\mathsf{T}}\big\}\,\big] + \widetilde{\mathcal{I}}(\lambda)}, \tag{B.9}$$

where $\widetilde{\mathcal{I}}(\lambda)$ is the Fisher information for the density λ, $\mathbb{E}_\lambda(\cdot)$ is the expectation operator with respect to λ, and $\mathbb{E}_b^\pi(\cdot)$ is the expectation operator given that parameter vector is b and the seller uses policy π. To obtain a lower bound on the right side of (B.9), we compute $\mathsf{tr}\{\mathcal{C}(b)\,(\partial\psi/\partial b)^\mathsf{T}\}$ as follows. Invoking the implicit function theorem on the first-order optimality condition $\widehat{\eta}(b,x) \;:=\; \widehat{D} + B^\mathsf{T}\widehat{p} + (B+B^\mathsf{T})\,x \;=\; 0$, we get

$$\frac{\partial\psi}{\partial b} \;=\; -\left(\frac{\partial\widehat{\eta}}{\partial x}\right)^{-1}\frac{\partial\widehat{\eta}}{\partial b} \;=\; -\left(B+B^\mathsf{T}\right)^{-1}\frac{\partial\widehat{\eta}}{\partial b}.$$

Therefore, we have $\mathsf{tr}\{\mathcal{C}(b)\,(\partial\psi/\partial b)^\mathsf{T}\} = -\mathsf{tr}\{\mathcal{C}(b)\,(\partial\widehat{\eta}/\partial b)^\mathsf{T}(B+B^\mathsf{T})^{-1}\}$ because $B+B^\mathsf{T}$ has a symmetric inverse. Moreover, using the fact that $(\partial\widehat{\eta}/\partial b)^\mathsf{T} = \big(\widehat{p}+\psi(b)\big)^\mathsf{T}\otimes\mathbf{I}_n \;+\; \mathbf{I}_n\otimes\psi(b)$, we deduce that $\mathcal{C}(b)\,(\partial\widehat{\eta}/\partial b)^\mathsf{T} = (B+B^\mathsf{T})\,e_1\big[\widehat{\eta}\big(b,\psi(b)\big) - \widehat{D}\big]^\mathsf{T}$. By the first-order condition of optimality, we know that $\widehat{\eta}\big(b,\psi(b)\big) = 0$, which implies that $\mathcal{C}(b)\,(\partial\widehat{\eta}/\partial\theta)^\mathsf{T} = -(B+B^\mathsf{T})\,e_1\widehat{D}^\mathsf{T}$. So, we get

$$\mathsf{tr}\Big\{\mathcal{C}(b)\left(\frac{\partial\psi}{\partial b}\right)^\mathsf{T}\Big\} \;=\; \mathsf{tr}\big\{(B+B^\mathsf{T})\,e_1\widehat{D}^\mathsf{T}(B+B^\mathsf{T})^{-1}\big\}.$$

Since the trace of a matrix is similarity-invariant, we conclude that $\mathsf{tr}\{\mathcal{C}(b)\,(\partial\psi/\partial b)^\mathsf{T}\} = \mathsf{tr}\{e_1\widehat{D}^\mathsf{T}\} = \widehat{D}_1$. Hence, the numerator on the right side of inequality (B.9) equals $\widehat{D}_1^2 > 0$. To find a finite upper bound on the denominator, we note that

$$\begin{aligned}\mathsf{tr}\{\mathcal{C}(b)\,\widehat{\mathcal{I}}^\pi_{t-1}(b)\,\mathcal{C}(b)^\mathsf{T}\} &= \frac{1}{\sigma^2}\,\mathsf{tr}\Big\{\,\mathbb{E}_b^\pi\big[(B+B^\mathsf{T})e_1b^\mathsf{T}(\mathbf{I}_n\otimes\mathcal{J}_{t-1})be_1^\mathsf{T}(B+B^\mathsf{T})\big]\Big\}\\ &= \frac{1}{\sigma^2}\,\mathbb{E}_b^\pi\big[b^\mathsf{T}(\mathbf{I}_n\otimes\widehat{\mathcal{J}}_{t-1})b\big]\;\mathsf{tr}\big\{(B+B^\mathsf{T})e_1e_1^\mathsf{T}(B+B^\mathsf{T})\big\},\end{aligned}$$

because $b^\mathsf{T}(\mathbf{I}_n\otimes\widehat{\mathcal{J}}_{t-1})b$ is a scalar. Moreover, we get by elementary algebra that $\mathsf{tr}\{(B+B^\mathsf{T})e_1e_1^\mathsf{T}(B+B^\mathsf{T})\} = \sum_{i=1}^n(\beta_{1i}+\beta_{i1})^2$. Since $\mathfrak{B}$ is a compact set, there exists a finite positive constant $\beta_{\max}$ such that $|\beta_{ij}| \le \beta_{\max}$ for all i and j. This fact implies that $\mathsf{tr}\{(B+B^\mathsf{T})e_1e_1^\mathsf{T}(B+B^\mathsf{T})\} \le 4n\beta_{\max}^2$. Another consequence of the same fact is that $b^\mathsf{T}(\mathbf{I}_n\otimes\widehat{\mathcal{J}}_{t-1})b = \sum_{s=1}^{t-1}\sum_{i=1}^n\mathbb{E}_b^\pi(\beta_i^\mathsf{T}x_s)^2 \le \beta_{\max}^2\sum_{s=1}^{t-1}\sum_{i=1}^n\mathbb{E}_b^\pi(\mathbf{1}_n^\mathsf{T}x_s)^2 = n\beta_{\max}^2\mathbf{1}_n^\mathsf{T}\widehat{\mathcal{J}}_{t-1}\mathbf{1}_n$. As a result, we have

$$\mathsf{tr}\{\mathcal{C}(b)\,\widehat{\mathcal{I}}^\pi_{t-1}(b)\,\mathcal{C}(b)^\mathsf{T}\} \;\le\; \frac{4n^2\beta_{\max}^4}{\sigma^2}\,\mathbb{E}_b^\pi\big[\mathbf{1}_n^\mathsf{T}\widehat{\mathcal{J}}_{t-1}\mathbf{1}_n\big]\,.$$

Thus inequality (B.9) implies that

$$\mathbb{E}_\lambda\big\{\mathbb{E}_b^\pi\big\|x_t-\psi(\theta)\big\|^2\big\} \ge \frac{c_0}{c_1+\mathbb{E}_\lambda\big[\,\mathbb{E}_b^\pi\big[\mathbf{1}_n^\mathsf{T}\widehat{\mathcal{J}}_{t-1}\mathbf{1}_n\big]\big\}\,\big]}\,,$$

where $c_0 = (\sigma^2\widehat{D}_1^2)/(4n^2\beta_{\max}^4)$ and $c_1 = \big(\sigma^2\widetilde{\mathcal{I}}(\lambda)\big)/(4n^2\beta_{\max}^4)$. We conclude the proof by summing up the preceding inequality over $2,\ldots,T$, and using the monotonicity of $\mathbb{E}_\lambda$. ■

Proof of Theorem 7. In what follows, we will essentially repeat the proof of Theorem 3 with slight modification. Using Lemma 4, we deduce that

$$\sup_{b\in\mathfrak{B},\,\psi(b)\neq\mathbf{0}_n}\Big\{\sum_{t=2}^T\mathbb{E}_b^\pi\big\|x_t-\psi(b)\big\|^2\Big\} \;\ge\; \sum_{t=2}^T\frac{c_0}{c_1+n^2(u-\ell)^2(t-1)}\,,$$

because $\mathbf{1}_n^\mathsf{T} \widehat{\mathcal{J}}_s \mathbf{1}_n \le n^2(u-\ell)^2 s$ for all s. Moreover, recalling the fact that the eigenvalues of B are in $[b_{\min},\, b_{\max}]$, we further get

$$\begin{aligned}\Delta^\pi(T) &= \sup_{b\in\mathfrak{B},\,\psi(b)\neq\mathbf{0}_n} \Big\{ -\sum_{t=1}^{T} \mathbb{E}_b^\pi\big(x_t-\psi(b)\big)^\mathsf{T} B\big(x_t-\psi(b)\big)\Big\} \\ &\ge |b_{\max}| \sup_{b\in\mathfrak{B},\,\psi(b)\neq\mathbf{0}_n} \sum_{t=2}^{T} \frac{c_0}{c_1+n^2(u-\ell)^2(t-1)}\,.\end{aligned}$$

Letting $K_1 = |b_{\max}|c_0/\big(n(u-\ell)\big)^2$, $K_2 = c_1/\big(n(u-\ell)\big)^2$, we obtain

$$\Delta^\pi(T) \;\ge\; K_1 \sum_{t=1}^{T-1} \frac{1}{K_2+t} \;\ge\; c\sum_{t=1}^{T-1}\frac{1}{t} \;\ge\; c\log T\,,$$

where $c = K_1/(1+K_2)$. This concludes the proof. ■

Proof of Lemma 10. Because the unique eigenvalue of $\mathbf{I}_n$ is 1, $\widehat{\mu}_{\min}(t)$ is also the smallest eigenvalue of $\widehat{\mathcal{J}}_t$. Take an arbitrary unit vector $y \in \mathbb{R}^n$. Then,

$$y^\mathsf{T}\widehat{\mathcal{J}}_t y \;=\; \sum_{s=1}^{t}(y\cdot x_s)^2.$$

The orthogonal pricing condition implies that for each $\tau = 1,2,\ldots$ there exist $X_\tau \in \mathbb{R}^n$ and $v_\tau > 0$ such that $\widehat{V}_\tau := \big\{v_\tau^{-1}\big(x_s - X_\tau\big) \,:\, s = n(\tau-1)+1,\,\ldots,\,n(\tau-1)+n\big\}$ forms an orthonormal basis of $\mathbb{R}^n$, and $e\cdot X_\tau = n^{-1/2}v_\tau\|X_\tau\|$ for all $e\in V_\tau$. Thus,

$$\begin{aligned}\sum_{s=1}^{t}(y\cdot x_s)^2 \;&=\; \sum_{s=1}^{t}\big(y\cdot X_{\lceil s/n\rceil} \;+\; y\cdot(x_s - X_{\lceil s/n\rceil})\big)^2 \\ &=\; \sum_{\tau=1}^{\lfloor t/n\rfloor}\sum_{i=1}^{n}\big(y\cdot X_\tau \;+\; y\cdot(x_{n(\tau-1)+i}-X_\tau)\big)^2\,.\end{aligned}$$

Let $e_{i\tau} = v_\tau^{-1}\big(x_{n(\tau-1)+i} - X_\tau\big)$, and $y_{i\tau} = y\cdot e_{i\tau}$. Because $e_{i\tau}\cdot X_\tau = n^{-1/2}v_\tau\|X_\tau\|$ for all $i=1,\ldots,n$, the preceding inequality implies

$$\begin{aligned}\sum_{s=1}^{t}(y\cdot x_s)^2 \;&=\; \sum_{\tau=1}^{\lfloor t/n\rfloor}\sum_{i=1}^{n}\bigg(n^{-1/2}\|X_\tau\|\sum_{j=1}^{n}y_{j\tau} \;+\; v_\tau y_{i\tau}\bigg)^2 \\ &=\; \sum_{\tau=1}^{\lfloor t/n\rfloor}\sum_{i=1}^{n}\bigg(n^{-1}\|X_\tau\|^2\Big(\sum_{j=1}^{n}y_{j\tau}\Big)^2 \;+\; 2n^{-1/2}\|X_\tau\|\Big(\sum_{j=1}^{n}y_{j\tau}\Big)v_\tau y_{i\tau} \;+\; v_\tau^2 y_{i\tau}^2\bigg) \\ &=\; \sum_{\tau=1}^{\lfloor t/n\rfloor} v_\tau^2\sum_{i=1}^{n}y_{i\tau}^2 \;=\; \|y\|^2\sum_{\tau=1}^{\lfloor t/n\rfloor}v_\tau^2\,.\end{aligned}$$

Combining the preceding inequality with the definition of $\widehat{J}_t$ and the fact that $\|y\| = 1$, we deduce that $y^{\mathsf{T}} \widehat{\mathcal{J}}_t y \geq \widehat{J}_t$. Recalling that the unit vector y was selected arbitrarily, we conclude by the Rayleigh-Ritz theorem that $\widehat{\mu}_{\min}(t) \geq \widehat{J}_t$. This concludes the proof. ■

Proof of Lemma 11. Our approach in this proof is similar to that in the proofs of Lemmas 3, 5, and 8. We will first construct an exponential supermartingale, and then use Markov's inequality to characterize the estimation errors. Note the following for the least squares estimate $\widehat{b}_t$:

$$\widehat{b}_t - b = (\mathbf{I}_n \otimes \widehat{\mathcal{J}}_t^{-1})\, \widehat{\mathcal{M}}_t \,, \tag{B.10}$$

where $\widehat{\mathcal{M}}_t$ is the $n^2 \times 1$ vector $\widehat{\mathcal{M}}_t := \sum_{s=1}^{t} \epsilon_s \otimes x_s$. Given an $n^2 \times 1$ vector y with $\|y\| = \delta$, let $\{\widehat{\mathcal{Z}}_s^y,\, s = 1, 2, \ldots\}$ be a stochastic process such that $\widehat{\mathcal{Z}}_0^y = 1$ and

$$\widehat{\mathcal{Z}}_s^y := \exp\left\{ \frac{1}{\zeta^* \sigma^2} \Big(y \cdot \widehat{\mathcal{M}}_s - \frac{1}{2} y^{\mathsf{T}} (\mathbf{I}_n \otimes \widehat{\mathcal{J}}_s) y \Big) \right\} \quad \text{for all } s = 1, 2, \ldots \,, \tag{B.11}$$

where $\zeta^* = (1 \vee \delta)\big(\nu^* \vee (z^*/z_0)\big)$, $z^* = \max_{y \in \mathbb{R}^n,\, \|y\| \leq 1,\, \widehat{p}+x \in [\ell,u]^n} \left\{ |y \cdot x| / \sigma^2 \right\}$, $\nu^* = \max_{|z| \leq z_0} \{\nu(z)\}$, and $\nu(z) = \sum_{k=2}^{\infty} \big(2 z^{k-2} \mathbb{E}_b^{\pi}[\epsilon_{it}^k] \big) / (k! \sigma^2)$. Denoting $\mathcal{F}_s := \sigma(D_1, \ldots, D_s)$, we deduce by tower property the following:

$$\mathbb{E}_b^{\pi}[\widehat{\mathcal{Z}}_s^y | \mathcal{F}_{s-1}] \;=\; \exp\left\{ \frac{1}{\zeta^* \sigma^2} \Big(y \cdot \widehat{\mathcal{M}}_{s-1} - \frac{1}{2} y^{\mathsf{T}} (\mathbf{I}_n \otimes \widehat{\mathcal{J}}_s) y \Big) \right\} \mathbb{E}_b^{\pi}\Big[\exp\left\{ \frac{1}{\zeta^* \sigma^2}\, y \cdot (\widehat{\mathcal{M}}_s - \widehat{\mathcal{M}}_{s-1}) \right\} \Big| \mathcal{F}_{s-1} \Big],$$

where $\mathbb{E}_b^{\pi}(\cdot)$ is the expectation operator given that parameter vector is b and the seller uses policy π. Let $y^i = \big(y_{n(i-1)+1}, \ldots, y_{n(i-1)+n} \big) \in \mathbb{R}^n$. Using inequality (A.5), the independence of the ϵ_{it}, and the fact that $|y^i \cdot x_s| / (\zeta^* \sigma^2) \leq z_0 |y^i \cdot x_s| / (\delta z^* \sigma^2) \leq z_0$ for all $i = 1, \ldots, n$, and $\widehat{p} + x_s \in [\ell, u]^n$ as in the argument used to derive inequality (B.6), we compute an upper bound on the conditional expectation $\mathbb{E}_b^{\pi}[\exp\{\frac{1}{\zeta^* \sigma^2}\, y \cdot (\widehat{\mathcal{M}}_s - \widehat{\mathcal{M}}_{s-1})\} | \mathcal{F}_{s-1}]$, and get

$$\mathbb{E}_b^{\pi}[\widehat{\mathcal{Z}}_s^y | \mathcal{F}_{s-1}] \;\leq\; \exp\left\{ \frac{1}{\zeta^* \sigma^2} \Big(y \cdot \widehat{\mathcal{M}}_{s-1} - \frac{1}{2} y^{\mathsf{T}} (\mathbf{I}_n \otimes \widehat{\mathcal{J}}_{s-1}) y \Big) \right\} \;=\; \widehat{\mathcal{Z}}_{s-1}^y \,,$$

because $(\mathbf{I}_n \otimes x_s)(\mathbf{I}_n \otimes x_s^{\mathsf{T}}) = \mathbf{I}_n \otimes x_s x_s^{\mathsf{T}}$ by virtue of the mixed product property. Having shown that $(\widehat{\mathcal{Z}}_s^y, \mathcal{F}_s)$ is a supermartingale, we define the stochastic process $\{\widetilde{\mathcal{Z}}_s\}$ satisfying $\widetilde{\mathcal{Z}}_s = \widehat{\mathcal{Z}}_s^{w_s}$, where $w_s = \delta (\mathbf{I}_n \otimes \widehat{\mathcal{J}}_s^{-1})\, \widehat{\mathcal{M}}_s / \|(\mathbf{I}_n \otimes \widehat{\mathcal{J}}_s^{-1})\, \widehat{\mathcal{M}}_s\|$. Letting $A := \{\|\widehat{\mathcal{M}}_t\| \leq \xi t\}$, and recalling the argument we employed for deriving inequalities (A.7) and (B.7), we get

$$\mathbb{P}_b^{\pi}\big(\|\widehat{b}_t - b\| > \delta,\; \widehat{J}_t \geq m \big) \;\leq\; \mathbb{P}_b^{\pi}\big(\|(\mathbf{I}_n \otimes \widehat{\mathcal{J}}_t^{-1})\, \widehat{\mathcal{M}}_t\| > \delta,\; \widehat{J}_t \geq m,\; A \big) \;+\; \mathbb{P}_b^{\pi}\big(\widehat{J}_t \geq m,\; A^c \big)\,, \tag{B.12}$$

where $\mathbb{P}_b^{\pi}(\cdot)$ is the probability measure given that parameter vector is b and the seller uses policy π. Using the definition of $\widetilde{\mathcal{Z}}_t$, the Rayleigh-Ritz theorem, and Lemma 10, we deduce that

$$\mathbb{P}_b^{\pi}\big(\|(\mathbf{I}_n \otimes \widehat{\mathcal{J}}_t^{-1})\, \widehat{\mathcal{M}}_t\| > \delta,\; \widehat{J}_t \geq m,\; A \big) \;\leq\; \mathbb{P}_b^{\pi}\Big(\widetilde{\mathcal{Z}}_t \geq \exp\left\{ \frac{\gamma\, \delta^2 m}{2 \zeta^* \sigma^2} \right\},\; \widehat{J}_t \geq m,\; A \Big).$$

As argued in the proof of Lemma 8, we fix a set $\mathcal{S}$ of $\lceil \pi\delta t \rceil^{n^2-1}$ distinct points on the sphere $\{y \in \mathbb{R}^{n^2} : \|y\| = \delta\}$ to make sure that $\widehat{\mathcal{Z}}_t^y \geq e^{-(\xi+n^2(u-\ell)^2)/(\zeta^*\sigma^2)}\widetilde{\mathcal{Z}}_t$ for some $y \in \mathcal{S}$, and obtain the following by the union bound:

$$\mathbb{P}_b^\pi\big(\|(\mathbf{I}_n \otimes \widehat{\mathcal{J}}_t^{-1})\,\widehat{\mathcal{M}}_t\| > \delta,\ \widehat{J}_t \geq m,\ A\big) \ \leq \ \sum_{\widetilde{y}_j \in \mathcal{S}} \mathbb{P}_b^\pi\big(\mathcal{Z}_t^{\widetilde{y}_j} \geq K_1 \exp(\rho_0 \delta^2 m)\big)\,.$$

where $K_1 = e^{-(\xi+n^2(u-\ell)^2)/(\zeta^*\sigma^2)}$ and $\rho_0 = \gamma/(2\zeta^*\sigma^2)$. By Markov's inequality, and the fact that $\mathbb{E}_b^\pi[\widehat{\mathcal{Z}}_t^{\widetilde{y}_j}] \leq \widehat{\mathcal{Z}}_0^{\widetilde{y}_j} = 1$ for all $\widetilde{y}_j \in \mathcal{S}$, this further implies

$$\mathbb{P}_b^\pi\big(\|(\mathbf{I}_n \otimes \widehat{\mathcal{J}}_t^{-1})\,\widehat{\mathcal{M}}_t\| > \delta,\ \widehat{J}_t \geq m,\ A\big) \leq \tfrac{1+K_2 t^{n^2-1}}{K_1}\exp(-\rho_0\delta^2 m)\,,$$

where $K_2 = \big(\pi(1 \vee \delta)\big)^{n^2-1}$. So, we have found an upper bound on the first term on the right hand side of (B.12). To bound the second term, let $\xi = (1\vee\delta)\sigma(u-\ell)^2 n(2\zeta^*\rho_0)^{1/2} \geq \delta\sigma(u-\ell)^2 n(2\zeta^*\rho_0)^{1/2}$, and get $\mathbb{P}_b^\pi\big(\widehat{J}_t \geq m,\ A^c\big) \leq 2n^2\exp(-\rho_0\,\delta^2 m)$ as argued in the proof of Lemma 8. Choosing $\rho = (1 \vee \delta)\rho_0 = \frac{1}{2}\gamma\sigma^{-2}\big(\nu^* \vee (z^*/z_0)\big)^{-1}$ and $k = 2n^2 + (1+K_2)/K_1$ concludes the proof. ■

Proof of Theorem 8. This proof is a generalization of the proof of Theorem 4 to the case of multiple dimensions. Recalling that the two sources of revenue loss of an ILS variant are estimation errors and perturbations from the greedy ILS price, the conditional expectation of the squared pricing errors can be bounded from above as follows:

$$\mathbb{E}_b^\pi\big\|\psi(b) - x_{t+1}\big\|^2 \leq 2\mathbb{E}_b^\pi\big\|\psi(b) - \psi(b_t)\big\|^2 + 2\mathbb{E}_b^\pi\big\|\psi(b_t) - x_{t+1}\big\|^2\,, \tag{B.13}$$

where b_t denotes the truncated least squares estimate in period t, and $\mathbb{E}_b^\pi(\cdot)$ is the expectation operator given that parameter vector is b and the seller uses policy π. To analyze the first term on the right-hand side of this inequality, we define the sequence of events $\{A_t\}$ such that

$$A_t \ = \ \bigcap_{\tau=\lfloor \frac{t-n}{2n} \rfloor}^{\lfloor \frac{t-n}{n} \rfloor} \ \bigcap_{i=1}^{n} \ \left\{\big\|\psi(b) - \psi(b_{n\tau+i})\big\|^2 \leq \frac{\|\psi(b)\|^2}{12}\right\} \qquad \text{for all } t \geq n, \tag{B.14}$$

where n is the number of products. We know by the triangle inequality that

$$\|x_{n\tau+i}\|^2 \ \geq \ \frac{\|\psi(b)\|^2}{3} \ - \ \|\psi(b) - \psi(b_{n\tau+i})\|^2 \ - \ \|\psi(b_{n\tau+i}) - x_{n\tau+i+1}\|^2$$

for all τ and i. By the definition of A_t, we further deduce that the following holds on A_t:

$$\begin{aligned}\|x_{n\tau+i}\|^2 &\geq \frac{\|\psi(b)\|^2}{3} \ - \ \frac{\|\psi(b)\|^2}{12} \ - \ \|\psi(b_{n\tau+i}) - x_{n\tau+i+1}\|^2 \\ &= \frac{\|\psi(b)\|^2}{4} \ - \ \|\psi(b_{n\tau+i}) - x_{n\tau+i+1}\|^2 \\ &\geq \frac{\|\psi(b)\|^2}{4} \ - \ \sum_{i=1}^{n}\|\psi(b_{n\tau+i}) - x_{n\tau+i+1}\|^2\end{aligned}$$

for $\tau = \lfloor (t-n)/(2n) \rfloor, \ldots, \lfloor (t-n)/n \rfloor$ and $i = 1, \ldots, n$. Combining the preceding inequality with the definition of $\widehat{J}_t$, which is given in (4.12), we further get

$$\begin{aligned}
\widehat{J}_t &\geq \sum_{\tau=\lfloor \frac{t-n}{2n} \rfloor}^{\lfloor \frac{t-n}{n} \rfloor} \min_{1 \leq i \leq n} \{\|x_{n\tau+i}\|^2\} \\
&\geq \sum_{\tau=\lfloor \frac{t-n}{2n} \rfloor}^{\lfloor \frac{t-n}{n} \rfloor} \frac{\|\psi(b)\|^2}{4} - \sum_{\tau=\lfloor \frac{t-n}{2n} \rfloor}^{\lfloor \frac{t-n}{n} \rfloor} \sum_{i=1}^{n} \|\psi(b_{n\tau+i}) - x_{n\tau+i+1}\|^2 \\
&\geq \sum_{\tau=\lfloor \frac{t-n}{2n} \rfloor}^{\lfloor \frac{t-n}{n} \rfloor} \frac{\|\psi(b)\|^2}{4} - \sum_{s=0}^{t} \left\|\psi(b_s) - x_{s+1}\right\|^2. \qquad \text{(B.15)}
\end{aligned}$$

Using condition (ii) in the hypothesis, we can bound the second sum on the right side of (B.15), and conclude that

$$\widehat{J}_t \geq \left(\frac{\|\psi(b)\|^2}{8n} - \frac{\kappa_1 \log t}{t-n} \right) (t-n) \quad \text{on } A_t\,,$$

where $t \geq n$. The sequence $(s-n)^{-1} \log s$ converges to zero as s tends to ∞. So, we know that there is a natural number $N_0(b)$ large enough that for all $t \geq N_0(b)$ we have $\widehat{J}_t \geq \|\psi(b)\|^2 (t-n)/16$ on A_t, implying that the growth rate of $\widehat{J}_t$ is linear on A_t. To show that the events A_t occur "frequently," we will construct an upper bound on $\mathbb{P}_b^\pi(A_t^c)$ as in the proof of Theorem 4. First,

$$\mathbb{P}_b^\pi(A_t^c) \;\leq\; \sum_{\tau=\lfloor \frac{t-n}{2n} \rfloor}^{\lfloor \frac{t-n}{n} \rfloor} \sum_{i=1}^{n} \mathbb{P}_b^\pi \left\{ \left\|\psi(b) - \psi(b_{n\tau+i})\right\|^2 > \frac{\|\psi(b)\|^2}{12} \right\},$$

where $\mathbb{P}_b^\pi(\cdot)$ is the probability measure given that parameter vector is b and the seller uses policy π. Since $n\lfloor (t-n)/(2n) \rfloor + 1 \geq \lfloor (t-n)/(2n) \rfloor$, and $n\lfloor (t-n)/n \rfloor + n \leq t$ for all t, the preceding inequality leads to

$$\mathbb{P}_b^\pi(A_t^c) \;\leq\; \sum_{s=\lfloor \frac{t-n}{2n} \rfloor}^{t} \mathbb{P}_b^\pi \left\{ \left\|\psi(b) - \psi(b_s)\right\|^2 > \frac{\|\psi(b)\|^2}{12} \right\}.$$

As in the proof of Theorem 4, we use the mean value theorem to obtain

$$\mathbb{P}_b^\pi(A_t^c) \leq \sum_{s=\lfloor \frac{t-n}{2n} \rfloor}^{t} \mathbb{P}_b^\pi \left\{ \sqrt{n^3 K_0} \left\|b - b_s\right\| > \frac{\|\psi(b)\|}{\sqrt{12}} \right\},$$

where $K_0 = \max_{i,\,j,\,b} \left\{ (\partial \psi_i(b)/\partial b_j)^2 \right\}$. Invoking condition (i) in the hypothesis, and recalling the fact that $\left\|b - \widehat{b}_s\right\| \geq \left\|b - b_s\right\|$, we get

$$\mathbb{P}_b^\pi(A_t^c) \leq \sum_{s=\lfloor \frac{t-n}{2n} \rfloor}^{t} \mathbb{P}_b^\pi \left\{ \left\|b - \widehat{b}_s\right\| > \frac{\|\psi(b)\|}{\sqrt{12 n^3 K_0}},\; \widehat{J}_s \geq \kappa_0 \log s \right\}.$$

By Lemma 11, this implies that

$$\mathbb{P}^{\pi}_{b}(A^c_t) \;\le\; \sum_{s=\lfloor\frac{t-n}{2n}\rfloor}^{t} ks^{n^2-1}\exp\left(-\rho\kappa_0\left(\frac{\|\psi(b)\|}{\sqrt{12n^3K_0}}\wedge\frac{\|\psi(b)\|^2}{12n^3K_0}\right)\log s\right) \;\le\; \sum_{s=\lfloor\frac{t-n}{2n}\rfloor}^{t} ks^{-q}\,,$$

where $q := \frac{1}{12n^3}\rho\kappa_0\{\frac{\|\psi(b)\|}{\sqrt{K_0}}\wedge\frac{\|\psi(b)\|^2}{K_0}\} - n^2 + 1$. As in the univariate case, we pick κ_0 greater than $12(n^5+2n^3)\rho^{-1}\{\frac{\|\psi(b)\|}{\sqrt{K_0}}\wedge\frac{\|\psi(b)\|^2}{K_0}\}^{-1}$ to ensure that $q \ge 3$, and consequently get $\sum_{t=5n}^{\infty}\mathbb{P}^{\pi}_{b}(A^c_t) \le k\pi^2/3$.

Finally we turn our attention to our initial upper bound on squared pricing errors, which is displayed in (B.13). Letting $N(b) = \max\{N_0(b), 5n\}$, we deduce by the multivariate mean value theorem that for all $t \ge N(b)$ we have

$$\begin{aligned}\mathbb{E}^{\pi}_{b}\big\|\psi(b)-\psi(b_t)\big\|^2 &= \mathbb{E}^{\pi}_{b}\Big[\big\|\psi(b)-\psi(b_t)\big\|^2; A_t\Big] \;+\; \mathbb{E}^{\pi}_{b}\Big[\big\|\psi(b)-\psi(b_t)\big\|^2; A^c_t\Big] \\ &\le n^3K_0\,\mathbb{E}^{\pi}_{b}\big[\|b-b_t\|^2; A_t\big] \;+\; \mathbb{E}^{\pi}_{b}\Big[\big\|\psi(b)-\psi(b_t)\big\|^2; A^c_t\Big]. \qquad \text{(B.16)}\end{aligned}$$

Because $\|b-\widehat{b}_t\| \ge \|b-b_t\|$, the first term on the right side of inequality (B.16) is bounded above by $K_0\,\mathbb{E}^{\pi}_{b}\big[\|b-\widehat{b}_t\|^2; A_t\big]$. By the least squares equation (B.10), this further implies that

$$\begin{aligned}n^3K_0\,\mathbb{E}^{\pi}_{b}\big[\|b-b_t\|^2; A_t\big] &\le n^3K_0\,\mathbb{E}^{\pi}_{b}\Big[\big\|(\mathbf{I}_n\otimes\widehat{\mathcal{J}}_t^{-1})\,\widehat{\mathcal{M}}_t\big\|^2;\, A_t\Big] \\ &\le \frac{256n^3K_0}{\|\psi(b)\|^4(t-n)^2}\,\mathbb{E}^{\pi}_{b}[\|\widehat{\mathcal{M}}_t\|^2;\, A_t]\end{aligned}$$

because for all $t \ge N(b)$ we have $\widehat{\mu}_{\min}(t) \ge \widehat{\mathcal{J}}_t \ge \|\psi(b)\|^2(t-n)/16$ on A_t, where $\widehat{\mu}_{\min}(t)$ is the smallest eigenvalue of $\mathbf{I}_n\otimes\widehat{\mathcal{J}}_t^{-1}$. Since $\mathbb{E}^{\pi}_{b}[\|\widehat{\mathcal{M}}_t\|^2;\, A_t] \le \mathbb{E}^{\pi}_{b}[\|\widehat{\mathcal{M}}_t\|^2] = n\sigma^2\,\mathbb{E}^{\pi}_{b}[\,\mathsf{tr}(\widehat{\mathcal{J}}_t)\,] \le n^2\sigma^2(u-\ell)^2t$, we further have

$$n^3K_0\,\mathbb{E}^{\pi}_{b}\big[\|b-b_t\|^2; A_t\big] \le \frac{K_1}{t},$$

for all $t \ge N(b)$, where $K_1 = 512n^5K_0\sigma^2(u-\ell)^2\|\psi(b)\|^{-4}$. Note also that the second term on the right side of inequality (B.16) is smaller than $4n(u-\ell)^2\,\mathbb{P}^{\pi}_{\theta}(A^c_t)$, because $\widehat{p}+\psi(b)$ and $\widehat{p}+\psi(b_t)$ are in $[\ell,u]^n$. Therefore, inequality (B.16) implies that

$$\mathbb{E}^{\pi}_{b}\big\|\psi(b)-\psi(b_t)\big\|^2 \;\le\; K_1\,t^{-1} \;+\; K_2\,\mathbb{P}^{\pi}_{b}(A^c_t),$$

where $K_2 = 4n(u-\ell)^2$. Note that the preceding inequality is almost identical to inequality (A.19) in the proof of Theorem 4. Applying the final set of arguments in that proof, i.e., summing the preceding inequality over $t = N(b),\ldots,T$, and invoking condition (ii) in the hypothesis, we conclude that $\Delta^{\pi}_{b}(T) \le C_0 + C_1\log T$, where $C_0 = 4|b_{\min}|\big((u-\ell)^2N(b) + K_2k\pi^2/6\big)$ and $C_1 = |b_{\min}|(2K_1+\kappa_1)$. Letting $C = \max\{C_0, C_1\}$ concludes the proof. ■

References

Anderson, T. W. and Taylor, J. (1976), 'Some Experimental Results on the Statistical Properties of Least Squares Estimates in Control Problems', *Econometrica* **44**(6), 1289–1302.

Besbes, O. and Zeevi, A. (2009), 'Dynamic Pricing Without Knowing the Demand Function: Risk Bounds and Near-Optimal Algorithms', *Operations Research* **57**(6), 1407–1420.

Besbes, O. and Zeevi, A. (2011), 'On the Minimax Complexity of Pricing in a Changing Environment', *Operations Research* **59**(1), 66–79.

Besbes, O. and Zeevi, A. (2013), 'On the Surprising Sufficiency of Linear Models for Dynamic Pricing with Demand Learning'. Working paper, Columbia University, New York, NY.

Broder, J. and Rusmevichientong, P. (2012), 'Dynamic Pricing under a General Parametric Choice Model', *Operations Research* **60**(4), 965–980.

Carvalho, A. and Puterman, M. (2005), 'Learning and Pricing in an Internet Environment with Binomial Demands', *Journal of Revenue and Pricing Management* **3**(4), 320–336.

den Boer, A. (2013), 'Dynamic Pricing with Multiple Products and Partially Specified Demand Distribution'. Forthcoming in *Mathematics of Operations Research*.

den Boer, A. and Zwart, B. (2013), 'Simultaneously Learning and Optimizing using Controlled Variance Pricing'. Forthcoming in *Management Science*.

Gill, R. and Levit, B. (1995), 'Applications of the van Trees Inequality: A Bayesian Cramér-Rao Bound', *ISI and Bernoulli Society for Mathematical Statistics and Probability* **1**(1/2), 59–79.

Goldenshluger, A. and Zeevi, A. (2009), 'Woodroofe's One-armed Bandit Problem Revisited', *Annals of Applied Probability* **19**(4), 1603–1633.

Harrison, J., Keskin, N. and Zeevi, A. (2012), 'Bayesian Dynamic Pricing Policies: Learning and Earning Under a Binary Prior Distribution', *Management Science* **58**(3), 570–586.

Kleinberg, R. and Leighton, T. (2003), 'The Value of Knowing a Demand Curve: Bounds on Regret for Online Posted-Price Auctions', *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science* pp. 594–605.

Lai, T. and Robbins, H. (1979), 'Adaptive Design and Stochastic Approximation', *Annals of Statistics* **7**(6), 1196–1221.

Lai, T. and Robbins, H. (1982), 'Iterated Least Squares in Multiperiod Control', *Advances in Applied Mathematics* **3**(1), 50–73.

Lai, T. and Wei, C. (1982), 'Least Squares Estimates in Stochastic Regression Models with Applications to Identification and Control of Dynamic Systems', *The Annals of Statistics* **10**(1), 154–166.

Morel, P., Stalk, G., Stanger, P. and Wetenhall, P. (2003), 'Pricing Myopia'. *The Boston Consulting Group Perspectives.*

Phillips, R. (2005), *Pricing and Revenue Optimization*, Stanford University Press, Stanford, CA.

Phillips, R. (2010). Private communication.

Taylor, J. (1974), 'Asymptotic Properties of Multiperiod Control Rules in the Linear Regression Model', *International Economic Review* **15**(2), 472–484.