# Chasing Demand: Learning and Earning in a Changing Environment


N. Bora Keskin*
University of Chicago

Assaf Zeevi†
Columbia University


This version: December 25, 2013


**Abstract**

We consider a dynamic pricing problem in which a seller faces an unknown demand model that can change over time. We measure the amount of change over a time horizon of $T$ periods using a quadratic variation metric, and allow a finite "budget" for such changes. We first derive a lower bound on the expected performance gap between any pricing policy and a clairvoyant who knows a priori the temporal evolution of the underlying demand model, and then design families of near-optimal pricing policies, the revenue performance of which asymptotically matches said lower bound. We also show that the seller can achieve a substantially better revenue performance in demand environments that change in "bursts" than it would in a demand environment that changes "smoothly." Finally, we extend our analysis to the case of rapidly changing demand settings, and obtain a range of results that quantify the net effect of the volatility in the demand environment on the seller's revenue performance.



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


# 1 Introduction

## 1.1 Background and Overview

Pricing under demand uncertainty often involves a tradeoff between *learning* about customers sensitivity to price variations, and *earning* short-term revenues. As a motivating example, consider the practice of evaluating loan applications in the financial sector. Because the applications are evaluated and approved on an individual basis, commercial banks and other financial institutions that sell short-term loans can offer a different interest rate to every customer. As noted by Phillips (2005) this particular transaction structure in consumer lending, which is called *customized pricing*,


*Booth School of Business, e-mail: bora.keskin@chicagobooth.edu

†Graduate School of Business, e-mail: assaf@gsb.columbia.edu

offers relatively seamless opportunities for price experimentation to learn about customer behavior. Representing price sensitivity of customers in the form of a demand curve, a firm can use historical as well as real time sales data to form estimates of the latter, while concurrently accumulating revenues from new sales. A key question in this context concerns the "perishability" of useful sales data, primarily due to changes in the demand environment.

Studies on dynamic pricing with demand model uncertainty have, by and large, focused almost exclusively on stylized settings where the demand environment (which is to be explored) does not change over time. The main focus of this paper is to extend this literature by formulating and studying a *time-varying* demand environment, and identifying some qualitative insights that arise from the learning-and-earning tradeoff in that case.

The particular learning-and-earning problem we consider has the following key features: (a) there is a seller who can dynamically change the price of its product over time; (b) the seller can observe the demand for its product, which depends on price and some unknown demand parameters; and (c) the unknown demand parameters can change over time. The seller's goal is to accumulate maximal revenues over a given time horizon, which could be achieved either by focusing on immediate revenues, or by learning the demand parameters to increase future revenues, or some combination thereof. Problem feature (c), which is the novel element in this study, motivates the seller to keep track of changing market conditions. We quantify the total amount of change over the time horizon using a quadratic variation metric in the demand model parameters. We measure the performance of a dynamic pricing policy using the growth rate of its *regret*: the expected revenue loss of a policy, as a function of the time horizon $T$, compared to a clairvoyant that knows the changing demand parameters. As will be explained in detail later, we first derive a lower bound on the minimum achievable growth rate of the regret, which *must* be incurred by *any* admissible policy, and then construct policies which admit a matching upper bound, and are hence optimal in order.

## 1.2 Main Contributions and Qualitative Insights

**Summary of high level contributions.** This paper makes three main contributions to the literature on dynamic pricing with demand model uncertainty. First, we find a sharp difference between "smooth" and "bursty" changes in a demand environment, analyzing the best achievable revenue performance and the structure of asymptotically optimal policies in each case. Second, our analysis addresses a three-way tradeoff between learning, earning, and *forgetting*, which is present in dynamic pricing problems in changing environments. To see the role of forgetting in this tradeoff, we derive several new results that establish the complexity of the problem. Comparing our results with previous performance bounds obtained in recent studies on dynamic pricing with demand learning, we are able to quantify the net effect of a changing demand environment on the seller's aggregate revenue. Third, the policies we construct provide simple guidelines for experimentation in changing environments. In the case of smooth changes, we employ a weighted least squares estimation procedure that discounts older observations at an (asymptotically) optimal rate, whereas in the case of bursty changes we build a joint pricing and detection policy that repetitively tests if there has been a significant change in the environment.

**On smooth versus bursty changes.** In this paper we identify two families of changing demand environments that stand in stark contrast in terms of (a) what the best achievable revenue performance is, and (b) how firms should use pricing as a learning tool. The first family of demand environments is characterized by *smooth* changes (see the setting formulated in Section 2 and studied in Section 3), whereas the second family of demand environments is characterized by *bursty* changes (see Section 4). With regard to (a), the case of bursty changes seems to present a harsher environment at first glance, simply because at any given time the accumulated demand information can become worthless due to a swift change in the demand model. Somewhat surprisingly, our analysis proves the opposite of this claim. The essential intuition behind this observation is that gradual changes can practically be undetectable, and lead to substantial revenue loss in the case of smooth changes (see the proof and discussion of Theorem 1). With regard to (b), our analysis offers distinct ways to implement successful price experimentation in the two families of changing environments mentioned above. Knowing that undetected changes will lead to severe inaccuracies in estimation, the seller needs to discount the weight of older demand observations while estimating the demand curve in smoothly evolving environments. This gives rise to two practical price experimentation policies: moving windows; and decaying weights (akin to exponential smoothing). These maintain a near-optimal balance between learning, earning, and forgetting (see the proof and discussion of Theorem 3). In the case of bursty changes, we construct a novel detection policy that can simultaneously detect and learn changes, incurring significantly smaller regret than the one characterizing smooth changes (see the proof and discussion of Theorem 4).

**Information depreciation in changing environments.** A distinguishing feature of our analysis is the explicit tradeoff between learning a demand curve, earning immediate revenues, and forgetting obsolete sales data. While the dual tradeoff between learning and earning has been studied extensively in the literature, there is limited work on the three-way tradeoff between learning, earning, and forgetting. A key question here is whether a seller in a *changing* environment should collect information faster (or slower) than a seller in a *static* environment. More rapid information collection is desirable because the seller needs to constantly adapt to time-varying market conditions. On the other hand, slower information collection might also seem preferable because any piece of collected information will lose its value over time, implying that excessive attempts to accumulate information can cost more than its value is worth. The answer to these tradeoffs will depend on how "information" is defined. In static demand environments, the definition of information is fairly obvious; see, e.g., Keskin and Zeevi (2012). In changing environments, information collected up to a certain point in time only represents a *nominal* amount because part of this information becomes obsolete over time. It turns out that this notion can be quantified by considering the smallest eigenvalue of a suitably weighted Fisher information matrix, which measures the *relevant* amount of information in period $t$ (see Section 3.2). Based on these definitions, one can revisit the tradeoffs related to the rate of information collection: a seller facing temporal demand changes should collect a larger amount of nominal information, but maintain a smaller amount of relevant information than a seller facing no demand change. The gap between the nominal and relevant information describes the near-optimal *forgetting rate* in the various demand settings studied in this paper, and moreover, enables us to quantify the time value of information in changing demand environments. For example, if we consider a policy that recalls only the data observed within a moving window, the ratio of the window size to the time horizon describes how fast the policy forgets. In light of this, we introduce an *information depreciation factor*, denoted by $\boldsymbol{\delta}(T)$, defined as the ratio of the near-optimal moving window size in a given environment to time horizon $T$. On the extreme end of the spectrum, in a static environment, $\boldsymbol{\delta}(T)$ is equal to 1. This paper identifies the value of $\boldsymbol{\delta}(T)$ in some non-stationary settings, and hence quantifies the extent of information depreciation.

**Organization of the paper.** This section ends with a review of relevant literature. In Section 2 we formulate the problem, and in Section 3 we analyze it by first deriving a lower bound on the revenue loss of any given policy, and then designing near-optimal policies that achieve the loss rate in said lower bound. In Section 4 we consider demand environments that change in bursts, and construct a near-optimal policy whose performance is substantially better than the near-optimal performance observed in the case of smooth changes. Section 5 extends the results in Section 3 to the case of more rapidly changing demand environments, presenting a range of results that characterize the impact of the volatility in demand environment on the revenue performance. Section 6 contains some concluding remarks. Proofs of all results are in appendices, though proof sketches communicating key intuitive ideas are detailed in the main body.

### 1.3 Related Literature

In recent years, the tradeoff between learning and earning has become a prominent area of study in the literature on dynamic pricing and revenue management (see, e.g., Lobo and Boyd 2003, Araman and Caldentey 2009, Besbes and Zeevi 2009, Farias and van Roy 2010, Harrison, Keskin and Zeevi 2012, Broder and Rusmevichientong 2012, den Boer and Zwart 2012, Keskin and Zeevi 2012, Wang, Deng and Ye 2012), as well as in the broader operations management context (see, e.g., Harrison and Sunar 2013). However, the vast majority of the studies in this area focus on learning in static environments in the sense that the ambient problem setting is unknown but does not change over time. One of the goals of the present paper is to provide a fairly general treatment of learning and earning in dynamically evolving environments, to study its implications on the value of price experimentation, and to illustrate the design of dynamic pricing policies that perform well in such settings.

In the economics literature, there has been considerable effort towards characterizing optimal learning policies in the presence of Markovian shifts in the demand model. As part of that effort, Balvers and Cosimano (1990) and Beck and Wieland (2002) examine dynamic control problems with autoregressive changes in underlying market-response model, while Rustichini and Wolinsky (1995) and Keller and Rady (1999) focus on similar problems in which underlying demand parameters evolve according to a two-state Markov chain. In all of these studies, the decision-maker is assumed to know the transition rule for the time-varying (and unknown) parameters.

Another research stream that targets tracking problems is the statistics literature on change-point detection. As discussed in the survey papers by Lai (1995) and Shiryaev (2010), the essential motivation for change-point detection has been military and quality control applications, making these problems distinct from the tracking problems addressed in this study: in traditional change-point detection, the uncertainty is essentially about the time of change, and it is assumed that the decision-maker knows exactly which model structure will be in place before and after the change; and this literature typically considers only a “passive” observation process, namely, the decision maker cannot influence the measurements being taken. These assumptions do not hold in the dynamic pricing applications that motivate our work.

In the operations research and management science (OR/MS) literature, the most notable studies on tracking problems are those of Aviv and Pazgal (2005), Besbes and Zeevi (2011), Chen and Farias (2013) and den Boer (2013). Aviv and Pazgal (2005) consider a revenue management problem with finite initial inventory, and construct a near-optimal policy, assuming that the underlying demand environment evolve according to a discrete-state-space Markov chain, the transition structure of which is known to the seller. Besbes and Zeevi (2011) consider a dynamic pricing problem in which the demand environment changes at an unknown time in the sales horizon. Assuming that the seller has perfect knowledge about the demand curves before and after the change, they characterize an asymptotically optimal policy for jointly pricing and learning said change. More recently, Chen and Farias (2013) find near-optimal policies in a dynamic pricing problem such that the market size evolves in a Markovian fashion whereas the price-sensitivity of customers remains stationary, and den Boer (2013) studies well-performing pricing policies in a similar problem in which the market size is unknown and can change over time, but the price-sensitivity of customers is known with certainty. In contrast to the economics, statistics, and OR/MS studies mentioned above, we consider a fairly general formulation in which (i) the unknown demand model faced by the seller can change over time, resulting in unobservable changes in the market size *and* the price-sensitivity of consumers, and (ii) the decision maker does not know the particular dynamics of changes in the demand model.

Our work is partially related to time series forecasting methods such as moving average and exponential smoothing, first formulated by Brown (1956), and later analyzed by Holt (1957) and Winters (1960). The main goal of such forecasting methods is to predict future values of a time series, the past values of which are observed in a noisy environment. A distinctive feature of our work is that the time series data in our formulation, which is composed of demand quantities, depend on the seller's pricing policy, and this necessitates the implementation of price experimentation to facilitate active learning. Therefore, "learning" and "forgetting" need to be carried out simultaneously in our dynamic pricing problem. More importantly, assuming that the particular dynamic structure of the underlying price-response curve is unknown, we show that the seller can achieve an asymptotically optimal balance between learning and forgetting by discounting older information using a *polynomial* weighting scheme.

At a high level, our work is also related to the literature on sequential stochastic optimization problems, which are often tackled by means of stochastic approximation type methods. In a recent paper, Besbes, Gur and Zeevi (2013) have examined how methods in the stochastic approximation literature can be used in changing environments. Because such stochastic optimization methods are designed to operate sans any parametric assumptions, they are structurally different than the pricing policies we construct in this paper, though that paper also emphasizes the notion of cumulative variation in the underlying response functions. Insofar as the analysis of our policies, it directly quantifies the net contribution of learning, earning, and forgetting, and consequently, brings forth key insights on how a seller can implement successful experimentation under time-varying model uncertainty. In particular, our work constructs learn-and-earn policies that forget obsolete information at a near-optimal rate, either smoothly or abruptly depending on the nature of changes.

## 2 Problem Formulation

**Basic model elements.** Consider a firm, called *the seller*, that sells a product over a time horizon of $T$ periods. In each period $t = 1, 2, \ldots$ the seller chooses a price $p_t$ for its product from a given interval $[\ell, u]$, where $0 < \ell < u < \infty$. After setting the price $p_t$, the seller observes demand $D_t$, which is given by

$$D_t = \alpha_t + \beta_t p_t + \epsilon_t \qquad \text{for } t = 1, 2, \ldots \tag{2.1}$$

where $\alpha_t \in \mathbb{R}$, $\beta_t \in \mathbb{R}_-$ are the demand model parameters, which are unknown to the seller, and $\epsilon_t$ are unobservable demand shocks. Assume that $\{\epsilon_t\}$ are independent and identically distributed random variables with mean zero and variance $\sigma^2$, and that there exists a positive constant $x_0$ such that $\mathbb{E}[\exp(x\epsilon_t)] < \infty$ for all $|x| \leq x_0$. An important example is where $\epsilon_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$, but it is perhaps useful to note that the homogeneity assumption is not essential; it suffices that the variance of $\epsilon_t$ is bounded, and that the exponential moment condition holds uniformly. For notational brevity, we let $\theta_t = (\alpha_t, \beta_t)$ denote the vector of unknown demand parameters in period $t$, and $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots)$ denote the sequence of demand parameter vectors. Let $\Theta$ be a compact rectangle in $\mathbb{R} \times \mathbb{R}_-$, from which the values of $\theta_t$ are chosen. Given a parameter vector $\theta = (\alpha, \beta) \in \Theta$ and a price $p \in [\ell, u]$, the seller's expected single-period revenue function is

$$r(p, \theta) := p\,(\alpha + \beta p). \tag{2.2}$$

We denote by $\varphi(\theta)$ the feasible price that maximizes the function $r(\cdot, \theta)$, that is

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

**Changing demand environment: the original problem.** We measure the amount of change in $T$ periods with the following quadratic variation metric: define a partition of $\{1, \ldots, T\}$ as any set of periods $\{t_0, t_1, \ldots, t_K\}$ satisfying $1 \leq t_0 < \ldots < t_K \leq T$ for some $K = 1, 2, \ldots$, and denote by $\mathcal{P}$ the set of all partitions of $\{1, \ldots, T\}$. Given time horizon $T$, and a demand vector sequence $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_T)$, let

$$V_{\boldsymbol{\theta}}(T) \;:=\; \sup_{\{t_0, t_1, \ldots, t_K\} \in \mathcal{P}} \left\{ \sum_{k=1}^{K} \|\theta_{t_k} - \theta_{t_{k-1}}\|^2 \right\}. \tag{2.4}$$

The values of $\theta_t$ are chosen from $\Theta$ such that the demand vector sequence $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_T) \in \Theta^T$ satisfies

$$V_{\boldsymbol{\theta}}(T) \leq B \qquad \text{for } T = 1, 2, \ldots \tag{2.5}$$

where $B > 0$. For notational brevity, we denote the set of demand vector sequences $\boldsymbol{\theta}$ satisfying (2.5) as follows:

$$\mathcal{V}(T, B) = \{\boldsymbol{\theta} : V_{\boldsymbol{\theta}}(T) \leq B\}. \tag{2.6}$$

Inequality (2.5) describes a setting in which nature has a finite quadratic variation budget to change the demand parameters throughout the time horizon. We refer to this setting as the *original problem instance*. In Section 4, we analyze a special case of the original problem in which the changes occur in bursts, and in Section 5, we extend our results to the case of more rapidly changing environments where the upper bound in (2.5) can depend on (and increase with) $T$.

To ensure that the optimal price is always feasible, we assume without loss of generality that $\varphi(\theta)$ lies in the interior of $[\ell, u]$ for all $\theta = (\alpha, \beta) \in \Theta$. This implies that $\beta$ is strictly negative for all $(\alpha, \beta) \in \Theta$, and hence, the revenue-maximizing price equals $\varphi\big((\alpha, \beta)\big) = -\alpha/(2\beta)$.

**Pricing policies, induced probabilities, and performance metric.** Let $H_t$ denote the vectorized history of demands and prices observed through the end of period $t$, that is, $H_t = (D_1, p_1, \ldots, D_t, p_t)$. Define a *policy* as a sequence of functions $\pi = (\pi_1, \pi_2, \ldots)$, where $\pi_1$ is a constant function, and for all $t = 1, 2, \ldots,$ $\pi_{t+1}$ is a function from $\mathbb{R}^{2t}$ into $[\ell, u]$, mapping $H_t$ to the price that will be charged in period $t+1$. Any such policy $\pi$ constructs a nonanticipating price sequence $p = (p_1, p_2, \ldots)$, where $p_t$ is determined by the function $\pi_t$, and hence adapted to $H_{t-1}$.

Given a sequence of demand parameter vectors $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots)$ and a pricing policy $\pi$, we define a family of probability measures on the sample space of demand sequences $D = (D_1, D_2, \ldots)$ as follows. Let $\mathbb{P}^\pi_{\boldsymbol{\theta}}$ be a probability measure satisfying

$$\mathbb{P}^\pi_{\boldsymbol{\theta}}(D_1 \in d\xi_1, \ldots, D_T \in d\xi_T) \;=\; \prod_{t=1}^{T} \mathbb{P}_\epsilon(\alpha_t + \beta_t p_t + \epsilon_t \in d\xi_t) \quad \text{for } \xi_1, \ldots, \xi_T \in \mathbb{R}, \tag{2.7}$$

where $\mathbb{P}_\epsilon(\cdot)$ is the probability measure governing the random variables $\{\epsilon_t\}$, and $p = (p_1, p_2, \ldots)$ is the price sequence formed under policy $\pi$ and demand realization $D = (D_1, D_2, \ldots)$.

The performance metric we use in this paper is $T$-period *regret*, defined as

$$\mathcal{R}^\pi(T, B) \;=\; \sup\left\{\Delta^\pi_{\boldsymbol{\theta}}(T) : \boldsymbol{\theta} \in \mathcal{V}(T, B)\right\}, \tag{2.8}$$

where for $T = 1, 2, \ldots$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_T)$

$$\Delta^\pi_{\boldsymbol{\theta}}(T) \;=\; \mathbb{E}^\pi_{\boldsymbol{\theta}}\left\{\sum_{t=1}^{T}\left(1 - \frac{r(p_t, \theta_t)}{r^*(\theta_t)}\right)\right\}, \tag{2.9}$$

$\mathbb{E}^\pi_{\boldsymbol{\theta}}(\cdot)$ is the expectation operator associated with the probability measure $\mathbb{P}^\pi_{\boldsymbol{\theta}}(\cdot)$, $r^*(\theta) := r\big(\varphi(\theta), \theta\big)$ is the optimal single-period revenue function, and $\mathcal{V}(T, B)$ is as defined in (2.6). The regret of a policy is the *worst-case* expected normalized revenue loss relative to a clairvoyant policy that knows the value of $\theta_t$ in every period. Given the normalization in (2.9), this can also be interpreted as the expected number of lost sales opportunities due to not knowing the underlying demand model. Under either interpretation, when the regret of a policy is sublinear in $T$, the policy is long-run-average optimal, and more generally, smaller regret corresponds to uniformly better revenue performance.

# 3 Analysis of the Original Problem

## 3.1 A Lower Bound on Regret

Our first result is a theoretical lower bound on the minimum achievable regret of any pricing policy in the original problem setting described in the preceding section.

**Theorem 1 (lower bound on regret)** *There exists a finite positive constant $c$ such that $\mathcal{R}^\pi(T, B) \geq c\, B^{1/3}\, T^{2/3}$ for any pricing policy $\pi$ and time horizon $T \geq 3$.*

Theorem 1 shows that the $T$-period regret of any given policy is at least on the order of $T^{2/3}$. A policy $\pi$ that achieves the loss rate in Theorem 1, i.e., any policy $\pi$ such that $\mathcal{R}^\pi(T) = O(T^{2/3})$

will hereafter be called *first-order optimal*, and *rate optimal* if the dependence on *both* $B$ and $T$ match the lower bound.

**Rough proof sketch.** The main intuition behind this result is that nature can change the demand parameters in a gradual manner such that it is very costly to detect changes and learn the new demand curve after a change. By carefully choosing a parameter change with squared norm of order $T^{-1/3}$, nature makes sure that: either (i) no detection test can identify this change without incurring a loss of order $T^{1/3}$, or (ii) the cost of learning the new parameter vector is of order $T^{1/3}$. Within its change budget, nature can use $T^{1/3}$ such parameter changes, implying that any given policy must have a loss of order $T^{2/3}$, even if it is designed to simultaneously detect and learn. To prove arguments (i) and (ii), we use the Kullback-Leibler divergence to quantify the difference between the likelihood of events under the probability measures before and after a potential change. For a given policy, if the Kullback-Leibler divergence is smaller than a threshold $\eta$, then we derive argument (i) via hypothesis testing results in information theory. In particular, a Fano-type lower bound on the error probabilities in a detection problem (cf. Tsybakov 2009, Theorem 2.2) implies that there is a significant probability of not detecting the potential change, which consequently leads to a revenue loss of order $T^{-1/3}N$ in the $N$ periods following the potential change. On the other hand, if the Kullback-Leibler divergence is larger than the threshold $\eta$, then we note that, despite the small amount of change, the cost of gathering information on the new demand parameters is bounded away from zero. This implies that the revenue loss until the next change will be of order $T^{1/3}$, as expressed in argument (ii). If there are $N$ periods between two changes, then arguments (i) and (ii) imply that the revenue loss between these two changes is at least of order $(T^{-1/3}N) \wedge T^{1/3}$. We therefore deduce that nature can cause a loss of order $T^{2/3}$ within $T$ periods, by choosing $N$ to be of order $T^{2/3}$ and spreading out potential changes throughout the time horizon. ■

**Discussion and key insights.** The derivation of Theorem 1 brings forth a key insight about the type of policies that could perform well in the setting described in the preceding section. Specifically, the seller can face a sequence of smooth changes that are virtually undetectable, making any effort to detect changes perform poorly. Therefore, successful policies in this environment should not focus on detecting every single change, but instead, they need to *forget* the past at some rate, with the hope that the negative effects of undetectable smooth changes will be filtered out sufficiently fast. The next subsection provides a general estimation procedure that implements this idea by assigning non-increasing weights to older demand observations.

### 3.2 A Weighted Least Squares Estimator

In what follows we describe a procedure to estimate $\theta_{t+1}$ given the history of demands and prices through the end of period $t$. Let $w^t = (w_1^t, \ldots, w_t^t)$ be a $t \times 1$ vector of nonnegative real numbers. Given history vector $H_t$ and weight vector $w^t$, set the *weighted least squares* estimator of $\theta_{t+1}$ to be

$$\hat{\theta}_{t+1} = \arg\min_{\theta}\{SSE_t(\theta, w^t)\}, \tag{3.1}$$

where $SSE_t(\theta, w^t) = \sum_{s=1}^{t} w_s^t (D_s - \alpha - \beta p_s)^2$ for $\theta = (\alpha, \beta)$. The solution of the weighted least squares problem (3.1) is

$$\hat{\theta}_{t+1} = \begin{bmatrix} \hat{\alpha}_t \\ \hat{\beta}_t \end{bmatrix} = \begin{bmatrix} \sum_{s=1}^{t} w_s^t & \sum_{s=1}^{t} w_s^t p_s \\ \sum_{s=1}^{t} w_s^t p_s & \sum_{s=1}^{t} w_s^t p_s^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{s=1}^{t} w_s^t D_s \\ \sum_{s=1}^{t} w_s^t D_s p_s \end{bmatrix}. \tag{3.2}$$

Let us now re-express (2.1) in the following compact form:

$$D_t \;=\; X_t \cdot \theta_t \;+\; \epsilon_t \qquad \text{for } t = 1, 2, \ldots \tag{3.3}$$

where $X_t = \left[\begin{smallmatrix} 1 \\ p_t \end{smallmatrix}\right]$. Then, (3.2) and (3.3) imply that

$$\begin{aligned} \hat{\theta}_{t+1} - \theta_{t+1} \;&=\; \left(\sum_{s=1}^{t} w_s^t X_s X_s^{\mathsf{T}}\right)^{-1} \left(\sum_{s=1}^{t} w_s^t X_s X_s^{\mathsf{T}} (\theta_s - \theta_{t+1}) \;+\; \sum_{s=1}^{t} w_s^t X_s \epsilon_s\right) \\ &=\; \left(\mathcal{J}_t^t\right)^{-1} \mathcal{W}_t^t \;+\; \left(\mathcal{J}_t^t\right)^{-1} \mathcal{M}_t^t \quad \text{ for all } t = 2, 3, \ldots, \end{aligned} \tag{3.4}$$

where $\mathcal{J}_s^t$ is the empirical Fisher information given by

$$\mathcal{J}_s^t \;=\; \sum_{q=1}^{s} w_q^t X_q X_q^{\mathsf{T}}, \tag{3.5}$$

$\mathcal{W}_s^t = \sum_{q=1}^{s} w_q^t X_q X_q^{\mathsf{T}} (\theta_q - \theta_{t+1})$, and $\mathcal{M}_s^t = \sum_{q=1}^{s} w_q^t X_q \epsilon_q$. The first term on the right hand side of (3.4) is the *estimation inaccuracy* due to the changing environment, and the second term is the *estimation error* due to noise.

## 3.3 Designing Rate Optimal Policies

We now construct policies whose regret grows at the rate given in Theorem 1. In this subsection, we first define a family of pricing policies that relies on (i) price experimentation with a carefully chosen frequency, and (ii) the use of weighted least squares estimation with a particular sequence of weights. Then we prove that this family of policies is rate optimal.

**Price experiments.** The policies we consider in this subsection conduct price tests with a certain frequency in the following manner: let $\kappa$ is a positive real number, and $x_1$, $x_2$ be two distinct test prices in $[\ell, u]$. To construct the set of periods, $\mathcal{X}_1$, $\mathcal{X}_2$, at which the test prices will be charged, let $n := \lceil \kappa B^{-1/3} T^{1/3} \rceil \geq 2$ and

$$\mathcal{X}_i \;:=\; \left\{ t = kn^2 + (i-1)n + q \;:\; k = 0, 1, \ldots, \lfloor T/n^2 \rfloor,\; q = 1, \ldots, n \right\} \tag{3.6}$$

for $i = 1, 2$. In period $t$, charge the price

$$p_t = \begin{cases} x_1 & \text{if } t \in \mathcal{X}_1 \\ x_2 & \text{if } t \in \mathcal{X}_2 \\ \varphi(\vartheta_t) & \text{otherwise,} \end{cases} \tag{3.7}$$

where $\vartheta_t$ is the truncated estimate that satisfies $\vartheta_t := \arg\min_{\vartheta \in \Theta}\{\|\vartheta - \hat{\theta}_t\|\}$. Note that, in the above experimentation scheme, the frequency of price tests is $2n/n^2 = 2/n$, which is of order $T^{-1/3}$.

**Moving windows.** Consider a policy that estimates the unknown demand vector using only the most recent price tests within a moving time window, forgetting all data outside said window. Given $B > 0$, the *moving window policy* with parameters $\kappa, x_1, x_2$, denoted by $M_B(\kappa, x_1, x_2)$, chooses prices according to (3.6) and (3.7), and uses a sequence of weight vectors $\{w^1, w^2, \ldots\}$ such that $w^t = (w_1^t, \ldots, w_t^t)$ for $t = 1, \ldots, T$, where

$$w_s^t = \begin{cases} 1 & \text{if } s \in \mathcal{X} \ \text{ and } \ s \geq t - n^2 \\ 0 & \text{otherwise,} \end{cases} \tag{3.8}$$

for $1 \leq s \leq t$, and $\mathcal{X} = \mathcal{X}_1 \cup \mathcal{X}_2$. Note that, for $M_B(\kappa, x_1, x_2)$, the empirical Fisher information in the estimation problem (3.1-3.2) has the following form:

$$\mathcal{J}_s^t \ = \ \mathfrak{X}\, \mathcal{I}_s^t \qquad \text{for } 1 \leq s \leq t\,, \tag{3.9}$$

where

$$\mathfrak{X} = \begin{bmatrix} 2 & x_1 + x_2 \\ x_1 + x_2 & x_1^2 + x_2^2 \end{bmatrix}, \tag{3.10}$$

and $\mathcal{I}_s^t = \frac{1}{2}\sum_{q=1}^{s} w_q^t \geq 0$ represents the "relevance" in period $t$ of the information in period $s$.

**Performance of the moving window policy.** We now show that the policy family defined above is rate optimal. In our first result, we derive upper bounds on the aggregate estimation inaccuracy due to changes in demand parameters.

**Lemma 1 (upper bound on aggregate estimation inaccuracy)** *There exists a finite positive constant $c_1$, such that under $M_B(\kappa, x_1, x_2)$*

$$\sum_{t=n}^{T-1} \big\| \big(\mathcal{J}_t^t\big)^{-1} \mathcal{W}_t^t \big\|^2 \ \leq \ c_1\, B^{1/3}\, T^{2/3} \tag{3.11}$$

*almost surely for all $T = 1, 2, \ldots$ and $\boldsymbol{\theta} \in \mathcal{V}(T, B)$.*

To derive Lemma 1, we first note that a change after periods $s$ contributes to the estimation inaccuracy in period $t$ if and only if $w_s^t > 0$, i.e., the demand observation in period $s$ has positive weight in period $t$. Therefore, in a changing environment, the seller needs to: (i) avoid giving excessive weight to past observations to limit the contribution of a parameter change to the estimation inaccuracy; and at the same time, (ii) give non-negligible weight to past observations to accumulate information. More formally, (i) can be viewed as a *forgetting* condition that guarantees that the norm of $\mathcal{W}_t^t$ grows sufficiently slowly. On the other hand, condition (ii), which can be interpreted as a *learning* condition, guarantees that the eigenvalues of $\mathcal{J}_t^t$ grow sufficiently fast. Moving windows resolve the tradeoff between these forgetting and learning conditions as follows: to obtain the bound in (3.11) for $M_B(\kappa, x_1, x_2)$, the size of the moving window should be small enough to meet the forgetting condition (i), but also large enough to meet the learning condition (ii). Lemma 1 states that the careful selection of window sizes in (3.8) leads to an $O(B^{1/3}\,T^{2/3})$ aggregate estimation inaccuracy, which grows at the minimum achievable rate described in Theorem 1.

Our second result characterizes how estimation errors due to noise decay over time.

**Lemma 2 (exponential decay of estimation error due to noise)** *Let $\pi$ be $M_B(\kappa, x_1, x_2)$. Then there exists a finite positive constant $\rho$ such that*

$$\mathbb{P}^{\pi}_{\boldsymbol{\theta}}\big\{\big\|\big(\mathcal{J}_t^t\big)^{-1}\mathcal{M}_t^t\big\| > z,\ \mathcal{I}_t^t > \gamma\big\} \ \leq\ 4\,e^{-\rho(z\wedge z^2)\gamma} \tag{3.12}$$

*for all $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_T)$, $z > 0$, $\gamma > 0$, and $t \geq 2$.*

Lemma 2 states that the tail probability of the estimation error $\big(\mathcal{J}_t^t\big)^{-1}\mathcal{M}_t^t$ decays exponentially, and the rate of this decay is determined by the amount of relevant information in period $t$, namely $\mathcal{I}_t^t = \frac{1}{2}\sum_{q=1}^{t} w_q^t \geq 0$. Using Lemmas 1 and 2 we obtain the following performance bound.

**Theorem 2 (rate optimality)** *Let $\pi$ be $M_B(\kappa, x_1, x_2)$. Then there exists a finite positive constant $C$ such that $\mathcal{R}^{\pi}(T, B) \leq C\,B^{1/3}\,T^{2/3}$ for all $T \geq 3$.*

**Discussion.** The preceding theorem establishes the rate optimality of the moving window policy constructed in this subsection. The main intuition behind this result is a careful balancing of three goals: (i) learning; (ii) earning; and (iii) forgetting. The experimentation scheme in (3.6-3.7), and the weights in (3.8) ensure that the information metric $\mathcal{I}_t^t$ is proportional to $B^{-1/3}T^{1/3}$, and as shown in Lemma 1, this leads to a rate optimal balance between learning and forgetting. In Theorem 2, we show that the same experimentation scheme and choice of weights also achieve a rate optimal balance between learning and earning: by maintaining the relevant amount of information in the order of $B^{-1/3}T^{1/3}$, the seller guarantees that the aggregate losses due to estimation inaccuracy and estimation error are $O(B^{1/3}T^{2/3})$. To keep the relevant amount of information at that level while forgetting the past at the rate given in Lemma 1, the seller conducts $O(B^{1/3}T^{2/3})$ price tests, implying that the cost of experimentation is also of order $B^{1/3}T^{2/3}$. These two relations between experimentation cost, estimation error, and estimation inaccuracy provide a fine balance between learning, earning, and forgetting, from which we derive the rate optimal performance bound in Theorem 2.

### 3.4 Adapting to Unknown Variation Budget: First-Order Optimal Policies

We now extend the analysis in the preceding subsection to the case where the seller does not know the variation budget $B$ at the outset. In this extension, we first modify the moving window policy family so that it no longer relies on the a priori knowledge of $B$, and then construct another policy family that gradually discounts the weights of older observations in the estimation procedure (3.1-3.2). We prove that these families of policies achieve $O(T^{2/3})$ regret in $T$ periods, and hence are first-order optimal.

**Price experiments.** To adapt to an unknown $B$ parameter, consider the following modification of the price experimentation scheme in Section 3.3. Given $\kappa > 0$, let $n := \lceil \kappa T^{1/3} \rceil$ and

$$\mathcal{X}_i \ := \ \left\{ t = kn + i \ : \ k = 0, 1, 2, \ldots, \lfloor T/n \rfloor \right\} \tag{3.13}$$

for $i = 1, 2$. The price to be charged in period $t$ is given by

$$p_t = \begin{cases} x_1 & \text{if } t \in \mathcal{X}_1 \\ x_2 & \text{if } t \in \mathcal{X}_2 \\ \varphi(\vartheta_t) & \text{otherwise,} \end{cases} \tag{3.14}$$

where $x_1$ and $x_2$ are two distinct test prices in $[\ell, u]$, and $\vartheta_t$ is the truncated estimate that satisfies $\vartheta_t := \arg\min_{\vartheta \in \Theta}\{\|\vartheta - \hat{\theta}_t\|\}$. This experimentation scheme, like its counterpart in the preceding subsection, conducts price tests with a frequency of $2/n$, which is of order $T^{-1/3}$.

**Moving windows and gradually decaying weights.** Our first policy in this subsection is obtained by modifying the definition of the moving window policy as follows: suppose that the *moving window policy* with parameters $\kappa, x_1, x_2$, denoted by $M(\kappa, x_1, x_2)$, charges the prices in (3.13-3.14), and uses the weight vectors $\{w^1, w^2, \ldots\}$ such that $w^t = (w_1^t, \ldots, w_t^t)$ for $t = 1, \ldots, T$, where

$$w_s^t = \begin{cases} 1 & \text{if } s \in \mathcal{X} \ \text{ and } \ s \geq t - n^2 \\ 0 & \text{otherwise,} \end{cases} \tag{3.15}$$

for $1 \leq s \leq t$, and $\mathcal{X} = \mathcal{X}_1 \cup \mathcal{X}_2$.

Our second policy puts decreasing weight on older observations in a gradually decaying manner. The *decaying weights policy* with parameters $\mu, \kappa, x_1, x_2$, denoted $W(\mu, \kappa, x_1, x_2)$, selects prices according to (3.13) and (3.14), and uses a sequence of weight vectors $\{w^1, w^2, \ldots\}$ such that $w^t = (w_1^t, \ldots, w_t^t)$ for $t = 1, \ldots, T$, where

$$w_s^t = \begin{cases} \left(1 - \dfrac{t-s}{n^2} + \dfrac{(t-s)^{1-\mu}}{n^2}\right)_+^{\frac{1}{\mu}} & \text{if } s \in \mathcal{X} \\ 0 & \text{otherwise,} \end{cases} \tag{3.16}$$

for $1 \leq s \leq t$, and $0 < \mu \leq 1$. Under $W(\mu, \kappa, x_1, x_2)$, the weight given to any observation decreases smoothly via the decay parameter $\mu$. An extreme choice for $\mu$ is 1, in which case the weights in

(3.16) become $w_s^t = \big(1-(t-s+1)/n^2\big)_+ \mathbb{I}\{s \in \mathcal{X}\}$ for $1 \le s \le t$, implying that weights decay linearly over time. As $\mu$ approaches zero, we achieve slower polynomial decay rates.

We note that, under both $M(\kappa, x_1, x_2)$ and $W(\mu, \kappa, x_1, x_2)$, the empirical Fisher information in the estimation problem (3.1-3.2) has the form given in (3.9). This allows us to generalize Lemmas 1 and 2 for $M(\kappa, x_1, x_2)$ and $W(\mu, \kappa, x_1, x_2)$, and consequently derive the following counterpart of Theorem 2.

**Theorem 3 (first-order optimality)** *Let $\pi$ be either $M(\kappa, x_1, x_2)$ or $W(\mu, \kappa, x_1, x_2)$. Then there exists a finite positive constant $C$ such that $\mathcal{R}^\pi(T, B) \le C\,T^{2/3}$ for all $T \ge 3$.*

Theorem 3 shows that the moving window and decaying weights policies described in this subsection achieve $O(T^{2/3})$ regret *without* relying on the a priori knowledge of $B$. The constant $C$ in Theorem 3 grows linearly in $B$, and the difference between this constant and the $O(B^{1/3})$ constant in Theorem 2 helps us quantify the "price" of adapting to an unknown variation budget.

# 4 Learning and Detection of Bursty Changes

In this section, we consider the case of bursty changes that are characterized by a positive minimum change constraint: suppose that there exists a positive constant $\delta$ satisfying

$$d_{\boldsymbol{\theta}} \;:=\; \inf\big\{\|\theta_t - \theta_s\| \,:\, \theta_t \neq \theta_s,\, 1 \le s < t \le T\big\} \;\ge\; \delta. \tag{4.1}$$

In contrast to gradual and potentially undetectable changes that can happen in the setting studied in the preceding section, condition (4.1) implies that changes happen in bursts; that is, whenever the demand vector changes, its Euclidean norm has to change by at least $\delta$. Combined with condition (2.5), this implies that there can be at most $\bar{C} = \lceil B/\delta^2 \rceil$ changes. An extreme example in the family of admissible changing environments described by conditions (2.5) and (4.1) is the case of a single change-point (over the entire time horizon). In comparison with traditional change-point detection problems, the distinguishing feature of our problem is the need to learn the demand parameters before and after the change, which makes it more difficult to detect the change-point. Another example is the case of switching back and forth between two distinct values of demand parameters. The repetitive nature of this example requires conducting multiple detection tests, which could lead to multiple false alarms before a non-spurious change-point is detected, unlike single change-point detection tests. In general, the demand sequence $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots)$ can take on $\bar{C}+1$ distinct values, all of which are initially unknown to the seller.

With the addition of the bursty change condition (4.1), we update our performance metric as follows: Let

$$\mathcal{R}^\pi(T, B, \delta) \;=\; \sup\big\{\Delta^\pi_{\boldsymbol{\theta}}(T) : \boldsymbol{\theta} \in \mathcal{V}(T, B, \delta)\big\}, \tag{4.2}$$

where $\Delta^\pi_{\boldsymbol{\theta}}(T)$ is as defined in (2.9), and $\mathcal{V}(T, B, \delta) = \{\boldsymbol{\theta} : V_{\boldsymbol{\theta}}(T) \le B, d_{\boldsymbol{\theta}} \ge \delta\}$. Note that $\mathcal{V}(T, B, \delta)$ is a subset of its counterpart in the preceding section, namely $\mathcal{V}(T, B)$, which is given in (2.6). One of the key questions we would like to investigate in this section is whether we can achieve significantly smaller regret by imposing the bursty change condition (4.1) on the set of admissible demand parameter sequences.

## 4.1 Dynamic Pricing with Simultaneous Learning and Detection

Assuming the seller knows that (4.1) holds, we design a well-performing pricing policy that detects change-points and learns unknown demand parameters simultaneously.

**Price experiments.** Let $\kappa$ and $\eta$ be two positive real numbers, and $x_1$, $x_2$ be two distinct test prices in $[\ell, u]$. The *detection policy* with parameters $\eta, \kappa, x_1, x_2$, denoted by $D(\eta, \kappa, x_1, x_2)$, divides the time horizon into cycles of $n := \lceil \kappa T^{1/2} \rceil$ periods, and conducts price experiments in the first $2m$ periods of every cycle, where $m := \lceil \kappa \log T \rceil$. To be precise, the sets of periods at which $D(\eta, \kappa, x_1, x_2)$ conducts price experiments are given by

$$\mathcal{X}_{ik} \;:=\; \big\{t = kn + (i-1)m + q \;:\; q = 1, 2, \ldots, m\big\} \tag{4.3}$$

for $i = 1, 2$, and $k = 0, 1, 2, \ldots, \lfloor T/n \rfloor$. In period $t$, $D(\eta, \kappa, x_1, x_2)$ charges the price

$$p_t = \begin{cases} x_1 & \text{if } t \in \mathcal{X}_1 \\ x_2 & \text{if } t \in \mathcal{X}_2 \\ \varphi(\vartheta_t) & \text{otherwise,} \end{cases} \tag{4.4}$$

where $\mathcal{X}_i = \bigcup_k \mathcal{X}_{ik}$ for $i = 1, 2$, and $\vartheta_t$ is the truncated estimate that satisfies $\vartheta_t := \arg\min_{\vartheta \in \Theta}\{\|\vartheta - \hat{\theta}_t\|\}$. In this experimentation scheme, the frequency of price tests is $2m/n$, which is of order $T^{-1/2} \log T$.

**Joint change-point detection and parameter estimation.** We will now describe a detection scheme that dynamically updates the weight vector sequence $\{w^t\}$ of the weighted least squares estimator in (3.2) by placing zero weight on periods that precede a detected change-point. Fix $k \in \{0, \ldots, \lfloor T/n \rfloor\}$. Denote by $\bar{D}_{ik}$ the average demand observed during the periods in $\mathcal{X}_{ik}$, that is

$$\bar{D}_{ik} \;:=\; m^{-1} \textstyle\sum_{t \in \mathcal{X}_{ik}} D_t. \tag{4.5}$$

Construct the binary-valued detection process $\chi := \{\chi_0, \chi_1, \ldots\}$ as follows: fix $\chi_0 = 1$, and define the *latest detection cycle* as $L(k) := \max\{\tau \le k : \chi_\tau = 1\}$. With this formalism a price experiment in period $s$ occurs after the latest detection (prior to period $t$) if and only if $s > nL(t/n)$. For every cycle $k = 0, 1, \ldots, \lfloor T/n \rfloor$, let

$$\chi_{k+1} = \begin{cases} 1 & \text{if } \sup_{i,k'} \big\{|\bar{D}_{ik} - \bar{D}_{ik'}| : i = 1, 2,\; L(k) \le k' < k\big\} > \eta \\ 0 & \text{otherwise,} \end{cases} \tag{4.6}$$

where the supremum of the empty set is taken to be $-\infty$. The detection test in (4.6) repeatedly checks whether there has been a change in the average demand observed since the latest detection. To make a comparison, it is necessary to compute at least one average demand estimate after each detection, and hence, it is not possible to have two consecutive detections: for any given cycle $k$ with $\chi_k = 1$, we have $L(k) = k$ and there is no $k'$ satisfying $L(k) \leq k' < k$, implying that $\chi_{k+1} = 0$.

In cycle $k = 0, 1, 2, \ldots, \lfloor T/n \rfloor$, the seller observes $\{\bar{D}_{ik'}\}_{i=1,2,\, k'=0,1,\ldots,k}$ by the end of period $(k+1)n$, which implies that $\chi_{k+1}$ is a function that maps demand in the first $(k+1)n$ periods, $D_1, D_2, \ldots, D_{(k+1)n}$, into $\{0, 1\}$. Based on the realization of the detection process $\chi$, $D(\eta, \kappa, x_1, x_2)$ uses the following weights for estimation:

$$w_s^t = \begin{cases} 1 & \text{if } s \in \mathcal{X} \text{ and } s > nL(t/n) \\ 0 & \text{otherwise,} \end{cases} \tag{4.7}$$

for $1 \leq s \leq t$, where $\mathcal{X} = \mathcal{X}_1 \cup \mathcal{X}_2$. In other words, $D(\eta, \kappa, x_1, x_2)$ recalls all available data as long as it does not detect a change, and forgets all past data immediately after it detects a change (hence it restarts learning whenever the value of the process $\chi$ switches from 0 to 1).

## 4.2 Performance of the Detection Policy

In this section, we prove that the regret of the detection policy described above is of order $T^{1/2} \log T$. To derive this result, we first define the random times at which detections happen. Using these random times, we decompose the regret according to different sources of loss, and then bound each of them.

Suppose that the unknown parameter sequence $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots)$ has $\mathcal{C}$ change-points in the first $T$ periods, and denote by $t_j^*$ the $j^{\text{th}}$ change-point. That is, let $1 = t_0^* < t_1^* < \ldots < t_{\mathcal{C}}^* < t_{\mathcal{C}+1}^* = T + 1$, where $t_j^* = \inf\{t \geq t_{j-1}^* : \theta_t \neq \theta_{t_{j-1}^*}\}$ for $j = 1, 2, \ldots, \mathcal{C}$. Recalling that $D(\eta, \kappa, x_1, x_2)$ divides the time horizon into cycles of $n = \lceil \kappa T^{1/2} \rceil$ periods, we let $\tau_j^* := \lfloor (t_j^* - 1)/n \rfloor$ be the cycle of the $j^{\text{th}}$ change-point, $\hat{\tau}_j^+$ be the cycle containing a correct detection subsequent to the $j^{\text{th}}$ change-point, and $\hat{\tau}_j^-$ be the cycle of first false detection after the $j^{\text{th}}$ change-point (if there is no detection between the $j^{\text{th}}$ and $(j+1)^{\text{st}}$ change-points then we set $\hat{\tau}_j^+ = \hat{\tau}_j^- = \tau_{j+1}^*$). More formally, define $\hat{\tau}_0^+ := 0$, and put

$$\begin{aligned} \hat{\tau}_j^+ &:= \inf\{\tau > \tau_j^* : \chi_\tau = 1\} \wedge \tau_{j+1}^* \quad \text{for } j = 1, 2, \ldots, \mathcal{C}, \\ \hat{\tau}_j^- &:= \inf\{\tau > \hat{\tau}_j^+ : \chi_\tau = 1\} \wedge \tau_{j+1}^* \quad \text{for } j = 0, 1, \ldots, \mathcal{C}, \end{aligned} \tag{4.8}$$

where the infimum of the empty set is taken to be $\infty$.

In the context of joint change-point detection and parameter estimation, the loss of a policy stems from four sources: (i) delay in correct detections; (ii) false alarms; (iii) estimation errors due to noise; and (iv) cost of experimentation. Let us first decompose the $T$-period regret with respect to losses due to (i-iii) and (iv). Because the cardinality of the experimentation set $\mathcal{X} = \mathcal{X}_1 \cup \mathcal{X}_2$ is at most $2m\lceil T/n \rceil \le 8T^{1/2}\log T$, we have

$$
\begin{aligned}
\Delta^{\pi}_{\boldsymbol{\theta}}(T) &= \mathbb{E}^{\pi}_{\boldsymbol{\theta}}\Big\{ \sum_{t=1}^{T} \Big(1 - \frac{r(p_t,\theta_t)}{r^*(\theta_t)}\Big)\Big\} \\
&= \mathbb{E}^{\pi}_{\boldsymbol{\theta}}\Big\{ \sum_{t=1}^{T} \Big(1 - \frac{r(p_t,\theta_t)}{r^*(\theta_t)}\Big)\mathbb{I}\{t \in \mathcal{X}\}\Big\} + \mathbb{E}^{\pi}_{\boldsymbol{\theta}}\Big\{ \sum_{t=1}^{T} \Big(1 - \frac{r(p_t,\theta_t)}{r^*(\theta_t)}\Big)\mathbb{I}\{t \notin \mathcal{X}\}\Big\} \\
&\le 2m\Big\lceil \frac{T}{n} \Big\rceil + \mathbb{E}^{\pi}_{\boldsymbol{\theta}}\Big\{ \sum_{t=1}^{T} \Big(1 - \frac{r(p_t,\theta_t)}{r^*(\theta_t)}\Big)\mathbb{I}\{t \notin \mathcal{X}\}\Big\} \\
&\le 8T^{1/2}\log T + \mathbb{E}^{\pi}_{\boldsymbol{\theta}}\Big\{ \sum_{t=1}^{T} \Big(1 - \frac{r(p_t,\theta_t)}{r^*(\theta_t)}\Big)\mathbb{I}\{t \notin \mathcal{X}\}\Big\}. \qquad (4.9)
\end{aligned}
$$

The first term on the right hand side above is the loss due to price experimentation, (iv), whereas the second term is the sum of losses due to the other three sources (i-iii). Now, let us decompose the second term to see the tradeoff between (i), (ii) and (iii):

$$
\begin{aligned}
\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\Big\{ \sum_{t=1}^{T} \Big(1 - \frac{r(p_t,\theta_t)}{r^*(\theta_t)}\Big)\mathbb{I}\{t \notin \mathcal{X}\}\Big\} &= \mathbb{E}^{\pi}_{\boldsymbol{\theta}}\Big\{ \sum_{j=0}^{\mathcal{C}} \sum_{s=n\tau_j^*+1}^{n\tau_{j+1}^*} \Big(1 - \frac{r(p_s,\theta_s)}{r^*(\theta_s)}\Big)\mathbb{I}\{s \notin \mathcal{X}\}\Big\} \\
&= \sum_{j=0}^{\mathcal{C}} \mathbb{E}^{\pi}_{\boldsymbol{\theta}}\Big\{ \sum_{s=n\tau_j^*+1}^{n\hat{\tau}_j^+} \Big(1 - \frac{r(p_s,\theta_s)}{r^*(\theta_s)}\Big)\mathbb{I}\{s \notin \mathcal{X}\} \\
&\qquad + \sum_{s=n\hat{\tau}_j^+ +1}^{n\hat{\tau}_j^-} \Big(1 - \frac{r(p_s,\theta_s)}{r^*(\theta_s)}\Big)\mathbb{I}\{s \notin \mathcal{X}\} \\
&\qquad + \sum_{s=n\hat{\tau}_j^- +1}^{n\tau_{j+1}^*} \Big(1 - \frac{r(p_s,\theta_s)}{r^*(\theta_s)}\Big)\mathbb{I}\{s \notin \mathcal{X}\}\Big\}. \qquad (4.10)
\end{aligned}
$$

The first, second, and third sums inside the expectation on the right hand side above are the losses due to delay of true detections, noise in estimation, and early false alarms, respectively. Our next task is to find upper bounds on these sums. In the analysis of the losses associated with delayed correct detections (or early false alarms), the following lemma is key.

**Lemma 3 (polynomial decay of detection error)** *Let $\pi$ be $D(\eta,\kappa,x_1,x_2)$ where $\kappa = c_\epsilon/(\eta\wedge\eta^2)$ and $c_\epsilon$ is a finite positive constant characterized by the distribution of $\{\epsilon_t\}$. For all $i$ and $k$, let $\bar{\epsilon}_{ik} := m^{-1}\sum_{t\in\mathcal{X}_{ik}}\epsilon_t$, with $m = \lceil \kappa \log T \rceil$. Then,*

$$\mathbb{P}^{\pi}_{\boldsymbol{\theta}}\big\{|\bar{\epsilon}_{ik}| \ge \tfrac{1}{2}\eta\big\} \le 2T^{-3/2}, \qquad (4.11)$$

*for all $T \ge 3$, $\boldsymbol{\theta} \in \Theta^T$, $i = 1,2$, and $k = 0,1,2,\ldots,\lfloor T/n \rfloor$.*

**Remark** The constant $c_\epsilon$, which appears in the above lemma as well as Lemmas 4, 5, and Theorem 4 below, is independent of $T$, $B$, $\delta$, and completely characterized by the exponential moment condition, $\mathbb{E}[\exp(x\epsilon_t)] < \infty$ for all $|x| \le x_0$. For example, in the case $\epsilon_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$, we have $c_\epsilon = 12\sigma^2$. A general expression of $c_\epsilon$ is provided in the first paragraph in the proof of Lemma 3.

Using Lemma 3, we first obtain the following upper bound on the loss due to detection delay.

**Lemma 4 (loss due to delay of true detections)** *Let $\pi$ be $D(\eta, \kappa, x_1, x_2)$ with $\eta = \frac{|x_1 - x_2|}{8(1 \vee x_1 \vee x_2)}\delta$ and $\kappa = c_\epsilon/(\eta \wedge \eta^2)$, where $c_\epsilon$ is the constant given in Lemma 3. Then there exists a finite positive constant $C_1$ such that*

$$\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\left\{ \sum_{s=n\tau_j^*+1}^{n\hat{\tau}_j^+} \left(1 - \frac{r(p_s, \theta_s)}{r^*(\theta_s)}\right) \mathbb{I}\{s \notin \mathcal{X}\} \right\} \le C_1\sqrt{T} \tag{4.12}$$

*for all $T \ge 3$ and $\boldsymbol{\theta} \in \mathcal{V}(T, B, \delta)$.*

In contrast to results in the single change-point detection literature, where the only uncertainty is about the time of change, the preceding lemma is proven without prior knowledge of the environment pre- and post-change. In this lemma, to estimate the expected loss due to detection delay we first analyze the detection test in (4.6), which repeatedly compares the average demand estimates in the current cycle $k$ with the ones in cycles $L(k), \ldots, k-1$, where $L(k) = \max\{\tau \le k : \chi_\tau = 1\}$ denotes the latest detection cycle before $k$. As long as the demand parameter vector in cycle $\tau_j^* + 1$ is significantly different than one of the demand parameter vectors in cycles $L(\tau_j^*), \ldots, \tau_j^*$, there is a high probability of detecting the $j^{\text{th}}$ change-point. On the other hand, if almost all of the demand parameter vectors in cycles $L(\tau_j^*), \ldots, \tau_j^*$ are the same as the demand parameter vector in cycle $\tau_j^* + 1$, this means that the unknown demand parameter sequence must have switched back to a value that was prevalent in the cycles that occurred after $L(\tau_j^*)$. In that case, it is unlikely that the detection test (4.6) will identify the $j^{\text{th}}$ change-point, but this would not lead to substantial loss because almost all of the information accumulated since cycle $L(\tau_j^*)$ will be relevant in cycle $\tau_j^* + 1$. We formalize this argument in the proof of Lemma 4, and show that the loss due to detection delay is of order $\sqrt{T}$ under our detection policy.

Our next result builds on Lemma 3 to show that the loss due to false alarms is of order $\sqrt{T}$.

**Lemma 5 (loss due to false alarms)** *Let $\pi$ be $D(\eta, \kappa, x_1, x_2)$ with $\eta = \frac{|x_1 - x_2|}{8(1 \vee x_1 \vee x_2)}\delta$ and $\kappa = c_\epsilon/(\eta \wedge \eta^2)$, where $c_\epsilon$ is the constant given in Lemma 3. Then there exists a finite positive constant $C_2$ such that*

$$\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\left\{ \sum_{s=n\hat{\tau}_j^-+1}^{n\tau_{j+1}^*} \left(1 - \frac{r(p_s, \theta_s)}{r^*(\theta_s)}\right) \mathbb{I}\{s \notin \mathcal{X}\} \right\} \le C_2\sqrt{T} \tag{4.13}$$

*for all $T \ge 3$ and $\boldsymbol{\theta} \in \mathcal{V}(T, B, \delta)$.*

It is worth noting that the setting studied in this section might include multiple false alarms because there is potentially more than one bursty change, and accordingly, the detection test (4.6) is repeated throughout the time horizon. Lemma 5 provides an upper bound on the revenue loss because of all such false alarms between the $j^{\text{th}}$ and $(j+1)^{\text{st}}$ change-points.

Having found $O(\sqrt{T})$ upper bounds on the losses due to false detections, we prove in the following lemma; the loss due to estimation noise is also $O(\sqrt{T})$.

**Lemma 6 (loss due to estimation noise)** *Let $\pi$ be $D(\eta, \kappa, x_1, x_2)$. Then there exists a finite positive constant $C_3$ such that*

$$\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\left\{ \sum_{s=n\hat{\tau}_j^{+}+1}^{n\hat{\tau}_j^{-}} \left(1 - \frac{r(p_s, \theta_s)}{r^*(\theta_s)}\right)\mathbb{I}\{s \notin \mathcal{X}\}\right\} \;\leq\; C_3\sqrt{T} \tag{4.14}$$

*for all $T \geq 3$ and $\boldsymbol{\theta} \in \Theta^T$.*

The preceding lemma provides an upper bound on the revenue loss incurred between the true detection after the $j^{\text{th}}$ change-point and the first false detection before the $(j+1)^{\text{st}}$ change-point. During this time interval, there are no changes in demand parameters and no detections, meaning that there is no estimation inaccuracy due to changes, and the revenue loss is entirely caused by estimation error due to noisy demand observations. By a straightforward modification of Lemma 2, the price experimentation scheme in (4.3-4.4) implies that the loss due to estimation noise in this case is at most of order $\sqrt{T}$.

In the final result of this section, we combine Lemmas 4, 5, and 6 with inequality (4.10) to obtain the following performance bound.

**Theorem 4 (near-optimality of the pricing-detection policy)** *Let $\pi$ be $D(\eta, \kappa, x_1, x_2)$ with $\eta = \frac{|x_1 - x_2|}{8(1 \vee x_1 \vee x_2)}\,\delta$ and $\kappa = c_\epsilon/(\eta \wedge \eta^2)$, where $c_\epsilon$ is the constant given in Lemma 3. Then there exists a finite positive constant $C$ such that $\mathcal{R}^{\pi}(T, B, \delta) \leq C\,T^{1/2}\log T$ for all $T \geq 3$.*

**Remark** We note that our detection policy, $D(\eta, \kappa, x_1, x_2)$, uses the knowledge of $\delta$ in the choice of parameters $\eta$ and $\kappa$, but does not require the knowledge of $B$.

According to Theorem 4, the $T$-period regret of our detection policy is in the order of $T^{1/2}\log T$. To put this result in perspective, we refer readers to two existing lower bounds: Keskin and Zeevi (2012) derive a lower bound of order $T^{1/2}$ in a learning-and-earning problem in a static demand environment. Besbes and Zeevi (2011) obtain another lower bound of order $T^{1/2}$ in a single change-point detection problem in which demand curves before and after the change-point are known. In light of these results, the policy in Theorem 4 is near-optimal in order (up to logarithmic terms in $T$).

Keskin and Zeevi (2012) and Besbes and Zeevi (2011) also provide policies that have $O(T^{1/2}\log T)$ regret in their settings, but because they focus on either learning or detection in isolation, they do not address the challenges arising in simultaneous learning and detection, such as the occurrence of a change-point prior to forming an informative (i.e., not-too-noisy) estimate of the average demand reference point for the detection test. To address such challenges, we employ a repeated detection test with a carefully chosen frequency to obtain the key results in Lemmas 3 and 4. Interestingly, while the seller in our setting is facing a more difficult problem compared to the ones studied by Keskin and Zeevi (2012) and Besbes and Zeevi (2011), the policy we design achieves, nonetheless, the near-optimal revenue performance; to that end, in the aforementioned papers policies were designed to deal exclusively with either learning or detection.

# 5 Rapidly Changing Demand Environments

We now generalize the original problem formulation in Section 2 to more rapidly changing environments, where the change budget given in condition (2.5) is increasing in $T$. To be precise, take $\nu \in [0,1]$, and assume that $\theta_t$ are chosen from the rectangle $\Theta \subseteq \mathbb{R} \times \mathbb{R}_-$ such that

$$V_{\boldsymbol{\theta}}(T) \;\leq\; B\,T^{\nu} \quad \text{for } T = 1, 2, \ldots \tag{5.1}$$

where $B > 0$, and $V_{\boldsymbol{\theta}}(T)$ is the quadratic variation of $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots)$ in $T$ periods, defined in (2.4). In condition (5.1), the parameter $\nu$ represents the *volatility* of the changing demand environment: if $\nu = 0$ then we have the original problem formulation studied in preceding sections, whereas if $\nu = 1$ then the demand environment is extremely volatile in the sense that there can be a substantial change in every single period. For intermediate values of $\nu \in (0,1)$, we obtain a spectrum of demand environments where the scale of change is characterized by the volatility parameter $\nu$.

We incorporate the change budget in (5.1) into our performance metric as follows: Let

$$\mathcal{R}^{\pi}(T,\nu) \;=\; \sup\left\{\Delta^{\pi}_{\boldsymbol{\theta}}(T) : \boldsymbol{\theta} \in \mathcal{V}(T,\nu)\right\}, \tag{5.2}$$

where $\Delta^{\pi}_{\boldsymbol{\theta}}(T)$ is as defined in (2.9), and $\mathcal{V}(T,\nu) = \{\boldsymbol{\theta} : V_{\boldsymbol{\theta}}(T) \leq B\,T^{\nu}\}$. Here we note that $\mathcal{V}(T,\nu)$ is a superset of $\mathcal{V}(T,B)$, namely the set of admissible demand parameter sequences in the original problem. The main question we address in this section is how much the regret would increase when we expand the set of admissible demand parameter sequences. We have the following lower bound on regret under condition (5.1).

**Theorem 5 (lower bound on regret)** *There exists a finite positive constant $c$ such that $\mathcal{R}^{\pi}(T,\nu) \geq c\,T^{(2+\nu)/3}$ for any pricing policy $\pi$ and time horizon $T \geq 3$.*

Note that when $\nu = 1$ the revenue losses must grow linearly with the horizon, namely, the regret is no longer sublinear, and *no policy* is long-run-average optimal. To achieve the growth rate of regret in Theorem 5, we modify the moving window and decaying weights policies in Section 3.4 as follows.

**First-order optimal policies in rapidly changing environments.** As before, we consider policies that conduct price tests with a certain frequency, but due to increased volatility of the demand vector sequence $\boldsymbol{\theta}$, the frequency of price tests needs to be higher. That is, we let $n := \lceil \kappa T^{(1-\nu)/3} \rceil$ where $\kappa > 0$, and construct the set of periods at which the test prices will be charged as $\mathcal{X}_i := \{t = kn + i \; : \; k = 0, 1, 2, \ldots, \lfloor T/n \rfloor\}$ for $i = 1, 2$. Choosing two distinct test prices $x_1$ and $x_2$ in $[\ell, u]$, we let the price in period $t$ be

$$p_t = \begin{cases} x_1 & \text{if } t \in \mathcal{X}_1 \\ x_2 & \text{if } t \in \mathcal{X}_2 \\ \varphi(\vartheta_t) & \text{otherwise,} \end{cases} \tag{5.3}$$

where $\vartheta_t$ is the truncated estimate of $\theta_t$, which satisfies $\vartheta_t := \arg\min_{\vartheta \in \Theta}\{\|\vartheta - \hat{\theta}_t\|\}$. The frequency of price tests in the experimentation scheme (5.3) is $2/n$, namely of order $T^{-(1-\nu)/3}$.

In rapidly changing environments, the moving window policy needs to have a smaller window size, whereas the decaying weights policy needs to have a more significant rate of decay. In the original problem, the window size of $M(\kappa, x_1, x_2)$ was of order $T^{2/3}$. In this section, we choose a window size of order $T^{(2-2\nu)/3}$: under condition (5.1), the *rapidly moving window policy* with parameters $\kappa, x_1, x_2$, denoted by $M_\nu(\kappa, x_1, x_2)$, chooses prices according to (5.3), and uses the weights $w_s^t = \mathbb{I}\{s \in \mathcal{X},\, s \geq t - n^2\}$ for $1 \leq s \leq t$, where $\mathcal{X} = \mathcal{X}_1 \cup \mathcal{X}_2$ and $n = \lceil \kappa T^{(1-\nu)/3} \rceil$. Similarly, the *rapidly decaying weights policy* with parameters $\mu, \kappa, x_1, x_2$, denoted by $W_\nu(\mu, \kappa, x_1, x_2)$, chooses prices according to (5.3), and uses the weights $w_s^t = \left(1 - \frac{t-s}{n^2} + \frac{(t-s)^{1-\mu}}{n_\mu^2}\right)_+^{1/\mu} \mathbb{I}\{s \in \mathcal{X}\}$ for $1 \leq s \leq t$, where $0 < \mu \leq 1$, $n = \lceil \kappa T^{(1-\nu)/3} \rceil$, and $n_\mu = n T^{\mu\nu}$.

In our next result, we extend Lemma 1 to the case of rapidly changing demand environments.

**Lemma 7 (upper bound on aggregate estimation inaccuracy)** *There exists a finite positive constant $c_1$, such that under either $M_\nu(\kappa, x_1, x_2)$ or $W_\nu(\mu, \kappa, x_1, x_2)$*

$$\sum_{t=n^2+1}^{T-1} \left\| \left(\mathcal{J}_t^t\right)^{-1} \mathcal{W}_t^t \right\|^2 \leq c_1 T^{(2+\nu)/3} \tag{5.4}$$

*for all $T = 1, 2, \ldots$ and $\boldsymbol{\theta} \in \mathcal{V}(T, \nu)$.*

By Lemma 7 and a straightforward modification of Lemma 2, we generalize Theorem 3, and derive the following upper bound on the regret for $M_\nu(\kappa, x_1, x_2)$ and $W_\nu(\mu, \kappa, x_1, x_2)$.

**Theorem 6 (first-order optimality)** *Let $\pi$ be either $M_\nu(\kappa, x_1, x_2)$ or $W_\nu(\mu, \kappa, x_1, x_2)$. Then there exist a finite positive constant $C$ such that $\mathcal{R}^\pi(T, \nu) \leq C\, T^{(2+\nu)/3}$ for all $T \geq 3$.*

The preceding theorem provides a range of results for different degrees of change scales, quantifying how the volatility parameter $\nu$ influences the growth rate of regret. As $\nu$ increases, nature can cause larger estimation inaccuracy, and in response the seller needs to forget the past faster, either by choosing a smaller moving window size, or by faster weight decay. Roughly speaking, every quanta of $O(T^{\nu})$ in the change budget translates to an $O(T^{\nu/3})$ of regret.

# 6 Concluding Remarks

**Measuring information depreciation.** To compute the near-optimal forgetting rates in the settings analyzed in this paper, let us compare the sizes of the moving windows we constructed in these settings. In a static environment, we can use all past data within the entire time horizon; hence the size of the "moving" window is $O(T)$. In changing environments, we use moving windows of smaller order, such as the $O(T^{2/3})$ moving windows in Section 3. Given a particular demand environment, let $\boldsymbol{\delta}(T)$ be the ratio of the the near-optimal moving window size in the that environment to the nominal time horizon $T$. In static settings $\boldsymbol{\delta}(T)$ is of order 1 by definition, and in the time-varying settings of Sections 3 and 5, $\boldsymbol{\delta}(T)$ of our policies are of order $T^{-1/3}$ and $T^{-(1+2\nu)/3}$, respectively. In the case of bursty changes, the information is depreciated only when a change-point is detected, which implies that $\boldsymbol{\delta}(T)$ of our pricing-detection policy would be of order 1 unless there is an extremely late change-point.

**Structure of well-performing policies.** Our study presents three families of dynamic pricing policies designed to perform well in changing demand environments. The moving window and decaying weights policies in Section 3 are based on a weighted least squares estimator that discounts older observations at a certain rate. The detection policy in Section 4 uses the same weighted least squares estimator, but can reduce the weight of all past observations to zero upon detecting a change. All of these policies have near-optimal performance in their respective settings, but at the same time they use quite distinct rules for weighing past observations, which suggests that successful pricing policies in presence of smooth and bursty changes can have very different structures.

**Calibrating the volatility parameter.** To design successful dynamic pricing policies in rapidly changing environments, a seller needs to characterize the volatility in the demand environment, which is represented by parameter $\nu$ in the problem formulation in Section 5. The demand volatility can be characterized by first observing the demand response to a given incumbent price $\hat{p}$ over $N$ periods, and then measuring the average variation in expected demand as $v_N = \frac{1}{N}\sum_{t=1}^{N}(D_t - D_{t-1})^2 - 2\sigma^2$. In a demand environment with volatility parameter $\nu$, the average variation $v_N$ would be of order $N^{\nu}/N = N^{\nu-1}$. In light of this knowledge, the seller can run an ordinary least squares regression between $\log v_N$ and $\log N$, and calibrate the volatility parameter $\nu$.

**Linear demand assumption and asymptotically optimal semi-myopic policies.** Linear regression models are commonly used in econometrics to express reduced-form relationships between variables. In this paper, we model the relationship between price and demand in a similar linear fashion. In practice this relationship can be described by more general functional forms, in which expected demand is a smooth decreasing function of price. In such cases, the linear demand assumption leads to model misspecification, but as shown by Besbes and Zeevi (2013), such model misspecification can be mitigated by designing asymptotically optimal semi-myopic policies under the linear demand assumption. In essence, their argument is that as long as the expected demand function has a well-defined derivative around the optimal price, one can use a linear approximation to the expected demand function within a small neighborhood of the optimal price without incurring substantial loss. The essential principle behind asymptotically optimal semi-myopic policies, which are also employed in this paper as well as in Keskin and Zeevi (2012) and Besbes and Zeevi (2013), is using price experimentation to ensure that the myopic price is within a small neighborhood of the optimal price with high probability, and thereby limiting losses due to model misspecification.

**Data storage constraints.** Moving window and decaying weights policies designed in Section 3 differ sharply in terms of how they store price and sales data. While the decaying weights policy makes use of all historical observations, the moving window requires only a relatively small number of observations. Hence, data storage considerations favor moving window policy due to its more flexible data requirements.

## Appendix A: Proof of Theorem 1

Divide the time horizon into cycles of $N = \lceil k_0 T^{2/3} \rceil$ periods, where $k_0 = 4^{2/3} B^{-2/3}$, and consider the setting in which (i) $\epsilon_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$, (ii) the value of $\theta_t$ can change only in the first period of a cycle, and (iii) $\theta_t$ takes values in the set $\{y_0, y_1\}$, where $y_0 = (a_0, b_0) = (2, -1)$ and $y_1 = (a_1, b_1) = (2 + N^{-1/4}, -1 - N^{-1/4})$. Note that (ii), (iii), and the above choice of $N$ imply that

$$V_{\boldsymbol{\theta}}(T) \;\le\; \left(\frac{T}{N} + 1\right) \|y_0 - y_1\|^2 \;\overset{\text{(a)}}{\le}\; \frac{4T}{N^{3/2}} \;\le\; B, \tag{A.1}$$

where: (a) follows because $N \le T$ and $\|y_0 - y_1\|^2 = 2N^{-1/2}$. Therefore the setting described above satisfies the quadratic variation bound in (2.5).

Now, focus on a single cycle, which is composed of $N$ periods. Let $\mathbf{P}_i^\pi$ be a probability measure satisfying

$$\mathbf{P}_i^\pi(D_1 \le \xi_1, \ldots, D_N \le \xi_N) \;=\; \prod_{t=1}^{N} \Phi\left(\frac{\xi_t - a_i - b_i p_t}{\sigma}\right) \qquad \text{for } \xi_1, \ldots, \xi_T \in \mathbb{R}, \tag{A.2}$$

where $\Phi(\cdot)$ denotes the standard Gaussian cumulative distribution function, and $p = (p_1, p_2, \ldots)$ is the price sequence formed under policy $\pi$ and demand realization $D = (D_1, D_2, \ldots)$. Then, the

Kullback-Leibler divergence from $\mathbf{P}_0^\pi$ to $\mathbf{P}_1^\pi$ is

$$\mathcal{K}(\mathbf{P}_0^\pi, \mathbf{P}_1^\pi) \;:=\; \mathbf{E}_0^\pi \log\left(\frac{\prod_{t=1}^N \phi\left(\dfrac{D_t - a_0 - b_0 p_t}{\sigma}\right)}{\prod_{t=1}^N \phi\left(\dfrac{D_t - a_1 - b_1 p_t}{\sigma}\right)}\right), \tag{A.3}$$

where $\mathbf{E}_0^\pi$ is the expectation operator associated with $\mathbf{P}_0^\pi$, and $\phi(\cdot)$ denotes the standard Gaussian density. By elementary algebra, we can re-express (A.3) as follows:

$$\begin{aligned}\mathcal{K}(\mathbf{P}_0^\pi, \mathbf{P}_1^\pi) &= -\frac{1}{2\sigma^2}\,\mathbf{E}_0^\pi\left\{\sum_{t=1}^N \left[(D_t - a_0 - b_0 p_t)^2 - (D_t - a_1 - b_1 p_t)^2\right]\right\} \\ &= -\frac{1}{2\sigma^2}\,\mathbf{E}_0^\pi\left\{\sum_{t=1}^N \left[\epsilon_t^2 - \big(\epsilon_t + a_0 - a_1 + (b_0 - b_1)p_t\big)^2\right]\right\},\end{aligned} \tag{A.4}$$

because $D_t = a_0 + b_0 p_t + \epsilon_t$ under $\mathbf{P}_0^\pi$. Let $\delta = y_0 - y_1$ and $X_t = \left[\begin{smallmatrix}1\\ p_t\end{smallmatrix}\right]$. Then, the preceding identity becomes

$$\begin{aligned}\mathcal{K}(\mathbf{P}_0^\pi, \mathbf{P}_1^\pi) &= -\frac{1}{2\sigma^2}\,\mathbf{E}_0^\pi\left\{\sum_{t=1}^N \left[\epsilon_t^2 - (\epsilon_t - \delta\cdot X_t)^2\right]\right\} \\ &= -\frac{1}{2\sigma^2}\,\mathbf{E}_0^\pi\left\{\sum_{t=1}^N (2\epsilon_t - \delta\cdot X_t)\;\delta\cdot X_t\right\} \\ &\overset{(b)}{=} \frac{1}{2\sigma^2}\,\mathbf{E}_0^\pi\left\{\sum_{t=1}^N (\delta\cdot X_t)^2\right\} \\ &\overset{(c)}{=} \frac{1}{2\sigma^2}\,\mathbf{E}_0^\pi\left\{\sum_{t=1}^N N^{-1/2}\big(p_t - 1\big)^2\right\} \\ &\overset{(d)}{=} \frac{1}{2\sigma^2 N^{1/2}}\,\mathbf{E}_0^\pi\left\{\sum_{t=1}^N \big(p_t - \varphi(y_0)\big)^2\right\} \\ &\overset{(e)}{=} \frac{1}{2\sigma^2 N^{1/2}}\,\boldsymbol{\Delta}_0^\pi(N),\end{aligned} \tag{A.5}$$

where: $\boldsymbol{\Delta}_i^\pi(N)$ denotes the $N$-period regret given that policy $\pi$ is exercised and $\theta_t = y_i$ for all $t = 1, \ldots, N$ and $i = 0, 1$, (b) follows because the $\epsilon_t$ are independent and have zero mean, (c) follows because $\delta = (-N^{-1/4}, N^{1/4})$, (d) follows because $\varphi(y_0) = 1$, and (e) follows by the definition of regret in (5.2) and the fact that $b_0/r^*(y_0) = -1$.

We will consider two cases for the value of $\mathcal{K}(\mathbf{P}_0^\pi, \mathbf{P}_1^\pi)$. Let $\eta > 0$.

Case 1. $\mathcal{K}(\mathbf{P}_0^\pi, \mathbf{P}_1^\pi) > \eta$. By (A.5), we deduce that

$$\boldsymbol{\Delta}_0^\pi(N) \geq 2\sigma^2 \eta N^{1/2}. \tag{A.6}$$

Case 2. $\mathcal{K}(\mathbf{P}_0^\pi, \mathbf{P}_1^\pi) \le \eta$. Define $I_i := \left[\varphi(y_i) - \frac{1}{4}N^{-1/4}, \varphi(y_i) + \frac{1}{4}N^{-1/4}\right]$ for $i = 0, 1$, and let $\chi_t$ be a random variable such that

$$\chi_t = \begin{cases} 1 & \text{if } p_t \in I_0 \\ 0 & \text{otherwise,} \end{cases} \tag{A.7}$$

for all $t$. Then, we have

$$\begin{aligned} \mathbf{\Delta}_0^\pi(N) + \mathbf{\Delta}_1^\pi(N) \;&\ge\; \left(\frac{2b_0}{a_0}\right)^2 \sum_{t=1}^N \mathbf{E}_0^\pi\big(p_t - \varphi(y_0)\big)^2 + \left(\frac{2b_1}{a_1}\right)^2 \sum_{t=1}^N \mathbf{E}_1^\pi\big(p_t - \varphi(y_1)\big)^2 \\ &\overset{\text{(e)}}{\ge}\; k_1 N^{-1/2} \sum_{t=1}^N \Big(\mathbf{P}_0^\pi(p_t \notin I_0) + \mathbf{P}_1^\pi(p_t \notin I_1)\Big) \\ &\overset{\text{(f)}}{\ge}\; k_1 N^{-1/2} \sum_{t=1}^N \Big(\mathbf{P}_0^\pi(\chi_t = 0) + \mathbf{P}_1^\pi(\chi_t = 1)\Big), \end{aligned} \tag{A.8}$$

where: $k_1 = \frac{1}{4}\min\left\{(b_0/a_0)^2, (b_1/a_1)^2\right\}$, (e) follows because $\big(p_t - \varphi(y_i)\big)^2 > \frac{1}{16}N^{-1/2}$ a.s. on the event $\{p_t \notin I_i\}$ for $i = 0, 1$, and (f) follows because $p_t \notin I_1$ is implied by $\chi_t = 1$. By Tsybakov's bound on minimax probability of error for two hypotheses (Tsybakov 2009, p. 90, Theorem 2.2(iii)), we know that $\mathcal{K}(\mathbf{P}_0^\pi, \mathbf{P}_1^\pi) \le \eta$ implies $\mathbf{P}_0^\pi(\chi_t = 0) + \mathbf{P}_1^\pi(\chi_t = 1) \ge \frac{1}{4}\exp(-\eta)$. Therefore we deduce by (A.8) that

$$\max_{i=0,1}\{\mathbf{\Delta}_i^\pi(N)\} \ge \frac{1}{4}\, k_1 \exp(-\eta) N^{1/2}. \tag{A.9}$$

Combining (A.6) and (A.9), we get $\max_{i=0,1}\{\mathbf{\Delta}_i^\pi(N)\} \ge k_2 N^{1/2}$, where $k_2 = \max\{2\sigma^2\eta, \frac{1}{4}k_1\exp(-\eta)\}$. Therefore we conclude that

$$\begin{aligned} \sup\left\{\Delta_{\boldsymbol{\theta}}^\pi(T) \,:\, V_{\boldsymbol{\theta}}(T) \le B\right\} \;&\overset{\text{(g)}}{\ge}\; \left\lfloor \frac{T}{N} \right\rfloor \max_{i=0,1}\{\mathbf{\Delta}_i^\pi(N)\} \\ &\ge\; k_2 \left\lfloor \frac{T}{N} \right\rfloor N^{1/2} \\ &\ge\; \frac{1}{2} k_2 N^{-1/2} T \\ &\ge\; c\, B^{1/3}\, T^{2/3}, \end{aligned} \tag{A.10}$$

where: $c = \frac{1}{8}k_2$ and (g) follows because there are at least $\lfloor T/N \rfloor$ cycles in $T$ periods. ■

# Appendix B: Proof of the Results in Section 3

**Proof of Lemma 1.** For the choice of weights in (3.8), the estimation inaccuracy in period $t$ is

$$\left(\mathcal{J}_t^t\right)^{-1}\mathcal{W}_t^t \;=\; \left(\mathcal{J}_t^t\right)^{-1} \sum_{s=1}^t w_s^t X_s X_s^\mathsf{T}(\theta_s - \theta_{t+1}) \;=\; \left(\mathcal{J}_t^t\right)^{-1} \sum_{\substack{s \in \mathcal{X}_1 \cup \mathcal{X}_2 \\ t - n^2 \le s \le t}} X_s X_s^\mathsf{T}(\theta_s - \theta_{t+1}), \tag{B.1}$$

for all $t$. Now note that $\left(\mathcal{J}_t^t\right)^{-1} = \left(\mathcal{I}_t^t\right)^{-1}\mathfrak{X}^{-1} = n^{-1}\mathfrak{X}^{-1}$ for all $t \geq n$ under $M_B(\kappa, x_1, x_2)$. Moreover, for $i = 1, 2$, we have $X_s X_s^\mathsf{T} = \begin{bmatrix} 1 & x_i \\ x_i & x_i^2 \end{bmatrix}$ for all $s \in \mathcal{X}_i$. Plugging these expressions of $\left(\mathcal{J}_t^t\right)^{-1}$ and $X_s X_s^\mathsf{T}$ into (B.1), we get

$$\begin{aligned} \left(\mathcal{J}_t^t\right)^{-1}\mathcal{W}_t^t &= \sum_{\substack{s\in\mathcal{X}_1 \\ t-n^2\leq s\leq t}} \tfrac{1}{(x_1-x_2)n}\begin{bmatrix} -x_2 & -x_1x_2 \\ 1 & x_1 \end{bmatrix}(\theta_s - \theta_{t+1}) \\ &\quad + \sum_{\substack{s\in\mathcal{X}_2 \\ t-n^2\leq s\leq t}} \tfrac{1}{(x_1-x_2)n}\begin{bmatrix} -x_1 & -x_1x_2 \\ 1 & x_2 \end{bmatrix}(\theta_s - \theta_{t+1}). \end{aligned} \tag{B.2}$$

Therefore we have

$$\begin{aligned} \left\|\left(\mathcal{J}_t^t\right)^{-1}\mathcal{W}_t^t\right\| &\overset{(a)}{\leq} \sum_{\substack{s\in\mathcal{X}_1 \\ t-n^2\leq s\leq t}} \left\|\tfrac{1}{(x_1-x_2)n}\begin{bmatrix} -x_2 & -x_1x_2 \\ 1 & x_1 \end{bmatrix}(\theta_s - \theta_{t+1})\right\| + \sum_{\substack{s\in\mathcal{X}_2 \\ t-n^2\leq s\leq t}} \left\|\tfrac{1}{(x_1-x_2)n}\begin{bmatrix} -x_1 & -x_1x_2 \\ 1 & x_2 \end{bmatrix}(\theta_s - \theta_{t+1})\right\| \\ &\overset{(b)}{\leq} \frac{1}{n}\sum_{\substack{s\in\mathcal{X}_1\cup\mathcal{X}_2 \\ t-n^2\leq s\leq t}} \left\|\theta_s - \theta_{t+1}\right\| \\ &\leq 2\max_{t-n^2\leq s\leq t}\left\|\theta_s - \theta_{t+1}\right\|, \end{aligned} \tag{B.3}$$

for all $t \geq n$, where: (a) follows by triangle inequality, and (b) follows by (B.2) and the fact that the eigenvalues of $\frac{1}{(x_1-x_2)n}\begin{bmatrix} -x_2 & -x_1x_2 \\ 1 & x_1 \end{bmatrix}$ and $\frac{1}{(x_1-x_2)n}\begin{bmatrix} -x_1 & -x_1x_2 \\ 1 & x_2 \end{bmatrix}$ are 0 and $\pm n^{-1}$. Squaring and summing both sides of (B.3) over $t = n, \ldots, T-1$, we get

$$\begin{aligned} \sum_{t=n}^{T-1}\left\|\left(\mathcal{J}_t^t\right)^{-1}\mathcal{W}_t^t\right\|^2 &\leq 4\sum_{t=n}^{T-1}\max_{t-n^2\leq s\leq t}\left\|\theta_s - \theta_{t+1}\right\|^2 \\ &\overset{(c)}{\leq} 4\sum_{j=1}^{\lceil T/n^2\rceil}\sum_{i=1}^{n^2}\max_{(j-1)n^2+i\leq s\leq jn^2+i}\left\|\theta_s - \theta_{jn^2+i+1}\right\|^2 \\ &\overset{(d)}{=} 4\sum_{i=1}^{n^2}\sum_{j=1}^{\lceil T/n^2\rceil}\max_{(j-1)n^2+i\leq s\leq jn^2+i}\left\|\theta_s - \theta_{jn^2+i+1}\right\|^2 \\ &\overset{(e)}{\leq} 4\sum_{i=1}^{n^2}V_{\boldsymbol{\theta}}(T) \\ &= 4n^2V_{\boldsymbol{\theta}}(T), \end{aligned} \tag{B.4}$$

where: (c) follows by expressing the time index as $t = jn^2 + i$, (d) follows by changing the order of summation, and (e) follows by (2.4) because $\{t_j = jn^2 + i : j = 0, 1, \ldots, \lceil T/n^2\rceil\}$ is a partition of

$\{1,\ldots,T\}$ for all $i$. By (2.5) and the preceding inequality, we conclude that

$$\sum_{t=n}^{T-1} \big\|\big(\mathcal{J}_t^t\big)^{-1}\mathcal{W}_t^t\big\|^2 \;\le\; 4n^2 B \;\overset{\text{(f)}}{\le}\; 16\kappa^2 B^{1/3}T^{2/3}, \tag{B.5}$$

where (f) follows because $n=\lceil \kappa B^{-1/3}T^{1/3}\rceil$. We obtain (3.11) by letting $c_1 = 16\kappa^2$. ■

**Proof of Lemma 2.** Because $\{\epsilon_t\}$ have a light-tailed distribution with mean zero and variance $\sigma^2$, we know by elementary real analysis that there exists a constant $\nu_0$ such that $\mathbb{E}[\exp(x\epsilon_t)] \le \exp(\frac{1}{2}\nu_0\sigma^2x^2)$ for all $x$ satisfying $|x|\le x_0$ (see, e.g., Keskin and Zeevi 2012, for a standard derivation of this constant). For any given $t=1,2,\ldots$ and $y\in\mathbb{R}^2$ such that $\|y\|=z$, define a stochastic process $\{\mathcal{Z}_s^{y,t},\ s=1,2,\ldots\}$ such that $\mathcal{Z}_0^{y,t}=1$ and

$$\mathcal{Z}_s^{y,t} = \begin{cases} \exp\big\{\frac{1}{\zeta}\big(y\cdot\mathcal{M}_s^t - \frac{1}{2}y^\mathsf{T}\mathcal{J}_s^t y\big)\big\} & \text{if } s\le t \\ \mathcal{Z}_{s-1}^{y,t} & \text{otherwise,}\end{cases} \tag{B.6}$$

where $\zeta = (1\vee z)\big(\nu_0\vee(x^*/x_0)\big)\sigma^2$ and $x^* = \max_{\|y\|\le 1,\, p\in[\ell,u]}\{|y_1+y_2p|/\sigma^2\}$. Let $\mathcal{F}_s := \sigma(\epsilon_1,\ldots,\epsilon_s)$. Using the tower property and the fact that $\mathcal{Z}_s^{y,t}$ is integrable for all $s$, we get

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

for $s\le t$. To find an upper bound on the conditional expectation on the right hand side of the identity immediately above, note that $\mathcal{M}_s^t-\mathcal{M}_{s-1}^t = w_s^t X_s\epsilon_s$, and $|y\cdot(w_s^tX_s)|/\zeta = w_s^t\,|y_1+y_2p_s|/\zeta \le w_s^t\,|y_1+y_2p_s|x_0/(zx^*\sigma^2) \le w_s^t\,x_0 \le x_0$ for all $p_s\in[\ell,u]$, because $w_s^t\le 1$ for all $s\le t$ by definition of the weights in (3.8). As a result, the conditional expectation on the right hand side of the preceding identity satisfies

$$\mathbb{E}_{\boldsymbol{\theta}}^{\pi}\Big[\exp\Big\{\frac{1}{\zeta}\ y\cdot(\mathcal{M}_s^t-\mathcal{M}_{s-1}^t)\Big\}\Big|\mathcal{F}_{s-1}\Big] \;\le\; \exp\Big\{\frac{1}{2\zeta^2}\,\nu_0\sigma^2(w_s^t)^2\,y^\mathsf{T}X_sX_s^\mathsf{T}y\Big\} \;\le\; \exp\Big\{\frac{1}{2\zeta}\,w_s^t\,y^\mathsf{T}X_sX_s^\mathsf{T}y\Big\}.$$

Consequently we get

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

So $(\mathcal{Z}_s^{y,t},\mathcal{F}_s)$ is a supermartingale for any given $y\in\mathbb{R}^2$ and $t=1,2,\ldots$

To derive inequality (3.12), recall equation (3.9), which states that $\mathcal{J}_s^t = \mathfrak{X}\mathcal{I}_s^t$ where $\mathfrak{X} = \begin{bmatrix} 2 & x_1+x_2 \\ x_1+x_2 & x_1^2+x_2^2\end{bmatrix}$ and $\mathcal{I}_s^t = \frac{1}{2}\sum_{q=1}^s w_q^t \ge 0$. Let $\mathcal{V}\subset\mathbb{R}^2$ be the set of eigenvectors of $\mathfrak{X}$, and consider the eigendecomposition of $\mathfrak{X}$:

$$\mathfrak{X} = P\Lambda P^\mathsf{T},$$

where $P$ is an orthogonal matrix that has the eigenvectors of $\mathfrak{X}$ in its columns, and $\Lambda$ is a diagonal matrix that has the eigenvalues of $\mathfrak{X}$ in its diagonal entries. For $z>0$, define a set of four vectors

$\mathcal{S}_z := \{\pm\frac{1}{\sqrt{2}} zv : v \in \mathcal{V}\} = \{\pm\frac{1}{\sqrt{2}} zP_i : P_i \text{ is the } i^{\text{th}} \text{ column of } P\}$. Then we have

$$\begin{aligned}
\mathbb{P}_{\boldsymbol{\theta}}^{\pi}\big\{\big\|\big(\mathcal{J}_t^t\big)^{-1}\mathcal{M}_t^t\big\| > z,\ \mathcal{I}_t^t > \gamma\big\} &= \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\big\{(\mathcal{M}_t^t)^{\mathsf{T}}(\mathcal{J}_t^t)^{-2}\mathcal{M}_t^t > z^2,\ \mathcal{I}_t^t > \gamma\big\} \\
&= \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\big\{(\mathcal{M}_t^t)^{\mathsf{T}} P\Lambda^{-2}P^{\mathsf{T}}\mathcal{M}_t^t > z^2(\mathcal{I}_t^t)^2,\ \mathcal{I}_t^t > \gamma\big\}.
\end{aligned}$$

Letting $\psi := \Lambda^{-1}P^{\mathsf{T}}\mathcal{M}_t^t$ we deduce that

$$\begin{aligned}
\mathbb{P}_{\boldsymbol{\theta}}^{\pi}\big\{\big\|\big(\mathcal{J}_t^t\big)^{-1}\mathcal{M}_t^t\big\| > z,\ \mathcal{I}_t^t > \gamma\big\} &\le \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\big\{\|\psi\| > z\,\mathcal{I}_t^t,\ \mathcal{I}_t^t > \gamma\big\} \\
&\le \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\Big\{\bigcup_{i=1,2}\Big\{|\psi_i| > \tfrac{1}{\sqrt{2}} z\,\mathcal{I}_t^t,\ \mathcal{I}_t^t > \gamma\Big\}\Big\}.
\end{aligned}$$

Note that $|\psi_i| > \frac{1}{\sqrt{2}} z\,\mathcal{I}_t^t$ implies that $P_i^{\mathsf{T}}\mathcal{M}_t^t$, the $i^{\text{th}}$ component of $P^{\mathsf{T}}\mathcal{M}_t^t$, has an absolute value larger than $\frac{1}{\sqrt{2}} z\lambda_i\,\mathcal{I}_t^t$, where $\lambda_i = P_i^{\mathsf{T}}\mathfrak{X}P_i$ is the $i^{\text{th}}$ diagonal entry of $\Lambda$. Therefore, viewing $\mathcal{V}$ as a basis for $\mathbb{R}^2$, we have

$$\begin{aligned}
\mathbb{P}_{\boldsymbol{\theta}}^{\pi}\big\{\big\|\big(\mathcal{J}_t^t\big)^{-1}\mathcal{M}_t^t\big\| > z,\ \mathcal{I}_t^t > \gamma\big\} &\le \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\Big\{\bigcup_{i=1,2}\Big\{|P_i^{\mathsf{T}}\mathcal{M}_t^t| > \tfrac{1}{\sqrt{2}} z\lambda_i\,\mathcal{I}_t^t,\ \mathcal{I}_t^t > \gamma\Big\}\Big\} \\
&\overset{\text{(a)}}{=} \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\Big\{\bigcup_{i=1,2}\Big\{|P_i^{\mathsf{T}}\mathcal{M}_t^t| > \tfrac{1}{\sqrt{2}} zP_i^{\mathsf{T}}\mathcal{J}_t^t P_i,\ \mathcal{I}_t^t > \gamma\Big\}\Big\} \\
&= \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\Big\{\bigcup_{i=1,2}\Big\{\Big|\big(\tfrac{1}{\sqrt{2}} zP_i\big)^{\mathsf{T}}\mathcal{M}_t^t\Big| > \big(\tfrac{1}{\sqrt{2}} zP_i\big)^{\mathsf{T}}\mathcal{J}_t^t\,\big(\tfrac{1}{\sqrt{2}} zP_i\big),\ \mathcal{I}_t^t > \gamma\Big\}\Big\} \\
&\overset{\text{(b)}}{=} \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\Big\{\bigcup_{w\in\mathcal{S}_z}\big\{w\cdot\mathcal{M}_t^t > w^{\mathsf{T}}\mathcal{J}_t^t w,\ \mathcal{I}_t^t > \gamma\big\}\Big\} \\
&\overset{\text{(c)}}{\le} \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\Big\{\bigcup_{w\in\mathcal{S}_z}\big\{\mathcal{Z}_t^{w,t} \ge e^{\frac{1}{2\zeta}w^{\mathsf{T}}\mathcal{J}_t^t w},\ \mathcal{I}_t^t > \gamma\big\}\Big\} \\
&\overset{\text{(d)}}{\le} \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\Big\{\bigcup_{w\in\mathcal{S}_z}\big\{\mathcal{Z}_t^{w,t} \ge e^{\rho_0 z^2\mathcal{I}_t^t},\ \mathcal{I}_t^t > \gamma\big\}\Big\} \\
&\le \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\Big\{\bigcup_{w\in\mathcal{S}_z}\big\{\mathcal{Z}_t^{w,t} \ge e^{\rho_0 z^2\gamma}\big\}\Big\},
\end{aligned}$$

where: $\rho_0 = (\lambda_1 \wedge \lambda_2)/(4\zeta)$, (a) and (d) follow because $\lambda_i = P_i^{\mathsf{T}}\mathfrak{X}P_i$ and $\mathcal{J}_t^t = \mathfrak{X}\,\mathcal{I}_t^t$, (b) follows by the definition of $\mathcal{S}_z = \{\pm\frac{1}{\sqrt{2}} zP_i : i = 1, 2\}$, and (c) follows by the definition of $\mathcal{Z}_t^{w,t}$ in (B.6). We therefore have

$$\mathbb{P}_{\boldsymbol{\theta}}^{\pi}\big\{\big\|\big(\mathcal{J}_t^t\big)^{-1}\mathcal{M}_t^t\big\| > z,\ \mathcal{I}_t^t > \gamma\big\} \overset{\text{(e)}}{\le} \sum_{w\in\mathcal{S}_z}\mathbb{P}_{\boldsymbol{\theta}}^{\pi}\big\{\mathcal{Z}_t^{w,t} \ge e^{\rho_0 z^2\gamma}\big\} \overset{\text{(f)}}{\le} \sum_{w\in\mathcal{S}_z} e^{-\rho_0 z^2\gamma} \overset{\text{(g)}}{=} 4e^{-\rho_0 z^2\gamma},$$

where: (e) follows by the union bound, (f) follows by the Markov's inequality and the fact that $(\mathcal{Z}_s^{w,t}, \mathcal{F}_s)$ is a supermartingale, and (g) follows because the cardinality of $\mathcal{S}_z$ is 4. We conclude the proof by letting $\rho = (1 \vee z)\rho_0 = (\lambda_1 \wedge \lambda_2)/\big(4\big(\nu_0 \vee (x^*/x_0)\big)\sigma^2\big)$. ■

**Proof of Theorem 2.** In the context of Section 3, the loss of a policy stems from three sources: (i) estimation inaccuracy due to changes in demand parameters, (ii) estimation errors due to noise, and (iii) price experimentation. To separate the effect of (iii) from (i-ii), note that

$$
\begin{aligned}
\Delta_{\boldsymbol{\theta}}^{\pi}(T) &= \mathbb{E}_{\boldsymbol{\theta}}^{\pi}\Big\{ \sum_{t=1}^{T} \Big( 1 - \frac{r(p_t,\theta_t)}{r^*(\theta_t)} \Big) \Big\} \\
&= \mathbb{E}_{\boldsymbol{\theta}}^{\pi}\Big\{ \sum_{t=1}^{T} \Big( 1 - \frac{r(p_t,\theta_t)}{r^*(\theta_t)} \Big) \mathbb{I}\{t \in \mathcal{X}\} \Big\} \;+\; \mathbb{E}_{\boldsymbol{\theta}}^{\pi}\Big\{ \sum_{t=1}^{T} \Big( 1 - \frac{r(p_t,\theta_t)}{r^*(\theta_t)} \Big) \mathbb{I}\{t \notin \mathcal{X}\} \Big\} \\
&\le 2n \left\lceil \frac{T}{n^2} \right\rceil \;+\; \mathbb{E}_{\boldsymbol{\theta}}^{\pi}\Big\{ \sum_{t=n+1}^{T} \Big( 1 - \frac{r(p_t,\theta_t)}{r^*(\theta_t)} \Big) \mathbb{I}\{t \notin \mathcal{X}\} \Big\} \\
&\le 4B^{1/3}T^{2/3} \;+\; \mathbb{E}_{\boldsymbol{\theta}}^{\pi}\Big\{ \sum_{t=n+1}^{T} \Big( 1 - \frac{r(p_t,\theta_t)}{r^*(\theta_t)} \Big) \mathbb{I}\{t \notin \mathcal{X}\} \Big\},
\end{aligned} \tag{B.7}
$$

because $t \in \mathcal{X}$ for all $t \le n$, and $n = \lceil \kappa B^{-1/3} T^{1/3} \rceil$. To find an upper bound on the expected sum on the right hand side above, we further note that

$$
\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}}^{\pi}\Big\{ \sum_{t=n+1}^{T} \Big( 1 - \frac{r(p_t,\theta_t)}{r^*(\theta_t)} \Big) \mathbb{I}\{t \notin \mathcal{X}\} \Big\} &= \mathbb{E}_{\boldsymbol{\theta}}^{\pi}\Big\{ \sum_{t=n}^{T-1} \Big( 1 - \frac{r(p_{t+1},\theta_{t+1})}{r^*(\theta_{t+1})} \Big) \mathbb{I}\{t+1 \notin \mathcal{X}\} \Big\} \\
&= -\frac{\beta_{t+1}}{r^*(\theta_{t+1})} \sum_{t=n}^{T-1} \mathbb{E}_{\boldsymbol{\theta}}^{\pi}\Big\{ \big(\varphi(\theta_{t+1}) - p_{t+1}\big)^2 \, \mathbb{I}\{t+1 \notin \mathcal{X}\} \Big\} \\
&\le c_2 \sum_{t=n}^{T-1} \mathbb{E}_{\boldsymbol{\theta}}^{\pi}\Big\{ \big(\varphi(\theta_{t+1}) - p_{t+1}\big)^2 \, \mathbb{I}\{t+1 \notin \mathcal{X}\} \Big\},
\end{aligned} \tag{B.8}
$$

where $c_2 = \max_{(\alpha,\beta)\in\Theta}\{4\beta^2/\alpha^2\}$. Here (3.4) and (3.7) imply that

$$
\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}}^{\pi}\Big\{ \big(\varphi(\theta_{t+1}) - p_{t+1}\big)^2 \, \mathbb{I}\{t+1 \notin \mathcal{X}\} \Big\} &\le \mathbb{E}_{\boldsymbol{\theta}}^{\pi}\big(\varphi(\theta_{t+1}) - \varphi(\vartheta_{t+1})\big)^2 \\
&\le 2K_0 \, \mathbb{E}_{\boldsymbol{\theta}}^{\pi}\big\| \big(\mathcal{J}_t^t\big)^{-1} \mathcal{W}_t^t \big\|^2 \;+\; 2K_0 \, \mathbb{E}_{\boldsymbol{\theta}}^{\pi}\big\| \big(\mathcal{J}_t^t\big)^{-1} \mathcal{M}_t^t \big\|^2,
\end{aligned} \tag{B.9}
$$

for all $t \ge n$, where $\vartheta_t = \arg\min_{\vartheta\in\Theta}\{\|\vartheta - \hat{\theta}_t\|\}$ is the truncated least squares estimate of $\theta_t$, and $K_0 = \max_{j=1,2}\big\{ \max_{\theta}\{(\partial\varphi(\theta)/\partial\theta_j)^2\}\big\}$. We find an upper bound on the second term on the right hand side of inequality (B.9) as follows: under $\pi = M_B(\kappa, x_1, x_2)$, we have

$$\begin{aligned}
\mathbb{E}^{\pi}_{\boldsymbol{\theta}} \big\| \big(\mathcal{J}^t_t\big)^{-1} \mathcal{M}^t_t \big\|^2 &= \int_0^\infty \mathbb{P}^{\pi}_{\boldsymbol{\theta}} \big\{ \big\| \big(\mathcal{J}^t_t\big)^{-1} \mathcal{M}^t_t \big\|^2 > \xi \big\} \, d\xi \,, \\
&\overset{\text{(a)}}{=} \int_0^\infty \mathbb{P}^{\pi}_{\boldsymbol{\theta}} \big\{ \big\| \big(\mathcal{J}^t_t\big)^{-1} \mathcal{M}^t_t \big\|^2 > \xi, \, \mathcal{I}^t_t \geq n \big\} \, d\xi \,, \\
&\overset{\text{(b)}}{\leq} 4 \int_0^\infty e^{-\rho(\sqrt{\xi} \wedge \xi) n} \, d\xi \\
&= 4 \int_0^1 e^{-\rho n \xi} \, d\xi + 4 \int_1^\infty e^{-\rho n \sqrt{\xi}} \, d\xi \\
&\leq 12/(\rho n), \qquad \text{(B.10)}
\end{aligned}$$

for all $t \geq n$, where: (a) follows because $\mathcal{I}^t_t \geq n$ for all $t \geq n$ under $M_B(\kappa, x_1, x_2)$, and (b) follows by Lemma 2. Thus, inequality (B.9) becomes

$$\mathbb{E}^{\pi}_{\boldsymbol{\theta}} \Big\{ \big(\varphi(\theta_{t+1}) - p_{t+1}\big)^2 \, \mathbb{I}\{t+1 \notin \mathcal{X}\} \Big\} \; \leq \; 2K_0 \, \mathbb{E}^{\pi}_{\boldsymbol{\theta}} \big\| \big(\mathcal{J}^t_t\big)^{-1} \mathcal{W}^t_t \big\|^2 \; + \; \frac{24K_0}{\rho n} \,, \qquad \text{(B.11)}$$

for all $t \geq n$. Summing over $t = n, \ldots, T-1$, we obtain

$$\sum_{t=n}^{T-1} \mathbb{E}^{\pi}_{\boldsymbol{\theta}} \Big\{ \big(\varphi(\theta_{t+1}) - p_{t+1}\big)^2 \, \mathbb{I}\{t+1 \notin \mathcal{X}\} \Big\} \; \leq \; 2K_0 \sum_{t=n}^{T-1} \mathbb{E}^{\pi}_{\boldsymbol{\theta}} \big\| \big(\mathcal{J}^t_t\big)^{-1} \mathcal{W}^t_t \big\|^2 \; + \; \frac{24K_0}{\rho} \, n^{-1} T \,. \qquad \text{(B.12)}$$

Invoking Lemma 1, and recalling that $n \geq \kappa B^{-1/3} T^{1/3}$, we deduce that

$$\sum_{t=n}^{T-1} \mathbb{E}^{\pi}_{\boldsymbol{\theta}} \Big\{ \big(\varphi(\theta_{t+1}) - p_{t+1}\big)^2 \, \mathbb{I}\{t+1 \notin \mathcal{X}\} \Big\} \; \leq \; 2K_0 c_1 B^{1/3} T^{2/3} + \frac{24K_0}{\rho\kappa} \, B^{1/3} T^{2/3} \; = \; c_3 B^{1/3} T^{2/3},$$

where $c_3 = 2K_0 c_1 + 24K_0/(\rho\kappa)$. Combining the preceding inequality with inequalities (B.7-B.8), we conclude that $\Delta^{\pi}_{\boldsymbol{\theta}}(T) \leq C B^{1/3} T^{2/3}$ for all $\boldsymbol{\theta} \in \mathcal{V}(T, B)$, where $C = 4 + c_2 c_3$. ■

**Proof of Theorem 3.** As in the proof of Theorem 2, we first isolate the loss due to price experimentation from other losses by noting that

$$\Delta^{\pi}_{\boldsymbol{\theta}}(T) \; = \; \mathbb{E}^{\pi}_{\boldsymbol{\theta}} \Big\{ \sum_{t=1}^{T} \Big(1 - \frac{r(p_t, \theta_t)}{r^*(\theta_t)}\Big) \Big\} \; \leq \; 3n^{-1} T \; + \; \mathbb{E}^{\pi}_{\boldsymbol{\theta}} \Big\{ \sum_{t=1}^{T} \Big(1 - \frac{r(p_t, \theta_t)}{r^*(\theta_t)}\Big) \mathbb{I}\{t \notin \mathcal{X}\} \Big\}, \qquad \text{(B.13)}$$

because the cardinality of $\mathcal{X}$ is less than or equal to $2(T/n + 1) \leq 3T/n$. The expected sum on the right hand side above is bounded above as follows:

$$\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\Big\{\sum_{t=1}^{T}\Big(1-\frac{r(p_t,\theta_t)}{r^*(\theta_t)}\Big)\mathbb{I}\{t\notin\mathcal{X}\}\Big\}$$

$$= \mathbb{E}^{\pi}_{\boldsymbol{\theta}}\Big\{\sum_{t=0}^{T-1}\Big(1-\frac{r(p_{t+1},\theta_{t+1})}{r^*(\theta_{t+1})}\Big)\mathbb{I}\{t+1\notin\mathcal{X}\}\Big\}$$

$$= n^2 - \frac{\beta_{t+1}}{r^*(\theta_{t+1})}\sum_{t=n^2}^{T-1}\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\Big\{\big(\varphi(\theta_{t+1})-p_{t+1}\big)^2\,\mathbb{I}\{t+1\notin\mathcal{X}\}\Big\}$$

$$\leq n^2 + c_2\sum_{t=n^2}^{T-1}\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\Big\{\big(\varphi(\theta_{t+1})-p_{t+1}\big)^2\,\mathbb{I}\{t+1\notin\mathcal{X}\}\Big\}$$

$$\overset{\text{(a)}}{\leq} n^2 + 2K_0c_2\sum_{t=n^2}^{T-1}\Big(\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\|(\mathcal{J}^t_t)^{-1}\mathcal{W}^t_t\|^2+\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\|(\mathcal{J}^t_t)^{-1}\mathcal{M}^t_t\|^2\Big), \tag{B.14}$$

where $c_2=\max_{(\alpha,\beta)\in\Theta}\{4\beta^2/\alpha^2\}$, $K_0=\max_{j=1,2}\big\{\max_\theta\{(\partial\varphi(\theta)/\partial\theta_j)^2\}\big\}$, and (a) follows by invoking identity (3.4) for the price experimentation scheme (3.13-3.14). To characterize the magnitude of $\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\|(\mathcal{J}^t_t)^{-1}\mathcal{M}^t_t\|^2$, we use the following result whose proof is identical to that of Lemma 2.

**Lemma B.1 (exponential decay of estimation error due to noise)** *Let $\pi$ be either $M(\kappa,x_1,x_2)$ or $W(\mu,\kappa,x_1,x_2)$. Then there exists a finite positive constant $\rho$ such that $\mathbb{P}^{\pi}_{\boldsymbol{\theta}}\{\|(\mathcal{J}^t_t)^{-1}\mathcal{M}^t_t\|>z,\ \mathcal{I}^t_t>\gamma\}\leq 4\,e^{-\rho(z\wedge z^2)\gamma}$ for all $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_T)$, $z>0$, $\gamma>0$, and $t\geq 2$.*

Replacing Lemma 2 with Lemma B.1 in the argument used to derive (B.10), we deduce that $\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\|(\mathcal{J}^t_t)^{-1}\mathcal{M}^t_t\|^2\leq 12/(\rho n)$ for all $t\geq n^2$ under $\pi=M(\kappa,x_1,x_2)$. On the other hand, if $\pi=W(\mu,\kappa,x_1,x_2)$, then we note that

$$\mathcal{I}^t_t = \frac{1}{2}\sum_{q=1}^{t}w^t_q = \frac{1}{2}\sum_{\substack{q\in\mathcal{X}\\1\leq q\leq t}}\Big(1-\frac{t-q}{n^2}+\frac{(t-q)^{1-\mu}}{n^2}\Big)_+^{\frac{1}{\mu}}$$

$$\geq \frac{1}{2}\sum_{\substack{q\in\mathcal{X}\\1\leq q\leq t}}\Big(1-\frac{1}{n}\Big\lceil\frac{t-q}{n}\Big\rceil\Big)_+^{\frac{1}{\mu}}$$

$$\overset{\text{(b)}}{=} n^{-\frac{1}{\mu}}\sum_{k=1}^{n-1}(n-k)^{\frac{1}{\mu}}$$

$$\geq n^{-\frac{1}{\mu}}\int_0^{n-1}\xi^{\frac{1}{\mu}}\,d\xi$$

$$\geq c_\mu n\,, \tag{B.15}$$

for all $t \geq n^2$, where: $c_\mu = 2^{-(\frac{1}{\mu}+1)}/(\frac{1}{\mu}+1)$, and (b) follows by letting $k = \lceil (t-q)/n \rceil$. Thus, under $\pi = W(\mu, \kappa, x_1, x_2)$, we have $\mathbb{E}^\pi_{\boldsymbol{\theta}} \|(\mathcal{J}_t^t)^{-1}\mathcal{M}_t^t\|^2 \leq 12/(\rho c_\mu n)$ for all $t \geq n^2$. Letting $\tilde{c} = 1 \wedge c_\mu$, we therefore get $\mathbb{E}^\pi_{\boldsymbol{\theta}} \|(\mathcal{J}_t^t)^{-1}\mathcal{M}_t^t\|^2 \leq 12/(\rho \tilde{c} n)$ for all $t \geq n^2$. Using this inequality on the right hand side of (B.14) we get

$$\mathbb{E}^\pi_{\boldsymbol{\theta}} \left\{ \sum_{t=1}^{T} \left(1 - \frac{r(p_t, \theta_t)}{r^*(\theta_t)}\right) \mathbb{I}\{t \notin \mathcal{X}\} \right\} \;\leq\; n^2 + 2K_0 c_2 \sum_{t=n^2}^{T-1} \mathbb{E}^\pi_{\boldsymbol{\theta}} \|(\mathcal{J}_t^t)^{-1}\mathcal{W}_t^t\|^2 + \frac{24 K_0 T}{\rho \tilde{c} n}. \tag{B.16}$$

Next we use the following generalization of Lemma 1, whose proof is at the end of this section.

**Lemma B.2 (upper bound on aggregate estimation inaccuracy)** *There exists a finite positive constant $c_1$, such that under either $M(\kappa, x_1, x_2)$ or $W(\mu, \kappa, x_1, x_2)$, $\sum_{t=n^2}^{T-1} \|(\mathcal{J}_t^t)^{-1}\mathcal{W}_t^t\|^2 \leq c_1 T^{2/3}$ almost surely for all $T = 1, 2, \ldots$ and $\boldsymbol{\theta} \in \mathcal{V}(T, B)$.*

By Lemma B.2 and inequality (B.16), we deduce that $\mathbb{E}^\pi_{\boldsymbol{\theta}} \{ \sum_{t=1}^T (1 - r(p_t, \theta_t)/r^*(\theta_t)) \mathbb{I}\{t \notin \mathcal{X}\} \} \leq n^2 + 2K_0 c_1 c_2 T^{2/3} + 24 K_0 T/(\rho \tilde{c} n)$. Plugging the value of $n = \lceil \kappa T^{1/3} \rceil$ in this upper bound, we get $\mathbb{E}^\pi_{\boldsymbol{\theta}} \{ \sum_{t=1}^T (1 - r(p_t, \theta_t)/r^*(\theta_t)) \mathbb{I}\{t \notin \mathcal{X}\} \} \leq 4\kappa^2 T^{2/3} + 2K_0 c_1 c_2 T^{2/3} + 24 K_0 T^{2/3}/(\rho \tilde{c} \kappa)$. Combining this inequality with (B.13) we conclude that $\Delta^\pi_{\boldsymbol{\theta}}(T) \leq C T^{2/3}$ for all $\boldsymbol{\theta} \in \mathcal{V}(T, B)$, where $C = 3/\kappa + 4\kappa^2 + 2K_0 c_1 c_2 + 24 K_0/(\rho \tilde{c} \kappa)$. ■

**Proof of Lemma B.2.** The proof of the claim for $M(\kappa, x_1, x_2)$ follows by the arguments used in the proof of Lemma 1. For $W(\mu, \kappa, x_1, x_2)$ we first note that, because $(\mathcal{J}_t^t)^{-1} = (\mathcal{I}_t^t)^{-1} \mathfrak{X}^{-1}$ for all $t \geq n^2$, we have

$$\begin{aligned}
(\mathcal{J}_t^t)^{-1}\mathcal{W}_t^t \;=\; & \sum_{s \in \mathcal{X}_1} \tfrac{w_s^t}{(x_1 - x_2)\mathcal{I}_t^t} \begin{bmatrix} -x_2 & -x_1 x_2 \\ 1 & x_1 \end{bmatrix} (\theta_s - \theta_{t+1}) \\
& + \sum_{s \in \mathcal{X}_2} \tfrac{w_s^t}{(x_1 - x_2)\mathcal{I}_t^t} \begin{bmatrix} -x_1 & -x_1 x_2 \\ 1 & x_2 \end{bmatrix} (\theta_s - \theta_{t+1}).
\end{aligned} \tag{B.17}$$

Consequently, we get

$$\begin{aligned}
\|(\mathcal{J}_t^t)^{-1}\mathcal{W}_t^t\| \;&\stackrel{(a')}{\leq}\; \sum_{s \in \mathcal{X}_1} w_s^t \left\| \tfrac{1}{(x_1 - x_2)n} \begin{bmatrix} -x_2 & -x_1 x_2 \\ 1 & x_1 \end{bmatrix} (\theta_s - \theta_{t+1}) \right\| + \sum_{s \in \mathcal{X}_2} w_s^t \left\| \tfrac{1}{(x_1 - x_2)n} \begin{bmatrix} -x_1 & -x_1 x_2 \\ 1 & x_2 \end{bmatrix} (\theta_s - \theta_{t+1}) \right\| \\
&\stackrel{(b')}{\leq}\; (\mathcal{I}_t^t)^{-1} \sum_{s \in \mathcal{X}_1 \cup \mathcal{X}_2} w_s^t \|\theta_s - \theta_{t+1}\| \\
&\stackrel{(c')}{\leq}\; (\mathcal{I}_t^t)^{-1} n^{-2} \sum_{\substack{s \in \mathcal{X}_1 \cup \mathcal{X}_2 \\ 1 \leq s < t - n^2}} \|\theta_s - \theta_{t+1}\| + (\mathcal{I}_t^t)^{-1} \sum_{\substack{s \in \mathcal{X}_1 \cup \mathcal{X}_2 \\ t - n^2 \leq s \leq t}} \|\theta_s - \theta_{t+1}\| \\
&\leq\; \frac{2(t - n^2 - 1)}{n^3} (\mathcal{I}_t^t)^{-1} \max_{1 \leq s < t - n^2} \|\theta_s - \theta_{t+1}\| \;+\; \frac{2n^2}{n} (\mathcal{I}_t^t)^{-1} \max_{t - n^2 \leq s \leq t} \|\theta_s - \theta_{t+1}\| \\
&\leq\; \frac{2t}{n^3} (\mathcal{I}_t^t)^{-1} \max_{1 \leq s < t - n^2} \|\theta_s - \theta_{t+1}\| \;+\; 2n (\mathcal{I}_t^t)^{-1} \max_{t - n^2 \leq s \leq t} \|\theta_s - \theta_{t+1}\|,
\end{aligned} \tag{B.18}$$

for all $t \geq n^2$, where: (a′) follows by triangle inequality, (b′) follows by (B.17) and the fact that the eigenvalues of $\frac{1}{(x_1-x_2)\mathcal{I}_t^t}\begin{bmatrix} -x_2 & -x_1x_2 \\ 1 & x_1 \end{bmatrix}$ and $\frac{1}{(x_1-x_2)\mathcal{I}_t^t}\begin{bmatrix} -x_1 & -x_1x_2 \\ 1 & x_2 \end{bmatrix}$ are 0 and $\pm\left(\mathcal{I}_t^t\right)^{-1}$, and (c′) follows because $w_s^t \leq n^{-2}$ for all $s < t - n^2$, and $w_s^t \leq 1$ for all $s \leq t$. We square and sum both sides of (B.18) over $t = n^2, \ldots, T-1$ to obtain

$$\sum_{t=n^2}^{T-1} \left\|\left(\mathcal{J}_t^t\right)^{-1}\mathcal{W}_t^t\right\|^2 \;\leq\; 8n^{-6}\sum_{t=n^2}^{T-1} t^2\left(\mathcal{I}_t^t\right)^{-2} \max_{1\leq s<t-n^2}\left\|\theta_s - \theta_{t+1}\right\|^2$$

$$+\, 8n^2 \sum_{t=n^2}^{T-1} \left(\mathcal{I}_t^t\right)^{-2} \max_{t-n^2\leq s\leq t}\left\|\theta_s - \theta_{t+1}\right\|^2. \tag{B.19}$$

By the argument used to derive (B.4-B.5), and the fact that $\mathcal{I}_t^t \geq c_\mu n$ for all $t \geq n^2$, we deduce that the second term on the right hand side of (B.19) is bounded above by $32c_\mu^{-2}\kappa^2 BT^{2/3}$. To find an upper bound on the first term, we note that

$$8n^{-6}\sum_{t=n^2+1}^{T-1} t^2\left(\mathcal{I}_t^t\right)^{-2}\max_{1\leq s<t-n^2}\left\|\theta_s-\theta_{t+1}\right\|^2 \overset{(d')}{\leq} 8n^{-6}\sum_{t=n^2+1}^{T-1} t^2\left(\mathcal{I}_t^t\right)^{-2}B$$

$$\overset{(e')}{\leq} 8c_\mu^{-2}n^{-8}\sum_{t=n^2+1}^{T-1} t^2 B$$

$$\leq 8c_\mu^{-2}n^{-8}BT^3, \tag{B.20}$$

where: (d′) follows by (2.5), and (e′) follows because $\mathcal{I}_t^t > c_\mu n$ for all $t \geq n^2$. Furthermore, because $n = \lceil \kappa T^{1/3}\rceil$, the right hand side of the preceding inequality is less than or equal to $8c_\mu^{-2}\kappa^{-8}BT^{1/3}$. Thus the right hand side of (B.19) is bounded above by $8c_\mu^{-2}(\kappa^{-8}+4\kappa^2)BT^{2/3}$. ■

## Appendix C: Proof of the Results in Section 4

Note that if $\tau_j^* \geq \tau_{j+1}^* - 2$, then the expected value on the right hand side of (4.10) is less than $2n \leq 4\kappa\sqrt{T}$, which gives us an upper bound on the losses due to detection and estimation between $j^{\text{th}}$ and $(j+1)^{\text{st}}$ change-points. Therefore, we hereafter focus on the case $\tau_j^* < \tau_{j+1}^* - 2$ for any given $j = 0, \ldots, \mathcal{C}$.

**Proof of Lemma 3.** Recall that the $\epsilon_t$ have a light-tailed distribution, which implies that there exist finite constants $x_0$ and $\nu_0$ such that $\mathbb{E}_{\boldsymbol{\theta}}^{\pi}[\exp(x\epsilon_t)] \leq \exp(\frac{1}{2}\nu_0\sigma^2x^2)$ for all $x$ satisfying $|x| \leq x_0$. Choosing $c_\epsilon = (6/x_0) \vee (12\nu_0\sigma^2)$, we prove (4.11) for $\pi = D(\eta, \kappa, x_1, x_2)$ with $\kappa = c_\epsilon/(\eta \wedge \eta^2)$. Note that, for the case $\epsilon_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$, we have $c_\epsilon = 12\sigma^2$.

Given a real number $y$ with $|y| = \frac{1}{2}\eta$, let $\{Z_t^y, t = 1, 2, \ldots\}$ be a stochastic process satisfying $Z_0^y = 1$ and

$$Z_t^y := \exp\left\{\tfrac{1}{\zeta}\left(yS_t - \tfrac{1}{2}y^2t\right)\right\} \quad \text{for all } t = 1, 2, \ldots \tag{C.1}$$

where $\zeta = \left(\frac{\eta}{2x_0}\right) \vee (\nu_0\sigma^2)$ and $S_t = \sum_{q=1}^{t} \epsilon_q$. First note that $Z_t^y$ is integrable for all $t$. Let $\mathcal{F}_t := \sigma(\epsilon_1, \ldots, \epsilon_t)$. Then, we have

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}}^{\pi}[Z_t^y|\mathcal{F}_{t-1}] &= \exp\left\{\tfrac{1}{\zeta}\left(yS_{t-1} - \tfrac{1}{2}y^2t\right)\right\} \mathbb{E}_{\boldsymbol{\theta}}^{\pi}\left[\exp\left(\tfrac{1}{\zeta}\, y\epsilon_t\right)|\mathcal{F}_{t-1}\right] \\
&\overset{(a)}{\leq} \exp\left\{\tfrac{1}{\zeta}\left(yS_{t-1} - \tfrac{1}{2}y^2t\right)\right\} \exp\left\{\tfrac{1}{2\zeta^2}\,\nu_0\sigma^2y^2\right\} \\
&\overset{(b)}{\leq} \exp\left\{\tfrac{1}{\zeta}\left(yS_{t-1} - \tfrac{1}{2}y^2(t-1)\right)\right\} \\
&= Z_{t-1}^y. \qquad \text{(C.2)}
\end{aligned}$$

for all $t = 1, 2, \ldots,$ where: (a) follows because $|y/\zeta| = |\eta/(2\zeta)| \leq x_0$, and (b) follows because $\nu_0\sigma^2 \leq \zeta$. Thus $(Z_t^y, \mathcal{F}_t)$ is a supermartingale for all $y \in \mathbb{R}$. Now note that

$$\bar{\epsilon}_{10} = \frac{1}{m}\sum_{t\in\mathcal{X}_{10}} \epsilon_t = \frac{1}{m}\sum_{t=1}^{m} \epsilon_t = \frac{S_m}{m}. \qquad \text{(C.3)}$$

Therefore, we have

$$\mathbb{P}_{\boldsymbol{\theta}}^{\pi}\left\{\left|\bar{\epsilon}_{10}\right| \geq \tfrac{1}{2}\eta\right\} = \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\left\{\left|S_m\right| \geq \tfrac{1}{2}\eta m\right\}. \qquad \text{(C.4)}$$

Choosing $y = \frac{1}{2}\eta$, we get

$$\begin{aligned}
\mathbb{P}_{\boldsymbol{\theta}}^{\pi}\{S_m \geq \tfrac{1}{2}\eta m\} &= \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\{S_m \geq ym\} \\
&= \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\{yS_m - \tfrac{1}{2}y^2m \geq \tfrac{1}{2}y^2m\} \\
&\leq \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\{Z_m^y \geq \exp\left(\tfrac{1}{2\zeta}y^2m\right)\} \\
&\overset{(c)}{\leq} \mathbb{E}_{\boldsymbol{\theta}}^{\pi}\{Z_m^y\} \exp\left(-\tfrac{1}{2\zeta}y^2m\right) \\
&\overset{(d)}{\leq} \exp\left(-\tfrac{1}{8\zeta}\eta^2m\right) \\
&\overset{(e)}{\leq} \exp\left(-\tfrac{1}{8\zeta}\eta^2\kappa\log T\right) \\
&= T^{-\frac{1}{8\zeta}\eta^2\kappa} \\
&\overset{(f)}{\leq} T^{-3/2}, \qquad \text{(C.5)}
\end{aligned}$$

where: (c) follows from Markov's inequality, (d) follows because $y^2 = \frac{1}{4}\eta^2$ and $Z_t^y$ is a supermartingale with $Z_0^y = 1$, (e) follows because $m = \lceil\kappa\log T\rceil \geq \kappa\log T$, and (f) follows because $\kappa \geq \frac{6}{\eta x_0} \vee \frac{12\nu_0\sigma^2}{\eta^2} = 12\zeta\eta^{-2}$. Similarly, choosing $y = -\frac{1}{2}\eta$, we deduce by the argument used for deriving (C.5) that $\mathbb{P}_{\boldsymbol{\theta}}^{\pi}\{S_m \leq -\frac{1}{2}\eta m\} \leq T^{-3/2}$. Therefore, $\mathbb{P}_{\boldsymbol{\theta}}^{\pi}\{\left|\bar{\epsilon}_{10}\right| \geq \frac{1}{2}\eta\} = \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\{\left|S_m\right| \geq \frac{1}{2}\eta m\} \leq 2T^{-3/2}$. Because the experimentation sets $\mathcal{X}_{ik}$ are disjoint and $\{\epsilon_t\}$ are independent and identically distributed random variables, $\bar{\epsilon}_{10}$ has the same distribution as $\bar{\epsilon}_{ik}$ for all $i$ and $k$. Therefore, by the above argument, we have $\mathbb{P}^{\pi}_{\boldsymbol{\theta}}\{|\bar{\epsilon}_{ik}| \geq \frac{1}{2}\eta\} \leq 2T^{-3/2}$ for all $i$ and $k$. ■

**Proof of Lemma 4.** Define

$$A_j \ := \ \bigcup_{k=L(\tau_j^*)}^{\tau_j^*} \left\{\theta_s = \theta_t \neq \theta_{(\tau_j^*+1)n+1} \text{ for all } s,t \in \mathcal{X}_{1k} \cup \mathcal{X}_{2k}\right\}, \tag{C.6}$$

the event that there is at least one cycle $k$ between $L(\tau_j^*)$ and $\tau_j^*$ such that there is no change-point in $\mathcal{X}_{1k} \cup \mathcal{X}_{2k}$, and the value of the demand parameter vector during the periods in $\mathcal{X}_{1k} \cup \mathcal{X}_{2k}$ is different than the one after the $j^{\text{th}}$ change-point. We first calculate the loss due to detection delay on $A_j$. Let $\mathcal{D}_j = \hat{\tau}_j^+ - \tau_j^*$ be the delay of the true detection following the $j^{\text{th}}$ change-point. Assuming $\tau_j^* < \tau_{j+1}^* - 2$, we have

$$\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\left\{ \sum_{s=n\tau_j^*+1}^{n\hat{\tau}_j^+} \left(1 - \frac{r(p_s,\theta_s)}{r^*(\theta_s)}\right) \mathbb{I}_{\{s\notin\mathcal{X}\}\cap A_j} \right\} \ \leq \ n\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\{\mathcal{D}_j\, \mathbb{I}_{A_j}\}. \tag{C.7}$$

Note that $\mathcal{D}_0 = 0$ by definition, and for $j = 1, 2, \ldots, \mathcal{C}$, the expected value of $\mathcal{D}_j\,\mathbb{I}_{A_j}$ can be bounded above as follows:

$$\begin{aligned}
\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\{\mathcal{D}_j\,\mathbb{I}_{A_j}\} \ &= \ \sum_{d=1}^{\tau_{j+1}^*-\tau_j^*} \mathbb{P}^{\pi}_{\boldsymbol{\theta}}\{\mathcal{D}_j \geq d\,,\, A_j\} \\
&\overset{\text{(a)}}{\leq} \ 2 \ + \sum_{d=3}^{\tau_{j+1}^*-\tau_j^*} \mathbb{P}^{\pi}_{\boldsymbol{\theta}}\{\mathcal{D}_j \geq d\,,\, A_j\} \\
&\overset{\text{(b)}}{\leq} \ 2 \ + \sum_{d=3}^{\tau_{j+1}^*-\tau_j^*} \mathbb{P}^{\pi}_{\boldsymbol{\theta}}\{\chi_{\tau_j^*+d-1} = 0\,,\, A_j\},
\end{aligned} \tag{C.8}$$

where: (a) follows because $\mathbb{P}^{\pi}_{\boldsymbol{\theta}}(\cdot)$ is a probability measure, and (b) follows because $\mathcal{D}_j \geq d \geq 3$ implies that there was no detection in cycle $\tau_j^* + d - 2$. On $A_j$, the probability of no detection in cycle $k^* = \tau_j^* + 1, \ldots, \tau_{j+1}^* - 2$ is

$$\mathbb{P}^{\pi}_{\boldsymbol{\theta}}\{\chi_{k^*+1} = 0\,,\, A_j\} \ = \ \mathbb{P}^{\pi}_{\boldsymbol{\theta}}\Big\{\sup\nolimits_{i,k}\{|\bar{D}_{ik^*} - \bar{D}_{ik}| : L(\tau_j^*) \leq k < k^*\} \leq \eta\,,\, A_j\Big\}. \tag{C.9}$$

Note that, on the event $A_j$, there exists a cycle $k_0 = L(\tau_j^*), \ldots, \tau_j^*$ such that for all $s,t \in \mathcal{X}_{1k_0} \cup \mathcal{X}_{2k_0}$ we have $\theta_s = \theta_t \neq \theta_{n(\tau_j^*+1)+1}$. Let $y_0 := \theta_{nk_0+1}$ and $y^* := \theta_{n(\tau_j^*+1)+1}$. Then, by the preceding identity, we have the following for $k^* = \tau_j^* + 1, \ldots, \tau_{j+1}^* - 2$:

$$
\begin{aligned}
\mathbb{P}_{\boldsymbol{\theta}}^{\pi}\big\{\chi_{k^*+1} = 0\,,\, A_j\big\} \;&\leq\; \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\big\{\big|\bar{D}_{ik^*} - \bar{D}_{ik_0}\big| \leq \eta \text{ for } i = 1,2\,,\, A_j\big\} \\
&=\; \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\left\{\frac{1}{m}\left|\sum_{t\in\mathcal{X}_{ik^*}} D_t \;-\; \sum_{s\in\mathcal{X}_{ik_0}} D_s\right| \leq \eta \text{ for } i = 1,2\,,\, A_j\right\} \\
&\overset{\text{(c)}}{=}\; \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\left\{\frac{1}{m}\left|\sum_{t\in\mathcal{X}_{ik^*}} (\tilde{X}_i\cdot\theta_t + \epsilon_t) \;-\; \sum_{s\in\mathcal{X}_{ik_0}} (\tilde{X}_i\cdot\theta_s + \epsilon_s)\right| \leq \eta \text{ for } i = 1,2\,,\, A_j\right\} \\
&\overset{\text{(d)}}{=}\; \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\left\{\frac{1}{m}\left|\sum_{t\in\mathcal{X}_{ik^*}} (\tilde{X}_i\cdot y^* + \epsilon_t) \;-\; \sum_{s\in\mathcal{X}_{ik_0}} (\tilde{X}_i\cdot y_0 + \epsilon_s)\right| \leq \eta \text{ for } i = 1,2\,,\, A_j\right\} \\
&=\; \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\left\{\frac{1}{m}\left|m\,\tilde{X}_i\cdot(y_0 - y^*) + \sum_{t\in\mathcal{X}_{ik^*}} \epsilon_t \;-\; \sum_{s\in\mathcal{X}_{ik_0}} \epsilon_s\right| \leq \eta \text{ for } i = 1,2\,,\, A_j\right\} \\
&\overset{\text{(e)}}{\leq}\; \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\left\{\frac{1}{m}\left|\sum_{t\in\mathcal{X}_{ik^*}} \epsilon_t \;-\; \sum_{s\in\mathcal{X}_{ik_0}} \epsilon_s\right| \geq \big|\tilde{X}_i\cdot(y_0 - y^*)\big| - \eta \text{ for } i = 1,2\,,\, A_j\right\} \\
&=\; \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\Big\{\big|\bar{\epsilon}_{ik^*} - \bar{\epsilon}_{ik_0}\big| \geq \big|\tilde{X}_i\cdot(y_0 - y^*)\big| - \eta \text{ for } i = 1,2\,,\, A_j\Big\}, \qquad \text{(C.10)}
\end{aligned}
$$

where: $\tilde{X}_i := \big[\begin{smallmatrix}1\\ x_i\end{smallmatrix}\big]$ and $\bar{\epsilon}_{ik} = m^{-1}\sum_{t\in\mathcal{X}_{ik}} \epsilon_t$ for all $i,k$, (c) follows by (3.3) and the fact that $p_t = x_i$ for all $t \in \mathcal{X}_{ik}$, (d) follows because $\theta_t = y^*$ for all $t \in \mathcal{X}_{ik^*}$ and $\theta_s = y_0$ for all $s \in \mathcal{X}_{ik_0}$, and (e) follows by triangle inequality. Because $y_0 \neq y^*$, we know by condition (4.1) that $\|y_0 - y^*\| \geq \delta$. By elementary algebra, this implies that $\big|\tilde{X}_i\cdot(y_0 - y^*)\big| \geq a\delta$ for some $i_0 = 1,2$, where $a = \frac{|x_1 - x_2|}{4(1\vee x_1\vee x_2)}$. Recalling that $\eta = \frac{|x_1 - x_2|}{8(1\vee x_1\vee x_2)}\,\delta = \frac{1}{2}a\delta$, we have

$$
\begin{aligned}
\mathbb{P}_{\boldsymbol{\theta}}^{\pi}\big\{\chi_{k^*+1} = 0\,,\, A_j\big\} \;&\leq\; \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\big\{\big|\bar{\epsilon}_{i_0k^*} - \bar{\epsilon}_{i_0k_0}\big| \geq \tfrac{1}{2}a\delta\,,\, A_j\big\} \\
&=\; \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\big\{\big|\bar{\epsilon}_{i_0k^*} - \bar{\epsilon}_{i_0k_0}\big| \geq \eta\,,\, A_j\big\} \\
&\leq\; \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\big\{\big|\bar{\epsilon}_{i_0k^*}\big| \geq \tfrac{1}{2}\eta\,,\, A_j\big\} + \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\big\{\big|\bar{\epsilon}_{i_0k_0}\big| \geq \tfrac{1}{2}\eta\,,\, A_j\big\} \\
&\overset{\text{(f)}}{\leq}\; 2\mathbb{P}_{\boldsymbol{\theta}}^{\pi}\Big\{\textstyle\bigcup_{i,k}\big\{\big|\bar{\epsilon}_{ik}\big| \geq \tfrac{1}{2}\eta\big\}\,,\, A_j\Big\} \\
&\leq\; 2\mathbb{P}_{\boldsymbol{\theta}}^{\pi}\Big\{\textstyle\bigcup_{i,k}\big\{\big|\bar{\epsilon}_{ik}\big| \geq \tfrac{1}{2}\eta\big\}\Big\} \\
&\overset{\text{(g)}}{\leq}\; 2\textstyle\sum_{i,k}\mathbb{P}_{\boldsymbol{\theta}}^{\pi}\big\{\big|\bar{\epsilon}_{ik}\big| \geq \tfrac{1}{2}\eta\big\}, \qquad \text{(C.11)}
\end{aligned}
$$

for $k^* = \tau_j^* + 1, \ldots, \tau_{j+1}^* - 2$, where: (f) follows because for any given $i$ and $k$, $\big|\bar{\epsilon}_{ik}\big| \geq \frac{1}{4}a\delta$ implies $\bigcup_{i,k}\big\{\big|\bar{\epsilon}_{ik}\big| \geq \frac{1}{4}a\delta\big\}$, and (g) follows by the union bound. Using the bound in Lemma 3 on the right hand side of (C.11), we get

$$\mathbb{P}^{\pi}_{\boldsymbol{\theta}}\big\{\chi_{k^*+1} = 0\,,\ A_j\big\} \ \le\ 4\textstyle\sum_{i,k} T^{-3/2} \ \le\ 8T^{-3/2}\lfloor T/n\rfloor \ \le\ 8n^{-1}T^{-1/2}\,, \tag{C.12}$$

for $k^* = \tau_j^* + 1, \ldots, \tau_{j+1}^* - 2$. Combining (C.8) and (C.12), we deduce that

$$\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\big\{\mathcal{D}_j\,\mathbb{I}_{A_j}\big\} \ \le\ 2+8\sum_{d=3}^{\tau_{j+1}^*-\tau_j^*} n^{-1}T^{-1/2} \ \le\ 2+8n^{-1}T^{-1/2}\lfloor T/n\rfloor \ \le\ 2+8n^{-2}T^{1/2} \ \le\ 2+8\kappa^{-2}T^{-1/2}.$$

Recalling (C.7), we conclude that

$$\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\Bigg\{\sum_{s=n\tau_j^*+1}^{n\hat{\tau}_j^+}\left(1-\frac{r(p_s,\theta_s)}{r^*(\theta_s)}\right)\mathbb{I}_{\{s\notin\mathcal{X}\}\cap A_j}\Bigg\} \ \le\ (2+8\kappa^{-2}T^{-1/2})n \ \le\ 4\kappa\sqrt{T}+16\kappa^{-1}. \tag{C.13}$$

Now we find an upper bound on loss due to detection delay on the event $A_j^c$. Assuming $\tau_j^* < \tau_{j+1}^*-2$, we have the following on $A_j^c$:

$$\begin{aligned}
&\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\Bigg\{\sum_{s=n\tau_j^*+1}^{n\hat{\tau}_j^+}\left(1-\frac{r(p_s,\theta_s)}{r^*(\theta_s)}\right)\mathbb{I}_{\{s\notin\mathcal{X}\}\cap A_j^c}\Bigg\}\\
&\quad\le\ n+\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\Bigg\{\sum_{s=n(\tau_j^*+1)+1}^{n\tau_{j+1}^*}\left(1-\frac{r(p_s,\theta_s)}{r^*(\theta_s)}\right)\mathbb{I}_{\{s\notin\mathcal{X},\,s\le n\hat{\tau}_j^+\}\cap A_j^c}\Bigg\}\\
&\quad\overset{(\mathrm{a}')}{\le}\ n+c_1\sum_{s=n(\tau_j^*+1)+1}^{n\tau_{j+1}^*}\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\Big\{\big(\varphi(\theta_s)-p_s\big)^2\,\mathbb{I}_{\{s\notin\mathcal{X},\,s\le n\hat{\tau}_j^+\}\cap A_j^c}\Big\}\\
&\quad\le\ n+c_2\sum_{s=n(\tau_j^*+1)+1}^{n\tau_{j+1}^*}\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\Big\{\|\theta_s-\hat{\theta}_s\|^2\,\mathbb{I}_{\{s\notin\mathcal{X},\,s\le n\hat{\tau}_j^+\}\cap A_j^c}\Big\}\\
&\quad=\ n+c_2\sum_{s=n(\tau_j^*+1)}^{n\tau_{j+1}^*-1}\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\Big\{\|\theta_{s+1}-\hat{\theta}_{s+1}\|^2\,\mathbb{I}_{\{s+1\notin\mathcal{X},\,s<n\hat{\tau}_j^+\}\cap A_j^c}\Big\},
\end{aligned} \tag{C.14}$$

where: $c_1 = \max_{(\alpha,\beta)\in\Theta}\{4\beta^2/\alpha^2\}$, $c_2 = c_1\max_{i=1,2}\big\{\max_\theta\{(\partial\varphi(\theta)/\partial\theta_i)^2\}\big\}$, and (a$'$) follows by definitions of $r(\cdot,\cdot)$, $r^*(\cdot)$, and $\varphi(\cdot)$. By (3.4), we have

$$\hat{\theta}_{s+1}-\theta_{s+1} \ =\ \big(\mathcal{J}_s^s\big)^{-1}\sum_{q=1}^{s} w_q^s X_q X_q^{\mathsf{T}}(\theta_q-\theta_{s+1}) \ +\ \big(\mathcal{J}_s^s\big)^{-1}\mathcal{M}_s^s \quad \text{for all } s. \tag{C.15}$$

The preceding identity implies the following for $s = n(\tau_j^*+1), \ldots, n\tau_{j+1}^* - 1$ satisfying $s+1 \notin \mathcal{X}$:

$$\|\hat{\theta}_{s+1} - \theta_{s+1}\| \overset{(\mathrm{b}')}{\leq} \big(\mathcal{I}_s^s\big)^{-1} \sum_{q=1}^{s} w_q^s \,\|\theta_q - \theta_{s+1}\| \;+\; \big\|\big(\mathcal{J}_s^s\big)^{-1}\mathcal{M}_s^s\big\|$$

$$\overset{(\mathrm{c}')}{\leq} \big(\mathcal{I}_s^s\big)^{-1} \sum_{q=nL(\tau_j^*)+1}^{n(\tau_j^*+1)} w_q^s \,\|\theta_q - \theta_{s+1}\| \;+\; \big\|\big(\mathcal{J}_s^s\big)^{-1}\mathcal{M}_s^s\big\|, \qquad \text{(C.16)}$$

where: (b′) follows by triangle inequality and the fact that the eigenvalues of $\big(\mathcal{J}_s^s\big)^{-1} X_q X_q^{\mathsf{T}}$ are 0 and $\pm\big(\mathcal{I}_s^s\big)^{-1}$ for all $q \in \mathcal{X}$, and (c′) follows because $w_q^s = 0$ for $q \leq nL(\tau_j^*) \leq n(\tau_j^*+1) \leq s$ and $\|\theta_q - \theta_{s+1}\| = 0$ for $(\tau_j^*+1)n + 1 \leq q \leq s \leq \tau_{j+1}^* n - 1$. By the definition of $A_j$ in (C.6) we have the following on $A_j^c$: for all cycles $k$ between $L(\tau_j^*)$ and $\tau_j^*$, either (i) there is a change-point in $\mathcal{X}_{1k} \cup \mathcal{X}_{2k}$, or (ii) the value of the demand parameter vector during the periods in $\mathcal{X}_{1k} \cup \mathcal{X}_{2k}$ is exactly the same as the one after the $j^{\text{th}}$ change-point. Letting $\mathcal{K}_j^{(\mathrm{i})}$ and $\mathcal{K}_j^{(\mathrm{ii})}$ be the sets of cycles between $L(\tau_j^*)$ and $\tau_j^*$ that satisfy conditions (i) and (ii), respectively, we re-express (C.16) as follows:

$$\|\hat{\theta}_{s+1} - \theta_{s+1}\| \;\leq\; \big(\mathcal{I}_s^s\big)^{-1} \Bigg( \sum_{k \in \mathcal{K}_j^{(\mathrm{i})}} + \sum_{k \in \mathcal{K}_j^{(\mathrm{ii})}} \Bigg) \sum_{q=nk+1}^{n(k+1)} w_q^s \,\|\theta_q - \theta_{s+1}\| \;+\; \big\|\big(\mathcal{J}_s^s\big)^{-1}\mathcal{M}_s^s\big\|$$

$$\overset{(\mathrm{d}')}{\leq} \big(\mathcal{I}_s^s\big)^{-1} \sum_{k \in \mathcal{K}_j^{(\mathrm{i})}} \sum_{q=nk+1}^{n(k+1)} w_q^s \,\|\theta_q - \theta_{s+1}\| \;+\; \big\|\big(\mathcal{J}_s^s\big)^{-1}\mathcal{M}_s^s\big\|$$

$$\overset{(\mathrm{e}')}{\leq} \big(\mathcal{I}_s^s\big)^{-1} c_3 \sum_{k \in \mathcal{K}_j^{(\mathrm{i})}} \sum_{q=nk+1}^{n(k+1)} w_q^s \;+\; \big\|\big(\mathcal{J}_s^s\big)^{-1}\mathcal{M}_s^s\big\|$$

$$\overset{(\mathrm{f}')}{\leq} \big(\mathcal{I}_s^s\big)^{-1} c_3 \,\mathcal{C} m \;+\; \big\|\big(\mathcal{J}_s^s\big)^{-1}\mathcal{M}_s^s\big\|, \qquad \text{(C.17)}$$

for $s = n(\tau_j^*+1), \ldots, n\tau_{j+1}^* - 1$ satisfying $s+1 \notin \mathcal{X}$, where: $c_3 = \max_{\theta,\theta' \in \Theta} \|\theta - \theta'\|$, (d′) follows because given $k \in \mathcal{K}_j^{(\mathrm{ii})}$, we have $w_q^s \|\theta_q - \theta_{s+1}\| = 0$ for $q = nk+1, \ldots, n(k+1)$, (e′) follows because $\|\theta_q - \theta_{s+1}\| \leq c_3$, and (f′) follows because $w_q^s = 0$ if $s \notin \mathcal{X}$, and the cardinality of $\mathcal{K}_j^{(\mathrm{i})}$ is less than or equal to the number of change-points, $\mathcal{C}$. Squaring and taking the expectation of both sides of (C.17), we get

$$\mathbb{E}_{\boldsymbol{\theta}}^{\pi}\Big\{ \|\hat{\theta}_{s+1} - \theta_{s+1}\|^2 \, \mathbb{I}_{\{s+1 \notin \mathcal{X},\, s < n\hat{\tau}_j^+\} \cap A_j^c} \Big\}$$

$$\leq \mathbb{E}_{\boldsymbol{\theta}}^{\pi}\Big\{ \Big( 2\big(\mathcal{I}_s^s\big)^{-2} c_3^2 \,\mathcal{C}^2 m^2 + 2\big\|\big(\mathcal{J}_s^s\big)^{-1}\mathcal{M}_s^s\big\|^2 \Big) \, \mathbb{I}_{\{s+1 \notin \mathcal{X},\, s < n\hat{\tau}_j^+\} \cap A_j^c} \Big\}. \qquad \text{(C.18)}$$

On $A_j^c$, we know that $\mathcal{I}_s^s = 2\big(\lceil s/n \rceil - L(\tau_j^*) + 1\big)m \geq 2\big(\lceil s/n \rceil - \tau_j^* + 1\big)m$ for $s = n(\tau_j^*+1), \ldots, n\hat{\tau}_j^+ - 1$

satisfying $s+1 \notin \mathcal{X}$. Thus, (C.18) implies that

$$\begin{aligned}
&\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\Big\{\|\hat{\theta}_{s+1}-\theta_{s+1}\|^2\,\mathbb{I}_{\{s+1\notin\mathcal{X},\,s<n\hat{\tau}_j^+\}\cap A_j^c}\Big\}\\
&\qquad\le\ \frac{c_3^2\,\mathcal{C}^2}{2\big(\lceil s/n\rceil-\tau_j^*+1\big)^2}\ +\ 2\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\Big\{\big\|\big(\mathcal{J}_s^s\big)^{-1}\mathcal{M}_s^s\big\|^2\,\mathbb{I}_{\{s+1\notin\mathcal{X},\,s<n\hat{\tau}_j^+\}\cap A_j^c}\Big\}\\
&\qquad\overset{(\mathrm{g}')}{\le}\ \frac{c_3^2\,\mathcal{C}^2}{2\big(\lceil s/n\rceil-\tau_j^*+1\big)^2}\ +\ \frac{12}{\rho\big(\lceil s/n\rceil-\tau_j^*+1\big)m}\,,
\end{aligned}\tag{C.19}$$

where (g′) follows by the arguments used to prove inequality (B.10) and Lemma 2. Summing both sides of (C.19) over $s=n(\tau_j^*+1),\ldots,n\tau_{j+1}^*-1$, we deduce that

$$\begin{aligned}
&\sum_{s=n(\tau_j^*+1)}^{n\tau_{j+1}^*-1}\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\Big\{\|\theta_{s+1}-\hat{\theta}_{s+1}\|^2\,\mathbb{I}_{\{s+1\notin\mathcal{X},\,s\le\hat{\tau}_j^+n\}\cap A_j^c}\Big\}\\
&\qquad\le\ \sum_{s=n(\tau_j^*+1)}^{n\tau_{j+1}^*-1}\left(\frac{c_3^2\,\mathcal{C}^2}{2\big(\lceil s/n\rceil-\tau_j^*+1\big)^2}\ +\ \frac{12}{\rho\big(\lceil s/n\rceil-\tau_j^*+1\big)m}\right)\\
&\qquad\overset{(\mathrm{h}')}{\le}\ n\sum_{q=2}^{\tau_{j+1}^*-\tau_j^*+1}\left(\frac{c_3^2\,\mathcal{C}^2}{2q^2}\ +\ \frac{12}{\rho q m}\right)\\
&\qquad\le\ n\left(\frac{c_3^2\,\mathcal{C}^2\,\pi^2}{12}\ +\ \frac{12}{\rho m}\log(\tau_{j+1}^*-\tau_j^*+1)\right)\\
&\qquad\overset{(\mathrm{i}')}{\le}\ n\left(\frac{c_3^2\,\mathcal{C}^2\,\pi^2}{12}\ +\ \frac{6}{\rho\kappa}\right)\\
&\qquad\overset{(\mathrm{j}')}{\le}\ \left(\frac{c_3^2\,\mathcal{C}^2\,\pi^2\kappa}{6}\ +\ \frac{12}{\rho}\right)\sqrt{T},
\end{aligned}\tag{C.20}$$

for $T\ge 3$, where: (h′) follows by expressing the time index as $s=(\tau_j^*+q-1)n+i$, (i′) follows because $m\ge\kappa\log T\ge 2\kappa\log(\tau_{j+1}^*-\tau_j^*+1)$ for $T\ge 3$, and (j′) follows because $n\le 2\kappa\sqrt{T}$. By inequalities (C.14) and (C.20), we deduce that $\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\big\{\sum_{s=n\tau_j^*+1}^{n\hat{\tau}_j^+}\big(1-r(p_s,\theta_s)/r^*(\theta_s)\big)\mathbb{I}_{\{s\notin\mathcal{X}\}\cap A_j^c}\big\}\le c_4\sqrt{T}$, where $c_4=2\kappa+c_2\big(c_3^2\mathcal{C}^2\pi^2\kappa/6+12/\rho\big)$. Combining this result with (C.13), we conclude that

$$\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\left\{\sum_{s=n\tau_j^*+1}^{n\hat{\tau}_j^+}\left(1-\frac{r(p_s,\theta_s)}{r^*(\theta_s)}\right)\mathbb{I}_{\{s\notin\mathcal{X}\}}\right\}\le C_1\sqrt{T},\tag{C.21}$$

where $C_1=4(\kappa+4\kappa^{-1})\vee c_4$. ■

**Proof of Lemma 5.** Assume that $\tau_j^*<\tau_{j+1}^*-2$, and that there exists at least one false detection between the $j^{\text{th}}$ and $(j+1)^{\text{st}}$ change-points. Let $\mathcal{E}_j=\tau_{j+1}^*-\hat{\tau}_j^-$ be the earliness of the first false detection after the $j^{\text{th}}$ change-point. Then, we have

$$\mathbb{E}_{\boldsymbol{\theta}}^{\pi}\left\{\sum_{s=n\hat{\tau}_j^-+1}^{n\tau_{j+1}^*}\left(1-\frac{r(p_s,\theta_s)}{r^*(\theta_s)}\right)\mathbb{I}_{\{s\notin\mathcal{X}\}}\right\} \;\le\; n\big(1+\mathbb{E}_{\boldsymbol{\theta}}^{\pi}\{\mathcal{E}_j\}\big). \tag{C.22}$$

For $j=0,1,\ldots,\mathcal{C}$, the expected earliness of false detections before the $(j+1)^{\text{st}}$ change-point is given by

$$\mathbb{E}_{\boldsymbol{\theta}}^{\pi}\{\mathcal{E}_j\} \;=\; \sum_{\varepsilon=1}^{\tau_{j+1}^*-\tau_j^*-2}\mathbb{P}_{\boldsymbol{\theta}}^{\pi}\{\mathcal{E}_j\ge\varepsilon\} \;=\; \sum_{\varepsilon=1}^{\tau_{j+1}^*-\tau_j^*-2}\;\sum_{q=\varepsilon}^{\tau_{j+1}^*-\tau_j^*-2}\mathbb{P}_{\boldsymbol{\theta}}^{\pi}\{\mathcal{E}_j=q\}. \tag{C.23}$$

By definition, $\tau_j^*<\hat{\tau}_j^+<\hat{\tau}_j^-\le\tau_{j+1}^*$, i.e., if there is a false detection between the $j^{\text{th}}$ and $(j+1)^{\text{st}}$ change-points then it must be preceded by the true detection after the $j^{\text{th}}$ change-point. Therefore, for all $q=1,\ldots,\tau_{j+1}^*-\tau_j^*-2$, the event $\mathcal{E}_j=q$ implies that the true detection after the $j^{\text{th}}$ change-point is between cycles $\tau_j^*+1$ and $\tau_{j+1}^*-q-2$, and that there is a false detection in cycle $\tau_{j+1}^*-q-1$. More formally, $\{\mathcal{E}_j=q\}\subseteq\big\{L(\tau_{j+1}^*-q-1)\ge\tau_j^*+1\big\}\cap\big\{\chi_{\tau_{j+1}^*-q}=1\big\}$, where $L(k)=\max\{\tau\le k:\chi_\tau=1\}$ is the latest detection cycle that precedes cycle $k$. Letting $B_{jq}:=\big\{L(\tau_{j+1}^*-q-1)\ge\tau_j^*+1\big\}$, we have the following by (C.23):

$$\mathbb{E}_{\boldsymbol{\theta}}^{\pi}\{\mathcal{E}_j\} \;\le\; \sum_{\varepsilon=1}^{\tau_{j+1}^*-\tau_j^*-2}\;\sum_{q=\varepsilon}^{\tau_{j+1}^*-\tau_j^*-2}\mathbb{P}_{\boldsymbol{\theta}}^{\pi}\big\{\chi_{\tau_{j+1}^*-q}=1\,,\,B_{jq}\big\}. \tag{C.24}$$

For $k^*=\tau_{j+1}^*-q-1$, the definition of the detection test (4.6) implies

$$\begin{aligned}
\mathbb{P}_{\boldsymbol{\theta}}^{\pi}\big\{\chi_{k^*+1}=1\,,\,B_{jq}\big\} \;&=\; \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\Big\{\sup\nolimits_{i,k}\big\{\big|\bar{D}_{ik^*}-\bar{D}_{ik}\big|:L(k^*)\le k<k^*\big\}>\eta\,,\,B_{jq}\Big\}\\
&\overset{\text{(a)}}{\le}\; \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\Big\{\sup\nolimits_{i,k}\big\{\big|\bar{D}_{ik^*}-\bar{D}_{ik}\big|:\tau_j^*+1\le k<k^*\big\}>\eta\,,\,B_{jq}\Big\}\\
&=\; \mathbb{P}_{\boldsymbol{\theta}}^{\pi}\left\{\bigcup_{i=1}^{2}\;\bigcup_{k=\tau_j^*+1}^{k^*-1}\big\{\big|\bar{D}_{ik^*}-\bar{D}_{ik}\big|>\eta\big\}\,,\,B_{jq}\right\}\\
&\overset{\text{(b)}}{\le}\; \sum_{i=1}^{2}\;\sum_{k=\tau_j^*+1}^{k^*-1}\mathbb{P}_{\boldsymbol{\theta}}^{\pi}\big\{\big|\bar{D}_{ik^*}-\bar{D}_{ik}\big|>\eta\big\},
\end{aligned} \tag{C.25}$$

where: (a) follows because $L(k^*)=L(\tau_{j+1}^*-q-1)\ge\tau_j^*+1$ on $B_{jq}$, and (b) follows by the union bound. Note that there are no change-points between cycles $\tau_j^*+1$ and $k^*=\tau_{j+1}^*-q-1$. Letting

$y^* := \theta_{n(\tau_j^*+1)+1}$, we therefore have

$$\bar{D}_{ik^*} - \bar{D}_{ik} \;=\; \frac{1}{m} \sum_{t \in \mathcal{X}_{ik^*}} D_t \;-\; \frac{1}{m} \sum_{s \in \mathcal{X}_{ik}} D_s$$

$$= \; \frac{1}{m} \sum_{t \in \mathcal{X}_{ik^*}} (\tilde{X}_i \cdot y^* + \epsilon_t) \;-\; \frac{1}{m} \sum_{s \in \mathcal{X}_{ik}} (\tilde{X}_i \cdot y^* + \epsilon_s)$$

$$= \; \bar{\epsilon}_{ik^*} - \bar{\epsilon}_{ik}, \tag{C.26}$$

for all $k = \tau_j^* + 1, \ldots, k^*$, where: $\tilde{X}_i = \left[\begin{smallmatrix} 1 \\ x_i \end{smallmatrix}\right]$ and $\bar{\epsilon}_{ik} = m^{-1} \sum_{t \in \mathcal{X}_{ik}} \epsilon_t$ for all $i, k$. Thus, (C.25) implies

$$\mathbb{P}^{\pi}_{\boldsymbol{\theta}}\big\{\chi_{k^*+1} = 1\,,\, B_{jq}\big\} \;\leq\; \sum_{i=1}^{2} \sum_{k=\tau_j^*+1}^{k^*-1} \mathbb{P}^{\pi}_{\boldsymbol{\theta}}\big\{\big|\bar{\epsilon}_{ik^*} - \bar{\epsilon}_{ik}\big| > \eta\big\}$$

$$\leq \; \sum_{i=1}^{2} \sum_{k=\tau_j^*+1}^{k^*-1} \Big(\mathbb{P}^{\pi}_{\boldsymbol{\theta}}\big\{\big|\bar{\epsilon}_{ik^*}\big| > \tfrac{1}{2}\eta\big\} + \mathbb{P}^{\pi}_{\boldsymbol{\theta}}\big\{\big|\bar{\epsilon}_{ik}\big| > \tfrac{1}{2}\eta\big\}\Big)$$

$$\overset{\text{(c)}}{\leq} \; 4 \sum_{i=1}^{2} \sum_{k=\tau_j^*+1}^{k^*-1} T^{-3/2}$$

$$\leq \; 8T^{-3/2} \left\lfloor \frac{T}{n} \right\rfloor$$

$$\leq \; 8n^{-1}T^{-1/2}, \tag{C.27}$$

for $k^* = \tau_{j+1}^* - q - 1$, where (c) follows by Lemma 3. Combining (C.24) and (C.27), we get

$$\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\big\{\mathcal{E}_j\big\} \;\leq\; 2 + 8 \sum_{\varepsilon=1}^{\tau_{j+1}^* - \tau_j^* - 2} \; \sum_{q=\varepsilon}^{\tau_{j+1}^* - \tau_j^* - 2} n^{-1}T^{-1/2}$$

$$\leq \; 2 + 8n^{-1}T^{-1/2} \left\lfloor \frac{T}{n} \right\rfloor^2$$

$$\leq \; 2 + 8n^{-3}T^{3/2}$$

$$\leq \; 2 + 8\kappa^{-3}. \tag{C.28}$$

Recalling (C.22), we conclude that

$$\mathbb{E}^{\pi}_{\boldsymbol{\theta}}\Bigg\{ \sum_{s=n\hat{\tau}_j^-+1}^{n\tau_{j+1}^*} \left(1 - \frac{r(p_s, \theta_s)}{r^*(\theta_s)}\right) \mathbb{I}_{\{s \notin \mathcal{X}\}} \Bigg\} \;\leq\; (3 + 8\kappa^{-3})n \;\leq\; C_2\sqrt{T}, \tag{C.29}$$

where $C_2 = 6\kappa + 16\kappa^{-2}$. ∎

**Proof of Lemma 6.** Assuming $\tau_j^* < \tau_{j+1}^* - 2$, we have

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}}^{\pi}\Bigg\{\sum_{s=n\hat{\tau}_j^+ +1}^{n\hat{\tau}_j^-}\left(1-\frac{r(p_s,\theta_s)}{r^*(\theta_s)}\right)\mathbb{I}_{\{s\notin\mathcal{X}\}}\Bigg\} &\overset{(a)}{\le} c_1\sum_{s=n(\tau_j^*+1)+1}^{n\tau_{j+1}^*}\mathbb{E}_{\boldsymbol{\theta}}^{\pi}\Big\{\big(\varphi(\theta_s)-p_s\big)^2\,\mathbb{I}_{\{s\notin\mathcal{X},\,n\hat{\tau}_j^+<s\le n\hat{\tau}_j^-\}}\Big\} \\
&\le c_2\sum_{s=n(\tau_j^*+1)+1}^{n\tau_{j+1}^*}\mathbb{E}_{\boldsymbol{\theta}}^{\pi}\Big\{\|\theta_s-\hat{\theta}_s\|^2\,\mathbb{I}_{\{s\notin\mathcal{X},\,n\hat{\tau}_j^+<s\le n\hat{\tau}_j^-\}}\Big\} \\
&= c_2\sum_{s=n(\tau_j^*+1)}^{n\tau_{j+1}^*-1}\mathbb{E}_{\boldsymbol{\theta}}^{\pi}\Big\{\|\theta_{s+1}-\hat{\theta}_{s+1}\|^2\,\mathbb{I}_{\{s+1\notin\mathcal{X},\,n\hat{\tau}_j^+\le s<n\hat{\tau}_j^-\}}\Big\},
\end{aligned} \tag{C.30}$$

where: $c_1 = \max_{(\alpha,\beta)\in\Theta}\{4\beta^2/\alpha^2\}$, $c_2 = c_1\max_{i=1,2}\big\{\max_\theta\{(\partial\varphi(\theta)/\partial\theta_i)^2\}\big\}$, and (a) follows by definitions of $r(\cdot,\cdot)$, $r^*(\cdot)$, and $\varphi(\cdot)$. Recalling (3.4), we have

$$\hat{\theta}_{s+1}-\theta_{s+1} \;=\; \big(\mathcal{J}_s^s\big)^{-1}\sum_{q=1}^{s} w_q^s X_q X_q^{\mathsf{T}}(\theta_q-\theta_{s+1}) \;+\; \big(\mathcal{J}_s^s\big)^{-1}\mathcal{M}_s^s \quad \text{for all } s. \tag{C.31}$$

For any given $s = n\hat{\tau}_j^+,\ldots,n\hat{\tau}_j^- - 1$ satisfying $s+1\notin\mathcal{X}$, we know that $w_q^s = 0$ for $1\le q\le n\hat{\tau}_j^+\le s$, and that $\theta_q - \theta_{s+1} = 0$ for $n(\tau_j^*+1)\le q\le s\le n\tau_{j+1}^*$. Because $\tau_j^* < \hat{\tau}_j^+ < \hat{\tau}_j^- \le \tau_{j+1}^*$, we deduce that

$$\hat{\theta}_{s+1}-\theta_{s+1} \;=\; \big(\mathcal{J}_s^s\big)^{-1}\mathcal{M}_s^s \quad \text{for } s = n\hat{\tau}_j^+,\ldots,n\hat{\tau}_j^- - 1 \text{ satisfying } s+1\notin\mathcal{X}. \tag{C.32}$$

Note that $\mathcal{I}_s^s = 2\big(\lceil s/n\rceil - L(s/n)+1\big)m = 2\big(\lceil s/n\rceil - \hat{\tau}_j^+ + 1\big)m$ for $s = n\hat{\tau}_j^+,\ldots,n\hat{\tau}_j^- - 1$ satisfying $s+1\notin\mathcal{X}$. Hence, (C.32) implies that

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}}^{\pi}\big\{\|\hat{\theta}_{s+1}-\theta_{s+1}\|^2\,\mathbb{I}_{\{s+1\notin\mathcal{X},\,n\hat{\tau}_j^+\le s<n\hat{\tau}_j^-\}}\big\} &\le \mathbb{E}_{\boldsymbol{\theta}}^{\pi}\Big\{\big\|\big(\mathcal{J}_s^s\big)^{-1}\mathcal{M}_s^s\big\|^2\,\mathbb{I}_{\{s+1\notin\mathcal{X},\,n\hat{\tau}_j^+\le s<n\hat{\tau}_j^-\}}\Big\} \\
&\le \mathbb{E}_{\boldsymbol{\theta}}^{\pi}\Bigg\{\frac{6}{\rho\big(\lceil s/n\rceil-\hat{\tau}_j^+ + 1\big)m}\,\mathbb{I}_{\{s+1\notin\mathcal{X},\,n\hat{\tau}_j^+\le s<n\hat{\tau}_j^-\}}\Bigg\},
\end{aligned} \tag{C.33}$$

by the arguments used to prove inequality (B.10) and Lemma 2. Summing both sides of (C.33) over $s = n(\tau_j^* + 1), \ldots, n\tau_{j+1}^* - 1$, we deduce that

$$
\begin{aligned}
\sum_{s=n(\tau_j^*+1)}^{n\tau_{j+1}^*-1} \mathbb{E}_{\boldsymbol{\theta}}^{\pi}\Big\{\|\theta_{s+1} - \hat{\theta}_{s+1}\|^2 \, \mathbb{I}_{\{s+1 \notin \mathcal{X},\, n\hat{\tau}_j^+ \le s < n\hat{\tau}_j^-\}}\Big\} & \\
&\le \sum_{s=n(\tau_j^*+1)}^{n\tau_{j+1}^*-1} \mathbb{E}_{\boldsymbol{\theta}}^{\pi}\Big\{\frac{6}{\rho\big(\lceil s/n \rceil - \hat{\tau}_j^+ + 1\big)m} \, \mathbb{I}_{\{s+1 \notin \mathcal{X},\, n\hat{\tau}_j^+ \le s < n\hat{\tau}_j^-\}}\Big\} \\
&\le \mathbb{E}_{\boldsymbol{\theta}}^{\pi}\Big\{\sum_{s=n\hat{\tau}_j^+}^{n\hat{\tau}_j^- -1} \frac{6}{\rho\big(\lceil s/n \rceil - \hat{\tau}_j^+ + 1\big)m}\Big\} \\
&\overset{\text{(b)}}{\le} n \sum_{q=2}^{\tau_{j+1}^* - \tau_j^* + 1} \frac{6}{\rho q m} \\
&\le \frac{6n}{\rho m} \log(\tau_{j+1}^* - \tau_j^* + 1) \\
&\overset{\text{(c)}}{\le} 6\rho^{-1}\sqrt{T}, \qquad \text{(C.34)}
\end{aligned}
$$

for $T \ge 3$, where: (b) follows by expressing the time index as $s = (\tau_j^* + q - 1)n + i$ and $\tau_j^* < \hat{\tau}_j^+ < \hat{\tau}_j^- \le \tau_{j+1}^*$, and (c) follows because $m \ge \kappa \log T \ge 2\kappa \log(\tau_{j+1}^* - \tau_j^* + 1)$ for $T \ge 3$, and $n \le 2\kappa\sqrt{T}$. Combining (C.30) and (C.34), we conclude that

$$
\mathbb{E}_{\boldsymbol{\theta}}^{\pi}\Bigg\{\sum_{s=n\hat{\tau}_j^+ + 1}^{n\hat{\tau}_j^-} \bigg(1 - \frac{r(p_s, \theta_s)}{r^*(\theta_s)}\bigg)\mathbb{I}_{\{s \notin \mathcal{X}\}}\Bigg\} \le 6c_2\rho^{-1}\sqrt{T}. \blacksquare \qquad \text{(C.35)}
$$

**Proof of Theorem 4.** By (4.10) and Lemmas 4, 5, and 6, we have

$$
\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}}^{\pi}\Bigg\{\sum_{t=1}^{T} \bigg(1 - \frac{r(p_t, \theta_t)}{r^*(\theta_t)}\bigg)\mathbb{I}_{\{t \notin \mathcal{X}\}}\Bigg\} &\le (\mathcal{C} + 1)(C_1 + C_2 + C_3)T^{1/2} \\
&\le (\bar{C} + 1)(C_1 + C_2 + C_3)T^{1/2}, \qquad \text{(C.36)}
\end{aligned}
$$

for all $T \ge 3$. Therefore, (4.9) implies $\Delta_{\boldsymbol{\theta}}^{\pi}(T) \le CT^{1/2}\log T$ for all $T \ge 3$, where $C = 8 + (\bar{C} + 1)(C_1 + C_2 + C_3)$. $\blacksquare$

## Appendix D: Proof of the Results in Section 5

**Proof of Theorem 5.** In the proof of Theorem 1, let $N = \lceil k_0 T^{2(1-\nu)/3} \rceil$, instead of $N = \lceil k_0 T^{2/3} \rceil$, where $k_0 = 4^{2/3}B^{-2/3}$. Repeating the same arguments from (A.1) to (A.10), deduce that

$\sup\{\Delta^\pi_{\boldsymbol\theta}(T) : V_{\boldsymbol\theta}(T) \le BT^\nu\} \ge \frac{1}{2}k_2 N^{-1/2}T$ for a certain constant $k_2$ independent of $T$, $B$, and $\nu$. We therefore conclude that $\sup\{\Delta^\pi_{\boldsymbol\theta}(T) : V_{\boldsymbol\theta}(T) \le BT^\nu\} \ge cT^{(2+\nu)/3}$ where $c = \frac{1}{8}k_2 B^{1/3}$. ■

**Proof of Lemma 7.** For $M_\nu(\kappa, x_1, x_2)$, using the arguments in the proof of Lemma 1 to obtain (B.4), we get $\sum_{t=n^2}^{T-1} \|(\mathcal{J}_t^t)^{-1}\mathcal{W}_t^t\|^2 \le 4n^2 V_{\boldsymbol\theta}(T)$. Under condition (5.1), this implies that $\sum_{t=n^2}^{T-1} \|(\mathcal{J}_t^t)^{-1}\mathcal{W}_t^t\|^2 \le 4n^2 BT^\nu \le 16\kappa^2 BT^{(2+\nu)/3}$. Letting $c_1 = 16\kappa^2 B$, we get (5.4).

For $W_\nu(\mu, \kappa, x_1, x_2)$, consider (B.17) in the proof of Lemma B.2, which still holds under condition (5.1):

$$\begin{aligned}(\mathcal{J}_t^t)^{-1}\mathcal{W}_t^t \;=\; & \sum_{s\in\mathcal{X}_1} \frac{w_s^t}{(x_1-x_2)\mathcal{I}_t^t}\begin{bmatrix} -x_2 & -x_1x_2 \\ 1 & x_1\end{bmatrix}(\theta_s - \theta_{t+1}) \\ & + \sum_{s\in\mathcal{X}_2} \frac{w_s^t}{(x_1-x_2)\mathcal{I}_t^t}\begin{bmatrix} -x_1 & -x_1x_2 \\ 1 & x_2\end{bmatrix}(\theta_s - \theta_{t+1}).\end{aligned} \tag{D.1}$$

By the arguments used to derive (B.19) and the fact that $w_s^t \le n^{-2}T^{-2\nu}$ for all $s < t - n^2$, we get

$$\begin{aligned}\sum_{t=n^2}^{T-1} \left\|(\mathcal{J}_t^t)^{-1}\mathcal{W}_t^t\right\|^2 \;\le\; & 8n^{-6}T^{-4\nu}\sum_{t=n^2}^{T-1} t^2 \left(\mathcal{I}_t^t\right)^{-2} \max_{1\le s< t-n^2} \left\|\theta_s - \theta_{t+1}\right\|^2 \\ & + 8n^2 \sum_{t=n^2}^{T-1} \left(\mathcal{I}_t^t\right)^{-2} \max_{t-n^2\le s\le t} \left\|\theta_s - \theta_{t+1}\right\|^2.\end{aligned} \tag{D.2}$$

Under $W_\nu(\mu, \kappa, x_1, x_2)$, we have $\mathcal{I}_t^t > c_\mu n$ for all $t \ge n^2$, where $c_\mu$ is a constant independent of $T$, $B$, and $\nu$. Hence, the preceding inequality implies that

$$\begin{aligned}\sum_{t=n^2}^{T-1} \left\|(\mathcal{J}_t^t)^{-1}\mathcal{W}_t^t\right\|^2 \;\le\; & 8c_\mu^{-2}n^{-8}T^{-4\nu}\sum_{t=n^2}^{T-1} t^2 \max_{1\le s< t-n^2} \left\|\theta_s - \theta_{t+1}\right\|^2 \\ & + 8c_\mu^{-2} \sum_{t=n^2}^{T-1} \max_{t-n^2\le s\le t} \left\|\theta_s - \theta_{t+1}\right\|^2.\end{aligned} \tag{D.3}$$

Therefore, by condition (5.1) and the fact that $n = \lceil \kappa T^{(1-\nu)/3}\rceil$, the first term on the right hand side of the preceding inequality is bounded above by $8c_\mu^{-2}n^{-8}T^{-4\nu}\sum_{t=n^2}^{T-1} t^2 V_{\boldsymbol\theta}(T) \le 8c_\mu^{-2}n^{-8}BT^{3-3\nu} \le 8c_\mu^{-2}\kappa^{-8}BT^{(1-\nu)/3}$. Furthermore, by (B.4) and the fact that $n = \lceil \kappa T^{(1-\nu)/3}\rceil$, the second term is less than or equal to $8c_\mu^{-2}n^2 V_{\boldsymbol\theta}(T) \le 32c_\mu^{-2}\kappa^2 BT^{(2+\nu)/3}$. Thus, the right hand side of (D.3) is bounded above by $8c_\mu^{-2}(\kappa^{-8} + 4\kappa^2)BT^{(2+\nu)/3}$. ■

**Proof of Theorem 6.** Note that inequalities (B.13) and (B.16) in the proof of Theorem 3 are valid under condition (5.1), implying that

$$\Delta^\pi_{\boldsymbol\theta}(T) \;\le\; \frac{3T}{n} \;+\; n^2 \;+\; 2K_0c_2\sum_{t=n^2}^{T-1} \mathbb{E}^\pi_{\boldsymbol\theta}\|(\mathcal{J}_t^t)^{-1}\mathcal{W}_t^t\|^2 \;+\; \frac{24K_0T}{\rho\tilde{c}n}. \tag{D.4}$$

By Lemma 7, the preceding inequality leads to $\Delta^{\pi}_{\boldsymbol{\theta}}(T) \leq 3T/n + n^2 + 2K_0c_1c_2T^{(2+\nu)/3} + 24K_0T/(\rho\tilde{c}n)$. Because $n = \lceil \kappa T^{(1-\nu)/3} \rceil$, this implies $\Delta^{\pi}_{\boldsymbol{\theta}}(T) \leq C\,T^{(2+\nu)/3}$ for all $\boldsymbol{\theta} \in \mathcal{V}(T,B)$, where $C = 3/\kappa + 4\kappa^2 + 2K_0c_1c_2 + 24K_0/(\rho\tilde{c}\kappa)$. ■

## References


Araman, V. and Caldentey, R. (2009), 'Dynamic Pricing for Nonperishable Products with Demand Learning', *Operations Research* **57**(5), 1169–1188.

Aviv, Y. and Pazgal, A. (2005), 'A Partially Observed Markov Decision Process for Dynamic Pricing', *Management Science* **51**(9), 1400–1416.

Balvers, R. and Cosimano, T. (1990), 'Actively Learning About Demand and the Dynamics of Price Adjustment', *The Economic Journal* **100**(402), 882–898.

Beck, G. and Wieland, V. (2002), 'Learning and Control in a Changing Economic Environment', *Journal of Economic Dynamics and Control* **26**, 1359–1377.

Besbes, O., Gur, Y. and Zeevi, A. (2013), 'Non-stationary Stochastic Optimization'. Working paper, Columbia University, New York, NY.

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

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

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

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

Brown, R. (1956), 'Exponential Smoothing for Predicting Demand'. Presented at the Tenth National Meeting of the Operations Research Society of America, San Francisco, November 16, 1956.

Chen, Y. and Farias, V. (2013), 'Simple Policies for Dynamic Pricing with Imperfect Forecasts', *Operations Research* **61**(3), 612–624.

den Boer, A. (2013), 'Dynamic Pricing with Multiple Products and Partially Specified Demand Distribution'. Working Paper, CWI, Amsterdam, The Netherlands.

den Boer, A. and Zwart, B. (2012), 'Simultaneously Learning and Optimizing using Controlled Variance Pricing'. Working Paper, CWI, Amsterdam, The Netherlands.

Farias, V. and van Roy, B. (2010), 'Dynamic Pricing with a Prior on Market Response', *Operations Research* **58**(1), 16–29.

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

Harrison, J. and Sunar, N. (2013), 'Investment Timing with Incomplete Information and Multiple Means of Learning'. Working Paper, Stanford University, Stanford, CA.

Holt, C. (1957), 'Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages', *Office of Naval Research Memorandum* **52**.

Keller, G. and Rady, S. (1999), 'Optimal Experimentation in a Changing Environment', *The Review of Economic Studies* **66**(3), 475–507.

Keskin, N. and Zeevi, A. (2012), 'Dynamic Pricing with an Unknown Demand Model: Asymptotically Optimal Semi-myopic Policies'. Working Paper, University of Chicago, Chicago, IL.

Lai, T. (1995), 'Sequential Changepoint Detection in Quality Control and Dynamical Systems', *Journal of the Royal Statistical Society. Series B (Methodological)* **57**(4), 613–658.

Lobo, M. and Boyd, S. (2003), 'Pricing and Learning with Uncertain Demand'. Working Paper. Stanford University, Stanford, CA.

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

Rustichini, A. and Wolinsky, A. (1995), 'Learning About Variable Demand in the Long Run', *Journal of Economic Dynamics and Control* **19**, 1283–1292.

Shiryaev, A. (2010), 'Quickest Detection Problems: Fifty Years Later', *Sequential Analysis* **29**(4), 345–385.

Tsybakov, A. (2009), *Introduction to Nonparametric Estimation*, Springer, New York.

Wang, Z., Deng, S. and Ye, Y. (2012), 'Close the Gaps: A Learning-while-doing Algorithm for a Class of Single-product Revenue Management Problems'. Working Paper, Stanford University, Stanford, CA.

Winters, P. (1960), 'Forecasting Sales by Exponentially Weighted Moving Averages', *Management Science* **6**(3), 324–342.