Joint Learning and Optimization of Multi-Product Pricing with Finite Resource Capacity and Unknown Demand Parameters

Qi (George) Chen
London Business School, Regent's Park, London, NW1 4SA, gchen@london.edu

Stefanus Jasin, Izak Duenyas
Stephen M. Ross School of Business, University of Michigan, Ann Arbor, MI 75080, sjasin, duenyas@umich.edu

We consider joint learning and pricing in network revenue management (NRM) with multiple products, multiple resources with finite capacity, parametric demand model, and a continuum set of feasible price vectors. We study the setting with a general parametric demand model and the setting with a well-separated demand model. For the general parametric demand model, we propose a heuristic that is rate-optimal (i.e., its regret bound exactly matches the known theoretical lower bound under any feasible pricing control for our setting). This heuristic is the first rate-optimal heuristic for a NRM with a general parametric demand model and a continuum of feasible price vectors. For the well-separated demand model, we propose a heuristic that is close to rate-optimal (up to a multiplicative logarithmic term). Our second heuristic is the first in the literature that deals with the setting of a NRM with a well-separated parametric demand model and a continuum set of feasible price vectors.

Key words: network revenue management, exploration and exploitation, parametric demand models, well-separated demand models, heuristics, asymptotic approach

1. Introduction

We consider a canonical dynamic pricing problem in the network revenue management (NRM) setting: A seller sells n types of products during a finite selling season subject to constraints imposed by m limited resources which cannot be replenished during the selling season, and he needs to decide the price for each product at the beginning of every decision period throughout the selling horizon. To effectively manage price adjustment, the seller needs to have a good knowledge of the underlying demand function (i.e., the average demand as a function of price); but in practice, such information is not always readily available *a priori*, so the seller needs to learn it on the fly from noisy demand observations. How should a seller jointly learn the demand function and price his products in a way that maximizes his expected total revenue?

This problem is prevalent in many industries (Talluri and van Ryzin 2005) and has drawn extensive interest from the academic literature (see den Boer (2015) for an overview). What makes

this problem challenging is that the seller not only needs to balance the well-known exploration (i.e., focusing on learning demand function by experimenting with suboptimal prices) and exploitation (i.e., focusing on earning by using the optimal price under the estimated demand function) trade-off, but also needs to manage the resource allocation over time under demand arrival uncertainties. Moreover, when the seller is managing prices for many products whose availability is subject to many resource capacity constraints, the price optimization can be computationally time-consuming: For example, it could take several hours for a hotel company to complete its price optimization *once* (Koushik et al. 2012, Pekgun et al. 2013). Due to the difficulty of computing the optimal pricing policy, most of the recent studies have focused on developing computationally efficient heuristic pricing controls with analytically provable strong performance. To analytically compare performance among different heuristics, a widely used performance measure is the so-called *regret* which corresponds to the difference between a *revenue upper bound*, defined as the maximal revenue a clairvoyant (who knows the demand function) would have got if there were no randomness in the demand realizations, and the expected revenue under a heuristic pricing control. Since most applications of NRM (e.g., airlines, hotels) involve a lot of potential sales transaction opportunities, the literature has focused on investigating the large-scale setting, which is operationalized by proportionally scaling the potential number of customers and the initial capacity levels by the same multiplicative scaling factor k, and characterizing the *asymptotic order* of regret as a function of k. An important finding in the literature is that even for the simplest special case of the NRM setting where there is one product with linear demand model and unknown intercept and slope, and one *unlimited* resource, the best (i.e., smallest) regret a seller can hope for is $\Omega(\sqrt{k})$ (Broder and Rusmevichientong 2012).This implies that, for the general NRM setting, the optimal regret bound of any pricing control can be no better than $\sqrt{k}$. Thus, prior work in the literature has developed heuristics with near-optimal regret bound in various special cases of the NRM setting which we review below. (A summary is provided in Table 1).

One stream of the literature investigates the case with the restriction that the seller is only allowed to choose from a pre-determined finite set of prices (e.g., Ferreira et al. (2018), Badanidiyuru et al. (2018)), which is closely related to the celebrated multi-armed bandit problem studied in the computer science literature. This restriction simplifies the problem in two ways: First, instead of estimating the whole demand function, the seller only needs to estimate the expected demand under a *finite* number of prices; second, the maximal revenue a clairvoyant can get with no demand uncertainty becomes lower which makes it easier to attain a smaller regret. Strong regret upper bounds have been established in this literature: For example, Ferreira et al. (2018) showed that a Thompson sampling based algorithm achieves a performance bound of $\mathcal{O}(\sqrt{kM\log M})$ where M is the number of feasible price vectors.[1]

While strong analytical results have been established in the pre-determined finite feasible price case, in practice, the seller could potentially earn more revenue by using other prices. Therefore, there is also another stream of literature that investigates the case where the seller can choose from a continuum of prices. Most of the work in this stream has focused on the so-called *nonparametric* approach: The seller has no idea about the functional form of the demand function (Besbes and Zeevi 2009, Wang et al. 2014, Besbes and Zeevi 2012, Lei et al. 2014, Chen and Gallego 2019). Unfortunately, strong analytical results under the nonparametric approach are limited to special cases. When there is a single product and a single resource, Wang et al. (2014) developed an algorithm that attains $\mathcal{O}(\sqrt{k}\log^{\frac{9}{2}}(k))$; Chen and Gallego (2019) improved/generalized the result to the case of multiple products with *no demand substitution* (in the sense that the demand of product i is independent of the prices of other products) and a single resource, and developed a primal-dual learning method that attains a regret of $\mathcal{O}(\sqrt{k}\log^2(k))$. The problem becomes very difficult with substitutable demand and multiple resource constraints because the seller needs to estimate a multi-variate vector-valued demand function in order to optimally allocate multiple resources over time. Besbes and Zeevi (2012) proposed a heuristic with sub-linear regret of $\mathcal{O}(k^{\frac{n+2}{n+3}}\log^{\frac{1}{2}}(k))$, but the regret bound deteriorates when the number of product types n is large. To address this, they imposed smoothness conditions on the underlying demand function, and proposed a heuristic with regret $O(k^{\frac{2}{3}+\epsilon}\log^{\frac{1}{2}}(k))$, where $\epsilon>0$ depends on n and how "smooth" the demand function is: If the demand function is sufficiently smooth (i.e., all of its higher order partial derivatives are uniformly bounded), ϵ can be arbitrarily small; otherwise, ϵ can be large. This result is improved by Chen et al. (2019) to $O(k^{\frac{1}{2}+\epsilon}\log(k))$ using a different heuristic under the same smoothness condition; hence, when the demand function is sufficiently smooth, there exists a nonparametric approach whose regret is arbitrarily close to the theoretical regret lower bound $\Omega(\sqrt{k})$.

An alternative to the nonparametric approach is the so-called *parametric* approach: The seller knows the parametric form of the demand function, but not its parameters. This is a popular approach in practice where practitioners may be able to figure out the type of demand functions that fits the reality well based on their institutional knowledge and past experience. Although the nonparametric approaches can be readily applied in this setting, the benefit of a parametric approach is that the seller only needs to estimate a *finite* number of parameters which fully determine the underlying demand function. One would imagine that the parametric approach should achieve lower regret than the nonparametric approaches; quite surprisingly, to the best of our knowledge, the only such result is in the setting of linear demand function families with multiple products and multiple resource constraints: Ferreira et al. (2018) proposed a TS-linear algorithm that attains a regret of $\mathcal{O}(\sqrt{k}\log(k))$. It is not clear whether their approach can be modified for other commonly used parametric demand models (e.g., multinomial logit demand) and also achieve

Table 1 Best regret upper bounds for different cases of NRM in the literature prior to our paper

Setting				**Best existing regret upper bound**
Number of feasible prices	Network complexity	Demand model	Allow product substitution?	
Finite	$n \geq 1, m \geq 1$	n/a	n/a	$\mathcal{O}(\sqrt{k})^\star$
Continuum	$n \geq 1, m = 1$	Nonparametric	No	$\mathcal{O}(\sqrt{k}\log^2(k))$
	$n \geq 1, m \geq 1$	Nonparametric	Yes	$\mathcal{O}(k^{\frac{n+2}{n+3}}\log^{\frac{1}{2}}(k))$
	$n \geq 1, m \geq 1$	Nonparametric	Yes	$\mathcal{O}(k^{\frac{1}{2}+\epsilon}\log(k))^{\star\star}$
	$n \geq 1, m \geq 1$	Linear	Yes	$\mathcal{O}(\sqrt{k}\log(k))$

$\star$ The revenue upper bound in this setting requires the clairvoyant to only use finite number of feasible price vectors, so the resulting regret bound is not directly comparable to the other regret bounds in this table. $\star\star$ The ϵ depends on n and the level of smoothness of the demand function (see Remark 2).

lower regret bounds than the nonparametric approaches. This gap in the literature calls for a parametric approach that not only works for commonly used parametric demand models but also achieves better regret bounds.

Our contributions. The contributions of this paper can be summarized as follows:

1. We revisit the NRM setting with a general parametric demand model and a continuum set of feasible price vectors. We propose a heuristic called *Parametric Self-adjusting Control* (PSC) whose regret is $O(\sqrt{k})$. To the best of our knowledge, this is the first pricing control whose regret matches the theoretical lower bound $\Omega(\sqrt{k})$ for the NRM setting when there are multiple products and multiple resource constraints. Thus, PSC not only improves and generalizes the results of Ferreira et al. (2018) but also resolves an open research problem on existence of a rate-optimal heuristic for NRM with a parametric demand model and a continuum set of feasible price vectors.

2. In addition to the setting with a general parametric demand model, we also consider the setting where the demand model also satisfies the *well-separated* condition. This type of demand model is popularized by Broder and Rusmevichientong (2012) and covers many realistic practical scenarios, including the setting where either the market size or the parameters of customer's willingness-to-pay function is unknown. We propose a modification of PSC called *Accelerated Parametric Self-adjusting Control* (APSC) whose regret is $O(\log^2 k)$ for the setting with well-separated demand models. Several novelties are involved in the design and analysis of this heuristic: First, we generalize the notion of well-separated demand model with a single parameter (Broder and Rusmevichientong 2012) to multiple product and multiple parameter setting, and our proof on the speed of learning in the multiple products setting is new. Second, in APSC, the underlying demand parameters are re-estimated as more data become available; the dynamic pricing decisions rules are re-calibrated accordingly based on an idea derived from Newton's method (Boyd and Vandenberghe 2004).

2. Problem Formulation

Notation. The following notation will be used throughout the paper. (Other notation will be introduced when necessary.) Denote by $\mathbb{R}$ (resp. $\mathbb{Z}$), $\mathbb{R}_+$ (resp. $\mathbb{Z}_+$), and $\mathbb{R}_{++}$ (resp. $\mathbb{Z}_{++}$) the set of real (resp. integer), nonnegative real (resp. integer), and positive real (resp. integer) numbers. For column vectors $a=(a_1;\ldots;a_n)\in\mathbb{R}^n, b=(b_1;\ldots;b_n)\in\mathbb{R}^n$, denote by $a\succeq b$ if $a_i\geq b_i$ for all i, and by $a\succ b$ if $a_i>b_i$ for all i. Denote by $\otimes$ the tensor product of sets, by $'$ the transpose of a vector or a matrix, and by I (resp. $\mathbf{e}$) an identity matrix (resp. a vector of ones) with a proper dimension. For any vector $v=[v_j]\in\mathbb{R}^n$, $||v||_p:=(\sum_{j=1}^n|v_j|^p)^{1/p}$ is its p-norm ($1\leq p\leq\infty$) and, for any real matrix $M=[M_{ij}]\in\mathbb{R}^{n\times n}$, $||M||_p:=\sup_{||v||_p=1}||Mv||_p$ is its induced p-norm. For example, $||M||_2=$ the largest eigenvalue of $M'M$, and $||M||_\infty=\max_{1\leq i\leq n}\sum_{j=1}^n|M_{ij}|$. For any function $f:X\to Y$, denote by $||f(.)||_\infty:=\sup_{x\in X}||f(x)||_\infty$ the infinity-norm of f. We use ∇ to denote the usual derivative operator and use a subscript to indicate the variables which this operation is applied to. (No subscript ∇ means that the derivative is applied to all variables.) If $f:\mathbb{R}^n\to\mathbb{R}$, then $\nabla_x f=(\frac{\partial f}{\partial x_1};\ldots;\frac{\partial f}{\partial x_n})$; if, on the other hand, $f=(f_1;\ldots;f_n):\mathbb{R}^n\to\mathbb{R}^n$, then

$$\nabla_x f=\begin{bmatrix}\frac{\partial f_1}{\partial x_1}&\cdots&\frac{\partial f_n}{\partial x_1}\\ \vdots&\ddots&\vdots\\ \frac{\partial f_1}{\partial x_n}&\cdots&\frac{\partial f_n}{\partial x_n}\end{bmatrix}.$$

The model. We consider the classical price-based NRM setting in which a monopolist maximizes his expected revenue from selling his products to incoming customers during a finite selling season. There are n types of products, each of which is made up from a subset of m types of resources. Denote by $A=[A_{ij}]\in\mathbb{R}_+^{m\times n}$ the *resource consumption matrix*, which indicates that a single unit of product j requires A_{ij} units of resource i. Denote by $C\in\mathbb{R}_+^m$ the vector of initial capacity levels of all resources at the beginning of the selling season which cannot be replenished and have zero salvage value at the end of the selling season.

We consider a discrete-time model with T *decision* periods, indexed by $t=1,2,...,T$. At the beginning of period t, the seller first decides the price $p_t=(p_{t,1};\ldots;p_{t,n})$ for his products, where p_t is chosen from a convex and compact set $\mathcal{P}=\otimes_{l=1}^n[\underline{p}_l,\bar{p}_l]\subseteq\mathbb{R}^n$ of feasible price vectors. The posted price p_t induces a demand, or sale, for one of the products with a certain probability. Here, we implicitly assume that at most one sale for one product occurs in each period. (We have made this assumption and chosen to focus on discrete time model to simplify the presentation of the analysis. Our analysis can be extended to either a discrete-time model with bounded demand arrivals in each period or continuous-time model with a compound Poisson process.) We assume that the demand distribution under any price is unknown to the seller, but this relationship can be estimated using statistical methods. Specifically, let $\lambda(.;.):\mathcal{P}\times\Theta\to\mathbf{\Delta}^{n-1}$ denote the family of *demand functions*

where $\mathbf{\Delta}^{n-1} := \{(x_1;\ldots;x_n) \in \mathbb{R}^n | \sum_{i=1}^n x_i \leq 1$, and $x_i \geq 0$ for all $i\}$ is the standard $(n-1)$-simplex, Θ is a compact subset of $\mathbb{R}^q$ and $q \in \mathbb{Z}_{++}$ is the dimension of the unknown parameter vector. We denote by θ^* the true parameter vector for the underlying demand function. Under the parametric demand case, the seller knows the functional form of $\lambda(.;\theta)$ for any $\theta \in \Theta$, but he does not know θ^*. Let $\Lambda_\theta := \{\lambda(p;\theta) : p \in \mathcal{P}\}$ denote the set of feasible demand rates under some parameter vector $\theta \in \Theta$. We assume that Λ_θ is convex. (For most commonly used parametric function families such as linear, multi-nomial logit, and exponential demand, Λ_θ is convex for all $\theta \in \Theta$.)

Let $D_t(p_t) = (D_{t,1}(p_t);\ldots;D_{t,n}(p_t))$ denote the vector of demand realization in period t under price p_t. It should be noted that, although demands for different products in the same period are not necessarily independent, demands over different periods are assumed to be independent conditional on the price vectors used. By definition, we have $D_t(p_t) \in \mathcal{D} := \{D \in \{0,1\}^n : \sum_{j=1}^n D_j \leq 1\}$ and $\mathbf{E}_{\theta^*}[D_t(p_t)] = \lambda(p_t;\theta^*)$. This allows us to write $D_t(p_t) = \lambda(p_t;\theta^*) + \Delta_t(p_t)$, where $\Delta_t(p_t)$ is a zero-mean random vector. For notational simplicity, whenever it is clear from the context which price p_t is being used, we will simply write $D_t(p_t)$ and $\Delta_t(p_t)$ as D_t and Δ_t respectively. The one-period expected revenue function under θ is given by the revenue function defined as $r(p;\theta) := p'\lambda(p;\theta)$. We assume that for all $\theta \in \Theta$, $\lambda(p;\theta)$ is invertible (see parametric family assumptions below); so we can write $r(p;\theta) = p'\lambda(p;\theta) = \lambda' p(\lambda;\theta) = r(\lambda;\theta)$ by abuse of notation. We make the following regularity assumptions about the family of parametric demand functions which are standard in the literature and satisfied by many commonly used demand functions.

Parametric Family Assumptions. *There exist positive constants $\bar{r}, \underline{v}, \bar{v}, \omega, \underline{v}, \bar{v}$ such that for all $p \in \mathcal{P}$ and for all $\theta \in \Theta$:*

A1. *$\lambda(.;\theta) : \mathcal{P} \to \Lambda_\theta$ is in $\mathsf{C}^2(\mathcal{P})$ and it has an inverse function $p(.;\theta) : \Lambda_\theta \to \mathcal{P}$ that is in $\mathsf{C}^2(\Lambda_\theta)$. $\lambda(p;.) : \Theta \to \mathbf{\Delta}^{n-1}$ is in $\mathsf{C}^1(\Theta)$. For all $\lambda, \lambda' \in \Lambda_\theta$, $||p(\lambda;\theta) - p(\lambda';\theta)||_2 \leq \omega||\lambda - \lambda'||_2$.*

A2. *For all $1 \leq i,j \leq n$, $||\lambda(p;\theta) - \lambda(p;\theta^*)||_2 \leq \omega||\theta - \theta^*||_2$, $|\frac{\partial \lambda_j}{\partial p_i}(p;\theta) - \frac{\partial \lambda_j}{\partial p_i}(p;\theta^*)| \leq \omega||\theta - \theta^*||_2$.*

A3. *$||r(.;\theta)||_\infty \leq \bar{r}$ and $r(.;\theta)$ is strongly concave in λ, i.e., $-\bar{v}I \preceq \nabla^2_{\lambda\lambda} r(\lambda;\theta) \preceq -\underline{v}I$ for all $\lambda \in \Lambda_\theta$.*

A4. *There exists a set of turn-off prices $p_j^\infty \in \mathbb{R} \cup \{\infty\}$ for $j = 1,\ldots,n$ such that for any $p = (p_1;\ldots;p_n)$, $p_j = p_j^\infty$ implies that $\lambda_j(p;\theta) = 0$ for all $\theta \in \Theta$.*

A1 and **A2** are natural regularity assumptions satisfied by many demand functions, e.g., linear demand, multi-nomial logit demand, and exponential demand. In **A3**, the boundedness of $r(.;\theta)$ follows since Θ and Λ_θ are compact and $r(.;\theta)$ is continuous; the strong concavity of $r(.;\theta)$ as a function of λ is a standard assumption in the literature and is satisfied by many commonly used demand functions such as linear, exponential, and multi-nomial logit functions. It should be noted that although some of these functions, such as multi-nomial logit, do not naturally correspond to

a concave revenue function when viewed as a function of p, they are nevertheless concave when viewed as a function of λ. This highlights the benefit of treating revenue as a function of demand rate instead of as a function of price. **A4** is common in the literature (Besbes and Zeevi 2009) for modeling convenience. In particular, the turn-off prices p_j^∞ are needed to model the seller's action to remove the offering of any product (i.e., shut down its demand) whenever needed (e.g., in the case of stock-out).

Admissible controls and the induced probability measures. Let $D_{1:t} := (D_1, D_2, \ldots, D_t)$ and $p_{1:t} := (p_1, p_2, \ldots, p_t)$ denote respectively the observed vectors of demand and price realizations up to and including period t. Let $\mathcal{H}_t$ denote the σ-field generated by $D_{1:t}$ and $p_{1:t}$. We define a *control* π as a sequence of functions $\pi = (\pi_1, \pi_2, \ldots, \pi_T)$, where π_t is a $\mathcal{H}_{t-1}$-measurable mapping that maps the history $D_{1:t-1}$ and $p_{1:t-1}$ to a distribution of price vectors on $\mathcal{P} \cup \{p^\infty\}$ (here $\mathcal{H}_0$ should be interpreted as the collection of seller's information before the selling horizon starts: it includes A, C, T, Θ and the class of demand functions $\{\lambda(.;\theta)\}_{\theta\in\Theta}$). This class of controls is often referred to as *non-anticipating controls* because the decision in each period depends only on the information the seller observes up to the beginning of the period. Under policy π, the seller sets the price in period t equal to $p_t^\pi = \pi_t(D_{1:t-1}, p_{1:t-1})$. Let Π_θ denote the set of all *admissible controls* if the true demand parameters were some $\theta \in \Theta$. That is,

$$\Pi_\theta := \left\{ \pi : \sum_{t=1}^{T} AD_t(p_t^\pi;\theta) \preceq C \text{ almost surely, and } \ p_t^\pi = \pi_t(\mathcal{H}_{t-1}) \right\}.$$

(Although the true underlying parameter is θ^*, we define above the set of admissible controls for any $\theta \in \Theta$.) Note that we require the capacity constraint to hold almost surely for all $\pi \in \Pi_\theta$, which can be satisfied by using turn-off prices p^∞ in case of stock-out. Let $\mathbb{P}_t^{\pi,\theta}$ denote the induced probability measure of $D_{1:t}$ under an admissible control $\pi \in \Pi_\theta$. For any realization $D_{1:t} = d_{1:t} := (d_1, d_2, \ldots, d_t)$, where $d_s = (d_{s,j}) \in \mathcal{D}, s = 1, \ldots, t$, we have:

$$\mathbb{P}_t^{\pi,\theta}(d_{1:t}) = \prod_{s=1}^{t} \left[\left(1 - \sum_{j=1}^{n} \lambda_j(p_s^\pi;\theta)\right)^{\left(1-\sum_{j=1}^{n} d_{s,j}\right)} \prod_{j=1}^{n} \lambda_j(p_s^\pi;\theta)^{d_{s,j}} \right],$$

where $p_s^\pi = \pi_s(d_{1:s-1}, p_{1:s-1}^\pi)$ where $p_{1:s-1}^\pi \in \mathcal{P}^{s-1}$ is the collection of prices used in previous periods. (By definition of $\lambda(p;\theta)$, the term $1 - \sum_{j=1}^{n} \lambda_j(p_s^\pi;\theta)$ can be interpreted as the probability of no-purchase in period s under price p_s^π.) For notational simplicity, we will write $\mathbb{P}_\theta^\pi := \mathbb{P}_T^{\pi,\theta}$ and denote by $\mathbf{E}_\theta^\pi$ the expectation with respect to the probability measure $\mathbb{P}_\theta^\pi$. The seller's total expected revenue under $\pi \in \Pi_\theta$ is given by:

$$R_\theta^\pi = \mathbf{E}_\theta^\pi \left[\sum_{t=1}^{T} (p_t^\pi)' D_t(p_t^\pi;\theta) \right].$$

Whenever it is clear that the prices $p_{1:t} \in \mathcal{P}^t$ are generated by an admissible control π, it is also convenient to write $\mathbb{P}_t^{p_{1:t},\theta}(d_{1:t}) = \prod_{s=1}^{t}[(1-\sum_{j=1}^{n}\lambda_j(p_s;\theta))^{(1-\sum_{j=1}^{n} d_{s,j})}\prod_{j=1}^{n}\lambda_j(p_s;\theta)^{d_{s,j}}]$.

Maximum likelihood estimator and exploration prices. As noted earlier, the seller can estimate the unknown θ^* using statistical methods. In this paper, we will focus primarily on *Maximum Likelihood* (ML) estimation which not only has certain desirable theoretical properties but is also widely used in practice. As shown in the statistics literature, to guarantee the regular behavior of ML estimator, certain statistical conditions need to be satisfied. To formalize these conditions in our context, it is convenient to first consider the distribution of a sequence of demand realizations when a sequence of $\tilde{q} \in \mathbb{Z}_{++}$ *fixed* price vectors $\tilde{p} = (\tilde{p}^{(1)}, \tilde{p}^{(2)}, \ldots, \tilde{p}^{(\tilde{q})}) \in \mathcal{P}^{\tilde{q}}$ have been applied. For all $d_{1:\tilde{q}} \in \mathcal{D}^{\tilde{q}}$, we define

$$\mathbb{P}^{\tilde{p},\theta}(d_{1:\tilde{q}}) := \prod_{s=1}^{\tilde{q}}\left[\left(1-\sum_{j=1}^{n}\lambda_j(\tilde{p}^{(s)};\theta)\right)^{\left(1-\sum_{j=1}^{n} d_{s,j}\right)}\prod_{j=1}^{n}\lambda_j(\tilde{p}^{(s)};\theta)^{d_{s,j}}\right],$$

and denote by $\mathbf{E}_\theta^{\tilde{p}}$ the expectation with respect to $\mathbb{P}^{\tilde{p},\theta}$. We make the following assumption.

A5. (Statistical Conditions) *There exist constants* $0 < \lambda_{\min} < \lambda_{\max} < 1$, $c_f > 0$ *and a set of prices* $\tilde{p} = (\tilde{p}^{(1)}, \ldots, \tilde{p}^{(\tilde{q})}) \in \mathcal{P}^{\tilde{q}}$ *such that:*

i. $\mathbb{P}^{\tilde{p},\theta}(.) \neq \mathbb{P}^{\tilde{p},\theta'}(.)$ *whenever* $\theta \neq \theta'$;

ii. *For all* $\theta \in \Theta$, $1 \le k \le \tilde{q}$ *and* $1 \le j \le n$, $\lambda_j(\tilde{p}^{(k)};\theta) \ge \lambda_{\min}$ *and* $\sum_{j=1}^{n}\lambda_j(\tilde{p}^{(k)};\theta) \le \lambda_{\max}$.

iii. *For all* $\theta \in \Theta$, $\mathcal{I}(\tilde{p},\theta) \succeq c_f I$ *where* $\mathcal{I}(\tilde{p},\theta) := [\mathcal{I}_{i,j}(\tilde{p},\theta)] \in \mathbb{R}^{q\times q}$ *is a* q *by* q *matrix defined as*

$$\mathcal{I}_{i,j}(\tilde{p},\theta) = \mathbf{E}_\theta^{\tilde{p}}\left[-\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log\mathbb{P}^{\tilde{p},\theta}(D_{1:\tilde{q}})\right].$$

We call $\tilde{p}$ *exploration prices*. **A5** ensures that there exists a set of price vectors (e.g., $\tilde{p}$) which, when used repeatedly, would allow the seller to use ML estimator to statistically identify the true demand parameter. Specifically, **A5**-i and **A5**-ii are crucial to guarantee that the estimation problem is well-defined, i.e., the seller is able to identify the true parameter vector by observing sufficient demand realizations under the exploration prices $\tilde{p}$. (If this is not the case, then the estimation problem is ill-defined and there is no hope for learning the true parameter vector.) The symmetric matrix $\mathcal{I}(\tilde{p},\theta)$ defined in **A5**-iii is known as the *Fisher information* matrix in the literature, and it captures the amount of information that the seller obtains about the true parameter vector using the exploration prices $\tilde{p}$. **A5**-iii requires the Fisher matrix to be strongly positive definite; this is needed to guarantee that the seller's information about the underlying parameter vector strictly increases as he observes more demand realizations under $\tilde{p}$.

REMARK 1. We want to point out that it is easy to find exploration prices for the commonly used demand function families. For example, for linear and exponential demand function families, any $\tilde{q} = n+1$ price vectors $\tilde{p}^{(1)}, \ldots, \tilde{p}^{(n+1)}$ constitute a set of exploration prices if (a) they are all in the interior of $\mathcal{P}$ and (b) the vectors $(1; \tilde{p}^{(1)})$, $\ldots$, $(1; \tilde{p}^{(n+1)}) \in \mathbb{R}^{n+1}$ are linearly independent. For the multi-nomial logit demand function family, any $\tilde{q} = 2$ price vectors $\tilde{p}^{(1)}, \tilde{p}^{(2)}$ constitute a set of exploration prices if (a) they are both in the interior of $\mathcal{P}$ and (b) $\tilde{p}_i^{(1)} \neq \tilde{p}_i^{(2)}$ for all $i = 1, \ldots, n$. While different choices of exploration prices would result in the same asymptotic convergence rate of the ML estimator, empirically, they do exhibit different convergence speed. To improve the empirical convergence speed of ML estimator, it is possible to choose the set of exploration prices that "maximizes" information accumulation in each observation, but it is beyond the scope of this paper. Interested readers are referred to the literature of optimum experimental design (e.g., Pronzato and Pázman (2013)).

The deterministic formulation and performance metric. It is common in the literature to consider the deterministic analog of the stochastic problem. Specifically, for any $\theta \in \Theta$, define:

$$\text{(P}(\theta)\text{)} \qquad J_\theta^D := \max_{p \in \mathcal{P}} \left\{ \sum_{t=1}^T r(p_t; \theta) : \sum_{t=1}^T A\lambda(p_t; \theta) \preceq C \right\},$$

$$\text{or equivalently, (P}_\lambda(\theta)\text{)} \qquad J_\theta^D := \max_{\lambda_t \in \Lambda_\theta} \left\{ \sum_{t=1}^T r(\lambda_t; \theta) : \sum_{t=1}^T A\lambda_t \preceq C \right\}.$$

By **A3**, $\mathrm{P}_\lambda(\theta)$ is a convex program and is computationally easier to solve than $\mathrm{P}(\theta)$. When $\mathrm{P}(\theta^*)$ is feasible, it can be shown that $J_{\theta^*}^D$ is in fact an upper bound for the expected revenue of any admissible control for the original stochastic problem: $R_{\theta^*}^\pi \leq J_{\theta^*}^D$ for all $\pi \in \Pi_{\theta^*}$. (See Besbes and Zeevi (2012) for proof.) This allows us to use $J_{\theta^*}^D$ as a benchmark to quantify the performance of any admissible pricing control. In this paper, we follow the convention and define the *regret* of an admissible control $\pi \in \Pi_{\theta^*}$ as $\rho^\pi := J_{\theta^*}^D - R_{\theta^*}^\pi$. Denote by $p^D(\theta)$ (resp. $\lambda^D(\theta)$) the optimal solution of $\mathrm{P}(\theta)$ (resp. $\mathrm{P}_\lambda(\theta)$). In addition, denote by $\mu^D(\theta)$ the optimal dual solution corresponding to the capacity constraints of $\mathrm{P}(\theta)$. (Note that $\mu^D(\theta)$ is also the optimal dual solution corresponding to the capacity constraints of $\mathrm{P}_\lambda(\theta)$.) Let $\mathrm{Ball}(x, r)$ denote a closed Euclidean ball centered at x with radius r. We state our last parametric assumption below:

A6. $\mathrm{P}(\theta)$ *is feasible for all* $\theta \in \Theta$ *and there exists* $\phi > 0$ *such that* $\mathrm{Ball}(p^D(\theta^*), \phi) \subseteq \mathcal{P}$.

Note that **A6** is sufficiently mild and is satisfied by most problem instances. Intuitively, it states that the deterministic optimal price should neither be too low that it attracts too much demand nor too high that it induces no demand.

Asymptotic setting. Since most revenue management applications (e.g., airlines, hotels) can be categorized as either moderate or large size, following the convention in the literature (Besbes and Zeevi 2009), we will consider a sequence of increasing problems where the the initial capacity levels and the number of decision periods in the selling season and are both scaled by a factor of $k = 1, 2, \ldots$, i.e., in the k^{th} problem, the selling horizon is divided into kT decision periods, and the initial capacity levels are given by kC. (Note that since the demand function in each period remains the same, as k increases, we effectively proportionally increase the initial capacity levels and the potential demand. In practical terms, one can interpret k as a proxy for the *size* of the problem. For example, $kC = 500$ could correspond to a flight with a capacity of 500 seats.) Note that the optimal deterministic solution of the deterministic analog of the k^{th} problem is still $\lambda^D(\theta^*)$ and the optimal dual solution is still $\mu^D(\theta^*)$. Let $\rho^\pi(k)$ denote the regret under admissible control $\pi \in \Pi_{\theta^*}$ in the k^{th} problem. We are primarily interested in the order of $\rho^\pi(k)$ as a function of k for large k. For the remainder of the paper, our goal is to develop heuristic pricing controls whose regrets grow slowly with respect to k.

3. General Demand Function Family

In this section, we introduce a heuristic called PSC whose regret exactly matches best achievable regret (up to constant multiplicative constants) for the NRM setting with multiple products, multiple resource constraints, a general parametric demand model that satisfies **A5**, and a continuum of feasible prices. Next, we introduce PSC, and then discuss the theoretical and practical implications.

3.1. Parametric Self-adjusting Control (PSC)

In PSC, the selling season is divided into an *exploration* stage followed by an *exploitation* stage. The exploration stage lasts for L periods (L is a tuning parameter to be selected by the seller) where the seller alternates among exploration prices (see **A5** and Remark 1 for their definitions and how to select them) to learn the demand function. At the end of the exploration stage, the seller computes his ML estimate of θ^*, denoted by $\hat{\theta}_L$ (in case the maximum of the likelihood function is not unique, take any maximum as the ML estimate), based on all his observations so far, and solves $\mathrm{P}_\lambda(\hat{\theta}_L)$ for its solution $\lambda^D(\hat{\theta}_L)$ as an estimate of the deterministically optimal demand rate $\lambda^D(\theta^*)$. Then, for the remaining $(T-L)$-period exploitation stage, the seller uses price vectors according to a simple adaptive rule which we explain in more detail below. Define $\hat{\Delta}_t(p_t; \hat{\theta}_L) := D_t - \lambda(p_t; \hat{\theta}_L)$ (we will suppress the dependency of $\hat{\Delta}$ on p_t and $\hat{\theta}_L$ when there is no confusion) and let C_t denote the remaining capacity at the *end* of period t. The complete PSC procedure is given below.

Parametric Self-adjusting Control (PSC)

Tuning Parameter: L

Stage 1 (Exploration)

a. Determine the exploration prices $\{\tilde{p}^{(1)}, \tilde{p}^{(2)}, ..., \tilde{p}^{(\tilde{q})}\}$.

b. For $t = 1$ to L, do:

- If $C_{t-1} \succeq A_j$ for all j, apply price $p_t = \tilde{p}^{(\lfloor (t-1)\tilde{q}/L \rfloor + 1)}$ in period t.
- Otherwise, apply price $p_{t',j} = p_j^\infty$ for all j and $t' \geq t$; then terminate PSC.

c. At the end of period L:

- Compute the ML estimate $\hat{\theta}_L$ based on $p_{1:L}$ and $D_{1:L}$.
- Solve $\mathrm{P}_\lambda(\hat{\theta}_L)$ for $\lambda^D(\hat{\theta}_L)$.

Stage 2 (Exploitation)

For $t = L+1$ to T, compute:

$$\hat{p}_t = p\left(\lambda^D(\hat{\theta}_L) - \sum_{s=L+1}^{t-1} \frac{\hat{\Delta}_s(p_s; \hat{\theta}_L)}{T-s}; \hat{\theta}_L\right). \tag{1}$$

- If $C_{t-1} \succeq A_j$, and $\hat{p}_t \in \mathcal{P}$, apply price $p_t = \hat{p}_t$ in period t
- Otherwise, for product $j = 1$ to n, do:
  - If $C_{t-1} \prec A_j$, apply price $p_{t,j} = p_j^\infty$.
  - Otherwise, apply price $p_{t,j} = p_{t-1,j}$.

What is the intuition behind the self-adjusting pricing rule in (1)? The idea seems fairly intuitive if the estimate of the parameter vector is accurate, a setting studied in Jasin (2014). In that setting, $\hat{\Delta}_t$ equals the stochastic variability in demand arrivals Δ_t, and the pricing rule in (1) reduces to adjusting the prices in each period t to achieve a *target demand rate*, i.e., $\lambda^D(\theta^*) - \sum_{s=L+1}^{t-1} \frac{\Delta_s}{T-s}$. The first part of this expression, $\lambda^D(\theta^*)$, is the optimal demand rate if there were no stochastic variability, and we use it as a *base rate*; the second part of the expression, on the other hand, works as a fine adjustment to the base rate in order to mitigate the observed stochastic variability. To see how such adjustment works, consider the case with a single product: If there is more demand than what the seller expects in period s, i.e., $\Delta_s > 0$, then the pricing rule automatically accounts for it by reducing the target demand rate for all remaining $(T-s)$-period; moreover, the target demand rate adjustment is made *uniformly* across all $(T-s)$-period so as to minimize unnecessary price variations. Jasin (2014) has shown that the ability to *accurately* mitigate the stochastic variability allows this self-adjusting pricing rule be effective *when the parameter vector is known*. However, as one can imagine, such *precise* adjustment is not possible when the parameter vector is subject to estimation error. Indeed, when $\hat{\theta}_L \neq \theta^*$, the seller can only adjust target demand rate based on an estimate of Δ_s, i.e., $\hat{\Delta}_s$; moreover, the seller can no longer correctly find out the price vector

that accurately induces (on average) the target demand rate since the inverse demand function is also subject to estimation error. Can this pricing rule work well when the underlying demand parameter is subject to estimation error? The answer is yes, and the key observation is that these two sources of systematic biases push the price decisions on opposing directions and their impact is thus reduced. To see that, consider a single product case where the seller over-estimates demand for all prices, i.e., $\lambda(p;\hat{\theta}_L) > \lambda(p;\theta^*)$ for all p: On the one hand, since the seller would underestimate the stochastic variation that he needs to adjust (i.e., $\hat{\Delta}_s = D_s - \lambda(p_s;\hat{\theta}_L) < D_s - \lambda(p_s;\theta^*) = \Delta_s$), this would push up the target demand rate (which would push down the price) than if there were no estimation error; on the other hand, since $p(\lambda;\hat{\theta}_L) > p(\lambda;\theta^*)$, for a given target demand rate, the presence of estimation error would push the price up. Quite interestingly, these opposing mechanisms are sufficient for PSC to achieve the optimal rate of regret.

THEOREM 1. (RATE-OPTIMALITY OF PSC) *Suppose that **A1**-**A6** hold. Set $L = \lceil \sqrt{kT} \rceil$. Then, there exists a constant $M_1 > 0$ independent of $k \geq 1$ such that $\rho^{PSC}(k) \leq M_1\sqrt{k}$ for all $k \geq 1$.*

To the best of our knowledge, PSC is the first heuristic that achieves $\mathcal{O}(\sqrt{k})$ in a setting with multiple products, multiple resource constraints, and a continuum of feasible prices. And it leverages the fact that the demand model is fully determined by a *finite dimensional* vector θ^*, which can be efficiently estimated by ML estimation:

LEMMA 1. (BOUNDS FOR ML ESTIMATOR WITH I.I.D OBSERVATIONS) *Suppose that **A5** holds. There exist positive constants η_1, η_2, η_3 independent of $L > 0$ and $\theta \in \Theta$, such that for all $\delta > 0$, we have $\mathbb{P}^{\pi}_{\theta}(||\theta - \hat{\theta}_L||_2 > \delta) \leq \eta_1 \exp(-\eta_2 L\delta^2)$ and $\mathbf{E}^{\pi}_{\theta}[||\theta - \hat{\theta}_L||_2^2]^{1/2} \leq \eta_3/\sqrt{L}$.*

The only heuristic we are aware of that achieves comparable regret is the TS-linear proposed in Section 4.1 in Ferreira et al. (2018) which achieves a slightly worse regret bound of $\mathcal{O}(\sqrt{k}\log k)$. Compared to PSC, TS-linear has two limitations. First, TS-linear is designed for a special case of linear demand model, while PSC can be applied to a much broader range of parametric demand models. Second, one of TS-linear's critical step in each decision period is to sample a random parameter vector $\tilde{\theta}$ from the posterior distribution of the parameters and then re-optimize $\mathrm{P}(\tilde{\theta})$: When the posterior does not permit closed-form, *in each period*, the seller needs to use Markov chain Monte Carlo methods to sample $\tilde{\theta}$ and then optimize $\mathrm{P}(\tilde{\theta})$ which can be computationally very time-consuming. This may limit the practicability of TS-linear in many large-scale revenue management applications where there are a lot of transaction opportunities (i.e., large k) and the seller needs to make frequent price adjustments. In contrast, PSC is much more applicable for those large-scale applications as it only requires a single convex optimization and one maximum likelihood estimation throughout the entire selling season, and all other steps of PSC can be easily

Table 2 Performance comparison between PSC and TS-Linear

Market "size"	$k=10^2$				$k=10^3$				$k=10^4$			
Policy	TS-Linear		PSC		TS-Linear		PSC		TS-Linear		PSC	
	$R^{\pi}_{\theta^*}/J^D_{\theta^*}$ (%)	time (sec.)	$R^{\pi}_{\theta^*}/J^D_{\theta^*}$ (%)	time (sec.)	$R^{\pi}_{\theta^*}/J^D_{\theta^*}$ (%)	time (sec.)	$R^{\pi}_{\theta^*}/J^D_{\theta^*}$ (%)	time (sec.)	$R^{\pi}_{\theta^*}/J^D_{\theta^*}$ (%)	time (sec.)	$R^{\pi}_{\theta^*}/J^D_{\theta^*}$ (%)	time (sec.)
$C=(3,5,7)$	67.0	1.7	79.7	4.0e-2	89.7	34.6	94.4	4.5e-2	89.8	2.6e3	98.5	2.0e-1
$C=(15,12,30)$	79.9	2.2	82.8	3.2e-2	82.5	37.4	94.3	4.1e-2	82.5	2.7e3	99.0	1.5e-1

In these examples, $n=2$, $m=3$, $A=[1,1;3,1;0,5]$, the demand function family is $\lambda(p_1,p_2;a_1,a_2,b_{11},b_{12},b_{21},b_{22}) = (a_1+b_{11}p_1+b_{12}p_2, a_2+b_{21}p_1+b_{22}p_2)'$ and the true parameter vector is $(8,9,-1.5,0,0,-3)$. For each setting, we conduct 500 independent sample runs for each policy, take the sample average of the revenues for $R^{\pi}_{\theta^*}$ and the sample average of the policy running time for "time". For TS-Linear, we use uniform distribution as the prior of the parameter vector; since the posterior of the parameter vector does not have a closed-form, we use the standard Metropolis-Hastings algorithm to draw demand parameter vector from the posterior.

computed. To further highlight the practical benefit of PSC, we compare PSC and TS-linear using the numerical examples in Ferreira et al. (2018) and summarize the percent of optimal revenue achieved (i.e., $R^{\pi}_{\theta^*}/J^D_{\theta^*}$) and computational time in Table 2. In all these examples, PSC not only performs significantly better than TS-linear but also achieves a computational time that is orders of magnitudes smaller. This suggests that PSC not only has a very strong analytical regret bound, but is also a scalable heuristic control that has strong empirical performance.

Remark 2. In a closely related work, Chen et al. (2019) developed a nonparametric approach named NSC. In fact, one can view PSC as a tailored version of NSC specifically designed for the parametric setting to leverage the knowledge of the parametric form of the demand model. While the analysis of PSC is analogous to NSC, the improvement of PSC's regret bound is quite significant. The regret bound of NSC, with properly chosen tuning parameters, is $\mathcal{O}(k^{\frac{1}{2}+\epsilon(n,\bar{s})})$ where $\epsilon(n,\bar{s}) = \frac{1}{2}\frac{n+2}{2\bar{s}+n-2}$, n is the number of products, and $\bar{s}$ is the highest degree below which the partial derivatives of the demand function are uniformly bounded by a constant. Note that for a given n, if $\bar{s}$ is sufficiently large (i.e., the demand function is sufficiently smooth), then $\epsilon(n,\bar{s})$ can be very close to zero, so NSC's regret can be very close to $\mathcal{O}(\sqrt{k})$; but, when NSC is blindly applied in our parametric setting, **A1** implies that $\bar{s}=3$, so the regret bound is $\mathcal{O}(k^{1-\frac{1}{n+4}})$, which is far worse than PSC's regret bound $\mathcal{O}(\sqrt{k})$. This means that while theoretically, NSC may attain a regret bound close to $\sqrt{k}$ when the underlying demand function is sufficiently smooth, it may not necessarily perform well in practice. This observation also raises an open question: Can $\mathcal{O}(\sqrt{k})$ regret be attained in the nonparametric setting with multiple products, multiple resources, and a continuum of prices? The most general case in the literature where regret bounds that are only up to logarithmic multiplicative terms larger than $O(\sqrt{k})$ is achievable is the case of single resource and multiple independent products (Chen and Gallego 2019). Two features of this setting greatly simplify the analysis: First, it has the nice structure that the optimal solution is either binding

or non-binding at the resource constraint; second, due to separable demand, the nonparametric estimation problem is effectively single-dimensional. This neat structure breaks down when there are multiple resource constraints and when demands are substitutable: To identify binding constraints, the seller may need to learn the whole multi-variate demand function which is equivalent to estimating the average demand on an uncountably many number of price vectors. Thus, we suspect that some new ideas in identifying the binding constraints without necessarily estimating the whole nonparametric demand model is necessary to develop nonparametric approaches with $\mathcal{O}(\sqrt{k})$ regret for NRM problem with a continuum of prices.

4. Well-Separated Demand Function Family

The joint learning and pricing problem studied in Section 3 is very general: It allows both a general parametric demand model and an arbitrary finite number of unknown parameters. This generality makes the learning problem difficult because *not all prices are equally informative*. For example, as illustrated in Figure 1 in Broder and Rusmevichientong (2012), when the true demand function belongs to the class of demand functions $\lambda(p;\theta) = 1/2 + \theta - \theta p$, the price $p = 1$ is an "uninformative price" since using it cannot help the seller statistically identify the true demand function: when $p = 1$, any θ would result in 50% chance of observing a demand. Therefore, in order to learn the true demand function, the seller needs to avoid uninformative prices by actively experimenting with informative prices; the need to conduct such costly price experiment is the reason why the regret lower bound is $\Omega(\sqrt{k})$ in general. In practice, however, there may be additional institutional knowledge that the seller can use to impose more structure into the class of parametric demand models in order to simplify learning. For example, suppose a seller has a good understanding of the distribution of customers' reservation prices but is uncertain about the potential market size; he may choose a parametric demand function as $\lambda(p;\theta) = \theta F(p)$ where F is a known function that captures the percentage of customers whose reservation price is above p. (Similar settings have been studied in Aviv and Pazgal (2005), Araman and Caldentey (2009), Farias and van Roy (2010) and Chen and Farias (2013).) Then, all prices such that $F(p) > 0$ are informative because under the same p, different θ would result in different probability of observing a demand. This observation motivates us to consider a special class of *well-separated demand functions* (to be formally defined below) which enables the seller to gather information without necessarily engaging in costly price experimentations. Can the seller achieve lower regret than $\Omega(\sqrt{k})$? If so, can such regret be achieved without computationally extensive re-optimizations? To address these questions, we first introduce the formal definition of well-separated demand, and then propose a modification of PSC that attains logarithmic-squared regret.

4.1. Well-separated demand

Define a collection of prices $\mathcal{W}(\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max}) := \{p \in \mathcal{P} : \sum_{j=1}^{n} \lambda_j(p;\theta) \le \tilde{\lambda}_{\max}, \lambda_j(p;\theta) \ge \tilde{\lambda}_{\min}, j = 1,\dots,n, \text{ for all } \theta \in \Theta\}$, where

$$\tilde{\lambda}_{\min} := \min\left\{\lambda_{\min}, \min_{\substack{p \in \mathsf{Ball}\left(p^D(\theta^*), \frac{7\phi}{8}\right)\\ \theta\in\Theta}} \min_{1\le j\le n} \lambda_j(p;\theta)\right\}, \tilde{\lambda}_{\max} := \max\left\{\lambda_{\max}, \max_{\substack{p \in \mathsf{Ball}\left(p^D(\theta^*), \frac{7\phi}{8}\right)\\ \theta\in\Theta}} \sum_{j=1}^{n} \lambda_j(p;\theta)\right\}.$$

Note that ϕ is defined in **A6** (in fact, for the results in this section to go through, one can replace $\mathsf{Ball}\left(p^D(\theta^*), \frac{7\phi}{8}\right)$ in the definition of $\tilde{\lambda}_{\min}$ and $\tilde{\lambda}_{\max}$ by $\mathsf{Ball}\left(p^D(\theta^*), l\right)$ with any $l \in (0,\phi)$); so by **A5**-ii and **A6**, one can easily verify that $0 < \tilde{\lambda}_{\min} < \tilde{\lambda}_{\max} < 1$, and $p^D(\theta^*) \in \mathcal{W}(\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max})$. All the results in this section require the following well-separated assumption to hold.

A7. (Well-Separated Assumption) *There exists $c_f > 0$ such that:*

i. *For all $p \in \mathcal{W}(\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max})$, $\mathbb{P}^{p,\theta}(.) \ne \mathbb{P}^{p,\theta'}(.)$ whenever $\theta \ne \theta'$;*

ii. *For all $\theta \in \Theta$, $p \in \mathcal{W}(\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max})$, $I(p,\theta) \succeq c_f I$ for $I(p,\theta) := [I_{i,j}(p,\theta)] \in \mathbb{R}^{q\times q}$ defined as*

$$[I(p,\theta)]_{i,j} = \mathbf{E}_\theta^p\left[-\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log\mathbb{P}^{p,\theta}(D)\right] = \mathbf{E}_\theta^p\left[-\frac{\partial}{\partial\theta_i}\log\mathbb{P}^{p,\theta}(D)\,\frac{\partial}{\partial\theta_j}\log\mathbb{P}^{p,\theta}(D)\right].$$

iii. *For any $p_{1:t} = (p_1,\dots,p_t) \in \mathcal{W}(\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max})^t$, $\log\mathbb{P}_t^{p_{1:t},\theta}(D_{1:t})$ is concave in θ on Θ.*

The idea of well-separated demand functions has been proposed by Broder and Rusmevichientong (2012) for the single product single parameter case. **A7** generalizes the idea of well-separated demand to the multiple product multiple parameters case, and ensures that any *single* price vector p that is not too far away from $p^D(\theta^*)$ (i.e., $p \in \mathcal{W}(\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max})$) are "informative": the seller can learn the true parameter vector by observing the demand realizations under that price vector. A necessary condition for **A7**-i to hold is that $n < q$; otherwise, under any price $p \in \mathcal{P}$, even with a perfect observation of $\lambda(p;\theta^*)$, the seller cannot uniquely identify a q-dimensional parameter vector θ^*. Note that **A7**-ii is analogous to **A5**-iii and it ensures that the seller's information about the parameter vector strictly increases as he observes more demand realizations under any $p \in \mathcal{W}(\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max})$. The last condition W3 requires the log-likelihood function to behave nicely. This is easily satisfied by many commonly used demand functions such as linear, multi-nomial logit, and exponential demand functions. Similar to Remark 4.1 in Broder and Rusmevichientong (2012), **A7** implies that there exists some constant $c_d > 0$ such that for any $\theta, \theta' \in \Theta$ and $p \in \mathcal{W}(\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max})$, $||\lambda(p;\theta) - \lambda(p;\theta')||_2 \ge c_d||\theta - \theta'||_2$ (see Section EC.5 of the Online Appendix for proof); this inequality and the fact that the demand function is continuous in p imply that, for all $p \in \mathcal{W}(\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max})$, the corresponding demand curves do not intersect each other, and are thus "well-separated". Note that this well-separated condition is not overly restrictive as it permits, for

example general demand functions with unknown additive market size (i.e., for each product j, its demand is $\lambda_j(p) = a_j + g_j(p)$ where the market size a_j is unknown and $g_j : \mathcal{P} \to [0,1]$ is a known function) and general demand functions with unknown multiplicative market size (i.e., for each product j, its demand is $\lambda_j(p) = a_j g_j(p)$ where the market size a_j is unknown and $g_j : \mathcal{P} \to [0,1]$ is a known function). (For more examples of well-separated demand in the single product single parameter setting, see Broder and Rusmevichientong (2012).) A nice implication of **A7** is that a tight estimation error bound similar to Lemma 1 would hold for *non-i.i.d. observations*, which is formalized in Lemma 2 below.

Lemma 2. (Estimation Error of ML Estimator with Non-I.I.D Observations) *Suppose that **A5** and **A7** hold. For any $q \in \mathbb{Z}_+$ and any admissible control π which satisfies $p_s = \pi_s(D_{1:s-1}) \in \mathcal{W}(\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max})$ for all $1 \le s \le t$, there exist constants $\eta_4, \eta_5, \eta_6 > 0$ independent of t and $\theta \in \Theta$, such that $\forall \delta > 0, \mathbb{P}_\theta^\pi(||\theta - \hat{\theta}_t||_2 > \delta) \le \eta_4 t^{q-1} \exp(-\eta_5 t \delta^2)$ and $\mathbf{E}_\theta^\pi[||\theta - \hat{\theta}_t||_2^2]^{1/2} \le \eta_6 \sqrt{[(q-1)\log t + 1]/t}$.*

Remark 3. The special case of Lemma 2 where $q = 1$ has been established in Theorem 4.7 in Broder and Rusmevichientong (2012). Generalizing their result to $q > 1$ is non-trivial. When $q = 1$, Θ lies on the real line, so the event that the estimation error is larger than δ implies that either $\hat{\theta}_t > \theta + \delta$ or $\hat{\theta}_t < \theta - \delta$. Hence, by **A7**-iii, $\mathbb{P}_\theta^\pi(||\theta - \hat{\theta}_t||_2 > \delta)$ is bounded from above by the probability that the likelihood of θ is smaller than *two* parameters: $\theta - \delta$ and $\theta + \delta$. In contrast, when $q > 1$, $\mathbb{P}_\theta^\pi(||\theta - \hat{\theta}_t||_2 > \delta)$ is bounded from above by the probability that the likelihood of θ is smaller than the maximum of the likelihood of an *uncountably many* number of parameter vectors (i.e., the boundary of $\mathsf{Ball}(\theta, \delta) \subseteq \mathbb{R}^q$). In our proof, we approximate the largest likelihood on the boundary of $\mathsf{Ball}(\theta, \delta)$ by a carefully chosen finite set of parameter vectors to reduce the challenge of deriving an upper bound of the sum of uncountably many probabilities to a finite number of probabilities.

In contrast to the setting in Section 3, Lemma 2 shows that in the well-separated demand setting, conducting price experimentation with suboptimal exploration prices is not the only way to learn θ^*; in fact, as long as the prices used are in $\mathcal{W}(\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max})$, the seller is also guaranteed to learn more information about θ^*. This implies that PSC, which solely relies on price experimentation during exploration stage to learn θ^* and neglects any learning opportunity during exploitation stage, may not be ideal for the well-separated setting. In fact, during exploitation stage, PSC may be enhanced by occasionally re-estimating θ^* based on all of the past observations and adjusting the pricing rule accordingly. We explain how this can be done below.

4.2. Accelerated Parametric Self-adjusting Control (APSC)

In this section, we develop APSC which incorporates re-estimation and re-calibration features into the exploitation stage of PSC to leverage the benefit of passive learning for well-separated demand. In a nutshell, APSC re-estimates the parameter vector at certain decision periods during exploitation, and uses the latest estimate to re-calibrate the estimate of the base rate $\lambda^D(\theta^*)$ and the price adjustment rule. We first provide the full description of the APSC below and then explain the details and the intuitions of the two main design features of this pricing control.

Accelerated Parametric Self-adjusting Control (APSC)

Tuning Parameters: L, η

Stage 1 (Exploration)

a - c in Stage 1 of PSC.

d. (Re-estimation Initialization) Compute $\mathcal{T} := \{t_z, 1 \le z \le Z+1\}$ where

$$t_1 = L, t_2 = L+1, t_{Z+1} = T, t_z = \lceil \frac{t_{z+1} - L}{2} \rceil + L, \forall 2 \le z \le Z. \tag{2}$$

e. (Re-calibration Initialization) Set $\mathcal{B} = \emptyset, \mathcal{N} = \{1, \ldots, m\}$, and do:

- For all i, compute the slack $s_i = C_i/T - (A\lambda^D(\hat{\theta}_{t_1}))_i$ of the i^{th} constraint.
- While $\mathcal{N} \neq \emptyset$, do: Let $i = \min_{i \in \mathcal{N}} s_i$; remove i from $\mathcal{N}$; add i to $\mathcal{B}$ if

$$\{(A\lambda^D(\hat{\theta}_{t_1}))_j\}_{j \in \mathcal{B} \cup \{i\}} \text{ are linearly independent, and } s_i \le \eta. \tag{3}$$

- Set auxiliary matrices $C_B := C(\mathcal{B})$, $B := A(\mathcal{B}, :)$, and variables

$$x_1^{NT} := \lambda^D(\hat{\theta}_{t_1}), \nu_1^{NT} := (BB')^{-1} B \nabla_\lambda r(x_1^{NT}; \hat{\theta}_{t_1}). \tag{4}$$

- Compute the estimate of the base rate: $\lambda_1^{NT} := x_1^{NT}$.

Stage 2 (Exploitation)

For $t = L+1$ to T:

a. Let z be such that $t_z < t \le t_{z+1}$, and compute

$$\hat{p}_t = p\left(\lambda_z^{NT} - \textstyle\sum_{s=L+1}^{t-1} \frac{\hat{\Delta}_s}{T-s}; \hat{\theta}_{t_z}\right), \tag{5}$$

where for all $s \in (L,t) \cap (t_{z'}, t_{z'+1}]$, $\hat{\Delta}_s := D_s - \lambda(p_s; \hat{\theta}_{t_{z'}})$.

- If $C_{t-1} \succeq A_j$, and $\hat{p}_t \in \mathcal{W}(\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max})$, apply price $p_t = \hat{p}_t$ in period t
- Otherwise, for product $j = 1$ to n, do:
  - If $C_{t-1} \prec A_j$, apply price $p_{t,j} = p_j^\infty$.
  - Otherwise, apply price $p_{t,j} = p_{t-1,j}$.

If $t = t_{z+1}$, then do the following two additional steps:

b. (Re-estimation step) Compute ML estimate $\hat{\theta}_{t_{z+1}}$ based on $p_{1:t_{z+1}}$ and $D_{1:t_{z+1}}$.

c. (Re-calibration step) Compute the new estimate of base rate, λ_{z+1}^{NT}, as follows:

- Compute x_{z+1}^{NT} and ν_{z+1}^{NT} as follows

$$\begin{bmatrix} x_{z+1}^{NT} \\ \nu_{z+1}^{NT} \end{bmatrix} := \begin{bmatrix} x_{z}^{NT} \\ \nu_{z}^{NT} \end{bmatrix} + \begin{bmatrix} -\nabla^2_{\lambda\lambda} r(x_z^{NT}; \hat{\theta}_{t_{z+1}}) & B' \\ B & O \end{bmatrix}^{-1} \begin{bmatrix} \nabla_\lambda r(x_z^{NT}; \hat{\theta}_{t_{z+1}}) - B'\nu_z^{NT} \\ C_B - Bx_z^{NT} \end{bmatrix} \tag{6}$$

- Let $\lambda_{z+1}^{NT} := x_{z+1}^{NT}$.

4.2.1. Re-estimation. Similar to PSC, APSC divides the selling season into the L-period exploration and $(T-L)$-period exploitation stages; the new feature in APSC is that θ^* is re-estimated (Stage 2 Step b) at re-estimation points in $\mathcal{T}$ defined in Stage 1 Step d. (Note that given T and L, both Z and $\mathcal{T}$ are well-defined.) The re-estimation points are designed in such a way that the estimate of θ^* is updated more frequently earlier in the selling season when it is still highly inaccurate; as the accuracy improves APSC reduces the frequency of re-estimation. These re-estimation points naturally divide the exploitation stage into Z segments where the z^{th} segment contains all the periods in $(t_z, t_{z+1}] := \{t_z+1, t_z+2, ..., t_{z+1}\}$. At the end of each segment z, based on all the observations, APSC uses ML estimation to get a re-estimate $\hat{\theta}_{t_{z+1}}$ which is used for the pricing decisions in the next segment. In contrast to the setting in the previous section, **A7** ensures that as long as the prices used in segment z are in $\mathcal{W}(\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max})$, the new data observed in segment z can always provide useful information about θ^*, so $\hat{\theta}_{t_{z+1}}$ would be a more accurate estimator of θ^* to improve pricing decisions in the next segment. Next we discuss how APSC incorporates the latest estimate into the pricing.

4.2.2. Re-calibration. Consider a period t where $t_z < t \le t_{z+1}$. To use the latest estimate $\hat{\theta}_{t_z}$, one could replace $\hat{\theta}_L$ in (1) with $\hat{\theta}_{t_z}$, resulting in the following pricing rule:

$$\hat{p}_t = p\left(\lambda^D(\hat{\theta}_{t_z}) - \textstyle\sum_{s=L+1}^{t-1} \frac{\hat{\Delta}_s}{T-s}; \hat{\theta}_{t_z}\right).$$

The practical challenge of applying the pricing rule above is that $\lambda^D(\hat{\theta}_{t_z})$ requires solving a constrained optimization problem $\mathrm{P}_\lambda(\hat{\theta}_{t_z})$ at every re-estimation point. As reported in (Koushik et al. 2012) and Pekgun et al. (2013), re-optimizations can be quite computationally expensive for problems with many products and many resource constraints. The re-calibration scheme in APSC which we explain in more detail below is designed to address this challenge. The high-level idea is the following. Note that the main reason of using $\lambda^D(\hat{\theta}_{t_z})$ is because it serves as an approximation of the base rate $\lambda^D(\theta^*)$; since some approximation error is incurred by using $\lambda^D(\hat{\theta}_{t_z})$ anyway, it is not necessary for the seller to exactly solve $\mathrm{P}_\lambda(\hat{\theta}_{t_z})$ for the optimal solution $\lambda^D(\hat{\theta}_{t_z})$, *as long as he can find a reasonably good approximation of* $\lambda^D(\theta^*)$, i.e., this is the ultimate goal the re-calibration

scheme tries to achieve. Motivated by this observation, instead of re-optimizing $\mathrm{P}_\lambda(\hat{\theta}_{t_z})$, APSC uses the sequence of re-estimates $\{\hat{\theta}_{t_z}\}_z$ to obtain, according to a *re-calibration scheme based on the Newton's method* (i.e., Stage 1 Step e and Stage 2 Step c), a sequence of increasingly more accurate approximations of $\lambda^D(\theta^*)$, i.e., $\{\lambda_z^{NT}\}_z$, which are used in (5).

To explain how the re-calibration scheme in APSC achieves the goal of finding good approximations of $\lambda^D(\theta^*)$ (without re-optimization) and its connection to Newton's method, the following observation is critical: Note if the seller knew which resource constraints are binding at the optimal solution in $\mathrm{P}_\lambda(\theta^*)$, then, due to strong concavity of the objective function, $\lambda^D(\theta^*)$ coincides with the optimal solution of another optimization problem with only equality constraints:

$$\max_{x\in\mathbb{R}^n}\{r(x;\theta^*) : B(\theta^*)x = C_B(\theta^*)/T\}, \tag{7}$$

where $B(\theta^*)$ and $C_B(\theta^*)$ correspond to the sub-matrices of A and C with only rows that correspond to the binding constraints of $\mathrm{P}_\lambda(\theta^*)$, and we denote by $x^D(\theta^*)$ (resp. $\nu^D(\theta^*)$) the optimal primal (resp. dual) solution to (7). This observation connects our problem to the celebrated Newton's method because it is a very effective iterative method to find the optimal solution of a concave optimization with *only equality constraints*: Take a pair of primal and dual feasible solutions (x,ν) to (7) that are in the neighborhood of $(x^D(\theta^*),\nu^D(\theta^*))$, one can get a new pair of primal and dual solutions $(\tilde{x},\tilde{\nu})$ such that $\tilde{x}$ is a much better approximation of $x^D(\theta^*)$ than x by applying one *Newton's iteration*,

$$\begin{bmatrix}\tilde{x}\\ \tilde{\nu}\end{bmatrix} := \begin{bmatrix}x\\ \nu\end{bmatrix} + \begin{bmatrix}-\nabla^2_{\lambda\lambda} r(x;\theta^*) & B(\theta^*)'\\ B(\theta^*) & O\end{bmatrix}^{-1}\begin{bmatrix}\nabla_\lambda r(x;\theta^*) - B(\theta^*)'\nu\\ C_B(\theta^*) - B(\theta^*)x\end{bmatrix}. \tag{8}$$

While (8) can be used to identify a sequence of solutions that converges to $\lambda^D(\theta^*)$ (since $x^D(\theta^*) = \lambda^D(\theta^*)$), the seller cannot use it since he does not know $\nabla_\lambda r(.;\theta^*), \nabla_{\lambda\lambda} r(.;\theta^*), B(\theta^*), C_B(\theta^*)$. But, as the re-estimates of θ^* gets more and more accurate, (8) can be approximated reasonably close so as to generate useful approximations of $\lambda^D(\theta^*)$. Specifically, at the end of the exploration stage, APSC ensures that $B(\theta^*)$ and $C_B(\theta^*)$ can be correctly identified with high probability (Stage 1 Step e); during exploitation stage, after each re-estimation, APSC replaces θ^* in the Hessian and Jacobian matrix of the revenue function in (8) by its latest estimate, and applies the Newton's iteration to get a new approximation of $\lambda^D(\theta^*)$. Below, we explain these two steps in more detail.

Re-calibration initialization (Stage 1 Step e). The goal of this initialization procedure is two-fold: (i) Identify the binding constraints for $\mathrm{P}_\lambda(\theta^*)$, and (ii) obtain a pair of primal and dual solutions of (7) as initial points for a sequence of Newton's iterations which mimic (8). Since $\mathrm{P}_\lambda(\theta^*)$ is unknown and cannot be solved, to achieve (i), APSC solves $\mathrm{P}_\lambda(\hat{\theta}_L)$ as an approximation to determine which resource constraints should be added to the set of *estimated* binding constraints $\mathcal{B}$. In identifying $\mathcal{B}$,

all constraints are considered sequentially based on their slacks at the optimal solution of $\mathrm{P}_\lambda(\hat{\theta}_L)$: Intuitively, if $\hat{\theta}_L$ is a good estimate of θ^*, one would imagine that constraints with smaller slack at the optimal solution of $\mathrm{P}_\lambda(\hat{\theta}_L)$ are more likely to be binding in $\mathrm{P}_\lambda(\theta^*)$. The criterion for whether or not to add a constraint i is based on (3): the first part ensures constraint i is not added if it is "redundant", and the second part requires the slack of constraint i to be smaller than a threshold η which is chosen by the seller. This threshold can be thought of as a form of "protection" against excluding true binding constraints from $\mathcal{B}$ due to the fact that we use $\mathrm{P}_\lambda(\hat{\theta}_L)$ as an approximation of $\mathrm{P}_\lambda(\theta^*)$. When η is chosen appropriately, it can be shown that the constraints in $\mathcal{B}$ coincide with the binding constraints in $\mathrm{P}_\lambda(\theta^*)$ (and hence $B = B(\theta^*)$ and $C_B = C_B(\theta^*)$) with a very high probability as $k \to \infty$. (We address how to chose η in Theorem 2 below.) Using $\mathcal{B}$ as the estimate of the binding constraints of $\mathrm{P}_\lambda(\theta^*)$, we can focus on the *Equality Constrained Problem* (ECP) below per our previous discussion on Newton's method:

$$\mathrm{ECP}(\theta) \quad \max_{x\in\mathbb{R}^n} \{r(x;\theta):\ Bx = C_B/T\},$$

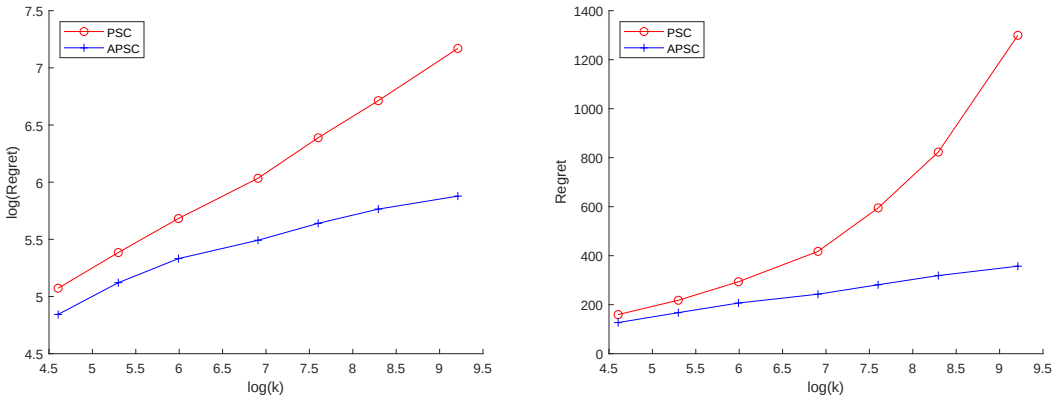
whose optimal primal and dual solutions are denoted by $x^D(\theta)$ and $\nu^D(\theta)$ respectively. Not surprisingly, to achieve (ii), we use $\lambda^D(\hat{\theta}_{t_1})$ as our initial primal solution x_1^{NT}, and use the formula proposed in Boyd and Vandenberghe (2004) to compute an initial dual solution ν_1^{NT}. (Naturally, since $\nabla_\lambda r(x^D(\hat{\theta}_L);\hat{\theta}_L) = B'\nu^D(\hat{\theta}_L)$ must hold at the optimal primal and dual solution of $\mathrm{ECP}(\hat{\theta}_L)$, this suggests that we use $\nu_1^{NT} = (BB')^{-1}B\nabla_\lambda r(x_1^{NT};\hat{\theta}_L)$.)

Re-calibration step (Stage 2 Step c). At the end of each segment $z \geq 1$, based on the new estimate $\hat{\theta}_{t_{z+1}}$, the re-calibration step is conducted to compute an approximation of $\lambda^D(\theta^*)$, i.e., λ_{z+1}^{NT}: It takes the primal and dual solution in the current segment x_z^{NT} and μ_z^{NT} as input and apply (6) to get a new pair of primal and dual solution for the next segment x_{z+1}^{NT} and μ_{z+1}^{NT} (ideally, the seller would like to use (8); but since it is not observable, APSC uses (6) as an approximation), and then uses $\lambda_{z+1}^{NT} := x_{z+1}^{NT}$ as the new approximation of $\lambda^D(\theta^*)$. Note that this step does not require any optimization. Despite its simplicity, our next theorem shows that APSC achieves logarithmic-squared regret when demand is well-separated.

THEOREM 2. (LOGARITHMIC-SQUARED REGRET OF APSC) *Suppose that* ***A1-A7*** *hold. Fix any* $\epsilon > 0$ *and set* $L = \lceil \log^{1+\epsilon}(kT) \rceil$ *and* $\eta = \log^{-\epsilon/4} k$. *There exists a constant* $M_2 > 0$ *independent of* $k \geq 3$ *such that* $\rho^{APSC}(k) \leq M_2 [\log^{1+\epsilon} k + (q-1)\log^2 k]$.

The result shows that in the NRM setting, when the conditions for effective demand learning becomes less restrictive due to the well-separated structure, the seller can leverage this via APSC and obtain significantly smaller regret. As shown in Figure 1, APSC does perform much better

Figure 1 Regret comparison between PSC and APSC



Note: The setting of the example that generates the plots above is the same as in the example in Table 2 with initial inventory of $(3,5,7)\times k$, except that the only unknown parameters are the intercepts (i.e., a well-separated demand setting). The exploration prices for PSC and APSC are the same as in examples in Table 2. For APSC, we use $\epsilon=0.5$ and determine tuning parameters L,η according to Theorem 2. The line of best fit in the log-log plot (left) of regret versus k of PSC has slope 0.45, indicating that the regret of PSC is approximately in the order of $\sqrt{k}$. The regret of APSC versus log(k) is approximately linear, indicating that the regret of APSC is approximately logarithmic.

than PSC when the underlying demand function is well-separated; moreover, the near-linear trend of the regret of APSC (resp. PSC) in the Regret$-\log(k)$ (resp. $\log(\text{Regret})-\log(k)$) plot is also consistent with the analytical regret bounds we derive for APSC (resp. PSC).

Remark 4. Broder and Rusmevichientong (2012) have established that, under the well-separated case with one unknown parameter, the best achievable lower bound on the performance of any admissible pricing control in the *uncapacitated* single product case is $\Omega(\log k)$ and this bound is achievable by a heuristic called MLE-GREEDY. An open research question is whether this bound is also achievable in the NRM case with multiple resource constraints and well-separated demand. Our result gives a partial answer. We show that the regret of APSC is worse than $\mathcal{O}(\log k)$ by a factor of $\log k$. However, when $q=1$ (i.e., single unknown parameter), the regret of APSC is $\mathcal{O}(\log^{1+\epsilon} k)$. Since ϵ can be chosen to be arbitrarily small, APSC almost attains the best achievable performance bound for the special case when $q=1$.

5. Closing Remarks

We develop two pricing heuristics to solve the problem of joint learning and pricing for NRM with multiple products and multiple resources. By establishing a $\mathcal{O}(\sqrt{k})$ regret bound for our first heuristic PSC, we show, for the first time in the literature, that it is possible to construct a rate-optimal heuristic for the NRM setting with a general parametric demand model and a continuum set of feasible price vectors. Our second heuristic APSC is the first heuristic in the literature

that deals with the NRM setting with a parametric demand model which also satisfies an extra well-separated condition; APSC achieves a much sharper $\mathcal{O}(\log^2 k)$ regret bound, which is close to rate-optimal (up to a multiplicative logarithmic term). These strong analytical bounds indicate that the design features in our proposed heuristics can be powerful ideas for developing effective pricing policies in practice, and also highlights the potential benefit of leveraging structural properties of underlying demand models to achieve better performance. It would be interesting to see whether our algorithms can also be extended to a more general setting with non-stationary or dynamically changing parameters. We leave this for future research.

Endnotes

1. Note that all the performance bounds established in Ferreira et al. (2018) are in terms of Bayesian regret which is the average regret for all $\theta \in \Theta$ weighted by the seller's prior on Θ. This means that the Bayesian regret is always no larger than the worst-case regret over all $\theta \in \Theta$.

Acknowledgments

The authors thank the area editor, the anonymous associate editor, and three anonymous referees for their insightful comments and suggestions that have significantly improved this paper.

References

Araman V, Caldentey R (2009) Dynamic pricing for nonperishable products with demand learning. *Oper. Res.* 57:1169–1188.

Aviv Y, Pazgal A (2005) Dynamic pricing of short life-cycle products through active learning. *Working paper* .

Badanidiyuru A, Kleinberg R, Slivkins A (2018) Bandits with knapsacks. *Journal of the ACM* 65(3):13.

Besbes O, Zeevi A (2009) Dynamic pricing without knowing the demand function: Risk bound and near-optimal algorithms. *Oper. Res.* 57:1407–1420.

Besbes O, Zeevi A (2012) Blind network revenue management. *Oper. Res.* 60:1537–1550.

Boyd S, Vandenberghe L (2004) *Convex Optimization* (Cambridge University Press).

Broder J, Rusmevichientong P (2012) Dynamic pricing under a general parametric choice model. *Oper. Res.* 60:965–980.

Chen N, Gallego G (2019) A primal-dual learning algorithm for personalized dynamic pricing with an inventory constraint. *Working Paper* .

Chen QG, Jasin S, Duenyas I (2019) Nonparametric self-adjusting control for joint learning and optimization of multi-product pricing with finite resource capacity. *Math. Oper. Res.* 44:601–631.

Chen Y, Farias VF (2013) Simple policies for dynamic pricing with imperfect forecasts. *Oper. Res.* 61:612–624.

den Boer AV (2015) Dynamic pricing and learning: Historical origins, current research, and new directions. *Surveys in Operations Research and Management Science* 20:1–18.

Farias VF, van Roy B (2010) Dynamic pricing with a prior on market response. *Oper. Res.* 58:16–29.

Ferreira KJ, Simchi-Levi D, Wang H (2018) Online network revenue management using thompson sampling. *Oper. Res.* 66:1586–1602.

Jasin S (2014) Reoptimization and self-adjusting price control for network revenue management. *Oper. Res.* 62:1168–1178.

Koushik D, Higbie JA, Eister C (2012) Retail price optimization at intercontinental hotels group. *Interface* 42:45 – 57.

Lei Y, Jasin S, Sinha A (2014) Near-optimal bisection search for nonparametric dynamic pricing with inventory constraint. *Working paper* .

Pekgun P, Menich RP, Acharya S, Finch PG, Deschamps F, Mallery K, van Sistine J, Christianson K, Fuller J (2013) Carlson rezidor hotel group maximizes revenue through improved demand management and price optimization. *Interface* 43:21 – 36.

Pronzato L, Pázman A (2013) *Design of experiments in nonlinear models : asymptotic normality, optimality criteria and small-sample properties* (Springer).

Talluri K, van Ryzin G (2005) *The theory and Practice of Revenue Management* (Springer).

Wang Z, Deng S, Ye Y (2014) Closing the gaps: A learning-while-doing algorithm for single-product revenue management problems. *Oper. Res.* 62:318–331.

Author Biographies

Qi (George) Chen is an Assistant Professor of Management Science and Operations at London Business School. His research focuses on the design of pricing strategies and mechanisms under both non-strategic uncertainty and strategic interactions, with applications in revenue management and pricing analytics, strategic sourcing and supply chain management, and online marketplaces.

Stefanus Jasin is an Associate Professor of Technology and Operations at the Ross School of Business, University of Michigan. He is interested in real-time pricing, e-commerce order fulfillment, assortment optimization, delivery consolidation, inventory optimization, joint learning and optimization, and optimization in on-demand market.

Izak Duenyas is the Herrick Professor of Business, a Professor of Technology and Operations at the Ross School of Business, and a Professor of Industrial and Operations Engineering at the University of Michigan. He is interested in supply chain management and coordination, revenue management in a variety of industries, evaluation of investment decisions in capacity, and in modeling and control of production systems.

Online Appendix to: *Joint Learning and Optimization of Multi-Product Pricing with Finite Resource Capacity and Unknown Demand Parameters*

Qi (George) Chen
London Business School, Regent's Park, London, NW1 4SA, gchen@london.edu

Stefanus Jasin, Izak Duenyas
Stephen M. Ross School of Business, University of Michigan, Ann Arbor, MI 75080, sjasin, duenyas@umich.edu

EC.1. Proof of Lemma 1

Proof: This proof is a multi-product extension of Lemma 3.7 in Broder and Rusmevichientong (2012). We first state an existing result in the literature below.

THEOREM EC.1. (TAIL INEQUALITY FOR MLE BASED ON IID SAMPLES, THEOREM 36.3 IN BOROVKOV (1999)) *Let $\Theta \in \mathbb{R}^q$ be compact and convex, and let $\{\mathbb{P}^\theta : \theta \subseteq \Theta\}$ be a family of distributions on a discrete sample space $\mathcal{Y}$. Suppose Y is a random variable taking value in $\mathcal{Y}$ with distribution $\mathbb{P}^\theta$, and the following conditions hold:*

(i) $\mathbb{P}^\theta \neq \mathbb{P}^{\theta'}$ whenever $\theta \neq \theta'$;

(ii) For some $r > q$, $\sup_{\theta\in\Theta} \mathbf{E}_\theta[||\nabla_\theta \log \mathbb{P}^\theta(Y)||_2^r] = \gamma < \infty$;

(iii) The function $\theta \to \sqrt{\mathbb{P}^\theta(Y)}$ is differentiable on Θ for any $Y \in \mathcal{Y}$;

(iv) The Fisher information matrix, whose $(i,j)^{th}$ entry is given by $\mathbf{E}_\theta\left[-\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log \mathbb{P}^\theta(Y)\right]$, is positive definite.

If $Y_1, Y_2, \ldots$ is a sequence of i.i.d. random variables taking value in $\mathcal{Y}$ with distribution $\mathbb{P}^\theta$, and $\hat{\theta}(t) = \arg\max_{\theta\in\Theta} \prod_{l=1}^t \mathbb{P}^\theta(Y_l)$ is the maximum likelihood estimate based on t i.i.d. samples, then, there exist constants $\eta_1 > 0$ and $\eta_2 > 0$ depending only on r, q, $\mathbb{P}^\theta$ and Θ such that for all $t \geq 1$ and all $\delta \geq 0, \mathbb{P}^\theta(||\hat{\theta}(t) - \theta||_2 > \delta) \leq \eta_1 \exp(-t\eta_2\delta^2)$.

To apply Theorem EC.1 to our setting, we simply need to verify conditions (i)-(iv). First, note that Θ is a compact subset of $\mathbb{R}^q$ and $\mathcal{D}^{\tilde{q}}$ is a discrete-valued sample space. Conditions (i) and (iv) are immediately satisfied because of **A5**-i and **A5**-iii. As for conditions (ii) and (iii), recall that

$$\begin{aligned}
\left|\left|\nabla_\theta \log \mathbb{P}^{\tilde{p},\theta}(D_{1:\tilde{q}})\right|\right|_2 &= \left|\left|\sum_{s=1}^{\tilde{q}}\left[\left(1-\sum_{j=1}^n D_{s,j}\right)\nabla_\theta \log\left(1-\sum_{j=1}^n \lambda_j(\tilde{p}^{(s)};\theta)\right) + \sum_{j=1}^n D_{s,j}\nabla_\theta \log \lambda_j(\tilde{p}^{(s)};\theta)\right]\right|\right|_2 \\
&\leq \sum_{s=1}^{\tilde{q}}\left(\left|\left|\nabla_\theta \log\left(1-\sum_{j=1}^n \lambda_j(\tilde{p}^{(s)};\theta)\right)\right|\right|_2 + \sum_{j=1}^n \left|\left|\nabla_\theta \log \lambda_j(\tilde{p}^{(s)};\theta)\right|\right|_2\right).
\end{aligned}$$

By **A1** and **A5**-ii, for all $1 \le s \le \tilde{q}$ and $1 \le j \le n$, $\lambda_j(\tilde{p}^{(s)};.) \in \mathsf{C}^1(\Theta)$ and is bounded away from zero, and $\sum_{j=1}^n \lambda_j(\tilde{p}^{(s)};.) \in \mathsf{C}^1(\Theta)$ is also bounded away from one. These imply that $||\nabla_\theta \log\left(1-\sum_{j=1}^n \lambda_j(\tilde{p}^{(s)};.)\right)||_2$ and $||\nabla_\theta \log \lambda_j(\tilde{p}^{(s)};.)||_2$, $j=1,\ldots,n$, are both continuous functions of θ for $s=1,\ldots,\tilde{q}$ and are, due to compactness of Θ, bounded. So, (ii) follows. As for (iii), note that $\mathbb{P}^{\tilde{p},\theta}(D_{1:\tilde{q}})$ is continuous in θ and it is also bounded away from zero. (In fact, $\mathbb{P}^{\tilde{p},\theta}(D_{1:\tilde{q}}) \ge [\lambda_{\min}^n(1-\lambda_{\max})]^{\tilde{q}}$ by **A5**-ii.) So, $\theta \to \sqrt{\mathbb{P}^{\tilde{p},\theta}(D_{1:\tilde{q}})}$ is differentiable on Θ for all $D_{1:\tilde{q}} \in \mathcal{D}^{\tilde{q}}$. We have thus verified all the conditions of Theorem EC.1. Then, a direct application of Theorem EC.1 leads to $\mathbb{P}_\theta^\pi(||\theta-\hat{\theta}_L||_2 > \delta) \le \eta_1 \exp(-\eta_2 L\delta^2)$. Also, since $\mathbf{E}_\theta^\pi\left[||\theta-\hat{\theta}_L||_2^2\right] = \int_0^\infty \mathbb{P}_\theta^\pi(||\theta-\hat{\theta}_L||_2^2 \ge x)dx = \int_0^\infty \mathbb{P}_\theta^\pi(||\theta-\hat{\theta}_L||_2 \ge \sqrt{x})dx \le \int_0^\infty \eta_1 e^{-\eta_2 L x}dx = \frac{\eta_1}{\eta_2 L}$, the result follows by taking $\eta_3 = \sqrt{\eta_1/\eta_2}$. □

EC.2. Proof of Theorem 1

Proof: We first establish a stability result on the optimization $P(\theta)$ stated below which we prove at the end of this Section.

Lemma EC.1. *There exist constants $\kappa > 0$ and $\bar{\delta} > 0$ independent of $k > 0$, such that for all $\theta \in \mathsf{Ball}(\theta^*, \bar{\delta})$,*

a. $p^D(\theta) \in \mathsf{Ball}(p^D(\theta^*), \phi/2)$, $\mathsf{Ball}(p^D(\theta), \phi/2) \subseteq \mathcal{P}$ and $||\lambda^D(\theta^*) - \lambda^D(\theta)||_2 \le \kappa||\theta^* - \theta||_2$,

b. there exists an optimal dual solution $\mu^D(\theta^*)$ of $P_\lambda(\theta^*)$, such that $A_i\lambda^D(\theta) = C_i$ for all $i \in \{j : \mu_j^D(\theta^*) > 0\}$.

Fix $\pi = PSC$. Theorem 1 can be established by a similar argument as in the proof of Theorem 1 in Chen et al. (2019) by replacing Proposition 1, Lemma 2 and Lemma 3 in Chen et al. (2019) by Lemma EC.1 in this paper, and replacing Lemma 1 in Chen et al. (2019) by Lemma 1 in this paper. All the proof arguments in Chen et al. (2019) can essentially be either simplified or directly followed with minor changes. Let $\epsilon(L) := \mathbf{E}_{\theta^*}^\pi[||\theta^* - \hat{\theta}_L||_2^2]^{1/2}$. Then, the last inequality in Section 5.1 in Chen et al. (2019) reduces to the following:

$$\rho^\pi(k) \le M_8\left(\epsilon(L)^2 k + \epsilon(L)^{-1}\log k + \epsilon(L)^{-2} + \bar{r}L\right) \le M_8\left(\frac{\eta_3^2 k}{L} + \frac{\sqrt{L}}{\eta_3}\log k + \frac{L}{\eta_3^2} + \bar{r}L\right).$$

Setting $L = \lceil\sqrt{kT}\rceil$ leads to $\rho^\pi(k) \le M_1\sqrt{k}$ for some M_1 independent of k. □

Proof of Lemma EC.1. Let $\bar{\delta} = \min\{\delta_1, \delta_2\}$ where δ_1 and δ_2 are strictly positive constants to be defined shortly. We will prove each part of the lemma in turn.

Proof of part (a). This is an immediate corollary of Proposition 1 in Chen et al. (2019). Note that, by **A2**, we have $||\lambda(p;\theta^*) - \lambda(p;\theta)||_\infty \le ||\lambda(p;\theta^*) - \lambda(p;\theta)||_2 \le \omega||\theta^* - \theta||_2$ and $||(\nabla\lambda(p;\theta^*) - \nabla\lambda(p;\theta))'||_\infty = \max_{1\le i\le n}\sum_{j=1}^n |\frac{\partial\lambda_i}{\partial p_j}(p;\theta) - \frac{\partial\lambda_i}{\partial p_j}(p;\theta^*)| \le n\omega||\theta^* - \theta||_2$ for all $\theta \in \Theta, p \in \mathcal{P}$. Hence,

$||\lambda(.;\theta^*) - \lambda(.;\theta)||_\infty = \sup_{p\in\mathcal{P}} ||\lambda(p;\theta^*) - \lambda(p;\theta)||_\infty \le \omega||\theta^* - \theta||_2$ and $||(\nabla\lambda(.;\theta^*) - \nabla\lambda(.;\theta))'||_\infty = \sup_{p\in\mathcal{P}} ||(\nabla\lambda(.;\theta^*) - \nabla\lambda(.;\theta))'||_\infty \le n\omega||\theta^* - \theta||_2$. Therefore, by Proposition 1 in Chen et al. (2019), there exists some $K_1 > 0$ independent of $k > 0$ such that $||p^D(\theta^*) - p^D(\theta)||_\infty \le K_1\omega||\theta^* - \theta||_2$. Let $\delta_1 = \phi(2n^{1/2}K_1)^{-1}$. For all θ satisfying $||\theta - \theta^*||_2 \le \bar{\delta} \le \delta_1$, we have $||p^D(\theta^*) - p^D(\theta)||_2 \le n^{1/2}||p^D(\theta^*) - p^D(\theta)||_\infty \le n^{1/2}K_1\delta_1 \le \phi/2$. Hence, $p^D(\theta) \in \mathsf{Ball}(p^D(\theta^*), \phi/2)$. Since $\mathsf{Ball}(p^D(\theta^*), \phi) \subseteq \mathcal{P}$ by **A6**, we conclude that $\mathsf{Ball}(p^D(\theta), \phi/2) \subseteq \mathcal{P}$. Since $\lambda(.;\theta^*)$ is continuously differentiable with respect to $p \in \mathcal{P}$ as implied by **A1**, and $\mathcal{P}$ is compact, there exists a constant $K_2 > 0$ independent of $k > 0$ such that

$$\begin{aligned}
||\lambda^D(\theta^*) - \lambda^D(\theta)||_2 &= ||\lambda(p^D(\theta^*);\theta^*) - \lambda(p^D(\theta);\theta)||_2 \\
&\le ||\lambda(p^D(\theta^*);\theta^*) - \lambda(p^D(\theta);\theta^*)||_2 + ||\lambda(p^D(\theta);\theta^*) - \lambda(p^D(\theta);\theta)||_2 \\
&\le K_2||p^D(\theta^*) - p^D(\theta)||_2 + \omega||\theta^* - \theta||_2 \le (\omega + n^{1/2}K_1K_2)||\theta^* - \theta||_2,
\end{aligned}$$

where the second inequality also follows by **A2**. Part (a) follows by letting $\kappa = \omega + n^{1/2}K_1K_2$.

Proof of part (b). Take any $\theta \in \Theta$. Recall that strong duality holds since $\mathrm{P}_\lambda(\theta)$ is a strongly concave optimization problem, and $\lambda^D(\theta)$ is its unique optimal solution. By the Karush-Kuhn-Tucker (KKT) condition, there exists a (not necessarily unique) dual solution $\mu \in \mathbb{R}^m_+$ such that

$$\mathbf{KKT}(\theta): \quad \nabla r(\lambda^D(\theta);\theta) = A'\mu \tag{EC.1}$$

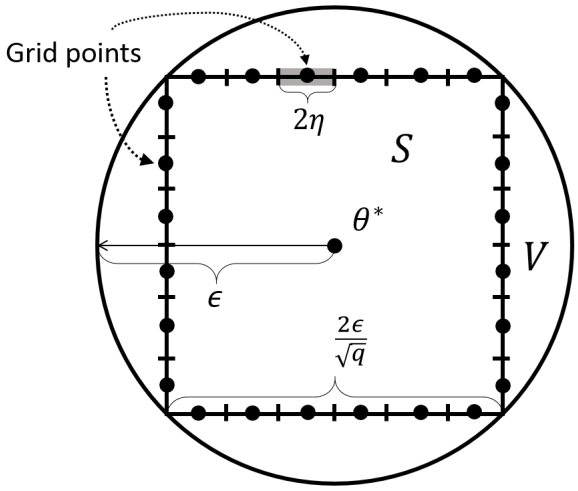
$$\mu_i(A_i\lambda^D(\theta) - C_i) = 0, \forall i = 1,\dots,m. \tag{EC.2}$$

Define $\mathcal{B}^* := \{j : A_j\lambda^D(\theta^*) = C_j\}$, $\mathcal{N}^* := \{j : A_j\lambda^D(\theta^*) < C_j\} = \{1,\dots,n\} - \mathcal{B}^*$, and $\Omega := \{\mathcal{I} \in \bar{\Omega} : \exists \mu \in \mathbb{R}^m_+, s.t., \nabla r(\lambda^D(\theta^*);\theta^*) = A'\mu, \text{ and } \mu_i = 0, \forall i \notin \mathcal{I}\}$ where $\bar{\Omega} := \{\mathcal{I} \subseteq \mathcal{B}^* : \{A_i\}_{i\in\mathcal{I}} \text{ are linearly independent}\}$. It can be easily verified that for any $\mathcal{I} \in \Omega$, there exists a dual solution μ of $\mathbf{KKT}(\theta^*)$ such that $\mu_i > 0$ only if $i \in \mathcal{I}$. Thus, to prove Part (b), we only need to show that there exists some $\delta_2 > 0$ such that for all $\theta \in \mathsf{Ball}(\theta^*, \delta_2)$, there exists $\mathcal{I} \in \Omega$ such that $A_i\lambda^D(\theta) = C_i$ for all $i \in \mathcal{I}$, which we prove below. To that end, we first state a claim.

Claim EC.1. *There exists some constant δ_3 such that for all $\theta \in \mathsf{Ball}(\theta^*, \delta_3)$, there exist a dual solution $\mu^\star(\theta)$ to $\boldsymbol{KKT}(\theta)$ such that: (i) $\{i : \mu^\star_i(\theta) > 0\} \subseteq \mathcal{B}^*$ and (ii) $\{A_i\}_{i:\mu^\star_i(\theta)>0}$ are linearly independent.*

Note that (i) holds since by Part (a), there exists some constant $\delta_3 \in (0, \delta_1]$ such that for all $\theta \in \mathsf{Ball}(\theta^*, \delta_3)$, $A_i\lambda^D(\theta) < C_i$ for all $i \in \mathcal{N}^*$, and hence, by (EC.2), any dual solution μ to $\mathbf{KKT}(\theta)$ must satisfy that $\mu_i = 0$ for $i \in \mathcal{N}^*$; (ii) follows since, if not, one can reduce one of $\mu^\star(\theta)$'s strictly positive component to zero while keeping the conditions (EC.1)-(EC.2) satisfied. This claim implies that for all $\theta \in \mathsf{Ball}(\theta^*, \delta_3)$, there exists solution $\mu^\star(\theta)$ to $\mathbf{KKT}(\theta)$ such that the only strictly

Figure EC.1 Geometric illustration of Lemma 2



Note: This illustrates the case when there are two parameters to estimate ($q=2$). V denotes the disk (ball) centered at θ^* with radius ϵ. Note that the event of $||\theta^* - \hat{\theta}_t||_2 > \epsilon$ corresponds to the event when $\hat{\theta}_t$ lies in the exterior of V. In this example, the surface of the rectangle(hypercube) S consists of four edges.

positive components of $\mu^\star(\theta)$ correspond to a subset of linearly independent rows in $\mathcal{B}^*$. Now, by Farkas' Lemma, for any $\mathcal{I} \in \bar{\Omega} - \Omega$, there exists a vector $v \in \mathbb{R}^n$ such that $v'\nabla r(\lambda^D(\theta^*);\theta^*) < 0$ and $A_i v \geq 0$ for all $i \in \mathcal{I}$; since $\nabla r(\lambda^D(\theta);\theta)$ is continuous in θ, we conclude that there exists a constant $\delta_2 \in (0, \delta_3)$ such that for all $\theta \in \mathsf{Ball}(\theta^*, \delta_2)$, $v'\nabla r(\lambda^D(\theta);\theta) < 0$. Then, by Farkas' Lemma and (ii) of the claim, for any $\theta \in \mathsf{Ball}(\theta^*, \delta_2)$, there exists a dual solution μ of $\mathbf{KKT}(\theta)$ such that $\{i; \mu_i > 0\} \in \Omega$. The result follows by (EC.2). $\square$

EC.3. Proof of Lemma 2

Proof. We first illustrate the main idea behind the proof using Figure EC.1. Note that $||\theta - \hat{\theta}_t||_2 > \epsilon$ is equivalent to the event that ML estimate $\hat{\theta}_t$ is outside of the ball $V := \mathsf{Ball}(\theta, \epsilon)$. In addition, under the concavity assumption of the log-likelihood, $\hat{\theta}_t \neq \mathsf{Ball}(\theta, \epsilon)$ implies that at least one point on the surface of a hypercube S, which is centered at θ and is a subset of V, has a larger log-likelihood than the log-likelihood at θ. The probability of this event is a valid upper bound of $\mathbb{P}_\theta^\pi(||\theta - \hat{\theta}_t||_2 > \epsilon)$. However, the challenge is that there are a continuum of such potential points. The idea of the proof is to consider a grid of points on the surface of that hypercube S, and the granularity of the grid is set to be fine enough so that any point on the surface of that hypercube can be closely approximated by one point on the grid. We will show that the existence of a point on the surface of S with a higher log-likelihood than the true parameter vector θ is extremely unlikely. We now rigorously prove this lemma.

Step 1

Fix some $0 < \tilde{\lambda}_{\min} < \tilde{\lambda}_{\max} < 1$. First, we will show that for all $D \in \mathcal{D}$, for all $p \in \mathcal{W}(\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max})$ and for all $\theta \in \Theta$, $\nabla_\theta \log \mathbb{P}_1^{p,\theta}(D)$ is jointly continuous in θ and p. Recall that $\nabla_\theta \log \mathbb{P}_1^{p,\theta}(D) = ((\partial/\partial\theta_1)\log \mathbb{P}_1^{p,\theta}(D); \ldots; (\partial/\partial\theta_n)\log \mathbb{P}_1^{p,\theta}(D))$ where for all $1 \le k \le n$,

$$\frac{\partial \log \mathbb{P}_1^{p,\theta}(D)}{\partial \theta_k} = -\frac{(1-\sum_{j=1}^n D_j)\log\left(1-\sum_{j=1}^n \lambda_j(p;\theta)\right)}{1-\sum_{j=1}^n \lambda_j(p;\theta)}\left(\sum_{j=1}^n \frac{\partial \lambda_j(p;\theta)}{\partial \theta_k}\right) + \sum_{j=1}^n \frac{D_j \log\left(\lambda_j(p;\theta)\right)}{\lambda_j(p;\theta)}\frac{\partial \lambda_j(p;\theta)}{\partial \theta_k}.$$

Since $\lambda_j(p;.) \in \mathsf{C}^1(\Theta)$ and $\lambda(.;\theta) \in \mathsf{C}^2(\mathcal{P})$ by **A1** and the denominators are strictly greater than zero, $\nabla_\theta \log \mathbb{P}_1^{p,\theta}(D)$ is jointly continuous in θ and p.

Step 2

Since Θ and $\mathcal{W}(\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max})$ are compact, $\mathcal{D}$ is finite and $\nabla_\theta \log \mathbb{P}_1^{p,\theta}(D)$ is jointly continuous in θ and p for all $D \in \mathcal{D}$, there exists a constant $c_g > 0$ independent of θ, p, D such that for all $\theta \in \Theta$, $p \in \mathcal{W}(\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max})$, and $v \in \mathbb{R}^q$ satisfying $||v||_2 = 1$, $(\nabla_\theta \log \mathbb{P}_1^{p,\theta}(D))'v < c_g$. Therefore, for any $v, ||v||_2 = 1$, if $p_s^\pi \in \mathcal{W}(\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max})$ for $1 \le s \le t$, then we have:

$$(\nabla_\theta \log \mathbb{P}_t^{\pi,\theta}(D_{1:t}))'v = \sum_{s=1}^t (\nabla_\theta \log \mathbb{P}_1^{p_s^\pi,\theta}(D_s))'v < c_g t. \tag{EC.3}$$

Now, fix $\epsilon > 0$ and consider a hypercube $S := \{x \in \mathbb{R}^q : -\epsilon \le x_i\sqrt{q} \le \epsilon, \forall i\}$ centered at the origin with edge $2\epsilon/\sqrt{q}$; we denote its surface by $\partial S := \cup_{j=1}^q \{x \in \mathbb{R}^q : -\epsilon \le x_i\sqrt{q} \le \epsilon, \forall i \ne j, |x_j| = \epsilon/\sqrt{q}\}$. Consider a subset of ∂S defined as $\cup_{j=1}^q \{x \in \mathbb{R}^q : |x_j| = \epsilon/\sqrt{q}, \text{ and } x_i = 2k_i\eta, \forall i \ne j, \forall k_i \in \mathbb{Z} \cap [-\frac{\epsilon}{2\sqrt{q}\eta}, \frac{\epsilon}{2\sqrt{q}\eta}]\}$: Note that this is a finite set, so we denote by N its cardinality and by v_j, $j = 1, \ldots, N$ all of its elements. Note that $N \le 2q(\epsilon/(\sqrt{q}\eta))^{q-1}$. Then, it is easy to verify that for any $x \in \partial S$, $\min_{j=1,\ldots,N} ||x - v_j||_2 \le \sqrt{q}\eta$. By W3, we have that for any $\theta' \in S + \theta$ and any $j = 1, \ldots, N$,

$$\log \mathbb{P}_t^{\pi,\theta'}(D_{1:t}) - \log \mathbb{P}_t^{\pi,\theta+v_j}(D_{1:t}) \le (\nabla_\theta \log \mathbb{P}_t^{\pi,\theta+v_j}(D_{1:t}))'(\theta' - \theta - v_j)$$

Let $j^*(\theta) = \arg\min_{j=1,\ldots,N} ||\theta - \theta - v_j||_2$. We then have

$$\log \mathbb{P}_t^{\pi,\theta'}(D_{1:t}) - \log \mathbb{P}_t^{\pi,\theta+v_{j^*(\theta')}}(D_{1:t}) \le c_g t||\theta' - \theta - v_{j^*(\theta')}||_2 \le c_g\sqrt{q}\eta t. \tag{EC.4}$$

where the first inequality follows by (EC.3). The following is the key argument for this proof:

$$\left\{||\hat{\theta}_t - \theta||_2 > \epsilon\right\} \subseteq \left\{||\hat{\theta}_t - \theta||_\infty > \frac{\epsilon}{\sqrt{q}}\right\}$$

$$
\begin{aligned}
&\subseteq \left\{\log \mathbb{P}_t^{\pi,\theta+v}(D_{1:t}) \geq \log \mathbb{P}_t^{\pi,\theta}(D_{1:t}), \text{for some } v \text{ with } ||v||_\infty = \tfrac{\epsilon}{\sqrt{q}}\right\} \\
&\subseteq \left\{\log \mathbb{P}_t^{\pi,\theta+v_{j^*(\theta+v)}}(D_{1:t}) + c_g\sqrt{q}\eta t \geq \log \mathbb{P}_t^{\pi,\theta}(D_{1:t}), \text{for some } v \text{ with } ||v||_\infty = \tfrac{\epsilon}{\sqrt{q}}\right\} \\
&\subseteq \cup_{j=1}^N \left\{\log \mathbb{P}_t^{\pi,\theta+v_j}(D_{1:t}) + c_g\sqrt{q}\eta t \geq \log \mathbb{P}_t^{\pi,\theta}(D_{1:t})\right\} \\
&= \cup_{j=1}^N \left\{Z_t^\pi(v_j, D_{1:t}) \geq \exp(-c_g\sqrt{q}\eta t)\right\}, \qquad \text{(EC.5)}
\end{aligned}
$$

where $Z_t^{\pi,\theta}(u, D_{1:t}) := \mathbb{P}_t^{\pi,\theta+u}(D_{1:t})/\mathbb{P}_t^{\pi,\theta}(D_{1:t})$ is the likelihood ratio for any $u \in \Theta - \theta$. The first inclusion follows by norm inequality, the second inclusion follows by the concavity of the log-likelihood function and the definition of ML estimator, the third inclusion follows by (EC.4), the fourth inequality follows because by definition $j^*(\theta + v) \in \{1, \ldots, N\}$ for all v.

Step 3

To use (EC.5) to prove Lemma 2, we state a lemma below which we prove at the end of this section:

LEMMA EC.2. *Fix some $0 < \tilde{\lambda}_{\min} < \tilde{\lambda}_{\max} < 1$. Suppose that an admissible control π satisfies $p_s = \pi_s(D_{1:s-1}) \in \mathcal{W}(\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max})$ for all $1 \leq s \leq t$. Then there exists a constant $c_h > 0$ independent of t and $\theta \in \Theta$ such that for all π and for all $u \in \Theta - \theta$, $\mathbf{E}_\theta^\pi[\sqrt{Z_t^{\pi,\theta}(u, D_{1:t})}] \leq \exp(-c_h||u||_2^2 t/2)$.*

By (EC.5) and Lemma EC.2, the following holds

$$
\begin{aligned}
\mathbb{P}_\theta^\pi\left(||\hat{\theta}_t - \theta||_2 > \epsilon\right) &\leq \sum_{j=1}^N \mathbb{P}_\theta^\pi\left(Z_t^{\pi,\theta}(v_j, D_{1:t}) \geq \exp(-c_g\sqrt{q}\eta t)\right) \\
&\leq \sum_{j=1}^N \exp\left(\frac{c_g\sqrt{q}\eta t}{2}\right) \mathbf{E}_\theta^\pi\left[\sqrt{Z_t^{\pi,\theta}(v_j, D_{1:t})}\right] \\
&\leq \sum_{j=1}^N \exp\left(\frac{c_g\sqrt{q}\eta t}{2} - \frac{c_h||v_j||_2^2 t}{2}\right) \\
&\leq 2q\left(\frac{\epsilon}{\sqrt{q}\eta}\right)^{q-1} \exp\left(-\frac{c_h\epsilon^2 t}{2q} + \frac{c_g\sqrt{q}\eta t}{2}\right),
\end{aligned}
$$

where the second inequality follows by the Markov's inequality, the third inequality follows by Lemma EC.2, and the last inequality follows because $N \leq 2q(\epsilon/(\sqrt{q}\eta))^{q-1}$ and $\min_{j=1,\ldots,N}||v_j||_2 \geq \min_{j=1,\ldots,N}||v_j||_\infty \geq \epsilon/\sqrt{q}$. Now, let $\eta = \epsilon/t$, then we have

$$
\mathbb{P}_\theta^\pi\left(||\hat{\theta}_t - \theta||_2 > \epsilon\right) \leq \min\left\{1, 2q^{\frac{3-q}{2}} t^{q-1} \exp\left(-\frac{c_h\epsilon^2 t}{2q} + \frac{c_g\sqrt{q}\epsilon}{2}\right)\right\}.
$$

Note that when $\epsilon \leq 1, \exp((-c_h\epsilon^2 q^{-1}t + c_g\sqrt{q}\epsilon)/2) \leq \exp(c_g\sqrt{q}/2)\exp(-c_h\epsilon^2 q^{-1}t/4)$. Note also that when $\epsilon > 1$, there exists $M > 0$ independent of ϵ such that $\exp((-c_h\epsilon^2 q^{-1}t + c_g\sqrt{q}\epsilon)/2) \leq \exp(-c_h\epsilon^2 q^{-1}t/4), \forall t > M$. With these two observations, we consider two cases below.

Case 1: $t > M$. In this case, we have $\mathbb{P}_\theta^\pi\left(||\hat{\theta}_t - \theta||_2 > \epsilon\right) \leq \tilde{\eta}_4 t^{q-1}\exp(-\eta_5 t\epsilon^2)$, where $\tilde{\eta}_4 = 2q^{(3-q)/2}\max\{1, \exp(c_g\sqrt{q}/2)\}$, and $\eta_5 = c_h q^{-1}/4$.

Case 2: $t \le M$. Let $\bar{\theta}$ be the largest distance between any two points in Θ. ($\bar{\theta} < \infty$ because Θ is bounded.) Then, we claim that for this case, $\mathbb{P}_\theta^\pi\left(||\hat{\theta}_t - \theta||_2 > \epsilon\right) \le \bar{\eta}_4 t^{q-1}\exp(-\eta_5 t\epsilon^2)$ where η_5 is defined as in Case 1 and $\bar{\eta}_4 = \exp(\eta_5 M\bar{\theta}^2)$. The claim is true because: if $\epsilon > \bar{\theta}$, $\mathbb{P}_\theta^\pi\left(||\hat{\theta}_t - \theta||_2 > \epsilon\right) = 0$, so the bound holds; if $\epsilon \le \bar{\theta}$, $\mathbb{P}_\theta^\pi\left(||\hat{\theta}_t - \theta||_2 > \epsilon\right) \le 1 = \bar{\eta}_4\exp(-\eta_5 M\bar{\theta}^2) \le \bar{\eta}_4 t^{q-1}\exp(-\eta_5 t\epsilon^2)$.

Combining the two cases above yields $\mathbb{P}_\theta^\pi\left(||\hat{\theta}_t - \theta||_2 > \epsilon\right) \le \min\{1, \eta_4 t^{q-1}\exp(-\eta_5 t\epsilon^2)\}$ where $\eta_4 = \max\{\tilde{\eta}_4, \bar{\eta}_4\}$. Hence,

$$\begin{aligned}
\mathbf{E}_\theta^\pi\left[||\hat{\theta}_t - \theta||_2^2\right] &= \int_0^\infty \mathbb{P}_\theta^\pi\left(||\hat{\theta}_t - \theta||_2^2 \ge x\right) dx \\
&= \int_0^\infty \min\left\{1, \eta_4 t^{q-1}\exp\left(-\eta_5 t x\right)\right\} dx \\
&\le \int_0^{\frac{2(q-1)\log t}{\eta_5 t}} dx + \int_{\frac{2(q-1)\log t}{\eta_5 t}}^\infty \left[\eta_4 t^{q-1}\exp\left(-\frac{\eta_5 t x}{2}\right)\right]\exp\left(-\frac{\eta_5 t x}{2}\right) dx \\
&\le \frac{2(q-1)\log t}{\eta_5 t} + \eta_4 \int_{\frac{2(q-1)\log t}{\eta_5 t}}^\infty \exp\left(-\frac{\eta_5 t x}{2}\right) dx \\
&\le \frac{2(q-1)\log t}{\eta_5 t} + \frac{2\eta_4}{\eta_5 t} \\
&\le \frac{2\max\{1,\eta_4\}}{\eta_5}\frac{(q-1)\log t + 1}{t}
\end{aligned}$$

where the first inequality holds because for all $x \ge \frac{2(q-1)\log t}{\eta_5 t}$, $t^{q-1}\exp\left(-\frac{\eta_5 t x}{2}\right) \le 1$. We complete the proof by letting $\eta_6 = \sqrt{2\max\{1,\eta_4\}/\eta_5}$. □

Proof of Lemma EC.2. Recall that $\mathcal{D} = \{D \in \{0,1\}^n : \sum_{j=1}^n D_j \le 1\}$. We define the conditional Hellinger distance as follows:

$$H_t^\pi(\theta_1, \theta_2, D_t | D_{1:t-1}) := \sum_{D_t \in \mathcal{D}}\left(\sqrt{\mathbb{P}_t^{\pi,\theta_1}(D_t|D_{1:t-1})} - \sqrt{\mathbb{P}_t^{\pi,\theta_2}(D_t|D_{1:t-1})}\right)^2.$$

We state a lemma and postpone its proof to the end of the proof of Lemma EC.2.

Lemma EC.3. *Fix some $0 < \tilde{\lambda}_{\min} < \tilde{\lambda}_{\max} < 1$. Suppose that an admissible control π satisfies $p_s = \pi_s(D_{1:s-1}) \in \mathcal{W}(\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max})$ for all $1 \le s \le t$. Then there exists a positive constant c_h independent of t such that $H_t^\pi(\theta_1, \theta_2, D_t|D_{1:t-1}) \ge c_h||\theta_1 - \theta_2||_2^2$ for all $\theta_1, \theta_2 \in \Theta$.*

For $u \in \Theta - \theta^*$, define $Z_t^{\pi,\theta}(u, D_t|D_{1:t-1}) := \mathbb{P}_t^{\pi,\theta+u}(D_t|D_{1:t-1})/\mathbb{P}_t^{\pi,\theta}(D_t|D_{1:t-1})$. By Lemma EC.3, we can derive a bound for its moment as follows:

$$\begin{aligned}
\mathbf{E}_\theta^\pi\left[\sqrt{Z_t^{\pi,\theta}(u, D_t|D_{1:t-1})}\right] &= \sum_{D_t\in\mathcal{D}}\sqrt{\frac{\mathbb{P}_t^{\pi,\theta+u}(D_t|D_{1:t-1})}{\mathbb{P}_t^{\pi,\theta}(D_t|D_{1:t-1})}}\,\mathbb{P}_t^{\pi,\theta}(D_t|D_{1:t-1}) \\
&= \sum_{D_t\in\mathcal{D}}\sqrt{\mathbb{P}_t^{\pi,\theta+u}(D_t|D_{1:t-1})\mathbb{P}_t^{\pi,\theta}(D_t|D_{1:t-1})}
\end{aligned}$$

$$= 1 - \frac{H_t^\pi(\theta, \theta+u, D_t|D_{1:t-1})}{2}$$
$$\leq \exp\left(-\frac{H_t^\pi(\theta, \theta+u, D_t|D_{1:t-1})}{2}\right) \leq \exp\left(-\frac{c_h||u||_2^2}{2}\right).$$

The result of Lemma EC.2 can now be proven by repeated conditioning: by definition,

$$\begin{aligned}
\mathbf{E}_\theta^\pi\left[\sqrt{Z_t^{\pi,\theta}(u, D_{1:t})}\right] &= \mathbf{E}_\theta^\pi\left[\mathbf{E}_\theta^\pi\left[\sqrt{Z_t^{\pi,\theta}(u, D_{1:t})}\middle| D_{1:t-1}\right]\right] \\
&= \mathbf{E}_\theta^\pi\left[\sqrt{Z_{t-1}^{\pi,\theta}(u, D_{1:t-1})}\,\mathbf{E}_\theta^\pi\left[\sqrt{Z_t^{\pi,\theta}(u, D_t|D_{1:t-1})}\right]\right] \\
&\leq \mathbf{E}_\theta^\pi\left[\sqrt{Z_{t-1}^{\pi,\theta}(u, D_{1:t-1})}\right]\exp\left(-\frac{c_h||u||_2^2}{2}\right) \\
&\leq \exp\left(-\frac{c_h||u||_2^2 t}{2}\right). \quad \square
\end{aligned}$$

Proof of Lemma EC.3. Note that, for any $\theta_1, \theta_2 \in \Theta, \theta_1 \neq \theta_2$, by Fatou's lemma, we have

$$\begin{aligned}
\liminf_{\theta'\to\theta_1, \theta''\to\theta_2} \frac{H_t^\pi(\theta', \theta'', D_t|D_{1:t-1})}{||\theta'-\theta''||_2^2} &= \liminf_{\theta'\to\theta_1, \theta''\to\theta_2} \sum_{D_t\in\mathcal{D}} \frac{\left(\sqrt{\mathbb{P}_t^{\pi,\theta'}(D_t|D_{1:t-1})} - \sqrt{\mathbb{P}_t^{\pi,\theta''}(D_t|D_{1:t-1})}\right)^2}{||\theta'-\theta''||_2^2} \\
&\geq \sum_{D_t\in\mathcal{D}} \liminf_{\theta'\to\theta_1, \theta''\to\theta_2} \frac{\left(\sqrt{\mathbb{P}_t^{\pi,\theta'}(D_t|D_{1:t-1})} - \sqrt{\mathbb{P}_t^{\pi,\theta''}(D_t|D_{1:t-1})}\right)^2}{||\theta'-\theta''||_2^2} \\
&= \frac{H_t^\pi(\theta_1, \theta_2, D_t|D_{1:t-1})}{||\theta_1-\theta_2||_2^2} \;>\; 0, \qquad \text{(EC.6)}
\end{aligned}$$

where the last inequality follows by **A7**-i. Let $\underline{\sigma}(.)$ denote the smallest eigenvalues of a real symmetric matrix. If we now set $\theta_1 = \theta_2 = \theta$, since $\sqrt{\mathbb{P}_t^{\pi,\theta}(D_t|D_{1:t-1})}$ is continuously differentiable in θ, there exists $\tilde{\theta}$ on the line segment connecting θ' and θ'' such that

$$\begin{aligned}
&\liminf_{\theta'\to\theta, \theta''\to\theta} \frac{H_t^\pi(\theta', \theta'', D_t|D_{1:t-1})}{||\theta'-\theta''||_2^2} \\
&\quad\geq \sum_{D_t\in\mathcal{D}} \liminf_{\theta'\to\theta, \theta''\to\theta} \left[\left(\frac{\partial}{\partial\theta}\sqrt{\mathbb{P}_t^{\pi,\tilde{\theta}}(D_t|D_{1:t-1})}\right)' \frac{\theta'-\theta''}{||\theta'-\theta''||_2}\right]^2 \\
&\quad= \sum_{D_t\in\mathcal{D}} \liminf_{\theta'\to\theta, \theta''\to\theta} \frac{(\theta'-\theta'')'}{||\theta'-\theta''||_2}\left(\frac{\partial}{\partial\theta}\sqrt{\mathbb{P}_t^{\pi,\tilde{\theta}}(D_t|D_{1:t-1})}\right)\left(\frac{\partial}{\partial\theta}\sqrt{\mathbb{P}_t^{\pi,\tilde{\theta}}(D_t|D_{1:t-1})}\right)' \frac{\theta'-\theta''}{||\theta'-\theta''||_2} \\
&\quad\geq \sum_{D_t\in\mathcal{D}} \liminf_{\theta'\to\theta, \theta''\to\theta} \underline{\sigma}\left(\left(\frac{\partial}{\partial\theta}\sqrt{\mathbb{P}_t^{\pi,\tilde{\theta}}(D_t|D_{1:t-1})}\right)\left(\frac{\partial}{\partial\theta}\sqrt{\mathbb{P}_t^{\pi,\tilde{\theta}}(D_t|D_{1:t-1})}\right)'\right) \\
&\quad= \sum_{D_t\in\mathcal{D}} \underline{\sigma}\left(\left(\frac{\partial}{\partial\theta}\sqrt{\mathbb{P}_t^{\pi,\theta}(D_t|D_{1:t-1})}\right)\left(\frac{\partial}{\partial\theta}\sqrt{\mathbb{P}_t^{\pi,\theta}(D_t|D_{1:t-1})}\right)'\right) \\
&\quad= \sum_{D_t\in\mathcal{D}} \frac{\underline{\sigma}\left(\left(\frac{\partial}{\partial\theta}\mathbb{P}_t^{\pi,\theta}(D_t|D_{1:t-1})\right)\left(\frac{\partial}{\partial\theta}\mathbb{P}_t^{\pi,\theta}(D_t|D_{1:t-1})\right)'\right)}{4\,\mathbb{P}_t^{\pi,\theta}(D_t|D_{1:t-1})}
\end{aligned}$$

$$= \frac{1}{4} \sum_{D_t \in \mathcal{D}} \underline{\sigma} \left(\left(\frac{\partial}{\partial \theta} \log \mathbb{P}_t^{\pi,\theta}(D_t|D_{1:t-1}) \right) \left(\frac{\partial}{\partial \theta} \log \mathbb{P}_t^{\pi,\theta}(D_t|D_{1:t-1}) \right)' \right) \mathbb{P}_t^{\pi,\theta}(D_t|D_{1:t-1})$$
$$\geq \frac{c_f}{4} > 0 \tag{EC.7}$$

where the first inequality follows by Fatou's Lemma as in (EC.6) and the Mean Value Theorem, and the third equality follows because

$$\frac{\partial}{\partial \theta} \sqrt{\mathbb{P}_t^{\pi,\theta}(D_t|D_{1:t-1})} = \frac{\frac{\partial}{\partial \theta} \mathbb{P}_t^{\pi,\theta}(D_t|D_{1:t-1})}{2\sqrt{\mathbb{P}_t^{\pi,\theta}(D_t|D_{1:t-1})}}$$

(by chain rule) and the last two inequalities follow by the definition of Fisher information and **A7**-ii. To prove Lemma EC.3, it suffices to show that, for any $\theta_1, \theta_2 \in \Theta, H_t^{\pi}(\theta_1, \theta_2, D_t|D_{1:t-1})/||\theta_1 - \theta_2||_2^2 \geq c_h$ for some $c_h > 0$ independent of θ_1, θ_2. (If $\theta_1 = \theta_2$, the ratio is to be understood as its limit.) Suppose not, since the ratio is always non-negative, there exist two sequences $\theta_1^n \to \theta_1, \theta_2^n \to \theta_2$ such that $\liminf_{n\to\infty} H_t^{\pi}(\theta_1^n, \theta_2^n, D_t|D_{1:t-1})/||\theta_1^n - \theta_2^n||_2^2 = 0$. But, this contradicts with (EC.6) when $\theta_1 \neq \theta_2$ and with (EC.7) when $\theta_1 = \theta_2$. □

EC.4. Proof of Theorem 2

Proof. Fix $\pi =$ APSC and let $k \geq 3$ throughout the proof. Without loss of generality, we will assume that $T = 1$ throughout. For national simplicity, we let $E(t) := ||\theta^* - \hat{\theta}_t||_2$. Set $L = \lceil (\log k)^{1+\epsilon} \rceil$ and $\eta = (\log k)^{-\epsilon/4}$. We first state an analog of Lemma EC.1(a) for ECP(θ) below whose proof is similar to the proof of Lemma EC.1 and so is omitted.

LEMMA EC.4. *There exist $\tilde{\delta} > 0$ and $\tilde{\kappa} > 0$ independent of $k > 0$ such that for all $\theta \in \mathsf{Ball}(\theta^*, \tilde{\delta})$, $||x^D(\theta^*) - x^D(\theta)||_2 \leq \tilde{\kappa}||\theta^* - \theta||_2$.*

We now proceed to prove Theorem 2 in several steps.

Step 1

We first show that the event $\mathcal{E} := \{\mathcal{B} \subseteq \{i : A_i\lambda^D(\theta^*) = C_i\}, \{A_i\}_{i\in\mathcal{B}}$ is a basis of $\{A_i\}_{\{i:A_i\lambda^D(\theta^*)=C_i\}}\}$ occurs with a very high probability, where $\mathcal{B}$ is identified in Stage 1 Step e of APSC. By construction of $\mathcal{B}$, it can be easily verified that $\mathcal{E}^c \subseteq \cup_{i=1}^m \mathcal{E}_i$ where $\mathcal{E}_i := \{C_i = A_i\lambda^D(\theta^*), C_i - A_i\lambda^D(\hat{\theta}_{t_1}) > \eta\} \cup \{C_i > A_i\lambda^D(\theta^*), C_i - A_i\lambda^D(\hat{\theta}_{t_1}) \leq \eta\}$. By definition of η,

$$\begin{aligned}
&\mathbb{P}^{\pi}\left(C_i = A_i\lambda^D(\theta^*), C_i - A_i\lambda^D(\hat{\theta}_{t_1}) > \eta\right) \\
&= \mathbb{P}^{\pi}\left(A_i\lambda^D(\theta^*) - A_i\lambda^D(\hat{\theta}_{t_1}) > \eta\right) \leq \mathbb{P}^{\pi}\left(\kappa||A||_2 E(t_1) > \eta\right) \\
&\leq \eta_1 \exp\left(-\eta_2 t_1 \frac{\eta^2}{\kappa^2||A||_2^2}\right) \leq \eta_1 \exp\left(-\frac{\eta_2}{\kappa^2||A||_2^2}(\log k)^{1+\frac{\epsilon}{2}}\right),
\end{aligned}$$

where the first inequality follows by Lemma EC.1(a), the second inequality follows by Lemma 1, and the last inequality holds by definition of t_1 and η. Define $\underline{s} := \min\{C_i - A_i\lambda^D(\theta^*) : C_i - A_i\lambda^D(\theta^*) > 0, i = 1, \ldots, m\}$. Since $\underline{s}$ does not scale with k, there exists a constant $\Omega_0 > 0$ such that $\eta < \underline{s}/2$ for all $k \geq \Omega_0$. So, for $k \geq \Omega_0$, by Lemma EC.1(a) and Lemma 1, we can bound:

$$\begin{aligned}
\mathbb{P}^\pi\left(C_i > A_i\lambda^D(\theta^*),\, C_i - A_i\lambda^D(\hat{\theta}_{t_1}) \leq \eta\right) &= \mathbb{P}^\pi\left(C_i \geq A_i\lambda^D(\theta^*) + \underline{s},\, C_i - A_i\lambda^D(\hat{\theta}_{t_1}) \leq \eta\right) \\
&\leq \mathbb{P}^\pi\left(A_i\lambda^D(\hat{\theta}_{t_1}) - A_i\lambda^D(\theta^*) \geq \underline{s} - \eta\right) \\
&\leq \mathbb{P}^\pi\left(\kappa||A||_2 E(t_1) \geq \underline{s} - \eta\right) \\
&\leq \eta_1 \exp\left(-\eta_2 t_1 \frac{(\underline{s}-\eta)^2}{\kappa^2||A||_2^2}\right) \;\leq\; \eta_1 \exp\left(-\frac{\eta_2\, \underline{s}^2}{4\kappa^2||A||_2^2}\log^{1+\epsilon} k\right).
\end{aligned}$$

Putting the above two bounds together, for $k \geq \Omega_0$, we have

$$\begin{aligned}
\mathbb{P}^\pi\left(\mathcal{E}^c\right) &\leq \sum_{i=1}^m \mathbb{P}^\pi\left(\mathcal{E}_i\right) \leq \sum_{i=1}^m \left[\mathbb{P}^\pi\left(C_i = (A\lambda^D(\theta^*))_i,\, i \notin \mathcal{B}\right) + \mathbb{P}^\pi\left(C_i > (A\lambda^D(\theta^*))_i,\, i \in \mathcal{B}\right)\right] \\
&\leq m\,\eta_1 \left[\exp\left(-\frac{\eta_2}{\kappa^2||A||_2^2}(\log k)^{1+\frac{\epsilon}{2}}\right) + \exp\left(-\frac{\eta_2\, \underline{s}^2}{4\kappa^2||A||_2^2}(\log k)^{1+\epsilon}\right)\right]. \qquad \text{(EC.8)}
\end{aligned}$$

Step 2

For all $t \geq t_1$, let $z(t)$ be the unique integer z such that $t \in [t_z + 1, t_{z+1}]$. Define $S_t := \sum_{s=1}^t D_t$. Let τ be the minimum of k and the first time $t \geq t_1 + 1$ such that the following condition (C1) is violated:

$$(C1)\;\; \psi > \left|\left|\sum_{s=t_1+1}^{t} \frac{\hat{\Delta}_s}{k-s}\right|\right|_2 + \left|\left|\frac{S_L - L\tilde{\lambda}_{\min}\mathbf{e}}{k-t}\right|\right|_2, \;\text{ where } \psi := \frac{\min\{\phi, 2\tilde{\lambda}_{\min}\}}{\max\{2, 4\omega\}}$$

and $\hat{\Delta}_s = D_s - \lambda(p_s; \hat{\theta}_{t_{z(s)}})$. Define $\mathcal{A} := \mathcal{E} \cap \{\cap_{z:t_z<\tau} \mathcal{A}_z\}$ where $\mathcal{A}_z := \{E(t_z) \leq \min\{\hat{\delta}, (\log t_z)^{-\epsilon/4}\}\}$ and $\hat{\delta} = \min\{\bar{\delta}, \tilde{\delta}, \phi/(2\omega\kappa)\}$ and $\bar{\delta}$ and $\tilde{\delta}$ are as defined in Lemma EC.1 and Lemma EC.4 respectively. (Event $\mathcal{A}$ can be interpreted as the event where a sufficient number of binding constraints are correctly identified and the size of all subsequent estimation errors are sufficiently small.)

Note that one immediate observation is that for $t_z < \tau$, $\lambda^D(\theta^*) \in \Lambda_{\hat{\theta}_{t_z}}$ on $\mathcal{A}$. This is because $||p(\lambda^D(\theta^*); \hat{\theta}_{t_z}) - p(\lambda^D(\hat{\theta}_{t_z}); \hat{\theta}_{t_z})||_2 \leq \omega||\lambda^D(\theta^*) - \lambda^D(\hat{\theta}_{t_z})||_2 \leq \omega\kappa||\theta^* - \hat{\theta}_{t_z}||_2 \leq \phi/2$, where the first inequality follows by **A1**, the second inequality follows by Lemma EC.1 (a) and the fact that $\hat{\delta} \leq \bar{\delta}$, and the last inequality follows since $\hat{\delta} \leq \phi/(2\omega\kappa)$. We then have $\lambda^D(\theta^*) \in \Lambda_{\hat{\theta}_{t_z}}$ since $p(\lambda^D(\theta^*); \hat{\theta}_{t_z}) \in \mathsf{Ball}(p^D(\hat{\theta}_{t_z}), \phi/2) \subseteq \mathcal{P}$, where the last inclusion follows by Lemma EC.1 (a).

Define $\hat{\lambda}_t := \lambda^{NT}_{z(t)} - \sum_{s=t_1+1}^{t-1} \frac{\hat{\Delta}_s}{k-s}$ and $\lambda_t := \lambda(p_t; \theta^*)$. The two important lemmas below, which we prove at the end, establish the approximation error of the re-calibration procedure and some important properties of APSC before the stopping time τ.

Lemma EC.5. *There exist positive constants γ and ξ independent of $\theta \in \Theta$ such that if $||x^D(\theta) - x^{NT}_{z-1}||_2 \leq \gamma$, then $||x^D(\theta) - x^{NT}_z||_2 \leq \xi||x^D(\theta) - x^{NT}_{z-1}||_2^2$.*

LEMMA EC.6. *There exist constants $\Omega_1 > 0$, and Γ_1 and Γ_2 independent of $k \geq \Omega_1$, such that for all $k \geq \Omega_1$ and all sample paths on $\mathcal{A}$:*

(a) $||x^D(\hat{\theta}_{t_z}) - x_z^{NT}||_2^2 \leq \Gamma_1 (\log t_z)^{-\epsilon/2}$ *for* $t_z < \tau$.

(b) $C_t \succ 0$, $p_t = \hat{p}_t \in \mathsf{Ball}(p^D(\theta^*), 7\phi/8) \subseteq \mathcal{W}(\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max})$ *and* $\hat{\lambda}_t \in \Lambda_{\hat{\theta}_{t_z}}$ *for all* $t \in (t_z, t_{z+1}] \cap [t_1, \tau)$.

(c) $\mathbf{E}^\pi[||x^D(\hat{\theta}_{t_z}) - x_z^{NT}||_2^2 \mathbf{1}_{\{t_z < \tau\}} \mid \mathcal{A}] \leq \Gamma_2 / t_z$

Lemma EC.5 essentially establishes a *uniform* locally quadratic convergence of the Newton's method for solving $\mathrm{ECP}(\hat{\theta}_{t_z})$ for all z. This result is used for proving Lemma EC.6 (a) and (c) which establish the approximation errors between x_z^{NT} and $x^D(\hat{\theta}_{t_z})$ under APSC. Note that Lemma EC.6 (b) is also very important as it shows that conditioning on $\mathcal{A}$, before τ, the seller still has positive capacity on hand, and the pricing decision under APSC is $p_t = \hat{p}_t$ which is given by an explicit expression that we can use to analyze APSC later. Note also that $\mathcal{A}$ happens with very high probability. Indeed, there exists a constant $\Omega_2 \geq \max\{\Omega_0, \Omega_1\}$ such that, for all $k \geq \Omega_2$,

$$
\begin{aligned}
k\mathbb{P}^\pi(\mathcal{A}^c) &\leq k \sum_{z=1}^{Z} \mathbb{P}^\pi\left(\mathcal{A}_z^c\right) + k\,\mathbb{P}^\pi\left(\cup_{i=1}^m \mathcal{E}_i\right) \\
&\leq k \sum_{z=1}^{Z} \left[\mathbb{P}^\pi(E(t_z) > \hat{\delta}) + \mathbb{P}^\pi(E(t_z) > (\log t_z)^{-\frac{\epsilon}{4}})\right] + k\,\mathbb{P}^\pi\left(\cup_{i=1}^m \mathcal{E}_i\right) \\
&\leq k \sum_{z=1}^{Z} \eta_4 t_z^{q-1} \left[\exp\left(-\eta_5 t_z \hat{\delta}^2\right) + \exp\left(-\frac{\eta_5 t_z}{(\log t_z)^{\frac{\epsilon}{2}}}\right)\right] + k\,\mathbb{P}^\pi\left(\cup_{i=1}^m \mathcal{E}_i\right) \\
&\leq 2k(\log_2 k) \left[\exp\left(-\frac{\eta_5 (\log k)^{1+\epsilon} \hat{\delta}^2}{2}\right) + \exp\left(-\frac{\eta_5 (\log k)^{1+\epsilon}}{2(\log k)^{\frac{\epsilon}{2}}}\right)\right] + k\,\mathbb{P}^\pi\left(\cup_{i=1}^m \mathcal{E}_i\right) \\
&\leq 2k(\log_2 k) \left[\exp\left(-\frac{\eta_5 (\log k)^{1+\epsilon} \hat{\delta}^2}{2}\right) + \exp\left(-\frac{\eta_5 (\log k)^{1+\frac{\epsilon}{2}}}{2}\right)\right] \\
&\quad + \; m\,\eta_1 k \left[\exp\left(-\frac{\eta_2}{\kappa^2 ||A||_2^2} (\log k)^{1+\frac{\epsilon}{2}}\right) + \exp\left(-\frac{\eta_2\, \underline{s}^2}{4\kappa^2 ||A||_2^2} (\log k)^{1+\epsilon}\right)\right] \;\leq\; \frac{1}{2},
\end{aligned}
$$

where the third inequality follows by Lemma 2, the fourth inequality follows by a combination of $\eta_4 t_z^{q-1} \exp(-\eta_5 t_z \hat{\delta}^2/2) \to 0$ and $\eta_4 t_z^{q-1} \exp(-\eta_5 t_z (\log t_z)^{-\epsilon/2}/2) \to 0$ as $k \to \infty$, $t_z \geq t_1 \geq (\log k)^{1+\epsilon}$ for $z \geq 1$, and $Z \leq \lceil \log_2 k \rceil \leq 2 \log_2 k$, the fifth inequality follows by (EC.8), and the last inequality follows because the formula after the fourth inequality goes to zero as $k \to \infty$. Note that the above inequality also implies $\mathbb{P}^\pi(\mathcal{A}) > \frac{1}{2}$ when $k \geq \Omega_2$.

Define $\Psi_\epsilon := \sum_{t=t_1+1}^{k-1} \left(\sum_{s=t_1+1}^{t-1} \frac{\bar{\epsilon}(s)}{k-s}\right)^2$ and $\Phi_\epsilon := \sum_{t=t_1+1}^{k-1} \bar{\epsilon}(s)^2$, where $\bar{\epsilon}(s) := \eta_6 \sqrt{[(q-1)\log t_z + 1]/t_z}$ for all $s \in [t_z + 1, t_{z+1}]$. By Lemma 2, $\mathbf{E}^\pi[||\hat{\theta}_{t_{z(t)}} - \theta^*||_2^2 \mathbf{1}_{\{t<\tau\}} | \mathcal{A}] \leq \mathbf{E}^\pi[||\hat{\theta}_{t_{z(t)}} - \theta^*||_2^2 \mathbf{1}_{\{t<\tau\}} | \mathcal{A}]/\mathbb{P}^\pi(\mathcal{A}) \leq 2\mathbf{E}^\pi[||\hat{\theta}_{t_{z(t)}} - \theta^*||_2^2 \mathbf{1}_{\{t<\tau\}} | \mathcal{A}] \leq 2\bar{\epsilon}(t)^2$. The next result, which we prove at the end, is useful to derive our bounds later.

LEMMA EC.7. *Under APSC, there exists a constant $K_3 > 0$ independent of $k \geq 1$ such that $\Psi_\epsilon < K_3(1+(q-1)\log k)$ and $\Phi_\epsilon < K_3[1+\log k+(q-1)(\log k)^2]$.*

Step 3

Let $K = \max\{\Omega_0, \Omega_1, \Omega_2, 3\}$, where Ω_0 (resp. Ω_1, Ω_2) is defined in Step 1 (resp. Step 2). If $k < K$, the total regret can be bounded by $K\bar{r} = \Theta(1)$. So, in the remainder of Step 3, we will focus on the case $k \geq K$. Define $r^D(\theta^*) := r(\lambda^D(\theta^*); \theta^*)$ and let R_t^π denote the revenue earned during period t under policy π. Then $\rho^\pi(k) \leq L\bar{r} + \sum_{t=L+1}^{k}(r^D(\theta^*) - \mathbf{E}^\pi[R_t^\pi])$. Let $\bar{\Delta}_t := R_t^\pi - r(\lambda_t; \theta^*)$. We have:

$$
\begin{aligned}
&\sum_{t=L+1}^{k}\left(r^D(\theta^*) - \mathbf{E}^\pi[R_t^\pi]\right) \\
&= \mathbf{E}^\pi\left[\sum_{t=L+1}^{\tau-1}\left(r^D(\theta^*) - R_t^\pi\right) + \sum_{t=\tau}^{k}\left(r^D(\theta^*) - R_t^\pi\right)\right] \\
&= \mathbf{E}^\pi\left[\sum_{t=L+1}^{\tau-1}\left(r^D(\theta^*) - r(\lambda_t;\theta^*)\right) + \sum_{t=\tau}^{k}\left(r^D(\theta^*) - R_t^\pi\right)\right] - \mathbf{E}^\pi\left[\sum_{t=L+1}^{\tau-1}\bar{\Delta}_t\right] \\
&\leq \mathbf{E}^\pi\left[\sum_{t=L+1}^{\tau-1}\left(r^D(\theta^*) - r(\lambda_t;\theta^*)\right) + \sum_{t=\tau}^{k}\left(r^D(\theta^*) - R_t^\pi\right)\middle|\mathcal{A}\right]\mathbb{P}^\pi(\mathcal{A}) + \bar{r}k\mathbb{P}^\pi(\mathcal{A}^c) - \mathbf{E}^\pi\left[\sum_{t=L+1}^{\tau-1}\bar{\Delta}_t\right] \\
&\leq \mathbf{E}^\pi\left[\sum_{t=L+1}^{\tau-1}\left(r^D(\theta^*) - r(\lambda_t;\theta^*)\right) + \sum_{t=\tau}^{k}\left(r^D(\theta^*) - R_t^\pi\right)\middle|\mathcal{A}\right] + \bar{r}k\mathbb{P}^\pi(\mathcal{A}^c) + \bar{r} \\
&\leq \mathbf{E}^\pi\left[\sum_{t=L+1}^{\tau-1}(\nabla r(\lambda^D(\theta^*);\theta^*))'(\lambda^D(\theta^*) - \lambda_t)\middle|\mathcal{A}\right] + \frac{\bar{v}}{2}\mathbf{E}^\pi\left[\sum_{t=L+1}^{\tau-1}||\lambda^D(\theta^*) - \lambda_t||_2^2\middle|\mathcal{A}\right] \\
&\qquad + \bar{r}\mathbf{E}^\pi[k-\tau+1|\mathcal{A}] \;+\; \bar{r}k\mathbb{P}^\pi(\mathcal{A}^c) \;+\; \bar{r} \\
&= \mathbf{E}^\pi\left[\sum_{t=t_1+1}^{\tau-1}\mu^D(\theta^*)'A(\lambda^D(\theta^*) - \lambda_t)\middle|\mathcal{A}\right] + \frac{\bar{v}}{2}\mathbf{E}^\pi\left[\sum_{t=t_1+1}^{\tau-1}||\lambda^D(\theta^*) - \lambda_t||_2^2\middle|\mathcal{A}\right] \\
&\qquad + \bar{r}\,\mathbf{E}^\pi[k-\tau|\mathcal{A}] \;+\; 2\,\bar{r} \;+\; \bar{r}\,k\mathbb{P}^\pi(\mathcal{A}^c). \qquad \text{(EC.9)}
\end{aligned}
$$

where $\mu^D(\theta^*)$ is the optimal dual solution $\mu^D(\theta^*)$ of $\mathrm{P}_\lambda(\theta^*)$ such that $A_i\lambda^D(\hat{\theta}_{t_1}) = C_i$ for all $i \in \{j : \mu_j^D(\theta^*) > 0\}$ (its existence is established by Lemma EC.1 Part (b)). The first inequality follows because $\bar{r}$ is the upper bound on revenue rate for each period, which is also the maximum possible regret for a single period on average. As for the second inequality, note that $\{\bar{\Delta}_t\}_{t=L+1}^{k-1}$ is a martingale difference sequence with respect to $\{\mathcal{H}_t\}_{t=L+1}^{k-1}$. Thus, by the Optional Stopping Time Theorem (see, for example, Williams (1991)), we have $-\mathbf{E}^\pi\left[\sum_{t=L+1}^{\tau-1}\bar{\Delta}_t\right] = -\mathbf{E}^\pi\left[\sum_{t=L+1}^{\tau}\bar{\Delta}_t\right] + \mathbf{E}^\pi\left[\bar{\Delta}_\tau\right] \leq \bar{r}$, so the second inequality holds. The third inequality follows by Taylor expansion and **A3**. The last equality follows by the KKT condition and that $L = t_1$ by definition. Note that conditioning on $\mathcal{A}$,

$$\mu^D(\theta^*)'A\lambda^D(\theta^*) = \mu^D(\theta^*)'A\lambda_z^{NT} \qquad \text{(EC.10)}$$

for all z such that $t_z < \tau$. Indeed, when $z = 1$, (EC.10) holds because $A_i \lambda^D(\theta^*) = C_i = A_i \lambda^D(\hat{\theta}_{t_1}) = A_i \lambda_1^{NT}$ for $i \in \{j : \mu_j^D(\theta^*) > 0\}$ where the first equality holds by complementary slackness, the second equality holds by Lemma EC.1 Part (b). When $z > 1$, (EC.10) holds since for any i, $A_i \lambda_1^{NT} = C_i$ implies that $A_i \lambda_z^{NT} = C_i$ (indeed, by construction of $\mathcal{B}$ and the definition of $\mathcal{E}$, any binding constraint of $\mathrm{P}_\lambda(\hat{\theta}_{t_1})$ that is not included in $\mathcal{B}$ is implied by the binding constraints of $\mathrm{P}_\lambda(\hat{\theta}_{t_1})$ that are included in $\mathcal{B}$; moreover, any binding constraints of $\mathrm{P}_\lambda(\hat{\theta}_{t_1})$ that are included in $\mathcal{B}$ are also binding at λ_z^{NT} by the Newton's iteration in (6)).

We now bound the first term in (EC.9). By Lemma EC.6 (b), conditioning on $\mathcal{A}$, $p_t = \hat{p}_t$ and $\hat{\lambda}_t = \lambda_{z(t)}^{NT} - \sum_{s=t_1+1}^{t-1} \frac{\hat{\Delta}_s}{k-s} \in \Lambda_{\hat{\theta}_{t_z}}$ for all $t_1 \le t < \tau$. Also, note that $\hat{\Delta}_t = D_t - \hat{\lambda}_t = \Delta_t + \lambda_t - \hat{\lambda}_t$. So, we can write the first term in (EC.9) as follows:

$$
\begin{aligned}
&\mathbf{E}^\pi \left[\sum_{t=t_1+1}^{\tau-1} \mu^D(\theta^*)' A \left(\lambda^D(\theta^*) - \hat{\lambda}_t + \hat{\lambda}_t - \lambda_t \right) \middle| \mathcal{A} \right] \\
&= \mathbf{E}^\pi \left[\sum_{t=t_1+1}^{\tau-1} \mu^D(\theta^*)' \left(A\lambda^D(\theta^*) - A\lambda_{z(t)}^{NT} + \sum_{s=t_1+1}^{t-1} \frac{A\hat{\Delta}_s}{k-s} + A\Delta_t - A\hat{\Delta}_t \right) \middle| \mathcal{A} \right] \\
&= \mathbf{E}^\pi \left[\sum_{t=t_1+1}^{\tau-1} \mu^D(\theta^*)' \left(A\lambda^D(\theta^*) - A\lambda_{z(t)}^{NT} \right) \middle| \mathcal{A} \right] + \mathbf{E}^\pi \left[\sum_{t=t_1+1}^{\tau-1} \mu^D(\theta^*)' \left(\sum_{s=t_1+1}^{t-1} \frac{A\hat{\Delta}_s}{k-s} + A\Delta_t - A\hat{\Delta}_t \right) \middle| \mathcal{A} \right] \\
&= \mathbf{E}^\pi \left[\sum_{t=t_1+1}^{\tau-1} \mu^D(\theta^*)' \left(\sum_{s=t_1+1}^{t-1} \frac{A\hat{\Delta}_s}{k-s} + A\Delta_t - A\hat{\Delta}_t \right) \middle| \mathcal{A} \right] \\
&= \mathbf{E}^\pi \left[\sum_{t=t_1+1}^{\tau-1} \mu^D(\theta^*)' A\Delta_t \middle| \mathcal{A} \right] + \mathbf{E}^\pi \left[\sum_{t=t_1+1}^{\tau-1} \left(\frac{\tau-t-1}{k-t} - 1 \right) \mu^D(\theta^*)' A\hat{\Delta}_t \middle| \mathcal{A} \right], \qquad \text{(EC.11)}
\end{aligned}
$$

where the second to the last equality follows by (EC.10). Since $\{\Delta_t\}_{t=t_1+1}^{k-1}$ is a martingale difference sequence with respect to $\{\mathcal{H}_t\}_{t=t_1+1}^{k-1}$, we can bound the first term of (EC.11) by:

$$
\begin{aligned}
\mathbf{E}^\pi \left[\sum_{t=t_1+1}^{\tau-1} \mu^D(\theta^*)' A\Delta_t \middle| \mathcal{A} \right] &= \frac{\mu^D(\theta^*)' A}{\mathbb{P}^\pi(\mathcal{A})} \left\{ \mathbf{E}^\pi \left[\sum_{t=t_1+1}^{\tau-1} \Delta_t \right] - \mathbf{E}^\pi \left[\sum_{t=t_1+1}^{\tau-1} \Delta_t \middle| \mathcal{A}^c \right] \mathbb{P}^\pi(\mathcal{A}^c) \right\} \\
&\le \mu^D(\theta^*)' A\mathbf{e} \, \frac{1 + k\mathbb{P}^\pi(\mathcal{A}^c)}{1 - \mathbb{P}^\pi(\mathcal{A}^c)} \le 3\, \mu^D(\theta^*)' A\mathbf{e},
\end{aligned}
$$

where the first inequality follows because $\mathbf{E}^\pi[\sum_{t=t_1+1}^{\tau-1} \Delta_t] = \mathbf{E}^\pi[\sum_{t=t_1+1}^{\tau} \Delta_t] - \mathbf{E}^\pi[\Delta_\tau] \prec \mathbf{e}$ (by Optional Stopping Time Theorem) and the fact that $|\Delta_t| \prec \mathbf{e}$. As for the second term of (EC.11), note that, by (C1) in the definition of τ,

$$
\begin{aligned}
&\mathbf{E}^\pi \left[\sum_{t=t_1+1}^{\tau-1} \left(\frac{\tau-t-1}{k-t} - 1 \right) \mu^D(\theta^*)' A\hat{\Delta}_t \,\middle|\, \mathcal{A} \right] \le \mathbf{E}^\pi \left[(k-\tau+1) \left| \mu^D(\theta^*)' \sum_{t=t_1+1}^{\tau-1} \frac{A\hat{\Delta}_t}{k-t} \right| \,\middle|\, \mathcal{A} \right] \\
&\le \mathbf{E}^\pi \left[(k-\tau+1) \left|\left| \mu^D(\theta^*) \right|\right|_2 ||A||_2 \left|\left| \sum_{t=t_1+1}^{\tau-1} \frac{\hat{\Delta}_t}{k-t} \right|\right|_2 \,\middle|\, \mathcal{A} \right] \le \psi \left|\left| \mu^D(\theta^*) \right|\right|_2 ||A||_2 \left(\mathbf{E}^\pi \left[k-\tau \middle| \mathcal{A} \right] + 1 \right).
\end{aligned}
$$

Hence, we can bound (EC.11) (i.e., the first term in (EC.9)) with $K_4\mathbf{E}^\pi[k-\tau+1|\mathcal{A}]$ where $K_4 := 3\mu^D(\theta^*)'A\mathbf{e}+\psi||\mu^D(\theta^*)||_2||A||_2$ is independent of $k \geq K$.

As for the second term in (EC.9), note that $\frac{\bar{v}}{2}\mathbf{E}^\pi[\sum_{t=t_1+1}^{\tau-1}||\lambda^D(\theta^*)-\lambda_t||_2^2|\mathcal{A}] \leq \bar{v}\mathbf{E}^\pi[\sum_{t=t_1+1}^{\tau-1}||\hat{\lambda}_t-\lambda_t||_2^2|\mathcal{A}] + \bar{v}\mathbf{E}^\pi[\sum_{t=t_1+1}^{\tau-1}||\lambda^D(\theta^*)-\hat{\lambda}_t||_2^2|\mathcal{A}]$ where the first term on the right hand side can be bounded by $\bar{v}\mathbf{E}^\pi\left[\sum_{t=t_1+1}^{\tau-1}||\hat{\lambda}_t-\lambda_t||_2^2\middle|\mathcal{A}\right] = \bar{v}\mathbf{E}^\pi\left[\sum_{t=t_1+1}^{\tau-1}||\lambda(\hat{p}_t;\hat{\theta}_{t_{z(t)}})-\lambda(\hat{p}_t;\theta^*)||_2^2\middle|\mathcal{A}\right] = \bar{v}\sum_{t=t_1+1}^{k-1}\mathbf{E}^\pi\left[\omega^2||\hat{\theta}_t-\theta^*||_2^2\mathbf{1}_{\{t<\tau\}}\middle|\mathcal{A}\right] \leq 2\bar{v}\omega^2\sum_{t=t_1+1}^{k-1}\bar{\epsilon}(t)^2 \leq 2\bar{v}\omega^2\Phi_\epsilon$ (by Lemma EC.6 (b) and **A2**), and the second term on the right hand side can be bounded as:

$$\begin{aligned}
&\bar{v}\mathbf{E}^\pi\left[\sum_{t=t_1+1}^{\tau-1}||\lambda^D(\theta^*)-\hat{\lambda}_t||_2^2\middle|\mathcal{A}\right] \\
&\quad\leq 2\bar{v}\mathbf{E}^\pi\left[\sum_{t=t_1+1}^{\tau-1}||\lambda^D(\theta^*)-\lambda^{NT}_{z(t)}||_2^2\middle|\mathcal{A}\right] + 2\bar{v}\mathbf{E}^\pi\left[\sum_{t=t_1+1}^{\tau-1}\left|\left|\sum_{s=t_1+1}^{t-1}\frac{\hat{\Delta}_s}{k-s}\right|\right|_2^2\middle|\mathcal{A}\right]. \quad \text{(EC.12)}
\end{aligned}$$

We now bound the two terms in (EC.12) starting from the second term.

$$\begin{aligned}
&2\bar{v}\mathbf{E}^\pi\left[\sum_{t=t_1+1}^{\tau-1}\left|\left|\sum_{s=t_1+1}^{t-1}\frac{\hat{\Delta}_s}{k-s}\right|\right|_2^2\middle|\mathcal{A}\right] \\
&\leq 4\bar{v}\left(\mathbf{E}^\pi\left[\sum_{t=t_1+1}^{\tau-1}\left|\left|\sum_{s=t_1+1}^{t-1}\frac{\Delta_s}{k-s}\right|\right|_2^2\middle|\mathcal{A}\right] + \mathbf{E}^\pi\left[\sum_{t=t_1+1}^{k-1}\left(\sum_{s=t_1+1}^{t-1}\frac{||\hat{\lambda}_s-\lambda_s||_2\mathbf{1}_{\{s<\tau\}}}{k-s}\right)^2\middle|\mathcal{A}\right]\right) \\
&\leq 4\bar{v}\left(\mathbf{E}^\pi\left[\sum_{t=t_1+1}^{\tau-1}\left|\left|\sum_{s=t_1+1}^{t-1}\frac{\Delta_s}{k-s}\right|\right|_2^2\middle|\mathcal{A}\right] + \mathbf{E}^\pi\left[\sum_{t=t_1+1}^{k-1}\left(\sum_{s=t_1+1}^{t-1}\frac{\omega E(s)\mathbf{1}_{\{s<\tau\}}}{k-s}\right)^2\middle|\mathcal{A}\right]\right) \\
&\leq 4\bar{v}\left(\frac{2}{\mathbb{P}^\pi(\mathcal{A})}\mathbf{E}^\pi\left[\sum_{t=t_1+1}^{k-1}\left|\left|\sum_{s=t_1+1}^{t-1}\frac{\Delta_s}{k-s}\right|\right|_2^2\right] + \sum_{t=t_1+1}^{k-1}\left[\sum_{s=t_1+1}^{t-1}\frac{\sqrt{\mathbf{E}^\pi\left[\omega^2E(s)^2\mathbf{1}_{\{s<\tau\}}\middle|\mathcal{A}\right]}}{k-s}\right]^2\right) \\
&= 4\bar{v}\left(\frac{2}{\mathbb{P}^\pi(\mathcal{A})}\mathbf{E}^\pi\left[\sum_{t=t_1+1}^{k-1}\sum_{s=t_1+1}^{t-1}\frac{||\Delta_s||_2^2}{(k-s)^2}\right] + \sum_{t=t_1+1}^{k-1}\left[\sum_{s=t_1+1}^{t-1}\frac{\sqrt{\mathbf{E}^\pi\left[\omega^2E(s)^2\mathbf{1}_{\{s<\tau\}}\middle|\mathcal{A}\right]}}{k-s}\right]^2\right) \\
&\leq 4\bar{v}\left(\frac{16}{\mathbb{P}^\pi(\mathcal{A})}\log k + \sum_{t=t_1+1}^{k-1}\left[\sum_{s=t_1+1}^{t-1}\frac{\sqrt{\mathbf{E}^\pi\left[\omega^2E(s)^2\mathbf{1}_{\{s<\tau\}}\middle|\mathcal{A}\right]}}{k-s}\right]^2\right) \\
&\leq 4\bar{v}(32\log k + 2\omega^2\Psi_\epsilon) \;\leq\; K_5(\Psi_\epsilon+\log k) \quad \text{(EC.13)}
\end{aligned}$$

for some constant $K_5 > 0$ independent of $k \geq K$, where the second inequality follows by Lemma EC.6 (b) and **A2**, the third inequality follows by applying the law of total probability to the first term and applying Cauchy-Swartz inequality to the second term (i.e., first expanding the square of the sum of the second term and then applying Cauchy-Swartz inequality to the cross-terms), the equality follows by the orthogonality of martingale differences $\{\Delta_s\}_s$, and the

fourth inequality holds by integral approximation (i.e., the first term after the third inequality can be bounded using $||\Delta_s||_2 = ||D_s - \lambda_s||_2 \le ||D_s||_2 + ||\lambda_s||_2 \le 2$ and $\sum_{t=t_1+1}^{k-1}\sum_{s=t_1+1}^{t-1}\frac{1}{(k-s)^2} \le \sum_{t=t_1+1}^{k-1}\frac{1}{k-t} \le 1+\log k \le 2\log k$ (recall that $k \ge 3$)), and the fifth inequality follows by the definition of Ψ_ϵ and $\mathbb{P}^\pi(\mathcal{A}) \ge 1/2$. We now bound the first term of (EC.12). Note that, conditioning on $\mathcal{A}$, for all $t < \tau$, we have

$$\left|\left|\lambda^D(\theta^*) - \lambda^{NT}_{z(t)}\right|\right|_2 = \left|\left|x^D(\theta^*) - x^{NT}_{z(t)}\right|\right|_2 \le \left|\left|x^D(\theta^*) - x^D(\hat{\theta}_{t_{z(t)}})\right|\right|_2 + \left|\left|x^D(\hat{\theta}_{t_{z(t)}}) - x^{NT}_z\right|\right|_2, \quad \text{(EC.14)}$$

where the first equality follows due to the observation that $\lambda^D(\theta^*) = x^D(\theta^*)$ on $\mathcal{A}$. This observation is true due to the following reasons: (1) since the objective function of $\mathrm{P}_\lambda(\theta^*)$ is strongly concave, removing the constraints that are non-binding at $\lambda^D(\theta^*)$ does not affect the optimal solution; (2) by definition of $\mathcal{E}$, the binding constraints at $\lambda^D(\theta^*)$ but are not included in $\mathcal{B}$ are redundant.

By Lemma EC.4,

$$\mathbf{E}^\pi\left[\sum_{s=t_1+1}^{\tau-1} ||x^D(\theta^*) - x^D(\hat{\theta}_{t_{z(s)}})||_2^2 \,\middle|\, \mathcal{A}\right] = \sum_{s=t_1+1}^{k-1} \mathbf{E}^\pi\left[\tilde{\kappa}^2||\theta^* - \hat{\theta}_{t_{z(s)}}||_2^2 \mathbf{1}_{\{s<\tau\}} \,\middle|\, \mathcal{A}\right] \le 2\tilde{\kappa}^2\Phi_\epsilon.$$

Furthermore, by Lemma EC.6 (a) and the fact that $t_{z+1} - t_z \le 2t_z$ for all z, we have

$$\begin{aligned}\mathbf{E}^\pi\left[\sum_{s=t_1+1}^{\tau-1} ||x^D(\hat{\theta}_{t_{z(s)}}) - x^{NT}_{z(s)}||_2^2 \,\middle|\, \mathcal{A}\right] &= \sum_{s=t_1+1}^{k-1} \mathbf{E}^\pi\left[||x^D(\hat{\theta}_{t_{z(s)}}) - x^{NT}_{z(s)}||_2^2 \mathbf{1}_{\{s<\tau\}} \,\middle|\, \mathcal{A}\right] \\ &\le \sum_{z=1}^{Z}(t_{z+1} - t_z)\frac{\Gamma_2}{t_z} \le 2Z\,\Gamma_2 \le 4\Gamma_2\log_2 k = \frac{4\Gamma_2}{\log_e 2}\log k.\end{aligned}$$

Combining the inequalities above, the second term of (EC.9) can be bounded as follows:

$$\begin{aligned}\frac{\bar{v}}{2}\mathbf{E}^\pi\left[\sum_{t=t_1+1}^{\tau-1} ||\lambda^D(\theta^*) - \lambda_t||_2^2\right] &\le 2\bar{v}\omega^2\Phi_\epsilon + K_5(\Psi_\epsilon + \log k) + 4\bar{v}\tilde{\kappa}^2\Phi_\epsilon + \frac{8\bar{v}\Gamma_2}{\log_e 2}\log_2 \\ &\le K_6(1 + \log k + (q-1)\log^2 k)\end{aligned}$$

for $K_6 = (2\bar{v}\omega^2 + 4\bar{v}\tilde{\kappa}^2 + K_5)K_3 + K_5 + \frac{8}{\log_e 2}\bar{v}\,\Gamma_2$. To bound the third term in (EC.9), the following lemma is useful which we prove at the end.

Lemma EC.8. *There exists a constant $K_7 > 0$ independent of $k \ge K$ such that for all $k \ge K$, $\mathbf{E}^\pi[k - \tau|\mathcal{A}] \le K_7(\log k + L)$.*

Combining all the above and recalling that $L = \lceil(\log k)^{1+\epsilon}\rceil$, for all $k \ge K$, we have:

$$\begin{aligned}\rho^\pi(k) &\le 2\bar{r}(\log k)^{1+\epsilon} + (K_4 + \bar{r})(\mathbf{E}^\pi[k-\tau|\mathcal{A}] + 1) + K_6(1 + \log k + (q-1)\log^2 k) + \frac{5}{2}\bar{r} \\ &\le \left(2\bar{r} + 2K_4K_7 + 2\bar{r}K_7 + K_6 + \frac{5}{2}\bar{r}\right)\left[1 + (\log k)^{1+\epsilon} + (q-1)\log^2 k\right] \\ &\le K_8[(\log k)^{1+\epsilon} + (q-1)\log^2 k],\end{aligned}$$

for some constant K_8 independent of $k \geq K$. The result of Theorem 2 follows by using $M_2 = \max\{\bar{r}K, K_8\}$. $\square$

Next, we prove the intermediate results Lemma EC.5-EC.8 below.

Proof of Lemma EC.5. Fix $\theta \in \Theta$. Note that ECP(θ) is a convex optimization with linear equality constraints. Let m_B denote the number of rows of B, and define F to be an n by $n - m_B$ matrix whose columns are unit orthogonal basis vectors and $BF = 0$. (In case there are multiple matrices that satisfy this condition, pick any one of them.) Then $\{x : Bx = C_B/T\} = \{x : x = Fz + \hat{x}, z \in \mathbb{R}^{n-m_B}\}$ where $\hat{x}$ satisfies $B\hat{x} = C_B/T$. Hence, ECP(θ) is equivalent to an unconstrained optimization problem $\max_{z \in \mathbb{R}^{n-m_B}} g(z;\theta) := r(Fz + \hat{x};\theta)$ in the sense that there is a one-to-one mapping between the optimizer of ECP(θ) $x^D(\theta)$ and the optimizer of the unconstrained problem $z^D(\theta)$: (1) $x^D(\theta) = Fz^D(\theta) + \hat{x}$, and (2) $z^D(\theta) = F'(x^D(\theta) - \hat{x})$. In addition, by Section 10.2.3 in Boyd and Vandenberghe (2004), if a feasible point of ECP(θ) $x^{(k)}$ and a feasible point of the unconstrained problem $z^{(k)}$ satisfy $x^{(k)} = Fz^{(k)} + \hat{x}$, then the Newton steps for ECP(θ) (to obtain a new feasible point $x^{(k+1)}$) and the unconstrained problem (to obtain a new feasible point $z^{(k+1)}$) coincide in the sense that $x^{(k+1)} = Fz^{(k+1)} + \hat{x}$. This relationship enables us to analyze the behavior of $x^{(k)}$ by studying $z^{(k)}$ whose convergence behavior is characterized by the well-known result below.

Theorem EC.2. (Quadratic Convergence of Newton's Method for Convex Unconstrained Optimization Problems, Section 9.5.3 in Boyd and Vandenberghe (2004)) *Suppose $g(z)$ is a concave function whose unconstrained optimizer is x^*. Let $\{x^{(k)}\}_{k=1}^{\infty}$ be a sequence of points obtained by Newton's method. Assume there exist positive constants m, M, L such that*

(i) $||\nabla^2 g(z) - \nabla^2 g(y)||_2 \leq L||z - y||_2$, *and*

(ii) $-MI \preceq \nabla^2 g(z) \preceq -mI$.

Then, there exists constant $\eta = \min\{1, 3(1-2\alpha)\}m^2/L$ where $\alpha \in (0, 0.5)$ such that if $||\nabla g(x^{(k)})||_2 < \eta$, then $||\nabla g(x^{(k+1)})||_2 \leq \frac{L}{2m^2}||\nabla g(x^{(k)})||_2^2$.

Before applying Theorem EC.2, we first show that the conditions in Theorem EC.2 hold. Note that since Λ_θ is compact, the linear transformation of it, $\mathcal{Z}_\theta := \{z : z = F'(x - \hat{x}), x \in \Lambda_\theta\}$ is also compact. Also note that since $p(.;\theta) \in \mathsf{C}^2(\Lambda_\theta)$ by **A1**, $r(.;\theta) \in \mathsf{C}^2(\Lambda_\theta)$ and $g(.;\theta) \in \mathsf{C}^2(\mathcal{Z}_\theta)$. Hence condition (i) holds: there exists some constant L such that $||\nabla^2_{zz} g(z;\theta) - \nabla^2_{zz} g(y;\theta)||_2 \leq L||z - y||_2$. Since the columns of F consist of unit orthogonal basis vectors, $\nabla^2_{zz} g(z;\theta) = F'\nabla^2_{\lambda\lambda} r(Fz + \hat{x};\theta)F$ and $-MI \preceq \nabla^2_{\lambda\lambda} r(Fz + \hat{x};\theta) \preceq -mI$ by **A3**, we conclude that (ii) holds: $-MI \preceq \nabla^2_{zz} g(z;\theta) \preceq -mI$. Then, by Theorem EC.2, there exists a constant $\eta = \min\{1, 3(1-2\alpha)\}\bar{m}^2/L$ for some $\alpha \in (0, 0.5)$ independent of θ such that if $||\nabla_z g(z^{(k)};\theta)||_2 < \eta$, then $||\nabla_z g(z^{(k+1)};\theta)||_2 < \frac{L}{2m^2}||\nabla_z g(z^{(k)};\theta)||_2^2$. Note that by strong convexity of $g(.;\theta)$, $M^{-1}||\nabla_z g(z;\theta)||_2 \leq ||z - z^D(\theta)||_2 \leq 2m^{-1}||\nabla_z g(z;\theta)||_2$. Also note

that for $x = Fz + \hat{x}$, $||x - x^D(\theta)||_2 \le ||F||_2||z - z^D(\theta)||_2 = ||z - z^D(\theta)||_2 \le ||F'||_2||x - x^D(\theta)||_2 = ||x - x^D(\theta)||_2$, so $||x - x^D(\theta)||_2 = ||z - z^D(\theta)||_2$. Therefore,

$$\begin{aligned}||x^{(k+1)} - x^D(\theta)||_2 = ||z^{(k+1)} - z^D(\theta)||_2 &\le 2m^{-1}||\nabla_z g(z^{(k+1)};\theta)||_2 \le Lm^{-3}||\nabla_z g(z^{(k)};\theta)||_2^2 \\ &\le Lm^{-3}M^2||z^{(k)} - z^D(\theta)||_2^2 = Lm^{-3}M^2||x^{(k)} - x^D(\theta)||_2^2\end{aligned}$$

Let $\gamma = \eta$ and $\xi = Lm^{-3}M^2$. Note that they are both independent of θ. The result follows by letting $x^{(k+1)} = x_z^{NT}$ and $x^{(k)} = x_{z-1}^{NT}$. □

Proof of Lemma EC.6. Let $\Omega_1 = \max_{i=1,..,4}\{V_i\}$, where V_i's are positive constants to be defined later. We prove the results one by one.

(a) Let $\bar{\kappa} = \max\{\kappa, \tilde{\kappa}\}$ where κ and $\tilde{\kappa}$ are defined in Lemma EC.1 and Lemma EC.4 respectively. Let $\Gamma_1 = \max\{1, 4\bar{\kappa}^2\}$. We proceed by induction. If $t_1 \ge \tau$, there is nothing to prove, so we consider the case when $t_1 < \tau$. Recall that $x_1^{NT} = \lambda^D(\hat{\theta}_{t_1})$ and $x^D(\theta^*) = \lambda^D(\theta^*)$ on $\mathcal{A}$. Thus, when $t_1 < \tau$

$$\begin{aligned}||x^D(\hat{\theta}_{t_1}) - x_1^{NT}||_2^2 &= ||x^D(\hat{\theta}_{t_1}) - \lambda^D(\hat{\theta}_{t_1})||_2^2 \\ &\le \left(||x^D(\hat{\theta}_{t_1}) - x^D(\theta^*)||_2 + ||\lambda^D(\theta^*) - \lambda^D(\hat{\theta}_{t_1})||_2\right)^2 \\ &\le 4\bar{\kappa}^2 E(t_1)^2 \ \le \ \Gamma_1(\log t_1)^{-\frac{\epsilon}{2}}\end{aligned}$$

where the last inequality follows by the definition of $\mathcal{A}$. This is our base case. We now do the inductive step. Suppose that $t_{z-1} < \tau$ and $||x^D(\hat{\theta}_{t_{z-1}}) - x_{z-1}^{NT}||_2^2 \le \Gamma_1(\log t_{z-1})^{-\epsilon/2}$. If $t_z \ge \tau$ there is nothing to prove. If $t_z < \tau$, then we need to show that $||x^D(\hat{\theta}_{t_z}) - x_z^{NT}||_2^2 \le \Gamma_1(\log t_z)^{-\epsilon/2}$ also holds. Let $V_1 > 0$ be the smallest integer satisfying $\lceil(\log V_1)^{1+\epsilon}\rceil > e^2$. Then, for $k \ge \Omega_1 \ge V_1$, we have

$$\begin{aligned}\left|\left|x^D(\hat{\theta}_{t_z}) - x_{z-1}^{NT}\right|\right|_2^2 &\le 3\left|\left|x^D(\hat{\theta}_{t_z}) - x^D(\theta^*)\right|\right|_2^2 + 3\left|\left|x^D(\theta^*) - x^D(\hat{\theta}_{t_{z-1}})\right|\right|_2^2 + 3\left|\left|x^D(\hat{\theta}_{t_{z-1}}) - x_{z-1}^{NT}\right|\right|_2^2 \\ &\le \frac{3\bar{\kappa}^2}{(\log t_z)^{\frac{\epsilon}{2}}} + \frac{3\bar{\kappa}^2}{(\log t_{z-1})^{\frac{\epsilon}{2}}} + \frac{3\Gamma_1}{(\log t_{z-1})^{\frac{\epsilon}{2}}} \\ &\le \frac{3\bar{\kappa}^2}{(\log t_z)^{\frac{\epsilon}{2}}} + \frac{3\bar{\kappa}^2}{(\log \sqrt{t_z})^{\frac{\epsilon}{2}}} + \frac{3\Gamma_1}{(\log \sqrt{t_z})^{\frac{\epsilon}{2}}} \\ &\le 3\left[\bar{\kappa}^2 + 2^{\frac{\epsilon}{2}}(\bar{\kappa}^2 + \Gamma_1)\right]\frac{1}{(\log t_z)^{\frac{\epsilon}{2}}},\end{aligned}$$

where the second inequality follows by definition of $\mathcal{A}$ and induction hypothesis, the third inequality follows because $t_{z-1} \ge \frac{t_z}{2} \ge \sqrt{t_z} \ge \sqrt{t_1} = \sqrt{\lceil(\log k)^{1+\epsilon}\rceil} > e$ when $k \ge \Omega_1 \ge V_1$. Let $V_2 \ge V_1$ be such that for all $k \ge V_2$ and $z = 1, \ldots, Z$, the following hold: (1) $(\log t_z)^{\epsilon/2} \ge 3\gamma^{-2}\left[\bar{\kappa}^2 + 2^{\epsilon/2}(\bar{\kappa}^2 + \Gamma_1)\right]$ and (2) $9\xi^2\left[\bar{\kappa}^2 + 2^{\epsilon/2}(\bar{\kappa}^2 + \Gamma_1)\right]^2(\log t_z)^{-\epsilon/2} \le 1 \le \Gamma_1$. (Recall that γ and ξ are the constants for the locally quadratic convergence of Newton's method defined in Lemma EC.5.) Inequality (1) ensures that $||x^D(\hat{\theta}_{t_z}) - x_{z-1}^{NT}||_2 \le \gamma$ for all $k \ge \Omega_1 \ge V_2$ and inequality (2) ensures, by the locally quadratic

convergence of the Newton's method, that $||x^D(\hat{\theta}_{t_z}) - x_z^{NT}||_2^2 \le \xi^2 ||x^D(\hat{\theta}_{t_z}) - x_{z-1}^{NT}||_2^4 \le \Gamma_1 (\log t_z)^{-\epsilon/2}$. This completes the induction.

(b) Note that $\hat{\lambda}_t \in \Lambda_{\hat{\theta}_t}$ is equivalent to $\hat{p}_t \in \mathcal{P}$ which is immediately satisfied if $\hat{p}_t \in \mathsf{Ball}(p^D(\theta^*), 7\phi/8) \subseteq \mathsf{Ball}(p^D(\theta^*), \phi) \subseteq \mathcal{P}$ (the last inclusion follows by **A6**); also, since $\mathsf{Ball}(p^D(\theta^*), 7\phi/8) \subseteq \mathcal{W}(\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max})$, when $\hat{p}_t \in \mathsf{Ball}(p^D(\theta^*), 7\phi/8)$, it also implies that $\hat{p}_t = p_t$. Hence, we only need to show $C_t \succ 0$ and $\hat{p}_t \in \mathsf{Ball}(p^D(\theta^*), 7\phi/8)$ for $t_1 \le t < \tau$. Let $V_3 \ge V_2$ be such that for all $k \ge V_3$ and $z = 1, \ldots, Z$, $\left(2\sqrt{\Gamma_1} + 3\bar{\kappa}\right)(\log t_z)^{-\epsilon/4} < \phi/(8\omega)$. We now prove the result by induction. If $\tau \le t_1 + 1$, then there is nothing to prove. Suppose that $\tau > t_1 + 1$. Since $E(t_1) \le \bar{\delta}$ on $\mathcal{A}$, by Lemma EC.1 (a), $p^D(\hat{\theta}_{t_1}) \in \mathsf{Ball}(p^D(\theta^*), \phi/2)$. For $t = t_1 + 1$, we then have $||\hat{p}_{t_1+1} - p^D(\theta^*)||_2 = ||p^D(\hat{\theta}_1) - p^D(\theta^*)||_2 \le \phi/2$, so $\hat{p}_{t_1+1} \in \mathcal{P}$. In addition, note that

$$\begin{aligned}
C_{t_1+1} = C_{t_1} - AD_{t_1+1} &= kC - AS_{t_1} - A\left(\hat{\lambda}_{t_1+1} + \hat{\Delta}_{t_1+1}\right) \\
&= kC - t_1 C + t_1 C - AS_{t_1} - A\left(\lambda_1^{NT} + \hat{\Delta}_{t_1+1}\right) \\
&\succeq (k - t_1 - 1)C + t_1 C - AS_{t_1} - A\hat{\Delta}_{t_1+1} \\
&\succeq (k - t_1 - 1)A\tilde{\lambda}_{\min}\mathbf{e} + t_1 A\tilde{\lambda}_{\min}\mathbf{e} - AS_{t_1} - A\hat{\Delta}_{t_1+1} \\
&= (k - t_1 - 1)A\left(\tilde{\lambda}_{\min}\mathbf{e} - \frac{S_{t_1} - t_1\tilde{\lambda}_{\min}\mathbf{e}}{k - t_1 - 1} - \frac{\hat{\Delta}_{t_1+1}}{k - t_1 - 1}\right) \\
&\succeq (k - t_1 - 1)A\left(\tilde{\lambda}_{\min}\mathbf{e} - \left|\left|\frac{S_{t_1} - t_1\tilde{\lambda}_{\min}\mathbf{e}}{k - t_1 - 1}\right|\right|_2 \mathbf{e} - \left|\left|\frac{\hat{\Delta}_{t_1+1}}{k - t_1 - 1}\right|\right|_2 \mathbf{e}\right) \\
&\succ (k - t_1 - 1)\left(\tilde{\lambda}_{\min} - \psi\right)A\mathbf{e} \succeq 0, \qquad \text{(EC.15)}
\end{aligned}$$

(recall that $S_t = \sum_{s=1}^{t} D_s$) where the first inequality follows because $A\lambda_1^{NT} \preceq C$, the second inequality follows because $A\tilde{\lambda}_{\min}\mathbf{e} \preceq C$ by definition of $\tilde{\lambda}_{\min}$, the fourth (strict) inequality follows by (C1) and $A\mathbf{e} \succ 0$, and the last inequality follows by the definition of ψ. This is the base case. Now suppose $C_s \succ 0, \hat{p}_s \in \mathsf{Ball}(p^D(\theta^*), 7\phi/8)$ for all $s \le t-1$ for some $t - 1 < \tau$ with $t - 1 \in [t_z, t_{z+1} - 1]$. If $t \ge \tau$, there is nothing to prove. So we only need to show that $C_t \succ 0, \hat{p}_t \in \mathsf{Ball}(p^D(\theta^*), 7\phi/8)$ when $t < \tau$. Note that when $t < \tau$, we have $t_z \le t < \tau$. Hence, by definition of $\mathcal{A}$, we have

$$\begin{aligned}
||\hat{p}_t - p^D(\theta^*)||_2 &\le ||\hat{p}_t - p(\lambda_z^{NT}; \hat{\theta}_{t_z})||_2 + ||p(\lambda_z^{NT}; \hat{\theta}_{t_z}) - p^D(\hat{\theta}_{t_z})||_2 + ||p^D(\hat{\theta}_{t_z}) - p^D(\theta^*)||_2 \\
&\le w\left|\left|\sum_{s=t_1+1}^{t-1} \frac{\hat{\Delta}_s}{k-s}\right|\right|_2 + ||p(\lambda_z^{NT}; \hat{\theta}_{t_z}) - p(\lambda^D(\hat{\theta}_{t_z}); \hat{\theta}_{t_z})||_2 + \frac{\phi}{2} \\
&\le \frac{\phi}{4} + \omega||\lambda_z^{NT} - \lambda^D(\hat{\theta}_{t_z})||_2 + \frac{\phi}{2} \le \frac{\phi}{4} + \frac{\phi}{8} + \frac{\phi}{2} = \frac{7\phi}{8}
\end{aligned}$$

where the second inequality follows by Lemma EC.1 (a) and the fact that $E(t_z) < \bar{\delta}$ on $\mathcal{A}$, the third inequality follows by **A1** and (C1) the last inequality results from the following inequality

$$\left|\left|\lambda_z^{NT} - \lambda^D(\hat{\theta}_{t_z})\right|\right|_2 \le \left|\left|\lambda_z^{NT} - \lambda^D(\theta^*)\right|\right|_2 + \left|\left|\lambda^D(\theta^*) - \lambda^D(\hat{\theta}_{t_z})\right|\right|_2$$

$$\begin{aligned}
&\le \left|\left|x_z^{NT} - x^D(\hat{\theta}_{t_z})\right|\right|_2 + \left|\left|x^D(\hat{\theta}_{t_z}) - x^D(\theta^*)\right|\right|_2 + \left|\left|\lambda^D(\theta^*) - \lambda^D(\hat{\theta}_{t_z})\right|\right|_2 \\
&\le 2\sqrt{\Gamma_1}(\log t_z)^{-\frac{\epsilon}{4}} + 3\bar{\kappa}E(t_z) \\
&\le \left(2\sqrt{\Gamma_1} + 3\bar{\kappa}\right)(\log t_z)^{-\frac{\epsilon}{4}} \;<\; \frac{\phi}{8\omega},
\end{aligned}$$

where the second inequality follows by (EC.14) and the fourth inequality follows by the definition of $\mathcal{A}$. Hence, $\hat{p}_t \in \mathsf{Ball}(p^D(\theta^*), 7\phi/8)$. For C_t, by a similar argument to the derivations in (EC.15), we have $C_t = kC - tC + tC - AS_{t_1} - \sum_{s=t_1+1}^{t} A(\lambda_{z(s)}^{NT} - \sum_{v=t_1+1}^{s-1} \frac{\hat{\Delta}_v}{k-v} + \hat{\Delta}_s) \succeq (k-t)C + t_1 C - AS_{t_1} - \sum_{s=t_1+1}^{t}(A\hat{\Delta}_s - \sum_{v=t_1+1}^{s-1} \frac{A\hat{\Delta}_v}{k-v}) \succeq (k-t)C + t_1 A\tilde{\lambda}_{\min}\mathbf{e} - AS_{t_1} - \sum_{s=t_1+1}^{t} \frac{A\hat{\Delta}_s(k-t)}{k-s} = (k-t)A(\tilde{\lambda}_{\min}\mathbf{e} - \frac{S_{t_1} - t_1\tilde{\lambda}_{\min}\mathbf{e}}{k-t} - \sum_{s=t_1+1}^{t} \frac{A\hat{\Delta}_s}{k-s}) \succ (k-t)A(\tilde{\lambda}_{\min} - \psi)\mathbf{e} \succeq 0$. This completes the induction.

(c) Let $V_4 \ge V_3$ be such that $27\xi^2\left(5\bar{\kappa}^4\left[8\eta_4 + 4(q-1)^2(\log t_z)^2\right]/(\eta_5^2 t_z) + 2\Gamma_1\Gamma_2/(\log t_{z-1})^{\frac{\epsilon}{2}}\right) < 1$ for all $k \ge V_4$ and $z = 1, \ldots, Z$, where $\Gamma_2 = \max\{1, 4\bar{\kappa}^2\eta_3^2\}$, η_4 and η_5 are as in Lemma 2. Again, we show by induction. For $z = 1$, we have:

$$\begin{aligned}
&\mathbf{E}^\pi[||x^D(\hat{\theta}_{t_1}) - x_1^{NT}||_2^2 \mathbf{1}_{\{t_1<\tau\}}|\mathcal{A}] = \mathbf{E}^\pi[||x^D(\hat{\theta}_{t_1}) - \lambda^D(\hat{\theta}_{t_1})||_2^2 \mathbf{1}_{\{t_1<\tau\}}|\mathcal{A}] \\
&\le 2\,\mathbf{E}^\pi[||x^D(\hat{\theta}_{t_1}) - x^D(\theta^*)||_2^2 \mathbf{1}_{\{t_1<\tau\}}|\mathcal{A}] + 2\,\mathbf{E}^\pi[||\lambda^D(\hat{\theta}_{t_1}) - \lambda^D(\theta^*)||_2^2 \mathbf{1}_{\{t_1<\tau\}}|\mathcal{A}] \\
&\le 4\bar{\kappa}^2\frac{\eta_3^2}{t_1} \;\le\; \frac{\Gamma_2}{t_1},
\end{aligned}$$

where the second to the last inequality follows by Lemma 1. This is our base case. We now do the inductive step. Suppose that $\mathbf{E}^\pi[||x^D(\hat{\theta}_{t_s}) - x_s^{NT}||_2^2 \mathbf{1}_{\{t_s<\tau\}}|\mathcal{A}] \le \Gamma_2 t_s^{-1}$ holds for $s = z-1$, we need to show that same thing holds for $s = z$. Then, for $k \ge \Omega_1 \ge V_4$, we have:

$$\begin{aligned}
&\mathbf{E}^\pi\left[\left|\left|x^D(\hat{\theta}_{t_z}) - x_z^{NT}\right|\right|_2^2 \mathbf{1}_{\{t_z<\tau\}}\middle|\mathcal{A}\right] \le \xi^2\mathbf{E}^\pi\left[\left|\left|x^D(\hat{\theta}_{t_z}) - x_{z-1}^{NT}\right|\right|_2^4 \mathbf{1}_{\{t_z<\tau\}}\middle|\mathcal{A}\right] \\
&\le 27\xi^2\left\{\mathbf{E}^\pi\left[\left|\left|x^D(\hat{\theta}_{t_z}) - x^D(\theta^*)\right|\right|_2^4 \mathbf{1}_{\{t_z<\tau\}}\middle|\mathcal{A}\right] + \mathbf{E}^\pi\left[\left|\left|x^D(\theta^*) - x^D(\hat{\theta}_{t_{z-1}})\right|\right|_2^4 \mathbf{1}_{\{t_z<\tau\}}\middle|\mathcal{A}\right]\right. \\
&\qquad \left. + \mathbf{E}^\pi\left[\left|\left|x^D(\hat{\theta}_{t_{z-1}}) - x_{z-1}^{NT}\right|\right|_2^4 \mathbf{1}_{\{t_z<\tau\}}\middle|\mathcal{A}\right]\right\} \\
&\le 27\xi^2\left\{\bar{\kappa}^4\mathbf{E}^\pi[E(t_z)^4\mathbf{1}_{\{t_z<\tau\}}|\mathcal{A}] + \bar{\kappa}^4\mathbf{E}_{\theta^*}^\pi[E(t_{z-1})^4\mathbf{1}_{\{t_z<\tau\}}|\mathcal{A}] + \frac{\Gamma_1}{(\log t_{z-1})^{\frac{\epsilon}{2}}}\frac{\Gamma_2}{t_{z-1}}\right\} \\
&\le 27\xi^2\left\{\frac{8\eta_4 + 4(q-1)^2(\log t_z)^2}{\eta_5^2 t_z^2}\bar{\kappa}^4 + \frac{8\eta_4 + 4(q-1)^2(\log t_{z-1})^2}{\eta_5^2 t_{z-1}^2}\bar{\kappa}^4 + \frac{\Gamma_1}{(\log t_{z-1})^{\frac{\epsilon}{2}}}\frac{2\Gamma_2}{t_z}\right\} \\
&\le 27\xi^2\left\{\frac{5\bar{\kappa}^4\left[8\eta_4 + 4(q-1)^2(\log t_z)^2\right]}{\eta_5^2 t_z} + \frac{2\Gamma_1\Gamma_2}{(\log t_{z-1})^{\frac{\epsilon}{2}}}\right\}\frac{1}{t_z} \\
&\le \frac{1}{t_z} \;\le\; \frac{\Gamma_2}{t_z},
\end{aligned}$$

where the first inequality follows by Lemma EC.6 (a), the third inequality follows by Lemma EC.4, Lemma EC.6 (a) and the induction hypothesis, and the fourth inequality holds because

Lemma EC.6 (b) shows that $p_s \in \mathcal{W}(\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max})$ for $s < \tau$ which means that the condition for Lemma 2 is satisfied, so

$$\begin{aligned}
&\mathbf{E}^\pi\left[||\theta - \hat{\theta}_t||_2^4 \mathbf{1}_{\{t<\tau\}} \Big| \mathcal{A}\right] \\
&\quad\le \int_0^\infty \mathbb{P}^\pi\left(||\hat{\theta}_t - \theta^*||_2^4 \mathbf{1}_{\{t<\tau\}} \ge x | \mathcal{A}\right) dx \\
&\quad\le \int_0^\infty \min\left\{1, \eta_4 t^{q-1} \exp\left(-\eta_5 t\sqrt{x}\right)\right\} dx \\
&\quad\le \int_0^{\left(\frac{2(q-1)\log t}{\eta_5 t}\right)^2} dx + \int_{\left(\frac{2(q-1)\log t}{\eta_5 t}\right)^2}^\infty \left[\eta_4 t^{q-1} \exp\left(-\frac{\eta_5 t\sqrt{x}}{2}\right)\right] \exp\left(-\frac{\eta_5 t\sqrt{x}}{2}\right) dx \\
&\quad\le \frac{4(q-1)^2(\log t)^2}{\eta_5^2 t^2} + \eta_4 \int_{\left(\frac{2(q-1)\log t}{\eta_5 t}\right)^2}^\infty \exp\left(-\frac{\eta_5 t\sqrt{x}}{2}\right) dx \\
&\quad\le \frac{4(q-1)^2(\log t)^2}{\eta_5^2 t^2} + \eta_4 \int_0^\infty \exp\left(-\frac{\eta_5 t\sqrt{x}}{2}\right) dx \\
&\quad\le \frac{8\eta_4 + 4(q-1)^2(\log t)^2}{\eta_5^2 t^2}.
\end{aligned}$$

This completes the induction. □

Proof of Lemma EC.7. We first derive a bound for Φ_ϵ. By definition $t_z = \lceil (t_{z+1} - L)/2 \rceil + L$ for $z > 1$, so $t_z - L \ge (t_{z+1} - L)/2$. This implies that $t_{z+1} - t_z \le t_z$ for all $z > 1$. For $z = 1$, we also have $t_2 - t_1 = 1 \le L = t_1$. Recall that $Z \le \lceil \log_2 k \rceil \le 2\log_2 k = \frac{2}{\log_e 2} \log k$. Thus, we can bound Φ_ϵ as follows:

$$\begin{aligned}
\Phi_\epsilon = \sum_{s=t_1+1}^{k-1} \bar{\epsilon}(s)^2 \;=\; \sum_{z=1}^{Z} (t_{z+1} - t_z) \bar{\epsilon}(t_z)^2 &\le \sum_{z=1}^{Z} (t_{z+1} - t_z)\, \eta_6^2\, \frac{(q-1)\log t_z + 1}{t_z} \\
&\le \eta_6^2 Z[(q-1)\log k + 1] \\
&\le K_\Phi[1 + \log k + (q-1)\log^2 k]
\end{aligned}$$

for some positive constant K_Φ independent of $k \ge 1$.

We now derive a bound for Ψ_ϵ. To do that, we first show that there exists a constant $K > 3$ such that for all $k \ge K$, (1)$(\log k)^{1+\epsilon}/k < 1/19$, (2)$Z \ge 3$ and (3)$t_{Z-2} \le k/3$. Note that as $k \to \infty$, we have $(\log k)^{1+\epsilon}/k \to 0$, $Z \to \infty$ and $t_{z+1} - L \to \infty$ for $z = Z-2, Z-1, Z$. This implies that $t_z - L = \lceil (t_{z+1} - L)/2 \rceil \le 2(t_{z+1} - L)/3$ for $z = Z-2, Z-1, Z$ when k is large. Therefore, there exists a constant $K > 3$ such that for all $k \ge K$, we have $(\log k)^{1+\epsilon}/k < 1/19$, $Z \ge 3$ and $t_{Z-2} \le \frac{8}{27}(t_{Z+1} - L) + L = \frac{8}{27}k + \frac{19}{27}(\log k)^{1+\epsilon} < \frac{k}{3}$.

Since $\bar{\epsilon}(t_z) = \eta_6 \sqrt{[(q-1)\log t_z + 1]/t_z} \le \eta_6\sqrt{q}$, we conclude that for $k < K$, $\Psi_\epsilon \le K(K\eta_6\sqrt{q})^2 = K^3\eta_6^2 q$. We now focus on the case when $k \ge K$. Note that,

$$\Psi_\epsilon = \sum_{t=t_1+1}^{k-1} \left(\sum_{s=t_1+1}^{t-1} \frac{\bar{\epsilon}(s)}{k-s}\right)^2 \le 2\sum_{t=t_1+1}^{k-1} \left(\sum_{s=t_1+1}^{t_{Z-2}} \frac{\bar{\epsilon}(s)}{k-s}\right)^2 + 2\sum_{t=t_{Z-2}+1}^{k-1} \left(\sum_{s=t_{Z-2}+1}^{t-1} \frac{\bar{\epsilon}(s)}{k-s}\right)^2 \quad \text{(EC.16)}$$

Since $t_{Z-2} > k/4$ (recall that $t_{z+1} \le 2t_z$ and $t_{Z+1} = k$), we have $\bar{\epsilon}(s) < \eta_6\sqrt{4[(q-1)\log k+1]/k}$ for all $s > t_{Z-2}$. So, for all $k \ge K$, the second term in (EC.16) can be bounded by

$$\frac{8\eta_6^2[1+(q-1)\log k]}{k}\sum_{t=t_{Z-2}+1}^{k-1}\left(\sum_{s=t_{Z-2}+1}^{t-1}\frac{1}{k-s}\right)^2 \le \frac{8\eta_6^2[1+(q-1)\log k]}{k}3k \le K_{\Psi,2}[1+(q-1)\log k],$$

for some positive constant $K_{\Psi,2} = 24\eta_6^2$ independent of $k \ge K$, where the first inequality follows since when $k \ge K > 3$ the following holds:

$$\begin{aligned}\sum_{t=t_1+1}^{k-1}\left(\sum_{s=t_1+1}^{t-1}\frac{1}{k-s}\right)^2 &\le \sum_{t=1}^{k-1}\left(\sum_{s=1}^{t-1}\frac{1}{k-s}\right)^2 \le \sum_{t=1}^{k-1}\left(\int_1^t\frac{1}{k-s}ds\right)^2 \le \sum_{t=1}^{k-1}\log^2\left(\frac{k}{k-t}\right)\\ &\le \log^2 k + \int_1^{k-1}\log^2\left(\frac{k}{k-t}\right)dt \le \log^2 k + 2k \le 3k.\end{aligned}$$

As for the first term in (EC.16), for all $k \ge K$, we have

$$\begin{aligned}2\sum_{t=t_1+1}^{k-1}\left(\sum_{s=t_1+1}^{t_{Z-2}}\frac{\bar{\epsilon}(s)}{k-s}\right)^2 &\le 2k\left(\sum_{s=t_1+1}^{t_{Z-2}}\frac{\bar{\epsilon}(s)}{k-s}\right)^2\\ &\le 2k\left(\sum_{z=1}^{Z-3}\frac{t_{z+1}-t_z}{k-t_{z+1}}\eta_6\sqrt{\frac{1+(q-1)\log t_z}{t_z}}\right)^2\\ &\le 4k\eta_6^2\left(\sum_{z=1}^{Z-3}\frac{t_{z+1}-t_z}{k-t_{z+1}}\sqrt{\frac{1+(q-1)\log k}{t_{z+1}}}\right)^2\\ &\le 4k\eta_6^2[1+(q-1)\log k]\left(\int_1^{t_{Z-2}}\frac{1}{k-x}\sqrt{\frac{1}{x}}dx\right)^2\\ &\le 4k\eta_6^2[1+(q-1)\log k]\left(\frac{2\log(\frac{\sqrt{2}}{\sqrt{2}-1})}{\sqrt{k}}\right)^2 \le K_{\Psi,1}[1+(q-1)\log k]\end{aligned}$$

where $K_{\Psi,1} = 16\eta_6^2\log^2(\frac{\sqrt{2}}{\sqrt{2}-1})$. The second inequality follows by Lemma 2. The third inequality follows because $t_{z+1} \le 2t_z$. Note that the function $f(x) = \frac{1}{(k-x)\sqrt{x}}$ is decreasing when $x < \frac{k}{3}$. Since $t_{Z-2} < \frac{k}{3}$, the fourth inequality holds by integral approximation. The fifth inequality follows by

$$\begin{aligned}\int_1^{t_{Z-2}}\frac{1}{k-x}\sqrt{\frac{1}{x}}dx &= \frac{1}{\sqrt{k}}\int_1^{t_{Z-2}}\left(\frac{1}{\sqrt{k}-\sqrt{x}}+\frac{1}{\sqrt{k}+\sqrt{x}}\right)d\sqrt{x}\\ &\le \frac{2}{\sqrt{k}}\log\left(\frac{2\sqrt{k}}{\sqrt{k}-\sqrt{t_{Z-2}}}\right) \le \frac{2\log(\frac{2\sqrt{3}}{\sqrt{3}-1})}{\sqrt{k}}.\end{aligned}$$

Thus, we conclude that there exists some positive constant K_Ψ independent of $k \ge 1$ such that $\Psi_\epsilon \le \max\{(K_{\Psi,1}+K_{\Psi,2})[1+(q-1)\log k], K^3\eta_6^2 q\} \le K_\Psi[1+(q-1)\log k]$. The result follows by letting $K_3 = \max\{K_\Phi, K_\Psi\}$. $\square$

Proof of Lemma EC.8. Because τ is non-negative, we can write $\mathbf{E}^\pi[k-\tau|\mathcal{A}] = k - \sum_{t=0}^{k-1}\mathbb{P}^\pi(\tau > t|\mathcal{A}) = \sum_{t=1}^{k-1}\mathbb{P}^\pi(\tau \le t|\mathcal{A})$. We now bound $\mathbb{P}^\pi(\tau \le t|\mathcal{A})$. To that end, let $\tilde{\tau}$ be the minimum of k and the first time $t \ge t_1+1$ such that the following condition (C2) is violated:

$$(C2)\ \ \psi > \left|\left|\sum_{s=t_1+1}^{t}\frac{\Delta_s}{k-s}\right|\right|_2 + \sum_{s=t_1+1}^{t}\frac{||\lambda_s-\hat{\lambda}_s||_2\mathbf{1}_{\{s\le\tau\}}}{k-s} + \left|\left|\frac{S_L - L\tilde{\lambda}_{\min}\mathbf{e}}{k-t}\right|\right|_2.$$

Note that $\tilde{\tau} \ge \tau$ on every sample path, so $\mathbb{P}^\pi(\tau \le t|\mathcal{A}) \le \mathbb{P}^\pi(\tilde{\tau} \le t|\mathcal{A})$. Then, by the union bound,

$$\begin{aligned}
\mathbb{P}^\pi(\tau \le t|\mathcal{A}) &\le \mathbb{P}^\pi\left(\max_{L+1\le s\le t}\left\{\left|\left|\frac{S_L - L\tilde{\lambda}_{\min}\mathbf{e}}{k-s}\right|\right|_2 + \left|\left|\sum_{v=L+1}^{s}\frac{\Delta_v}{k-v}\right|\right|_2 + \sum_{v=L+1}^{s}\frac{||\lambda_v-\hat{\lambda}_v||_2\mathbf{1}_{\{v\le\tau\}}}{k-v}\right\} \ge \psi\middle|\mathcal{A}\right)\\
&\le \mathbb{P}^\pi\left(\max_{L+1\le s\le t}\left|\left|\frac{S_L - L\tilde{\lambda}_{\min}\mathbf{e}}{k-s}\right|\right|_2 \ge \frac{\psi}{2}\middle|\mathcal{A}\right) + \mathbb{P}^\pi\left(\max_{L+1\le s\le t}\left|\left|\sum_{v=L+1}^{s}\frac{\Delta_v}{k-v}\right|\right|_2 \ge \frac{\psi}{4}\middle|\mathcal{A}\right)\\
&\quad + \mathbb{P}^\pi\left(\max_{L+1\le s\le t}\sum_{v=L+1}^{s}\frac{||\lambda_v-\hat{\lambda}_v||_2\mathbf{1}_{\{v\le\tau\}}}{k-v} \ge \frac{\psi}{4}\middle|\mathcal{A}\right) \qquad \text{(EC.17)}
\end{aligned}$$

We now bound the three terms in (EC.17) starting from the first term. By Markov's inequality:

$$\begin{aligned}
\mathbb{P}^\pi\left(\max_{L+1\le s\le t}\left|\left|\frac{S_L - L\tilde{\lambda}_{\min}\mathbf{e}}{k-s}\right|\right|_2 \ge \frac{\psi}{2}\middle|\mathcal{A}\right) &\le \mathbb{P}^\pi\left(\frac{\left|\left|S_L - L\tilde{\lambda}_{\min}\mathbf{e}\right|\right|_2^2}{(k-t)^2} \ge \frac{\psi^2}{4}\middle|\mathcal{A}\right)\\
&\le \min\left\{1, \frac{4}{\psi^2}\mathbf{E}^\pi\left[\frac{\left|\left|S_L - L\tilde{\lambda}_{\min}\mathbf{e}\right|\right|_2^2}{(k-t)^2}\middle|\mathcal{A}\right]\right\}\\
&\le \min\left\{1, \frac{4n(1+\tilde{\lambda}_{\min})^2L^2}{\psi^2(k-t)^2}\right\},
\end{aligned}$$

where the last inequality follows because $\left|\left|S_L - L\tilde{\lambda}_{\min}\mathbf{e}\right|\right|_2 \le \left|\left|L\mathbf{e} + L\tilde{\lambda}_{\min}\mathbf{e}\right|\right|_2 = \sqrt{n}(1+\tilde{\lambda}_{\min})L$. For the second term in (EC.17), we have

$$\begin{aligned}
\mathbb{P}^\pi\left(\max_{L+1\le s\le t}\left|\left|\sum_{v=L+1}^{s}\frac{\Delta_v}{k-v}\right|\right|_2 \ge \frac{\psi}{4}\middle|\mathcal{A}\right) &= \mathbb{P}^\pi\left(\max_{L+1\le s\le t}\left|\left|\sum_{v=L+1}^{s}\frac{\Delta_v}{k-v}\right|\right|_2^2 \ge \frac{\psi^2}{16}\middle|\mathcal{A}\right)\\
&\le \frac{1}{\mathbb{P}^\pi(\mathcal{A})}\,\mathbb{P}^\pi\left(\max_{L+1\le s\le t}\left|\left|\sum_{v=L+1}^{s}\frac{\Delta_v}{k-v}\right|\right|_2^2 \ge \frac{\psi^2}{16}\right)\\
&\le \frac{16}{\psi^2\mathbb{P}^\pi(\mathcal{A})}\,\mathbf{E}^\pi\left[\left|\left|\sum_{s=L+1}^{t}\frac{\Delta_s}{k-s}\right|\right|_2^2\right] = \frac{16}{\psi^2\mathbb{P}^\pi(\mathcal{A})}\,\mathbf{E}^\pi\left[\sum_{s=L+1}^{t}\frac{||\Delta_s||_2^2}{(k-s)^2}\right]\\
&\le \frac{16}{\psi^2\mathbb{P}^\pi(\mathcal{A})}\left[\frac{4}{(k-t)^2} + \frac{4}{k-t}\right] \le \frac{128}{\psi^2\mathbb{P}^\pi(\mathcal{A})}\frac{1}{k-t},
\end{aligned}$$

where the first inequality follows by the law of total probability, the second inequality follows by the Doob's sub-martingale inequality, the second inequality follows by the orthogonality of

martingale differences, and the third inequality follows by the same integral approximation bound as in deriving (EC.13). We now bound the last term in (EC.17):

$$\begin{aligned}
&\mathbb{P}^\pi\left(\max_{L+1\le s\le t}\sum_{v=L+1}^{s}\frac{||\lambda_v-\hat{\lambda}_v||_2\mathbf{1}_{\{v\le\tau\}}}{k-v}\ge\frac{\psi}{4}\middle|\mathcal{A}\right)\le\frac{16}{\psi^2}\left(\sum_{s=L+1}^{t}\frac{\sqrt{\mathbf{E}^\pi[||\lambda_s-\hat{\lambda}_s||_2^2\mathbf{1}_{\{\tau\le s\}}|\mathcal{A}]}}{k-s}\right)^2\\
&\le\frac{16}{\psi^2}\left(\sum_{s=L+1}^{t}\frac{\sqrt{\mathbf{E}^\pi[||\lambda_s-\hat{\lambda}_s||_2^2\mathbf{1}_{\{\tau<s\}}|\mathcal{A}]}+\sqrt{\mathbf{E}^\pi[||\lambda_s-\hat{\lambda}_s||_2^2\mathbf{1}_{\{\tau=s\}}|\mathcal{A}]}}{k-s}\right)^2\\
&\le\frac{32}{\psi^2}\left(\sum_{s=L+1}^{t}\frac{\sqrt{\mathbf{E}^\pi[||\lambda_s-\hat{\lambda}_s||_2^2\mathbf{1}_{\{\tau<s\}}|\mathcal{A}]}}{k-s}\right)^2+\frac{32}{\psi^2}\left(\sum_{s=L+1}^{t}\frac{\sqrt{\mathbf{E}^\pi[||\lambda_s-\hat{\lambda}_s||_2^2\mathbf{1}_{\{\tau=s\}}|\mathcal{A}]}}{k-s}\right)^2\\
&\le\frac{32\omega^2}{\psi^2}\left(\sum_{s=L+1}^{t}\frac{2\bar{\epsilon}(s)}{k-s}\right)^2+\frac{32}{\psi^2}\left(\sum_{s=L+1}^{t}\frac{\sqrt{2}\sqrt{\mathbf{E}^\pi[\mathbf{1}_{\{\tau=s\}}|\mathcal{A}]}}{k-s}\right)^2\le\frac{128\omega^2}{\psi^2}\left(\sum_{s=L+1}^{t}\frac{\bar{\epsilon}(s)}{k-s}\right)^2+\frac{128}{\psi^2}\left(\frac{1}{k-t}\right)
\end{aligned}$$

where the first inequality follows by first applying Markov's inequality and then applying Cauchy-Schwartz inequality (as in the derivation of (EC.13)), the fourth inequality follows by Lemma 2 and the fact that for any two points $x_1,x_2\in\mathbf{\Delta}^{n-1}$ we have $||x_1-x_2||_2^2\le 2$, and the last inequality follows because by Cauchy-Schwartz inequality,

$$\left(\sum_{s=L+1}^{t}\frac{\sqrt{\mathbf{E}^\pi[\mathbf{1}_{\{\tau=s\}}|\mathcal{A}]}}{k-s}\right)^2\le\left(\sum_{s=L+1}^{t}\frac{1}{(k-s)^2}\right)\left(\sum_{s=L+1}^{t}\mathbf{E}^\pi[\mathbf{1}_{\{\tau=s\}}|\mathcal{A}]\right)\le\frac{1}{(k-t)^2}+\frac{1}{k-t}\le\frac{2}{k-t}.$$

Finally, we have for all $k\ge K\ge\Omega_2\ge 3$,

$$\begin{aligned}
\mathbf{E}^\pi[k-\tau|\mathcal{A}]&=\sum_{t=1}^{k-1}\mathbb{P}^\pi(\tau\le t|\mathcal{A})\le\sum_{t=1}^{k-1}\frac{128}{\psi^2\mathbb{P}^\pi(\mathcal{A})}\frac{1}{k-t}+\sum_{t=1}^{k-1}\min\left\{1,\frac{4n(1+\tilde{\lambda}_{\min})^2L^2}{\psi^2(k-t)^2}\right\}\\
&\quad+\frac{128\omega^2}{\psi^2}\sum_{t=1}^{k-1}\left(\sum_{s=L+1}^{t}\frac{\bar{\epsilon}(s)}{k-s}\right)^2+\frac{128}{\psi^2}\sum_{t=1}^{k-1}\left(\frac{1}{k-t}\right)\\
&\le\frac{128}{\psi^2}\frac{\log k}{\mathbb{P}^\pi(\mathcal{A})}+L+\sum_{t=1}^{k-L}\frac{4nL^2(1+\tilde{\lambda}_{\min})^2}{\psi^2(k-t)^2}+\frac{128\omega^2}{\psi^2}\sum_{t=L+1}^{k-1}\left(\sum_{s=L+1}^{t-1}\frac{\bar{\epsilon}(s)}{k-s}\right)^2+\frac{128}{\psi^2}\sum_{t=1}^{k-1}\left(\frac{1}{k-t}\right)\\
&\le\frac{256}{\psi^2}\log k+\left(\frac{4n(1+\tilde{\lambda}_{\min})^2}{\psi^2}+1\right)L+\frac{128K_3\omega^2q}{\psi^2}\log k+\frac{128}{\psi^2}\log k\le K_7(\log k+L),
\end{aligned}$$

where $K_7=384/\psi^2+128K_3\omega^2q/\psi^2+4n(1+\tilde{\lambda}_{\min})^2/\psi^2+1$, the second inequality follows by integral approximation, and the third inequality follows by Lemma EC.7. □

EC.5. Implication of the Well-separated Demand Assumption A7

In this section, we prove the following claim we made when commenting on assumption **A1** in Section 4:

LEMMA EC.9. *There exists some constant $c_d > 0$ such that for any $\theta, \theta' \in \Theta$ and $p \in \mathcal{W}(\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max})$, $||\lambda(p;\theta) - \lambda(p;\theta')||_2 \geq c_d||\theta - \theta'||_2$*

Proof. There exists a constant c_h such that

$$\begin{aligned} c_h||\theta-\theta'||_2^2 &\leq \sum_{D\in\mathcal{D}} (\sqrt{\mathbb{P}^{p,\theta}(D)} - \sqrt{\mathbb{P}^{p,\theta'}(D)})^2 \\ &= \sum_{D\in\mathcal{D}} \left(\frac{\mathbb{P}^{p,\theta}(D) - \mathbb{P}^{p,\theta'}(D)}{\sqrt{\mathbb{P}^{p,\theta}(D)} + \sqrt{\mathbb{P}^{p,\theta'}(D)}} \right)^2 \leq \frac{\sum_{D\in\mathcal{D}} (\mathbb{P}^{p,\theta}(D) - \mathbb{P}^{p,\theta'}(D))^2}{4(\tilde{\lambda}_{\min})^n (1-\tilde{\lambda}_{\max})} \\ &= \frac{\sum_{i=1}^n [\lambda_i(\theta;p) - \lambda_i(\theta';p)]^2 + [(1-\sum_{j=1}^n \lambda_j(\theta;p)) - (1-\sum_{j=1}^n \lambda_j(\theta';p))]^2}{4(\tilde{\lambda}_{\min})^n (1-\tilde{\lambda}_{\max})} \\ &\leq \frac{(n+1)\sum_{i=1}^n [\lambda_i(\theta;p) - \lambda_i(\theta';p)]^2}{4(\tilde{\lambda}_{\min})^n (1-\tilde{\lambda}_{\max})} = \frac{(n+1)}{4(\tilde{\lambda}_{\min})^n (1-\tilde{\lambda}_{\max})} ||\lambda(\theta;p) - \lambda(\theta';p)||_2^2 \end{aligned}$$

for any $\theta, \theta' \in \Theta$ and any $p \in \mathcal{W}(\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max})$, where the first inequality follows by Lemma EC.3, the second inequality follows since $p \in \mathcal{W}(\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max})$, and the last inequality follows since $(\sum_{i=1}^n x_i)^2 \leq n\sum_{i=1}^n x_i^2$. Setting $c_d = 4c_h(\tilde{\lambda}_{\min})^n(1-\tilde{\lambda}_{\max})/(n+1)$ completes the proof. □

References

Borovkov AA (1999) *Mathematical Statistics* (CRC Press).

Boyd S, Vandenberghe L (2004) *Convex Optimization* (Cambridge University Press).

Broder J, Rusmevichientong P (2012) Dynamic pricing under a general parametric choice model. *Oper. Res.* 60:965–980.

Chen QG, Jasin S, Duenyas I (2019) Nonparametric self-adjusting control for joint learning and optimization of multi-product pricing with finite resource capacity. *Math. Oper. Res.* 44:601–631.

Williams D (1991) *Probability with martingales* (Cambridge university press).