Dynamic Pricing in an Evolving and Unknown Marketplace

Yiwei Chen
Fox School of Business, Temple University, Philadelphia, PA 19122, yiwei.chen@temple.edu

Zheng Wen
DeepMind, Mountain View, CA 94043, zhengwen@google.com

Yao Xie
School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, yao.xie@isye.gatech.edu

We consider a firm that sells a single type product on multiple local markets over a finite horizon via dynamically adjusted prices. To prevent price discrimination, prices posted on different local markets at the same time are the same. The entire horizon consists of one or multiple change-points. Each local market's demand function linearly evolves over time between any two consecutive change-points. Each change-point is classified as either a zeroth-order or a first-order change-point in terms of how smooth the demand function changes at this point. At a zeroth-order change-point, at least one local market's demand function has an abrupt change. At a first-order change-point, all local markets' demand functions continuously evolve over time, but at least one local market's demand evolution speed has an abrupt change. The firm has no information about any parameter that modulates the demand evolution process before the start of the horizon. The firm aims at finding a pricing policy that yields as much revenue as possible. We show that the regret under any pricing policy is lower bounded by $CT^{1/2}$ with $C > 0$, and the lower bound becomes as worse as $CT^{2/3}$ if at least one change-point is a first-order change-point.

We propose a *Joint Change-Point Detection and Time-adjusted Upper Confidence Bound (CU)* algorithm. This algorithm consists of two components: the change-point detection component and the exploration-exploitation component. In the change-point detection component, the firm uniformly samples each price for one time in each batch of the time interval with the same length. She uses sales data collected at the times that she uniformly samples prices to both detect whether a change occurs and judge whether it is a zeroth-order or a first-order change if it occurs. In the exploration-exploitation component, the firm implements a time-adjusted upper confidence bound (UCB) algorithm between two consecutive detected change-points. Because demand dynamically evolves between two consecutive change-points, we introduce a time factor into the classical UCB algorithm to correct the bias of using historic sales data to estimate demand at present. We theoretically show that our CU algorithm achieves the regret lower bounds (up to logarithmic factors). Our numerical study shows that our policy performs well in a wide range of market environments.

Key words: revenue management, dynamic pricing, online learning, multi-armed bandit, change-point detection, asymptotic optimality

1. Introduction

In many industries, demand is not always stable during the selling season. It is also typically quite challenging to precisely predict how demand evolves over time. For example, as pointed out by Besbes and Sauré (2014), "*the sales of NFL jersey replicas during playoff season, where sales of the jerseys of winning teams increase over time, until the moment the team loses and sales come to an almost abrupt halt (Parsons (2004))*".

As another example, as studied by Hu et al. (2017), the product life cycles of most electronic consumer products consist of growth, maturity, and decline stages. Demand evolves with different patterns in these three stages. Demand is linearly increasing over time in the growth stage, stable in the maturity stage, and linearly decreasing in the decline stage. In addition, before an electronic product is launched to the market, it is very difficult to predict the exact times at which this product's demand will enter the maturity and the decline stages. It is also very hard to tell how demand will precisely evolve in each stage.

As pointed out above, it is already very difficulty to predict the collection of times at which a product switches its stage in its life cycle. This issue becomes more serious if a product faces an external shock that cannot be predicted in advance. For example, Bauer et al. (2020) study the impact of the Covid-19 on demands of a wide range products in 2021, such as personal computers, consumer electronics, wire and wireless communications. Maylín-Aguilar and Montoro-Sánchez (2020) study the impact of the 2007-2008 economic downturn on companies in Spain. Notice that both Covid-19 and the 2007-2008 economic downturn have immediate and big shocks on the economy. This typically leads to a burst change of a product's demand. The interplay of the hard-to-predict stage switching times of a product life cycle and the unpredictable global turmoils make people very hard to have a good prediction of a product's demand process in this product's entire life cycle.

Therefore, we are motivated to study the following problem. For a product whose demand evolves over time, in the absence of having any knowledge of the demand evolution process, how a firm dynamically adjusts the product's prices over time, so as to collect as much revenue as possible during the selling season.

We consider a firm that sells a single type product on multiple local markets over a finite horizon with a length of T. To avoid price discrimination, the prices posted on different local markets at the same time have to be the same. Demand evolves in the following way. There exists a collection of change-points. Between any two consecutive change-points, the demand function on each local market linearly evolves over time and is modulated by stationary parameters. At each change-point, either a zeroth-order type change or a first-order type change occurs. A zeroth-order type change means that at least one local market's demand function has an abrupt change. A first-order type change means that all local markets' demand functions continuously evolve around this change-point, but at least one local market's demand function has an abrupt change in its evolving speed. Before the start of the season, the firm has no information on any parameter that modulates the demand evolution processes on all local markets, such as the number of change-points, the times that changes occur, or any parameter that modulates the demand functions between any two consecutive change-points. The firm aims at finding a pricing policy that maximizes her expected total revenue during the entire season.

For any pricing policy, we define "regret" associated with this policy as the difference between the optimal expected revenue the firm collects during the entire season if she perfectly knew the demand evolution process ex-ante and her expected revenue collected under this policy if she has no information of the demand evolution process ex-ante. In this paper, we use regret to measure the performance of a pricing policy.

1.1. Our Contributions

We make the following contributions in this paper.

1. We show that regret is always lower bounded by $CT^{1/2}$, with $C > 0$ (Theorem 1 Part 1). In addition, if at least one change-point is first-order (a first-order change happens at this change-point), then regret is always lower bounded by $CT^{2/3}$ (Theorem 1 Part 2).
2. We propose an easy-to-implement *Joint Change-point Detection and Time-adjusted Upper Confidence Bound (CU)* algorithm (see §4.1). This algorithm consists of two components. The first component is the *Change-point Detection* sub-algorithm (see §4.2). Since the latest detected change-point, the firm uniformly samples every price for one time in each batch of the time interval with the same length. She uses historic price and sales data collected at times and performs uniform sampling of all prices in the past with different lengths of time windows. She detects whether a change has happened and judges whether it is a zeroth-order or a first-order change once it has been detected. Therefore, our algorithm has a feature that we do not require the firm to have any prior knowledge of the types of change-points.

   The second component of our algorithm is the *Time-adjusted Upper Confidence Bound (UCB)* algorithm (see §4.3). Since the latest detected change-point, at each time that the firm does not perform uniform sampling for change-point detection, she makes the pricing decision based on the classical UCB algorithm to balance the exploration and exploitation. However, we cannot directly implement the classical UCB algorithm. Recall that this algorithm is designed for the demand process that is stationary over time. However, our model allows demand to dynamically evolve over time between any two consecutive change-points. Hence, for each given price, the empirical mean of the historic sales data under this price gives us a biased estimate of the expected demand under this price at the current point of time. Therefore, we introduce a time-adjusted factor into the classical UCB algorithm to correct this bias.

3. We show that the regret on the CU algorithm is upper bounded by $O\left(T^{2/3}(\log T)^{1/3}\right)$ if at least one change is a first-order change, and $O\left(T^{1/2}(\log T)^{1/2}\right)$ if all changes are zeroth-order changes (Theorem 2). These results and the aforementioned regret lower bounds imply that the CU algorithm achieves the regret lower bounds (up to logarithmic factors).
4. We propose a novel approach to establish the upper bounds on the regret of the CU algorithm. First, we classify all change-point detection events into well-detected and badly-detected events. Well-detected events mean that the firm successfully detects every change-point and its type in a short time window after it occurs. Moreover, she does not falsely detect any change-point if it does not exist. Badly-detected events refer to all events that are not well-detected events. Doing so allows us to separately analyze the revenue loss that arises from the change-point detection sub-algorithm and the revenue loss that arises from the time-adjusted UCB sub-algorithm (Lemma 3).

   Second, we upper bound the revenue loss that arises from the change-point detection sub-algorithm. We do so by establishing an upper bound on the probability that the badly-detected events happen (Lemma 4). We develop different approaches to upper bound the probability that a zeroth-order change-point is detected (Lemma 1 Part 1), the probability that a first-order change-point is detected (Lemma 1 Part 2), the probability that the firm falsely detects a zeroth-order change-point if it does not exist (Lemma 1 Part 3(a)) and the probability that the firm falsely detects a first-order change-point if it does not exist (Lemma 1 Part 3(b)).

   Third, we upper bound the revenue loss that arises from the time-adjusted UCB sub-algorithm. We do so by establishing an upper bound on the regret that is conditional on that all change-point events are well-detected events (Lemma 11). One key step to establishing this upper bound is to handle regret accumulated in each time interval that is between any two consecutively detected change-points. Depending on whether the ending time of each time interval is detected as a zeroth-order or a first-order change-point, we develop different approaches to establish regret upper bounds for this time interval (see Steps 3.3 and 3.4 in §5.2).

The rest of the paper is organized as follows. In §1.2, we provide literature review. In §2, we present our model. In §3, we establish lower bounds on the regret under any pricing policy. In §4, we present our CU algorithm. In §5, we present upper bounds on the regret under the CU algorithm. We provide an outline of the proof of these regret upper bounds. In §6, we In §7, we conclude our paper.

1.2. Literature Review

Three streams of literature have informed and inspired our research: revenue management and pricing with demand learning, statistical change-point detection, and multi-armed bandit (MAB) problems with change-point detection.

Revenue management and pricing with demand learning. There is a large body of literature that studies revenue management and pricing problems that the demand functions are stationary during the season, but the decision-makers lack the perfect information of the demand functions before the start of the season. Therefore, the decision-makers need to carefully balance exploration and exploitation over time, to collect as much revenue as possible. This stream of literature include but not limited to Araman and Caldentey (2009), Besbes and Zeevi (2009, 2012), Broder and Rusmevichientong (2012), Chen and Gallego (2018), Chen et al. (2019b), Chen and Shi (2019), Cheung et al. (2017), den Boer and Zwart (2015), Farias and Van Roy (2010), Ferreira et al. (2018), Harrison et al. (2012), Keskin and Zeevi (2014), Wang et al. (2014). The following papers jointly consider the pricing and inventory replenishment decisions in the presence of unknown demand functions: Chen and Chao (2017), Chen et al. (2019a), Chen and Chao (2019), Chen et al. (2015). The present paper is distinguished from these papers that we allow the demand function to be non-stationary over time.

In recent years, there is a growing number of revenue management and pricing literature that model the demand functions to be non-stationary over time and assume that the decision-makers lack the perfect information of the demand function evolution processes. Besbes and Zeevi (2011) study a setting that the demand function might have an abrupt change at an unknown time and needs to be detected. The paper assumes that the pre- and post-change demand functions are known in advance. Hence, the challenge is mainly the detection. The authors propose a dynamic pricing policy and show that it achieves the regret lower bound. Besbes and Sauré (2014) also study a setting that the demand function might have an abrupt change at an unknown time. The paper assumes that the decision-maker knows the pre-change demand function in advance, but not the post-change demand function. However, the decision-maker can detect the demand function change instantaneously at the time that the change occurs. She can also learn instantaneously and perfectly the post-change demand function. The authors study the effect of the inventory constraint on the optimal pricing policy. Chen and Farias (2013) study a setting that the market size process stochastically evolves over time and is unknown to the decision-maker. However, the customer's willingness-to-pay function is stationary over time and is perfectly known to the decision-maker. The authors propose a pricing policy that requires the decision-maker to repeatedly solve a deterministic optimization problem by using real-time information on the market size and the remaining inventory. The authors show that this policy achieves a constant factor performance guarantee relative to the optimal revenue. Den Boer (2015) study a setting that the demand function is in an additive form of a price-independent function that stochastically evolves over time and is unknown to the decision-maker; a price-dependent function whose evolution process is known to the decision-maker. All the papers mentioned above assume that the decision-makers know at least some parameters that modulate the demand evolution processes. By contrast, our paper allows the decision-maker to be agnostic to all these parameters. Chen et al. (2017) study a general problem that the cost function is non-stationary over time and is convex in each period. The authors divide the entire horizon into multiple batches. They re-run online gradient descent algorithm in the beginning of each batch. Cheung et al. (2018) study a non-stationary linear stochastic bandit problem. They propose a sliding Window Upper Confidence Bound algorithm and shows that it achieves the optimality. Zhou et al. (2020) study a multi-armed bandit problem that each arm's reward is modulated by the system's underlying state that evolves according to a Markov process. The decision maker does not observe the underlying state and has to learn the unknown transition probability matrix as well as the reward distribution. The authors propose and analyze an online learning algorithm that jointly uses the spectral method-of-moments estimations for hidden Markov models and upper confidence bound methods. Keskin and Li (2020) study a firm's dynamic pricing problem with unknown and time-varying heterogeneity in customers' preferences for quality. With unknown market transition structure, the authors design a simple and practically implementable policy, bounded learning policy.

The paper by Keskin and Zeevi (2016) is closely related to our present paper. The paper studies a setting that the demand function is linear in price at all points of times. The parameters of the linear demand function may evolve over time. The authors use a notion "budget" to measure the parameter changes during the season. This metric is analogous to the number of change-points in our model. The paper considers two types of demand evolution processes. One is that all changes are abrupt, which is analogous to the zeroth-order changes in our paper. The other is that all changes are smooth, which is analogous to the first-order changes in our paper. The authors develop different pricing algorithms for each type of demand evolution process and establish their respective regret upper bounds.

It is worth discussing the distinctions between this paper and our paper.

1. Keskin and Zeevi (2016) assume that the demand function at any time is linear in price. By contrast, our paper allows the demand function to take any form with respect to the prices.
2. Keskin and Zeevi (2016) propose two different pricing algorithms in terms of whether all demand function changes are smooth or abrupt. For the case that all demand function changes are smooth, each price used for the change-point detection is posted for one time in each

time interval with length $O\left(T^{1/3}\right)$ and the number of sales data under this price used for the detection is $O\left(T^{1/3}\right)$, where T is the length of the horizon (see §3.3.1 and §3.3.2 in Keskin and Zeevi (2016)). For the case that all demand function changes are abrupt, each price used for the change-point detection is posted for $O(\log T)$ times in each time interval with length $O\left(T^{1/2}\right)$ and the number of sales data under this price used for the detection is $O(\log T)$ (see §4.4.1 in Keskin and Zeevi (2016)). Implementing these algorithms requires the decision-maker to have a prior knowledge of which case the demand evolution process falls into.

By contrast, our paper proposes a unified pricing algorithm for both cases without requiring the decision-maker to have any prior knowledge of the types of demand function changes. Regardless of the types of the change-points, the decision maker always posts each price used for the change-point detection for one time in each time interval with length $O\left(T^{1/2}/\left(\log T\right)^{1/2}\right)$ (see Theorem 2 in our paper). However, the decision-maker uses different numbers of the sales data under this price to detect a change-point and its type. The number of data used to detect a smooth (first-order type) change is $O\left(T^{1/3}\left(\log T\right)^{2/3}\right)$ (see Theorem 2 in our paper). The number of data used to detect an abrupt (zeroth-order type) change is $O(\log T)$ (see Theorem 2 in our paper).

3. In Keskin and Zeevi (2016), in the case that all demand function changes are abrupt, the authors assume that the demand function between any two consecutive change-points is stationary. Hence, if the decision-maker knew the demand function, then her optimal price would be stationary. By contrast, our paper allows each local market's demand function between any two consecutive change-points to evolve over time linearly. Hence, even in an idealized scenario that the decision-maker perfectly knew the demand function in advance, the optimal prices might still change over time (see Example 1).

Statistical change-point detection. Change-point detection has been a long standing challenge in statistics. Classical sequential change-point detection Siegmund (1985), Basseville and Nikiforov (1993), Brodsky and Darkhovsky (1993), Chen and Gupta (2012), Veeravalli and Banerjee (2013), Tartakovsky et al. (2014), where one monitors *i.i.d.* univariate and low-dimensional multivariate observations observations from a single data stream is a well-developed area. Outstanding contributions include Shewhart's control chart Shewhart (1931), Page's CUSUM procedure Page (1954, 1955), Shiryaev-Roberts procedure Shiryaev (1963), Roberts (1966), Gordon's non-parametric procedure Gordon and Pollak (1994), and window-limited procedures Lai (1995). Various asymptotic optimality results have been established for these classical methods Lorden (1971), Pollak (1985, 1987), Moustakides (1986), Lai (1995, 1998). Here, we will use a simple sliding window two-sample based procedure for change-point detection, because it will simplify analysis and enough for us to show our main results.

Multi-armed bandit (MAB) problems with change-point detection. A large body literature study MAB problems in the settings with one or multiple change-points, which can be viewed as a type of bandit problems with structured rewards. The reward functions on all arms are stationary between any two consecutive change-points. At least one arm's reward has an abrupt change at a change-point. Before the start of the season, the decision-makers are agnostic to the reward evolution processes and the points of times at which changes occur. This stream of literature includes but not limited to Auer et al. (2002), Auer (2002), Besbes et al. (2014), Cao et al. (2018a), Garivier and Moulines (2011), Kocsis and Szepesvári (2006), Liu et al. (2018). The present paper is distinguished from these papers in the following two aspects. First, these papers assume that each arm's reward between any two consecutive change-points is constant. By contrast, our paper allows each arm's reward to evolve over time between any two consecutive change-points linearly. Second, these papers assume that a zeroth-order change occurs at each change-point, i.e., at least one arm's reward has a sudden jump at each change-point. By contrast, our paper also allows a first-order change to occur at a change-point. This means that all arms' rewards are continuous in time around a change-point, and at least one arm's reward evolving speed has a sudden jump at this change-point. Other types of structured rewards have also been considered, e.g., the so-called matroid bandits in Kveton et al. (2014) and cascading bandits in Cheung et al. (2019). Because the first-order changes will result in non-stationary and non-i.i.d. observations, the analysis will be significantly different from the existing works assuming the reward function on all arms are stationary between two consecutive change-points (see, e.g., the explanation and analysis for first-order change-point detection in Cao et al. (2018b)). Thus, we need new techniques for analyzing MAB under both zeroth-order and first-order changes.

2. Model

We consider a firm that sells a single type product on M geographically separated local markets over a finite horizon $\{1,\cdots,T\}$ via anonymous posted prices. The price posted on each local market in each period is selected from a finite set $\{p_1,\cdots,p_K\}$ with $K<\infty$[1]. We denote $\bar{p} \triangleq \max\{p_1,\cdots,p_K\}$. We assume that in each period, the prices posted for different local markets are the same, i.e.,

[1] The assumption that the number of prices is finite is consistent with the practice. For instance, it is pointed out by Ferreira et al. (2015) Page 77 that Rue La La chooses prices from a finite set: "*Rue La La typically chooses prices that end in 4.90 or 9.90 (i.e., \$24.90 or \$119.90). The set of possible prices is characterized by a lower bound and an upper bound and every increment of five dollars in between; for example, if the lower bound on a style's price is \$24.90 and the upper bound is \$44.90, then the set of possible prices is* $\mathcal{M}=\{\$24.90,\$29.90,\$34.90,\$39.90,\$44.90\}$."

there is no price discrimination over customers who purchase from different local markets. This is consistent with many companies' practices, such as Apple, Brooks Brothers, and Ikea. We denote by $\pi_t \in \{1, \cdots, K\}$ the index of the price posted on all local markets in period t.

On each local market m, at most one customer arrives in each period t, with probability $S_t^m \in [\underline{d}, 1]$ where $\underline{d} > 0$. We hereafter call S_t^m the market size. An arriving customer makes an immediate purchase if her willingness-to-pay is no less than the posted price. Otherwise, she immediately leaves the system. We assume that a customer's willingness-to-pay is a random variable, with the c.d.f. $F_t^m(\cdot)$. We define the complementary c.d.f. as $\bar{F}_t^m(\cdot) \triangleq 1 - F_t^m(\cdot)$. We assume $\bar{F}_t^m(p_k) \geq \underline{\theta} > 0$ for all t, m and k.

Market environment evolution processes. We allow the market environment to evolve over time. We assume that there exist J points of times, $1 < \tau_1 < \cdots < \tau_J < T$, such that at each of these times, at least one local market's environment has a change, such as an abrupt change of the market size, or an abrupt change of customer willingness-to-pay distribution function. These J points of times are hereafter called as *change-points*. We will formally define what happen at these change-points later. We make a convention that $\tau_0 = 1$ and $\tau_{J+1} = T + 1$.[2]

We assume that in each global market environment $j \in \{0, \cdots, J\}$ that corresponds to the time interval $\{\tau_j, \cdots, \tau_{j+1} - 1\}$ and each local market $m \in \{1, \cdots, M\}$, the market size evolves with a linear trend and the customer willingness-to-pay distribution functions are stationary, i.e., $S_t^m = a_j^m + b_j^m \frac{t}{T}$ and $\bar{F}_t^m(p_k) = \theta_j^m(k)$ for any $t \in \{\tau_j, \cdots, \tau_{j+1} - 1\}$.[3]

Now, we formally define what happens at each change-point. We categorize all change-points into two classes in terms of how smooth the changes are at these points: the *zeroth-order change-points* and the *first-order change-points*.

DEFINITION 1.

1. (Zeroth-order change-point) A change-point τ_j is a <u>zeroth-order change-point</u> if and only if

$$\left(a_j^m + b_j^m \frac{\tau_j}{T}\right) \theta_j^m(k) \neq \left(a_{j-1}^m + b_{j-1}^m \frac{\tau_j}{T}\right) \theta_{j-1}^m(k), \ \exists\ m, k. \tag{1}$$

2. (First-order change-point) A change-point τ_j is a <u>first-order change-point</u> if and only if

$$\left(a_j^m + b_j^m \frac{\tau_j}{T}\right) \theta_j^m(k) = \left(a_{j-1}^m + b_{j-1}^m \frac{\tau_j}{T}\right) \theta_{j-1}^m(k), \ \forall\ m, k \tag{2}$$

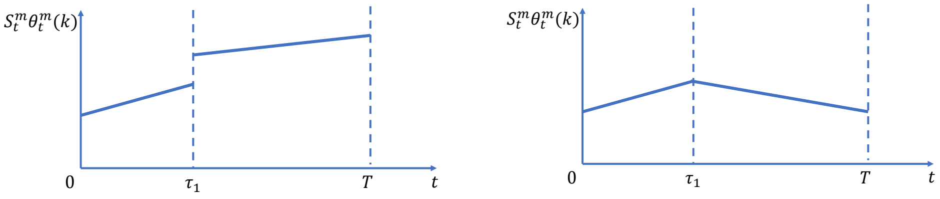
and

$$b_j^m \theta_j^m(k) \neq b_{j-1}^m \theta_{j-1}^m(k), \ \exists\ m, k. \tag{3}$$

[2] Our model allows $J = 0$. This entails that the system has no change-point.

[3] The reason that the trend term in the market size function S_t^m is scaled by T is as follows. We define the market size S_t^m as the probability that a customer arrives to local market m in period t. Hence, the value of S_t^m must fall between 0 and 1. By scaling the trend term by T, the sufficient conditions for $S_t^m \in [0,1]$ are $a_j^m \in [0,1]$ and $a_j^m + b_j^m \in [0,1]$. These conditions are independent of T. Therefore, these conditions guarantee $S_t^m \in [0,1]$ regardless of how large T is when we do the asymptotic analysis with respect to T.

The definitions above entail that at a zeroth-order change-point, at least one local market's demand function has an abrupt change. At a first-order change-point, although all local markets' demand functions change smoothly over time, there is at least one local market whose demand function evolution speed has an abrupt change. Figure 1(a) (resp. Figure 1(b)) illustrates a zeroth-order (resp. first-order) change that happens at time τ_1. We denote by $\mathcal{N}_0$ (resp. $\mathcal{N}_1$) the collection of all zeroth-order (resp. first-order) change-points. We make a convention that $J+1 \in \mathcal{N}_0$.



(a) Zeroth-order change at time τ_1 (b) First-order change at time τ_1

Figure 1 Two types of change-points.

It is worth noting that our model is general enough to characterize a wide range of market environment evolution processes that have been either theoretically studied or empirically verified. We give some examples below.

1. (Stationary processes) There is no change-point in this family of processes, i.e., $J = 0$. Therefore, the model is reduced to be a classical revenue management model with stationary demand functions that do not evolve over time.
2. (Piecewise-stationary processes) For this family of processes, the market size in each local market does not have a trend in each global market environment, i.e., $b_j^m = 0$. This model has been extensively studied in the literature. See, e.g., Besbes and Zeevi (2011).
3. (Continuous piecewise-linear processes) For this family of processes, each local market's market size evolution process is continuous and piecewise-linear in time, with the evolving speeds that only change at the change-points. Hu et al. (2017) provide empirical evidence that this model can be used to characterize the electronic products' market size evolution processes over the product life cycles.

It is also worth noting that we define a change-point as a time at which *at least* one local market's environment changes. This entails that our model allows multiple local markets' environments to change simultaneously. This may happen if an event that affects the global market occurs.

We define

$$\Delta_0 \triangleq \min_{j \in \mathcal{N}_0 \backslash \{J+1\}} \max_{m,k} \left| \left(a_j^m + b_j^m \frac{\tau_j}{T} \right) \theta_j^m(k) - \left(a_{j-1}^m + b_{j-1}^m \frac{\tau_j}{T} \right) \theta_{j-1}^m(k) \right|$$

and

$$\Delta_1 \triangleq \min_{j\in\mathcal{N}_1} \max_{m,k} \left| b_j^m \theta_j^m(k) - b_{j-1}^m \theta_{j-1}^m(k) \right|$$

and $\bar{b} \triangleq \max_{j,m} |b_j^m|$. Therefore, the definitions of the zeroth-order and the first-order change-points imply $\Delta_0 > 0$ and $\Delta_1 > 0$, respectively.

The firm's information structure. We assume that the firm does not know any market environmental parameter before the start of the selling horizon. To be specific, the firm does not know (1) the number of change-points J, (2) at which points of times changes occur $\{\tau_1, \cdots, \tau_J\}$, (3) the market size parameters $\left\{a_j^m, b_j^m : \forall\ j, m\right\}$, (4) the willingness-to-pay distribution parameters $\left\{\theta_j^m(k) : \forall\ j, m, k\right\}$. We denote by X_t^m the sales quantity on the local market m in period t. We denote by $\mathcal{F}_t = \sigma\left(X_s^m, \pi_s : \forall\ s \in \{1, \cdots, t\}, m \in \{1, \cdots, M\}\right)$ the filtration generated by the history of the sales and price processes up to time t. We denote by Π the family of all price processes $\{\pi_t : \forall\ t \in \{1, \cdots, T\}\}$ that satisfy the property that π_t is $\mathcal{F}_{t-1}$-measurable for all $t \in \{1, \cdots, T\}$.

Under policy $\pi \in \Pi$, the expected total revenue the firm garners is equal to

$$V^\pi(T) = \sum_{j=0}^{J} \sum_{t=\tau_j}^{\tau_{j+1}-1} \mathsf{E}\left[g_t(\pi_t)\right],$$

where

$$g_t(k) = p_k \sum_{m=1}^{M} \left(a_j^m + b_j^m \frac{t}{T}\right) \theta_j^m(k),$$
$$\forall\ j \in \{0, \cdots, J\}, t \in \{\tau_j, \cdots, \tau_{j+1}-1\}, k \in \{1, \cdots, K\}.$$

Regret. We denote by

$$k_t^* \in \arg\max_k g_t(k)$$

the firm's optimal pricing decision in period t if she had the full information of all system parameters before the start of the selling horizon.

We denote by

$$V^*(T) = \sum_{t=1}^{T} g_t(k_t^*)$$

the firm's optimal expected total revenue in the full information scenario.

We use *regret* to measure the performance of any pricing policy $\pi \in \Pi$:

$$\text{Regret}^\pi(T) \triangleq V^*(T) - V^\pi(T),$$

It is worth noting that even in the full information scenario, when the number of local markets is more than 1 ($M > 1$), it is not always the case that the optimal prices at different times that are in the same global market environment are the same. The following example illustrates this point.

EXAMPLE 1. We consider an instance with two local markets $M = 2$, two prices $p_1 = 2$ and $p_2 = 3$ and no change-point $J = 0$. Parameters for local market 1 are $a_0^1 = 0.5$, $b_0^1 = 0$, $\theta_0^1(1) = 0.5$ and $\theta_0^1(2) = 0.3$. Parameters for local market 2 are $a_0^2 = 0$, $b_0^2 = 0.5$, $\theta_0^2(1) = 0.6$ and $\theta_0^2(2) = 0.5$. Therefore, $k_t^* = 1$ if $t \le T/3$ and $k_t^* = 2$ if $t > T/3$.

3. Lower Bounds on Regret

In this section, we establish lower bounds on the regret for any pricing policy. The main results are as follows.

THEOREM 1.

1. *If there is no first-order change-point, i.e., $\mathcal{N}_1 = \emptyset$, then*

$$\text{Regret}^{\pi}(T) \ge \frac{1}{32e}\sqrt{\frac{KT}{3}}, \ \forall \ \pi \in \Pi.$$

2. *For $T \ge 46$, if there exists at least one first-order change-point, i.e., $\mathcal{N}_1 \ne \emptyset$, then*

$$\text{Regret}^{\pi}(T) \ge \frac{1}{1024e^2} K^{1/3} T^{2/3}, \ \forall \ \pi \in \Pi.$$

Part 1 is a standard result in the multi-armed bandit literature. To show Part 1, it is sufficient to analyze a special class of instances in which there is no change-point, and the market size does not evolve over time, i.e., the classical multi-armed bandit problems in the stationary environments.

Part 2 shows that the regret lower bound becomes worse ($O\left(T^{2/3}\right)$) if the market environment changes smoothly (first-order change). The intuition is as follows. At a first-order change-point, the change only happens on the demand function evolution speed (trend), and the demand function still continuously evolves at this change-point. Therefore, such a smooth change cannot be immediately observed. The firm needs to collect more data in a longer time window around this change-point to detect such change.

Now, we provide an outline of the proof of Part 2 of this theorem. First, we construct two instances that both have a single first-order change-point and no zeroth-order change-point, and the first-order changes in these two instances are at the same time. The firm has the same set of alternative prices to post in both instances. The second instance is almost the same as the first one, except that the c.d.f. of one sub-optimal price in the first instance has a slight perturbation, such that it is optimal in the second instance. Such a slight difference leads to different optimal pricing strategies if the firm knew which instance she played with ex-ante. Second, we categorize all times after the change-point into L groups with a tuning parameter L. In each group, the distance between any two consecutive points of time is L. For each group of times, We establish an upper bound on the Kullback-Leibler divergence between the two instances. We then use this result to prove that for any pricing policy and each group of times, the sum of the regrets accumulated over all times in this group over two instances must be bad enough. By using this result and carefully choosing the tuning parameter L, we show that for any pricing policy, the cumulative regret over the entire horizon must perform badly in at least one of the two instances.

4. Joint Change-point Detection and Time-adjusted Upper Confidence Bound (CU) Algorithm

In the previous section, we establish lower bounds on regret for any pricing algorithm. In this section, we propose an easy-to-implement algorithm, *Joint Change-point Detection, and Time-adjusted Upper Confidence Bound (CU)* algorithm. We will show that the regret of this algorithm achieves our established lower bounds (up to logarithmic factors).

4.1. The CU Algorithm

Before we formally present our CU algorithm, we provide a high-level intuition of this algorithm. Our CU algorithm is designed to achieve the following objectives:

1. **OBJECTIVE 1: Do accurate and efficient change-point detection.** Our CU algorithm shall be capable of correctly detecting every change-point and identifying its type in a short time window after this change occurs. In addition, it should avoid falsely alarming a change-point that does not exist.
2. **OBJECTIVE 2: Do efficient exploration-exploitation after a change occurs.** After a change occurs, the market enters a new environment. Hence, the firm has to re-learn the new market environment. Therefore, our CU algorithm shall be capable of helping the firm quickly learn the new market environment and exploit the learning outcome to make the optimal decision for the majority of times in this new market environment.

To achieve both of these two objectives, our CU algorithm is designed with two components, the *change-point detection component* and the *exploration-exploitation component*, that target at achieving two listed objectives above, respectively.

We introduce the following notation that facilitates us to present our algorithm. We denote by $\hat{j}_t$ the firm's estimate of the index of the global market environment that time t falls in. We denote by $\hat{\tau}_j$ the firm's estimate of the starting time of the jth global market environment.

Our CU algorithm is formally presented as follows.

Joint Change-point Detection and Time-adjusted Upper Confidence Bound (CU) Algorithm

1. ***(Initialize Parameters)***

   Before the start of the season, the firm determines the following parameters.

   (a) The size of data used to detect whether there is a zeroth-order change-point, w_0, with $\bmod(w_0, 2) = 0$.

   (b) The zeroth-order change-point detection tolerance error $\epsilon_0 \in \left(0, \frac{\Delta_0}{2}\right)$.

   (c) The size of data used to detect whether there is a first-order change-point, w_1, with $\bmod(w_1, 3) = 0$.

   (d) The first-order change-point detection tolerance error $\epsilon_1 \in \left(0, \frac{\Delta_1}{2}\right)$.

   (e) The time interval of doing uniform sampling of all prices, $L > K$.

   (f) The time-adjusted UCB confidence level $\beta > 0$.

   (g) $\hat{j}_1 = 0$ and $\hat{\tau}_0 = 1$.

2. In each period t,

   (a) If $t \in \left\{\hat{\tau}_{\hat{j}_t}, \cdots, \hat{\tau}_{\hat{j}_t} - 1 + 2K\right\}$, the firm posts a price with the following index

   $$\pi_t^{\mathrm{CU}} = \left\lceil \frac{t - \hat{\tau}_{\hat{j}_t} + 1}{2} \right\rceil.$$

   (b) ***(Detect change-points)*** Otherwise, if $\bmod\left(t - \hat{\tau}_{\hat{j}_t} + 1, L\right) \in \{1, \cdots, K\}$, then

   i. The firm does *uniform sampling* by posting a price with the following index

   $$\pi_t^{\mathrm{CU}} = \bmod\left(t - \hat{\tau}_{\hat{j}_t} + 1, L\right).$$

   The firm then observes the sales in period t in all local markets: $\{X_t^m : \forall\, m \in \{1, \cdots, M\}\}$.

   ii. The firm uses a *change-point detection algorithm* (specified later) to detect whether a change occurs during the time interval $\left\{\hat{\tau}_{\hat{j}_t}, \cdots, t\right\}$. If a change is detected, then the firm sets $\hat{j}_{t+1} = \hat{j}_t + 1$ and $\hat{\tau}_{\hat{j}_{t+1}} = t + 1$.

   (c) ***(Balance exploration and exploitation)*** Otherwise, the firm computes π_t^{CU} by using a *time-adjusted upper confidence bound (UCB) algorithm* (specified later).

Before we proceed to present the two sub-algorithms within this CU algorithm, the change-point detection algorithm, and the time-adjusted UCB algorithm, we first discuss the key features of the CU algorithm.

First, recall that the firm has no information about (1) the number of change-points, (2) the points of times those changes occur, and (3) the type of each change-point. Therefore, our policy has the *change-point detection component 2(b)ii* that facilitates the firm to detect whether a change has occurred.

Second, note that if a price has not been posted for a good number of periods, and a change occurs at this price, this change cannot be easily detected. This fact is due to the lack of enough sales data under this price. Therefore, to avoid this issue, in 2(b)i of the CU algorithm, we require the firm to uniformly sample all prices at the beginning of each batch of periods with size L. The uniform sampling ensures the firm to collect enough sales data under each price that enables her to do the change-point detection.

Third, recall that for each global market environment, the market is modulated by parameters that are stationary over time. However, the firm does not know the values of these parameters in advance. Therefore, our policy has the *exploration-exploitation component 2(c)* that enables the firm to effectively learn the unknown parameters and take the optimal actions in each global market environment.

Fourth, recall that in a given global market environment, if each local market's market size did not evolve over time, then the classical UCB algorithm can be directly applied and have been proven to perform well. However, our model has a key feature that each local market's market size may linearly evolve over time. Therefore, the classical UCB algorithm computed from using historic data gives us a biased estimate of the demand function at the present time. Therefore, we introduce a *time-adjustment factor* into the classical UCB algorithm to correct this bias. The detail will be presented in Section 4.3.

Fifth, in order to get a finite confidence bound of each price, we need to have sales data under each price. Hence, as shown in 2(a) of our CU algorithm, immediately after the firm detects a change-point, we require every price to be posted.

4.2. Sub-algorithm 1: Change-point Detection Algorithm

In this subsection, we present the first sub-algorithm in the CU algorithm, the *change-point detection algorithm*. We begin with providing an overview of the key idea of this sub-algorithm.

Recall from the CU algorithm Part 2(b)i that the firm uniformly samples all prices at the beginning of each batch of periods with a given size L. In each period that the firm does such uniform sampling, after the firm posts a price p, the firm tracks the most recent w_0 (resp. w_1) periods in which the firm did uniform sampling and also posted the same price p. The firm then uses the sales data under price p at these w_0 (resp. w_1) periods to detect if a zeroth-order (resp. first-order) change just happened at price p.

To detect if a zeroth-order change just happened, the firm divides w_0 sales data under the same price p into two groups that are with equal size. The firm computes the average sales quantity of each group and then computes their *first-order difference*. Note that this first-order difference should not deviate too much from 0 if there is no zeroth-order change at price p. Hence, the firm uses this intuition to check whether a zeroth-order change at price p just happened. If the first-order difference of the average sales quantity exceeds a threshold level, then the firm believes a zeroth-order change at price p just happened. Otherwise, the firm believes there was no zeroth-order change at price p and continues to detect if a first-order change at price p just happened.

The firm divides w_1 sales data under the same price p into three groups that are with equal size. The firm computes the average sales quantity of each group and then computes their *second-order difference*. Note that this second-order difference should not deviate too much from 0 if there is no first-order change at price p. Hence, the firm uses this intuition to check whether a first-order change at price p just happened. If the second-order difference of the average sales quantity exceeds a threshold level, then the firm believes a first-order change at price p just happened. Otherwise, the firm believes there was no first-order change at price p.

Now, we formally present our change-point detection sub-algorithm.

We introduce the following notation that facilitates us to present this algorithm. We denote by

$$\bar{X}_t^{m,0} = \frac{\sum_{i=0}^{w_0/2-1} X_{t-iL}^m}{w_0/2}, \qquad \bar{X}_t^{m,1} = \frac{\sum_{i=0}^{w_1/3-1} X_{t-iL}^m}{w_1/3}$$

two empirical mean sales on local market m with the sales data selected with time interval L. We will use $\bar{X}_t^{m,0}$ (resp. $\bar{X}_t^{m,1}$) to detect zeroth-order (resp. first-order) change-points.

The change-point detection algorithm is as follows.

Sub-algorithm 1: Change-point Detection Algorithm

The firm sets $\hat{j}_{t+1} = \hat{j}_t + 1$ and $\hat{\tau}_{\hat{j}_{t+1}} = t+1$ if and only if at least one of the following two events happens:

1. (Detect whether a zeroth-order change-point exists)

   If $t \geq \hat{\tau}_{\hat{j}_t} + (w_0 - 1)L + 2K$, then the firm computes whether there exists m, such that

$$\left| \bar{X}_t^{m,0} - \bar{X}_{t-w_0L/2}^{m,0} \right| > \epsilon_0. \tag{4}$$

2. (Detect whether a first-order change-point exists)

   If $t \geq \hat{\tau}_{\hat{j}_t} + (w_1 - 1)L + 2K$, then the firm computes whether there exists m, such that

$$\left| \frac{2T}{\left(\frac{w_1}{3} - 1\right)L} \left(\left(\bar{X}_t^{m,1} - \bar{X}_{t-w_1L/3}^{m,1} \right) - \left(\bar{X}_{t-w_1L/3}^{m,1} - \bar{X}_{t-2w_1L/3}^{m,1} \right) \right) \right| > \epsilon_1. \tag{5}$$

We make the following observations of this change-point detection sub-algorithm.

1. At each point of time that the firm detects a change-point, our algorithm allows the firm to simultaneously do both the zeroth-order change-point detection and the first-order change-point detection.
2. Our algorithm does not require the firm to have any prior knowledge about the number of zeroth-order change-points or the number of first-order change-points, or how these two types of change-points are distributed over time.
3. In Equation (4), we detect a zeroth-order change-point by computing the *first-order difference* of two average sales quantities computed from two groups of sales data under the same price. It is worth noting that this does not take into account the trend of any market environment. This is reasonable if the magnitude of a zeroth-order change is much more than the change that comes from the accumulation of the trend that evolves during the change-point detection time window, i.e., $\frac{w_0 L}{T} << \Delta_0$. Hence, to enable our algorithm to work, the selection of parameters w_0 and L should satisfy this condition.
4. In Equation (5), we detect a first-order change-point. Note that this detection is only performed after we complete the zeroth-order change-point detection and do not observe a zeroth-order change. In addition, if a first-order change occurs, then it comes from the change of the trend. Hence, we perform a first-order change-point detection by computing the *second-order difference* of three average sales quantities computed from three groups of sales data under the same price. Because the length of the first-order change-point detection time window is $w_1 L$, the magnitude of the first-order change is in the order of $O\left(\frac{w_1 L}{T}\right)$. Hence, to normalize the metric for the first-order change-point detection, we multiple the L.H.S. of Equation (5) by a term that is in the order of $O\left(\left(\frac{w_1 L}{T}\right)^{-1}\right)$.
5. We observe that the sample size for the zeroth-order (resp. first-order) detection, w_0 (resp. w_1), is critical for our change-point detection accuracy, i.e., whether the firm can successfully detect each change-point, identify its type, and avoid any false alarm of a change-point that does not actually exist. The following lemma helps us to understand this point in a rigorous and quantitative way.

LEMMA 1. *Consider period t with $\operatorname{mod}\left(t - \hat{\tau}_{\hat{j}_t} + 1, L\right) = k \in \{1, \cdots, K\}$, i.e., the firm does uniform sampling in period t. We assume $\hat{\tau}_{\hat{j}_t} \in \{\tau_j, \cdots, \tau_{j+1} - 1\}$ for some $j \in \{0, 1, \cdots, J\}$, i.e., $\hat{\tau}_{\hat{j}_t}$ is in the jth market environment. We assume $t \le \tau_{j+2} - 1$, i.e., t is in the jth or the $(j+1)$th market environment. We denote $\mathcal{G}_t = \{\pi_s^{\mathrm{CU}} : s \in \{1, \cdots, t\}\}$. Under our change-point detection algorithm, we have the following results:*

1. *Suppose* $\tau_{j+1} \in \mathcal{N}_0$, *with Equation* (1) *that holds for local market* m *and price* p_k. *If* $t \geq \hat{\tau}_{\hat{j}_t} + (w_0 - 1)\,L + 2K$ *and* $t - \frac{w_0}{2} L < \tau_{j+1} \leq t - \left(\frac{w_0}{2} - 1\right) L$, *then*

$$\mathbb{P}\left(\left|\bar{X}_t^{m,0} - \bar{X}_{t-w_0L/2}^{m,0}\right| > \epsilon_0 \middle| \mathcal{G}_t\right) \geq 1 - \underbrace{2\exp\left(-\frac{1}{2}\left(\left(\epsilon_0 - 2\bar{b}\frac{w_0 L}{T}\right)^+\right)^2 w_0\right)}_{U^0}.$$

2. *Suppose* $\tau_{j+1} \in \mathcal{N}_1$, *with Equation* (2) *that holds for all local markets and prices and* (3) *that holds for local market* m *and price* p_k. *If* $t \geq \hat{\tau}_{\hat{j}_t} + (w_1 - 1)\,L + 2K$ *and* $t - \frac{w_1}{3} L < \tau_{j+1} \leq t - \left(\frac{w_1}{3} - 1\right) L$, *then*

$$\mathbb{P}\left(\left|\tfrac{2T}{\left(\frac{w_1}{3}-1\right)L}\left(\left(\bar{X}_t^{m,1} - \bar{X}_{t-w_1L/3}^{m,1}\right) - \left(\bar{X}_{t-w_1L/3}^{m,1} - \bar{X}_{t-2w_1L/3}^{m,1}\right)\right)\right| > \epsilon_1 \middle| \mathcal{G}_t\right)$$
$$\geq 1 - \underbrace{2\exp\left(-\frac{1}{12}\left(\frac{1}{3} - \frac{1}{w_1}\right)^3 \epsilon_1^2 \frac{w_1^3 L^2}{T^2}\right)}_{U^1}.$$

3. *Suppose* $t \leq \tau_{j+1} - 1$.
   (a) *If* $t \geq \hat{\tau}_{\hat{j}_t} + (w_0 - 1)\,L + 2K$, *then*

$$\mathbb{P}\left(\left|\bar{X}_t^{m,0} - \bar{X}_{t-w_0L/2}^{m,0}\right| > \epsilon_0 \middle| \mathcal{G}_t\right) \leq U^0\ .$$

   (b) *If* $t \geq \hat{\tau}_{\hat{j}_t} + (w_1 - 1)\,L + 2K$, *then*

$$\mathbb{P}\left(\left|\frac{2T}{\left(\frac{w_1}{3}-1\right)L}\left(\left(\bar{X}_t^{m,1} - \bar{X}_{t-w_1L/3}^{m,1}\right) - \left(\bar{X}_{t-w_1L/3}^{m,1} - \bar{X}_{t-2w_1L/3}^{m,1}\right)\right)\right| > \epsilon_1 \middle| \mathcal{G}_t\right) \leq U^1.$$

Part 1 entails that when detecting a zeroth-order change-point that exists, if the size of data used for the detection (w_0) is large, and the time window for the detection is much shorter than the length of the entire horizon ($\frac{w_0 L}{T}$ is small), then our algorithm ensures a high probability of detection. Part 2 entails that when detecting a first-order change-point that exists, if the size of data used for the detection (w_1) is large, and the time window for the detection is not much shorter than the length of the entire horizon ($\frac{w_1 L}{T}$ is not too small), such that $\frac{w_1^3 L^2}{T^2}$ is large enough, then our algorithm ensures a high probability of detection. Part 3(a) entails that when detecting a zeroth-order change-point that does not exist, if the size of data used for the detection (w_0) is large, and the time window for the detection is much shorter than the length of the entire horizon ($\frac{w_0 L}{T}$ is small), then our algorithm guarantees a high probability of avoiding a false detection. Part 3(b) entails that while detecting a first-order change-point that does not exist, if the size of data used for the detection (w_1) is large and the time window for the detection is not much shorter than the length of the entire horizon ($\frac{w_1 L}{T}$ is not too small), such that $\frac{w_1^3 L^2}{T^2}$ is large enough, then our algorithm guarantees a high probability of avoiding a false detection.

4.3. Sub-algorithm 2: Time-adjusted UCB Algorithm

In this subsection, we present the second sub-algorithm in the CU algorithm, the *time-adjusted UCB algorithm*. We begin with providing an overview of the key idea of this sub-algorithm.

After detecting a change-point, the firm needs to restart from exploring the new market environment and exploit the learning outcome to take optimal actions. It has been extensively studied that the classical UCB algorithm is effective in doing real-time exploration and exploitation. However, we cannot directly implement the classical UCB algorithm. Recall that this algorithm is designed for the demand process that is stationary over time. However, our model allows demand to dynamically evolve over time between any two consecutive change-points. Hence, for each given price, the empirical mean of the historic sales data under this price gives us a *biased estimate* of the expected demand under this price in the current period. To correct this bias, we introduce a *time-adjusted factor* into the classical UCB algorithm.

Before we formally present our algorithm, we introduce the following notation that facilitates us to present this algorithm.

Consider the time interval $\left\{\hat{\tau}_{\hat{j}_t}, \cdots, t-1\right\}$. Denote by

$$N_{\hat{\tau}_{\hat{j}_t}, t-1}(k) = \sum_{t'=\hat{\tau}_{\hat{j}_t}}^{t-1} \mathbf{1}\left\{k_{t'} = k\right\}$$

the number of periods that price p_k is posted during this time interval.

Denote by

$$s_t(k) = \frac{1}{N_{\hat{\tau}_{\hat{j}_t}, t-1}(k)} \sum_{t'=\hat{\tau}_{\hat{j}_t}}^{t-1} t' \mathbf{1}\left\{\pi_{t'}^{\mathrm{CU}} = k\right\}$$

the average of the period indexes during this time interval that price p_k is posted.

Denote by

$$d_t^m(k) = \frac{1}{N_{\hat{\tau}_{\hat{j}_t}, t-1}(k)} \sum_{t'=\hat{\tau}_{\hat{j}_t}}^{t-1} X_{t'}^m \mathbf{1}\left\{\pi_{t'}^{\mathrm{CU}} = k\right\}$$

the average per-period sales under price p_k on each local market m during this time interval.

Denote by

$$\mathcal{T}_t^{(1)} = \left\{ t' \in \left\{\hat{\tau}_{\hat{j}_t}, \cdots, t-1\right\} : \pi_{t'}^{\mathrm{CU}} = k_t, N_{\hat{\tau}_{\hat{j}_t}, t'}(k_t) \leq \left\lfloor \frac{N_{\hat{\tau}_{\hat{j}_t}, t-1}(k_t)}{2} \right\rfloor \right\}$$

and

$$\mathcal{T}_t^{(2)} = \left\{ t' \in \left\{\hat{\tau}_{\hat{j}_t}, \cdots, t-1\right\} : \pi_{t'}^{\mathrm{CU}} = k_t, N_{\hat{\tau}_{\hat{j}_t}, t'}(k_t) > \left\lfloor \frac{N_{\hat{\tau}_{\hat{j}_t}, t-1}(k_t)}{2} \right\rfloor \right\}$$

the first and the second half of the collection of periods in $\left\{\hat{\tau}_{\hat{j}_t}, \cdots, t-1\right\}$ at which the price with index k_t is posted, respectively.

Now, we are ready to present our *time-adjusted UCB algorithm.*

Sub-algorithm 2: Time-adjusted UCB Algorithm

1. For each price p_k and the associated collection of times in $\left\{\hat{\tau}_{\hat{j}_t}, \cdots, t-1\right\}$ at which she posts this price, the firm computes $s_t(k)$ and $d_t^m(k)$.
2. For each price p_k and each local market m, the firm makes time adjustment on $d_t^m(k)$ to estimate the sales probability at time t.
   (a) The firm computes the index of a price that is posted for the most number of periods during $\left\{\hat{\tau}_{\hat{j}_t}, \cdots, t-1\right\}$:

$$k_t \in \arg\max_k N_{\hat{\tau}_{\hat{j}_t}, t-1}(k).$$

   (b) For any time s and each local market m, the firm computes an *estimated time adjustment factor* as

$$\hat{\delta}^m(s,t) = \frac{\frac{t-s_t^{(1)}}{s_t^{(2)}-s_t^{(1)}} d_t^{m,(2)} - \frac{t-s_t^{(2)}}{s_t^{(2)}-s_t^{(1)}} d_t^{m,(1)}}{\frac{s-s_t^{(1)}}{s_t^{(2)}-s_t^{(1)}} d_t^{m,(2)} - \frac{s-s_t^{(2)}}{s_t^{(2)}-s_t^{(1)}} d_t^{m,(1)}},$$

   where for $i \in \{1,2\}$,

$$s_t^{(i)} = \frac{1}{|\mathcal{T}_t^{(i)}|} \sum_{t' \in \mathcal{T}_t^{(i)}} t'$$

   and

$$d_t^{m,(i)} = \frac{1}{|\mathcal{T}_t^{(i)}|} \sum_{t' \in \mathcal{T}_t^{(i)}} X_{t'}^m.$$

   (c) The firm estimates the sales probability under each price p_k on each local market m at time t according to

$$D_t^m(k) = d_t^m(k)\,\hat{\delta}^m(s_t(k), t).$$

3. The firm implements a UCB algorithm.

(a) For each price p_k, the firm computes its UCB expected revenue at time t according to

$$U_t(k) = \max\left\{ p_k \sum_{m=1}^{M} D_t^m(k) + \beta M p_k \sqrt{\frac{\log T}{N_{\hat{\tau}_{\hat{j}_t}, t-1}(k)}} + 2\beta M p_k \sqrt{\frac{\log T}{N_{\hat{\tau}_{\hat{j}_t}, t-1}(k_t)}}, 0 \right\}. \tag{6}$$

(b) The firm posts a price that maximizes the UCB expected revenue at time t:

$$\pi_t^{\mathrm{CU}} \in \arg\max_k U_t(k).$$

We shall discuss several observations of this algorithm as follows. Note that our objective of developing this algorithm is to make a good exploration-exploitation balance when the global market environment is stationary. Therefore, in the following discussions, we restrict to the case that $\tau_j \le \hat{\tau}_{\hat{j}_t} \le t \le \tau_{j+1} - 1$ for some $j \in \{0, \cdots, J\}$, i.e., the global market environment does not change during the time interval that we use in this algorithm.

1. In Step 1, the average sales quantity has the property that $\mathsf{E}[d_t^m(k) | \mathcal{F}_t] = \left(a_j^m + b_j^m \frac{s_t(k)}{T}\right) \theta_j^m(k)$. Therefore, the firm may use $d_t^m(k)$ to estimate the sales probability on local market m under price p_k at time $s_t(k)$.
2. Recall that each local market's market size may evolve over time (b_j^m may not be equal to zero). Hence, $d_t^m(k)$ is a biased estimate of the sales probability at time t if $b_j^m \neq 0$. This bias that arises from the market size's trend over time can be corrected by multiplying $d_t^m(k)$ with a time adjustment factor, defined as the ratio of local market m's market size at time t to its market size at time $s_t(k)$: $\delta^m(s_t(k), t) = \frac{a_j^m + b_j^m \frac{t}{T}}{a_j^m + b_j^m \frac{s_t(k)}{T}}$.

   Since the firm does not know the values of the market size parameters a_j^m and b_j^m, she needs to estimate this time adjustment factor. The reason that $\hat{\delta}^m(s,t)$ is a good estimate of $\delta^m(s,t)$ is as follows. The expectation of the numerator in $\hat{\delta}^m(s,t)$ is equal to

$$\begin{aligned} &\mathsf{E}\left[\frac{t - s_t^{(1)}}{s_t^{(2)} - s_t^{(1)}} d_t^{m,(2)} - \frac{t - s_t^{(2)}}{s_t^{(2)} - s_t^{(1)}} d_t^{m,(1)} \middle| \mathcal{G}_t \right] \\ =& \frac{t - s_t^{(1)}}{s_t^{(2)} - s_t^{(1)}} \left(a_j^m + b_j^m \frac{s_t^{(2)}}{T} \right) \theta_j^m(k_t) - \frac{t - s_t^{(2)}}{s_t^{(2)} - s_t^{(1)}} \left(a_j^m + b_j^m \frac{s_t^{(1)}}{T} \right) \theta_j^m(k_t) \\ =& \left(a_j^m + b_j^m \frac{t}{T} \right) \theta_j^m(k_t). \end{aligned}$$

   Similarly, the expectation of the denominator in $\hat{\delta}^m(s,t)$ is equal to $\left(a_j^m + b_j^m \frac{s}{T}\right) \theta_j^m(k_t)$. Therefore, if the firm collects enough sales data under price p_{k_t}, then $\hat{\delta}^m(s,t)$ is very close to $\delta^m(s,t)$.

   Recall from Step 2(a) in this algorithm that the price used to estimate $\delta^m(s,t)$, p_{k_t}, is selected to be the one that has been posted for the most number of times. Hence, the number of sales data collected on each local market under price p_{k_t} is guaranteed to be no less than $\lceil \frac{t - \hat{\tau}_{\hat{j}_t}}{K} \rceil$. Therefore, this prevents a large estimation error of $\delta^m(s,t)$.

3. The presence of the estimated time adjustment factor $\hat{\delta}^m(s,t)$ ensures that $D_t^m(k)$ (defined in Step 3(a)) is a good estimate of the sales probability at time t. Therefore, the first term of the UCB function $U_t(\cdot)$ in Equation (6) is a good estimate of the firm's expected total revenue under price p_k at time t.
4. The UCB function $U_t(\cdot)$ in Equation (6) consists of two confidence interval terms, $\beta M p_k\sqrt{\frac{\log T}{N_{\hat{\tau}_{\hat{j}_t},t-1}(k)}}$ and $2\beta M p_k\sqrt{\frac{\log T}{N_{\hat{\tau}_{\hat{j}_t},t-1}(k_t)}}$. The intuition of having these two confidence interval terms is as follows. First, on each local market m, the firm uses the sales data under price p_k to estimate the sales probability under this price at time $s_t(k)$. Because the size of this dataset is $N_{\hat{\tau}_{\hat{j}_t},t-1}(k)$, the length of the confidence interval of this estimation is in the order of $\frac{1}{\sqrt{N_{\hat{\tau}_{\hat{j}_t},t-1}(k)}}$. Hence, we have the first confidence interval term. Second, on each local market m, the firm uses the sales data under price p_{k_t} to estimate the time adjustment factor $\delta^m(s,t)$. Because the size of this dataset is $N_{\hat{\tau}_{\hat{j}_t},t-1}(k_t)$, the length of the confidence interval of this estimation is in the order of $\frac{1}{\sqrt{N_{\hat{\tau}_{\hat{j}_t},t-1}(k_t)}}$. Hence, we have the second confidence interval term.

It is worth noting that implementing our proposed CU algorithm (including two sub-algorithms in the CD algorithm) only requires the firm to use the historic sales and price information. Therefore, our CU algorithm can be easily implemented in practice.

5. Performance of the CU Algorithm

In this section, we present and analyze the performance of the CU algorithm.

5.1. Main Result

In this subsection, we present the following main result of the regret under the CU algorithm.

THEOREM 2. *Suppose for all $j \in \{0,\cdots,J\}$,*

$$\tau_{j+1}-\tau_j \in \begin{cases} \Omega\left(T^{1/2}(\log T)^{1/2+\delta_i}\right) & \textit{if } j,j+1\in\mathcal{N}_0 \\ \Omega\left(T^{5/6}(\log T)^{1/6+\delta_i}\right) & \textit{otherwise} \end{cases},$$

with $\delta_i>0$. If the firm chooses $\epsilon_0=O(1)$, $\epsilon_1=O(1)$, $L=\max\left\{\lceil T^{1/2}/(\log T)^{1/2}\rceil, K+1\right\}$, $w_0=2\lceil 2\log T/\epsilon_0^2\rceil$, $w_1=3\lceil 3T^{1/3}(\log T)^{2/3}/\epsilon_1^{2/3}\rceil$,[4] *$\beta=36K/\underline{d\theta}$, then*

$$\begin{aligned}\text{Regret}^{\text{CU}}(T) &\le \begin{cases} O\left(\frac{K}{\Delta_1^{4/3}}T^{2/3}(\log T)^{1/3}|\mathcal{N}_1|\right) & \textit{if } \mathcal{N}_1\neq\emptyset \\ O\left(\left(K^{3/2}+\frac{J^{1/2}}{\Delta_0^2}\right)J^{1/2}T^{1/2}(\log T)^{1/2}\right) & \textit{if } \mathcal{N}_1=\emptyset \end{cases} \\ &= \begin{cases} O\left(T^{2/3}(\log T)^{1/3}\right) & \textit{if } \mathcal{N}_1\neq\emptyset \\ O\left(T^{1/2}(\log T)^{1/2}\right) & \textit{if } \mathcal{N}_1=\emptyset \end{cases}.\end{aligned}$$

[4] Parameters L, w_0 and w_1 are upper bounded by T.

The upper bounds on the regret under the CU algorithm that we establish above enjoy the following salient features.

1. Our established regret upper bounds attain the regret lower bounds established in Theorem 1 (up to logarithmic factors), both in the case that at least one change-point is a first-order change-point and in the case that all change-points are zeroth-order change-points. This entails that our upper bounds are tight, and our CU algorithm achieves the regret lower bounds.
2. Note that the number of data used to detect a first-order change-point, $w_1 = O\left(T^{1/3}(\log T)^{2/3}\right)$, is in a larger order than the number of data used to detect a zeroth-order change-point, $w_0 = O(\log T)$. In addition, the regret upper bound in the case that at least one change-point is a first-order change-point, $O\left(T^{2/3}(\log T)^{1/3}\right)$, is also in a larger order than the one in the case that all change-points are zeroth-order change-points, $O\left(T^{1/2}(\log T)^{1/2}\right)$.

   The intuition of these observations is as follows. Around each first-order change-point, the demand function smoothly evolves over time, and the change only occurs in its evolving speed. Such slow-speed change cannot be easily observed in a short time window. Hence, the firm needs to collect more sales data over a longer time window to make a successful detection. In addition, during the long detection time window, she still uses her old knowledge accumulated before this first-order change-point to do exploration-exploitation. Hence, she fails to quickly refresh her knowledge to restart to learn the new parameters that modulate the market environment that starts from this change-point. As a result, she loses much revenue due to the slow response to the appearance of a first-order change-point.

The following corollary studies the impact of the scale of the number of change-points with respect to the horizon on the regret upper bound.

COROLLARY 1. *Consider all conditions given in Theorem 2.*

1. *If the number of first-order change-points is scaled with T as $|\mathcal{N}_1| = O(T^\sigma)$ with $\sigma \in \left(0, \frac{1}{3}\right)$, then*
$$\text{Regret}^{\text{CU}}(T) \le O\left(T^{\frac{2}{3}+\sigma}(\log T)^{1/3}\right).$$
2. *If the system has no first-order change-point ($|\mathcal{N}_1| = 0$) and the number of change-points (the same as the number of zeroth-order change-points) is scaled with T as $J = O(T^\sigma)$ with $\sigma \in \left(0, \frac{1}{2}\right)$, then*
$$\text{Regret}^{\text{CU}}(T) \le O\left(T^{\frac{1}{2}+\sigma}(\log T)^{1/2}\right).$$

5.2. Performance Analysis

In this subsection, we provide an outline of the proof of Theorem 2.

We observe that for large T, we have the following properties.

LEMMA 2. *Suppose all conditions in Theorem 2 are satisfied. Then, there exists $\underline{T}$, such that for $T \geq \underline{T}$,*

1. *For any $j \in \{0, \cdots, J\}$,*

$$\begin{aligned}\tau_{j+1} - \tau_j \geq\ & \frac{w_0 L}{2} \mathbf{1}\{j \in \mathcal{N}_0\} + \frac{w_1 L}{3} \mathbf{1}\{j \in \mathcal{N}_1\} \\ & + \frac{w_0 L}{2} \mathbf{1}\{j+1 \in \mathcal{N}_0\} + \frac{2 w_1 L}{3} \mathbf{1}\{j+1 \in \mathcal{N}_1\} \\ & + 2K.\end{aligned}$$

2.

$$\frac{w_1 L}{3T} \leq \frac{d\theta}{8(2K+1)\bar{b}}.$$

Recall that Theorem 2 is for the regime that T is sufficiently large. Hence, in the rest of this subsection, we restrict to $T \geq \underline{T}$, such that all properties in Lemma 2 are satisfied.

To simplify the notation in our analysis, we hereafter drop the superscript "CU" from notation π^{CU}. Under the CU algorithm, the regret comes from two sources: revenue loss from change-point detection and revenue loss from exploration-exploitation due to the firm's no information on the parameters that modulate the system ex-ante. This reasoning motivates us to proceed with our proof in the following four steps.

Step 1: Establish an upper bound on $\mathrm{Regret}^{\mathrm{CU}}(T)$ that is expressed in terms of the aforementioned two sources of revenue losses.

To do so, we need to classify all change-point detection events into well-detected events and badly-detected events. We denote by

$$\mathcal{A}_j = \{\hat{\tau}_j \geq \tau_j\}, \ \forall\ j \in \{1, \cdots, J+1\}$$

the event that the firm does not falsely detect the jth change-point earlier than it occurs.

We denote by

$$\mathcal{B}_j = \left\{\hat{\tau}_j \leq \tau_j + \frac{w_0 L}{2} \mathbf{1}\{j \in \mathcal{N}_0\} + \frac{w_1 L}{3} \mathbf{1}\{j \in \mathcal{N}_1\}\right\}, \ \forall\ j \in \{0, \cdots, J+1\}$$

the event that the firm successfully detects the jth change-point in a given time window after this change occurs. We note that the length of the detection time window depends on whether the change-point is a zeroth-order or a first-order change-point.

Following from the definition of these events, we establish the following upper bound on the regret under policy π^{CU}:

Lemma 3. *We have*

$$\mathrm{Regret}^{\mathrm{CU}}(T) \leq \underbrace{M\bar{p}T\sum_{j=1}^{J+1}\mathbb{P}\left(\mathcal{A}_j^c\middle|\cap_{l=1}^{j-1}(\mathcal{A}_l\cap\mathcal{B}_l)\right)}_{H^1} + \underbrace{M\bar{p}T\sum_{j=1}^{J}\mathbb{P}\left(\mathcal{B}_j^c\middle|\cap_{l=1}^{j-1}(\mathcal{A}_l\cap\mathcal{B}_l)\cap\mathcal{A}_j\right)}_{H^2}$$
$$+\underbrace{\mathsf{E}\left[\sum_{t=1}^{T}g_t(k_t^*)-g_t(\pi_t)\middle|\cap_{l=1}^{J}(\mathcal{A}_l\cap\mathcal{B}_l)\cap\mathcal{A}_{J+1}\right]}_{H^3}.$$

The terms H^1 and H^2 measure two types of revenue losses that arise from the change-point detection. To be specific, the term H^1 provides an upper bound on the total revenue loss that arises from falsely detecting at least one change-point earlier than it occurs. The term H^2 provides an upper bound on the total revenue loss that arises from failing to detect at least one change-point in the given time window after this change occurs. The term H^3 provides an upper bound on the revenue loss that arises from exploration-exploitation. This is due to the firm's lack of information on the parameters, which modulate the demand functions ex-ante, conditional on that the firm does not incur the aforementioned two change-point detection errors.

Step 2: Establish upper bounds on the revenue losses that arise from change-point detection, H^1 and H^2.

We have the following results:

Lemma 4.

1.

$$H^1 \leq M\bar{p}K(J+1)\frac{T^2}{L}\left(U^0+U^1\right).$$

2.

$$H^2 \leq M\bar{p}T\left(U^0|\mathcal{N}_0|+U^1|\mathcal{N}_1|\right).$$

We provide a sketch of the proof of this lemma. For Part 1, we show that in each global market environment j, at each time that the firm performs the change-point detection to detect the $(j+1)$th change-point, she has a small probability of detecting a change falsely. For Part 2, we show that if a zeroth-order (resp. first-order) change occurs at time τ_j on some local market m under some price p_k, then there is a high probability that this change is successfully detected after the firm collects $w_0/2$ (resp. $w_1/3$) sales data from local market m under price p_k since time τ_{j+1}.

Step 3: Establish an upper bound on the revenue loss that arises from exploration-exploitation, H^3.

We need to proceed in five steps, Steps 3.1-3.5, to establish a tractable upper bound on H^3.

Step 3.1: Express H^3 in an equivalent way.

To facilitate the analysis, we introduce the following auxiliary algorithm, $\pi^{\mathrm{CU},t}$. This algorithm is almost the same as π^{CU}, except that the firm sets time t as a change-point and stops performing the change-point detection algorithm (sub-algorithm 1) since time t. The following lemma expresses H^3 in an equivalent way in terms of $\pi^{\mathrm{CU},t}$, rather than π^{CU}.

Lemma 5.

$$H^3 = \mathsf{E}_{\{\hat{\tau}_1,\cdots,\hat{\tau}_J\}}\left[\sum_{j=0}^{J} H_j^3 \middle| \cap_{l=1}^{J}\left(\mathcal{A}_l \cap \mathcal{B}_l\right) \cap \mathcal{A}_{J+1}\right],$$

where the subscript of the expectation $\{\hat{\tau}_1,\cdots,\hat{\tau}_J\}$ means we take expectation w.r.t. all possible change-point times, and

$$H_j^3 \triangleq \mathsf{E}\left[\sum_{t=\hat{\tau}_j}^{\hat{\tau}_{j+1}-1} g_t\left(k_t^*\right) - g_t\left(\pi_t^{\mathrm{CU},\hat{\tau}_j}\right) \middle| \hat{\tau}_j, \hat{\tau}_{j+1}\right], \ \forall\ j \in \{0,\cdots,J\},$$

with a convention that $\hat{\tau}_{J+1} = T+1$ if $\mathbf{1}\{\cap_{l=1}^{J}\left(\mathcal{A}_l \cap \mathcal{B}_l\right) \cap \mathcal{A}_{J+1}\} = 1$.

Recall that policy $\pi^{\mathrm{CU},t}$ does not perform change-point detection since time t. Therefore, this lemma entails that we can compute each function H_j^3 without worrying about the change-point detection outcome after time $\hat{\tau}_j$. This allows us to purely focus on analyzing the effect of exploration-exploitation on H_j^3. To lighten notation, in the rest of this subsection, we drop the superscript "CU" from notation $\pi^{\mathrm{CU},t}$. In addition, to further lighten notation, we drop the condition in $\mathsf{E}\left[\cdot|\hat{\tau}_j,\hat{\tau}_{j+1}\right]$ that defines H_j^3.

Recall that for any $j \in \{0,\cdots,J\}$, conditional on $\mathcal{B}_j \cap \mathcal{A}_{j+1} \cap \mathcal{B}_{j+1}$, Lemma 2 implies $\hat{\tau}_{j+1} - \hat{\tau}_j \geq 2K$. Therefore, we can decompose H_j^3 in the following way:

$$H_j^3 = \underbrace{\mathsf{E}\left[\sum_{t=\hat{\tau}_j}^{\hat{\tau}_j+2K-1} g_t\left(k_t^*\right) - g_t\left(\pi_t^{\hat{\tau}_j}\right)\right]}_{H_{j,1}^3} + \underbrace{\mathsf{E}\left[\sum_{t\in\mathcal{T}_j^{\mathrm{US}}} g_t\left(k_t^*\right) - g_t\left(\pi_t^{\hat{\tau}_j}\right)\right]}_{H_{j,2}^3} + \underbrace{\mathsf{E}\left[\sum_{t\in\mathcal{T}_j^{\mathrm{UCB}}} g_t\left(k_t^*\right) - g_t\left(\pi_t^{\hat{\tau}_j}\right)\right]}_{H_{j,3}^3},$$

where

$$\mathcal{T}_j^{\mathrm{US}} = \{t \in \{\hat{\tau}_j + 2K,\cdots,\hat{\tau}_{j+1}-1\} : \mathrm{mod}\,(t,L) \in \{1,\cdots,K\}\}$$

denotes the collection of times in $\{\hat{\tau}_j + 2K,\cdots,\hat{\tau}_{j+1}-1\}$ at which the firm performs the uniform sampling over all prices (Step 2(b) in the CU algorithm), and

$$\mathcal{T}_j^{\mathrm{UCB}} = \{t \in \{\hat{\tau}_j + 2K,\cdots,\hat{\tau}_{j+1}-1\} : \mathrm{mod}\,(t,L) \notin \{1,\cdots,K\}\}$$

denotes the collection of times in $\{\hat{\tau}_j + 2K, \cdots, \hat{\tau}_{j+1} - 1\}$ at which the firm performs the time-adjusted UCB algorithm (sub-algorithm 2).

The term $H^3_{j,1}$ denotes the revenue loss that arises from uniformly sampling each price twice immediately after the jth change-point is detected as $\hat{\tau}_j$. The term $H^3_{j,2}$ denotes the revenue loss that arises from uniformly sampling all prices during $\{\hat{\tau}_j + 2K, \cdots, \hat{\tau}_{j+1} - 1\}$. The term $H^3_{j,3}$ denotes the revenue loss that arises from implementing the time-adjusted UCB algorithm during $\{\hat{\tau}_j + 2K, \cdots, \hat{\tau}_{j+1} - 1\}$.

Step 3.2: Establish upper bounds on $H^3_{j,1}$ and $H^3_{j,2}$, respectively.

We have the following results:

LEMMA 6. *For any $j \in \{0, \cdots, J\}$, conditional on $\mathcal{B}_j \cap \mathcal{A}_{j+1} \cap \mathcal{B}_{j+1}$,*

1.

$$H^3_{j,1} \le 2M\bar{p}K.$$

2.

$$H^3_{j,2} \le M\bar{p}K\left(\frac{\hat{\tau}_{j+1} - \hat{\tau}_j}{L} + 1\right).$$

Step 3.3: Establish an upper bound on $H^3_{j,3}$ for the case that $j+1 \in \mathcal{N}_0$.

Now, we upper bound $H^3_{j,3}$. It is worth pointing out that we have to adopt two different approaches and thus establish two different upper bounds on $H^3_{j,3}$ with respect to whether the $(j+1)$th change-point τ_{j+1} is a zeroth-order or a first-order change-point. Hence, in this step, we establish an upper bound on $H^3_{j,3}$ for the case that time τ_{j+1} is a zeroth-order change-point, $j+1 \in \mathcal{N}_0$. We will analyze the other case that time τ_{j+1} is a first-order change-point, $j+1 \in \mathcal{N}_1$, in Step 3.4.

We begin with presenting the following intermediate result that decomposes the cumulative revenue loss during $\{\hat{\tau}_j, \cdots, \hat{\tau}_{j+1} - 1\}$ into two parts: revenue loss before the change-point τ_{j+1} (during $\{\hat{\tau}_j, \cdots, \tau_{j+1} - 1\}$) and the revenue loss after this change-point (during $\{\tau_{j+1}, \cdots, \hat{\tau}_{j+1} - 1\}$).

LEMMA 7. *For any $j+1 \in \mathcal{N}_0$, conditional on $\mathcal{B}_j \cap \mathcal{A}_{j+1} \cap \mathcal{B}_{j+1}$*[5]*,*

$$\begin{aligned} H^3_{j,3} \le M\bar{p} \sum_{t \in \mathcal{T}^{\mathrm{UCB}}_{j,<}} \mathbb{P}\left(g_t\left(k^*_t\right) \ge U_t\left(k^*_t\right)\right) + \sum_{t \in \mathcal{T}^{\mathrm{UCB}}_{j,<}} \mathsf{E}\left[U_t\left(\pi^{\hat{\tau}_j}_t\right) - L_t\left(\pi^{\hat{\tau}_j}_t\right)\right] \\ + M\bar{p} \sum_{t \in \mathcal{T}^{\mathrm{UCB}}_{j,<}} \mathbb{P}\left(L_t\left(\pi^{\hat{\tau}_j}_t\right) \ge g_t\left(\pi^{\hat{\tau}_j}_t\right)\right) + M\bar{p}\frac{w_0 L}{2}, \end{aligned} \tag{7}$$

[5] To lighten notation, we drop this condition in all probability $\mathbb{P}(\cdot)$ and expectation $\mathsf{E}[\cdot]$ that appear in this lemma.

where

$$\mathcal{T}_{j,<}^{\mathrm{UCB}} = \left\{ t < \tau_{j+1} : t \in \mathcal{T}_j^{\mathrm{UCB}} \right\}$$

and

$$L_t(k) = \min\left\{ p_k \sum_{m=1}^{M} D_t^m(k) - \beta M p_k \sqrt{\frac{\log T}{N_{\hat{\tau}_j, t-1}(k)}} - 2\beta M p_k \sqrt{\frac{\log T}{N_{\hat{\tau}_j, t-1}(k_t)}}, M\bar{p} \right\}, \ \forall\ k.$$

On the right-hand side of Equation (7), the first three terms jointly provide an upper bound on the revenue loss that arises from implementing the time-adjusted UCB algorithm (Step 2(c) in the CU algorithm) before the change-point τ_{j+1}. Note that we introduce a new function $L_t(\cdot)$ in the second and the third terms. Analogous to the upper confidence bound function $U_t(\cdot)$, we hereafter call this function the *lower confidence bound (LCB)* function.

The fourth term provides an upper bound on the revenue loss accumulated since the change-point τ_{j+1}. Because this change-point is a zeroth-order change-point, the demand function has an abrupt change at this point. The firm's knowledge of the parameters that modulate the jth market environment is gained before this change-point. Thus, it becomes useless after the system enters the $(j+1)$th market environment. Hence, at each time of transition, the firm's expected revenue garnered under the decision made by using the knowledge gained from the jth market environment may be far away from the optimal revenue. Such transition happens after the $(j+1)$th market environment change and before this change is detected at time $\hat{\tau}_{j+1}$, Note that $M\bar{p}$ is an upper bound on the expected revenue garnered at each time t, $g_t(\cdot)$. Therefore, we simply use this quantity as an upper bound on the revenue loss at each time during $\{\tau_{j+1}, \cdots, \hat{\tau}_{j+1} - 1\}$.

Now, we establish upper bounds on the first three terms on the right-hand side of Equation (7), respectively.

LEMMA 8. *For any $j+1 \in \mathcal{N}_0$, conditional on $\mathcal{B}_j \cap \mathcal{A}_{j+1} \cap \mathcal{B}_{j+1}$[6],*

1.

$$\sum_{t \in \mathcal{T}_{j,<}^{\mathrm{UCB}}} \mathbb{P}\left(g_t(k_t^*) \geq U_t(k_t^*)\right) \leq 6M\left(\hat{\tau}_{j+1} - \hat{\tau}_j\right) T^{-(\beta \underline{d\theta})^2/2\cdot 36^2 K^2} + \frac{2M}{1 - e^{-(\underline{d\theta})^2/72K^3}}.$$

2.

$$\sum_{t \in \mathcal{T}_{j,<}^{\mathrm{UCB}}} \mathsf{E}\left[U_t\left(\pi_t^{\hat{\tau}_j}\right) - L_t\left(\pi_t^{\hat{\tau}_j}\right)\right] = 12\beta M\bar{p}\sqrt{K\left(\hat{\tau}_{j+1} - \hat{\tau}_j\right)\log T}.$$

[6] To lighten notation, we drop this condition in all probability $\mathbb{P}(\cdot)$ and expectation $\mathsf{E}[\cdot]$ that appear in this lemma.

3.

$$\sum_{t\in\mathcal{T}^{\mathrm{UCB}}_{j,<}} \mathbb{P}\left(L_t\left(\pi_t^{\hat{\tau}_j}\right) \geq g_t\left(\pi_t^{\hat{\tau}_j}\right)\right) \leq 6M\left(\hat{\tau}_{j+1}-\hat{\tau}_j\right)T^{-(\beta\underline{d\theta})^2/2\cdot 36^2K^2} + \frac{2M}{1-e^{-(\underline{d\theta})^2/72K^3}}.$$

Part 1 entails that there is a high probability that the UCB function $U_t(\cdot)$ is indeed an upper bound on the expected revenue function $g_t(\cdot)$. Analogously, Part 3 entails that there is a high probability that the LCB function $L_t(\cdot)$ is indeed a lower bound on the expected revenue function $g_t(\cdot)$. Part 2 entails that the confidence intervals introduced in defining the UCB and LCB functions do not lead to too much revenue loss.

Step 3.4: Establish an upper bound on $H^3_{j,3}$ for the case that $j+1\in\mathcal{N}_1$.

Here, we establish an upper bound on $H^3_{j,3}$ for the case that time τ_{j+1} is a first-order change-point, $j+1\in\mathcal{N}_1$. Our approach is quite different from the one for the case that τ_{j+1} is a zeroth-order change-point. Recall from Step 3.3 that if time τ_{j+1} is a zeroth-order change-point, then we adopt an approach to separately upper bound the revenue losses accumulated before and after this time. Because the demand function has an abrupt change at a zeroth-order change-point, the firm does not need to spend too much time to detect the change successfully. With the parameters specified in Theorem 2, the length of time that the firm spends to detect a zeroth-order change-point after it occurs is in the order of $O(w_0L)=O\left(T^{1/2}(\log T)^{1/2}\right)$. Because the expected revenue garnered at each time is finitely bounded by $M\bar{p}$, the cumulative revenue loss after the $(j+1)$th change occurs is also upper bounded by $O\left(T^{1/2}(\log T)^{1/2}\right)$.

However, if we adopt the same proof technique to the current case that the change-point τ_{j+1} is a first-order change-point, then we will get a very loose upper bound on the cumulative revenue loss after the $(j+1)$th change occurs. The intuition is as follows. Because the demand function changes smoothly around a first-order change-point, the firm needs to spend quite a long time to detect the change successfully. With the parameters specified in Theorem 2, the time that the firm spends to detect a first-order change-point after it occurs is in the order of $O(w_1L)=O\left(T^{5/6}(\log T)^{1/6}\right)$. Hence, if we still use $M\bar{p}$ to upper bound the expected revenue garnered at each time since τ_{j+1}, then we get an upper bound on the cumulative revenue loss since τ_{j+1} that is in the order of $O\left(T^{5/6}(\log T)^{1/6}\right)$.

This upper bound is too loose that it makes us impossible to prove Theorem 2 that the regret is upper bounded by $O\left(T^{2/3}(\log T)^{1/3}\right)$. Therefore, we need to adopt a different approach to upper bound $H^3_{j,3}$ if time τ_{j+1} is a first-order change-point.

If time τ_{j+1} is a first-order change-point, then the demand function after τ_{j+1} does not deviate too much from the one before τ_{j+1}. Hence, at each time after τ_{j+1}, the expected revenue garnered under the time-adjusted UCB algorithm that is used for the jth market environment is still quite close to the optimal revenue. Therefore, this motivates us to jointly analyze the revenue losses both before and after τ_{j+1}, rather than analyzing them separately.

To proceed our analysis, we need to introduce the following notation. Denote by

$$\phi_j^m(k) = b_{j+1}^m \theta_{j+1}^m(k) - b_j^m \theta_j^m(k)$$

the first-order change at time τ_{j+1} on local market m under price p_k.

For $t \in \{\hat{\tau}_j, \cdots, \hat{\tau}_{j+1} - 1\}$, we denote by

$$\Delta s_t(k) = \frac{1}{N_{\hat{\tau}_j, t-1}(k)} \sum_{t'=\hat{\tau}_j}^{t-1} (t' - \hat{\tau}_{j+1})^+ \mathbf{1}\left\{\pi_{t'}^{\hat{\tau}_j} = k\right\}$$

the average of the excess times after the change-point τ_{j+1} at which price p_k is posted, and

$$\Delta s_t^{(i)} = \frac{1}{\mathcal{T}_t^{(i)}} \sum_{t' \in \mathcal{T}_t^{(i)}} (t' - \hat{\tau}_{j+1})^+, \forall\, i \in \{1, 2\}.$$

the average of the excess times among the ith half of times that price p_{k_t} is posted.

We denote

$$\bar{g}_t(k) = p_k \sum_{m=1}^{M} \bar{D}_t^m(k),$$

where

$$\begin{aligned} &\bar{D}_t^m(k) \\ &= \left(\left(a_j^m + b_j^m \frac{s_t(k)}{T}\right) \theta_j^m(k) + \phi_j^m(k) \frac{\Delta s_t(k)}{T}\right) \\ &\cdot \frac{\frac{t - s_t^{(1)}}{s_t^{(2)} - s_t^{(1)}} \left(\left(a_j^m + b_j^m \frac{s_t^{(2)}}{T}\right) \theta_j^m(k_t) + \phi_j^m(k_t) \frac{\Delta s_t^{(2)}}{T}\right) - \frac{t - s_t^{(2)}}{s_t^{(2)} - s_t^{(1)}} \left(\left(a_j^m + b_j^m \frac{s_t^{(1)}}{T}\right) \theta_j^m(k_t) + \phi_j^m(k_t) \frac{\Delta s_t^{(1)}}{T}\right)}{\frac{s_t(k) - s_t^{(1)}}{s_t^{(2)} - s_t^{(1)}} \left(\left(a_j^m + b_j^m \frac{s_t^{(2)}}{T}\right) \theta_j^m(k_t) + \phi_j^m(k_t) \frac{\Delta s_t^{(2)}}{T}\right) - \frac{s_t(k) - s_t^{(2)}}{s_t^{(2)} - s_t^{(1)}} \left(\left(a_j^m + b_j^m \frac{s_t^{(1)}}{T}\right) \theta_j^m(k_t) + \phi_j^m(k_t) \frac{\Delta s_t^{(1)}}{T}\right)}. \end{aligned}$$

Hence, if $t \in \{\hat{\tau}_j, \cdots, \tau_{j+1} - 1\}$, then $\bar{g}_t(k) = g_t(k)$. If $t \in \{\tau_{j+1}, \cdots, \hat{\tau}_{j+1} - 1\}$, then $\bar{g}_t(k)$ may deviate from $g_t(k)$ by balancing the demand function parameters in both the jth and the $(j+1)$th market environments, rather than being purely determined by the demand function parameters in the $(j+1)$th market environment as in $g_t(k)$. We will hereafter call $\bar{g}_t(\cdot)$ the *balanced revenue function*. In our subsequent analysis, we will use $\bar{g}_t(\cdot)$, not $g_t(\cdot)$, to establish an upper bound on $H_{j,3}^3$.

The key intuition of using $\bar{g}_t(\cdot)$ to facilitate the analysis is as follows. For $t \in \{\tau_{j+1}, \cdots, \hat{\tau}_{j+1} - 1\}$, computing the UCB function $U_t(\cdot)$ (resp. the LCB function $L_t(\cdot)$) requires the firm to use sales data both before and after the change-point τ_{j+1}. Recall that the balanced revenue function $\bar{g}_t(\cdot)$

is constructed by using the demand functions also both before and after the change-point τ_{j+1}. Therefore, quantifying the gap between $\bar{g}_t(\cdot)$ and $U_t(\cdot)$ (resp. $L_t(\cdot)$) is reduced to quantifying the gap between the actual sales and its mean value at any time that is either before or after the change-point. The result can be easily obtained from Hoeffding's inequality.

Now, we present the following intermediate result:

LEMMA 9. *For any $j+1 \in \mathcal{N}_1$, conditional on $\mathcal{B}_j \cap \mathcal{A}_{j+1} \cap \mathcal{B}_{j+1}$*[7],

$$\begin{aligned} H_{j,3}^3 \leq 33M\bar{p} \sum_{t \in \mathcal{T}_j^{\mathrm{UCB}}} \mathbb{P}\left(\bar{g}_t(k) \geq U_t(k)\right) + \sum_{t \in \mathcal{T}_j^{\mathrm{UCB}}} \mathsf{E}\left[U_t\left(\pi_t^{\hat{\tau}_j}\right) - L_t\left(\pi_t^{\hat{\tau}_j}\right)\right] \\ + 33M\bar{p} \sum_{t \in \mathcal{T}_j^{\mathrm{UCB}}} \mathbb{P}\left(L_t(k) \geq \bar{g}_t(k)\right) + 60M\bar{p}\bar{b}(K+1)\frac{w_1^2 L^2}{T}. \end{aligned} \tag{8}$$

On the right-hand side of Equation (8), the first three terms jointly provide an upper bound on the revenue loss by using the balanced revenue function $\bar{g}_t(\cdot)$ to replace the true revenue function $g_t(\cdot)$. The fourth term is an upper bound on the gap between $\bar{g}_t(\cdot)$ and $g_t(\cdot)$.

Now, we establish upper bounds on the first three terms on the R.H.S. of Equation (8), respectively.

LEMMA 10. *For any $j+1 \in \mathcal{N}_1$, conditional on $\mathcal{B}_j \cap \mathcal{A}_{j+1} \cap \mathcal{B}_{j+1}$*[8],

1.

$$\sum_{t \in \mathcal{T}_j^{\mathrm{UCB}}} \mathbb{P}\left(\bar{g}_t(k_t^*) \geq U_t(k_t^*)\right) \leq 6M\left(\hat{\tau}_{j+1} - \hat{\tau}_j\right) T^{-(\beta \underline{d\theta})^2/2\cdot 36^2 K^2} + \frac{2M}{1 - e^{-(\underline{d\theta})^2/72K^3}}.$$

2.

$$\sum_{t \in \mathcal{T}_j^{\mathrm{UCB}}} \mathsf{E}\left[U_t\left(\pi_t^{\hat{\tau}_j}\right) - L_t\left(\pi_t^{\hat{\tau}_j}\right)\right] = 12\beta M\bar{p}\sqrt{K\left(\hat{\tau}_{j+1} - \hat{\tau}_j\right)\log T}.$$

3.

$$\sum_{t \in \mathcal{T}_j^{\mathrm{UCB}}} \mathbb{P}\left(L_t\left(\pi_t^{\hat{\tau}_j}\right) \geq \bar{g}_t\left(\pi_t^{\hat{\tau}_j}\right)\right) \leq 6M\left(\hat{\tau}_{j+1} - \hat{\tau}_j\right) T^{-(\beta \underline{d\theta})^2/2\cdot 36^2 K^2} + \frac{2M}{1 - e^{-(\underline{d\theta})^2/72K^3}}.$$

Part 1 entails that there is a high probability that the UCB function $U_t(\cdot)$ is indeed an upper bound on the balanced revenue function $\bar{g}_t(\cdot)$. Analogously, Part 3 entails that there is a high probability that the LCB function $L_t(\cdot)$ is indeed a lower bound on the balanced revenue function $\bar{g}_t(\cdot)$. Part 2 entails that the confidence intervals introduced in defining the UCB and LCB functions do not lead to too much revenue loss.

[7] To lighten notation, we drop this condition in all probability $\mathbb{P}(\cdot)$ and expectation $\mathsf{E}[\cdot]$ that appear in this lemma.

[8] To lighten notation, we drop this condition in all probability $\mathbb{P}(\cdot)$ and expectation $\mathsf{E}[\cdot]$ that appear in this lemma.

Step 3.5: Establish a tractable upper bound on H^3.

By jointly using the results of the upper bounds on $H^3_{j,1}$ and $H^3_{j,2}$ (Lemma 6), the upper bound on $H^3_{j,3}$ in the case that $j+1 \in \mathcal{N}_0$ (Lemmas 7 and 8) and the upper bound on $H^3_{j,3}$ in the case that $j+1 \in \mathcal{N}_1$ (Lemmas 9 and 10), we establish the following upper bound on H^3:

Lemma 11.

$$H^3 \le 3M\bar{p}K(J+1) + M\bar{p}K\frac{T}{L} + 132M^2\bar{p}\left(3T^{1-\left(\beta\underline{d}^2\underline{\theta}^2\right)^2/2\cdot 36^2K^2} + \frac{J+1}{1-e^{-(\underline{d\theta})^2/72K^3}}\right)$$
$$+12\beta M\bar{p}\sqrt{K(J+1)T\log T} + M\bar{p}\frac{w_0L}{2}|\mathcal{N}_0| + 60M\bar{p}\bar{b}(K+1)\frac{w_1^2L^2}{T}|\mathcal{N}_1|.$$

Step 4: Complete the proof of Theorem 2.

Now, we are ready to jointly use the established upper bounds on H^1, H^2 and H^3 (Lemmas 4 and 11) to complete the proof of Theorem 2. The detailed proof can be found in Appendix C.

6. Numerical Examples

To demonstrate the validity of our theory, below we present a few examples to demonstrate the revenue as a function of time for a few instances.

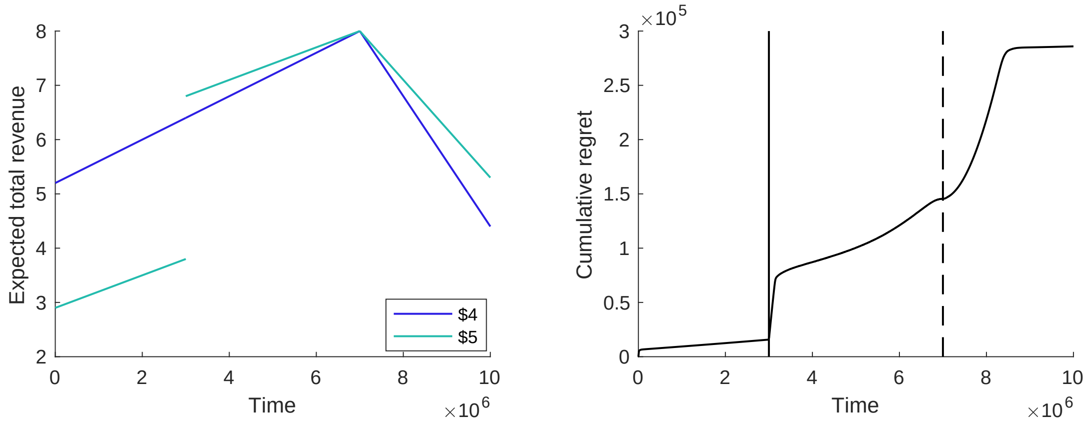
6.1. Benchmark Example

We first present an illustrative example with $M=2$, $K=2$, $J=2$. There is one zero-order and one first-order change in the model, respectively. We illustrate the cumulative regret of our algorithm as a function of time. From the results in Fig. 2, we can observe that whenever there is a change, the regret grows quickly before the change was detected, and the regret is plateaued, and the growth rate is significantly lower after the change has been detected. Moreover, we can observe that the zeroth-order change is much easier to detect than the first-order change, which is reflected in Fig. 2 that the time it takes much longer for the revenue to plateaued after the first-order change compared to the zeroth-order change.

6.2. Impact of the Environment Change

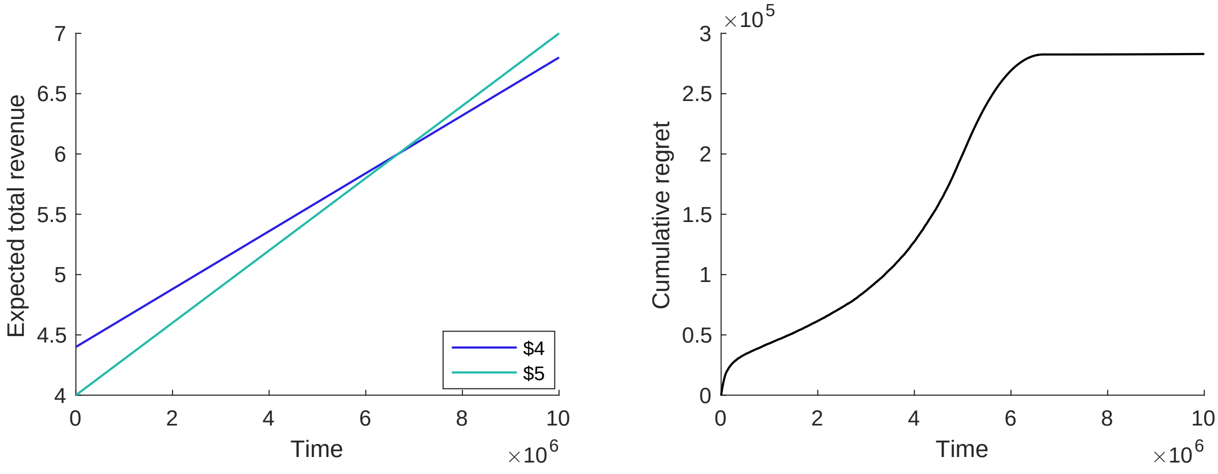
Next we present a few examples to demonstrate the effect of the environment change, with (i) $J=0$, no change, (ii) $J=1$, one zero-order change, (iii) $J=1$, one first-order change, (iv) $J=2$, both zero-order changes; and (v) $J=2$, both first-order changes. We can observe that the algorithm can be adapted to both types of changes.

We first present a benchmark result when $J=0$ there is no change. As shown in Fig. 3, note that the regret converges as time increases. We then study how our algorithm responds to $J=1$ one



(a) (b)

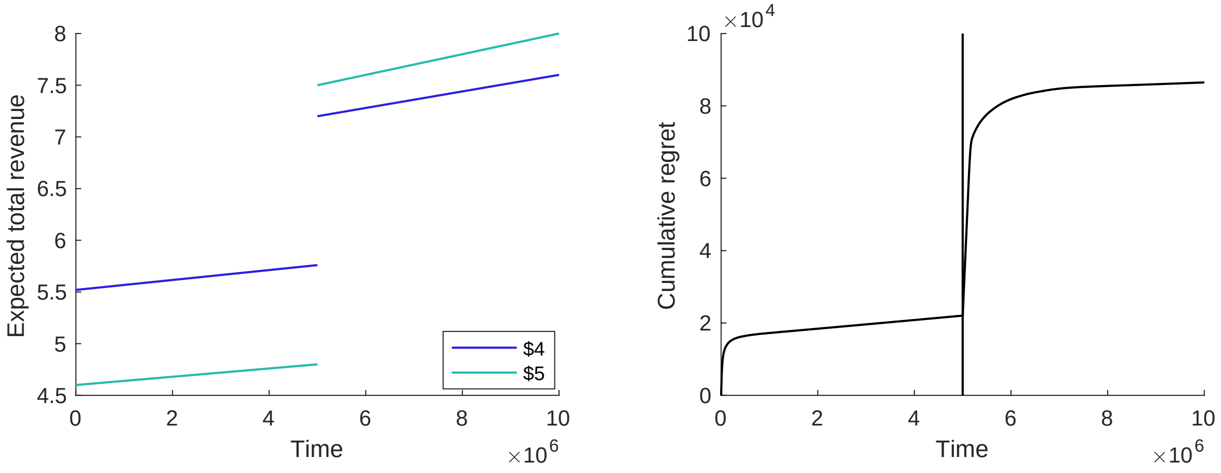
Figure 2 An example of cumulative regret of our algorithm with one zero-order and one first-order changes: $M=2$, $K=2$, $J=2$: (a) Illustration of the change in the model; (b) Cumulative regret of our algorithm: two changes are marked by solid and dashed lines.



(a) (b)

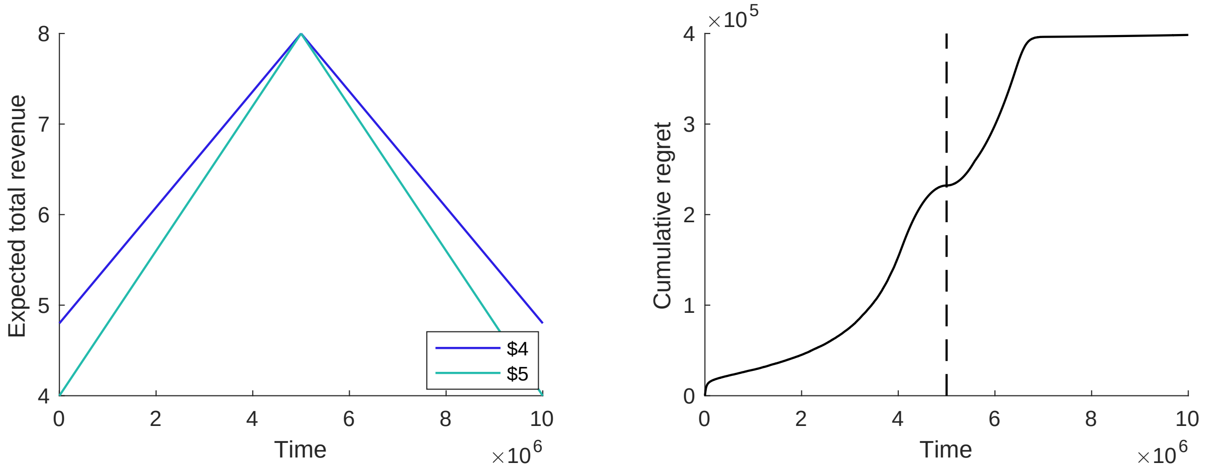
Figure 3 An example of cumulative regret of our algorithm with no change $J=0$: (a) Illustration of the model; (b) Cumulative regret of our algorithm.

zero-order change, shown in Fig. 4. Note that the algorithm adapts to the change and the regret converges as time increases. The next example study the algorithm responds to $J=1$, one first-order change, shown in Fig. 5. We observe that the algorithm responds to first-order change with a smoother change in regret than the zero-order change, which is expected since the first-change change is smooth and gradual whereas the zero-order change is abrupt. Finally, we study the case when there we are $J=2$ changes, when they are both zero-order change (shown in Fig. 6) and both first-order change (shown in Fig. 7). Similar observations can be made that the response to the first-order change results in a more smooth transition in regret. Moreover, in Fig. 7, we can observe that the response to the upward and downward changes can be different.



(a) (b)

Figure 4 An example of cumulative regret of our algorithm with one zero-order change $J = 1$: (a) Illustration of the change in the model; (b) Cumulative regret of our algorithm: the change is marked by a solid line.



(a) (b)

Figure 5 An example of cumulative regret of our algorithm with one first-order change $J = 1$: (a) Illustration of the change in the model; (b) Cumulative regret of our algorithm: the change is marked by a dashed line.

6.3. Impact of Different Number of Markets

In this setting, we study the impact of the number of markets by adding new markets to the benchmark model. For $M = 2, 3, 4$, the model is designed such that the difference between the expected revenue of the two price options remains the same for a fair comparison. From the simulation result shown in Fig. 8, we can see that the increase in the number of markets results in a slightly larger regret. We note that when M is larger, in general, the regret is higher, which is expected as predicted by our theory.

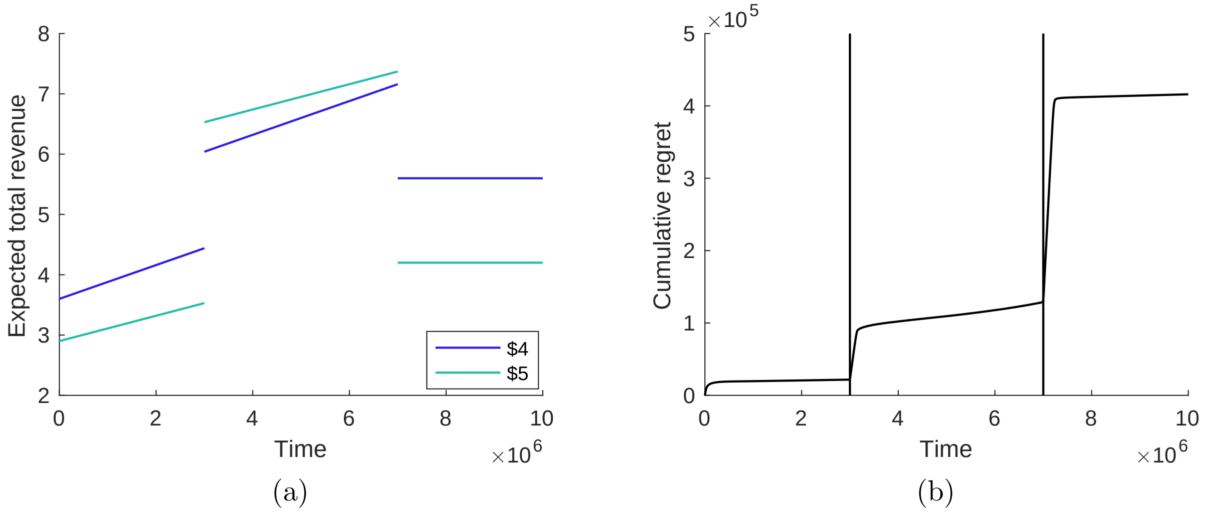


Figure 6 An example of cumulative regret of our algorithm with two zero-order change $J = 2$: (a) Illustration of the change in the model; (b) Cumulative regret of our algorithm: two changes are marked by solid lines.

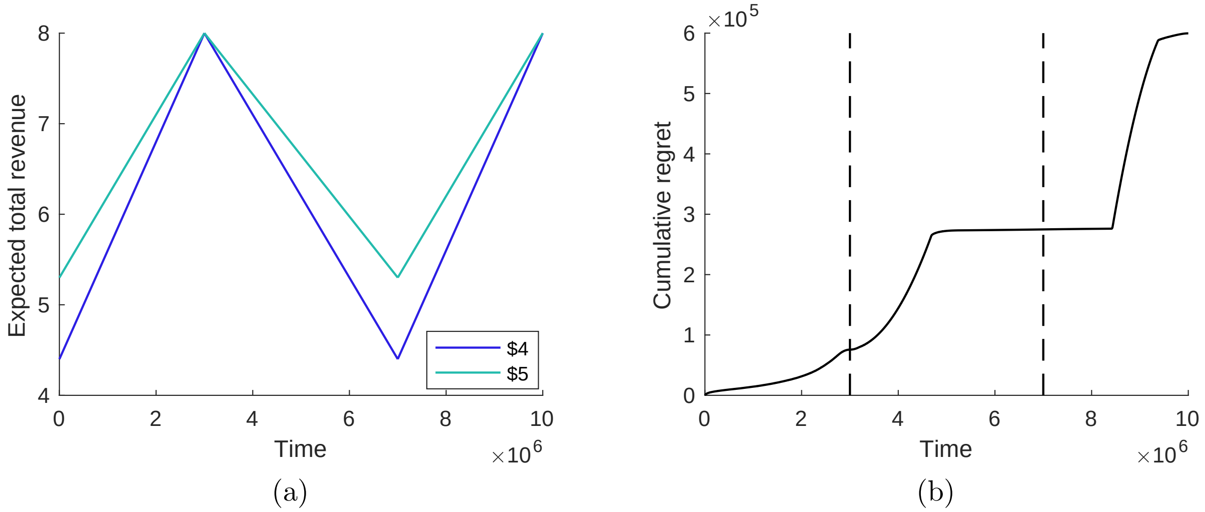


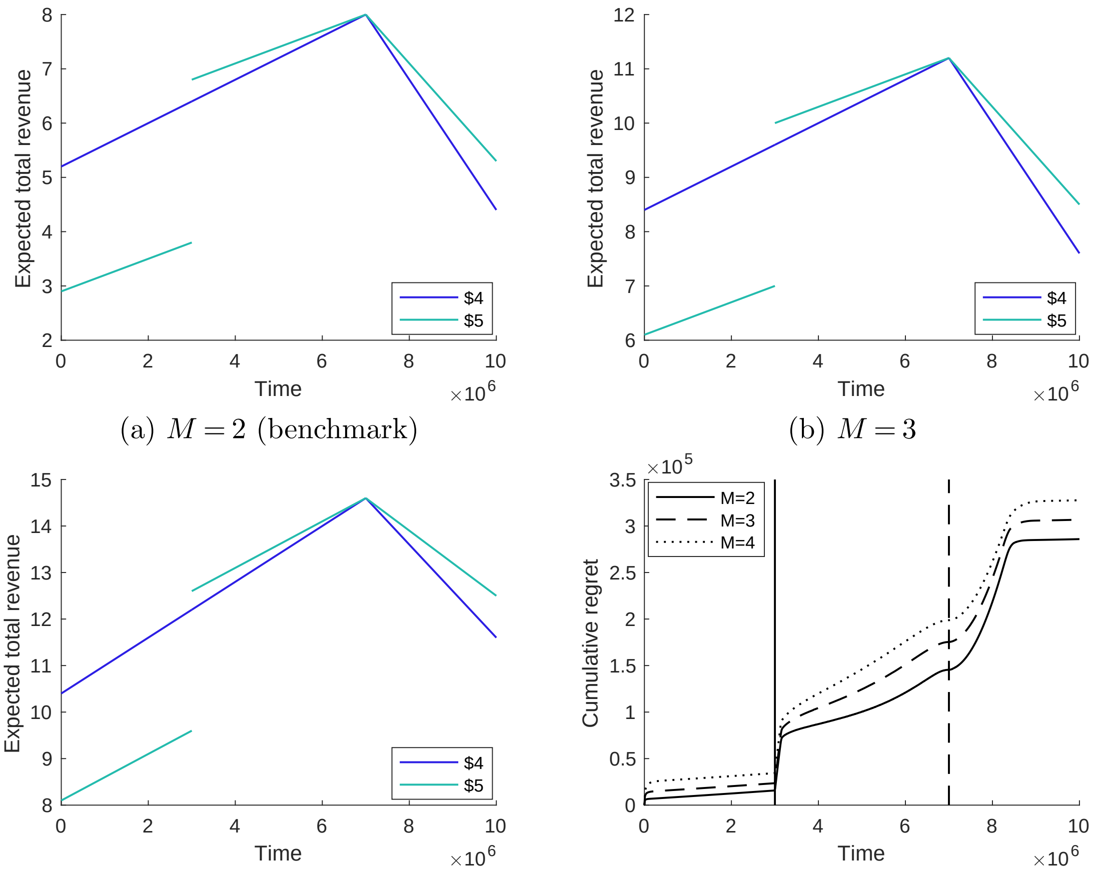
Figure 7 An example of cumulative regret of our algorithm with two first-order change $J = 2$: (a) Illustration of the change in the model; (b) Cumulative regret of our algorithm: two changes are marked by dashed lines.

6.4. Impact of Different Number of Prices

Now we also study the impact of the number of prices by adding new price options to the benchmark model. As shown in Fig. 9, a unit price at $6 and $3 are considered. By comparing the cumulative regret with $K = 2, 3, 4$, we observe that the increase in the number of prices results in larger regret, and the impact is much larger than that of the number of markets.

6.5. Impact of Policy Parameter Change

In this example, we consider different thresholds for change-point detection ϵ_0 and ϵ_1. As shown in Fig. 10, in this example, the choice of threshold ϵ has a significant impact on the growth of the regret, and a better choice is when $\epsilon_0 = 0.115$. We also show the impact of ϵ_1 using a simple two-



(a) $M = 2$ (benchmark)

(b) $M = 3$

(c) $M = 4$

(d) Cumulative regret

Figure 8 Cumulative regret under different number of markets: (a) $M = 2$; (b) $M = 3$; (c) $M = 4$; (d) the cumulative regret of the proposed algorithm respect to each of the settings: two changes are marked by solid and dashed lines.

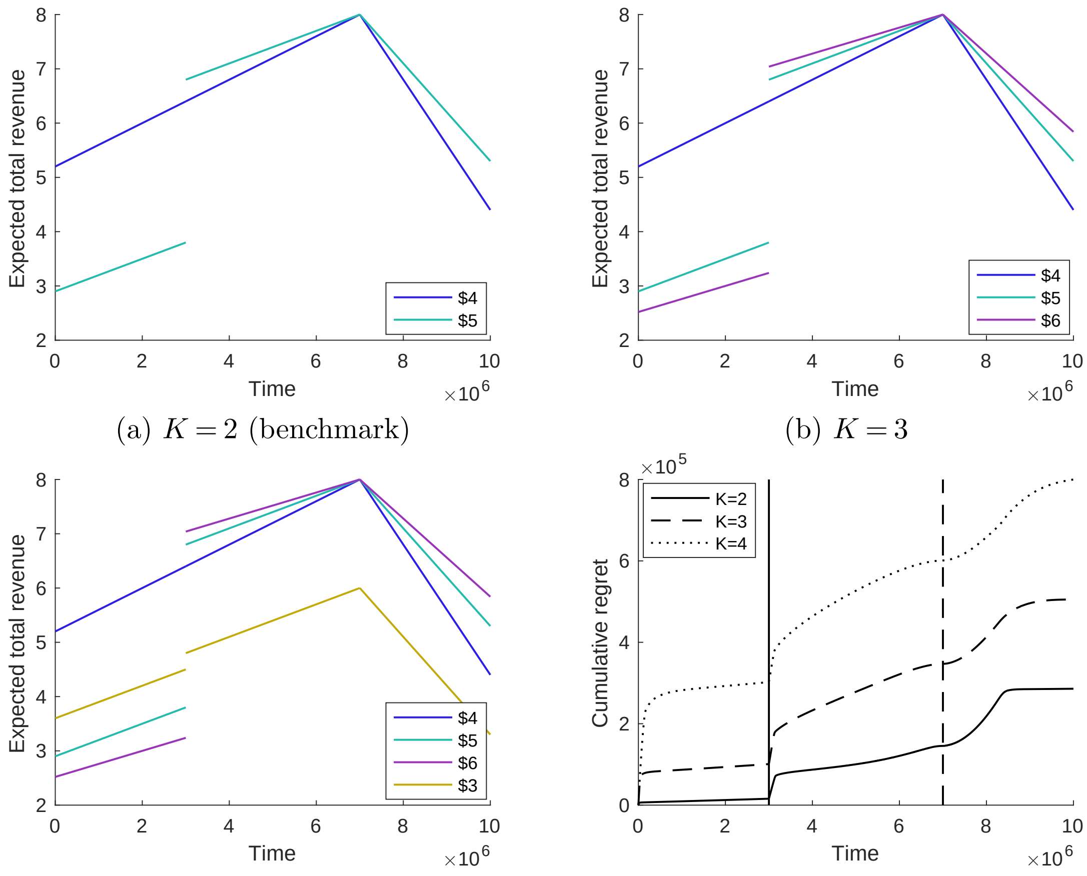
market two-price model, with one first-order change. Fig. 11 shows that with a moderate choice of $\epsilon_1 = 3$, the regret is minimal, while a small ϵ_1 will lead to a larger detection delay and thus larger regret, and an overly large ϵ_1 will fail to detect the change at all.

6.6. Policy Comparisons

Finally, we compare different policies regarding change-point detection: (i) the vanilla UCB algorithm without change-point detection and (ii) modification of our algorithm with zero-order change-point detection only. The results are shown in Fig. 12. We can observe that indeed our proposed algorithm achieves the smallest regret.

7. Concluding Remarks

We consider a firm that sells a single type product on multiple local markets over a finite horizon via dynamically adjusted prices. To prevent price discrimination, prices posted on different local markets at the same time are the same. The entire horizon consists of one or multiple change-points. Each local market's demand function linearly evolves over time between any two consecutive



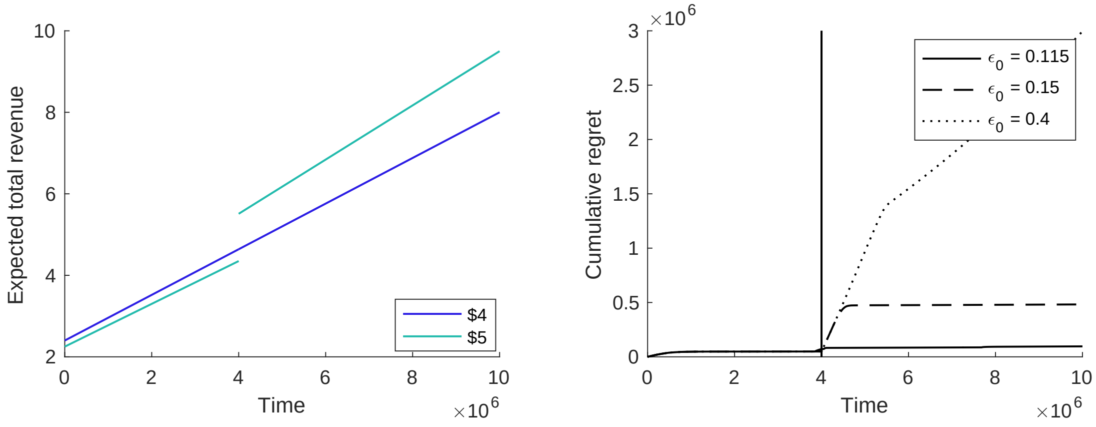
(a) $K = 2$ (benchmark)

(b) $K = 3$

(c) $K = 4$

(d) Cumulative regret

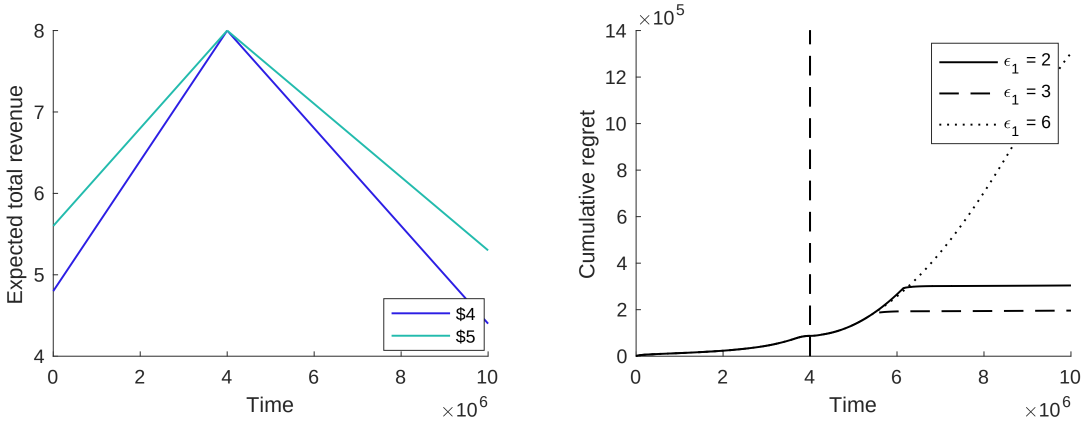
Figure 9 Cumulative regret under different number of prices. (a) $K = 2$**; (b)** $K = 3$**; (c)** $K = 4$**; (d) the cumulative regret of the proposed algorithm respect to each of the settings: two changes are marked by solid and dashed lines.**



(a) Expected Revenue

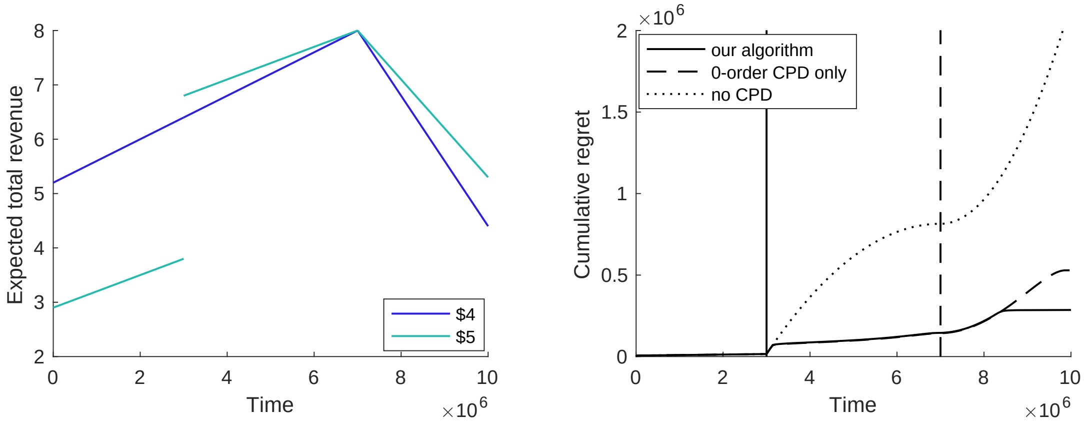
(b) Cumulative regret

Figure 10 Cumulative regret under different zero-order change-detection parameter ϵ_0**: (a) model setting; (b) regret for different** ϵ_0**.**



(a) Expected Revenue (b) Cumulative regret

Figure 11 Cumulative regret under different first-order change-detection parameter ϵ_1: (a) model setting; (b) regret for different ϵ_1.



(a) Expected Revenue (b) Cumulative regret

Figure 12 Cumulative regret under different policies regarding change-point detection (CPD): (a) model setting; (b) regret comparison of the proposed algorithm and the benchmarks.

change-points. Each change-point is either a zeroth-order or a first-order change-point. The firm has no information about any parameter that modulates the demand evolution process before the start of the horizon. The firm aims at finding a pricing policy that yields as much revenue as possible.

We show that whether there exists any first-order change-point leads to completely different results of the regret lower bounds. Under any pricing policy, the regret is always lower bounded by $CT^{1/2}$ with $C > 0$. However, it becomes as worse as $CT^{2/3}$ if at least one change-point is a first-order change-point.

In our CU algorithm, the change-point detection component allows us to use the uniformly sampled data to both detect whether a change-point occurs. Moreover, the algorithm judges whether it is a zeroth-order or a first-order change if it occurs. This entails that our algorithm does not require the firm to have any prior knowledge of the types of change-points. The exploration-exploitation component in the CU algorithm leverages the power of the classical UCB algorithm that balances exploration and exploitation. Note that our model allows demand to dynamically evolve over time between any two consecutive change-points. However, the classical UCB algorithm is designed for the demand process that is stationary over time. Hence, this algorithm gives us a biased estimate of the demand at the present time by using historic data. Therefore, we introduce a time adjustment factor onto the classical UCB algorithm to correct this bias. We show that the CU algorithm achieves the regret lower bounds (up to logarithmic factors).

Acknowledgments

The authors thank Jason Acimovic, Omar Besbes for their constructive and helpful suggestions, and Haoyun Wang for help with the numerical examples. The work of the 3rd author is supported by National Science Foundation (NSF) CAREER CCF-1650913, DMS-1830210, CMMI-1538746 and CCF-1442635.

References

Araman VF, Caldentey R (2009) Dynamic pricing for nonperishable products with demand learning. *Operations research* 57(5):1169–1188.

Auer P (2002) Using confidence bounds for exploitation-exploration trade-offs. *Journal of Machine Learning Research* 3(Nov):397–422.

Auer P, Cesa-Bianchi N, Freund Y, Schapire RE (2002) The nonstochastic multiarmed bandit problem. *SIAM journal on computing* 32(1):48–77.

Basseville M, Nikiforov IV (1993) *Detection of Abrupt Changes: Theory and Application* (Prentice Hall).

Bauer H, Burkacky O, Kenevan P, Mahindroo A, Patel M (2020) How the semiconductor industry can emerge stronger after the covid-19 crisis .

Besbes O, Gur Y, Zeevi A (2014) Stochastic multi-armed-bandit problem with non-stationary rewards. *Advances in neural information processing systems*, 199–207.

Besbes O, Sauré D (2014) Dynamic pricing strategies in the presence of demand shifts. *Manufacturing & Service Operations Management* 16(4):513–528.

Besbes O, Zeevi A (2009) Dynamic pricing without knowing the demand function: Risk bounds and near-optimal algorithms. *Operations Research* 57(6):1407–1420.

Besbes O, Zeevi A (2011) On the minimax complexity of pricing in a changing environment. *Operations research* 59(1):66–79.

Besbes O, Zeevi A (2012) Blind network revenue management. *Operations research* 60(6):1537–1550.

Broder J, Rusmevichientong P (2012) Dynamic pricing under a general parametric choice model. *Operations Research* 60(4):965–980.

Brodsky E, Darkhovsky BS (1993) *Nonparametric methods in change point problems* (Springer).

Cao Y, Wen Z, Kveton B, Xie Y (2018a) Nearly optimal adaptive procedure with change detection for piecewise-stationary bandit. *arXiv preprint arXiv:1802.03692* .

Cao Y, Xie Y, Gebraeel N (2018b) Multi-sensor slope change detection. *Annals of Operations Research* 263(1):163–189.

Chen B, Chao X (2017) Data-based dynamic pricing and inventory control with censored demand and limited price changes. *Available at SSRN 2700747* .

Chen B, Chao X (2019) Parametric demand learning with limited price explorations in a backlog stochastic inventory system. *IISE Transactions* 51(6):605–613.

Chen B, Chao X, Ahn HS (2019a) Coordinating pricing and inventory replenishment with nonparametric demand learning. *Operations Research* 67(4):1035–1052.

Chen B, Chao X, Shi C (2015) Nonparametric algorithms for joint pricing and inventory control with lost-sales and censored demand. *Mathematics of Operations Research, Major Revision* .

Chen J, Gupta AK (2012) *Parametric Statistical Change Point Analysis: With Applications to Genetics, Medicine, and Finance* (Birkhauser).

Chen N, Gallego G (2018) A primal-dual learning algorithm for personalized dynamic pricing with an inventory constraint. *Available at SSRN* .

Chen Q, Jasin S, Duenyas I (2019b) Nonparametric self-adjusting control for joint learning and optimization of multiproduct pricing with finite resource capacity. *Mathematics of Operations Research* .

Chen X, Wang Y, Wang Y (2017) Non-stationary stochastic optimization under l_ {p, q}-variation measures. *Available at SSRN 3014773* .

Chen Y, Farias VF (2013) Simple policies for dynamic pricing with imperfect forecasts. *Operations Research* 61(3):612–624.

Chen Y, Shi C (2019) Network revenue management with online inverse batch gradient descent method. *Available at SSRN* .

Cheung WC, Simchi-Levi D, Wang H (2017) Dynamic pricing and demand learning with limited price experimentation. *Operations Research* 65(6):1722–1731.

Cheung WC, Simchi-Levi D, Zhu R (2018) Learning to optimize under non-stationarity. *arXiv preprint arXiv:1810.03024* .

Cheung WC, Tan V, Zhong Z (2019) A Thompson sampling algorithm for cascading bandits. *Proceedings of Machine Learning Research*, volume 89 of *Proceedings of Machine Learning Research*, 438–447 (PMLR).

Den Boer AV (2015) Tracking the market: Dynamic pricing and learning in a changing environment. *European journal of operational research* 247(3):914–927.

den Boer AV, Zwart B (2015) Dynamic pricing and learning with finite inventories. *Operations research* 63(4):965–978.

Farias VF, Van Roy B (2010) Dynamic pricing with a prior on market response. *Operations Research* 58(1):16–29.

Ferreira KJ, Lee BHA, Simchi-Levi D (2015) Analytics for an online retailer: Demand forecasting and price optimization. *Manufacturing & Service Operations Management* 18(1):69–88.

Ferreira KJ, Simchi-Levi D, Wang H (2018) Online network revenue management using thompson sampling. *Operations research* 66(6):1586–1602.

Garivier A, Moulines E (2011) On upper-confidence bound policies for switching bandit problems. *International Conference on Algorithmic Learning Theory*, 174–188 (Springer).

Gordon L, Pollak M (1994) An efficient sequential nonparametric scheme for detecting a change of distribution. *The Annals of Statistics* 763–804.

Harrison JM, Keskin NB, Zeevi A (2012) Bayesian dynamic pricing policies: Learning and earning under a binary prior distribution. *Management Science* 58(3):570–586.

Hu K, Acimovic J, Erize F, Thomas DJ, Van Mieghem JA (2017) Forecasting product life cycle curves: Practical approach and empirical analysis .

Keskin NB, Li M (2020) Selling quality-deffereitiated products in a markovian market with unkonwn transition probabilities. *Working Paper* .

Keskin NB, Zeevi A (2014) Dynamic pricing with an unknown demand model: Asymptotically optimal semi-myopic policies. *Operations Research* 62(5):1142–1167.

Keskin NB, Zeevi A (2016) Chasing demand: Learning and earning in a changing environment. *Mathematics of Operations Research* 42(2):277–307.

Kocsis L, Szepesvári C (2006) Discounted ucb. *2nd PASCAL Challenges Workshop*, volume 2.

Kveton B, Wen Z, Ashkan A, Eydgahi H, Eriksson B (2014) Matroid bandits: Fast combinatorial optimization with learning. *Proceedings of the Thirtieth Conference on Uncertainty in Artificial Intelligence*, 420–429, UAI'14 (Arlington, Virginia, USA).

Lai TL (1995) Sequential changepoint detection in quality control and dynamical systems. *Journal of the Royal Statistical Society. Series B (Methodological)* 613–658.

Lai TL (1998) Information bounds and quick detection of parameter changes in stochastic systems. *Information Theory, IEEE Transactions on* 44(7):2917–2929.

Liu F, Lee J, Shroff N (2018) A change-detection based framework for piecewise-stationary multi-armed bandit problem. *Thirty-Second AAAI Conference on Artificial Intelligence*.

Lorden G (1971) Procedures for reacting to a change in distribution. *Annals of Mathematical Statistics* 42(6):1897–1908.

Maylín-Aguilar C, Montoro-Sánchez Á (2020) The industry life cycle in an economic downturn: Lessons from firm's behavior in spain, 2007–2012. *Journal of Business Cycle Research* 1–30.

Moustakides GV (1986) Optimal stopping times for detecting changes in distributions. *Ann. Statist.* 14:1379–1387.

Page ES (1954) Continuous inspection schemes. *Biometrika* 41(1/2):100–115.

Page ES (1955) A test for a change in a parameter occurring at an unknown point. *Biometrika* 42(3/4):523–527.

Parsons JCW (2004) *Using a newsvendor model for demand planning of NFL replica jerseys*. Ph.D. thesis, Massachusetts Institute of Technology.

Pollak M (1985) Optimal detection of a change in distribution. *Ann. Statist.* 13:206–227.

Pollak M (1987) Average run lengths of an optimal method of detecting a change in distribution. *Ann. Statist.* .

Roberts SW (1966) A comparison of some control chart procedures. *Technometrics* (8):411–430.

Shewhart AW (1931) Economic control of quality of manufactured product. *Preprinted by ASQC quality press* .

Shiryaev WA (1963) On optimal methods in quickest detection problems. *Theory Prob. Appl.* 8:22 – 46.

Siegmund DO (1985) *Sequential Analysis: Tests and Confidence Intervals*. Springer Series in Statistics (Springer).

Tartakovsky A, Nikiforov I, Basseville M (2014) *Sequential Analysis: Hypothesis Testing and Changepoint Detection* (Chapman and Hall/CRC).

Veeravalli VV, Banerjee T (2013) Quickest change detection. *E-Reference Signal Processing* (Elsevier).

Wang Z, Deng S, Ye Y (2014) Close the gaps: A learning-while-doing algorithm for single-product revenue management problems. *Operations Research* 62(2):318–331.

Zhou X, Chen N, Gao X, Xiong Y (2020) Regime switching bandits. *arXiv preprint arXiv:2001.09390* .

Appendix A: Proofs for §3

In this section, we denote by $\mathrm{KL}_B(\theta,\theta')$ the Kullback-Leibler (KL) divergence of two Bernoulli random variables with parameters θ and θ'. We begin with stating and proving the following result that will be used in the proof of Theorem 1.

LEMMA 12. *Consider any $\theta \in (0,1)$ and $\epsilon \in \left(-\frac{\theta}{2}, \frac{1-\theta}{2}\right]$. We have*

$$\mathrm{KL}_B(\theta,\theta+\epsilon) \le \frac{\epsilon^2}{\theta(1-\theta)}.$$

Proof of Lemma 12.

We have

$$\begin{aligned}\mathrm{KL}_B(\theta,\theta+\epsilon) &= \theta\log\frac{\theta}{\theta+\epsilon} + (1-\theta)\log\frac{1-\theta}{1-\theta-\epsilon}\\ &= -\theta\log\left(1+\frac{\epsilon}{\theta}\right) - (1-\theta)\log\left(1-\frac{\epsilon}{1-\theta}\right)\\ &\le -\theta\left(\frac{\epsilon}{\theta}-\frac{\epsilon^2}{\theta^2}\right) + (1-\theta)\left(\frac{\epsilon}{1-\theta}+\frac{\epsilon^2}{(1-\theta)^2}\right)\\ &= \frac{\epsilon^2}{\theta(1-\theta)}.\end{aligned}$$

The inequality follows from the properties that $\log(1+x) \ge x - x^2$ for $x \ge -\frac{1}{2}$, and $\epsilon \in \left(-\frac{\theta}{2}, \frac{1-\theta}{2}\right]$.

Q.E.D.

Proof of Theorem 1.

1. Consider any pricing policy π. We analyze the performance of π in two instances.

   The first instance is defined in the following way.

This instance has a single market $M = 1$ and K distinct prices with $p_k \in [2,3]$ for $k \in \{1, \cdots, K\}$. For each price p_k, the c.c.d.f.s of customer willingness-to-pays are stationary:

$$\bar{F}_t^{(1)}(p_k) = \begin{cases} \frac{1}{p_k} + \frac{1}{p_k}\epsilon & \text{if } k = 1 \\ \frac{1}{p_k} & \text{if } k \neq 1 \end{cases},$$

where $\epsilon = \frac{1}{4}\sqrt{\frac{K}{3T}}$.

The market size is stationary and is equal to 1.

Therefore, for any t and $k \neq 1$, we have

$$\begin{aligned} g_t^{(1)}(k) &= p_k \bar{F}_t^{(1)}(p_k) \\ &= p_k \frac{1}{p_k} \\ &= 1. \end{aligned}$$

For any t and $k = 1$, we have

$$\begin{aligned} g_t^{(1)}(1) &= p_1 \bar{F}_t^{(1)}(p_1) \\ &= p_1 \left(\frac{1}{p_1} + \frac{1}{p_1}\epsilon \right) \\ &= 1 + \epsilon. \end{aligned}$$

Therefore, for any t and $k \neq 1$,

$$g_t^{(1)}(1) - g_t^{(1)}(k) = \epsilon. \tag{9}$$

Therefore, for any t, $\pi_t^{*,(1)} = 1$.

Next, we construct the second instance.

Define

$$k' \in \arg\min_{k \neq 1} \mathsf{E}^{(1)} \left[\sum_{t=1}^{T} \mathbf{1}\{\pi_t = k\} \right].$$

We notice that

$$\begin{aligned} \sum_{k \neq 1} \mathsf{E}^{(1)} \left[\sum_{t=1}^{T} \mathbf{1}\{\pi_t = k\} \right] &\leq \sum_{k=1}^{K} \mathsf{E}^{(1)} \left[\sum_{t=1}^{T} \mathbf{1}\{\pi_t = k\} \right] \\ &= \mathsf{E}^{(1)} \left[\sum_{t=1}^{T} \sum_{k=1}^{K} \mathbf{1}\{\pi_t = k\} \right] \\ &= \mathsf{E}^{(1)} \left[\sum_{t=1}^{T} 1 \right] \\ &= T. \end{aligned}$$

Therefore, the definition of k' implies $k' \leq \frac{T}{K-1}$.

The second instance is almost the same as the first instance, except the definition of the c.c.d.f. of customer willingness-to-pays for price $p_{k'}$:

$$\bar{F}_t^{(2)}(p_{k'}) = \frac{1}{p_{k'}} + \frac{1}{p_{k'}} 2\epsilon.$$

Therefore, for any t and $k = k'$, we have

$$\begin{aligned} g_t^{(2)}(k') &= p_{k'} \bar{F}_t^{(2)}(p_{k'}) \\ &= p_{k'} \left(\frac{1}{p_{k'}} + \frac{1}{p_{k'}} 2\epsilon \right) \\ &= 1 + 2\epsilon. \end{aligned}$$

Therefore, for any t and $k \neq 1, k'$,

$$g_t^{(2)}(k') - g_t^{(2)}(k) = 2\epsilon. \tag{10}$$

For any t and $k = 1$,

$$g_t^{(2)}(k') - g_t^{(2)}(1) = \epsilon. \tag{11}$$

Therefore, for any t, $\pi_t^{*,(2)} = k'$.

Therefore,

$$\begin{aligned} \sum_{i=1}^{2} \text{Regret}^{\pi,(i)}(T) &= \sum_{i=1}^{2} \mathsf{E}^{(i)} \left[\sum_{t=1}^{T} \left(g_t^{(i)} \left(p_{\pi_t^{*,(i)}} \right) - g_t^{(i)}(p_{\pi_t}) \right) \right] \\ &\geq \mathsf{E}^{(1)} \left[\sum_{t=1}^{T} \epsilon \mathbf{1}\{\pi_t \neq 1\} \right] + \mathsf{E}^{(2)} \left[\sum_{t=1}^{T} \epsilon \mathbf{1}\{\pi_t \neq k'\} \right] \\ &= \epsilon \left(\mathsf{E}^{(1)} \left[\sum_{t=1}^{T} \mathbf{1}\{\pi_t \neq 1\} \right] + \mathsf{E}^{(2)} \left[\sum_{t=1}^{T} \mathbf{1}\{\pi_t \neq k'\} \right] \right). \end{aligned}$$

The inequality follows from Equations (9), (10), (11).

Define event A as

$$A = \left\{ \sum_{t=1}^{T} \mathbf{1}\{\pi_t \neq 1\} \geq \frac{T}{2} \right\}.$$

Hence, the complement of A, denoted as A^c, is

$$A^c = \left\{ \sum_{t=1}^{T} \mathbf{1}\{\pi_t = 1\} \geq \frac{T}{2} \right\}.$$

Therefore,

$$\sum_{i=1}^{2} \text{Regret}^{\pi,(i)}(T) \geq \epsilon \left(\mathsf{E}^{(1)} \left[\sum_{t=1}^{T} \mathbf{1}\{\pi_t \neq 1\} \right] + \mathsf{E}^{(2)} \left[\sum_{t=1}^{T} \mathbf{1}\{\pi_t \neq k'\} \right] \right)$$

$$
\begin{aligned}
&= \epsilon \mathsf{E}^{(1)}\left[\sum_{t=1}^{T}\mathbf{1}\{\pi_t \neq 1\}\middle| A\right]\mathbb{P}^{(1)}(A) + \epsilon \mathsf{E}^{(1)}\left[\sum_{t=1}^{T}\mathbf{1}\{\pi_t \neq 1\}\middle| A^c\right]\mathbb{P}^{(1)}(A^c) \\
&\qquad + \epsilon \mathsf{E}^{(2)}\left[\sum_{t=1}^{T}\mathbf{1}\{\pi_t \neq k'\}\middle| A\right]\mathbb{P}^{(2)}(A) + \epsilon \mathsf{E}^{(2)}\left[\sum_{t=1}^{T}\mathbf{1}\{\pi_t \neq k'\}\middle| A^c\right]\mathbb{P}^{(2)}(A^c) \\
&\geq \epsilon \mathsf{E}^{(1)}\left[\sum_{t=1}^{T}\mathbf{1}\{\pi_t \neq 1\}\middle| A\right]\mathbb{P}^{(1)}(A) + \epsilon \mathsf{E}^{(2)}\left[\sum_{t=1}^{T}\mathbf{1}\{\pi_t \neq k'\}\middle| A^c\right]\mathbb{P}^{(2)}(A^c) \\
&\geq \epsilon \mathsf{E}^{(1)}\left[\sum_{t=1}^{T}\mathbf{1}\{\pi_t \neq 1\}\middle| A\right]\mathbb{P}^{(1)}(A) + \epsilon \mathsf{E}^{(2)}\left[\sum_{t=1}^{T}\mathbf{1}\{\pi_t = 1\}\middle| A^c\right]\mathbb{P}^{(2)}(A^c) \\
&\geq \frac{T\epsilon}{2}\mathbb{P}^{(1)}(A) + \frac{T\epsilon}{2}\mathbb{P}^{(2)}(A^c) \\
&= \frac{T\epsilon}{2}\left(\mathbb{P}^{(1)}(A) + \mathbb{P}^{(2)}(A^c)\right) \\
&\geq \frac{T\epsilon}{4}\exp\left(-\mathrm{KL}^{\pi}(1,2)\right).
\end{aligned}
$$

The third inequality follows from the property that $k' \neq 1$. In the fifth inequality, $\mathrm{KL}^{\pi}(1,2)$ denotes the KL-divergence between the two instances defined above. Hence, this inequality follows from the Pinsker's inequality.

Now, we establish an upper bound of $\mathrm{KL}^{\pi}(1,2)$.

We have

$$
\begin{aligned}
\mathrm{KL}^{\pi}(1,2) &= \mathsf{E}^{(1)}\left[\sum_{t=1}^{T}\mathrm{KL}_B\left(\frac{1}{p_{k'}}, \frac{1}{p_{k'}} + \frac{1}{p_{k'}}2\epsilon\right)\mathbf{1}\{\pi_t = k'\}\right] \\
&\leq \mathsf{E}^{(1)}\left[\sum_{t=1}^{T}\frac{(2\epsilon)^2}{\frac{1}{p_{k'}}\left(1-\frac{1}{p_{k'}}\right)}\mathbf{1}\{\pi_t = k'\}\right] \\
&\leq \mathsf{E}^{(1)}\left[\sum_{t=1}^{T}\frac{(2\epsilon)^2}{\frac{1}{3}\left(1-\frac{1}{2}\right)}\mathbf{1}\{\pi_t = k'\}\right] \\
&= 24\epsilon^2 \mathsf{E}^{(1)}\left[\sum_{t=1}^{T}\mathbf{1}\{\pi_t = k'\}\right] \\
&\leq 24\epsilon^2\frac{T}{K-1} \\
&\leq 48\epsilon^2\frac{T}{K}.
\end{aligned}
$$

The first equality follows from the chain rule of the KL divergence and the property that under our constructed two instances, for any t, the probability measures of two instances are different only if $\pi_t = k'$. The first inequality follows from Lemma 12. The second inequality follows from the property that $p_{k'} \in [2,3]$. The third inequality follows from the property that $\mathsf{E}^{(1)}[\mathbf{1}\{\pi_t = k'\}] \leq \frac{T}{K-1}$. The fourth inequality follows from the property that for $K \geq 2$, $\frac{1}{K-1} \leq \frac{2}{K}$.

Therefore, we have

$$\begin{aligned}\sum_{i=1}^{2} \mathrm{Regret}^{\pi,(i)}(T) &\geq \frac{T\epsilon}{4}\exp\left(-\mathrm{KL}^{\pi}(1,2)\right)\\ &\geq \frac{T\epsilon}{4}\exp\left(-48\epsilon^2\frac{T}{K}\right)\\ &= \frac{1}{16e}\sqrt{\frac{KT}{3}}.\end{aligned}$$

Therefore, between the two instances defined above, there must exist at least one instance i, with

$$\begin{aligned}\mathrm{Regret}^{\pi,(i)}(T) &\geq \frac{1}{2}\frac{1}{16e}\sqrt{\frac{KT}{3}}\\ &= \frac{1}{32e}\sqrt{\frac{KT}{3}}.\end{aligned}$$

2. Consider any pricing policy π. We analyze the performance of π in two instances.

The first instance is defined in the following way.

This instance has two markets $M = 2$ and K distinct prices with $p_k \in [2,3]$ for $k \in \{1,\cdots,K\}$. For each market $m \in \{1,2\}$ and each price p_k, the c.c.d.f.s of customer willingness-to-pays are stationary:

$$\bar{F}_t^{m,(1)}(p_k) = \begin{cases} \frac{1}{p_k} + \frac{1}{4p_k}\frac{S}{T}(-1)^{m-1} & \text{if } k = 1 \\ \frac{1}{p_k} & \text{if } k \neq 1 \end{cases}.$$

There is a first-order change-point in period $T - S$, with $S = \lceil T^{8/9}K^{1/9} \rceil$. The market size evolution process for each market $m \in \{1,2\}$ is

$$S_t^{m,(1)} = \frac{1}{2} + \frac{1}{4}\frac{t-(T-S)}{T}\mathbf{1}\{m=1\}\mathbf{1}\{t \geq T-S\}.$$

Therefore, for $t \geq T - S + 1$ and $k \neq 1$, we have

$$\begin{aligned} g_t^{(1)}(k) &= p_k \sum_{m=1}^{2}\left(\frac{1}{2} + \frac{1}{4}\frac{t-(T-S)}{T}\mathbf{1}\{m=1\}\right)\frac{1}{p_k}\\ &= 1 + \frac{1}{4}\frac{t-(T-S)}{T}.\end{aligned}$$

For any $t \geq T - S + 1$ and $k = 1$, we have

$$\begin{aligned} g_t^{(1)}(1) &= p_1 \sum_{m=1}^{2}\left(\frac{1}{2} + \frac{1}{4}\frac{t-(T-S)}{T}\mathbf{1}\{m=1\}\right)\left(\frac{1}{p_1} + \frac{1}{4p_1}\frac{S}{T}(-1)^{m-1}\right)\\ &= 1 + \frac{1}{4}\frac{t-(T-S)}{T} + \frac{1}{16}\frac{t-(T-S)}{T}\frac{S}{T}.\end{aligned}$$

Therefore, for $t \geq T - S + 1$ and $k \neq 1$,

$$g_t^{(1)}(1) - g_t^{(1)}(k) = \frac{1}{16}\frac{t-(T-S)}{T}\frac{S}{T}. \tag{12}$$

Therefore, for $t \geq T-S+1$, $\pi_t^{*,(1)} = 1$.

Next, we construct the second instance.

Define

$$k' \in \arg\min_{k \neq 1} \mathsf{E}^{(1)} \left[\sum_{t=T-S+1}^{T} \mathbf{1}\left\{\pi_t = k\right\} \right].$$

We notice that

$$\begin{aligned}\sum_{k \neq 1} \mathsf{E}^{(1)} \left[\sum_{t=T-S+1}^{T} \mathbf{1}\left\{\pi_t = k\right\} \right] &\leq \sum_{k=1}^{K} \mathsf{E}^{(1)} \left[\sum_{t=T-S+1}^{T} \mathbf{1}\left\{\pi_t = k\right\} \right] \\ &= \mathsf{E}^{(1)} \left[\sum_{t=T-S+1}^{T} \sum_{k=1}^{K} \mathbf{1}\left\{\pi_t = k\right\} \right] \\ &= \mathsf{E}^{(1)} \left[\sum_{t=T-S+1}^{T} 1 \right] \\ &= S.\end{aligned}$$

Therefore, the definition of k' implies $k' \leq \frac{S}{K-1}$.

The second instance is almost the same as the first instance, except the definition of the c.c.d.f. of customer willingness-to-pays for price $p_{k'}$:

$$\bar{F}_t^{m,(2)}\left(p_{k'}\right) = \frac{1}{p_{k'}} + \frac{1}{2p_{k'}} \frac{S}{T} \left(-1\right)^{m-1}.$$

Therefore, for $t \geq T-S+1$ and $k = k'$, we have

$$\begin{aligned}g_t^{(2)}\left(k'\right) &= p_{k'} \sum_{m=1}^{2} \left(\frac{1}{2} + \frac{1}{4} \frac{t-(T-S)}{T} \mathbf{1}\left\{m=1\right\} \right) \left(\frac{1}{p_{k'}} + \frac{1}{2p_{k'}} \frac{S}{T} \left(-1\right)^{m-1} \right) \\ &= 1 + \frac{1}{4} \frac{t-(T-S)}{T} + \frac{1}{8} \frac{t-(T-S)}{T} \frac{S}{T}.\end{aligned}$$

Therefore, for $t \geq T-S+1$ and $k \neq 1, k'$,

$$g_t^{(2)}\left(k'\right) - g_t^{(2)}\left(k\right) = \frac{1}{8} \frac{t-(T-S)}{T} \frac{S}{T}. \tag{13}$$

For $t \geq T-S+1$ and $k = 1$,

$$g_t^{(2)}\left(k'\right) - g_t^{(2)}\left(1\right) = \frac{1}{16} \frac{t-(T-S)}{T} \frac{S}{T}. \tag{14}$$

Therefore, for $t \geq T-S+1$, $\pi_t^{*,(2)} = k'$.

Therefore,

$$\sum_{i=1}^{2} \mathrm{Regret}^{\pi,(i)}\left(T\right)$$

$$
\begin{aligned}
&= \sum_{i=1}^{2} \mathsf{E}^{(i)} \left[\sum_{t=1}^{T} \left(\pi_t^{*,(i)} \sum_{m=1}^{2} S_t^{m,(i)} \bar{F}_t^{m,(i)} \left(p_{\pi_t^{*,(i)}} \right) - \pi_t \sum_{m=1}^{2} S_t^{m,(i)} \bar{F}_t^{m,(i)} \left(p_{\pi_t} \right) \right) \right] \\
&\geq \sum_{i=1}^{2} \mathsf{E}^{(i)} \left[\sum_{t=T-S+1}^{T} \left(\pi_t^{*,(i)} \sum_{m=1}^{2} S_t^{m,(i)} \bar{F}_t^{m,(i)} \left(p_{\pi_t^{*,(i)}} \right) - \pi_t \sum_{m=1}^{2} S_t^{m,(i)} \bar{F}_t^{m,(i)} \left(p_{\pi_t} \right) \right) \right] \\
&\geq \mathsf{E}^{(1)} \left[\sum_{t=T-S+1}^{T} \frac{1}{16} \frac{t-(T-S)}{T} \frac{S}{T} \mathbf{1} \left\{ \pi_t \neq 1 \right\} \right] + \mathsf{E}^{(2)} \left[\sum_{t=T-S+1}^{T} \frac{1}{16} \frac{t-(T-S)}{T} \frac{S}{T} \mathbf{1} \left\{ \pi_t \neq k' \right\} \right] \\
&= \frac{1}{16} \frac{S}{T} \left(\mathsf{E}^{(1)} \left[\sum_{t=T-S+1}^{T} \frac{t-(T-S)}{T} \mathbf{1} \left\{ \pi_t \neq 1 \right\} \right] + \mathsf{E}^{(2)} \left[\sum_{t=T-S+1}^{T} \frac{t-(T-S)}{T} \mathbf{1} \left\{ \pi_t \neq k' \right\} \right] \right) \\
&\geq \frac{1}{16} \frac{K^{1/9}}{T^{1/9}} \left(\mathsf{E}^{(1)} \left[\sum_{t=T-S+1}^{T} \frac{t-(T-S)}{T} \mathbf{1} \left\{ \pi_t \neq 1 \right\} \right] + \mathsf{E}^{(2)} \left[\sum_{t=T-S+1}^{T} \frac{t-(T-S)}{T} \mathbf{1} \left\{ \pi_t \neq k' \right\} \right] \right).
\end{aligned}
$$

The first inequality follows from the property that $\pi_t^{*,(i)}$ is an optimal decision in period t, but π_t is not necessarily optimal. The second inequality follows from Equations (12), (13), (14).

We denote by $L = \lfloor T^{2/3} K^{-2/3} \rfloor$ the length the review time window. Hence, for $T \geq 46$ and $K \geq 2$, we have

$$
\frac{S}{L} \geq \frac{T^{8/9} K^{1/9}}{T^{2/3} K^{-2/3}} = T^{2/9} K^{7/9} \geq 46^{2/9} 2^{7/9} > 4.
$$

For any $l \in \{1, \cdots, L\}$, we denote $\mathcal{T}_l = \{ t \in \{T-S+1, \cdots, T\} : t-(T-S) \equiv l \pmod{L} \}$. Therefore,

$$
\begin{aligned}
&\sum_{i=1}^{2} \text{Regret}^{\pi,(i)}(T) \\
&\geq \frac{1}{16} \frac{K^{1/9}}{T^{1/9}} \sum_{l=1}^{L} \left(\underbrace{ \mathsf{E}^{(1)} \left[\sum_{t \in \mathcal{T}_l} \frac{t-(T-S)}{T} \mathbf{1} \left\{ \pi_t \neq 1 \right\} \right] + \mathsf{E}^{(2)} \left[\sum_{t \in \mathcal{T}_l} \frac{t-(T-S)}{T} \mathbf{1} \left\{ \pi_t \neq k' \right\} \right] }_{R_l} \right).
\end{aligned}
$$

Next, we establish a lower bound of R_l. Define event A_l as

$$
A_l = \left\{ \sum_{t \in \mathcal{T}_l} \frac{t-(T-S)}{T} \mathbf{1} \left\{ \pi_t \neq 1 \right\} \geq \frac{1}{2} \sum_{t \in \mathcal{T}_l} \frac{t-(T-S)}{T} \right\}.
$$

Hence, the complement of A_l, denoted as A_l^c, is

$$
A_l^c = \left\{ \sum_{t \in \mathcal{T}_l} \frac{t-(T-S)}{T} \mathbf{1} \left\{ \pi_t = 1 \right\} \geq \frac{1}{2} \sum_{t \in \mathcal{T}_l} \frac{t-(T-S)}{T} \right\}.
$$

Therefore,

$$
R_l = \mathsf{E}^{(1)} \left[\sum_{t \in \mathcal{T}_l} \frac{t-(T-S)}{T} \mathbf{1} \left\{ \pi_t \neq 1 \right\} \middle| A_l \right] \mathbb{P}^{(1)} \left(A_l \right)
$$

$$
\begin{aligned}
&\quad + \mathsf{E}^{(1)}\left[\sum_{t\in\mathcal{T}_l}\frac{t-(T-S)}{T}\mathbf{1}\left\{\pi_t\neq 1\right\}\middle| A_l^c\right]\mathbb{P}^{(1)}\left(A_l^c\right)\\
&\quad + \mathsf{E}^{(2)}\left[\sum_{t\in\mathcal{T}_l}\frac{t-(T-S)}{T}\mathbf{1}\left\{\pi_t\neq k'\right\}\middle| A_l\right]\mathbb{P}^{(2)}\left(A_l\right)\\
&\quad + \mathsf{E}^{(2)}\left[\sum_{t\in\mathcal{T}_l}\frac{t-(T-S)}{T}\mathbf{1}\left\{\pi_t\neq k'\right\}\middle| A_l^c\right]\mathbb{P}^{(2)}\left(A_l^c\right)\\
&\geq \mathsf{E}^{(1)}\left[\sum_{t\in\mathcal{T}_l}\frac{t-(T-S)}{T}\mathbf{1}\left\{\pi_t\neq 1\right\}\middle| A_l\right]\mathbb{P}^{(1)}\left(A_l\right)\\
&\quad + \mathsf{E}^{(2)}\left[\sum_{t\in\mathcal{T}_l}\frac{t-(T-S)}{T}\mathbf{1}\left\{\pi_t\neq k'\right\}\middle| A_l^c\right]\mathbb{P}^{(2)}\left(A_l^c\right)\\
&\geq \mathsf{E}^{(1)}\left[\sum_{t\in\mathcal{T}_l}\frac{t-(T-S)}{T}\mathbf{1}\left\{\pi_t\neq 1\right\}\middle| A_l\right]\mathbb{P}^{(1)}\left(A_l\right)\\
&\quad + \mathsf{E}^{(2)}\left[\sum_{t\in\mathcal{T}_l}\frac{t-(T-S)}{T}\mathbf{1}\left\{\pi_t= 1\right\}\middle| A_l^c\right]\mathbb{P}^{(2)}\left(A_l^c\right)\\
&\geq \frac{1}{2}\sum_{t\in\mathcal{T}_l}\frac{t-(T-S)}{T}\mathbb{P}^{(1)}\left(A_l\right)+\frac{1}{2}\sum_{t\in\mathcal{T}_l}\frac{t-(T-S)}{T}\mathbb{P}^{(2)}\left(A_l^c\right)\\
&= \frac{1}{2}\sum_{t\in\mathcal{T}_l}\frac{t-(T-S)}{T}\left(\mathbb{P}^{(1)}\left(A_l\right)+\mathbb{P}^{(2)}\left(A_l^c\right)\right)\\
&\geq \frac{1}{4}\sum_{t\in\mathcal{T}_l}\frac{t-(T-S)}{T}\exp\left(-\mathrm{KL}_l^{\pi}\left(1,2\right)\right)\\
&\geq \frac{1}{4}\sum_{n=0}^{\left\lfloor\frac{S}{L}\right\rfloor-1}\frac{nL}{T}\exp\left(-\mathrm{KL}_l^{\pi}\left(1,2\right)\right)\\
&= \frac{L}{8T}\left(\left\lfloor\frac{S}{L}\right\rfloor-1\right)\left\lfloor\frac{S}{L}\right\rfloor\exp\left(-\mathrm{KL}_l^{\pi}\left(1,2\right)\right)\\
&\geq \frac{L}{8T}\left(\frac{S}{2L}\right)^2\exp\left(-\mathrm{KL}_l^{\pi}\left(1,2\right)\right)\\
&= \frac{S^2}{32TL}\exp\left(-\mathrm{KL}_l^{\pi}\left(1,2\right)\right)\\
&\geq \frac{T^{7/9}K^{2/9}}{32L}\exp\left(-\mathrm{KL}_l^{\pi}\left(1,2\right)\right).
\end{aligned}
$$

The second inequality follows from the property that $k'\neq 1$. In the fourth inequality, $\mathrm{KL}_l^{\pi}\left(1,2\right)$ denotes the KL divergence between the two instances defined above over the collection of periods $\mathcal{T}_l$. Hence, this inequality follows from the Pinsker's inequality. The sixth inequality follows from the property that for $\frac{S}{L}\geq 4$, $\left\lfloor\frac{S}{L}\right\rfloor-1\geq\frac{S}{L}-1-1\geq\frac{S}{2L}$.

Now, we establish an upper bound of $\mathrm{KL}_l^{\pi}\left(1,2\right)$.

We have

$$\begin{aligned}
\mathrm{KL}_l^{\pi}(1,2) &= \mathsf{E}^{(1)}\left[\sum_{t\in\mathcal{T}_l}\left(\sum_{m=1}^{2}\mathrm{KL}_B\left(\left(\frac{1}{2}+\frac{1}{4}\frac{t-(T-S)}{T}\mathbf{1}\{m=1\}\right)\frac{1}{p_{k'}},\right.\right.\right.\\
&\qquad\left.\left.\left.\left(\frac{1}{2}+\frac{1}{4}\frac{t-(T-S)}{T}\mathbf{1}\{m=1\}\right)\frac{1}{p_{k'}}\left(1+\frac{S}{2T}(-1)^{m-1}\right)\right)\right)\mathbf{1}\{\pi_t=k'\}\right]\\
&= \mathsf{E}^{(1)}\left[\sum_{t\in\mathcal{T}_l}\left(\mathrm{KL}_B\left(\left(\frac{1}{2}+\frac{1}{4}\frac{t-(T-S)}{T}\right)\frac{1}{p_{k'}},\left(\frac{1}{2}+\frac{1}{4}\frac{t-(T-S)}{T}\right)\frac{1}{p_{k'}}\left(1+\frac{S}{2T}\right)\right)\right.\right.\\
&\qquad\left.\left.+\mathrm{KL}_B\left(\frac{1}{2p_{k'}},\frac{1}{2p_{k'}}\left(1-\frac{S}{2T}\right),\right)\right)\mathbf{1}\{\pi_t=k'\}\right]\\
&\le \mathsf{E}^{(1)}\left[\sum_{t\in\mathcal{T}_l}\left(\frac{\left(\frac{1}{2}+\frac{1}{4}\frac{t-(T-S)}{T}\right)^2\left(\frac{S}{2T}\frac{1}{p_{k'}}\right)^2}{\left(\frac{1}{2}+\frac{1}{4}\frac{t-(T-S)}{T}\right)\frac{1}{p_{k'}}\left(1-\left(\frac{1}{2}+\frac{1}{4}\frac{t-(T-S)}{T}\right)\frac{1}{p_{k'}}\right)}+\frac{\left(\frac{1}{2p_{k'}}\frac{S}{2T}\right)^2}{\frac{1}{2p_{k'}}\left(1-\frac{1}{2p_{k'}}\right)}\right)\mathbf{1}\{\pi_t=k'\}\right]\\
&= \mathsf{E}^{(1)}\left[\sum_{t\in\mathcal{T}_l}\left(\frac{\left(\frac{1}{2}+\frac{1}{4}\frac{t-(T-S)}{T}\right)\frac{1}{p_{k'}}\left(\frac{S}{2T}\right)^2}{1-\left(\frac{1}{2}+\frac{1}{4}\frac{t-(T-S)}{T}\right)\frac{1}{p_{k'}}}+\frac{\frac{1}{2p_{k'}}\left(\frac{S}{2T}\right)^2}{1-\frac{1}{2p_{k'}}}\right)\mathbf{1}\{\pi_t=k'\}\right]\\
&\le \mathsf{E}^{(1)}\left[\sum_{t\in\mathcal{T}_l}\left(\frac{\frac{3}{4}\cdot\frac{1}{2}\left(\frac{S}{2T}\right)^2}{1-\frac{3}{4}\cdot\frac{1}{2}}+\frac{\frac{1}{2\cdot 2}\left(\frac{S}{2T}\right)^2}{1-\frac{1}{2\cdot 2}}\right)\mathbf{1}\{\pi_t=k'\}\right]\\
&= \frac{7}{30}\frac{S^2}{T^2}\mathsf{E}^{(1)}\left[\sum_{t\in\mathcal{T}_l}\mathbf{1}\{\pi_t=k'\}\right]\\
&\le \frac{7}{30}\frac{K^{2/9}}{T^{2/9}}\mathsf{E}^{(1)}\left[\sum_{t\in\mathcal{T}_l}\mathbf{1}\{\pi_t=k'\}\right].
\end{aligned}$$

The first equality follows from the chain rule of the KL divergence and the property that under our constructed two instances, for any t, the probability measures of two instances are different only if $\pi_t=k'$. The first inequality follows from Lemma 12. The second inequality follows from the property that $\frac{1}{2}+\frac{1}{4}\frac{t-(T-S)}{T}\le\frac{1}{2}+\frac{1}{4}=\frac{3}{4}$ and $p_{k'}\ge 2$.

Therefore,

$$\begin{aligned}
\sum_{i=1}^{2}\mathrm{Regret}^{\pi,(i)}(T) &\ge \frac{1}{16}\frac{K^{1/9}}{T^{1/9}}\sum_{l=1}^{L}R_l\\
&\ge \frac{1}{16}\frac{K^{1/9}}{T^{1/9}}\sum_{l=1}^{L}\frac{T^{7/9}K^{2/9}}{32L}\exp\left(-\mathrm{KL}_l^{\pi}(1,2)\right)\\
&\ge \frac{1}{16}\frac{K^{1/9}}{T^{1/9}}\sum_{l=1}^{L}\frac{T^{7/9}K^{2/9}}{32L}\exp\left(-\frac{7}{30}\frac{K^{2/9}}{T^{2/9}}\mathsf{E}^{(1)}\left[\sum_{t\in\mathcal{T}_l}\mathbf{1}\{\pi_t=k'\}\right]\right)\\
&= \frac{K^{1/3}T^{2/3}}{512}\cdot\frac{1}{L}\sum_{l=1}^{L}\exp\left(-\frac{7}{30}\frac{K^{2/9}}{T^{2/9}}\mathsf{E}^{(1)}\left[\sum_{t\in\mathcal{T}_l}\mathbf{1}\{\pi_t=k'\}\right]\right)\\
&\ge \frac{K^{1/3}T^{2/3}}{512}\exp\left(-\frac{7}{30}\frac{K^{2/9}}{T^{2/9}}\frac{1}{L}\sum_{l=1}^{L}\mathsf{E}^{(1)}\left[\sum_{t\in\mathcal{T}_l}\mathbf{1}\{\pi_t=k'\}\right]\right)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{K^{1/3}T^{2/3}}{512} \exp\left(-\frac{7}{30}\frac{K^{2/9}}{T^{2/9}}\frac{1}{L}\mathsf{E}^{(1)}\left[\sum_{t=T-S+1}^{T}\mathbf{1}\left\{\pi_t = k'\right\}\right]\right) \\
&\geq \frac{K^{1/3}T^{2/3}}{512} \exp\left(-\frac{7}{30}\frac{K^{2/9}}{T^{2/9}}\frac{1}{L}\frac{S}{K-1}\right) \\
&\geq \frac{K^{1/3}T^{2/3}}{512} \exp\left(-\frac{7}{30}\frac{K^{2/9}}{T^{2/9}}\frac{1}{L}\frac{2S}{K}\right) \\
&= \frac{K^{1/3}T^{2/3}}{512} \exp\left(-\frac{7S}{15T^{2/9}LK^{7/9}}\right) \\
&\geq \frac{K^{1/3}T^{2/3}}{512} \exp\left(-\frac{7\left(2T^{8/9}K^{1/9}\right)}{15T^{2/9}\left(\frac{1}{2}T^{2/3}K^{-2/3}\right)K^{7/9}}\right) \\
&= \frac{K^{1/3}T^{2/3}}{512} \exp\left(-\frac{28}{15}\right) \\
&\geq \frac{1}{512e^2}K^{1/3}T^{2/3}.
\end{aligned}
$$

The fourth inequality follows from the property that $\exp(\cdot)$ is a convex function and Jensen's inequality. The fifth inequality follows from the property that $k' \leq \frac{S}{K-1}$. The sixth inequality follows from the property that for $K \geq 2$, $K-1 \geq \frac{K}{2}$.

Therefore, between two instances that we construct in this part, there must be at least one instance i, with

$$
\begin{aligned}
\text{Regret}^{\pi,(i)}(T) &\geq \frac{1}{2}\frac{1}{512e^2}K^{1/3}T^{2/3} \\
&= \frac{1}{1024e^2}K^{1/3}T^{2/3}.
\end{aligned}
$$

Q.E.D.

Appendix B: Proofs for §4

Proof of Lemma 1.

In this proof, For any $s \leq t$, we define $\Delta\bar{X}_s^{m,0} \triangleq \bar{X}_s^{m,0} - \mathsf{E}\left[\bar{X}_s^{m,0}|\mathcal{G}_t\right]$ and $\Delta\bar{X}_s^{m,1} \triangleq \bar{X}_s^{m,1} - \mathsf{E}\left[\bar{X}_s^{m,1}|\mathcal{G}_t\right]$.

First, we prove Part 1.

We have

$$
\begin{aligned}
&\left|\mathsf{E}\left[\bar{X}_t^{m,0} - \bar{X}_{t-w_0L/2}^{m,0}\middle|\mathcal{G}_t\right]\right| \\
=& \left|\frac{2}{w_0}\sum_{i=0}^{w_0/2-1}\left(a_{j+1}^m + b_{j+1}^m\frac{\tau_{t+1}+t-iL-\tau_{t+1}}{T}\right)\theta_{j+1}^m(k)\right. \\
&\left. -\frac{2}{w_0}\sum_{i=w_0/2}^{w_0-1}\left(a_j^m + b_j^m\frac{\tau_{t+1}+t-iL-\tau_{t+1}}{T}\right)\theta_j^m(k)\right|
\end{aligned}
$$

$$
\begin{aligned}
&= \Bigg| \left(a_{j+1}^m + b_{j+1}^m \frac{\tau_{t+1}}{T}\right) \theta_{j+1}^m(k) - \left(a_j^m + b_j^m \frac{s}{T}\right) \theta_j^m(k) \\
&\quad + \frac{2}{w_0} \sum_{i=0}^{w_0/2-1} b_{j+1}^m \frac{t-iL-\tau_{t+1}}{T} \theta_{j+1}^m(k) - \frac{2}{w_0} \sum_{i=w_0/2}^{w_0-1} b_j^m \frac{t-iL-\tau_{t+1}}{T} \theta_j^m(k) \Bigg| \\
&\geq \left| \left(a_{j+1}^m + b_{j+1}^m \frac{\tau_{t+1}}{T}\right) \theta_{j+1}^m(k) - \left(a_j^m + b_j^m \frac{s}{T}\right) \theta_j^m(k) \right| \\
&\quad - \left| \frac{2}{w_0} \sum_{i=0}^{w_0/2-1} b_{j+1}^m \frac{t-iL-\tau_{t+1}}{T} \theta_{j+1}^m(k) \right| - \left| \frac{2}{w_0} \sum_{i=w_0/2}^{w_0-1} b_j^m \frac{t-iL-\tau_{t+1}}{T} \theta_j^m(k) \right| \\
&\geq \Delta_0 - \left| \frac{2}{w_0} \sum_{i=0}^{w_0/2-1} b_{j+1}^m \frac{t-iL-\tau_{t+1}}{T} \theta_{j+1}^m(k) \right| - \left| \frac{2}{w_0} \sum_{i=w_0/2}^{w_0-1} b_j^m \frac{t-iL-\tau_{t+1}}{T} \theta_j^m(k) \right| \\
&\geq \Delta_0 - \left(|b_{j+1}^m| \theta_{j+1}^m(k) + |b_j^m| \theta_j^m(k) \right) \frac{w_0 L}{T} \\
&\geq \Delta_0 - 2\bar{b} \frac{w_0 L}{T} \\
&> 2\epsilon_0 - 2\bar{b} \frac{w_0 L}{T}. \qquad (15)
\end{aligned}
$$

The first inequality follows from the triangle inequality. The second inequality follows from the definition of Δ_0. The fourth inequality follows from the definition of $\bar{b}$ and the property that $\theta_{j+1}^m(k), \theta_j^m(k) \in [0,1]$. The fifth inequality follows from the property that $\epsilon_0 < \Delta_0/2$.

Therefore,

$$
\begin{aligned}
\mathbb{P}\left(\left| \bar{X}_t^{m,0} - \bar{X}_{t-w_0L/2}^{m,0} \right| \leq \epsilon_0 \middle| \mathcal{G}_t \right) &= \mathbb{P}\left(\left| \mathsf{E}\left[\bar{X}_t^{m,0} - \bar{X}_{t-w_0L/2}^{m,0} \middle| \mathcal{G}_t \right] + \Delta\bar{X}_t^{m,0} - \Delta\bar{X}_{t-w_0L/2}^{m,0} \right| \leq \epsilon_0 \middle| \mathcal{G}_t \right) \\
&\leq \mathbb{P}\left(\left| \mathsf{E}\left[\bar{X}_t^{m,0} - \bar{X}_{t-w_0L/2}^{m,0} \middle| \mathcal{G}_t \right] \right| - \left| \Delta\bar{X}_t^{m,0} - \Delta\bar{X}_{t-w_0L/2}^{m,0} \right| \leq \epsilon_0 \middle| \mathcal{G}_t \right) \\
&\leq \mathbb{P}\left(\left| \Delta\bar{X}_t^{m,0} - \Delta\bar{X}_{t-w_0L/2}^{m,0} \right| \geq 2\epsilon_0 - 2\bar{b}\frac{w_0L}{T} - \epsilon_0 \middle| \mathcal{G}_t \right) \\
&\leq 2\exp\left(-\frac{1}{2} \left(\left(\epsilon_0 - 2\bar{b}\frac{w_0L}{T} \right)^+ \right)^2 w_0 \right) \\
&= U^0.
\end{aligned}
$$

The second inequality follows from Equation (**??**). The third inequality follows from the Hoeffding's inequality.

Therefore,

$$
\begin{aligned}
\mathbb{P}\left(\left| \bar{X}_t^{m,0} - \bar{X}_{t-w_0L/2}^{m,0} \right| > \epsilon_0 \middle| \mathcal{G}_t \right) &= 1 - \mathbb{P}\left(\left| \bar{X}_t^{m,0} - \bar{X}_{t-w_0L/2}^{m,0} \right| \leq \epsilon_0 \middle| \mathcal{G}_t \right) \\
&\geq 1 - U^0.
\end{aligned}
$$

Next, we prove Part 2.

We have

$$
\begin{aligned}
&\left|\mathsf{E}\left[\frac{2T}{\left(\frac{w_1}{3}-1\right)L}\left(\left(\bar{X}_t^{m,1}-\bar{X}_{t-w_1L/3}^{m,1}\right)-\left(\bar{X}_{t-w_1L/3}^{m,1}-\bar{X}_{t-2w_1L/3}^{m,1}\right)\right)\middle|\mathcal{G}_t\right]\right| \\
={}& \frac{2T}{\left(\frac{w_1}{3}-1\right)L}\left|\frac{3}{w_1}\sum_{i=0}^{w_1/3-1}\left(a_{j+1}^m+b_{j+1}^m\frac{\tau_{t+1}+t-iL-\tau_{t+1}}{T}\right)\theta_{j+1}^m(k)\right. \\
&\qquad -\frac{3}{w_1}\sum_{i=w_1/3}^{2w_1/3-1}\left(a_j^m+b_j^m\frac{t-iL}{T}\right)\theta_j^m(k) \\
&\qquad -\frac{3}{w_1}\sum_{i=w_1/3}^{2w_1/3-1}\left(a_j^m+b_j^m\frac{t-iL}{T}\right)\theta_j^m(k) \\
&\qquad \left.+\frac{3}{w_1}\sum_{i=2w_1/3}^{w_1-1}\left(a_j^m+b_j^m\frac{t-iL}{T}\right)\theta_j^m(k)\right| \\
={}& \frac{2T}{\left(\frac{w_1}{3}-1\right)L}\left|\frac{3}{w_1}\sum_{i=0}^{w_1/3-1}\left(a_{j+1}^m+b_{j+1}^m\frac{\tau_{t+1}+t-iL-\tau_{t+1}}{T}\right)\theta_{j+1}^m(k)\right. \\
&\qquad -\frac{3}{w_1}\sum_{i=0}^{w_1/3-1}\left(a_j^m+b_j^m\frac{\tau_{t+1}+t-iL-\tau_{t+1}}{T}\right)\theta_j^m(k) \\
&\qquad +\frac{3}{w_1}\sum_{i=0}^{w_1/3-1}\left(a_j^m+b_j^m\frac{t-iL}{T}\right)\theta_j^m(k) \\
&\qquad -\frac{3}{w_1}\sum_{i=w_1/3}^{2w_1/3-1}\left(a_j^m+b_j^m\frac{t-iL}{T}\right)\theta_j^m(k) \\
&\qquad -\frac{3}{w_1}\sum_{i=w_1/3}^{2w_1/3-1}\left(a_j^m+b_j^m\frac{t-iL}{T}\right)\theta_j^m(k) \\
&\qquad \left.+\frac{3}{w_1}\sum_{i=2w_1/3}^{w_1-1}\left(a_j^m+b_j^m\frac{t-iL}{T}\right)\theta_j^m(k)\right| \\
={}& \frac{2T}{\left(\frac{w_1}{3}-1\right)L}\left|\frac{3}{w_1}\sum_{i=0}^{w_1/3-1}\left(a_{j+1}^m+b_{j+1}^m\frac{\tau_{t+1}+t-iL-\tau_{t+1}}{T}\right)\theta_{j+1}^m(k)\right. \\
&\qquad \left.-\frac{3}{w_1}\sum_{i=0}^{w_1/3-1}\left(a_j^m+b_j^m\frac{\tau_{t+1}+t-iL-\tau_{t+1}}{T}\right)\theta_j^m(k)\right| \\
={}& \frac{2T}{\left(\frac{w_1}{3}-1\right)L}\left|\frac{3}{w_1}\sum_{i=0}^{w_1/3-1}b_{j+1}^m\frac{t-iL-\tau_{t+1}}{T}\theta_{j+1}^m(k)-\frac{3}{w_1}\sum_{i=0}^{w_1/3-1}b_j^m\frac{t-iL-\tau_{t+1}}{T}\theta_j^m(k)\right| \\
={}& \frac{2T}{\left(\frac{w_1}{3}-1\right)L}\frac{3}{w_1}\sum_{i=0}^{w_1/3-1}\frac{t-iL-\tau_{t+1}}{T}\left|b_{j+1}^m\theta_{j+1}^m-b_j^m\theta_j^m\right| \\
\geq{}& \frac{2T}{\left(\frac{w_1}{3}-1\right)L}\frac{3}{w_1}\sum_{i=0}^{w_1/3-1}\frac{iL}{T}\left|b_{j+1}^m\theta_{j+1}^m-b_j^m\theta_j^m\right|
\end{aligned}
$$

$$
\begin{aligned}
&= \left| b^m_{j+1}\theta^m_{j+1} - b^m_j\theta^m_j \right| \\
&\geq \Delta_1 \\
&> 2\epsilon_1. \qquad (16)
\end{aligned}
$$

The fourth equality follows from Equation (2). The second inequality follows from the definition of Δ_1. The third inequality follows from the property that $\epsilon_1 < \Delta_1/2$.

Therefore,

$$
\begin{aligned}
&\mathbb{P}\left(\left| \frac{2T}{\left(\frac{w_1}{3}-1\right)L} \left(\left(\bar{X}^{m,1}_t - \bar{X}^{m,1}_{t-w_1L/3} \right) - \left(\bar{X}^{m,1}_{t-w_1L/3} - \bar{X}^{m,1}_{t-2w_1L/3} \right) \right) \right| \leq \epsilon_1 \middle| \mathcal{G}_t \right) \\
={}& \mathbb{P}\Bigg(\Bigg| \mathsf{E}\left[\frac{2T}{\left(\frac{w_1}{3}-1\right)L} \left(\left(\bar{X}^{m,1}_t - \bar{X}^{m,1}_{t-w_1L/3} \right) - \left(\bar{X}^{m,1}_{t-w_1L/3} - \bar{X}^{m,1}_{t-2w_1L/3} \right) \right) \middle| \mathcal{G}_t \right] \\
&\qquad + \frac{2T}{\left(\frac{w_1}{3}-1\right)L} \left(\left(\Delta\bar{X}^{m,1}_t - \Delta\bar{X}^{m,1}_{t-w_1L/3} \right) - \left(\Delta\bar{X}^{m,1}_{t-w_1L/3} - \Delta\bar{X}^{m,1}_{t-2w_1L/3} \right) \right) \Bigg| \leq \epsilon_1 \Bigg| \mathcal{G}_t \Bigg) \\
\leq{}& \mathbb{P}\Bigg(\Bigg| \mathsf{E}\left[\frac{2T}{\left(\frac{w_1}{3}-1\right)L} \left(\left(\bar{X}^{m,1}_t - \bar{X}^{m,1}_{t-w_1L/3} \right) - \left(\bar{X}^{m,1}_{t-w_1L/3} - \bar{X}^{m,1}_{t-2w_1L/3} \right) \right) \middle| \mathcal{G}_t \right] \Bigg| \\
&\qquad - \frac{2T}{\left(\frac{w_1}{3}-1\right)L} \left| \left(\Delta\bar{X}^{m,1}_t - \Delta\bar{X}^{m,1}_{t-w_1L/3} \right) - \left(\Delta\bar{X}^{m,1}_{t-w_1L/3} - \Delta\bar{X}^{m,1}_{t-2w_1L/3} \right) \right| \leq \epsilon_1 \Bigg| \mathcal{G}_t \Bigg) \\
\leq{}& \mathbb{P}\left(\frac{2T}{\left(\frac{w_1}{3}-1\right)L} \left| \left(\Delta\bar{X}^{m,1}_t - \Delta\bar{X}^{m,1}_{t-w_1L/3} \right) - \left(\Delta\bar{X}^{m,1}_{t-w_1L/3} - \Delta\bar{X}^{m,1}_{t-2w_1L/3} \right) \right| \geq 2\epsilon_1 - \epsilon_1 \middle| \mathcal{G}_t \right) \\
\leq{}& 2\exp\left(-\frac{w_1}{9} \left(\frac{\left(\frac{w_1}{3}-1\right)L}{2T} \right)^2 \epsilon_1^2 \right) \\
\leq{}& 2\exp\left(-\frac{1}{12} \left(\frac{1}{3} - \frac{1}{w_1} \right)^3 \epsilon_1^2 \frac{w_1^3 L^2}{T^2} \right) \\
\leq{}& 2\exp\left(-\frac{1}{48} \left(\frac{1}{3} - \frac{1}{w_1} \right)^3 \Delta_1^2 \frac{w_1^3 L^2}{T^2} \right) \\
={}& U^1.
\end{aligned}
$$

The second inequality follows from Equation (16). The third inequality follows from the Hoeffding's inequality.

Therefore,

$$
\begin{aligned}
&\mathbb{P}\left(\left| \frac{2T}{\left(\frac{w_1}{3}-1\right)L} \left(\left(\bar{X}^{m,1}_t - \bar{X}^{m,1}_{t-w_1L/3} \right) - \left(\bar{X}^{m,1}_{t-w_1L/3} - \bar{X}^{m,1}_{t-2w_1L/3} \right) \right) \right| > \epsilon_1 \middle| \mathcal{G}_t \right) \\
={}& 1 - \mathbb{P}\left(\left| \frac{2T}{\left(\frac{w_1}{3}-1\right)L} \left(\left(\bar{X}^{m,1}_t - \bar{X}^{m,1}_{t-w_1L/3} \right) - \left(\bar{X}^{m,1}_{t-w_1L/3} - \bar{X}^{m,1}_{t-2w_1L/3} \right) \right) \right| \leq \epsilon_1 \middle| \mathcal{G}_t \right) \\
\geq{}& 1 - 2\exp\left(-\frac{1}{48} \left(\frac{1}{3} - \frac{1}{w_1} \right)^3 \Delta_1^2 \frac{w_1^3 L^2}{T^2} \right).
\end{aligned}
$$

Next, we prove Part 3.

If $t \geq \hat{\tau}_{\hat{j}_t} + (w_0 - 1)L$, then

$$\begin{aligned}
\left|\mathsf{E}\left[\bar{X}_t^{m,0} - \bar{X}_{t-w_0L/2}^{m,0}|\mathcal{G}_t\right]\right| &= \left|\frac{2}{w_0}\sum_{i=0}^{w_0/2-1}\left(a_j^m + b_j^m\frac{s+t-iL-s}{T}\right)\theta_j^m(k)\right. \\
&\qquad \left. - \frac{2}{w_0}\sum_{i=w_0/2}^{w_0-1}\left(a_j^m + b_j^m\frac{s+t-iL-s}{T}\right)\theta_j^m(k)\right| \\
&= \left|\frac{2}{w_0}\sum_{i=0}^{w_0/2-1} b_j^m\frac{t-iL-s}{T}\theta_j^m(k) - \frac{2}{w_0}\sum_{i=w_0/2}^{w_0-1} b_j^m\frac{t-iL-s}{T}\theta_j^m(k)\right| \\
&= |b_j^m|\theta_j^m(k)\frac{w_0L}{2T} \\
&\leq \bar{b}\frac{w_0L}{2T} \\
&< 2\bar{b}\frac{w_0L}{T}.
\end{aligned} \tag{17}$$

The first inequality follows from the definition of $\bar{b}$ and the property that $\theta_j^m(k) \in [0,1]$.

Therefore,

$$\begin{aligned}
\mathbb{P}\left(\left|\bar{X}_t^{m,0} - \bar{X}_{t-w_0L/2}^{m,0}\right| > \epsilon_0 \middle| \mathcal{G}_t\right) &= \mathbb{P}\left(\left|\mathsf{E}\left[\bar{X}_t^{m,0} - \bar{X}_{t-w_0L/2}^{m,0}\middle|\mathcal{G}_t\right] + \Delta\bar{X}_t^{m,0} - \Delta\bar{X}_{t-w_0L/2}^{m,0}\right| > \epsilon_0 \middle| \mathcal{G}_t\right) \\
&\leq \mathbb{P}\left(\left|\mathsf{E}\left[\bar{X}_t^{m,0} - \bar{X}_{t-w_0L/2}^{m,0}\middle|\mathcal{G}_t\right]\right| + \left|\Delta\bar{X}_t^{m,0} - \Delta\bar{X}_{t-w_0L/2}^{m,0}\right| > \epsilon_0 \middle| \mathcal{G}_t\right) \\
&\leq \mathbb{P}\left(\left|\Delta\bar{X}_t^{m,0} - \Delta\bar{X}_{t-w_0L/2}^{m,0}\right| > \epsilon_0 - 2\bar{b}\frac{w_0L}{T} \middle| \mathcal{G}_t\right) \\
&\leq 2\exp\left(-\frac{1}{2}\left(\left(\epsilon_0 - 2\bar{b}\frac{w_0L}{T}\right)^+\right)^2 w_0\right) \\
&= U^0.
\end{aligned}$$

The second inequality follows from Equation (17). The third inequality follows from the Hoeffding's inequality.

If $t \geq \hat{\tau}_{\hat{j}_t} + (w_1 - 1)L$, then

$$\begin{aligned}
&\left|\mathsf{E}\left[\frac{2T}{\left(\frac{w_1}{3}-1\right)L}\left(\left(\bar{X}_t^{m,1} - \bar{X}_{t-w_1L/3}^{m,1}\right) - \left(\bar{X}_{t-w_1L/3}^{m,1} - \bar{X}_{t-2w_1L/3}^{m,1}\right)\right)\middle|\mathcal{G}_t\right]\right| \\
&= \frac{2T}{\left(\frac{w_1}{3}-1\right)L}\left|\frac{3}{w_1}\sum_{i=0}^{w_1/3-1}\left(a_j^m + b_j^m\frac{t-iL}{T}\right)\theta_j^m(k)\right. \\
&\qquad - \frac{3}{w_1}\sum_{i=w_1/3}^{2w_1/3-1}\left(a_j^m + b_j^m\frac{t-iL}{T}\right)\theta_j^m(k) \\
&\qquad - \frac{3}{w_1}\sum_{i=w_1/3}^{2w_1/3-1}\left(a_j^m + b_j^m\frac{t-iL}{T}\right)\theta_j^m(k)
\end{aligned}$$

$$
\begin{aligned}
&\qquad\qquad +\frac{3}{w_1}\sum_{i=2w_1/3}^{w_1-1}\left(a_j^m+b_j^m\frac{t-iL}{T}\right)\theta_j^m\left(k\right)\Bigg|\\
&=0. \qquad\qquad (18)
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\mathbb{P}\left(\left|\frac{2T}{\left(\frac{w_1}{3}-1\right)L}\left(\left(\bar{X}_t^{m,1}-\bar{X}_{t-w_1L/3}^{m,1}\right)-\left(\bar{X}_{t-w_1L/3}^{m,1}-\bar{X}_{t-2w_1L/3}^{m,1}\right)\right)\right|>\epsilon_1\middle|\mathcal{G}_t\right)\\
&=\mathbb{P}\Bigg(\Bigg|\mathsf{E}\left[\frac{2T}{\left(\frac{w_1}{3}-1\right)L}\left(\left(\bar{X}_t^{m,1}-\bar{X}_{t-w_1L/3}^{m,1}\right)-\left(\bar{X}_{t-w_1L/3}^{m,1}-\bar{X}_{t-2w_1L/3}^{m,1}\right)\right)\middle|\mathcal{G}_t\right]\\
&\qquad\quad+\frac{2T}{\left(\frac{w_1}{3}-1\right)L}\left(\left(\Delta\bar{X}_t^{m,1}-\Delta\bar{X}_{t-w_1L/3}^{m,1}\right)-\left(\Delta\bar{X}_{t-w_1L/3}^{m,1}-\Delta\bar{X}_{t-2w_1L/3}^{m,1}\right)\right)\Bigg|>\epsilon_1\Bigg|\mathcal{G}_t\Bigg)\\
&\le\mathbb{P}\Bigg(\Bigg|\mathsf{E}\left[\frac{2T}{\left(\frac{w_1}{3}-1\right)L}\left(\left(\bar{X}_t^{m,1}-\bar{X}_{t-w_1L/3}^{m,1}\right)-\left(\bar{X}_{t-w_1L/3}^{m,1}-\bar{X}_{t-2w_1L/3}^{m,1}\right)\right)\middle|\mathcal{G}_t\right]\Bigg|\\
&\qquad\quad+\frac{2T}{\left(\frac{w_1}{3}-1\right)L}\left|\left(\Delta\bar{X}_t^{m,1}-\Delta\bar{X}_{t-w_1L/3}^{m,1}\right)-\left(\Delta\bar{X}_{t-w_1L/3}^{m,1}-\Delta\bar{X}_{t-2w_1L/3}^{m,1}\right)\right|>\epsilon_1\Bigg|\mathcal{G}_t\Bigg)\\
&=\mathbb{P}\left(\frac{2T}{\left(\frac{w_1}{3}-1\right)L}\left|\left(\Delta\bar{X}_t^{m,1}-\Delta\bar{X}_{t-w_1L/3}^{m,1}\right)-\left(\Delta\bar{X}_{t-w_1L/3}^{m,1}-\Delta\bar{X}_{t-2w_1L/3}^{m,1}\right)\right|>\epsilon_1\middle|\mathcal{G}_t\right)\\
&\le 2\exp\left(-\frac{w_1}{9}\left(\frac{\left(\frac{w_1}{3}-1\right)L\epsilon_1}{2T}\right)^2\right)\\
&\le 2\exp\left(-\frac{1}{12}\left(\frac{1}{3}-\frac{1}{w_1}\right)^3\epsilon_1^2\frac{w_1^3L^2}{T^2}\right)\\
&=U^1.
\end{aligned}
$$

The second equality follows from Equation (18). The second inequality follows from the Hoeffding's inequality.

Q.E.D.

Appendix C: Proofs for §5

Proof of Lemma 2.

Following from the condition given in Theorem 2, we have $w_0L=O\left(T^{1/2}\left(\log T\right)^{1/2}\right)$ and $w_1L=O\left(T^{5/6}\left(\log T\right)^{1/6}\right)$.

1. Consider any $j\in\{0,\cdots,J\}$. Consider the first case that $j,j+1\in\mathcal{N}_0$.

   Following from the condition given in Theorem 2, we have $\tau_{j+1}-\tau_j\in\Omega\left(T^{1/2}\left(\log T\right)^{1/2+\delta_i}\right)$. Because $w_0L=O\left(T^{1/2}\left(\log T\right)^{1/2}\right)=o\left(T^{1/2}\left(\log T\right)^{1/2+\delta_i}\right)$, there exists $\underline{T}$, such that for $T\ge\underline{T}$, Part 1 of this lemma (for this case) holds.

   Consider the second case that at least one out of j and $j+1$ is in $\mathcal{N}_1$.

Following from the condition given in Theorem 2, we have $\tau_{j+1}-\tau_j \in \Omega\left(T^{5/6}\left(\log T\right)^{1/6+\delta_i}\right)$. Because $w_1 L = O\left(T^{5/6}\left(\log T\right)^{1/6}\right) = o\left(T^{5/6}\left(\log T\right)^{1/6+\delta_i}\right)$, there exists $\underline{T}$, such that for $T \geq \underline{T}$, Part 1 of this lemma (for this case) holds.

2. Because $w_1 L = O\left(T^{5/6}\left(\log T\right)^{1/6}\right)$, we have $\frac{w_1 L}{T} = O\left(T^{-1/6}\left(\log T\right)^{1/6}\right) = o\left(1\right)$. Therefore, there exists $\underline{T}$, such that for $T \geq \underline{T}$, Part 1 of this lemma (for this case) holds.

Q.E.D.

Proof of Lemma 3.

We have

$$\begin{aligned}
&\text{Regret}^{\text{CU}}\left(T\right)\\
&= \sum_{t=1}^{T} g_t\left(k_t^*\right) - \mathsf{E}\left[\sum_{t=1}^{T} g_t\left(\pi_t\right)\right]\\
&= \mathsf{E}\left[\sum_{t=1}^{T} g_t\left(k_t^*\right) - g_t\left(\pi_t\right)\right]\\
&= \mathsf{E}\left[\sum_{t=1}^{T} g_t\left(k_t^*\right) - g_t\left(\pi_t\right)\middle| \left\{\cap_{l=1}^{J}\left(\mathcal{A}_l \cap \mathcal{B}_l\right) \cap \mathcal{A}_{J+1}\right\}^c\right] \mathbb{P}\left(\left\{\cap_{l=1}^{J}\left(\mathcal{A}_l \cap \mathcal{B}_l\right) \cap \mathcal{A}_{J+1}\right\}^c\right)\\
&\quad + \mathsf{E}\left[\sum_{t=1}^{T} g_t\left(k_t^*\right) - g_t\left(\pi_t\right)\middle| \cap_{l=1}^{J}\left(\mathcal{A}_l \cap \mathcal{B}_l\right) \cap \mathcal{A}_{J+1}\right] \mathbb{P}\left(\cap_{l=1}^{J}\left(\mathcal{A}_l \cap \mathcal{B}_l\right) \cap \mathcal{A}_{J+1}\right)\\
&\leq M\bar{p}T\mathbb{P}\left(\left\{\cap_{l=1}^{J}\left(\mathcal{A}_l \cap \mathcal{B}_l\right) \cap A_{l+1}\right\}^c\right) + \underbrace{\mathsf{E}\left[\sum_{t=1}^{T} g_t\left(k_t^*\right) - g_t\left(\pi_t\right)\middle| \cap_{l=1}^{J}\left(\mathcal{A}_l \cap \mathcal{B}_l\right) \cap \mathcal{A}_{J+1}\right]}_{H^3}.
\end{aligned}$$

The inequality follows from the properties that $a_j^m + b_j^m \frac{t}{T} \in [0,1]$, $p_k \leq \bar{p}$ and $\theta_j^m\left(\cdot\right) \in [0,1]$ and the property that $\mathbb{P}\left(\cap_{l=1}^{J}\left(\mathcal{A}_l \cap \mathcal{B}_l\right) \cap \mathcal{A}_{J+1}\right) \leq 1$.

Now, we bound $\mathbb{P}\left(\left\{\cap_{l=1}^{J}\left(\mathcal{A}_l \cap \mathcal{B}_l\right) \cap \mathcal{A}_{J+1}\right\}^c\right)$. We have

$$\begin{aligned}
&\mathbb{P}\left(\left\{\cap_{l=1}^{J}\left(\mathcal{A}_l \cap \mathcal{B}_l\right) \cap \mathcal{A}_{J+1}\right\}^c\right)\\
&= \sum_{j=1}^{J+1} \mathbb{P}\left(\cap_{l=1}^{j-1}\left(\mathcal{A}_l \cap \mathcal{B}_l\right) \cap \mathcal{A}_j^c\right) + \sum_{j=1}^{J} \mathbb{P}\left(\cap_{l=1}^{j-1}\left(\mathcal{A}_l \cap \mathcal{B}_l\right) \cap \mathcal{A}_j \cap \mathcal{B}_j^c\right)\\
&\leq \sum_{j=1}^{J+1} \mathbb{P}\left(\mathcal{A}_j^c \middle| \cap_{l=1}^{j-1}\left(\mathcal{A}_l \cap \mathcal{B}_l\right)\right) + \sum_{j=1}^{J} \mathbb{P}\left(\mathcal{B}_j^c \middle| \cap_{l=1}^{j-1}\left(\mathcal{A}_l \cap \mathcal{B}_l\right) \cap \mathcal{A}_j\right).
\end{aligned}$$

Therefore,

$$\text{Regret}^{\text{CU}}\left(T\right) \leq M\bar{p}T \sum_{j=1}^{J+1} \mathbb{P}\left(\mathcal{A}_j^c \middle| \cap_{l=1}^{j-1}\left(\mathcal{A}_l \cap \mathcal{B}_l\right)\right) + M\bar{p}T \sum_{j=1}^{J} \mathbb{P}\left(\mathcal{B}_j^c \middle| \cap_{l=1}^{j-1}\left(\mathcal{A}_l \cap \mathcal{B}_l\right) \cap \mathcal{A}_j\right) + H^3.$$

Q.E.D.

Proof of Lemma 4.

We introduce the following notation before we prove this lemma.

We denote

$$\hat{\mathcal{S}}_{j-1} = \left\{ \hat{\tau}_{j-1} \in \left\{ \tau_{j-1}, \cdots, \tau_{j-1} + \frac{w_0 L}{2} \mathbf{1}\left\{ j \in \mathcal{N}_0 \right\} + \frac{w_1 L}{3} \mathbf{1}\left\{ j \in \mathcal{N}_1 \right\} \right\} \right\}.$$

We denote

$$D_t^0 = \left| \bar{X}_t^{m,0} - \bar{X}_{t - w_0 L/2}^{m,0} \right|$$

and

$$D_t^1 = \left| \frac{2T}{\left(\frac{w_1}{3} - 1 \right) L} \left(\left(\bar{X}_t^{m,1} - \bar{X}_{t - w_1 L/3}^{m,1} \right) - \left(\bar{X}_{t - w_1 L/3}^{m,1} - \bar{X}_{t - 2 w_1 L/3}^{m,1} \right) \right) \right|.$$

We denote

$$\mathcal{S}_j^0 = \left\{ t \in \left\{ \hat{\tau}_{j-1} + \left(w_0 - 1 \right) L, \cdots, \tau_j - 1 \right\} : \operatorname{mod}\left(t - \hat{\tau}_{j-1}, L \right) \in \left\{ 1, \cdots, K \right\} \right\}$$

and

$$\mathcal{S}_j^1 = \left\{ t \in \left\{ \hat{\tau}_{j-1} + \left(w_1 - 1 \right) L, \cdots, \tau_j - 1 \right\} : \operatorname{mod}\left(t - \hat{\tau}_{j-1}, L \right) \in \left\{ 1, \cdots, K \right\} \right\}.$$

We denote

$$\bar{\mathcal{S}}_{j-1} = \left\{ \hat{\tau}_{j-1} \leq \tau_j - \frac{w_0 L}{2} \mathbf{1}\left\{ j \in \mathcal{N}_0 \right\} - \frac{2 w_1 L}{3} \mathbf{1}\left\{ j \in \mathcal{N}_1 \right\} \right\}.$$

Without loss of generality, we assume that the change at time τ_j happens at price p_k. We denote

$$s = \min\left\{ t \geq \tau_j : \operatorname{mod}\left(t - \hat{\tau}_{j-1}, L \right) = k \right\}.$$

Now, we prove Part 1. We have

$$\begin{aligned}
\mathbb{P}\left(\mathcal{A}_j^c \middle| \cap_{l=1}^{j-1} \left(\mathcal{A}_l \cap \mathcal{B}_l \right) \right) &= \mathbb{P}\left(\hat{\tau}_j \leq \tau_j - 1 \middle| \hat{\mathcal{S}}_{j-1} \right) \\
&= \mathbb{P}\left(D_t^0 > \epsilon_0, \ \exists\ t \in \mathcal{S}_j^0, \ \text{or } D_t^1 > \epsilon_1, \ \exists\ t \in \mathcal{S}_j^1 \middle| \hat{\mathcal{S}}_{j-1} \right) \\
&\leq \sum_{t \in \mathcal{S}_j^0} \mathbb{P}\left(D_t^0 > \epsilon_0 \middle| \hat{\mathcal{S}}_{j-1} \right) + \sum_{t \in \mathcal{S}_j^1} \mathbb{P}\left(D_t^1 > \epsilon_1 \middle| \hat{\mathcal{S}}_{j-1} \right) \\
&\leq |\mathcal{S}_j^0| U^0 + |\mathcal{S}_j^1| U^1 \\
&\leq K \frac{T}{L} U^0 + K \frac{T}{L} U^1 \\
&= K \frac{T}{L} \left(U^0 + U^1 \right).
\end{aligned}$$

The second inequality follows from Lemma 1 Part 3.

Therefore,

$$H^1 \le M\bar{p}K\left(J+1\right)\frac{T^2}{L}\left(U^0+U^1\right).$$

Next, we prove Part 2. We have

$$\begin{aligned}
\mathbb{P}\left(\mathcal{B}_j^c \middle| \cap_{l=1}^{j-1}\left(\mathcal{A}_l\cap\mathcal{B}_l\right)\cap\mathcal{A}_j\right) &= \mathbb{P}\left(\hat{\tau}_j > \tau_j + \frac{w_0 L}{2}\mathbf{1}\left\{j\in\mathcal{N}_0\right\} + \frac{w_1 L}{3}\mathbf{1}\left\{j\in\mathcal{N}_1\right\} \middle| \hat{\mathcal{S}}_{j-1}, \hat{\tau}_j \ge \tau_j\right) \\
&= \mathbb{P}\left(\hat{\tau}_j > \tau_j + \frac{w_0 L}{2}\mathbf{1}\left\{j\in\mathcal{N}_0\right\} + \frac{w_1 L}{3}\mathbf{1}\left\{j\in\mathcal{N}_1\right\} \middle| \bar{\mathcal{S}}_{j-1}, \hat{\tau}_j \ge \tau_j\right) \\
&\le 1 - \mathbb{P}\left(D^0_{s+(w_0/2-1)L} > \epsilon_0 \text{ or } D^1_{s+(w_1/3-1)L} > \epsilon_1 \middle| \bar{\mathcal{S}}_{j-1}, \hat{\tau}_j \ge \tau_j\right) \\
&\le 1 - \left(1-U^0\right)\mathbf{1}\left\{j\in\mathcal{N}_0\right\} - \left(1-U^1\right)\mathbf{1}\left\{j\in\mathcal{N}_1\right\} \\
&= U^0\mathbf{1}\left\{j\in\mathcal{N}_0\right\} + U^1\mathbf{1}\left\{j\in\mathcal{N}_1\right\}.
\end{aligned}$$

The second equality follows from Lemma 2. The second inequality follows from Lemma 1 Parts 1 and 2.

Therefore,

$$H^2 \le M\bar{p}T\left(U^0|\mathcal{N}_0| + U^1|\mathcal{N}_1|\right).$$

Q.E.D.

Proof of Lemma 5.

We have

$$\begin{aligned}
H^3 &= \mathsf{E}\left[\sum_{j=0}^{J}\sum_{t=\hat{\tau}_j}^{\hat{\tau}_{j+1}-1} g_t\left(k_t^*\right) - g_t\left(\pi_t^{\mathrm{CU}}\right) \middle| \cap_{l=1}^{J}\left(\mathcal{A}_l\cap\mathcal{B}_l\right)\cap\mathcal{A}_{J+1}\right] \\
&= \mathsf{E}\left[\sum_{j=0}^{J}\sum_{t=\hat{\tau}_j}^{\hat{\tau}_{j+1}-1} \left(g_t\left(k_t^*\right) - g_t\left(\pi_t^{\mathrm{CU}}\right)\right)\mathbf{1}\left\{\cap_{l=1}^{J}\left(\mathcal{A}_l\cap\mathcal{B}_l\right)\cap\mathcal{A}_{J+1}\right\}\right] \\
&\quad\cdot\mathbb{P}\left(\cap_{l=1}^{J}\left(\mathcal{A}_l\cap\mathcal{B}_l\right)\cap\mathcal{A}_{J+1}\right)^{-1} \\
&= \mathsf{E}\left[\sum_{j=0}^{J}\sum_{t=\hat{\tau}_j}^{\hat{\tau}_{j+1}-1} \left(g_t\left(k_t^*\right) - g_t\left(\pi_t^{\mathrm{CU},\hat{\tau}_j}\right)\right)\mathbf{1}\left\{\cap_{l=1}^{J}\left(\mathcal{A}_l\cap\mathcal{B}_l\right)\cap\mathcal{A}_{J+1}\right\}\right] \\
&\quad\cdot\mathbb{P}\left(\cap_{l=1}^{J}\left(\mathcal{A}_l\cap\mathcal{B}_l\right)\cap\mathcal{A}_{J+1}\right)^{-1} \\
&= \mathsf{E}\left[\sum_{j=0}^{J}\sum_{t=\hat{\tau}_j}^{\hat{\tau}_{j+1}-1} g_t\left(k_t^*\right) - g_t\left(\pi_t^{\mathrm{CU},\hat{\tau}_j}\right) \middle| \cap_{l=1}^{J}\left(\mathcal{A}_l\cap\mathcal{B}_l\right)\cap\mathcal{A}_{J+1}\right] \\
&= \mathsf{E}_{\{\hat{\tau}_1,\cdots,\hat{\tau}_J\}}\left[\mathsf{E}\left[\sum_{j=0}^{J}\sum_{t=\hat{\tau}_j}^{\hat{\tau}_{j+1}-1} g_t\left(k_t^*\right) - g_t\left(\pi_t^{\mathrm{CU},\hat{\tau}_j}\right) \middle| \hat{\tau}_1,\cdots,\hat{\tau}_J\right] \middle| \cap_{l=1}^{J}\left(\mathcal{A}_l\cap\mathcal{B}_l\right)\cap\mathcal{A}_{J+1}\right]
\end{aligned}$$

$$
\begin{aligned}
&= \mathsf{E}_{\{\hat{\tau}_1,\cdots,\hat{\tau}_J\}}\left[\sum_{j=0}^{J}\mathsf{E}\left[\sum_{t=\hat{\tau}_j}^{\hat{\tau}_{j+1}-1} g_t\left(k_t^*\right)-g_t\left(\pi_t^{\mathrm{CU},\hat{\tau}_j}\right)\middle|\hat{\tau}_j,\hat{\tau}_{j+1}\right]\middle|\cap_{l=1}^{J}\left(\mathcal{A}_l\cap\mathcal{B}_l\right)\cap\mathcal{A}_{J+1}\right]\\
&= \sum_{j=0}^{J}\mathsf{E}_{\{\hat{\tau}_1,\cdots,\hat{\tau}_J\}}\left[\mathsf{E}\left[\sum_{t=\hat{\tau}_j}^{\hat{\tau}_{j+1}-1} g_t\left(k_t^*\right)-g_t\left(\pi_t^{\mathrm{CU},\hat{\tau}_j}\right)\middle|\hat{\tau}_j,\hat{\tau}_{j+1}\right]\middle|\cap_{l=1}^{J}\left(\mathcal{A}_l\cap\mathcal{B}_l\right)\cap\mathcal{A}_{J+1}\right].
\end{aligned}
$$

The third equality follows from the definition of $\pi^{\mathrm{CU},\hat{\tau}_j}$.

Q.E.D.

Proof of Lemma 6.

For Part 1, we have

$$H_{j,1}^3 \leq 2M\bar{p}K.$$

The inequality follows from the properties that $a_j^m + b_j^m \frac{t}{T} \in [0,1]$, $p_k \leq \bar{p}$ and $\theta_j^m(\cdot) \in [0,1]$.

For Part 2, we have

$$
\begin{aligned}
H_{j,2}^3 &\leq M\bar{p}K\left\lceil\frac{\hat{\tau}_{j+1}-\hat{\tau}_j-2K}{L}\right\rceil\\
&\leq M\bar{p}K\left(\frac{\hat{\tau}_{j+1}-\hat{\tau}_j-2K}{L}+1\right)\\
&\leq M\bar{p}K\left(\frac{\hat{\tau}_{j+1}-\hat{\tau}_j}{L}+1\right).
\end{aligned}
$$

The first inequality follows from the properties that $a_j^m + b_j^m \frac{t}{T} \in [0,1]$, $p_k \leq \bar{p}$ and $\theta_j^m(\cdot) \in [0,1]$.

Q.E.D.

Proof of Lemma 7.

We have

$$
\begin{aligned}
H_{j,3}^3 &= \mathsf{E}\left[\sum_{t\in\mathcal{T}_{j,<}^{\mathrm{UCB}}} g_t\left(k_t^*\right)-g_t\left(\pi_t^{\hat{\tau}_j}\right)\right]+\mathsf{E}\left[\sum_{t\in\mathcal{T}_j^{\mathrm{UCB}}\setminus\mathcal{T}_{j,<}^{\mathrm{UCB}}} g_t\left(k_t^*\right)-g_t\left(\pi_t^{\hat{\tau}_j}\right)\right]\\
&\leq \mathsf{E}\left[\sum_{t\in\mathcal{T}_{j,<}^{\mathrm{UCB}}} g_t\left(k_t^*\right)-g_t\left(\pi_t^{\hat{\tau}_j}\right)\right]+\mathsf{E}\left[\sum_{t=\tau_{j+1}}^{\hat{\tau}_{j+1}-1} g_t\left(k_t^*\right)-g_t\left(\pi_t^{\hat{\tau}_j}\right)\right]\\
&\leq \mathsf{E}\left[\sum_{t\in\mathcal{T}_{j,<}^{\mathrm{UCB}}} g_t\left(k_t^*\right)-g_t\left(\pi_t^{\hat{\tau}_j}\right)\right]+M\bar{p}\left(\hat{\tau}_{j+1}-\tau_{j+1}\right)\\
&\leq \mathsf{E}\left[\sum_{t\in\mathcal{T}_{j,<}^{\mathrm{UCB}}} g_t\left(k_t^*\right)-g_t\left(\pi_t^{\hat{\tau}_j}\right)\right]+M\bar{p}\frac{w_0L}{2}
\end{aligned}
$$

$$
\begin{aligned}
&= \mathsf{E}\left[\sum_{t\in\mathcal{T}_{j,<}^{\mathrm{UCB}}} g_t\left(k_t^*\right)-U_t\left(\pi_t^{\hat{\tau}_j}\right)+U_t\left(\pi_t^{\hat{\tau}_j}\right)-L_t\left(\pi_t^{\hat{\tau}_j}\right)+L_t\left(\pi_t^{\hat{\tau}_j}\right)-g_t\left(\pi_t^{\hat{\tau}_j}\right)\right]\\
&\quad +M\bar{p}\frac{w_0L}{2}\\
&\leq \mathsf{E}\left[\sum_{t\in\mathcal{T}_{j,<}^{\mathrm{UCB}}} g_t\left(k_t^*\right)-U_t\left(k_t^*\right)+U_t\left(\pi_t^{\hat{\tau}_j}\right)-L_t\left(\pi_t^{\hat{\tau}_j}\right)+L_t\left(\pi_t^{\hat{\tau}_j}\right)-g_t\left(\pi_t^{\hat{\tau}_j}\right)\right]\\
&\quad +M\bar{p}\frac{w_0L}{2}\\
&= \sum_{t\in\mathcal{T}_{j,<}^{\mathrm{UCB}}}\mathsf{E}\left[g_t\left(k_t^*\right)-U_t\left(k_t^*\right)\right]+\sum_{t\in\mathcal{T}_{j,<}^{\mathrm{UCB}}}\mathsf{E}\left[U_t\left(\pi_t^{\hat{\tau}_j}\right)-L_t\left(\pi_t^{\hat{\tau}_j}\right)\right]\\
&\quad +\sum_{t\in\mathcal{T}_{j,<}^{\mathrm{UCB}}}\mathsf{E}\left[L_t\left(\pi_t^{\hat{\tau}_j}\right)-g_t\left(\pi_t^{\hat{\tau}_j}\right)\right]+M\bar{p}\frac{w_0L}{2}.
\end{aligned}
$$

The second inequality follows from the properties that $a_j^m+b_j^m\frac{t}{T}\in[0,1]$, $p_k\leq\bar{p}$ and $\theta_j^m(\cdot)\in[0,1]$. The third inequality follows from the definition of $\mathcal{B}_{j+1}$ and the property that $j+1\in\mathcal{N}_0$. The fourth inequality follows from the definition of $U_t(\cdot)$.

For any $t\in\mathcal{T}_{j,<}^{\mathrm{UCB}}$ and $k\in\{1,\cdots,K\}$, we have

$$
\begin{aligned}
\mathsf{E}\left[g_t(k)-U_t(k)\right] &= \mathsf{E}\left[\left(g_t(k)-U_t(k)\right)\mathbf{1}\left\{g_t(k)\geq U_t(k)\right\}\right]\\
&\quad +\mathsf{E}\left[\left(g_t(k)-U_t(k)\right)\mathbf{1}\left\{g_t(k)<U_t(k)\right\}\right]\\
&\leq \mathsf{E}\left[\left(g_t(k)-U_t(k)\right)\mathbf{1}\left\{g_t(k)\geq U_t(k)\right\}\right]\\
&\leq M\bar{p}\mathbb{P}\left(g_t(k)\geq U_t(k)\right).
\end{aligned}
$$

The second inequality follows from the properties that $a_j^m+b_j^m\frac{t}{T}\in[0,1]$, $p_k\leq\bar{p}$ and $\theta_j^m(\cdot)\in[0,1]$ and the property that $U_t(\cdot)\geq 0$.

For any $t\in\mathcal{T}_{j,<}^{\mathrm{UCB}}$ and $k\in\{1,\cdots,K\}$, we have

$$
\begin{aligned}
\mathsf{E}\left[L_t(k)-g_t(k)\right] &= \mathsf{E}\left[\left(L_t(k)-g_t(k)\right)\mathbf{1}\left\{L_t(k)\geq g_t(k)\right\}\right]\\
&\quad +\mathsf{E}\left[\left(L_t(k)-g_t(k)\right)\mathbf{1}\left\{L_t(k)<g_t(k)\right\}\right]\\
&\leq \mathsf{E}\left[\left(L_t(k)-g_t(k)\right)\mathbf{1}\left\{L_t(k)\geq g_t(k)\right\}\right]\\
&\leq M\bar{p}\mathbb{P}\left(L_t(k)\geq g_t(k)\right).
\end{aligned}
$$

The second inequality follows from the properties that $a_j^m+b_j^m\frac{t}{T}\in[0,1]$ and $\theta_j^m(\cdot)\in[0,1]$ and the property that $L_t(\cdot)\leq M\bar{p}$.

Therefore,

$$
H_{j,3}^3\leq M\bar{p}\sum_{t\in\mathcal{T}_{j,<}^{\mathrm{UCB}}}\mathbb{P}\left(g_t\left(k_t^*\right)\geq U_t\left(k_t^*\right)\right)+\sum_{t\in\mathcal{T}_{j,<}^{\mathrm{UCB}}}\mathsf{E}\left[U_t\left(\pi_t^{\hat{\tau}_j}\right)-L_t\left(\pi_t^{\hat{\tau}_j}\right)\right]
$$

$$+M\bar{p}\sum_{t\in\mathcal{T}_{j,<}^{\mathrm{UCB}}}\mathbb{P}\left(L_t\left(\pi_t^{\hat{\tau}_j}\right)\geq g_t\left(\pi_t^{\hat{\tau}_j}\right)\right)+M\bar{p}\frac{w_0L}{2}.$$

Q.E.D.

Proof of Lemma 8.

We define

$$\Delta X_t^m\triangleq X_t^m-\mathsf{E}\left[X_t^m|\mathcal{G}_t\right]$$

and

$$v_t^m\left(k\right)\triangleq\frac{\sum_{t'=\hat{\tau}_j}^{t-1}\Delta X_{t'}^m\mathbf{1}\left\{\pi_{t'}=k\right\}}{N_{\hat{\tau}_j,t-1}\left(k\right)}$$

and

$$w_t^m\left(s\right)\triangleq\frac{s-s_t^{(1)}}{s_t^{(2)}-s_t^{(1)}}\frac{\sum_{t'\in\mathcal{T}_t^{(2)}}\Delta X_{t'}^m}{\left\lceil\frac{N_{\hat{\tau}_j,t-1}(k_t)}{2}\right\rceil}-\frac{s-s_t^{(2)}}{s_t^{(2)}-s_t^{(1)}}\frac{\sum_{t'\in\mathcal{T}_t^{(1)}}\Delta X_{t'}^m}{\left\lfloor\frac{N_{\hat{\tau}_j,t-1}(k_t)}{2}\right\rfloor},\ \forall\ s\in\left\{s_t\left(k\right),t\right\}.$$

Therefore,

$$v_t^m\left(k\right)=d_t^m\left(k\right)-\left(a_j^m+b_j^m\frac{s_t\left(k\right)}{T}\right)\theta_j^m\left(k\right)$$

and

$$w_t^m\left(s\right)=\frac{s-s_t^{(1)}}{s_t^{(2)}-s_t^{(1)}}d_t^{m,(2)}-\frac{s-s_t^{(2)}}{s_t^{(2)}-s_t^{(1)}}d_t^{m,(1)}-\left(a_j^m+b_j^m\frac{s}{T}\right)\theta_j^m\left(k_t\right),\ \forall\ s\in\left\{s_t\left(k\right),t\right\}.$$

We define

$$\mathcal{W}_t^m\triangleq\left\{\left|w_t^m\left(s_t\left(k\right)\right)\right|\leq\frac{\underline{d\theta}}{2}\right\}.$$

Hence, conditional on $\mathcal{W}_t^m$,

$$\begin{aligned}\left(a_j^m+b_j^m\frac{s_t\left(k\right)}{T}\right)\theta_j^m\left(k_t\right),\left(a_j^m+b_j^m\frac{s_t\left(k\right)}{T}\right)\theta_j^m\left(k_t\right)+w_t^m\left(s_t\left(k\right)\right)&\in\left[\underline{d\theta}-\frac{\underline{d\theta}}{2},1+\frac{\underline{d\theta}}{2}\right]\\&\subseteq\left[\frac{\underline{d\theta}}{2},\frac{3}{2}\right].\end{aligned}$$

Therefore, conditional on $\mathcal{W}_t^m$,

$$\begin{aligned}&\left|D_t^m\left(k\right)-\left(a_j^m+b_j^m\frac{t}{T}\right)\theta_j^m\left(k\right)\right|\\&=\left|\left(\left(a_j^m+b_j^m\frac{s_t\left(k\right)}{T}\right)\theta_j^m\left(k\right)+v_t^m\left(k\right)\right)\frac{\left(a_j^m+b_j^m\frac{t}{T}\right)\theta_j^m\left(k_t\right)+w_t^m\left(t\right)}{\left(a_j^m+b_j^m\frac{s_t(k)}{T}\right)\theta_j^m\left(k_t\right)+w_t^m\left(s_t\left(k\right)\right)}\right.\\&\qquad\left.-\left(\left(a_j^m+b_j^m\frac{s_t\left(k\right)}{T}\right)\theta_j^m\left(k\right)\right)\frac{\left(a_j^m+b_j^m\frac{t}{T}\right)\theta_j^m\left(k_t\right)}{\left(a_j^m+b_j^m\frac{s_t(k)}{T}\right)\theta_j^m\left(k_t\right)}\right|\\&\leq\frac{4}{\underline{d}^2\underline{\theta}^2}\left(\frac{3}{2}\left|v_t^m\left(k\right)\right|+\frac{3}{2}\left|w_t^m\left(s_t\left(k\right)\right)\right|+\left|w_t^m\left(t\right)\right|\right)\\&\leq\frac{6}{\underline{d}^2\underline{\theta}^2}\left(\left|v_t^m\left(k\right)\right|+\left|w_t^m\left(s_t\left(k\right)\right)\right|+\left|w_t^m\left(t\right)\right|\right).\end{aligned}\tag{19}$$

The first inequality follows from Lemma 14.

Now, we prove Part 1. For any $t \in \mathcal{T}_{j,<}^{\mathrm{UCB}}$ and $k \in \{1, \cdots, K\}$, we have

$$
\begin{aligned}
&\mathbb{P}\left(g_t(k) \geq U_t(k)\right) \\
&\leq \mathbb{P}\left(g_t(k) \geq p_k \sum_{m=1}^{M} D_t^m(k) + \beta M p_k \sqrt{\frac{\log T}{N_{\hat{\tau}_j,t-1}(k)}} + 2\beta M p_k \sqrt{\frac{\log T}{N_{\hat{\tau}_j,t-1}(k_t)}}\right) \\
&\leq \mathbb{P}\left(\left|g_t(k) - p_k \sum_{m=1}^{M} D_t^m(k)\right| \geq \beta M p_k \sqrt{\frac{\log T}{N_{\hat{\tau}_j,t-1}(k)}} + 2\beta M p_k \sqrt{\frac{\log T}{N_{\hat{\tau}_j,t-1}(k_t)}}\right) \\
&= \mathbb{P}\left(\left|p_k \sum_{m=1}^{M} \left(a_j^m + b_j^m \frac{t}{T}\right)\theta_j^m(k) - p_k \sum_{m=1}^{M} D_t^m(k)\right| \geq \beta M p_k \sqrt{\frac{\log T}{N_{\hat{\tau}_j,t-1}(k)}} + 2\beta M p_k \sqrt{\frac{\log T}{N_{\hat{\tau}_j,t-1}(k_t)}}\right) \\
&\leq \sum_{m=1}^{M} \mathbb{P}\left(\left|\left(a_j^m + b_j^m \frac{t}{T}\right)\theta_j^m(k) - D_t^m(k)\right| \geq \beta \sqrt{\frac{\log T}{N_{\hat{\tau}_j,t-1}(k)}} + 2\beta \sqrt{\frac{\log T}{N_{\hat{\tau}_j,t-1}(k_t)}}\right) \\
&= \sum_{m=1}^{M} \mathbb{P}\left(\left|\left(a_j^m + b_j^m \frac{t}{T}\right)\theta_j^m(k) - D_t^m(k)\right| \geq \beta \sqrt{\frac{\log T}{N_{\hat{\tau}_j,t-1}(k)}} + 2\beta \sqrt{\frac{\log T}{N_{\hat{\tau}_j,t-1}(k_t)}}, \mathcal{W}_t^m\right) \\
&\quad + \sum_{m=1}^{M} \mathbb{P}\left(\left|\left(a_j^m + b_j^m \frac{t}{T}\right)\theta_j^m(k) - D_t^m(k)\right| \geq \beta \sqrt{\frac{\log T}{N_{\hat{\tau}_j,t-1}(k)}} + 2\beta \sqrt{\frac{\log T}{N_{\hat{\tau}_j,t-1}(k_t)}}, \mathcal{W}_t^{m,c}\right) \\
&\leq \sum_{m=1}^{M} \mathbb{P}\left(\frac{6}{\underline{d}^2\underline{\theta}^2}\left(|v_t^m(k)| + |w_t^m(s_t(k))| + |w_t^m(t)|\right) \geq \beta \sqrt{\frac{\log T}{N_{\hat{\tau}_j,t-1}(k)}} + 2\beta \sqrt{\frac{\log T}{N_{\hat{\tau}_j,t-1}(k_t)}}, \mathcal{W}_t^m\right) \\
&\quad + \sum_{m=1}^{M} \mathbb{P}\left(\mathcal{W}_t^{m,c}\right) \\
&\leq \sum_{m=1}^{M} \mathbb{P}\left(|v_t^m(k)| \geq \beta \frac{\underline{d}^2\underline{\theta}^2}{6} \sqrt{\frac{\log T}{N_{\hat{\tau}_j,t-1}(k)}}, \mathcal{W}_t^m\right) + \sum_{m=1}^{M} \mathbb{P}\left(|w_t^m(s_t(k))| \geq \beta \frac{\underline{d}^2\underline{\theta}^2}{6} \sqrt{\frac{\log T}{N_{\hat{\tau}_j,t-1}(k_t)}}, \mathcal{W}_t^m\right) \\
&\quad + \sum_{m=1}^{M} \mathbb{P}\left(|w_t^m(t)| \geq \beta \frac{\underline{d}^2\underline{\theta}^2}{6} \sqrt{\frac{\log T}{N_{\hat{\tau}_j,t-1}(k_t)}}, \mathcal{W}_t^m\right) + \sum_{m=1}^{M} \mathbb{P}\left(\mathcal{W}_t^{m,c}\right) \\
&\leq \sum_{m=1}^{M} \underbrace{\mathbb{P}\left(|v_t^m(k)| \geq \beta \frac{\underline{d}^2\underline{\theta}^2}{6} \sqrt{\frac{\log T}{N_{\hat{\tau}_j,t-1}(k)}}\right)}_{Q_1^m} + \sum_{m=1}^{M} \underbrace{\mathbb{P}\left(|w_t^m(s_t(k))| \geq \beta \frac{\underline{d}^2\underline{\theta}^2}{6} \sqrt{\frac{\log T}{N_{\hat{\tau}_j,t-1}(k_t)}}\right)}_{Q_2^m} \\
&\quad + \sum_{m=1}^{M} \underbrace{\mathbb{P}\left(|w_t^m(t)| \geq \beta \frac{\underline{d}^2\underline{\theta}^2}{6} \sqrt{\frac{\log T}{N_{\hat{\tau}_j,t-1}(k_t)}}\right)}_{Q_3^m} + \sum_{m=1}^{M} \underbrace{\mathbb{P}\left(\mathcal{W}_t^{m,c}\right)}_{Q_4^m}.
\end{aligned} \tag{20}
$$

The first inequality follows from the property that $\mathbb{P}(X \geq \max\{Y, 0\}) \leq \mathbb{P}(X \geq Y)$. The second inequality follows from the property that $\mathbb{P}(X \geq Y) \leq \mathbb{P}(|X| \geq Y)$. The first equality follows from the definition of $g_t(k)$. The third and the fifth inequalities follow from the property that $\mathbb{P}(X + Y \geq a + b) \leq \mathbb{P}(X \geq a) + \mathbb{P}(Y \geq b)$. The fourth and the sixth inequalities follow from the property that $\mathbb{P}(X, Y) \leq \mathbb{P}(X)$.

Next, we establish upper bounds on Q_1^m, Q_2^m, Q_3^m and Q_4^m, respectively.

For Q_1^m, we have

$$\begin{aligned} Q_1^m &= \mathbb{P}\left(\left|\frac{\sum_{t'=\hat{\tau}_j}^{t-1}\Delta X_{t'}^m \mathbf{1}\{\pi_{t'}=k\}}{N_{\hat{\tau}_j,t-1}(k)}\right| \ge \beta\frac{\underline{d}^2\underline{\theta}^2}{6}\sqrt{\frac{\log T}{N_{\hat{\tau}_j,t-1}(k)}}\right) \\ &\le 2\exp\left(-\frac{1}{2\cdot 3^2}\left(\beta\underline{d}^2\underline{\theta}^2\right)^2\log T\right) \\ &= 2T^{-\left(\beta\underline{d}^2\underline{\theta}^2\right)^2/2\cdot 3^2}. \end{aligned} \tag{21}$$

The inequality follows from Hoeffding's inequality.

Next, we bound Q_2^m, Q_3^m and Q_m^4. We denote

$$\alpha^{(2)} = \frac{s - s_t^{(1)}}{s_t^{(2)} - s_t^{(1)}}$$

and

$$\alpha^{(1)} = \frac{s - s_t^{(2)}}{s_t^{(2)} - s_t^{(1)}} \frac{\left\lceil \frac{N_{\hat{\tau}_j,t-1}(k_t)}{2} \right\rceil}{\left\lfloor \frac{N_{\hat{\tau}_j,t-1}(k_t)}{2} \right\rfloor}.$$

Hence, for $s \in \{s_t(k), t\}$, Lemma 15 implies $|\alpha^{(2)}| \le 2K+1 \le 6K$ and $|\alpha^{(1)}| \le 2(2K+1) \le 6K$. Therefore, for $s \in \{s_t(k), t\}$ and $a > 0$,

$$\begin{aligned} &\mathbb{P}\left(|w_t^m(\tau)| \ge a\right) \\ &= \mathbb{P}\left(\left|\frac{s - s_t^{(1)}}{s_t^{(2)} - s_t^{(1)}} \frac{\sum_{t' \in \mathcal{T}_t^{(2)}} \Delta X_{t'}^m}{\left\lceil \frac{N_{\hat{\tau}_j,t-1}(k_t)}{2} \right\rceil} - \frac{s - s_t^{(2)}}{s_t^{(2)} - s_t^{(1)}} \frac{\sum_{t' \in \mathcal{T}_t^{(1)}} \Delta X_{t'}^m}{\left\lfloor \frac{N_{\hat{\tau}_j,t-1}(k_t)}{2} \right\rfloor}\right| \ge a\right) \\ &= \mathbb{P}\left(\left|\frac{1}{N_{\hat{\tau}_j,t-1}(k_t)}\left(\alpha^{(2)}\sum_{t' \in \mathcal{T}_t^{(2)}} \Delta X_{t'}^m - \alpha^{(1)}\sum_{t' \in \mathcal{T}_t^{(1)}} \Delta X_{t'}^m\right)\right| \ge a\frac{\left\lceil \frac{N_{\hat{\tau}_j,t-1}(k_t)}{2} \right\rceil}{N_{\hat{\tau}_j,t-1}(k_t)}\right) \\ &\le \mathbb{P}\left(\left|\frac{1}{N_{\hat{\tau}_j,t-1}(k_t)}\left(\alpha^{(2)}\sum_{t' \in \mathcal{T}_t^{(2)}} \Delta X_{t'}^m - \alpha^{(1)}\sum_{t' \in \mathcal{T}_t^{(1)}} \Delta X_{t'}^m\right)\right| \ge \frac{a}{2}\right) \\ &\le 2\exp\left(-\frac{N_{\hat{\tau}_j,t-1}(k_t)^2 a^2/2}{|\mathcal{T}_t^{(1)}|\left(\alpha^{(1)}\right)^2 + |\mathcal{T}_t^{(2)}|\left(\alpha^{(2)}\right)^2}\right) \\ &\le 2\exp\left(-\frac{N_{\hat{\tau}_j,t-1}(k_t)\,a^2}{72K^2}\right). \end{aligned}$$

The second inequality follows from Hoeffding's inequality. The third inequality follows from the property that $|\alpha^{(1)}|, |\alpha^{(2)}| \le 6K$.

Therefore, for Q_2^m and Q_3^m, we have

$$\begin{aligned} Q_2^m, Q_3^m &\leq 2\exp\left(-\frac{\left(\beta \underline{d}^2 \underline{\theta}^2\right)^2 \log T}{2\cdot 36^2 K^2}\right) \\ &= 2T^{-\left(\beta \underline{d}^2 \underline{\theta}^2\right)^2/2\cdot 36^2 K^2}. \end{aligned} \tag{22}$$

For Q_4^m, we have

$$\begin{aligned} Q_4^m &\leq 2\exp\left(-\frac{N_{\hat{\tau}_j,t-1}\left(k_t\right)\left(\underline{d\theta}\right)^2}{72K^2}\right) \\ &\leq 2\exp\left(-\frac{\left(t-\hat{\tau}_j\right)\left(\underline{d\theta}\right)^2}{72K^3}\right). \end{aligned} \tag{23}$$

The second inequality follows from the property that $N_{\hat{\tau}_j,t-1}\left(k_t\right) \geq \lceil \frac{t-\hat{\tau}_j}{K} \rceil \geq \frac{t-\hat{\tau}_j}{K}$.

Therefore,

$$\begin{aligned} &\sum_{t\in\mathcal{T}_{j,<}^{\text{UCB}}} \mathbb{P}\left(g_t\left(k_t^*\right) \geq U_t\left(k_t^*\right)\right) \\ \leq\ & \sum_{t\in\mathcal{T}_{j,<}^{\text{UCB}}}\sum_{m=1}^{M} Q_1^m + \sum_{t\in\mathcal{T}_{j,<}^{\text{UCB}}}\sum_{m=1}^{M} Q_2^m + \sum_{t\in\mathcal{T}_{j,<}^{\text{UCB}}}\sum_{m=1}^{M} Q_3^m + \sum_{t\in\mathcal{T}_{j,<}^{\text{UCB}}}\sum_{m=1}^{M} Q_4^m \\ \leq\ & 2M|\mathcal{T}_{j,<}^{\text{UCB}}|T^{-\left(\beta \underline{d}^2 \underline{\theta}^2\right)^2/2\cdot 3^2} + 2M|\mathcal{T}_{j,<}^{\text{UCB}}|T^{-\left(\beta \underline{d}^2 \underline{\theta}^2\right)^2/2\cdot 36^2 K^2} + 2M|\mathcal{T}_{j,<}^{\text{UCB}}|T^{-\left(\beta \underline{d}^2 \underline{\theta}^2\right)^2/2\cdot 36^2 K^2} \\ &\quad + 2M \sum_{t\in\mathcal{T}_{j,<}^{\text{UCB}}} \exp\left(-\frac{\left(t-\hat{\tau}_j\right)\left(\underline{d\theta}\right)^2}{72K^3}\right) \\ \leq\ & 6M|\mathcal{T}_{j,<}^{\text{UCB}}|T^{-\left(\beta \underline{d}^2 \underline{\theta}^2\right)^2/2\cdot 36^2 K^2} + 2M\sum_{t=\hat{\tau}_j}^{\infty} \exp\left(-\frac{\left(t-\hat{\tau}_j\right)\left(\underline{d\theta}\right)^2}{72K^3}\right) \\ =\ & 6M|\mathcal{T}_{j,<}^{\text{UCB}}|T^{-\left(\beta \underline{d}^2 \underline{\theta}^2\right)^2/2\cdot 36^2 K^2} + \frac{2M}{1-e^{-\left(\underline{d\theta}\right)^2/72K^3}} \\ \leq\ & 6M\left(\hat{\tau}_{j+1}-\hat{\tau}_j\right)T^{-\left(\beta \underline{d}^2 \underline{\theta}^2\right)^2/2\cdot 36^2 K^2} + \frac{2M}{1-e^{-\left(\underline{d\theta}\right)^2/72K^3}}. \end{aligned}$$

The first inequality follows from Equation (20). The second inequality follows from Equations (21), (22) and (23). The fourth inequality follows from the property that $|\mathcal{T}_{j,<}^{\text{UCB}}| \leq \hat{\tau}_{j+1}-\hat{\tau}_j$.

Next, we prove Part 2.

We define $\mathcal{S}_j\left(k\right) \triangleq \left\{t\in\{\hat{\tau}_j,\cdots,\tau_{j+1}-1\} : \pi_t^{\hat{\tau}_j} = k\right\}$. Therefore,

$$\begin{aligned} \sum_{t\in\mathcal{T}_{j,<}^{\text{UCB}}} \mathsf{E}\left[U_t\left(\pi_t^{\hat{\tau}_j}\right) - L_t\left(\pi_t^{\hat{\tau}_j}\right)\right] &\leq \sum_{t\in\mathcal{T}_{j,<}^{\text{UCB}}} \mathsf{E}\left[2\beta M p_k\sqrt{\frac{\log T}{N_{\hat{\tau}_j,t-1}\left(\pi_t^{\hat{\tau}_j}\right)}} + 4\beta M p_k\sqrt{\frac{\log T}{N_{\hat{\tau}_j,t-1}\left(k_t\right)}}\right] \\ &\leq \sum_{t\in\mathcal{T}_{j,<}^{\text{UCB}}} \mathsf{E}\left[6\beta M p_k\sqrt{\frac{\log T}{N_{\hat{\tau}_j,t-1}\left(\pi_t^{\hat{\tau}_j}\right)}}\right] \end{aligned}$$

$$
\begin{aligned}
&\leq 6\beta M\bar{p}\sqrt{\log T}\sum_{t\in\mathcal{T}_{j,<}^{\mathrm{UCB}}}\mathsf{E}\left[\frac{1}{\sqrt{N_{\hat{\tau}_j,t-1}\left(\pi_t^{\hat{\tau}_j}\right)}}\right]\\
&\leq 6\beta M\bar{p}\sqrt{\log T}\sum_{t=\hat{\tau}_j}^{\tau_{j+1}-1}\mathsf{E}\left[\frac{1}{\sqrt{N_{\hat{\tau}_j,t-1}\left(\pi_t^{\hat{\tau}_j}\right)}}\right]\\
&= 6\beta M\bar{p}\sqrt{\log T}\mathsf{E}\left[\sum_{k=1}^{K}\sum_{y=1}^{|\mathcal{S}_j(k)|}\frac{1}{\sqrt{y}}\right]\\
&\leq 6\beta M\bar{p}\sqrt{\log T}\mathsf{E}\left[\sum_{k=1}^{K}\int_{y=0}^{|\mathcal{S}_j(k)|}\frac{1}{\sqrt{y}}dy\right]\\
&= 12\beta M\bar{p}\sqrt{\log T}\mathsf{E}\left[\sum_{k=1}^{K}\sqrt{|\mathcal{S}_j\left(k\right)|}\right]\\
&\leq 12\beta M\bar{p}\sqrt{\log T}\mathsf{E}\left[K\sqrt{\frac{\sum_{k=1}^{K}|\mathcal{S}_j\left(k\right)|}{K}}\right]\\
&= 12\beta M\bar{p}\sqrt{\log T}K\sqrt{\frac{\tau_{j+1}-\hat{\tau}_j}{K}}\\
&\leq 12\beta M\bar{p}\sqrt{K\left(\hat{\tau}_{j+1}-\hat{\tau}_j\right)\log T}.
\end{aligned}
$$

The first inequality follows from the definitions of $U_t\left(\cdot\right)$ and $L_t\left(\cdot\right)$. The second inequality holds since the definition of k_t implies $N_{\hat{\tau}_j,t-1}\left(\pi_t^{\hat{\tau}_j}\right)\leq N_{\hat{\tau}_j,t-1}\left(k_t\right)$. The third inequality follows from the property that $p_k\leq\bar{p}$. The first equality follows from the definition of $|\mathcal{S}_j\left(k\right)|$. The fourth inequality follows from the property that $\sqrt{\cdot}$ is a concave function and Jensen's inequality. The third equality follows from the property that $\sum_{k=1}^{K}|\mathcal{S}_j\left(k\right)|=\tau_{j+1}-\hat{\tau}_j$. The fifth inequality follows from the property that $\hat{\tau}_{j+1}\geq\tau_{j+1}$.

For Part 3, by following the similar proof of Part 1, we have

$$
\sum_{t\in\mathcal{T}_{j,<}^{\mathrm{UCB}}}\mathbb{P}\left(L_t\left(\pi_t^{\hat{\tau}_j}\right)\geq g_t\left(\pi_t^{\hat{\tau}_j}\right)\right)\leq 6M\left(\hat{\tau}_{j+1}-\hat{\tau}_j\right)T^{-\left(\beta\underline{d}^2\underline{\theta}^2\right)^2/2\cdot 36^2K^2}+\frac{2M}{1-e^{-(\underline{d\theta})^2/72K^3}}.
$$

Q.E.D.

Proof of Lemma 9.

We have

$$
H_{j,3}^3=\mathsf{E}\left[\sum_{t\in\mathcal{T}_j^{\mathrm{UCB}}}\bar{g}_t\left(k_t^*\right)-\bar{g}_t\left(\pi_t^{\hat{\tau}_j}\right)\right]+\mathsf{E}\left[\sum_{t\in\mathcal{T}_j^{\mathrm{UCB}}}g_t\left(k_t^*\right)-\bar{g}_t\left(k_t^*\right)\right]
$$

$$
\begin{aligned}
&+\mathsf{E}\left[\sum_{t\in\mathcal{T}_j^{\mathrm{UCB}}} \bar{g}_t\left(\pi_t^{\hat{\tau}_j}\right)-g_t\left(\pi_t^{\hat{\tau}_j}\right)\right]\\
\le\ & \mathsf{E}\left[\sum_{t\in\mathcal{T}_j^{\mathrm{UCB}}} \bar{g}_t\left(k_t^*\right)-\bar{g}_t\left(\pi_t^{\hat{\tau}_j}\right)\right]+2\max_k \mathsf{E}\left[\sum_{t\in\mathcal{T}_j^{\mathrm{UCB}}}\left|\bar{g}_t\left(k\right)-g_t\left(k\right)\right|\right]\\
=\ & \mathsf{E}\left[\sum_{t\in\mathcal{T}_j^{\mathrm{UCB}}} \bar{g}_t\left(k_t^*\right)-\bar{g}_t\left(\pi_t^{\hat{\tau}_j}\right)\right]+2\max_k \mathsf{E}\left[\sum_{t\in\mathcal{T}_j^{\mathrm{UCB}}\setminus\mathcal{T}_{j,<}^{\mathrm{UCB}}}\left|\bar{g}_t\left(k\right)-g_t\left(k\right)\right|\right]\\
\le\ & \mathsf{E}\left[\sum_{t\in\mathcal{T}_j^{\mathrm{UCB}}} \bar{g}_t\left(k_t^*\right)-\bar{g}_t\left(\pi_t^{\hat{\tau}_j}\right)\right]+2\max_k \mathsf{E}\left[\sum_{t=\tau_{j+1}}^{\hat{\tau}_{j+1}-1}\left|\bar{g}_t\left(k\right)-g_t\left(k\right)\right|\right]\\
=\ & \mathsf{E}\left[\sum_{t\in\mathcal{T}_j^{\mathrm{UCB}}} \bar{g}_t\left(k_t^*\right)-U_t\left(\pi_t^{\hat{\tau}_j}\right)+U_t\left(\pi_t^{\hat{\tau}_j}\right)-L_t\left(\pi_t^{\hat{\tau}_j}\right)+L_t\left(\pi_t^{\hat{\tau}_j}\right)-\bar{g}_t\left(\pi_t^{\hat{\tau}_j}\right)\right]\\
&+2\max_k \mathsf{E}\left[\sum_{t=\tau_{j+1}}^{\hat{\tau}_{j+1}-1}\left|\bar{g}_t\left(k\right)-g_t\left(k\right)\right|\right]\\
\le\ & \mathsf{E}\left[\sum_{t\in\mathcal{T}_j^{\mathrm{UCB}}} \bar{g}_t\left(k_t^*\right)-U_t\left(k_t^*\right)+U_t\left(\pi_t^{\hat{\tau}_j}\right)-L_t\left(\pi_t^{\hat{\tau}_j}\right)+L_t\left(\pi_t^{\hat{\tau}_j}\right)-\bar{g}_t\left(\pi_t^{\hat{\tau}_j}\right)\right]\\
&+2\max_k \mathsf{E}\left[\sum_{t=\tau_{j+1}}^{\hat{\tau}_{j+1}-1}\left|\bar{g}_t\left(k\right)-g_t\left(k\right)\right|\right]\\
=\ & \sum_{t\in\mathcal{T}_j^{\mathrm{UCB}}} \mathsf{E}\left[\bar{g}_t\left(k_t^*\right)-U_t\left(k_t^*\right)\right]+\sum_{t\in\mathcal{T}_j^{\mathrm{UCB}}} \mathsf{E}\left[U_t\left(\pi_t^{\hat{\tau}_j}\right)-L_t\left(\pi_t^{\hat{\tau}_j}\right)\right]\\
&+\sum_{t\in\mathcal{T}_j^{\mathrm{UCB}}} \mathsf{E}\left[L_t\left(\pi_t^{\hat{\tau}_j}\right)-\bar{g}_t\left(\pi_t^{\hat{\tau}_j}\right)\right]+2\max_k \mathsf{E}\left[\sum_{t=\tau_{j+1}}^{\hat{\tau}_{j+1}-1}\left|\bar{g}_t\left(k\right)-g_t\left(k\right)\right|\right].
\end{aligned}
$$

The second equality follows from the property that $\bar{g}_t(\cdot)=g_t(\cdot)$ for $t\le\hat{\tau}_{j+1}-1$. The third inequality follows from the definition of $U_t(\cdot)$.

Now, we bound $\left|\bar{D}_t^m(k)-\left(a_j^m+b_j^m\frac{t}{T}\right)\theta_j^m(k_t)\right|$. We denote

$$X=\left(a_j^m+b_j^m\frac{s_t(k)}{T}\right)\theta_j^m(k),\qquad \Delta X=\phi_j^m(k)\frac{\Delta s_t(k)}{T},$$

and

$$Y=\left(a_j^m+b_j^m\frac{t}{T}\right)\theta_j^m(k_t),\qquad \Delta Y=\phi_j^m(k_t)\left(\frac{t-s_t^{(1)}}{s_t^{(2)}-s_t^{(1)}}\frac{\Delta s_t^{(2)}}{T}-\frac{t-s_t^{(2)}}{s_t^{(2)}-s_t^{(1)}}\frac{\Delta s_t^{(1)}}{T}\right),$$

and

$$Z=\left(a_j^m+b_j^m\frac{s_t\left(k\right)}{T}\right)\theta_j^m\left(k_t\right),\qquad \Delta Z=\phi_j^m\left(k_t\right)\left(\frac{s_t\left(k\right)-s_t^{(1)}}{s_t^{(2)}-s_t^{(1)}}\frac{\Delta s_t^{(2)}}{T}-\frac{s_t\left(k\right)-s_t^{(2)}}{s_t^{(2)}-s_t^{(1)}}\frac{\Delta s_t^{(1)}}{T}\right).$$

First, the properties that $a_j^m+b_j^m\frac{s}{T}\in[\underline{d},1]$ for any $s\in\{s_t\left(k\right),t\}$ and $\theta_j^m\left(k\right)\in[\underline{\theta},1]$ for any k jointly imply $X,Y,Z\in[\underline{d\theta},1]$. Second, the properties that $|b_j^m|\le\bar{b}$ and $\theta_j^m\left(k\right)\in[\underline{\theta},1]$ for any j, m, and k imply $|\phi_j^m\left(k\right)|\le|b_{j+1}^m||\theta_{j+1}^m\left(k\right)|+|b_j^m||\theta_j^m\left(k\right)|\le2\bar{b}$. In addition, we have $\frac{\Delta s_t^{(2)}}{T}\le\frac{\hat{\tau}_{j+1}-\tau_{j+1}}{T}\le\frac{w_1L}{3T}\le\frac{\underline{d\theta}}{8(2K+1)\bar{b}}$, where the first inequality follows from the definition on $\Delta s_t^{(2)}$, the second inequality follows from the condition that $j+1\in\mathcal{N}_1$ and the definition of $\mathcal{B}_{j+1}$, the third inequality follows from Lemma 2 Part 2. These results and Lemma 15 jointly imply $|\Delta X|,|\Delta Y|,|\Delta Z|\le 4\bar{b}\left(2K+1\right)\frac{w_1L}{3T}\le\frac{1}{2\underline{d\theta}}$. Third, the properties above jointly imply $X+\Delta X\in\left[\frac{1}{2\underline{d\theta}},\frac{3}{2\underline{d\theta}}\right]$ and $Z+\Delta Z\ge\frac{1}{2\underline{d\theta}}$.

Therefore, we have

$$\begin{aligned}\left|\bar{D}_t^m\left(k\right)-\left(a_j^m+b_j^m\frac{t}{T}\right)\theta_j^m\left(k\right)\right| &= \left|\frac{\left(X+\Delta X\right)\left(Y+\Delta Y\right)}{Z+\Delta Z}-\frac{XY}{Z}\right|\\ &\le 4\left(\underline{d\theta}\right)^2\left(|\Delta X|+\frac{3}{2\underline{d\theta}}|\Delta Y|+\frac{3}{2\underline{d\theta}}|\Delta Z|\right)\mathbf{1}\left\{t\ge\tau_{j+1}\right\}\\ &= \left(\underline{d\theta}\right)\left(4|\Delta X|+6\left(\underline{d\theta}\right)|\Delta Y|+6\left(\underline{d\theta}\right)|\Delta Z|\right)\mathbf{1}\left\{t\ge\tau_{j+1}\right\}\\ &\le \left(4|\Delta X|+6|\Delta Y|+6|\Delta Z|\right)\mathbf{1}\left\{t\ge\tau_{j+1}\right\}\\ &\le \bar{b}\left(2K+1\right)\frac{64w_1L}{3T}\mathbf{1}\left\{t\ge\tau_{j+1}\right\}.\end{aligned}\tag{24}$$

The first inequality follows from the three properties above and Lemma 14. The second inequality follows from the properties that $\underline{d}\le1$ and $\underline{\theta}\le1$. The third inequality follows from the second property above.

Next, we use Equation 24 to bound $\bar{g}_t\left(k\right)$. We have

$$\begin{aligned}\left|\bar{g}_t\left(k\right)-p_k\sum_{m=1}^M\left(a_j^m+b_j^m\frac{t}{T}\right)\theta_j^m\left(k\right)\right| &= \left|p_k\sum_{m=1}^M\bar{D}_t^m\left(k\right)-p_k\sum_{m=1}^M\left(a_j^m+b_j^m\frac{t}{T}\right)\theta_j^m\left(k\right)\right|\\ &\le p_k\sum_{m=1}^M\left|\bar{D}_t^m\left(k\right)-\left(a_j^m+b_j^m\frac{t}{T}\right)\theta_j^m\left(k\right)\right|\\ &\le \bar{p}\sum_{m=1}^M\left|\bar{D}_t^m\left(k\right)-\left(a_j^m+b_j^m\frac{t}{T}\right)\theta_j^m\left(k\right)\right|\\ &\le M\bar{p}\bar{b}\left(2K+1\right)\frac{64w_1L}{3T}\mathbf{1}\left\{t\ge\tau_{j+1}\right\}.\end{aligned}\tag{25}$$

The first equality follows from the property that $j+1\in\mathcal{N}_1$. The first inequality follows from the triangle inequality. The second inequality follows from the property that $p_k\le\bar{p}$. The third inequality follows from Equation (24).

Therefore,

$$\begin{aligned}
\left|\bar{g}_t(k) - g_t(k)\right| &= \left|\bar{g}_t(k) - p_k \sum_{m=1}^{M}\left(\left(a_j^m + b_j^m \frac{t}{T}\right)\theta_j^m(k) + \phi_j^m(k)\frac{(t-\tau_{j+1})^+}{T}\right)\right| \\
&\le \left|\bar{g}_t(k) - p_k \sum_{m=1}^{M}\left(a_j^m + b_j^m \frac{t}{T}\right)\theta_j^m(k)\right| + \left|p_k \sum_{m=1}^{M}\phi_j^m(k)\frac{(t-\tau_{j+1})^+}{T}\right| \\
&\le M\bar{p}\bar{b}(2K+1)\frac{64w_1L}{3T}\mathbf{1}\{t \ge \tau_{j+1}\} + 2M\bar{p}\bar{b}\frac{(t-\tau_{j+1})^+}{T} \\
&\le 2M\bar{p}\left(\bar{b}(2K+1)\frac{32w_1L}{3T} + \bar{b}\frac{\hat{\tau}_{j+1}-\tau_{j+1}}{T}\right)\mathbf{1}\{t \ge \tau_{j+1}\} \\
&\le 2M\bar{p}\left(\bar{b}(2K+1)\frac{64w_1L}{3T} + \bar{b}\frac{w_1L}{3T}\right)\mathbf{1}\{t \ge \tau_{j+1}\} \\
&\le \frac{256}{3}M\bar{p}\bar{b}(K+1)\frac{w_1L}{T}\mathbf{1}\{t \ge \tau_{j+1}\}. \qquad (26) \\
&\le 32M\bar{p}(K+1)\frac{d\theta}{2K+1}\mathbf{1}\{t \ge \tau_{j+1}\} \\
&\le 32M\bar{p}\underline{d}\underline{\theta}\mathbf{1}\{t \ge \tau_{j+1}\} \\
&\le 32M\bar{p}. \qquad (27)
\end{aligned}$$

The first inequality follows from the triangle inequality. The second inequality follows from Equation (25) and the properties that $p_k \le \bar{p}$ and $|\phi_j^m(k)| \le 2\bar{b}$. The third inequality follows from the property that $t < \hat{\tau}_{j+1}$. The fourth inequality follows from the property that $j+1 \in \mathcal{N}_1$ and the definition of $\mathcal{B}_{j+1}$. The sixth inequality follows from Equation (26) and Lemma 2 Part 2 that $\frac{w_1 L}{3T} \le \frac{d\theta}{8(2K+1)\bar{b}}$ jointly imply The eighth inequality follows from the properties that $\underline{d} \le 1$ and $\underline{\theta} \le 1$.

Therefore, for any k,

$$\begin{aligned}
\mathsf{E}\left[\sum_{t=\tau_{j+1}}^{\hat{\tau}_{j+1}-1}\left|\bar{g}_t(k) - g_t(k)\right|\right] &\le \sum_{t=\tau_{j+1}}^{\hat{\tau}_{j+1}-1}\frac{256}{3}M\bar{p}\bar{b}(K+1)\frac{w_1L}{T} \\
&= \frac{256}{3}M\bar{p}\bar{b}(K+1)\frac{w_1L}{T}(\hat{\tau}_{j+1}-\tau_{j+1}) \\
&\le \frac{256}{9}M\bar{p}\bar{b}(K+1)\frac{w_1^2L^2}{T} \\
&\le 30M\bar{p}\bar{b}(K+1)\frac{w_1^2L^2}{T}.
\end{aligned}$$

The first inequality follows from Equation (26). The second inequality follows from the property that $j+1 \in \mathcal{N}_1$ and the definition of $\mathcal{B}_{j+1}$.

Therefore, for any $t \in \mathcal{T}_j^{\mathrm{UCB}}$ and $k \in \{1, \cdots, K\}$, we have

$$\begin{aligned}
\mathsf{E}[\bar{g}_t(k) - U_t(k)] &= \mathsf{E}[(\bar{g}_t(k) - U_t(k))\mathbf{1}\{\bar{g}_t(k) \ge U_t(k)\}] \\
&\quad + \mathsf{E}[(\bar{g}_t(k) - U_t(k))\mathbf{1}\{\bar{g}_t(k) < U_t(k)\}] \\
&\le \mathsf{E}[(\bar{g}_t(k) - U_t(k))\mathbf{1}\{\bar{g}_t(k) \ge U_t(k)\}]
\end{aligned}$$

$$
\begin{aligned}
&\leq \bar{g}_t(k)\,\mathbb{P}\left(\bar{g}_t(k) \geq U_t(k)\right) \\
&\leq \left(g_t(k) + \left|\bar{g}_t(k) - g_t(k)\right|\right)\mathbb{P}\left(\bar{g}_t(k) \geq U_t(k)\right) \\
&\leq \left(M\bar{p} + 32M\bar{p}\right)\mathbb{P}\left(\bar{g}_t(k) \geq U_t(k)\right) \\
&= 33M\bar{p}\mathbb{P}\left(\bar{g}_t(k) \geq U_t(k)\right).
\end{aligned}
$$

The second inequality follows from the property that $U_t(\cdot) \geq 0$. The third inequality follows from the triangle inequality. The fourth inequality follows from the properties that $a_j^m + b_j^m \frac{t}{T} \in [0,1]$, $p_k \leq \bar{p}$ and $\theta_j^m(\cdot) \in [0,1]$ and Equation (27).

For any $t \in \mathcal{T}_j^{\mathrm{UCB}}$ and $k \in \{1, \cdots, K\}$, we have

$$
\begin{aligned}
\mathsf{E}\left[L_t(k) - \bar{g}_t(k)\right] &= \mathsf{E}\left[\left(L_t(k) - \bar{g}_t(k)\right)\mathbf{1}\left\{L_t(k) \geq \bar{g}_t(k)\right\}\right] \\
&\quad + \mathsf{E}\left[\left(L_t(k) - \bar{g}_t(k)\right)\mathbf{1}\left\{L_t(k) < \bar{g}_t(k)\right\}\right] \\
&\leq \mathsf{E}\left[\left(L_t(k) - \bar{g}_t(k)\right)\mathbf{1}\left\{L_t(k) \geq \bar{g}_t(k)\right\}\right] \\
&= \left(L_t(k) - \bar{g}_t(k)\right)\mathbb{P}\left(L_t(k) \geq \bar{g}_t(k)\right) \\
&\leq \left(L_t(k) - g_t(k) + \left|\bar{g}_t(k) - g_t(k)\right|\right)\mathbb{P}\left(L_t(k) \geq \bar{g}_t(k)\right) \\
&\leq \left(L_t(k) + \left|\bar{g}_t(k) - g_t(k)\right|\right)\mathbb{P}\left(L_t(k) \geq \bar{g}_t(k)\right) \\
&\leq \left(M\bar{p} + 32M\bar{p}\right)\mathbb{P}\left(L_t(k) \geq \bar{g}_t(k)\right) \\
&= 33M\bar{p}\mathbb{P}\left(L_t(k) \geq \bar{g}_t(k)\right).
\end{aligned}
$$

The second inequality follows from the triangle inequality. The third inequality follows from the property that $a_j^m + b_j^m \frac{t}{T} \in [0,1]$ and $\theta_j^m(\cdot) \in [0,1]$ jointly imply $g_t(k) \geq 0$. The fourth inequality follows from the property that $L_t(\cdot) \leq M\bar{p}$ and Equation (27).

All results derived in this proof jointly imply

$$
\begin{aligned}
H_{j,3}^3 \leq\ & 33M\bar{p} \sum_{t \in \mathcal{T}_j^{\mathrm{UCB}}} \mathbb{P}\left(\bar{g}_t(k) \geq U_t(k)\right) + \sum_{t \in \mathcal{T}_j^{\mathrm{UCB}}} \mathsf{E}\left[U_t\left(\pi_t^{\hat{\tau}_j}\right) - L_t\left(\pi_t^{\hat{\tau}_j}\right)\right] \\
& + 33M\bar{p} \sum_{t \in \mathcal{T}_j^{\mathrm{UCB}}} \mathbb{P}\left(L_t(k) \geq \bar{g}_t(k)\right) + 60M\bar{p}\bar{b}(K+1)\frac{w_1^2 L^2}{T}.
\end{aligned}
$$

Q.E.D.

Proof of Lemma 10.

The proof of this lemma is analogous to the proof of Lemma 8. Therefore, we omit the proof.

Q.E.D.

Proof of Lemma 11.

We have

$$\begin{aligned}
H^3 &= \mathsf{E}_{\{\hat{\tau}_1,\cdots,\hat{\tau}_J\}}\left[\sum_{j=0}^{J} H^3_{j,1}+H^3_{j,2}+H^3_{j,3}\middle|\cap_{l=1}^{J}\left(\mathcal{A}_l\cap\mathcal{B}_l\right)\cap\mathcal{A}_{J+1}\right]\\
&\leq \mathsf{E}_{\{\hat{\tau}_1,\cdots,\hat{\tau}_J\}}\Bigg[\sum_{j=0}^{J} 2M\bar{p}K+M\bar{p}K\left(\frac{\hat{\tau}_{j+1}-\hat{\tau}_j}{L}+1\right)\\
&\qquad +33M\bar{p}\left(6M\left(\hat{\tau}_{j+1}-\hat{\tau}_j\right)T^{-\left(\beta\underline{d}^2\underline{\theta}^2\right)^2/2\cdot 36^2K^2}+\frac{2M}{1-e^{-(\underline{d\theta})^2/72K^3}}\right)\\
&\qquad +12\beta M\bar{p}\sqrt{K\left(\hat{\tau}_{j+1}-\hat{\tau}_j\right)\log T}\\
&\qquad +33M\bar{p}\left(6M\left(\hat{\tau}_{j+1}-\hat{\tau}_j\right)T^{-\left(\beta\underline{d}^2\underline{\theta}^2\right)^2/2\cdot 36^2K^2}+\frac{2M}{1-e^{-(\underline{d\theta})^2/72K^3}}\right)\\
&\qquad +M\bar{p}\frac{w_0L}{2}\mathbf{1}\left\{j+1\in\mathcal{N}_0\right\}+60M\bar{p}\bar{b}\left(K+1\right)\frac{w_1^2L^2}{T}\mathbf{1}\left\{j+1\in\mathcal{N}_1\right\}\Bigg|\cap_{l=1}^{J}\left(\mathcal{A}_l\cap\mathcal{B}_l\right)\cap\mathcal{A}_{J+1}\Bigg]\\
&= 3M\bar{p}K\left(J+1\right)+M\bar{p}K\frac{T}{L}+132M^2\bar{p}\left(3T^{1-\left(\beta\underline{d}^2\underline{\theta}^2\right)^2/2\cdot 36^2K^2}+\frac{J+1}{1-e^{-(\underline{d\theta})^2/72K^3}}\right)\\
&\qquad +12\beta M\bar{p}\mathsf{E}_{\{\hat{\tau}_1,\cdots,\hat{\tau}_J\}}\left[\sum_{j=0}^{J}\sqrt{K\left(\hat{\tau}_{j+1}-\hat{\tau}_j\right)\log T}\middle|\cap_{l=1}^{J}\left(\mathcal{A}_l\cap\mathcal{B}_l\right)\cap\mathcal{A}_{J+1}\right]\\
&\qquad +M\bar{p}\frac{w_0L}{2}\left|\mathcal{N}_0\right|+60M\bar{p}\bar{b}\left(K+1\right)\frac{w_1^2L^2}{T}\left|\mathcal{N}_1\right|\\
&\leq 3M\bar{p}K\left(J+1\right)+M\bar{p}K\frac{T}{L}+132M^2\bar{p}\left(3T^{1-\left(\beta\underline{d}^2\underline{\theta}^2\right)^2/2\cdot 36^2K^2}+\frac{J+1}{1-e^{-(\underline{d\theta})^2/72K^3}}\right)\\
&\qquad +12\beta M\bar{p}\sqrt{K\left(J+1\right)T\log T}+M\bar{p}\frac{w_0L}{2}\left|\mathcal{N}_0\right|+60M\bar{p}\bar{b}\left(K+1\right)\frac{w_1^2L^2}{T}\left|\mathcal{N}_1\right|.
\end{aligned}$$

The first inequality follows from Lemmas 6, 7, 8, 9 and 10. The second inequality follows from the property that $\sqrt{\cdot}$ is a concave function and Jensen's inequality.

Q.E.D.

Proof of Theorem 2.

First, for the parameters specified in this theorem, we have $w_0L=O\left(T^{1/2}\left(\log T\right)^{1/2}\right)=o\left(T\right)$ and $w_1L=O\left(T^{5/6}\left(\log T\right)^{1/6}\right)=o\left(T\right)$. Because $\tau_{j+1}-\tau_j=O\left(T\right)$ for all $j\in\{0,\cdots,J\}$, Lemma 2 Part 1 is satisfied for large T.

Second, we have $\frac{w_1L}{T}=O\left(T^{-1/6}\left(\log T\right)^{1/6}\right)=o\left(1\right)$. Hence, Lemma 2 Part 2 is satisfied for large T.

For large T, we have

$$U^0=2\exp\left(-\frac{1}{2}\left(\left(\epsilon_0-2\bar{b}\frac{w_0L}{T}\right)^+\right)^2w_0\right)$$

$$\begin{aligned}
&\leq 2\exp\left(-2\left(\left(1-2\bar{b}\frac{w_0 L}{\epsilon_0 T}\right)^{+}\right)^{2}\log T\right)\\
&\leq 2\exp\left(-\frac{3}{2}\log T\right)\\
&= 2T^{-3/2}. \qquad (28)
\end{aligned}$$

The first equality follows from the definition of U^0. The first inequality follows from the property that $w_0 = 2\lceil 2\log T/\epsilon_0^2\rceil \geq 4\log T/\epsilon_0^2$. The second inequality holds since the properties that $\epsilon_0 = O(1)$ and $\frac{w_0 L}{T} = O\left(T^{-1/2}(\log T)^{1/2}\right)$ imply that for larger T, $\frac{w_0 L}{T} \leq \frac{\epsilon_0}{2\bar{b}}\left(1-\frac{\sqrt{3}}{2}\right)$.

We have

$$\begin{aligned}
U^1 &= 2\exp\left(-\frac{1}{12}\left(\frac{1}{3}-\frac{1}{w_1}\right)^{3}\epsilon_1^2\frac{w_1^3 L^2}{T^2}\right)\\
&\leq 2\exp\left(-\frac{9}{4}\left(1-\frac{3}{w_1}\right)^{3}\log T\right)\\
&\leq 2\exp\left(-\frac{3}{2}\log T\right)\\
&= 2T^{-3/2}. \qquad (29)
\end{aligned}$$

The first equality follows from the definition of U^1. The first inequality follows from the properties that $w_1 = 3\lceil 3T^{1/3}(\log T)^{2/3}/\epsilon_1^{2/3}\rceil \geq 9T^{1/3}(\log T)^{2/3}/\epsilon_1^{2/3}$ and $L = \max\left\{\lceil T^{1/2}/(\log T)^{1/2}\rceil, K+1\right\} \geq T^{1/2}/(\log T)^{1/2}$. The second inequality holds since the properties that $\epsilon_1 = O(1)$ and $w_1 = O\left(T^{1/3}(\log T)^{2/3}\right)$ imply that for larger T, $w_1 \geq \frac{3}{1-(2/3)^{1/3}}$.

Therefore, for large T,

$$\begin{aligned}
H^1 + H^2 &\leq O\left(\left(\frac{T^2}{L}+T\right)\left(U^0+U^1\right)\right)\\
&= O\left(T^{3/2}(\log T)^{1/2}\left(U^0+U^1\right)\right)\\
&\leq O\left(T^{3/2}(\log T)^{1/2}T^{-3/2}\right)\\
&= O\left((\log T)^{1/2}\right).
\end{aligned}$$

The first inequality follows from Lemma 4. The first equality follows from the property that $L = O\left(T^{1/2}/(\log T)^{1/2}\right)$. The second inequality follows from Equations (28) and (29).

Next, we analyze H^3. For large T,

$$\begin{aligned}
H^3 \leq\ & 3M\bar{p}K(J+1) + M\bar{p}K\frac{T}{L} + 132M^2\bar{p}\left(3T^{1-\left(\beta\underline{d}^2\underline{\theta}^2\right)^2/2\cdot 36^2K^2} + \frac{J+1}{1-e^{-(\underline{d\theta})^2/72K^3}}\right)\\
&+12\beta M\bar{p}\sqrt{K(J+1)T\log T} + M\bar{p}\frac{w_0 L}{2}|\mathcal{N}_0| + 60M\bar{p}\bar{b}(K+1)\frac{w_1^2L^2}{T}|\mathcal{N}_1|\\
=\ & O(KJ) + O\left(KT^{1/2}(\log T)^{1/2}\right) + O\left(T^{1/2}\right) + O\left(K^{3/2}J^{1/2}T^{1/2}(\log T)^{1/2}\right)
\end{aligned}$$

$$
\begin{aligned}
&+O\left(T^{1/2}\left(\log T\right)^{1/2}|\mathcal{N}_0|\right)+O\left(\frac{K}{\Delta_1^{4/3}}T^{2/3}\left(\log T\right)^{1/3}|\mathcal{N}_1|\right)\\
=\;&O\left(\left(K^{3/2}J^{1/2}+\frac{|\mathcal{N}_0|}{\Delta_0^2}\right)MT^{1/2}\left(\log T\right)^{1/2}\right)+O\left(\frac{K}{\Delta_1^{4/3}}T^{2/3}\left(\log T\right)^{1/3}|\mathcal{N}_1|\right).
\end{aligned}
$$

The first inequality follows from Lemma 11. The first equality follows from the properties that $\frac{T}{L}=O\left(T^{1/2}\left(\log T\right)^{1/2}\right)$, $\beta=36K/\underline{d\theta}$, $w_0L=O\left(\frac{1}{\Delta_0^2}T^{1/2}\left(\log T\right)^{1/2}\right)$ and $\frac{w_1^2L^2}{T}=O\left(\frac{1}{\Delta_1^{4/3}}T^{2/3}\left(\log T\right)^{1/3}\right)$.

Therefore,

$$
\begin{aligned}
\text{Regret}^{\text{CU}}(T)&\le H^1+H^2+H^3\\
&\le O\left(\left(\log T\right)^{1/2}\right)+O\left(\left(K^{3/2}J^{1/2}+\frac{|\mathcal{N}_0|}{\Delta_0^2}\right)T^{1/2}\left(\log T\right)^{1/2}\right)+O\left(\frac{K}{\Delta_1^{4/3}}T^{2/3}\left(\log T\right)^{1/3}|\mathcal{N}_1|\right)\\
&=O\left(\left(K^{3/2}J^{1/2}+\frac{|\mathcal{N}_0|}{\Delta_0^2}\right)T^{1/2}\left(\log T\right)^{1/2}\right)+O\left(\frac{K}{\Delta_1^{4/3}}T^{2/3}\left(\log T\right)^{1/3}|\mathcal{N}_1|\right)\\
&=\begin{cases} O\left(\frac{K}{\Delta_1^{4/3}}T^{2/3}\left(\log T\right)^{1/3}|\mathcal{N}_1|\right) & \text{if } \mathcal{N}_1\neq\emptyset\\ O\left(\left(K^{3/2}J^{1/2}+\frac{J}{\Delta_0^2}\right)T^{1/2}\left(\log T\right)^{1/2}\right) & \text{if } \mathcal{N}_1=\emptyset\end{cases}\\
&=\begin{cases} O\left(\frac{K}{\Delta_1^{4/3}}T^{2/3}\left(\log T\right)^{1/3}|\mathcal{N}_1|\right) & \text{if } \mathcal{N}_1\neq\emptyset\\ O\left(\left(K^{3/2}+\frac{J^{1/2}}{\Delta_0^2}\right)J^{1/2}T^{1/2}\left(\log T\right)^{1/2}\right) & \text{if } \mathcal{N}_1=\emptyset\end{cases}\\
&=\begin{cases} O\left(T^{2/3}\left(\log T\right)^{1/3}\right) & \text{if } \mathcal{N}_1\neq\emptyset\\ O\left(T^{1/2}\left(\log T\right)^{1/2}\right) & \text{if } \mathcal{N}_1=\emptyset\end{cases}.
\end{aligned}
$$

Q.E.D.

Appendix D: Auxiliary Results

LEMMA 13. *Suppose $X_l^{(i)}\ge 0$ for $i\in\{1,2\}$ and $l\in\{1,\cdots,L\}$. We have*

$$
\max\left\{\sum_{l=1}^{L}X_l^{(1)},\sum_{l=1}^{L}X_l^{(2)}\right\}\ge\frac{L}{2}\min_{l\in\{1,\cdots,L\}}\max\left\{X_l^{(1)},X_l^{(2)}\right\}.
$$

Proof of Lemma 13.

We denote $\mathcal{L}^{(1)}=\left\{l:X_l^{(1)}\ge X_l^{(2)}\right\}$ and $\mathcal{L}^{(2)}=\left\{l:X_l^{(1)}<X_l^{(2)}\right\}$. Therefore,

$$
\begin{aligned}
\max\left\{\sum_{l=1}^{L}X_l^{(1)},\sum_{l=1}^{L}X_l^{(2)}\right\}&\ge\sum_{l=1}^{L}X_l^{(1)}\\
&\ge\sum_{l\in\mathcal{L}^{(1)}}X_l^{(1)}\\
&\ge|\mathcal{L}^{(1)}|\min_{l\in\mathcal{L}^{(1)}}X_l^{(1)}
\end{aligned}
$$

$$
\begin{aligned}
&= |\mathcal{L}^{(1)}| \min_{l \in \mathcal{L}^{(1)}} \max\left\{X_l^{(1)}, X_l^{(2)}\right\} \\
&\geq |\mathcal{L}^{(1)}| \min_{l \in \{1,\cdots,L\}} \max\left\{X_l^{(1)}, X_l^{(2)}\right\}.
\end{aligned}
$$

The second inequality follows from the property that $X_l^{(1)} \geq 0$. The equality follows from the definition of $\mathcal{L}^{(1)}$.

By following the similar argument, we have

$$
\max\left\{\sum_{l=1}^{L} X_l^{(1)}, \sum_{l=1}^{L} X_l^{(2)}\right\} \geq |\mathcal{L}^{(2)}| \min_{l \in \{1,\cdots,L\}} \max\left\{X_l^{(1)}, X_l^{(2)}\right\}.
$$

Therefore,

$$
\begin{aligned}
\max\left\{\sum_{l=1}^{L} X_l^{(1)}, \sum_{l=1}^{L} X_l^{(2)}\right\} &\geq \max\left\{|\mathcal{L}^{(1)}|, |\mathcal{L}^{(2)}|\right\} \min_{l \in \{1,\cdots,L\}} \max\left\{X_l^{(1)}, X_l^{(2)}\right\} \\
&\geq \frac{L}{2} \min_{l \in \{1,\cdots,L\}} \max\left\{X_l^{(1)}, X_l^{(2)}\right\}.
\end{aligned}
$$

The second inequality follows from the property that $\max\left\{|\mathcal{L}^{(1)}|, |\mathcal{L}^{(2)}|\right\} \geq \left(|\mathcal{L}^{(1)}| + |\mathcal{L}^{(2)}|\right)/2 = T/2$.

Q.E.D.

LEMMA 14. *If $X, X+\Delta X \in [0,a]$, $Y \in [0,b]$, $Z, Z+\Delta Z \geq \underline{c} > 0$ and $Z \leq \bar{c}$, then*

$$
\left|\frac{(X+\Delta X)(Y+\Delta Y)}{Z+\Delta Z} - \frac{XY}{Z}\right| \leq \frac{1}{\underline{c}^2}\left(b\bar{c}|\Delta X| + a\bar{c}|\Delta Y| + ab|\Delta Z|\right).
$$

Proof of Lemma 14.

We have

$$
\begin{aligned}
\left|\frac{(X+\Delta X)(Y+\Delta Y)}{Z+\Delta Z} - \frac{XY}{Z}\right| &= \left|\frac{(X+\Delta X)(Y+\Delta Y)Z - XY(Z+\Delta Z)}{Z(Z+\Delta Z)}\right| \\
&\leq \frac{1}{\underline{c}^2}\left|(X+\Delta X)(Y+\Delta Y)Z - XY(Z+\Delta Z)\right| \\
&\leq \frac{1}{\underline{c}^2}\left(\left|(X+\Delta X)(Y+\Delta Y)Z - XYZ\right| + \left|XYZ - XY(Z+\Delta Z)\right|\right) \\
&\leq \frac{1}{\underline{c}^2}\left(\bar{c}\left|(X+\Delta X)(Y+\Delta Y) - XY\right| + ab|\Delta Z|\right) \\
&\leq \frac{1}{\underline{c}^2}\left(\bar{c}\left|Y\Delta X\right| + \bar{c}\left|(X+\Delta X)\Delta Y\right| + ab|\Delta Z|\right) \\
&\leq \frac{1}{\underline{c}^2}\left(b\bar{c}|\Delta X| + a\bar{c}|\Delta Y| + ab|\Delta Z|\right).
\end{aligned}
$$

The first inequality follows from the property that $Z, Z+\Delta Z \geq c > 0$. The second and the fourth inequalities follow from the triangle inequality. The third inequality follows from the properties that $X \in [0,a]$, $Y \in [0,b]$ and $Z \in [0,\bar{c}]$. The fifth inequality follows from the properties that $X+\Delta X \in [0,a]$, $Y \in [0,b]$.

Q.E.D.

LEMMA 15. *Under policy* $\pi^{\mathrm{CU},\tau}$, *for* $t \geq \tau + 2K$,

$$\frac{|s - s_t^{(i)}|}{s_t^{(2)} - s_t^{(1)}} \leq 2K + 1.$$

Proof of Lemma 15.

First, the definition of k_t implies $N_{\tau,t-1}(k_t) \geq \lceil \frac{t-\tau}{K} \rceil$. Second, the definitions of $\mathcal{T}_t^{(1)}$ and $\mathcal{T}_t^{(2)}$ jointly imply $s_t^{(2)} - s_t^{(1)} \geq \frac{N_{\tau,t-1}(k_t)}{2}$. Hence, these two results jointly imply $s_t^{(2)} - s_t^{(1)} \geq \frac{1}{2}\lceil \frac{t-\tau}{K} \rceil \geq \frac{t-\tau}{2K}$. Therefore, for any $i \in \{1,2\}$ and $s \in \{\tau, \cdots, t\}$,

$$\begin{aligned} \frac{|s - s_t^{(i)}|}{s_t^{(2)} - s_t^{(1)}} &\leq \frac{t - \tau + 1}{\frac{t-\tau}{2K}} \\ &\leq \frac{2K+1}{2K} 2K \\ &\leq 2K + 1. \end{aligned}$$

The second inequality follows from the property that $t - \tau \geq 2K$. **Q.E.D.**