Dynamic Pricing Under Customer Review Effects

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Ruizhi Shi

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Advisor: Prof. Zizhuo Wang

June, 2019

Acknowledgements

First of all, I would like to show my deepest gratitude to my advisor Prof. Zizhuo Wang, who is the role model both for my academic pursuit as well as my life. He is always very productive and creative. Whenever I am stuck at some problem, he can always provide helpful suggestions and insightful critics. He always reads my draft word by word and gives detailed instruction and feedback promptly. He cares for me not only my academic achievement but also my career advancement. In my second year, he encouraged me to work as an intern, which helped me lay a foundation in my industrial development. He showed me how to be a better person.

I also want to thank the rest of my dissertation committee: Prof. William L. Cooper, Prof. Ankur Mani, Prof. Tony Cui, for their helpful comments and suggestions. I want to thank them for allocating their time to my dissertation defense. Many of them have a very tight schedule, but still squeeze time to attend the defense.

I want to thank my colleagues at UMN: Zeyang Wu, Xiaochen Zhang, Xiang Gao, Xiaobo Li, Guiyun Feng, Yuanchen Su, Behrooz Pourghannad. It has always been a great pleasure to hang out with them. Were it not for their companion, I would not get through my PhD life joyfully and happily.

Lastly, I want to thank my parents, for their unconditional help and care throughout my life. I am thankful for their everlasting effort of raising me up, caring for me and guiding me. They do not only give me a helping hand whenever I am in need but are always on my back, support me, lead me and comfort me. I also want to thank my wife Xiangyi, who is always by my side through my hard times. Her companion is the source of happiness, serendipity, and joy in my life.

Dedication

To my parents, for their love and support.

Abstract

A dramatic development in e-commerce has been made in the past decade. For customers shopping online, other customers' review is an important source of information. There are already studies finding that product with good review can have higher demand. However, a review may be subjective and even a good quality product may still have low rating reviews. These negative reviews could be a problem when a seller just starts the business with only a few reviews, in which case a negative review will drag the average review significantly. It could be a seller's nightmare in a highly competitive market since customers will be unwilling to purchase a product with negative reviews whatever the quality is. To resolve this issue, the seller could simply lower the price to attract more purchases and get more reviews. However, keeping a low price may hurt the profit in the short run.

In this dissertation, we study the trade off between using a low price to attract more purchases and more reviews, and using a high price for a higher profit. We consider a monopolist selling a single product to a sequence of customers. Each customer will make a purchase decision based on the current review and the price. If a customer purchases the product, he will post a review, which is drawn from a normal distribution centered at the true quality. Only the seller knows the true quality, therefore, the customers have to use the review as a reference of the true quality. We derive the optimal policy on how the seller should adjust the price to maximize the expected revenue. We also derive upper bounds on the best performance of any policy and further extend the results to multiple price policies. After that, we then consider a discrete model where the review distribution can be a general distribution under some mild assumptions. The results of this dissertation highlight the trade off between short term and long term revenue,

provide insights on how to design a good pricing policy and enable sellers to make a better decision.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In the past several years, e-commerce has developed dramatically. In 2013, US e-commerce sales were 263 billion dollars, accounting for only 5.8% of the total retail sales in that year [1]. In 2018, US e-commerce sales grew to 513 billion dollars, accounting for 9.7% of total retail sales [2]. The growth of e-commerce in China is even more dramatic. The online retail sales soared from 270 billion dollars in 2013 [3] to 1 trillion dollars in 2017 [4].

For a consumer, the main advantage of buying products online is convenience. Instead of having to drive to a store and wait in a queue for checking out, consumers can click their mouse to make an order 24/7 at home. E-commerce also benefits the sellers significantly, for it gives them access to a much larger market at a low operational cost. For the entire society, it lowers the market entry barrier and enhances competition, thus increasing social welfare.

When shopping online, it may be sometimes difficult for customers to evaluate the quality of a product just based on the descriptions listed by the seller. One reason is the lack of credibility. Few sellers will admit the flaw of their products. It is often the case that a customer gets a product that does not match the description online or even a fake one. In this case, the customer has to either spend much time and effort to return the product or accept the loss. Another reason is that the description online is not informative enough. For example, sellers sometimes use quantitative parameters to describe their products. Although it is necessary to have some quantitative measurements, customers may have no sense of the numbers at all. For example, a camera producer may market their product as a portable camera with the detail parameters of its size, weight and battery life, while what customers really want to know is if it fits in a sports bag, if it is easy to carry for hiking or if it is necessary to prepare some backup batteries.

Facing such a situation, customers tend to refer to other buyers' review to estimate the true quality of the product. Indeed, reviews play an important role for online customers to make a purchase decision. According to a survey [5], a restaurant can get $5-9\%$ revenue increment if its Yelp rating gets one extra star. However, a review is usually a personal opinion which varies from one to another, even for the same product. For example, many high rating products on Amazon.com still have 1-star or 2-star reviews. These low reviews could be more problematic for a newly-launched product with only a few reviews, in which case a low rating review will drag the average review down significantly. In a highly competitive market, this could be a disaster for the seller since customers will be deterred to buy the product whatever the actual product quality is. In this case, should there be more purchases, the review will be likely to go back to a higher level. However, the dilemma is that without a good review, there would probably be no more purchases. One solution, in this case, is to make the price lower. With some discounts, the seller may be able to attract more purchases from the customers, which hopefully will reinstate the review to a good level. However, keeping a low price too long may hurt the profit of the seller. In this dissertation, we are interested in the trade off between using a low price to lure customers into purchasing the product and writing reviews and using a high price for a higher profit. In particular, we would like to answer the following questions:

(1) How should a seller set the price when considering the customer review effect? Should the seller give a discount at the beginning?

(2) If the seller can set multiple prices during the selling season, what is the value of the flexibility of using more prices?

(3) What is the value of adaptively adjusting prices based on the current review, compared with predetermining prices?

(4) How does the review distribution influence the result? Is there a significant difference if the customer rate the item with one to five stars instead of just a click of "like" or "dislike"?

To answer these questions, we consider a monopolist selling a single product to a sequence of customers. Only the seller knows the true quality of the product initially. However, each customer who buys the product will leave a review drawn from a normal distribution centered at the true quality. Subsequent customers use the current review score as an estimation of the true quality of the product. A customer's utility is the difference between the estimated quality and price. When a new customer arrives, if his utility is positive, he will purchase the product and leave a new review. Otherwise, the customer will leave without a purchase. In our model, the product's review is a simple average of all the previous review scores. There will also be an initial review score indicating the initial impression of this product. The seller's goal is to maximize his expected revenue.

Under this model, we consider different types of pricing policies.

- A fixed two-price policy. In the fixed two-price policy, the seller has to determine the prices he will use, and a switching time. All these decisions must be made at the beginning of the sales period. This policy is commonly used by restaurants and hotels. For example, when a new restaurant or hotel opens, it will offer some discount to customers. The discount expires after a certain period after which a normal price will be used. Also, all the information will be announced at the beginning.

- A fixed multi-price policy. In this policy, the seller can use multiple prices. All the prices and switching times have to be predetermined.

- A partially adaptive policy. In this policy, the seller only announces the switching times at the beginning of the sales. All the prices (except for the first price) can be determined adaptively at the switching time. This gives the seller some flexibility in adjusting the price.

- A fully adaptive two-price policy. In the fully adaptive two-price policy, the seller has maximum flexibility. He only needs to decide the first price at the beginning of the sales and can switch to the second price at any time. The second price is also determined adaptively based on the review level at the switching time.

For any policy, we evaluate the performance by "regret", which is the difference between the expected revenue of the policy and the optimal revenue if the customer knows the true quality of the product at the beginning (we will give a formal definition later). For the fixed two-price policy, we provide a policy that has a sublinear regret. We also prove that our proposed policy achieves the best regret asymptotically. The policy consists of a low initial price and a subsequent high price with a sublinear switching time, which means most of the time is spent on the second price. This is consistent with the real world practice, where a seller will give some discounts at the beginning and switch to the normal price after customers become familiar with the product. We then extend our result to the fixed multi-price case. We show that as the number of prices increases, the regret converges to the square root of the duration of the sales. After that, we show that the partially adaptive policy has the same order of regret as the predetermined fixed price policy, which means the extra flexibility on pricing does not have a significant value. Finally, we consider the fully adaptive two-price policy and prove it still has the same regret as the regret of the fixed two-price policy, which means the flexibility on the switching time does not help neither.

In addition to the model where customers give reviews based on a normal distribution, we also consider a discrete model, in which customers come at a discrete time with a general review distribution. We show that no matter what review metric the customers are using, five stars, ten stars, or "like" and "dislike", our analysis still holds.

The rest of the dissertation is organized in the following way: Chapter 2 reviews related literature. Chapter 3 discusses the basic setting of the model and studies the performance of different policies. Chapter 4 considers a discrete model. We conclude the dissertation and point out some future research directions in Chapter 6.

Chapter 2

Literature Review

In this section, we review related literature. Our work falls into the field of revenue management. Revenue management is a filed that concerns using disciplined tactics to optimize sales decisions (including product availability, prices, etc) to increase revenue growth of firms. It first arose from the airline industry, where airlines started adjusting fares to increase their revenue. For example, to win over customers from its competitors, American Airlines provided a low fare to those flights predicted to have empty seats with a requirement of advanced purchasing and non-refunding [6]. Later, the hospitality industry also adopted the idea of adjusting the price based on the inventory level, for it shared a similar business model with advanced booking and perishable inventory [7]. From there revenue management started to become a common practice and has been widely applied across various industries. At the heart of revenue management is the ability to provide different prices to customers with different level of willingness to pay. We refer readers to find more details about revenue management in [8], [9] and [10].

There is a growing research interest in revenue management in customer behavior, in which the purchase decision is not only determined by the current price but also by some other factors. For example, [11] propose the multinomial logit choice model

(MNL), where customers choose between multiple alternatives. [12] first apply the MNL model in the context of revenue management. They study the assortment problem that what product set should be offered to maximize the revenue and derive the structure of the optimal policy. [13] consider the cross-selling problem where a customer can choose between the requested item and a package including the requested item. [14] study the inventory control problem when a customer can choose between flights with the same origin and destination. [15] discuss the spiral-down effect when the seller ignored the customers' strategic behavior of choosing a lower fare product. [16] study a model that a customer's utility is influenced by past prices. [17] consider patient customers who can wait for the best price. [18] study the case where customers' utility is based on other people's usage of the product in his network. In this work, we study the pricing policy when there is a customer review effect.

The effect of customer review is part of the "word of mouth" effect. There is plenty of related literature on the effect of customer review in marketing. One area is the empirical study of how customer review influences sales. For example, [19] examine the relationship between box office revenue and its review. They argue that it is the number of the review, rather than the review itself plays an important role. [20] study the relation between book sales and its review on different online shopping platforms. They show that an increase in the review score can lead to an increase in the sales. [21] show that the number of blogs or posts showing an intention of purchasing before the release of a product is vital to the final sales. [22] study the influence of customer review on the sales of DVD players and find that while customer review has a substantial impact on the sales for those weak brands (not recognized by customers), it does not have a substantial impact on the sales of strong brands (famous brands with a good reputation). [23] study the relationship between the digital camera and its review. They find that for a high-involvement product, for which customers do some research before making a purchase decision, it is the review on external websites, not the review on the website that sells the digital camera that have a significant impact on the sales. In contrast to those empirical works, in this dissertation, we focus on how the seller should set the prices when aware of the review effect.

Our work is also related to dynamic pricing. As a way of price discrimination, dynamic pricing is often used to optimize revenue when the demand function is changing over time. [24] study the case when customers' arrival is based on a Poisson Process and the goods are perishable. They derive structural properties and get a closed-form solution when the demand satisfies some conditions. [25] consider a similar problem, but the price can only be chosen from a discrete set. They show the structure of the optimal policy and derive the closed-form solution. [26] consider a dynamic pricing problem across different products. They give a closed-form solution when the demand is deterministic and lower and upper bounds when the demand is stochastic. [27] study the dynamic pricing problem when the demand is not homogeneous. They show that the optimal price changes monotonically as the inventory level goes down. In our work, customers' willingness to pay changes as the review score changes. Therefore there is a need for dynamic pricing to optimize the revenue.

Besides price discrimination, another reason for using dynamic pricing is demand learning. The seller has to dynamically adjust the price to collect information about the demand when it is not fully revealed. One class of such study is to assume the demand function follows a specific parametric form. [28] first study the performance of iterative least square method (ILS). In the ILS, demand is a linear function of the price. The seller uses least square estimation to estimate the parameters, then uses myopic optimal price every step based on the estimated parameters. They show this strategy has an $O(T)$ regret, where T is the length of the selling period. [29] provide a strategy in fixed length price cycles. In each cycle, the seller first uses two test prices then switches to a myopic best price. The time spent on test prices is exponentially decreasing cycle by cycle. Moreover, they prove that under this strategy, the regret is $O(\sqrt{T})$. They also prove that for any policy, there are instances where the regret is $\Omega(\sqrt{T})$. [30] extend the work to multi-product setting and prove a much broader class of strategies that achieve the best regret order. They also consider the case where the seller knows the expected demand under some price. In this case, the regret of ILS is $O(\log T)$. [31] consider a variant of the problem, in which the true demand function is one of the two known functions to the seller. The seller has a prior belief of the distribution of the demand being each of the known function and updates that belief every step. They provide a policy and a comprehensive analysis of its regret. Another class of learning problems is non-parametric learning, which does not assume a parametric form of the demand function. [32] study a strategy that first considers a set of candidate prices, then tests the performance of each price in the learning phase, and finally uses the best price for the remaining time. They show that this strategy has a regret of $O(T^{1/4})$. [33] further improve the result by providing a new dynamic learning strategy achieving $O(T^{1/2})$ regret. Because of the randomness of the demand function in our setting, some methods are similar with demand learning problems. However, in our work, it is the customers who learn the quality of the product.

Finally, there are several papers that consider the pricing problem with review effect. For example, similar to our paper [34] also consider a pricing problem for the seller when customers use review to infer the quality of the product. However, in their setting, the review is related to both the quality of the product and its price. In our setting, the review only depends on the quality of the product. [35] also study a pricing problem under review effect. They assume that the review is drawn from a normal distribution, centered at the true quality of the product and only the seller knows the true quality. The main difference is that in their work, the customers utility function is stochastic, which means there is always a positive probability for a customer to buy the product regardless of the current price and review. This ensures that the learning process will continue, and customers will learn the true quality of the product eventually. Moreover, they simplify the model by converting the problem to a deterministic problem, using the probability of purchase as the demand rate and customers will always post the same review score. While in our setting, the utility function is deterministic, which means the learning process may stop under specific price and review unless the seller changes the price. This imposes additional constraints to the seller, and therefore the seller has to be more conservative about pricing. Another difference is that in their model customers use a Bayesian method to update their belief, while in our model customers use the simple average review score, which is commonly used on major online shopping websites. These differences lead to different results between our model and theirs.

Chapter 3

Continuous Model

In this chapter we introduce the basic model setting and discuss various policies. Consider a seller that sells one product to a sequence of customers in a discrete time horizon indexed by $t = 0, 1, 2, ..., T$. The product has a mean quality of q and for each customer who purchases this product, he will experience a valuation of the product drawn from a normal distribution $N(q, \sigma^2)$. We assume that only the seller knows q and σ but not the customers.

In our model, we assume that at each time period, there is a review score for this product which affects the purchase decision of customers. In particular, we use q_t to denote the current review score at time period t, then we assume that the customer's utility of purchasing the product is $u_t = q_t - p$ where p is the price in that period. If u_t is positive, then the customer will purchase the product, in which case he will leave a review with the true valuation r_t he experienced (drawn from $N(q, \sigma^2)$). In our model, we assume that the review score is the average of all past reviews. Therefore, the average review for the product in the next period will be

$$q_{t+1} = \frac{tq_t + r_t}{t+1}.$$

In our model, we assume that the starting time is t_0 and the initial review level at t_0 is q_0. One can view that this initial set of reviews are obtained by some die-hard customers that will buy and review the products anyway. They could also be professional product reviewers. Given the above-defined dynamics, we have the review at time t must follow:

$$tq_t = t_0 q_0 + r_{t_0+1} + r_{t_0+2} + \cdots + r_t,$$

where $r_{t_0+1}, \ldots, r_t \sim N(q, \sigma^2)$. To simplify the discussion, we consider a diffusion approximation of the above process. In particular, we use a Brownian motion to replace the sum of normal random variables. Let G_t be the approximate total review level at time t. Starting from (t_0, q_0), G_t is a Brownian motion with drift:

$$G_t = t_0 q_0 + q(t - t_0) + \sigma \mathbb{B}_{t-t_0},\ t \geq t_0,$$

where $\mathbb{B}_s$ is a standard Brownian motion. Obviously, when the review process G_t falls below pt, customers will no longer purchase the product, and the sales stop until the seller switches to another price. The seller's goal is to maximize his expected revenue. We use π to denote a non-anticipating policy, in which a price to be used at time t is determined only by the history information $\{G_s, s \leq t\}$. Denote $J_T(\pi)$ the expected revenue of the seller using policy π. In this chapter we will focus on four policies

- A fixed two-price policy. The seller can use at most two prices during the sales. He needs to determind the two prices he will use and the time to swithc the price beforehand.

- A fixed multi-price policy. The seller can set multiple prices. All the prices and swithcing times have to be predetermined.

- A partially adaptive policy. In this policy, only the switching times are predetermined. The seller can decide all the prices (except the first one) adaptively at the switching time.

- A fully adaptive two-price policy. In this policy, only the first price is pretermined. The seller can switch to the second price at any time. The second price is also determined adaptively

Note that if the customers know the true quality q in the beginning, then the seller will set a price $p = q$ for the entire selling season, and the revenue until time T, will be $q(T - t_0)$. Therefore, we define the regret of the seller when using policy π by

$$R_T(\pi) = q \cdot (T - t_0) - J_T(\pi).$$

The seller's goal is to minimize his regret.

The rest of this chapter is organized as follows. In Section 3.1, we discuss a fixed two-price policy. In Section 3.2, we consider a fixed multi-price policy. Those are all fixed price policies, where the seller has to specify all the policy details at the start of the sales. In Section 3.3, we discuss some adaptive policies, in which the seller does not need to decide everything in the beginning of the sales.

3.1 Fixed Two-Price Case

We first consider a case in which the seller can set at most two prices during the entire selling horizon. The seller's policy consists of an initial price p_1, a stopping time t_1 adapted to the process G_t and a second price p_2 which is a function of the history of G_t before t_1.

Consider a fixed two-price policy, in which the switching time t_1 and the second price p_2 are chosen in advance rather than determined adaptively based on the process G_t. To be more precise, in a fixed price policy π_{t_1,p_1,p_2}, the price p_1 is used from t_0 until t_1 or when the sales stop, whichever comes first. Then the seller switches the price to p_2 at t_1. We call $[t_0, t_1]$ the first period and $[t_1, T]$ the second period.

Now we analyze the performance of this policy. Let $\tau_1 = \inf\{t : G_t < p_1 t\} \wedge t_1$ be the time the sales stop during the first period. Similarly, let $\tau_2 = \inf\{t : G_t < p_2 t, t \geq \tau_1\} \wedge (\tau_1 + T - t_1)$ be the time the sales stop during the second period.

With the above notations, we represent the seller's revenue for policy π_{t_1,p_1,p_2} until time T as:

$$J_T(\pi_{t_1,p_1,p_2}) = \mathbb{E}_\tau[p_1(\tau_1 - t_0) + p_2(\tau_2 - \tau_1)].$$

The regret for a fixed two-price policy is

$$R_T(\pi_{t_1,p_1,p_2}) = q \cdot (T - t_0) - J_T(\pi_{t_1,p_1,p_2}).$$

Next, we show that by properly choosing the parameters t_1, p_1 and p_2, one can achieve an asymptotically optimal revenue (or equivalently, a sublinear regret). We have the following theorem:

Theorem 1. *Let* $t_1 = T^{2/3}$, $p_1 = -\log T$ *and* $p_2 = q - T^{-1/3}\log T$. *Then* $R_T(\pi_{t_1,p_1,p_2}) = O(T^{2/3}\log T)$.

Before we prove this theorem, we make a few comments. One may notice that p_1 is negative in the statement of Theorem 1. This is because we assume the review of each customer follows a normal distribution. If we restrict p_1 to be positive, then there is a constant probability that the sales terminate in the first period, in which case the regret would be $O(T)$ (one can see the detail of this argument in the proof of Theorem 2). In practice, the review of products are usually positive (for example a number between 0 and 10). In such a case, the seller can just set $p_1 = 0$ to make sure that the sales can continue, and Theorem 1 would also hold. We will discuss this case in detail in Chapter 4. The general idea of the policy matches what we have seen in the real practice, that a seller provides a promotion (a low price) in the beginning of the sales for a short period, and then switches to a normal price for the rest of the sales horizon.

Proof. Recall that

$$J_T(\pi_{t_1,p_1,p_2}) = \mathbb{E}_{\tau_1,\tau_2}[p_1(\tau_1 - t_0) + p_2(\tau_2 - \tau_1)].$$

Since $p_1 = -\log T < 0$ and $\tau_1 \leq t_1$, we have

$$p_1 \cdot \mathbb{E}_\tau[t_1 \wedge \tau - t_0] \geq p_1(t_1 - t_0).$$

Similarly, since $p_2 > 0$, $\mathbb{E}_{\tau_1,\tau_2}[\tau_2 - \tau_1] \geq (T - t_1)\mathbb{P}(\tau_2 = \tau_1 + T - t_1)$, we have

$$p_2 \cdot \mathbb{E}_{\tau_1,\tau_2}[\tau_2 - \tau_1] \geq p_2 \cdot (T - t_1)\mathbb{P}(\tau_2 = \tau_1 + T - t_1).$$

In the following, we denote $A_1 = \{\tau_1 < t_1\}$ and $A_2 = \{\tau_2 < \tau_1 + T - t_1\}$. Therefore $A_1 \cup A_2 = \{\tau_2 < T\}$. And we have:

$$\begin{aligned}
J_T(\pi_{t_1,p_1,p_2}) &\geq p_1(t_1 - t_0) + p_2(T - t_1) \cdot \mathbb{P}(\tau_2 = T) \\
&= p_1(t_1 - t_0) + p_2(T - t_1) \cdot (1 - \mathbb{P}(A_1 \cup A_2)) \\
&\geq p_1(t_1 - t_0) + p_2(T - t_1) \cdot (1 - \mathbb{P}(A_1) - \mathbb{P}(A_2)).
\end{aligned}$$

Therefore,

$$\begin{aligned} R_T(\pi_{t_1,p_1,p_2}) &= q(T-t_0) - J_T(\pi_{t_1,p_1,p_2}) \\ &\le q(T-t_0) - [p_1(t_1-t_0) + p_2(T-t_1)(1-\mathbb{P}(A_1)-\mathbb{P}(A_2))] \\ &= (q-p_2)T + (p_2-p_1)t_1 + p_2(T-t_1)\mathbb{P}(A_1) + p_2(T-t_1)\mathbb{P}(A_2) - (q-p_1)t_0 \\ &\le (q-p_2)T + (p_2-p_1)t_1 + p_2T\mathbb{P}(A_1) + p_2T\mathbb{P}(A_2). \end{aligned} \tag{1.1}$$

In the following, we will bound each of the four terms in (1.1). For the first term, by the definition of p_2, we have

$$(q-p_2)\cdot T = T^{2/3}\log T.$$

For the second term, we have

$$(p_2-p_1)t_1 = (q - T^{-1/3}\log T)\cdot T^{2/3} + T^{2/3}\log T = O(T^{2/3}\log T).$$

For the third term, by Lemma 7, we have:

$$\mathbb{P}(A_1) \le \mathbb{P}(\cup_{t_0\le t\le\infty}\{G_t \le p_1 t\}) = \exp\left(-\frac{2(q-p_1)(q_0-p_1)t_0}{\sigma^2}\right) \le \exp\left(-\frac{2t_0\log^2 T}{\sigma^2}\right).$$

Therefore, when $T \ge \exp(\sigma^2/2t_0)$, $\mathbb{P}(A_1) \le 1/T$. Thus $p_2T\mathbb{P}(A_1) = O(1)$.

For the last term, we first bound $\mathbb{P}(A_2)$. Conditioning $\mathbb{P}(A_2)$ on the realization of

G_{t_1}, we have:

$$\begin{aligned}\mathbb{P}(A_2) &= \int_y \mathbb{P}(A_2|G_{t_1}=y)\mathbb{P}(G_{t_1}=y)dy \\ &\leq \Phi\left(\frac{p_2t_1 - q_0t_0 - q(t_1-t_0)}{\sigma\sqrt{t_1-t_0}}\right) \\ &+ \int_{p_2t_1}^{\infty} \frac{1}{\sigma\sqrt{t_1-t_0}}\phi\left(\frac{y-q_0t_0-q(t_1-t_0)}{\sigma\sqrt{t_1-t_0}}\right)\exp\left(-\frac{2(q-p_2)(y-p_2t_1)}{\sigma^2}\right)dy.\end{aligned} \tag{1.2}$$

where $\phi(x) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}x^2)$ and $\Phi(x) = \int_{-\infty}^{x}\phi(y)dy$ are the density and cumulative distribution function for standard normal distribution. Here the first term corresponds to the case where $G_{t_1} \leq p_2t_1$ and the second term uses the fact that because of Lemma 7,

$$\mathbb{P}(A_2|G_{t_1}=y) = \exp(-2(q-p_2)(y-p_2t_1)/\sigma^2).$$

For the first term in (1.2), we note that when T is sufficiently large,

$$p_2t_1 - q_0t_0 - q(t_1-t_0) = -T^{1/3}\log T + (q-q_0)t_0 < 0.$$

Therefore, by the fact that for any $x < 0$, $\Phi(x) \leq \frac{1}{\sqrt{2\pi}|x|}\exp(-x^2/2)$ [36] , we have:

$$\Phi\left(\frac{p_2t_1 - q_0t_0 - q(t_1-t_0)}{\sigma\sqrt{t_1-t_0}}\right) \leq \frac{\sigma T^{1/3}}{\sqrt{2\pi}(T^{1/3}\log T - (q-q_0)t_0)}\exp\left(-\frac{(T^{1/3}\log T - (q-q_0)t_0)^2}{2\sigma^2(T^{2/3}-t_0)}\right).$$

When T is large,

$$\frac{\sigma T^{1/3}}{\sqrt{2\pi}(T^{1/3}\log T - (q-q_0)t_0)} \leq 1 \text{ and } -\frac{(T^{1/3}\log T - (q-q_0)t_0)^2}{2\sigma^2(T^{2/3}-t_0)} \leq -\log T.$$

Thus,

$$\Phi\left(\frac{p_2t_1 - q_0t_0 - q(t_1-t_0)}{\sigma\sqrt{t_1-t_0}}\right) = O\left(\frac{1}{T}\right).$$

For the second term in (1.2), we have:

$$\begin{aligned}
&\int_{p_2t_1}^{\infty} \exp\left(-\frac{2(q-p_2)(y-p_2t_1)}{\sigma^2}\right)\frac{1}{\sigma\sqrt{t_1-t_0}}\phi\left(\frac{y-q_0t_0-q(t_1-t_0)}{\sigma\sqrt{t_1-t_0}}\right)dy\\
=&\exp\left(-\frac{2(q-p_2)(q_0-p_2)t_0}{\sigma^2}\right)\\
&*\int_{p_2t_1}^{\infty}\frac{1}{\sqrt{2\pi}\sigma\sqrt{t_1-t_0}}\exp\left(-\frac{(y-q_0t_0-q(t_1-t_0)+2(q-p_2)(t_1-t_0))^2}{2\sigma^2(t_1-t_0)}\right)dy\\
=&\exp\left(-\frac{2(q-p_2)(q_0-p_2)t_0}{\sigma^2}\right)\left[1-\Phi\left(\frac{(q-p_2)(t_1-t_0)-(q_0-p_2)t_0}{\sigma\sqrt{t_1-t_0}}\right)\right]\\
=&\exp\left(-\frac{2(q-p_2)(q_0-p_2)t_0}{\sigma^2}\right)\Phi\left(\frac{(q_0-p_2)t_0-(q-p_2)(t_1-t_0)}{\sigma\sqrt{t_1-t_0}}\right),
\end{aligned}$$

When T is sufficiently large, $-2(q-p_2)(q_0-p_2)t_0/\sigma^2 \le 2q(q_0+q)t_0/\sigma^2$ and $(q_0 - p_2)t_0 - (q-p_2)(t_1-t_0) = q_0t_0 - qt_0 + 2T^{-1/3}\log T t_0 - T^{1/3}\log T < 0$. Again, we apply that when $x<0$, $\Phi(x) \le \frac{1}{\sqrt{2\pi}|x|}\exp(-x^2/2)$. We have

$$\begin{aligned}
&\Phi\left(\frac{(q_0-p_2)t_0-(q-p_2)(t_1-t_0)}{\sigma\sqrt{t_1-t_0}}\right)\\
\le&\frac{\sigma T^{1/3}}{\sqrt{2\pi}(T^{1/3}\log T - 2T^{-1/3}\log T - q_0t_0 + qt_0)}\exp\left(\frac{-(T^{1/3}\log T - 2T^{-1/3}\log T - q_0t_0+qt_0)^2}{2\sigma^2(T^{2/3}-t_0)}\right).
\end{aligned}$$

When T is large,

$$\begin{aligned}
\frac{\sigma T^{1/3}}{\sqrt{2\pi}(T^{1/3}\log T - 2T^{-1/3}\log T - q_0t_0 + qt_0)} &\le 1,\\
\frac{-(T^{1/3}\log T - 2T^{-1/3}\log T - q_0t_0+qt_0)^2}{2\sigma^2(T^{2/3}-t_0)} &\le -\log T.
\end{aligned}$$

Thus, we have proved that $\mathbb{P}(A_2) = O(\frac{1}{T})$ and by combining these four terms, we have:

$$R_T(\pi_{t_1,p_1,p_2}) = O(T^{2/3} \log T).$$

□

Next, we will show that a regret of $O(T^{2/3})$ is the best one can do among all fixed two-price policies. We first introduce a lemma to bound the expected stopping time for a single price policy.

Lemma 1. *Let* $\tau_p = \inf\{t : G_t < pt, t \geq t_0\} \wedge T$*. For any* $T > 0$*, if* $p \leq q_0$*, then*

$$\begin{aligned}
&\mathbb{E}_{\tau_p}[\tau_p - t_0] \\
&\leq (T - t_0)\left[1 - \exp\left(-\frac{2(q_0 - p)(q - p)t_0}{\sigma^2}\right)\right]^+ \\
&+ \frac{4(q_0 - p)t_0}{\sigma} \min\left(1, \exp\left(-\frac{2(q_0 - p)(q - p)t_0}{\sigma^2}\right)\right)\sqrt{T - t_0}.
\end{aligned}$$

If $p > q_0$*, then* $\mathbb{E}_{\tau_p}[\tau_p - t_0] = 0$*.*

Proof. Let $a = (q_0 - p)t_0/\sigma, b = (q - p)/\sigma$. Applying Lemma 8 we have when $p \leq q_0$,

$$\mathbb{E}_{\tau_p}[\tau_p - t_0] \leq (T - t_0)[1 - \exp(-2ab)]^+ + 4a \min\{1, \exp(-2ab)\}\sqrt{T - t_0}.$$

Thus the lemma holds. □

With Lemma 1, we can now bound the revenue generated after switching the price.

Lemma 2. *Denote* $G_t = t_0 q_0 + q(t - t_0) + \sigma \mathbb{B}_{t-t_0}$*,* $t \geq t_0$*. Let* $\tau_{p,s} = \inf\{t : G_t < pt, t \geq s\} \wedge (T - s)$ *and* $S = q(T - s) - p\mathbb{E}_{\tau_{p,s}}[\tau_{p,s} - s]$*. For any* $T > 0$*,* $\epsilon > 0$*, if* $p > q + T^{-(1/2-\epsilon)}$

and $s \geq T^{1-\epsilon}$*, then*

$$S = \Omega(T - s),$$

if $s \leq T^{1-\epsilon}$*,*

$$S = \Omega\left(\frac{T - s}{\sqrt{s - t_0}}\right).$$

Lemma 2 shows the relationship between the switching time and the regret. The seller will have at least $\Omega(T/\sqrt{t})$ regret if he switches the price at t. The intuition is that, the standard deviation for G_{t_1} at time t_1 is $\sigma\sqrt{t_1}$. To make sure the sale process continues, the price should satisfy $q - p_2 = \Omega(\log T t_1^{-1/2})$ at time t. Otherwise, if $p_2 >= q - \sigma/\sqrt{t_1}$, the sales will stop at t_1 with a constant probability. With the above two lemmas, we state the second theorem to bound the total regret of a general fixed two-price policy.

Theorem 2. *For any fixed two-price policy* π_{t_1,p_1,p_2}*,* $R_T(\pi_{t_1,p_1,p_2}) = \Omega(T^{2/3})$*.*

Theorem 2 shows that our policy in Theorem 1 achieves the best regret order. The proof first bounds the regret for the first price, which is linear in the switching time. This is mainly because if the seller charges a high first price when the process just starts, the sales have a high probability to stop since the initial review q_0 and t_0 are constants. To reduce the risk, the seller has to set a low price thus leading to a linear regret.

The next step is to use Lemma 2 to bound the regret in the second period. If the switching time is sublinear to T, we can directly use Lemma 2 to bound the regret. On the other hand, if the switching time is close to T, then the regret in the first period will be high since it is linear on the switching time. Combining the two cases we can prove the theorem.

Proof. The revenue of a fixed two-price policy π_{t_1,p_1,p_2} is

$$J_T(\pi_{t_1,p_1,p_2}) = \mathbb{E}_{\tau_1,\tau_2}[p_1(\tau_1 - t_0) + p_2(\tau_2 - \tau_1)].$$

Therefore,

$$\begin{aligned} R_T(\pi_{t_1,p_1,p_2}) =& q(T - t_0) - J_T(\pi_{p_1,p_2,t_1}) \\ \geq& \underbrace{q(t_1 - t_0) - p_1\mathbb{E}_{\tau_1}[\tau_1 - t_0]}_{S_1} + \underbrace{q(T - t_1) - p_2\mathbb{E}_{\tau_1,\tau_2}[\tau_2 - \tau_1]}_{S_2}, \end{aligned}$$

We first bound S_1. We consider four possible cases.

1. If $p_1 \leq 0$, then $p_1\mathbb{E}_{\tau_1}[\tau_1 - t_0] \leq 0$, and $S_1 \geq q(t_1 - t_0)$.

2. If $p_1 \geq q_0$, then $p_1\mathbb{E}_{\tau_1}[\tau_1 - t_0] = 0$, and $S_1 = q(t_1 - t_0)$.

3. If $0 \leq p_1 \leq \min(q, q_0)$, then by Lemma 1,

$$p_1\mathbb{E}_{\tau_1}[\tau_1 - t_0)] \leq q(t_1 - t_0)[1 - \exp(-2q_0qt_0/\sigma^2)] + \frac{4qq_0t_0}{\sigma}\sqrt{t_1 - t_0},$$

   and thus

$$S_1 \geq q(t_1 - t_0)\exp(-2q_0qt_0/\sigma^2) - \frac{4qq_0t_0}{\sigma}\sqrt{t_1 - t_0}.$$

4. If $q < p_1 \leq q_0$, then by Lemma 1,

$$p_1\mathbb{E}_{\tau_1}[\tau_1 - t_0] \leq \frac{4p_1(q_0 - p_1)t_0}{\sigma}\sqrt{t_1 - t_0} \leq \frac{q_0^2t_0}{\sigma}\sqrt{t_1 - t_0},$$

   and

$$S_1 \geq q(t_1 - t_0) - \frac{q_0^2t_0}{\sigma}\sqrt{t_1 - t_0}.$$

Therefore in all cases, $S_1 = \Omega(t_1)$. Now let's consider S_2,

$$S_2 = q(T-t_1) - p_2\mathbb{E}_{\tau_1,\tau_2}[\tau_2-\tau_1] = \mathbb{E}_{\tau_1}[q(T-t_1) - p_2\mathbb{E}_{\tau_2}[\tau_2-\tau_1|\tau_1]].$$

If $p_2 \le q + T^{-3/8}$ and $t_1 \ge T^{7/8}$,

$$S_2 \ge q(T-t_1) - (q+T^{-3/8})(T-t_1) \ge -T^{5/8}, S_1 = \Omega(t_1) = \Omega(T^{7/8}).$$

The total regret is $\Omega(T^{7/8})$.

Let $\epsilon = 1/8$, from lemma 2, if $p_2 \ge q + T^{-3/8}$ and $t_1 \ge T^{7/8}$,

$$S_2 = \mathbb{E}_{\tau_1}[\Omega(T-\tau_1)] \ge \Omega(T-t_1).$$

The total regret is

$$S_1 + S_2 = \Omega(t_1) + \Omega(T-t_1) = \Omega(T).$$

If $t_1 < T^{7/8}$, from Lemma 2, we have

$$S_2 = \mathbb{E}_{\tau_1}\left[\Omega\left(\frac{T-t_1}{\sqrt{\tau_1-t_0}}\right)\right] \ge \Omega\left(\frac{T-t_1}{\sqrt{t_1-t_0}}\right).$$

In this case,

$$\begin{aligned} R_T(\pi_{p_1,p_2,t_1}) = S_1 + S_2 &\ge \Omega(t_1-t_0) + \Omega\left(\frac{T-t_1}{\sqrt{t_1-t_0}}\right) \\ &\ge c_1(t_1-t_0) + \frac{c_2(T-t_1)}{2\sqrt{t_1-t_0}} + \frac{c_2(T-t_1)}{2\sqrt{t_1-t_0}} \\ &\ge \left[\frac{c_1c_2^2(T-t_1)^2}{4}\right]^{1/3} \\ &= \Omega(T^{2/3}). \end{aligned}$$

□

To summarize, in this section we provide a fixed two-price policy with a sublinear regret $O(T^{2/3}\log T)$. We also prove this policy achieves the best performance asymptotically for any fixed two-price policy.

3.2 Fixed Multi-Price Case

Theorem 2 proves all the fixed two-price policy have a regret of $\Omega(T^{2/3}/\log T)$. The question then arises: how can we do better? Since the main cause of the regret is that the price needs to follow $q-p=\Omega(\log T t^{-1/2})$, one obvious improvement is to let the seller set a fixed multi-price policy so the price can keep increasing as the sales continues. Such a policy is discussed in this session.

We consider a general setting where the seller can set K different prices instead of two. We use $\pi_{T,K}(t_1,t_2,\ldots,t_{K-1};p_1,p_2,\ldots,p_K)$ to denote a fixed price policy in which price p_i is used from t_{i-1} to t_i. We still use the notation $\tau_0=t_0, t_K=T$, $\tau_i=\inf\{t:G_t<p_it, t\geq\tau_{i-1}\}\wedge(\tau_{i-1}+t_i-t_{i-1})$.

Next we define the seller's revenue for policy $\pi_{T,K}(t_1,t_2,\ldots,t_{K-1};p_1,p_2,\ldots,p_K)$. We have

$$J_T(\pi_{T,K}(t_1,t_2,\ldots,t_{K-1};p_1,p_2,\ldots,p_K))=\mathbb{E}_\tau\left[\sum_{i=1}^{K}p_i(\tau_i-\tau_{i-1})\right].$$

We also define the regret of a policy as

$$R_T(\pi_{T,K}(t_1,t_2,\ldots,t_{K-1};p_1,p_2,\ldots,p_K))=qT-J_T(\pi_{T,K}(t_1,t_2,\ldots,t_{K-1};p_1,p_2,\ldots,p_K)).$$

Now we propose a policy in the K prices setting. Let $a_i=\frac{1}{2}+\frac{1}{2(2^i-1)},\ i=1,\ldots,K$.

	p_1	t_1	p_2	t_2	p_3	t_3	p_4	t_4	Regret
$K=2$	$-\log T$	$T^{\frac{2}{3}}$	$q-\log T/T^{\frac{1}{3}}$	T	N/A	N/A	N/A	N/A	$O(T^{\frac{2}{3}}\log T)$
$K=3$	$-\log T$	$T^{\frac{4}{7}}$	$q-\log T/T^{\frac{2}{7}}$	$T^{\frac{6}{7}}$	$q-\log T/T^{\frac{3}{7}}$	T	N/A	N/A	$O(T^{\frac{4}{7}}\log T)$
$K=4$	$-\log T$	$T^{\frac{8}{15}}$	$q-\log T/T^{\frac{4}{15}}$	$T^{\frac{4}{5}}$	$q-\log T/T^{\frac{2}{5}}$	$T^{\frac{14}{15}}$	$q-\log T/T^{\frac{7}{15}}$	T	$O(T^{\frac{8}{15}}\log T)$

Table 3.1: Illustrations of Policy $\hat{\pi}_{T,k}$.

We define

$$\hat{\pi}_{T,K} = \pi_{T,K}(\hat{t}_1, \hat{t}_2, \ldots, \hat{t}_{K-1}; \hat{p}_1, \hat{p}_2, \ldots, \hat{p}_K),$$

where

$$\hat{t}_K = T,\ \hat{t}_i = \hat{t}_{i+1}^{2/(1+2a_i)} = \hat{t}_{i+1}^{1-1/(2^{i+1}-1)},\ i = 1, 2, \ldots, K-1, \tag{2.3}$$

and

$$\hat{p}_1 = -\log T,\ \hat{p}_i = q - \frac{\log T}{\sqrt{\hat{t}_{i-1}}},\ i = 2, \ldots, K.$$

We have the following result about $\hat{\pi}_{T,K}$.

Theorem 3. *$R_T(\hat{\pi}_{T,K}) = O(T^{a_K}\log T)$, for $K \geq 2$.*

Before we prove Theorem 3, we show some examples of $\hat{\pi}_{T,K}$ for different K in Table 3.1 to illustrate the algorithm. The regret decreases and converges to $O(T^{1/2}\log T)$ as K increases.

One can see that in figure 3.1, in policy $\hat{\pi}_{T,K}$, the seller updates the prices less frequently toward the end of the selling horizon, and the price is closer and closer to q.

Proof. We prove this theorem by induction. In Theorem 1, we have already proved that for $K = 2$, $R_T(\hat{\pi}_{T,2}) = O(T^{2/3}\log T)$, which is consistent with Theorem 3 for $K = 2$. Suppose the result holds for $K = k-1$, we now consider $K = k$.

In the following arguments, if not otherwise specified, the notation t_i, p_i, τ_i refer to the time stamp, price and stopping time in the policy $\hat{\pi}_{T,k}$.

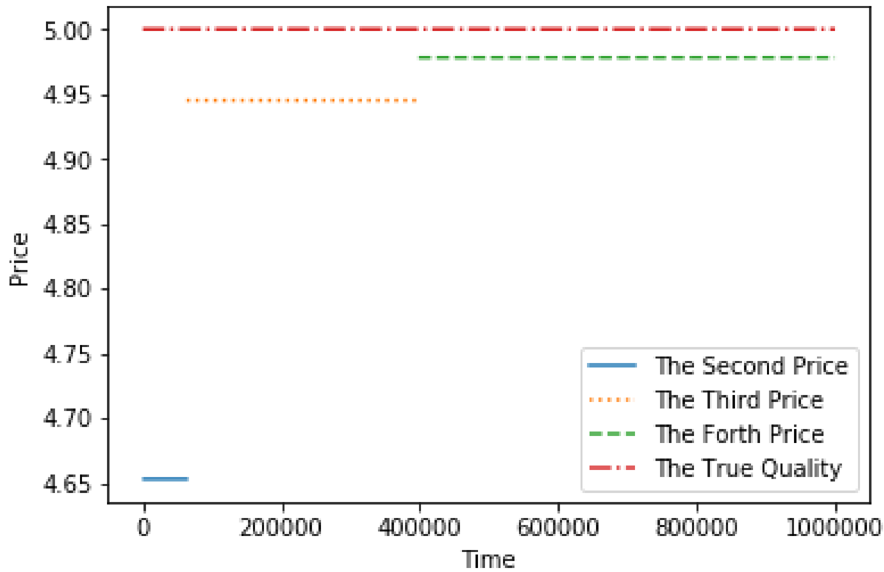

Figure 3.1: A Sample price path for a 4-price policy

The regret of $\hat{\pi}_{T,k}$ is:

$$
\begin{aligned}
R_T(\hat{\pi}_{T,k}) &= qT - J_T(\hat{\pi}_{T,k}) \\
&= qt_{k-1} - J_{t_{k-1}}(\hat{\pi}_{t_{k-1},k-1}) + J_{t_{k-1}}(\hat{\pi}_{t_{k-1},k-1}) - J_{t_{k-1}}(\hat{\pi}_{T,k}) \\
&\quad + q(T - t_{k-1}) - \mathbb{E}_\tau p_k(\tau_K - \tau_{K-1}),
\end{aligned}
$$

where $J_{t_{k-1}}(\hat{\pi}_{T,k}) = \mathbb{E}[\sum_{i=1}^{k-1} p_i(\tau_i - \tau_{i-1})]$, that is, the revenue of $\hat{\pi}_{T,k}$ until t_{k-1}.

Similar with Theorem 1, we denote $A_i = \{\tau_i < \tau_{i-1} + t_i - t_{i-1}\}$. And we have

$$\begin{aligned} R_T(\hat{\pi}_{T,k}) &\leq qt_{k-1} - J_{t_{k-1}}(\hat{\pi}_{t_{k-1},k-1}) + J_{t_{k-1}}(\hat{\pi}_{t_{k-1},k-1}) - J_{t_{k-1}}(\hat{\pi}_{T,k}) \\ &\quad + q(T - t_{k-1}) - p_k(T - t_{k-1}) \left(1 - \sum_{i=1}^{k} \mathbb{P}(A_i)\right) \\ &\leq qt_{k-1} - J_{t_{k-1}}(\hat{\pi}_{t_{k-1},k-1}) + J_{t_{k-1}}(\hat{\pi}_{t_{k-1},k-1}) - J_{t_{k-1}}(\hat{\pi}_{T,k}) \\ &\quad + (q - p_k)(T - t_{k-1}) + qT \sum_{i=1}^{k} \mathbb{P}(A_i). \end{aligned}$$

For the first term, by induction, we have

$$qt_{k-1} - J_{t_{k-1}}(\hat{\pi}_{t_{k-1},k-1}) = O(t_{k-1}^{a_{k-1}} \log t_{k-1}) = O(t_k^{2a_{k-1}/(1+2a_{k-1})} \log t_k) = O(T^{a_k} \log T),$$

where we use the fact that $a_k = 2a_{k-1}/(1 + 2a_{k-1})$ by definition.

For the second term, by definition, we have

$$\begin{aligned} &|J_{t_{k-1}}(\hat{\pi}_{t_{k-1},k-1}) - J_{t_{k-1}}(\hat{\pi}_{T,k})| \\ \leq &\sum_{i=2}^{k-1} t_i |(q - \log t_{k-1}/\sqrt{t_{i-1}}) - (q - \log T/\sqrt{t_{i-1}})| + t_1(\log T - \log t_{k-1}) \\ = &\sum_{i=2}^{k-1} t_i \log(T/t_{k-1})/\sqrt{t_{i-1}} + t_1 \log(T/t_{k-1}). \end{aligned}$$

By definition,

$$t_i/\sqrt{t_{i-1}} = t_i^{1-1/(1+2a_{i-1})} = t_i^{a_i} = t_{i+1}^{2a_i/(1+2a_i)} = t_{i+1}^{a_{i+1}} = T^{a_k}.$$

Because $a_1 = 1$, we also have $t_1 = t_1^{a_1} = T^{a_k}$. And by that $\hat{t}_{k-1} = T^{1-1/(2^k-1)}$,

$$\log(T/t_{k-1}) = \frac{1}{2^k - 1} \log T.$$

Combining the above argument, we have

$$|J_{t_{k-1}}(\hat{\pi}_{t_{k-1},k-1}) - J_{t_{k-1}}(\hat{\pi}_{T,k})| \leq \frac{k}{2^k - 1} \log T \cdot T^{a_k} = O(T^{a_k} \log T).$$

For the third term, by definition, we have

$$(q - p_k)(T - t_{k-1}) = O(T \log T / \sqrt{t_{k-1}}) = O(T^{2a_{k-1}/(1+2a_{k-1})} \log T) = O(T^{a_k} \log T).$$

For the last term, we first bound $\mathbb{P}(A_i)$. By similar arguments as in (1.2), we have

$$\begin{aligned}
\mathbb{P}(A_i) = &\int_{\mathbb{R}} \mathbb{P}(A_i | G_{t_{i-1}} = y) \mathbb{P}(G_{t_{i-1}} = y) dy \\
\leq &\, \Phi\left(\frac{p_i t_{i-1} - q_0 t_0 - q(t_{i-1} - t_0)}{\sigma\sqrt{t_{i-1} - t_0}}\right) \\
&+ \int_{p_i t_{i-1}}^{\infty} \frac{1}{\sigma\sqrt{t_{i-1} - t_0}} \phi\left(\frac{y - q_0 t_0 - q(t_{i-1} - t_0)}{\sigma\sqrt{t_{i-1} - t_0}}\right) \exp\left(-\frac{2(q - p_i)(y - p_i t_{i-1})}{\sigma^2}\right) dy,
\end{aligned} \tag{2.4}$$

where the last inequality is because of Lemma 7 and that $A_i \subseteq \{G_t \geq p_i t, \text{ for all } t \geq t_{i-1}\}$. For the first term in (2.4), when T is sufficiently large,

$$p_i t_{i-1} - q_0 t_0 - q(t_{i-1} - t_0) = -\sqrt{t_{i-1}} \log T + (q - q_0) t_0 < 0.$$

Now applying that when $x < 0$, $\Phi(x) \le \frac{1}{\sqrt{2\pi}|x|}\exp(-x^2/2)$, we have

$$\Phi\left(\frac{p_i t_{i-1} - q_0 t_0 - q(t_{i-1}-t_0)}{\sigma\sqrt{t_{i-1}-t_0}}\right) \le \frac{\sigma\sqrt{t_{i-1}}}{\sqrt{2\pi}(\sqrt{t_{i-1}}\log T - (q-q_0)t_0)}\exp\left(-\frac{(\sqrt{t_{i-1}}\log T - (q-q_0)t_0)^2}{2\sigma^2(t_{i-1}-t_0)}\right).$$

When T is large,

$$\frac{\sigma\sqrt{t_{i-1}}}{\sqrt{2\pi}(\sqrt{t_{i-1}}\log T - (q-q_0)t_0)} \le 1 \text{ and } -\frac{(\sqrt{t_{i-1}}\log T - (q-q_0)t_0)^2}{2\sigma^2(t_{i-1}-t_0)} \le -\log T.$$

Thus,

$$\Phi\left(\frac{p_i t_{i-1} - q_0 t_0 - q(t_{i-1}-t_0)}{\sigma\sqrt{t_{i-1}-t_0}}\right) = O\left(\frac{1}{T}\right).$$

For the second term in (2.4), we have:

$$\begin{aligned}
&\int_{p_i t_{i-1}}^{\infty} \exp\left(-\frac{2(q-p_i)(y-p_i t_{i-1})}{\sigma^2}\right)\frac{1}{\sigma\sqrt{t_{i-1}-t_0}}\phi\left(\frac{y-q_0t_0-q(t_{i-1}-t_0)}{\sigma\sqrt{t_{i-1}-t_0}}\right)dy\\
=&\exp\left(-\frac{2(q-p_i)(q_0-p_i)t_0}{\sigma^2}\right)\\
&*\int_{p_i t_{i-1}}^{\infty}\frac{1}{\sqrt{2\pi}\sigma\sqrt{t_{i-1}-t_0}}\exp\left(-\frac{(y-q_0t_0-q(t_{i-1}-t_0)+2(q-p_i)(t_{i-1}-t_0))^2}{2\sigma^2(t_{i-1}-t_0)}\right)dy\\
=&\exp\left(-\frac{2(q-p_i)(q_0-p_i)t_0}{\sigma^2}\right)\left[1-\Phi\left(\frac{(q-p_i)(t_{i-1}-t_0)-(q_0-p_i)t_0}{\sigma\sqrt{t_{i-1}-t_0}}\right)\right]\\
=&\exp\left(-\frac{2(q-p_i)(q_0-p_i)t_0}{\sigma^2}\right)\Phi\left(\frac{(q_0-p_i)t_0-(q-p_i)(t_{i-1}-t_0)}{\sigma\sqrt{t_{i-1}-t_0}}\right).
\end{aligned}$$

When T is sufficiently large, $-2(q-p_i)(q_0-p_i)t_0/\sigma^2 \le 2q(q_0+q)t_0/\sigma^2$ and $(q_0 - p_i)t_0 - (q-p_i)(t_{i-1}-t_0) = q_0t_0 - qt_0 + 2\log T t_0/\sqrt{t_{i-1}} - \sqrt{t_{i-1}}\log T < 0$. Again, we apply that when $x < 0$, $\Phi(x) \le \frac{1}{\sqrt{2\pi}|x|}\exp(-x^2/2)$. We have

$$\begin{aligned}
&\Phi\left(\frac{(q_0-p_i)t_0-(q-p_i)(t_{i-1}-t_0)}{\sigma\sqrt{t_{i-1}-t_0}}\right)\\
\le&\frac{\sigma\sqrt{t_{i-1}}}{\sqrt{2\pi}(\sqrt{t_{i-1}}\log T - 2\log T/\sqrt{t_{i-1}} - q_0t_0 + qt_0)}\exp\left(\frac{-(\sqrt{t_{i-1}}\log T - 2\log T/\sqrt{t_{i-1}} - q_0t_0+qt_0)^2}{2\sigma^2(t_{i-1}-t_0)}\right).
\end{aligned}$$

When T is large,

$$\frac{\sigma\sqrt{t_{i-1}}}{\sqrt{2\pi}(\sqrt{t_{i-1}}\log T - 2\log T/\sqrt{t_{i-1}} - q_0 t_0 + q t_0)} \leq 1$$
$$\frac{-(\sqrt{t_{i-1}}\log T - 2\log T/\sqrt{t_{i-1}} - q_0 t_0 + q t_0)^2}{2\sigma^2(t_{i-1} - t_0)} \leq -\log T.$$

Thus, we have proved that $\mathbb{P}(A_i) = O(\frac{1}{T})$. Since k is fixed, $qT\sum_{i=1}^{k}\mathbb{P}(A_i) = O(1)$. By combining these terms, we have:

$$R_T(\pi_{T,k}(t_1, t_2, \ldots, t_{k-1}; p_1, p_2, \ldots, p_k)) = O(T^{a_k}\log T).$$

Thus the theorem is proved. □

Similar to Theorem 2, next we will show that a regret of $O(T^{a_k}\log T)$ is the best one can do.

Theorem 4. *For any fixed price policy $\pi_{T,K}(t_1, t_2, \ldots, t_{K-1}; p_1, p_2, \ldots, p_K)$, we have*

$$R_T(\pi_{T,K}(t_1, t_2, \ldots, t_{K-1}; p_1, p_2, \ldots, p_K)) = \Omega(T^{a_K}).$$

The theorem shows that our policy in Theorem 3 achieves the best regret rate. The proof is completed by induction. Similar to the proof of Theorem 2, we call $(t_0, t_{k-1}]$ the first period and $(t_{k-1}, T]$ the second period. In Theorem 2 we prove the regret of the first period is $\Omega(t_1)$, linear in the switching time. This is because the seller can only use one price. While in the multiple price case, the regret of the first period is $\Omega(t_{k-1}^{a_{k-1}})$, because the seller can set $k-1$ prices in this period. This is the main cause of the improvement compared with a fixed two-price policy. The regret of the second period is still $\Omega(T/\sqrt{t_{k-1}})$, but since the regret for the first period decreases, the overall regret decreases.

Proof. We prove this theorem by induction. We already show that for $K = 2$, $R_T(\pi_{t_1,p_1,p_2}) = \Omega(T^{2/3})$. Suppose the result holds for $K = k-1$, we now consider $K = k$.

To simplify the discussion, we use $\pi_{T,k}$ to refer to a general fixed price policy with k prices. We have

$$R_T(\pi_{T,k}) = \underbrace{qt_{k-1} - J_{t_{k-1}}(\pi_{T,k})}_{S_1} + \underbrace{q(T - t_{k-1}) - p_k \mathbb{E}_{\tau_{k-1},\tau_k}[(\tau_k - \tau_{k-1})]}_{S_2}.$$

For S_1, by induction, we have

$$S_1 = qt_{k-1} - J_{t_{k-1}}(\pi_{T,k}) = \Omega(t_{k-1}^{a_{k-1}}).$$

Now consider the regret in the last period.

$$S_2 = q(T - t_{k-1}) - p_k \mathbb{E}_{\tau_{k-1},\tau_k}[(\tau_k - \tau_{k-1})] = \mathbb{E}_{\tau_{k-1}}[q(T - t_{k-1}) - p_k \mathbb{E}_{\tau_k}[(\tau_k - \tau_{k-1})|\tau_{k-1}]].$$

Similar to the proof of Theorem 2, now we consider three cases. Denote ϵ be a small positive number satisfying $0 < \epsilon < \min\{1 - a_k/a_{k-1}, a_{k-1} - 1/2\}$. If $t_{k-1} > T^{1-\epsilon}$, $p_k \le q + T^{-(1/2-\epsilon)}$,

$$S_1 = \Omega(t_{k-1}^{a_{k-1}}) = \Omega(T^{a_k}),\ S_2 \ge -T^{-(1/2-\epsilon)}(T - t_{k-1}) \ge -T^{1/2+\epsilon}.$$

Hence

$$S_1 + S_2 = \Omega(T^{a_k}).$$

From Lemma 2 we know, if $t_{k-1} > T^{1-\epsilon}$, $p_k > q + T^{-(1/2-\epsilon)}$,

$$S_2 = \mathbb{E}_{\tau_{k-1}}[q(T - t_{k-1}) - p_k \mathbb{E}_{\tau_k}[(\tau_k - \tau_{k-1})|\tau_{k-1}]] \ge \mathbb{E}_{\tau_{k-1}}[\Omega(T - \tau_{k-1})] = \Omega(T - t_{k-1}).$$

We have

$$S_1 + S_2 = \Omega(t_{k-1}^{a_{k-1}}) + \Omega(T - t_{k-1}) = \Omega(T^{a_{k-1}}).$$

If $t_1 \le T^{1-\epsilon}$,

$$S_2 = \mathbb{E}_{\tau_{k-1}}[q(T-t_{k-1}) - p_k \mathbb{E}_{\tau_k}[(\tau_k - \tau_{k-1})|\tau_{k-1}]] = \mathbb{E}_{\tau_{k-1}}\left[\Omega\left(\frac{T-t_{k-1}}{\sqrt{\tau_i - t_0}}\right)\right] = \Omega\left(\frac{T-t_{k-1}}{\sqrt{t_{k-1}-t_0}}\right).$$

Now we bound the regret,

$$\begin{aligned}
R_T(\pi_{T,k}) = S_1 + S_2 &\ge \Omega(t_{k-1}^{a_{k-1}}) + \Omega\left(\frac{T-t_{k-1}}{\sqrt{t_{k-1}-t_0}}\right) \\
&\ge c_1 t_{k-1}^{a_{k-1}} + \frac{2a_{k-1}c_2(T-t_{k-1})}{2a_{k-1}\sqrt{t_{k-1}-t_0}} \\
&\ge \left[\frac{c_1 c_2^{a_{k-1}}(T-t_{k-1})^{2a_{k-1}}}{2^{2a_{k-1}}}\right]^{1/(2a_{k-1}+1)} \\
&= \left[\frac{c_1 c_2^{a_{k-1}}}{2^{2a_{k-1}}}\right]^{1/(2a_{k-1}+1)} (T-t_{k-1})^{2a_{k-1}/(1+2a_{k-1})} \\
&= \Omega(T^{a_k}),
\end{aligned}$$

where the second inequality uses the fact that $x + 2a_{k-1}y \ge (xy^{2a_{k-1}})^{1/(2a_{k-1}+1)}$ for all $x, y \ge 0$, and that $a_{k-1} \le 1$. And the last inequality uses that $t_{k-1} \le T^{1-\epsilon}$. Thus the theorem is proved. □

Next we show some numerical results. In Figure 3.2, we choose $q = 5, q_0 = 5, t_0 = 1, \sigma = 2$. We plot the regret for T from 10000 to 200000 by increasing 10000 each time. It shows the regret - T curve for fixed price policy with 2, 3 and 4 prices. The regret increases sub-linearly. And the order of growth decreases when a policy uses more prices. Table 3.2 shows the regression coefficient of the model $\log \text{Regret} = \beta_0 + \beta_1 \log(\log T) + \beta_2 \log T$ when T increases from 10^4 to 10^7. The result is very close to our bound in Theorem 3.

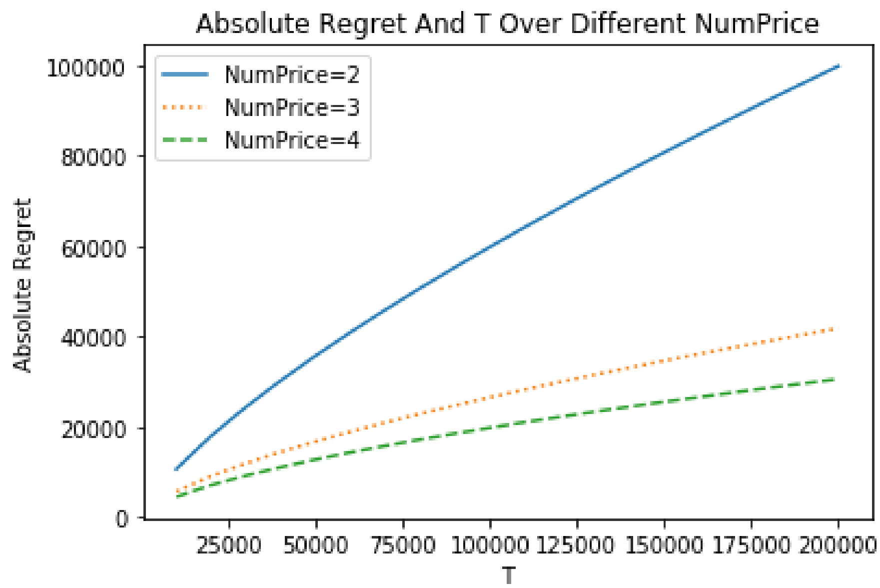

Figure 3.2: Regret for fixed price policy with 2, 3 and 4 prices

Number of prices	Theorem 3 value	β_2
2	$2/3 \approx 0.667$	0.671
3	$4/7 \approx 0.571$	0.570
4	$8/15 \approx 0.533$	0.526

Table 3.2: Coefficient Value

To summarize, in this session we extend our analysis to the fixed multi-price policy, where the seller can set multiple prices instead of two. We provide a fixed multi-price policy with a regret of $O(T^{a_k} \log T)$, where k is the number of prices the seller can use. And a_k is a decreasing sequence converging to $1/2$. Therefore as the number of prices increases, the regret converges to $O(\sqrt{T} \log T)$. We also prove that a regret of $\Omega((T^{a_k})$ is the best the seller can achieve for any fixed k-price policy.

3.3 Adaptive Price Case

In previous sections, we considered fixed-price policies in which the prices and switching times have to be determined beforehand. One may wonder if the performance of a policy can be improved by relaxing the fixed price assumption. In this section, we investigate into that direction by allowing the prices to be chosen adaptively based on the review history (but the switching time still needs to be fixed). We call such pricing policies partially adaptive policies. In the following, we study the optimal regret of such partially adaptive policies.

3.3.1 Partially Adaptive Two-Price Case

In a partially adaptive two-price policy, the seller still picks the first price p_1 and the switching time t_1. However, p_2 can be decided adaptively when the seller switches the price. Denote $p_2^*(x) = \arg\max_p p\mathbb{E}[(\tau_2 - \tau_1)|G_{\tau_1} = x]$, where the definitions of τ_1, τ_2, p_1 are the same as before. We define the revenue and the regret of a partially adaptive two-price policy π_{t_1,p_1},

$$J_T(\pi_{t_1,p_1}) = p_1\mathbb{E}[\tau_1 - t_0] + \mathbb{E}_{G_{\tau_1}}[p_2^*\mathbb{E}_{\tau_2}[\tau_2 - \tau_1|G_{\tau_1}]]$$

$$R_T(\pi_{t_1,p_1}) = q(T - t_0) - J_T(\pi_{t_1,p_1}).$$

Before stating the main result about the regret of the partially adaptive policy, we first introduce two lemmas.

The first lemma shows that the majority of the trajectories of G_t can be bounded by the curve $qs + \sqrt{s}\log T$.

Lemma 3. *The expected revenue of the trajectories in the set* $H = \{G_t | G_s > qs + \sqrt{s}\log T,\ \exists s \in [0,T]\}$ *is neglectable.*

$$\mathbb{E}[R \cdot 1(H)] = O(1).$$

To prove this, we split the sales horizon to many small intervals. Then we show that the probability G_t hits the curve at one interval is very small, therefore the expected revenue earned from those trajectories is also small. Summing the revenue over all the intervals, the value is still neglectable.

Now that we prove the expected revenue earned from trajectories that hit the $qs + \sqrt{s}\log T$ curve is neglectable, we can bound the revenue by the adaptive policy, similar to what we did in Lemma 2.

Lemma 4. *Denote* $G_t = t_0 q_0 + q(t - t_0) + \sigma\mathbb{B}_{t-t_0},\ t \geq t_0$. *Let* $\tau_{p,s} = \inf\{t : G_t < pt, t \geq s\} \wedge (T - s)$ *and* $S = q(T-s) - p^*\mathbb{E}_{\tau_{p^*,s}}[\tau_{p^*,s} - s]$, *where* $p^* = \operatorname{argmax} p\mathbb{E}_{\tau_{p,s}}[\tau_{p,s} - s | G_s]$. *If* $s \leq T^{1-\epsilon}$,

$$S = \Omega\left(\frac{T-s}{\sqrt{s}\log T}\right).$$

We prove the theorem by showing that given G_s, whatever price the seller uses, the regret is $T/\sqrt{s}$, similar to the case of the fixed price policy. To prove this, we consider three different cases based on the range of p. When $p > q$, we can prove the expected stopping time is too short so the total regret is $\Omega(T)$. When p is smaller than q but very close, we use Lemma 1 to prove the regret is $T/\sqrt{s}$. When p is small enough we can directly show the total regret is $T/\sqrt{s}$.

Comparing Lemma 4 with Lemma 2 we can see the extra flexibility in the adaptive policy does not yield higher revenue or a lower regret. Given additional information on the review at time s, the seller is still facing a regret of $\Omega(T/(\log T\sqrt{s}))$. The next theorem proves that a fixed two-price adaptive policy still has the $\Omega(T^{2/3})$ regret.

Theorem 5. *For any partially adaptive two-price policy* π_{t_1,p_1}*,* $R_T(\pi_{t_1,p_1}) = \Omega(T^{2/3}/\log T)$*.*

Before stating the main proof, we want to make some comments on the reason that the adaptive policy has the same order of regret as the regret of the fixed price policy. Consider the regret in the first period, when the seller decides the first price, the only information he has is the initial review level q_0 and t_0, as well as the switching time t_1. This is the same information he has when determining the first price and switching time in a fixed price policy. Therefore there is no advantage for the partially adaptive policy in the first period. In the second period, the additional information at G_{τ_1} does not help much. This is because from Lemma 3 the seller knows most of the trajectories of the cumulative review stay between $qs - \sqrt{s}\log T$ and $qs + \sqrt{s}\log T$. Therefore in a fixed price policy, although the seller does not know the exact review level when switching the price, he can still make a decision on the second price accordingly. On the other hand, even the review is high at τ_1, giving the seller some room to charge a higher price, he still needs to make sure the price satisfies $q - p = \Omega(\log Tt^{-1/2})$. Otherwise, the policy is very likely to stop halfway.

Proof. We have a regret formula

$$R_T(\pi_{t_1,p_1}) = \underbrace{q(t_1 - t_0) - p_1\mathbb{E}_\tau[\tau_1 - t_0]}_{S_1} + \underbrace{q(T - t_1) - \mathbb{E}_{G_{\tau_1}}[p_2^*\mathbb{E}_{\tau_2}[\tau_2 - \tau_1|G_{\tau_1}]]}_{S_2}.$$

We first bound S_1. We consider four possible cases.

1. If $p_1 \le 0$, then $p_1 \mathbb{E}_\tau[(t_1 - t_0) \wedge (\tau - t_0)] \le 0$, and $S_1 \ge q(t_1 - t_0)$.

2. If $p_1 \ge q_0$, then $p_1 \mathbb{E}_\tau[(t_1 - t_0) \wedge (\tau - t_0)] = 0$, and $S_1 = q(t_1 - t_0)$.

3. If $0 \le p_1 \le \min(q, q_0)$, then by Lemma 1,

$$p_1 \mathbb{E}_\tau[(t_1 - t_0) \wedge (\tau - t_0)] \le q(t_1 - t_0)[1 - \exp(-2q_0 q t_0/\sigma^2)] + \frac{4qq_0t_0}{\sigma}\sqrt{t_1 - t_0},$$

   and thus

$$S_1 \ge q(t_1 - t_0)\exp(-2q_0 q t_0/\sigma^2) - \frac{4qq_0t_0}{\sigma}\sqrt{t_1 - t_0}.$$

4. If $q < p_1 \le q_0$, then by Lemma 1,

$$p_1 \mathbb{E}_\tau[(t_1 - t_0) \wedge (\tau - t_0)] \le \frac{4p_1(q_0 - p_1)t_0}{\sigma}\sqrt{t_1 - t_0} \le \frac{q_0^2 t_0}{\sigma}\sqrt{t_1 - t_0},$$

   and

$$S_1 \ge q(t_1 - t_0) - \frac{q_0^2 t_0}{\sigma}\sqrt{t_1 - t_0}.$$

Therefore in all cases, $S_1 = \Omega(t_1)$.

Now consider the regret of S_2.

- $t_1 > T^{6/7}$

  In this case, $S_1 = \Omega(T^{6/7})$.

  Now consider the value of τ_1. If $\tau_1^{8/7} \log T \le T - t_1$, from Lemma 4, we have $S_2 = \Omega(\frac{T - t_1}{\sqrt{\tau_1}\log T})$. And

$$S = S_1 + S_2 = \Omega(T^{6/7}).$$

  If $\tau_1^{8/7} \log T > T - t_1$, we have $p_2 \le G_{\tau_1}/\tau_1 = q + \tau_1^{-1/2} \log T$.

$$S_2 \ge q(T - t_1) - p_2(T - t_1) \ge -\tau_1^{-1/2+8/7} \log T \ge -T^{9/14} \log T,$$

The last inequality is because $\tau_1 \leq T$. Since $9/14 < 2/3$, we also have

$$S = S_1 + S_2 = \Omega(T^{2/3}).$$

- $t_1 \leq T^{2/3}$

  In this case, from Lemma 4,

$$S_2 = \Omega(\frac{T - t_1}{\sqrt{t_1} \log T}).$$

  Combining with S_1 we have

$$\begin{aligned} R_T(\pi_{t_1,p_1}) &= S_1 + S_2 \\ &\geq \Omega(t_1) + \Omega\left(\frac{T - t_1}{\sqrt{t_1} \log T}\right) \\ &\geq c_1(t_1) + \frac{c_2(T - t_1)}{2\sqrt{t_1}} + \frac{c_2(T - t_1)}{2\sqrt{t_1}} \\ &\geq \left[\frac{c_1 c_2^2 (T - t_1)^2}{4}\right]^{1/3} \\ &= \Omega(T^{2/3}). \end{aligned}$$

  Thus the theorem is proved.

□

3.3.2 Partially Adaptive Multi-Price Case

In this section, we extend our previous result of the partially adaptive policy to multiple prices and show that a multi-price partially adaptive policy still has the same regret as a fixed multi-price policy.

Theorem 6. *For any partially adaptive multi-price policy* $\pi_{T,K}(t_1, t_2, \ldots, t_{K-1}; p_1)$*, we have*

$$R_T(\pi_{T,K}(t_1, t_2, \ldots, t_{K-1}; p_1)) = \Omega(T^{a_K}/\log T),$$

where a_k *is defined in Theorem 4.*

Proof. We prove this theorem by induction. We already show that for $K = 2$, $R_T(\pi_{t_1,p_1}) = \Omega(T^{2/3}/\log T)$. Suppose the result holds for $K = k - 1$, we now consider $K = k$.

To simplify the discussion, we use $\pi_{T,k}$ to refer to a general fixed price policy with k price.

$$R_T(\pi_{T,k}) = \underbrace{qt_{k-1} - J_{t_{k-1}}(\pi_{T,k})}_{S_1} + \underbrace{q(T - t_{k-1}) - p_k \mathbb{E}_{\tau_{k-1},\tau_k}[(\tau_k - \tau_{k-1})]}_{S_2}.$$

For S_1, by induction, we have

$$S_1 = qt_{k-1} - J_{t_{k-1}}(\pi_{T,k}) = \Omega(t_{k-1}^{a_{k-1}}).$$

Now consider the regret of S_2. Let $\epsilon = \min\{1 - a_k/a_{k-1}, 2a_k - 1\}$.

- $t_{k-1} > T^{1-\epsilon}$

  In this case, $S_1 = \Omega(T^{(1-\epsilon)a_{k-1}}) = \Omega(T^{a_k})$.

  Now consider the value of τ_{k-1}. If $\tau_{k-1}^{1+\epsilon} \log T \leq T - t_{k-1}$, from Lemma 4, we have $S_2 = \Omega(\frac{T - t_{k-1}}{\sqrt{\tau_{k-1}} \log T})$. And

$$S = S_1 + S_2 = \Omega(T^{a_k}).$$

If $\tau_{k-1}^{1+\epsilon} \log T > T - t_{k-1}$, we have $p_k \leq G_{\tau_{k-1}}/\tau_{k-1} = q + \tau_{k-1}^{-1/2} \log T$.

$$S_2 \geq q(T - t_{k-1}) - p_k(T - t_{k-1}) \geq -\tau_{k-1}^{1/2+\epsilon} \log T \geq -T^{1/2+\epsilon} \log T.$$

We also have

$$S = S_1 + S_2 = \Omega(T^{a_k}),$$

since by definition of ϵ, $a_k > 1/2 + \epsilon/2$.

- $t_{k-1} \leq T^{1-\epsilon}$

In this case, from Lemma 4,

$$S_2 = \Omega(\frac{T - t_{k-1}}{\sqrt{t_{k-1}} \log T}).$$

Combining with S_1 we have

$$\begin{aligned}
R_T(\pi_{T,k}) = S_1 + S_2 &\geq \Omega(t_{k-1}^{a_{k-1}}) + \Omega\left(\frac{T - t_{k-1}}{\sqrt{t_{k-1}} \log T}\right) \\
&\geq c_1 t_{k-1}^{a_{k-1}} + \frac{2a_{k-1} c_2 (T - t_{k-1})}{2a_{k-1} \sqrt{t_{k-1}} \log T} \\
&\geq \left[\frac{c_1 c_2^{a_{k-1}} (T - t_{k-1})^{2a_{k-1}}}{2^{2a_{k-1}} \log^{2a_k} T}\right]^{1/(2a_{k-1}+1)} \\
&\geq \left[\frac{c_1 c_2^{a_{k-1}}}{2^{2a_{k-1}}}\right]^{1/(2a_{k-1}+1)} (T - t_{k-1})^{2a_{k-1}/(1+2a_{k-1})} \log^{-1} T \\
&= \Omega(T^{a_k} \log^{-1} T),
\end{aligned}$$

where the second inequality uses the fact that $x + 2a_{k-1} y \geq (x y^{2a_{k-1}})^{1/(2a_{k-1}+1)}$ for all $x, y \geq 0$, and that $a_{k-1} \leq 1$. And the last inequality uses the fact that $t_{k-1} \leq T^{1-\epsilon}$. Thus the theorem is proved.

□

Theorem 6 states that, like in the two-price case, allowing adaptivity in multi-price case does not help with the asymptotic regret either. In other words, it is enough to consider a policy with fixed switching time and price to achieve asymptotic optimal regret.

3.3.3 Fully Adaptive Two-Price Case

In the previous sessions, we consider the partially adaptive policy and show that the policy has the same order of regret as the regret of the fixed price policy. A question one may ask is, is there a better policy that can improve the regret? In this section, we consider a policy with the even more flexibility, the fully adaptive two-price policy. In a fully adaptive two-price policy, the seller only decides the first price p_1. During the entire sales horizon, the seller can pick any time to make a price switch. We use π_{p_1} to denote a fully adaptive two-price policy. The revenue and regret are defined by

$$\begin{aligned} J_T(\tau_{p_1}) &= \mathbb{E}[q\tau_1 - p_1\tau_1] + \mathbb{E}[q(T-\tau_1) - p_2^*(\tau_2 - \tau_1)] \\ R_T(\tau_{p_1}) &= q(T - t_0) - J_T(\tau_{p_1}), \end{aligned}$$

where τ_1 and τ_2 are the stopping times for p_1 and p_2, determined by the seller's policy and the performance of G_t. We will show that the fully adaptive two-price policy also has the same order of regret as the regret of the fixed two-price policy.

Theorem 7. *For any fully adaptive two-price policy π_{p_1}, $R_T(\pi_{p_1}) = \Omega(T^{2/3}/\log T)$.*

We first discuss the intuition why the fully adaptive two-price policy still has the same order of regret. In the previous session, we argue that the review process G_t is very likely between $qt - \sqrt{t}\log T$ and $qt + \sqrt{t}\log T$ at time t, therefore, the review q_t is between $q - t^{-1/2}\log T$ and $q + t^{-1/2}\log T$ at time t. As t grows larger, q_t converges to q. Because of this convergence, the additional flexibility does not help much. Even though the seller could switch to any price at any time, he still needs to balance the long term effect, that when q_t converges to q, $p_2 \leq q_t$.

Proof. Denote τ the stopping time for the first price.

- $p_1 \leq q - \log^{-1} T, \tau \geq T^{3/4}$

  In this case,

  $$S_1 = (q - p_1)\tau \geq T^{3/4}/\log T.$$

  We also have

  $$S_2 \geq (q - p_2)(T - \tau) = -T^{5/8} \log T.$$

  $S_1 + S_2 = \Omega(T^{3/4}/\log T)$.

- $p_1 \leq q - \log^{-1} T, \tau < T^{3/4}$

  In this case, the regret for the first period

  $$S_1 = (q - p_1)\tau \geq \tau \log^{-1} T.$$

  From Lemma 4, $S_2 = \Omega((T - \tau)/(\sqrt{\tau} \log T))$. Therefore we have

  $$S_1 + S_2 = \tau \log^{-1} T + c\frac{T - \tau}{\sqrt{\tau} \log T} = \Omega\left(\frac{T^{2/3}}{\log T}\right).$$

- $q - \log^{-1} T < p_1 \leq q$

Since $p_1 \leq q$, we have $S_1 \geq 0$. From Lemma 7, we have

$$
\begin{aligned}
&\mathbb{P}(\tau > \sqrt{T}) \\
=&\Phi\left(\frac{(q_0-p_1)t_0}{\sigma T^{1/4}} - \frac{q-p_1}{\sigma}T^{1/4}\right) - \exp\left(-\frac{2(q_0-p_1)(q-p_1)t_0}{\sigma^2}\right)\Phi\left(-\frac{(q_0-p_1)t_0}{\sigma T^{1/4}} + \frac{q-p_1}{\sigma}T^{1/4}\right). \\
&\leq 1 - \exp\left(-\frac{2(q_0-p_1)(q-p_1)t_0}{\sigma^2}\right) + \frac{2(q_0-p_1)t_0}{T^{1/4}\sigma} \\
&\leq 1 - \exp\left(-\frac{2q_0t_0}{\sigma^2 \log T}\right) + \frac{2q_0t_0}{T^{1/4}\sigma} \\
&= O(\log^{-1} T),
\end{aligned}
$$

where the last equality is because $1 - \exp(-x) = O(x)$ when x goes 0. Similarly, we have

$$
\begin{aligned}
&\mathbb{P}(\tau > T^{3/4}) \\
=&\Phi\left(\frac{(q_0-p_1)t_0}{\sigma T^{3/8}} - \frac{q-p_1}{\sigma}T^{3/8}\right) - \exp\left(-\frac{2(q_0-p_1)(q-p_1)t_0}{\sigma^2}\right)\Phi\left(-\frac{(q_0-p_1)t_0}{\sigma T^{3/8}} + \frac{q-p_1}{\sigma}T^{3/8}\right). \\
&\leq 1 - \exp\left(-\frac{2(q_0-p_1)(q-p_1)t_0}{\sigma^2}\right) + \frac{2(q_0-p_1)t_0}{T^{3/8}\sigma} \\
&\leq 1 - \exp\left(-\frac{2q_0t_0}{\sigma^2 \log T}\right) + \frac{2q_0t_0}{T^{3/8}\sigma} \\
&= O(\log^{-1} T).
\end{aligned}
$$

$$
\begin{aligned}
S &= S_1 + S_2 \\
&= \mathbb{E}[(q-p_1)\tau + (q-p_2)(T-\tau)] \\
&\geq \mathbb{P}(\tau \leq \sqrt{T})[0 + \Omega(T/(T^{1/4}\log T))] + \mathbb{P}(T^{3/4} >= \tau > \sqrt{T})[0 + \Omega(T/(T^{3/8}\log T))] \\
&+ \mathbb{P}(T^{3/4} < \tau)[0 + \tau^{-1/2}\log T(T-\tau)] \\
&\geq (1 - c_1 \log^{-1} T)[c_2(T^{3/4}/\log T)] + c_1 \log^{-1} T[T^{5/8}\log T] \\
&= \Omega(T^{3/4}/\log T).
\end{aligned}
$$

- $p_1 > q$

From Lemma 7, we have

$$
\begin{aligned}
\mathbb{P}(\tau > \sqrt{T}) &\leq \frac{2(q_0 - p_1)t_0}{T^{1/4}\sigma} \\
&\leq \frac{2q_0 t_0}{T^{1/4}\sigma} \\
&= O(T^{-1/4})
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{P}(\tau > T^{3/4}) &\leq \frac{2(q_0 - p_1)t_0}{T^{3/8}\sigma} \\
&\leq \frac{2q_0 t_0}{T^{3/8}\sigma} \\
&= O(T^{-3/8}).
\end{aligned}
$$

Therefore we have

$$
\begin{aligned}
S &= S_1 + S_2 \\
&= \mathbb{E}[(q-p_1)\tau + (q-p_2)(T-\tau)] \\
&\geq \mathbb{P}(\tau \leq \sqrt{T})[-(q-q_0)\tau + \Omega(T/(T^{1/4}\log T))] + \mathbb{P}(T^{3/4} < \tau)[-(q-q_0)\tau + \tau^{-1/2}\log T(T-\tau)] \\
&+ \mathbb{P}(T^{3/4} >= \tau > \sqrt{T})[-(q-q_0)\tau + \Omega(T/(T^{3/8}\log T))] \\
&\geq (1 - c_2 T^{-3/8})[-(q-q_0)\sqrt{T} + \Omega(T^{3/4}/\log T))] + c_2 T^{-3/8}[-(q-q_0)T^{3/4} + \Omega(T^{5/8}/\log T))] \\
&+ c_2 T^{-3/8}[-(q-q_0)T - T^{5/8}\log T] \\
&= -(q-q_0)\sqrt{T} + c_3 T^{3/4}/\log T + c_2(q-q_0)T^{1/8} - c_2 c_3 T^{3/8}/\log T - c_2(q-q_0)T^{3/8} \\
&+ c_2 c_3 T^{1/4}/\log T - (q-q_0)c_2 T^{5/8} - c_2 T^{1/4}\log T \\
&= \Omega(T^{3/4}/\log T)
\end{aligned}
$$

where c_2 and c_3 are constants.

Combining all cases, we have that the regret is $\Omega(T^{2/3}/\log T)$.

□

In this chapter, we introduce the basic model setting and discuss four types of policies. For the fixed price policy, we provide some sample policies and prove their regret. We also prove this is the best among fixed price policies. We then discuss some adaptive policies and prove they have the same order of regret as the regret of the fixed price policy.

Chapter 4

Discrete Model

In the last chapter, we discuss different policies and prove their regret bounds. However, these results are all based on the assmuption that the review follows a normal distribution. While in practice, most review systems require customers input discretized rating. One may wonder if the underlying distribution changes, will the results remain the same? In this section, we will address this question by considering a discrete model where the customer review is a general distribution with expectation q and standard devitation σ. We also require the distribution is bounded between $[\underline{q}, \bar{q}]$. Since q is now a number between $\underline{q}$ and $\bar{q}$, we also assume the prior knowledge q_0 is between $\underline{q}$ and $\bar{q}$. Therefore we assume any price the seller could charge should also be in the same range.

Consider a customer comes at $t_0 + 1, t_0 + 2, \ldots, T$. Let $G_t = q_0 t_0 + \sum_{i=t_0+1}^{t} X_i$, where X_i is the review given by the ith customer. Similar to the fixed two-price policy, we use $\pi^D_{t_1,p_1,p_2}$ to denote a fixed two-price policy in discrete time setting. The stopping time also has a similar definition, $\tau_1 = \inf\{t : G_t < p_1 t, t \geq t_0, t \in Z\}$, $\tau_2 = \inf\{t : G_t < p_2 t, t \geq \tau_1, t \in Z\}$.

The revenue of a discrete policy is

$$J_T(\pi^D_{t_1,p_1,p_2}) = \mathbb{E}_\tau[p_1(\tau_1 - t_0) + p_2(\tau_2 - \tau_1)].$$

The regret is

$$R_T(\pi^D_{t_1,p_1,p_2}) = q(T - t_0) - J_T(\pi^D_{t_1,p_1,p_2}).$$

In the following, we show that we can use the same fixed two-price policy with a similar order of regret.

Theorem 8. *Let* $t_1 = \lfloor T^{3/2} \rfloor$, $p_1 = \underline{q}$, $p_2 = q - \log T/T^{1/3}$. $R_T(\pi^D_{t_1,p_1,p_2}) = O(T^{2/3}\log T)$.

Based on the law of large number, as the number of review grows, the average review rating q_t converges to q. Therefore we can use a policy similar to Theorem 1. We first use a low price to make sure the sales continue and at some time we switch to a high price.

The proof uses Hoeffding's inequality to bound the probability that q_t devitates from q.

Proof. In the following, we denote $A_1 = \{\tau_1 < t_1\}$ and $A_2 = \{\tau_2 < \tau_1 + T - t_1\}$. Therefore $A_1 \cup A_2 = \{\tau_2 < T\}$. And we have:

$$\begin{aligned} J_T(\pi^D_{t_1,p_1,p_2}) &\geq \underline{q}t_1 + p_2(T - t_1) \cdot \mathbb{P}(\tau_2 = T) \\ &= \underline{q}t_1 + p_2(T - t_1) \cdot (1 - \mathbb{P}(A_1 \cup A_2)) \\ \underline{q}t_1 + &\geq p_2(T - t_1) \cdot (1 - \mathbb{P}(A_1) - \mathbb{P}(A_2)). \end{aligned}$$

Since $p_1 = 0$ and $G_t \geq 0$, we have $\mathbb{P}(A_1) = 0$. For $\mathbb{P}(A_2)$,

$$\begin{aligned}
\mathbb{P}(A_2) &\leq \sum_{t=t_1+1}^{T} \mathbb{P}(G_t \leq p_2 t) \\
&= \sum_{t=t_1+1}^{T} \mathbb{P}\left(\sum_{t_1+1}^{t} X_i + q_0 t_0 \leq p_2 t\right) \\
&= \sum_{t=t_1+1}^{T} \mathbb{P}\left(\sum_{t_1+1}^{t} X_i/(t-t_0) \leq (p_2 t - q_0 t_0)/(t-t_0)\right) \\
&= \sum_{t=t_1+1}^{T} \mathbb{P}\left(\sum_{t_1+1}^{t} X_i/(t-t_0) \leq p_2 - (q_0 - p_2)t_0/(t-t_0)\right) \\
&\leq \sum_{t=t_1+1}^{T} \mathbb{P}\left(\sum_{t_1+1}^{t} X_i/(t-t_0) \leq q - T^{-1/3}\log T + \frac{(\underline{q}-\bar{q})t_0}{t-t_0}\right).
\end{aligned}$$

The last inequality comes from the fact that both p_2 and q_0 are between $\underline{q}$ and $\bar{q}$.

From Hoeffding's Inequality we have

$$\begin{aligned}
&\mathbb{P}\left(\frac{\sum_{t_1+1}^{t} X_i}{(t-t_0)} \leq q - T^{-1/3}\log T + \frac{t_0}{t-t_0}\right) \\
&\leq \exp\left(\frac{-2(t-t_0)}{(\bar{q}-\underline{q})^2}\left(T^{-1/3}\log T - \frac{t_0}{t-t_0}\right)^2\right) \\
&= \exp\left(\frac{-2(t-t_0)}{(\bar{q}-\underline{q})^2}\left(T^{-2/3}\log^2 T + \frac{t_0^2}{(t-t_0)^2} - 2T^{-1/3}\log T\frac{t_0}{(t-t_0)}\right)\right) \\
&\leq \exp\left(\frac{-2\log^2 T - \frac{2t_0^2}{(t-t_0)} + 4t_0 T^{-1/3}\log T}{(\bar{q}-\underline{q})^2}\right),
\end{aligned}$$

the last inequality is because $t \geq \lfloor T^{2/3} \rfloor$. The dominating term above is $\exp(-2\log^2 T/(\bar{q}-\underline{q})) = T^{-2\log T/(\bar{q}-\underline{q})}$. Therefore,

$$\mathbb{P}(A_2) \leq T \cdot T^{-2\log T/(\bar{q}-\underline{q})} = T^{1-2\log T/(\bar{q}-\underline{q})}.$$

Now we bound the regret,

$$
\begin{aligned}
R_T(\pi_{t_1,p_1,p_2}) &= q(T-t_0) - J_T(\pi_{t_1,p_1,p_2}) \\
&\leq q(T-t_0) - [p_1(t_1-t_0) + p_2(T-t_1)(1-\mathbb{P}(A_1)-\mathbb{P}(A_2))] \\
&= (q-p_2)(T-t_1) + (q-p_1)(t_1-t_0) + p_2(T-t_1)\mathbb{P}(A_2) - qt_0 \\
&\leq (q-p_2)(T-t_1) + (q-\underline{q})(t_1-t_0) + p_2 T\mathbb{P}(A_2) \\
&\leq T^{2/3}\log T + q\left\lfloor T^{2/3}\right\rfloor + qT^{2-2\log T/(\bar{q}-\underline{q})}.
\end{aligned}
$$

We have $R_T(\pi_{t_1,p_1,p_2}) = O(T^{2/3}\log T)$.

□

Next we show that for any partially adaptive two-price policy this is the best one can do. Before proving the main theorem, we first show two lemmas.

The first lemma bounds the possible review level.

Lemma 5. *The expected revenue of the trajectories in the set* $H^D = \{G_t | G_s > qs + \sqrt{s}\log T,\ \exists s \in [0,T]\}$ *is neglectable.*

$$\mathbb{E}[R^D \cdot 1(H^D)] = O(1).$$

Proof. Since the price p is bounded by the upper limit of the review distribution $\bar{q}$. We only need to show that the probability that any trajectories cross the $qs + \sqrt{s}\log T$ boundary is $O(1)$.

At any point s, the probability that G_s cross the boundary is

$$\begin{aligned}\mathbb{P}(G_s > qs + \sqrt{s}\log T) &= \mathbb{P}\left(\sum_{t_0+1}^{s} X_i + q_0 t_0 \geq qs + \sqrt{s}\log T\right) \\ &= \mathbb{P}\left(\sum_{t_0+1}^{s} (X_i - q)/(s-t_0) \geq (qt_0 - q_0 t_0 + \sqrt{s}\log T)/(s-t_0)\right).\end{aligned}$$

From Hoeffding's inequality we have

$$\begin{aligned}\mathbb{P}(G_s > qs + \sqrt{s}\log T) &\leq \exp\left(-\frac{2(s-t_0)^2[(qt_0 - q_0t_0 + \sqrt{s}\log T)/(s-t_0)]^2}{(s-t_0)(\bar{q}-\underline{q})^2}\right) \\ &= \exp\left(-\frac{2(qt_0 - q_0t_0 + \sqrt{s}\log T)^2}{(s-t_0)(\bar{q}-\underline{q})^2}\right) \\ &= \exp\left(-\frac{2(qt_0 - q_0t_0)^2 + 4(qt_0 - q_0t_0)\sqrt{s}\log T + 2s\log^2 T}{(s-t_0)(\bar{q}-\underline{q})^2}\right) \\ &= O(T^{-2\log T/(\bar{q}-\underline{q})^2}),\end{aligned}$$

the last equality is because that the dominating term is $2s\log^2 T/[(s-t_0)(\bar{q}-\underline{q})^2]$. Now we sum over s

$$\sum_s \mathbb{P}(G_s > qs + \sqrt{s}\log T) = O(T^{1-2\log T/(\bar{q}-\underline{q})^2}) = O(1).$$

□

In the following discussion, we assume $q_s \leq q + \log T/\sqrt{s}$. The second lemma bounds the regret of the last price.

Lemma 6. *Denote $G_t = G_t = q_0 t_0 + \sum_{i=t_0+1}^{t} X_i$, $t > t_0$. X_i is an i.i.d. random variable with range in $[\underline{q}, \bar{q}]$. Its expectation is q and its variance is σ. Let $\tau_{p,s} = \inf\{t : G_t < pt, t \geq s\} \wedge (T-s)$ and $S = q(T-s) - p^*\mathbb{E}_{\tau_{p^*,s}}[\tau_{p^*,s} - s]$, where $p^* =$*

$\operatorname{argmax} p\mathbb{E}_{\tau_{p,s}}[\tau_{p,s} - s|G_s]$. *If* $\exists\epsilon,\ s \le T^{1-\epsilon}$,

$$S = \Omega\left(\frac{T-s}{\sqrt{s}\log T}\right).$$

We have the following theorem.

Theorem 9. *For any partially adaptive two-price* $\pi_T^D(t_1;p_1)$*, we have*

$$R_T(\pi_T^D(t_1;p_1)) = \Omega(T^{2/3}/\log T).$$

Proof. The revenue of a partially adaptive two-price $\pi_T^D(t_1;p_1)$ is

$$J_T(\pi_T^D(t_1;p_1)) = \mathbb{E}_{\tau_1,\tau_2}[p_1(\tau_1 - t_0) + \max_{p_2} p_2(\tau_2 - \tau_1)].$$

Therefore,

$$\begin{aligned} R_T(\pi_T^D(t_1;p_1)) =& q(T - t_0) - J_T(\pi_T^D(t_1;p_1)) \\ \ge& \underbrace{q(t_1 - t_0) - p_1\mathbb{E}_{\tau_1}[\tau_1 - t_0]}_{S_1} + \underbrace{q(T - t_1) - \mathbb{E}_{\tau_1,\tau_2}[p_2^*(\tau_2 - \tau_1)]}_{S_2}. \end{aligned}$$

We consider several cases

- $t_1 > T^{2/3}$ From Lemma 6, we have

$$S_1 = \Omega\left(\frac{t_1 - t_0}{\sqrt{t_0}\log T}\right) = \Omega(t_1/\log T).$$

If $\tau_1 > T^{2/3}$, from Lemma 5 we have $q_{\tau_1} \le q + \log T/\tau_1$. We can bound S_2,

$$S_2 \ge q(T - t_1) - q_{\tau_1}(T - t_1) \ge -\log T \cdot (T - t_1)/\tau_1 \ge -\log T \cdot T^{1/3}.$$

Combining S_1 and S_2, we have

$$S_1 + S_2 = \Omega(t_1/\log T) - T^{1/3}\log T = \Omega(T^{2/3}/\log T).$$

Next we discuss the range of τ_1.

If $\tau_1 \leq T^{2/3}$, and $\tau^{1.1}\log T \leq T - t_1$, we can apply Lemma 6 to bound S_2. We have

$$S_2 = \Omega\left(\frac{T - t_1}{\sqrt{\tau_1 - t_0}\log T}\right) = \Omega\left(\tau_1^{0.6}\right) \geq 0.$$

Combining S_1 and S_2 we have

$$S_1 + S_2 \geq \Omega(t_1/\log T) = \Omega(T^{2/3}/\log T).$$

If $\tau_1 \leq T^{2/3}$, and $\tau^{1.1}\log T > T - t_1$, we have

$$S_2 \geq q(T - t_1) - p(T - t_1) \geq (q - (q + \log T/\sqrt{\tau_1}))(T - t_1) \geq -T^{2/5}\log^3 T.$$

Combining S_1 and S_2 we have

$$S_1 + S_2 = \Omega(t_1/\log T) - T^{2/5}\log^3 T = \Omega(T^{2/3}/\log T).$$

- $t_1 \leq T^{2/3}$ We first apply Lemma 6 to bound S_1, we have

$$S_1 = \Omega\left(\frac{t_1 - t_0}{\sqrt{t_0}\log t_1}\right) = \Omega\left(t_1/\log t_1\right).$$

Then we use Lemma 4 again to bound S_2, we have

$$S_2 = \Omega\left(\frac{T - t_1}{\sqrt{t_1}\log T}\right).$$

Combining S_1 and S_2, we have

$$S_1 + S_2 = \Omega\left(t_1/\log t_1\right) + \Omega\left(\frac{T - t_1}{\sqrt{t_1}\log T}\right) = \Omega(T^{2/3}/\log T).$$

$\square$

To summarize, in this chapter, we consider a discrete model and a general reivew distribution. We prove that under some mild assmuptions fixed two-price policy still has $O(T^{2/3}\log T)$ regret. We also prove that any partially adaptive two-price policy has at least $\Omega(T^{2/3}/\log T)$ regret. Since the result does not require a specific review distribution, it also answers the question we address in the introduction, that the review system - no matter 'like' or 'dislike', or 5 stars, or 10 stars - does no have any impact on the regret asymptotically.

Chapter 5

Numerical Experiments

In this section, we conduct numerical experiments to test the performance of the fixed price policy and the partially adaptive policy. By these numerical experiments, we aim to address the following two questions in these two policies

- How each parameter (q, q_0, t_0, σ) influences the performance of the regret?
- How much benefit one can get from additional price changes?

Furthermore, we would like to test the performance of the fixed price policy in the discrete model setting.

5.1 Fixed Price Policy

In this section, we test the performance of the fixed price policy proposed in Theorem 1 and Theorem 3. We start by introducing the setup of the experiments. In our experiments, we consider a problem with $q = 5$, $q_0 = 5$, $\sigma = 2$, $t_0 = 1$ as the basic model setting. And we let T grow from 10000 to 200000 by 10000 each time to draw the regret-T curve.

Next, we discuss how we conduct the simulation. We simulate sample paths of sales and reviews given any policy. We discretize the time interval into $t_0+1, t_0+2, \ldots, T$. In each time period, if the current review is higher than the current price, then the customer will make a purchase. Otherwise there is no purchase at that time period. When a customer purchases the product, the product also gets a new review generated from a normal distribution with expectation q and standard deviation σ and the average review updates accordingly.

In each of our simulated result, we run 10000 sample paths using the above method and use the average revenue as the expected revenue the seller can earn for the corresponding policy.

We first verify the asymptotic behavior of the regret. We plot the regret-T curve for fixed price policy with 2, 3 and 4 prices. Figure 5.1 shows that as the number of prices increases, the regret decreases. In the right figure of Figure 5.1, we also show the relative regret, which is defined to be the regret divided by $q(T-t_0)$. As T increases, we can see that the relative regret decreases. Furthermore, one may notice that as the number of prices increases, the marginal benefit of having one more price to change decreases. As in the figure, the difference between the fixed two-price policy and the three-price policy is larger than the difference between the three-price policy and the four-price policy. This observation is consistent with the theoretical results for the multi-price policies. Recall that by Theorem 3 a k-price policy has a regret of $\Omega(T^{a_k})$, where $a_k = \frac{1}{2} + \frac{1}{2(2^k-1)}$. As $k \to \infty$, the difference between a_k and a_{k-1} decreases.

Next, we discuss the effect of q on the regret for the fixed price policy. Recall q is the expectation of the review distribution. Figure 5.2 shows the regret-T curves of the fixed two-price policy with q changing from 5 to 10 . It shows that increasing q will increase the absolute regret, while decreasing the relative regret. The increase in the absolute regret is because the regret is multiplied by a factor. And the decrease in the

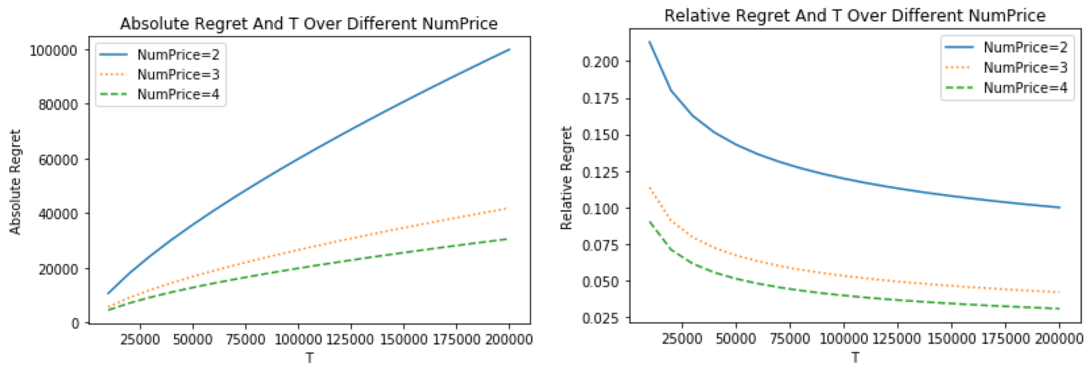

Figure 5.1: Fixed Price Policy, Regret-Time for different number of price used

relative regret is because of the fact that most of the regret is made in the first period where $p_1 = -\log T$. And increasing q decreases the relative regret for the first period .

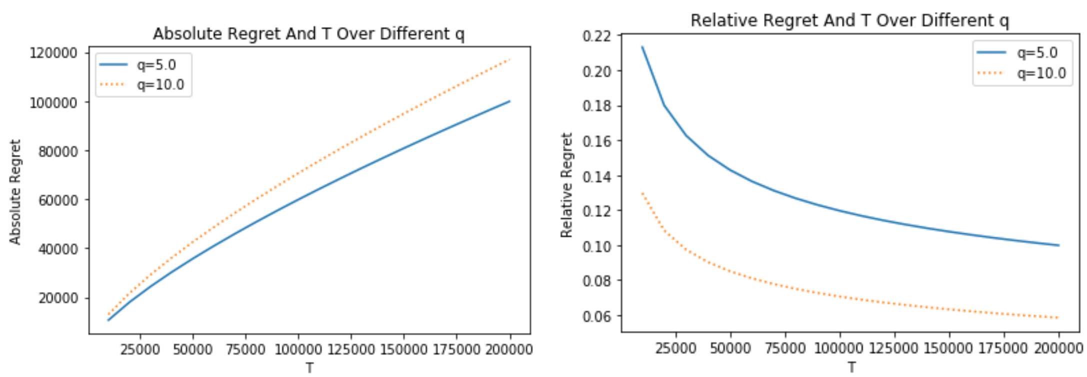

Figure 5.2: Fixed Price Policy, Regret-Time For Different q

Next we discuss the impact of t_0 on the regret of fixed price policy. Recall t_0 is the initial number of reviews (with average review q_0). Figure 5.3 shows the regret-T curves of the fixed two-price policy with t_0 changing from 1 to 100. It shows that as t_0 increases, the regret decreases. This is mainly because for a fixed T increasing t_0 reduces the effective selling horizon, therefore reduce the total regret.

Next we consider the effect of σ on the regret of fixed price policy. Recall σ is the standard deviation for the customers review distribution. Figure 5.4 shows the regret-T

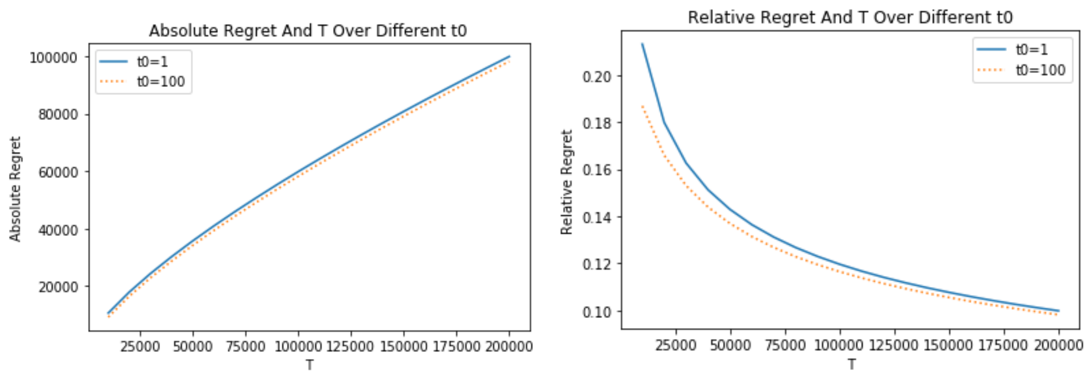

Figure 5.3: Fixed Price Policy, Regret-Time For Different t_0

curves of the fixed two-price with σ increasing from 2 to 5. As shown in Figure 5.4, when σ increases, the regret increases. This is because when σ increases, the review distribution has a fatter tail and the review process has a higher probability falling below pt. Although σ does not affect the asymptotic results as shown in Theorem 3, it actually has a significant impact on the regret.

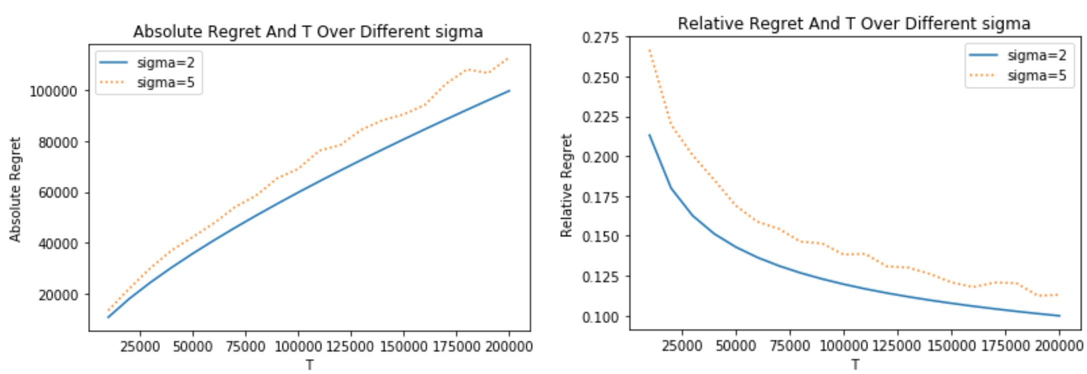

Figure 5.4: Fixed Price Policy, Regret-Time for Different σ

5.2 Partially Adaptive Policy

In this section we show numerical test results of the partially adaptive policy. We first illustrate how we implement the experiments.

We use backward induction to solve a dynamic programming. First we discretize the time interval into $t_0 + 1, t_0 + 2, \ldots, T$. Denote $V_t(r_i|p)$ the expected revenue the seller can earn for the rest of the time when the current review is r_i, the current price is p and the switching times are proposed in (2.3). Note that $V_t(r_i|p)$ is not the optimal partially adaptive policy since the switching times are not optimal. When t is not a switching time, we have if $r_i \geq p$,

$$V_t(r_i|p) = p + \sum_j p_{ij} V_{t+1}(r_j|p),$$

where p_{ij} is the probability the review becomes to r_j at time $t+1$. If $r_i < p$, we have

$$V_t(r_i|p) = V_t(r_{i+1}|p).$$

When t is a switching time, we have

$$V_t(r_i|p) = \sum_j p_{ij} \max_{p'} V_{t+1}(r_j|p').$$

And the boundary condition is

$$V_T(r_i|p) = 0, \forall r_i, p.$$

In the following experiments, without specification, the default parameters are: $q = 5$, $q_0 = 5$, $t_0 = 1$, $\sigma = 2$. T is increased from 10000 to 200000 with an increment of 10000 each time.

We first verify the regret growth for the partially adaptive policy. Figure 5.5 shows the performance of the partially adaptive policy with a different number of prices. Similar to the fixed price policy, as the number of price used increases, the regret decreases. Also the relative regret decreases to 0, which means the partially adaptive policy has a sublinear regret.

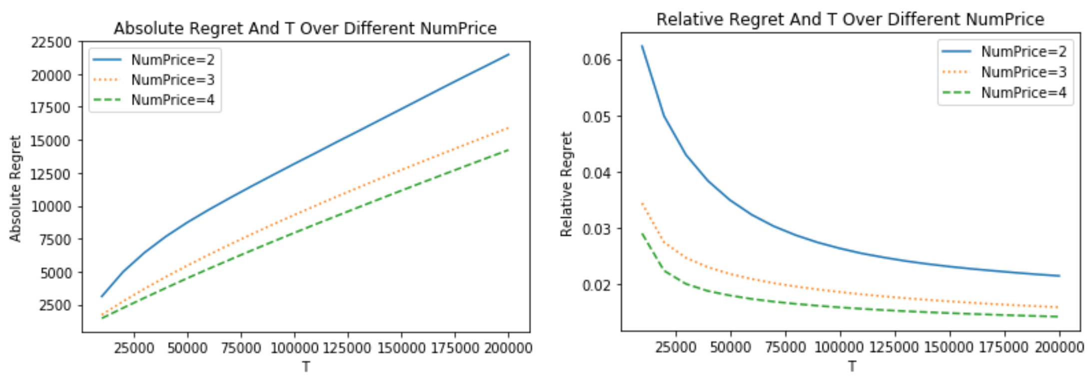

Figure 5.5: Adaptive Policy, Regret-Time For Different Number Of Price Used

Figure 5.6 shows the impact of q on adaptive policy. As q increases, the absolute regret increases while the relative regret drops. Similar to the fixed price case, increasing q makes the relative regret in the first period smaller, which leads to the overall smaller regret through the horizon

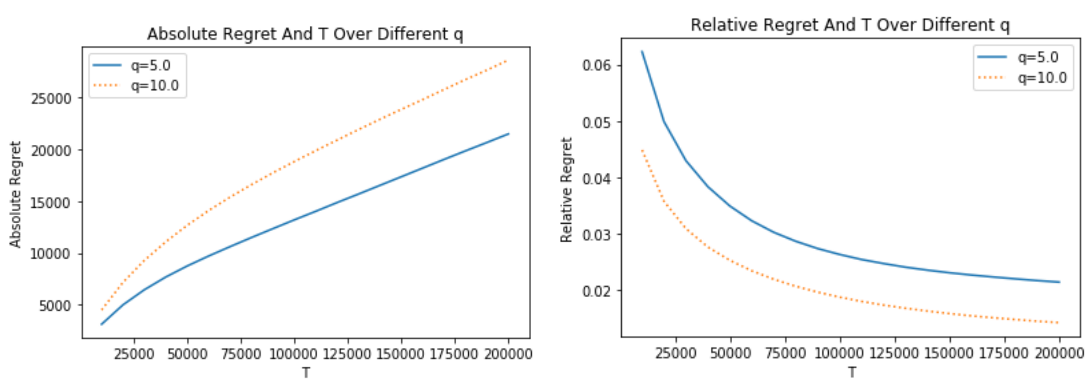

Figure 5.6: Partially Adaptive Policy, Regret-Time, For Different q

The next plot Figure 5.7 shows the impact of q_0 on the regret. We find that increasing q_0 could decrease the regret of partially adaptive policy. This is mainly because the partially adaptive policy could find the optimal price for different review level, which could take advantage of a higher initial review q_0.

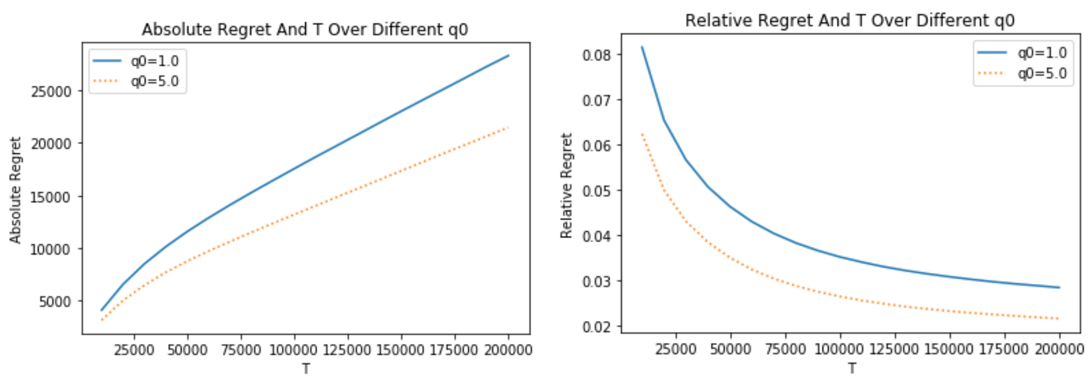

Figure 5.7: Partially Adaptive Policy, Regret-Time, For Different q_0

The next plot Figure 5.8 shows increasing t_0 could reduce the regret. Compared with the case in fixed price policy, we can see that the adaptive policy takes more advantage of increasing t_0. This is because in the partially adaptive policy, the price is optimized based on the parameters. Therefore increasing t_0 can make the seller price more aggressively.

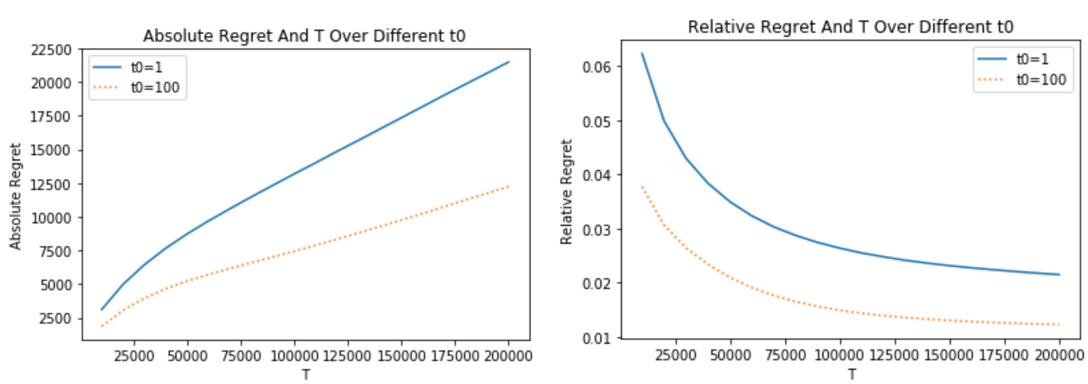

Figure 5.8: Partially Adaptive Policy, Regret-Time, For Different t_0

The next plot Figure 5.9 shows the impact of σ. To one's surprise, increasing σ will cause a huge increase in the regret for the partially adaptive policy compared with the fixed adaptive policy. This result can be viewed similarly as Figure 5.8, that decreasing σ can decrease the regret for the adaptive policy because the adaptive policy can optimize the price based on the σ while fixed price policy is very conservative and does not take full advantage when σ is small. Therefore increasing σ will have a bigger impact on the partially adaptive policy.

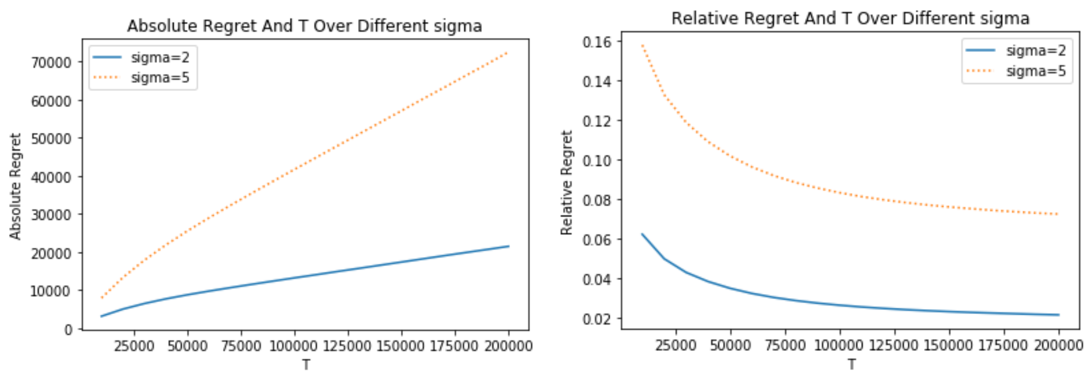

Figure 5.9: Partially Adaptive Policy, Regret-Time, For Different σ

Figure 5.10 shows the second period optimal price used in the partially adaptive policy. The optimal price increases as the review level increases.

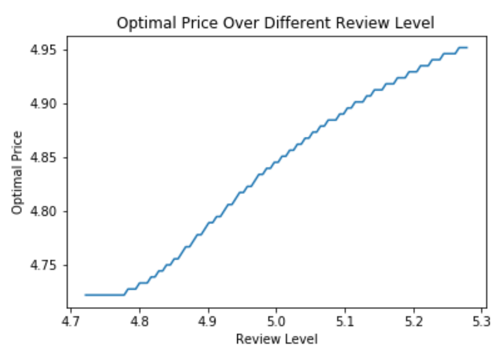

Figure 5.10: Optimal Price Over Different Review Level

5.3 Comparison with fully adaptive policy

Before moving to the discrete model section, we compare the performance of fixed two-price policy, partially adaptive two-price policy and the fully adaptive two-price policy.

First we describe how we conduct the experiment. For the fully adaptive policy, we use backward induction to calculate the expected revenue. Discretize the time interval into $t_0+1, t_0+2, \ldots, T$. We first calculate the revenue earned in the second period. Denote $V_t'(r_i|p)$ the expected revenue the seller can earn for the rest of the time when the current review is r_i, the current price is p and the seller cannot switch to another price. If $r_i \geq p$,

$$V_t'(r_i|p) = p + \sum_j p_{ij} V_{t+1}'(r_j|p),$$

where p_{ij} is the probability the review becomes to r_j at time $t+1$. If $r_i < p$,

$$V_t'(r_i|p) = V_{t+1}'(r_i|p).$$

The boundary condition is

$$V_T'(r_i|p) = 0, \forall r_i, p.$$

Next we calculate the expected revenue before switching the price. Denote $V_t(r_i|p)$ the expected revenue the seller can earn for the rest of the time when the current review is r_i, the current price is p and the seller is still able to switch to another price. We have if $r_i \geq p$,

$$V_t(r_i|p) = \max(p + \sum_j p_{ij} V_{t+1}(r_j|p), \max_{p'} V_t'(r_i|p')),$$

where p_{ij} is the probability the review becomes to r_j at time $t+1$. If $r_i < p$, we have

$$V_t(r_i|p) = \max_{p'} V_t'(r_i|p').$$

The boundary condition is

$$V_T(r_i|p) = 0, \forall r_i, p.$$

Figure 5.11 shows that the performance of partially adaptive policy is close to the performance of fully adaptive policy. It means that being able to adaptively choose the price is more important to the seller than being able to determine the switching time adaptively.

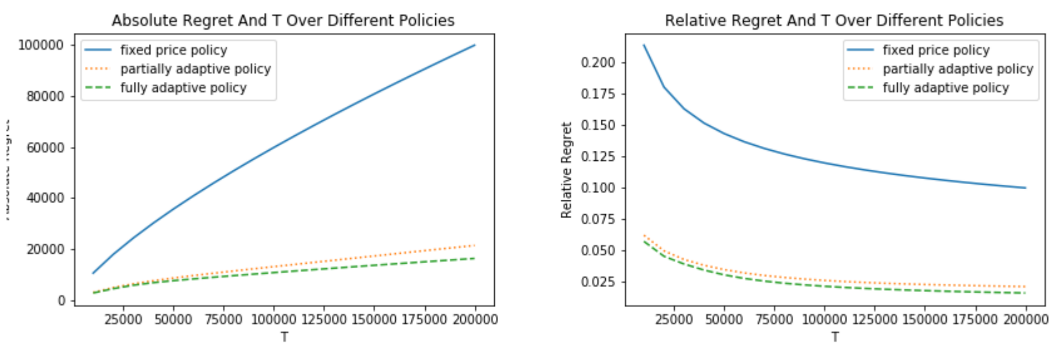

Figure 5.11: Performance Comparison, Regret-Time

5.4 Discrete Model

In this section, we show the result for the discrete model, that a customer is coming at $t_0+1, t_0+2, \ldots, T$. In the previous two sections, the review follows a normal distribution. In this section, we assume the review follows a binary distribution with the probability of giving a good review q. The simulation is the same as the one in Section 5.1, except that the review now follows a binary distribution.

And the basic model setting is $q = 0.6$, $q_0 = 0.6$, $t_0 = 100$ and $T = 10000, 20000, \ldots, 200000$. We use the policy proposed in Theorem 8. We also test a four-price policy modified from Theorem 3.

We first verify the regret growth in the discrete model setting. Figure 5.12 shows the regret-T curve for fixed price policy with two prices and four prices. The relative regret decreases to 0 as T increases. It means the previous result still holds for the discrete model setting and with a binary review distribution.

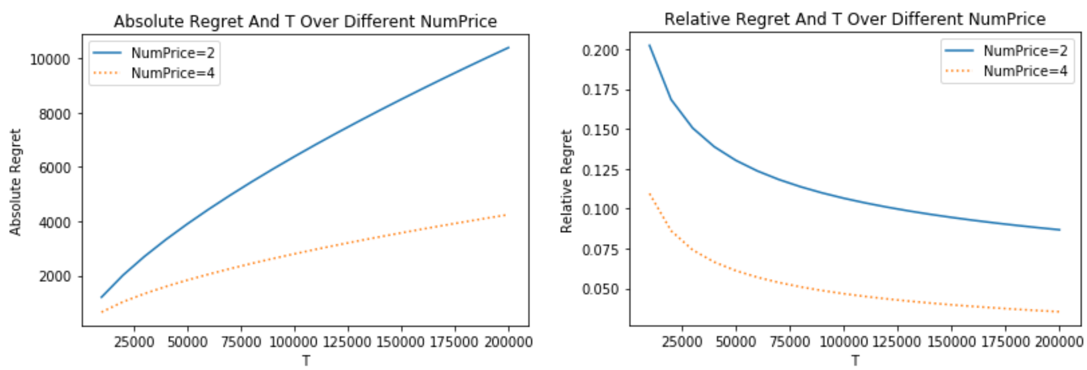

Figure 5.12: Discrete Model, Regret-Time, for discrete policy with different number of prices

Next we check the impact of q, which is the probability of giving a good review. We draw three curves with $q = 0.3$, $q = 0.6$, $q = 0.8$ and the rest of the parameters are the same with the basic model setting. Figure 5.13 shows that increasing q increases the absolute regret but decreases the relative regret. This is consistent with the result we have in Section 5.1.

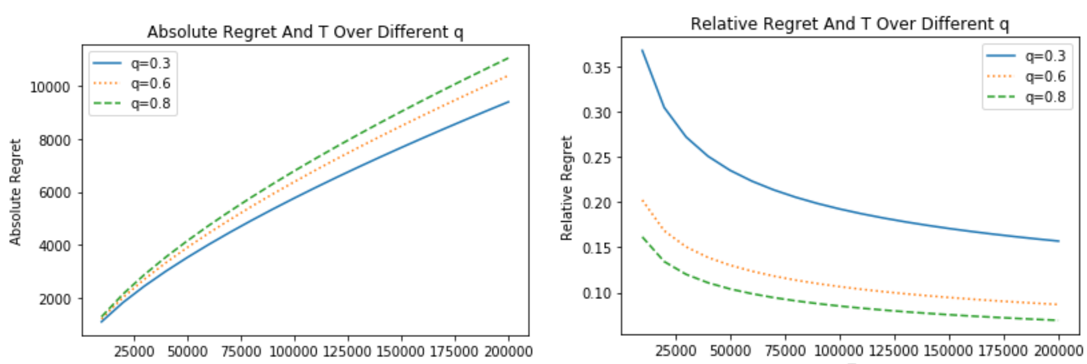

Figure 5.13: Discrete Model, Regret-Time, for discrete policy with different q

Next, we check the impact of t_0, which is the initial review level. We draw two curves with $t_0 = 10$ and $t_0 = 100$, and the rest of the parameters are the same with the basic model setting. Figure 5.14 shows that increasing t_0 slightly decreases the regret. This is mainly because increases t_0 reduces the total sales horizon when T is fixed.

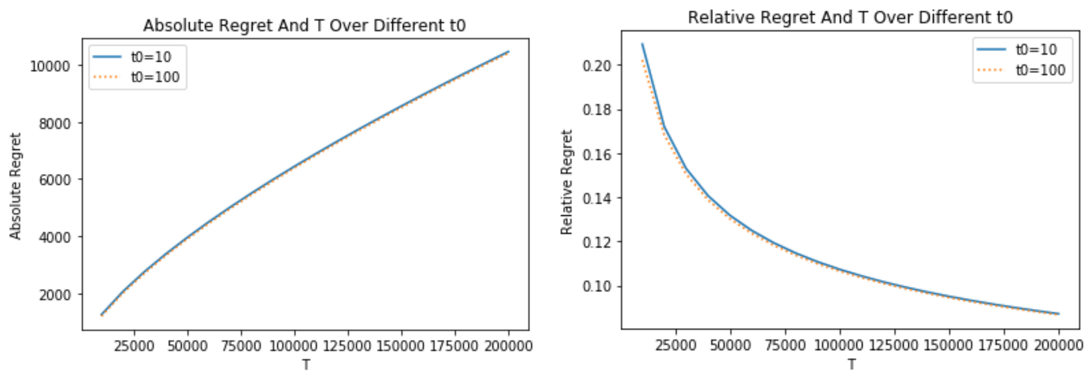

Figure 5.14: Discrete Model, Regret-Time, for discrete policy with different t_0

All the plots show that the performance of the fixed price policy in the discrete model and a binary review distribution is similar to the performance of the fixed price policy in the continuous model and a normal review distribution. This verifies that our result can be applied to a general review distribution.

Chapter 6

Conclusion

In this dissertation, we consider a dynamic pricing problem when customers' utility is determined by both the price and the current review level. We first discuss a fixed two-price policy, where both the price and the switching time are given at the start of the sales horizon. We provide a policy and prove that it can achieve $O(T^{2/3} \log T)$ regret. The policy is similar to what we observe in practice, in which the seller first gives a discount, then switches to a higher price. We also prove that the $\Omega(T^{2/3}/\log T)$ regret is the best one can do for any two price fixed price policy. After that, we extend the result to fixed multi-price policy. By increasing the number of prices to use, the seller can achieve a better regret $O(T^{a_k} \log T)$, where k is the number of prices to use and a_k is a decreasing sequence converging to 1/2. We then prove that this regret is the best the seller can get under any k price fixed price policy.

After that, we consider another type of policy — adaptive policy. For the partially adaptive policy, the seller only needs to specify the switching time and the first price at the beginning. We prove that the partially adaptive policy still has the same order of regret, $\Omega(T^{2/3}/\log T)$ for two-price and $\Omega(T^{a_k}/\log T)$ for k price. The extra flexibility of adaptively determining the price does not bring the seller any advantage asymptotically.

We then consider the fully adaptive two-price policy, for which the seller only needs to determine the first price at the beginning of the sales. The seller can decide when to switch to the second price at any time, and the price is also determined by the seller adaptively. We show that the fully adaptive two-price policy still has a $\Omega(T^{2/3}/\log T)$ regret.

Finally, we consider a discrete model and the customer review is a general distribution. We prove that our proposed policy can still achieve a $O(T^{2/3}\log T)$ regret and this is also asymptotic optimal. Therefore the asymptotic optimal performance is independent of the distribution of the customer review.

Finally, we comment that although our results are all based on a simple utility function $u = q - p$, the results in this dissertation can be extended to general utility functions. To briefly discuss this, consider a general utility function $u(q, p)$. Assume $u(q, p)$ is strictly increasing on q and strictly decreasing on p. We also assume that for any q, there exists p such that $u(q, p) = 0$. Denote $f(q)$ that $u(q, f(q)) = 0$. It is easy to see the for a given q, $f(q)$ is unique. Under this setting, if the customers know the true quality of the product, the seller can set a price $p = f(q)$ and his revenue will be $f(q)(T - t_0)$. Therefore we define the regret in this setting by

$$R_T(\pi) = f(q)(T - t_0) - J_T(\pi).$$

We can modify our proposed policy, by letting $p_1 = f(-\log T)$, $t_1 = T^{2/3}$ and $p_2 = f(q - \log T/T^{1/3})$. If f is a L-Lipschitz continuous function then we can prove that the regret of the modified policy is still $O(T^{2/3}\log T)$. The proof is similar to the proof of Theorem 1 with an additional argument that

$$f(q) - f(q - \log T/T^{1/3}) \leq L \cdot \log T/T^{1/3}.$$

Therefore our work extends to the general utility function setting. Lastly, we list some future research directions for this work:

- A fully adaptive policy with multiple prices. We can extend the fully adaptive two-price policy to multi-price setting. In the fully adaptive multi-price setting, the seller can switch to up to K prices. Only the first price is predetermined. The rest $K-1$ prices and all the switching times are determined adaptively.

- Customer review based on both the valuation of the product and the price of the product. Our model assumes that the review r_t is drawn from a normal distribution $N(q, \sigma^2)$ that does not depend on the price. One can consider a new model that the review is depending on both the valuation and the price charged, for example, r_t is drawn from $N(q - p_t, \sigma^2)$, where p_t is the price paid by the tth customer.

- Utility functions that include the number of purchases. Right now the utility function only depends on the average review and the price. One can consider the case that customers are risk averse and their utility functions are also negatively correlated with the estimated variance of the reviews.

- Customer arrival rate depending on the price. For example, we can consider a model where customer arrival rate is a Poisson Process with rate $\lambda(p)$, where p is the current price.

- Customers are heterogeneous. In this model, we assume customers are homogeneous. We can consider the case where there are two types of customers, for example, high valuation low price sensitivity and low valuation high price sensitivity. The seller may want to use a high price to select high valuation customers and keep the review high.

- The quality of the product q is a variable controlled by the seller. In the current model, we assume the quality is a fixed parameter. We can consider a model when the seller can determine the quality of the product. A higher quality product has a higher cost.

- There are multiple sellers in the market. In the current model, we consider only one seller selling the product. We can extend this to multiple seller setting. The customer's purchase decision is based on the price and review of each seller's product.

- The quality of the product is not revealed to the seller. Currently in the model we assume only the seller knows the true quality of the product. We can consider a model that the seller does not know the true quality of the product, but has to use customer's review to infer the quality.

References

[1] Quarterly retail e-commerce sales report 4th quarter 2013. Technical report, The Census Bureau of the Department of Commerce.

[2] Quarterly retail e-commerce sales report 4th quarter 2018. Technical report, The Census Bureau of the Department of Commerce.

[3] E-commerce in china 2013. Technical report, Ministry of Commerce of the People's Republic of China.

[4] E-commerce in china 2017. Technical report, Ministry of Commerce of the People's Republic of China.

[5] Michael Luca. Reviews, reputation, and revenue: The case of yelp. com. 2016.

[6] Robert G Cross. *Revenue Management: Hard-core Tactics for Market Domination.* Crown Business, 1998.

[7] Richard D Hanks, Robert G Cross, and R Paul Noland. Discounting in the hotel industry: A new approach. *Cornell Hotel and Restaurant Administration Quarterly*, 43(4):94–103, 2002.

[8] Kalyan T Talluri and Garrett J Van Ryzin. *The theory and practice of revenue management*, volume 68. Springer Science & Business Media, 2006.

[9] Jeffrey I McGill and Garrett J Van Ryzin. Revenue management: Research overview and prospects. *Transportation Science*, 33(2):233–256, 1999.

[10] Gabriel Bitran and René Caldentey. An overview of pricing models for revenue management. *Manufacturing & Service Operations Management*, 5(3):203–229, 2003.

[11] Daniel McFadden et al. Conditional logit analysis of qualitative choice behavior. pages 105–142, 1973.

[12] Kalyan Talluri and Garrett Van Ryzin. Revenue management under a general discrete choice model of consumer behavior. *Management Science*, 50(1):15–33, 2004.

[13] Serguei Netessine, Sergei Savin, and Wenqiang Xiao. Revenue management through dynamic cross selling in e-commerce retailing. *Operations Research*, 54(5):893–913, 2006.

[14] Dan Zhang and William L Cooper. Revenue management for parallel flights with customer-choice behavior. *Operations Research*, 53(3):415–431, 2005.

[15] William L Cooper, Tito Homem-de Mello, and Anton J Kleywegt. Models of the spiral-down effect in revenue management. *Operations Research*, 54(5):968–987, 2006.

[16] Ioana Popescu and Yaozhong Wu. Dynamic pricing strategies with reference effects. *Operations Research*, 55(3):413–429, 2007.

[17] Hyun-soo Ahn, Mehmet Gümüş, and Philip Kaminsky. Pricing and manufacturing decisions when demand is a function of prices in multiple periods. *Operations Research*, 55(6):1039–1057, 2007.

[18] Ozan Candogan, Kostas Bimpikis, and Asuman Ozdaglar. Optimal pricing in networks with externalities. *Operations Research*, 60(4):883–905, 2012.

[19] Wenjing Duan, Bin Gu, and Andrew B Whinston. Do online reviews matter? An empirical investigation of panel data. *Decision support systems*, 45(4):1007–1016, 2008.

[20] Judith A Chevalier and Dina Mayzlin. The effect of word of mouth on sales: Online book reviews. *Journal of marketing research*, 43(3):345–354, 2006.

[21] Guiyang Xiong and Sundar Bharadwaj. Prerelease buzz evolution patterns and new product performance. *Marketing Science*, 33(3):401–421, 2014.

[22] Nga N Ho-Dac, Stephen J Carson, and William L Moore. The effects of positive and negative online customer reviews: Do brand strength and category maturity matter? *Journal of Marketing*, 77(6):37–53, 2013.

[23] Bin Gu, Jaehong Park, and Prabhudev Konana. Research notethe impact of external word-of-mouth sources on retailer sales of high-involvement products. *Information Systems Research*, 23(1):182–196, 2012.

[24] Guillermo Gallego and Garrett Van Ryzin. Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Management Science*, 40(8):999–1020, 1994.

[25] Youyi Feng and Baichun Xiao. A continuous-time yield management model with multiple prices and reversible price changes. *Management Science*, 46(5):644–657, 2000.

[26] Guillermo Gallego and Garrett Van Ryzin. A multiproduct dynamic pricing problem and its applications to network yield management. *Operations Research*, 45(1):24–41, 1997.

[27] Wen Zhao and Yu-Sheng Zheng. Optimal dynamic pricing for perishable assets with nonhomogeneous demand. *Management Science*, 46(3):375–388, 2000.

[28] Tze Leung Lai and Herbert Robbins. Asymptotically efficient adaptive allocation rules. *Advances in applied mathematics*, 6(1):4–22, 1985.

[29] Josef Broder and Paat Rusmevichientong. Dynamic pricing under a general parametric choice model. *Operations Research*, 60(4):965–980, 2012.

[30] N Bora Keskin and Assaf Zeevi. Dynamic pricing with an unknown demand model: Asymptotically optimal semi-myopic policies. *Operations Research*, 62(5):1142–1167, 2014.

[31] J Michael Harrison, N Bora Keskin, and Assaf Zeevi. Bayesian dynamic pricing policies: Learning and earning under a binary prior distribution. *Management Science*, 58(3):570–586, 2012.

[32] Omar Besbes and Assaf Zeevi. Blind network revenue management. *Operations Research*, 60(6):1537–1550, 2012.

[33] Zizhuo Wang, Shiming Deng, and Yinyu Ye. Close the gaps: A learning-while-doing algorithm for single-product revenue management problems. *Operations Research*, 62(2):318–331, 2014.

[34] Davide Crapis, Bar Ifrach, Costis Maglaras, and Marco Scarsini. Monopoly pricing in the presence of social learning. *Management Science*, 63(11):3586–3608, 2016.

[35] Dongwook Shin and Assaf Zeevi. Dynamic pricing and learning with online product reviews. Technical report, Working paper, Columbia University, 2017.

[36] Rick Durrett. *Probability: Theory and Examples*. Cambridge university press, 2010.

[37] Thomas H Scheike. A boundary-crossing result for brownian motion. *Journal of Applied Probability*, 1992.

Appendix A

Proof of Theorems

Proof of Lemma 7

Lemma 7. *Let $\mathbb{B}_t$ be a standard Brownian motion. For $a, b \geq 0$, the probability that B_t goes beyond the line $a + bt$ is:*

$$\mathbb{P}(\cup_{0 \leq t \leq \infty}\{\mathbb{B}_t \geq a + bt\}) = \exp(-2ab).$$

Proof. This result is from [37]:

□

Proof of Lemma 8

Lemma 8. *Let $\mathbb{B}_t$ be a standard Brownian motion. For $a \geq 0$, the probability that $\mathbb{B}_t$ goes beyond the line $a + bt$ is:*

$$\mathbb{P}(\cup_{0 \leq t \leq T}\{\mathbb{B}_t \geq a + bt\}) = 1 - \Phi(a/\sqrt{T} + b\sqrt{T}) + \exp(-2ab)\Phi(-a/\sqrt{T} + b\sqrt{T}). \quad (0.1)$$

Furthermore, let $\tau = \inf\{t : \mathbb{B}_t \geq a + bt\} \wedge T$*. Then*

$$\mathbb{E}_{\tau}[\tau] \leq [1 - \exp(-2ab)]^{+}T + 4a \cdot \min\{1, \exp(-2ab)\}\sqrt{T}, \tag{0.2}$$

where $[x]^{+} = \max\{x, 0\}$.

Proof. First, (0.1) follows from Proposition 2 in [37]. For (0.2), we have by (0.1),

$$\mathbb{E}[\tau] = \int_0^T \Big(\Phi(a/\sqrt{t} + b\sqrt{t}) - \exp(-2ab)\Phi(-a/\sqrt{t} + b\sqrt{t})\Big)\, dt.$$

When $b \geq 0$, we have

$$\begin{aligned}
&\Phi(a/\sqrt{t} + b\sqrt{t}) - \exp(-2ab)\Phi(-a/\sqrt{t} + b\sqrt{t}) \\
=&(1 - \exp(-2ab))\Phi(a/\sqrt{t} + b\sqrt{t}) + \exp(-2ab)[\Phi(a/\sqrt{t} + b\sqrt{t}) - \Phi(-a/\sqrt{t} + b\sqrt{t})] \\
\leq&1 - \exp(-2ab) + \frac{2a\exp(-2ab)}{\sqrt{t}}.
\end{aligned}$$

When $b < 0$, we have

$$\Phi(a/\sqrt{t}+b\sqrt{t}) - \exp(-2ab)\Phi(-a/\sqrt{t}+b\sqrt{t}) \leq \Phi(a/\sqrt{t}+b\sqrt{t}) - \Phi(-a/\sqrt{t}+b\sqrt{t}) \leq \frac{2a}{\sqrt{t}}.$$

Combining the two different cases together, we have the result claimed in the lemma.

□

Proof of Lemma 2.

Proof. At time s, if we know the exact review level G_s, then we can view the second period revenue as a new selling process starting from $\hat{q}_0 = y/s$, $\hat{t}_0 = s$. Conditioning

$\mathbb{E}_{\tau_{p,s}}[\tau_{p,s} - s]$ on the realization of G_s, we have

$$p\mathbb{E}_{\tau_{p,s}}[\tau_{p,s} - s] = p\int_{y\geq ps} \mathbb{E}_{\tau_{p,s}}[\tau_{p,s} - s|G_s = y]\mathbb{P}(G_s = y)dy. \tag{0.3}$$

1. The Price Is Higher Than The True Quality

   If $p > q$, then by applying Lemma 1, we have for $y \geq ps$,

$$p\mathbb{E}_{\tau_{p,s}}[\tau_{p,s} - s|G_s = y] \leq \frac{4p(y - ps)}{\sigma}\sqrt{T - s}.$$

Hence,

$$\begin{aligned}
& p\int_{y\geq ps} \mathbb{E}_{\tau_{p,s}}[\tau_{p,s} - s|G_s = y]\mathbb{P}(G_s = y)dy. \\
\leq & \frac{4p}{\sigma}\sqrt{T-s}\int_{y\geq ps} \frac{1}{\sigma\sqrt{s-t_0}}\phi\left(\frac{y - q(s-t_0) - q_0t_0}{\sigma\sqrt{s-t_0}}\right)(y - ps)dy \\
= & \frac{4p}{\sigma}\sqrt{T-s}\int_{y\geq ps} \frac{1}{\sigma\sqrt{s-t_0}}\phi\left(\frac{y - q(s-t_0) - q_0t_0}{\sigma\sqrt{s-t_0}}\right)(y - q(s-t_0) - q_0t_0)dy \\
+ & \frac{4p}{\sigma}\sqrt{T-s}\int_{y\geq ps} \frac{1}{\sigma\sqrt{s-t_0}}\phi\left(\frac{y - q(s-t_0) - q_0t_0}{\sigma\sqrt{s-t_0}}\right)(q(s-t_0) + q_0t_0 - ps)dy \\
= & \frac{4p\sqrt{T-s}\sqrt{s-t_0}}{\sqrt{2\pi}}\exp\left(-\frac{(ps - q(s-t_0) - q_0t_0)^2}{2\sigma^2(s-t_0)}\right) \\
+ & \frac{4p}{\sigma}((q-p)(s-t_0) + (q_0-p)t_0)\sqrt{T-s}\left(1 - \Phi\left(\frac{ps - q(s-t_0) - q_0t_0}{\sigma\sqrt{s-t_0}}\right)\right) \\
\leq & \frac{4p\sqrt{T-s}\sqrt{s-t_0}}{\sqrt{2\pi}}\exp\left(-\frac{((p-q)(s-t_0) + (p-q_0)t_0)^2}{2\sigma^2(s-t_0)}\right) + \frac{q_0^2t_0}{\sigma}\sqrt{T-s}.
\end{aligned}$$

Now we consider three possible cases.

1. $p > 2q + q_0$. In this case

$$\begin{aligned} &p\sqrt{s-t_0}\exp\left(-\frac{((p-q)(s-t_0)+(p-q_0)t_0)^2}{2\sigma^2(s-t_0)}\right) \\ \le&(p-q)\sqrt{s-t_0}\exp\left(-\frac{(p-q)^2(s-t_0)}{2\sigma^2}\right)+q\sqrt{s-t_0}\exp\left(-\frac{q^2(s-t_0)}{2\sigma^2}\right)=O(1). \end{aligned}$$

   where the last equality is because $x\exp(-x^2/2\sigma^2)$ is uniformly bounded by a constant. Thus in this case $S \ge q(T-s) - O(\sqrt{T-s}) = \Omega(T-s)$.

2. $p \le 2q + q_0$ and $s < T^{1-\epsilon}$. In this case

$$\frac{4p\sqrt{T-s}\sqrt{s-t_0}}{\sqrt{2\pi}} \le \frac{4(2q+q_0)}{\sqrt{2\pi}}T^{11/12}.$$

   Thus $S = \Omega(T)$.

3. If $q + T^{-(1/2-\epsilon)} < p \le 2q + q_0$ and $s \ge T^{1-\epsilon}$. In this case

$$\begin{aligned} &p\sqrt{s-t_0}\exp\left(-\frac{((p-q)(s-t_0)+(p-q_0)t_0)^2}{2\sigma^2(s-t_0)}\right) \\ \le&(2q+q_0)T^{1/2}\exp\left(-\frac{T^\epsilon-t_0}{2\sigma^2}+\frac{2q(q+q_0)t_0}{\sigma^2}\right)=O(1). \end{aligned}$$

   Thus $S = \Omega(T-s)$.

From 1, 3 we conclude that if $p > q + T^{-(1/2-\epsilon)}$, $s \ge T^{1-\epsilon}$,

$$S = \Omega(T-s).$$

From 1, 2 we conclude that if $q \le p$ and $s < T^{1-\epsilon}$,

$$S = \Omega\left(\frac{T-s}{\sqrt{s-t_0}}\right).$$

2. The Price Is Not Higher Than The True Quality And The switching time To Use The Price Is Bounded.

When $p \leq q/2$, $S \geq q(T-s) - p(T-s) \geq q(T-s)/2 = \Omega(T-s)$.

Next, we assume $p > q/2$. Applying Lemma 1, we have

$$\begin{aligned} & p\mathbb{E}_{\tau_{p,s}}[\tau_{p,s} - s | G_s = y] \\ \leq & p(T-s)[1 - \exp(-2yq/\sigma^2)] + \frac{4p(y-ps)}{\sigma} \exp\left(-\frac{2(q-p)(y-ps)}{\sigma^2}\right)\sqrt{T-s}. \end{aligned}$$

Hence,

$$\begin{aligned} & p\mathbb{E}_{\tau_{p,s}}[\tau_{p,s} - s | G_s = y] \\ =& p\int_{y\geq ps} \mathbb{P}(G_s = y)\cdot \mathbb{E}_{\tau_{p,s}}[\tau_{p,s} - s|G_s = y]dy \\ \leq & \int_{y \geq ps} \frac{1}{\sigma\sqrt{s-t_0}}\phi\left(\frac{y - q(s-t_0) - q_0 t_0}{\sigma\sqrt{s-t_0}}\right) \\ & \cdot \left[p(T-s) - p(T-s)\exp(-2yq/\sigma^2) + \frac{4p(y-ps)}{\sigma}\exp\left(-\frac{2(q-p)(y-ps)}{\sigma^2}\right)\sqrt{T-s}\right] dy. \end{aligned} \tag{0.4}$$

For the first term in (0.4), we have

$$\begin{aligned} & \int_{y\geq ps} \frac{1}{\sigma\sqrt{s-t_0}}\phi\left(\frac{y-q(s-t_0)-q_0t_0}{\sigma\sqrt{s-t_0}}\right)p(T-s)dy \\ =& p(T-s)\left(1 - \Phi\left(\frac{ps - q(s-t_0) - q_0 t_0}{\sigma\sqrt{s-t_0}}\right)\right). \end{aligned}$$

For the second term in (0.4) , we have

$$
\begin{aligned}
&-p(T-s)\int_{y\geq ps}\frac{1}{\sigma\sqrt{s-t_0}}\phi\left(\frac{y-q(s-t_0)-q_0t_0}{\sigma\sqrt{s-t_0}}\right)\exp\left(\frac{-2yq}{\sigma^2}\right)dy\\
=&-p(T-s)\exp\left(-\frac{2qq_0t_0}{\sigma^2}\right)\int_{ps}^{\infty}\frac{1}{\sqrt{2\pi}\sigma\sqrt{s-t_0}}\exp\left(-\frac{(y-q_0t_0+q(s-t_0))^2}{2\sigma^2(s-t_0)}\right)dy\\
=&-p(T-s)\exp\left(-\frac{2qq_0t_0}{\sigma^2}\right)\left[1-\Phi\left(\frac{ps-q_0t_0+q(s-t_0)}{\sigma\sqrt{s-t_0}}\right)\right].
\end{aligned}
$$

For the third term in (0.4), we have

$$
\begin{aligned}
&\int_{y\geq ps}\frac{1}{\sigma\sqrt{s-t_0}}\phi\left(\frac{y-q(s-t_0)-q_0t_0}{\sigma\sqrt{s-t_0}}\right)\frac{4p(y-ps)}{\sigma}\exp\left(-\frac{2(q-p)(y-ps)}{\sigma^2}\right)\sqrt{T-s}dy\\
=&\exp\left(-\frac{2(q-p)(q_0-p)t_0}{\sigma^2}\right)\frac{4p\sqrt{T-s}}{\sigma}\\
&\cdot\int_{ps}^{\infty}\frac{y-q_0t_0+q(s-t_0)-2p(s-t_0)}{\sqrt{2\pi}\sigma\sqrt{s-t_0}}\exp\left(-\frac{(y-q_0t_0-q(s-t_0)+2(q-p)(s-t_0))^2}{2\sigma^2(s-t_0)}\right)dy\\
&+\frac{4p((p-q)(s-t_0)+(q_0-p)t_0)}{\sigma}\sqrt{T-s}\int_{y\geq ps}\frac{1}{\sigma\sqrt{s-t_0}}\\
&\cdot\phi\left(\frac{y-q(s-t_0)-q_0t_0}{\sigma\sqrt{s-t_0}}\right)\exp\left(-\frac{2(q-p)(y-ps)}{\sigma^2}\right)dy\\
\leq&\exp\left(-\frac{2(q-p)(q_0-p)t_0}{\sigma^2}\right)\frac{2\sqrt{2}p\sqrt{T-s}\sqrt{s-t_0}}{\sqrt{\pi}}\exp\left(-\frac{((p-q_0)t_0+(q-p)(s-t_0))^2}{2\sigma^2(s-t_0)}\right)\\
&+\frac{4p}{\sigma}|q_0-p|t_0\sqrt{T-s}.\\
\leq&\exp\left(-\frac{2(q-p)(q_0-p)t_0}{\sigma^2}\right)\frac{2\sqrt{2}pT^{1-\epsilon/2}}{\sqrt{\pi}}\exp\left(-\frac{((p-q_0)t_0+(q-p)(s-t_0))^2}{2\sigma^2(s-t_0)}\right)+\frac{4p}{\sigma}|q_0-p|t_0\sqrt{T-s},
\end{aligned}
$$

where the last inequality is because $s \leq T^{1-\epsilon}$. Now go back to the regret, we have

$$
\begin{aligned}
S \geq\ & q(T-s) - p(T-s)\left(1 - \Phi\left(\frac{ps - q(s-t_0) - q_0 t_0}{\sigma\sqrt{s-t_0}}\right)\right) \\
& + p(T-s)\exp\left(-\frac{2qq_0t_0}{\sigma^2}\right)\left[1 - \Phi\left(\frac{ps - q_0t_0 + q(s-t_0)}{\sigma\sqrt{s-t_0}}\right)\right] \\
& - \exp\left(-\frac{2(q-p)(q_0-p)t_0}{\sigma^2}\right)\frac{2\sqrt{2}pT^{1-\epsilon/2}}{\sqrt{\pi}}\exp\left(-\frac{((p-q_0)t_0 + (q-p)(s-t_0))^2}{2\sigma^2(s-t_0)}\right) \\
-& \frac{4p}{\sigma}|q_0 - p|t_0\sqrt{T-s} \\
\geq\ & (q-p)(T-s) + p(T-s)\Phi\left(\frac{ps - q(s-t_0) - q_0t_0}{\sigma\sqrt{s-t_0}}\right) \\
& + p(T-s)\exp\left(-\frac{2qq_0t_0}{\sigma^2}\right)\left[1 - \Phi\left(\frac{ps - q_0t_0 + q(s-t_0)}{\sigma\sqrt{s-t_0}}\right)\right] \\
& - \exp\left(-\frac{2(q-p)(q_0-p)t_0}{\sigma^2}\right)\frac{2\sqrt{2}pT^{1-\epsilon/2}}{\sqrt{\pi}}\exp\left(-\frac{((p-q_0)t_0 + (q-p)(s-t_0))^2}{2\sigma^2(s-t_0)}\right) - O(\sqrt{T}).
\end{aligned}
$$

Denote $(q-p)\sqrt{s-t_0} = c$, we have

$$
\begin{aligned}
S \geq\ & (q-p)(T-s) + p(T-s)\Phi\left(-\frac{c}{\sigma} + \frac{(p-q_0)t_0}{\sigma\sqrt{s-t_0}}\right) \\
& + p(T-s)\exp\left(-\frac{2qq_0t_0}{\sigma^2}\right)\left[1 - \Phi\left(\frac{ps - q_0t_0 + q(s-t_0)}{\sigma\sqrt{s-t_0}}\right)\right] - O(\sqrt{T}) \\
& - \exp\left(-\frac{2(q-p)(q_0-p)t_0}{\sigma^2}\right)\frac{2\sqrt{2}pT^{1-\epsilon/2}}{\sqrt{\pi}}\exp\left(-\frac{c^2}{2\sigma^2} - \frac{(p-q_0)^2t_0^2}{2\sigma^2(s-t_0)} - \frac{(q-p)(p-q_0)t_0}{\sigma^2}\right).
\end{aligned}
$$

We consider the following cases:

1. $c > \log T$. In this case,

$$\begin{aligned}
&\exp\left(-\frac{2(q-p)(q_0-p)t_0}{\sigma^2}\right)\frac{2\sqrt{2}pT^{1-\epsilon/2}}{\sqrt{\pi}}\exp\left(-\frac{c^2}{2\sigma^2}-\frac{(p-q_0)^2t_0^2}{2\sigma^2(s-t_0)}-\frac{(q-p)(p-q_0)t_0}{\sigma^2}\right)\\
\leq &\exp\left(-\frac{2(q-p)(q_0-p)t_0}{\sigma^2}\right)\frac{2\sqrt{2}pT^{1-\epsilon/2}}{\sqrt{\pi}}\exp\left(-\frac{\log^2 T}{2\sigma^2}-\frac{(q-p)(p-q_0)t_0}{\sigma^2}\right)\\
\leq &\exp\left(-\frac{2(q-p)(q_0-p)t_0}{\sigma^2}\right)\frac{2\sqrt{2}p}{\sqrt{\pi}}\exp\left(-\frac{(q-p)(p-q_0)t_0}{\sigma^2}\right)T^{1-\epsilon/2-\log T/(2\sigma^2)} = O(1),
\end{aligned}$$

where the last equality is because $p \leq q$. Thus,

$$S \geq (q-p)(T-s) - O(\sqrt{T}) \geq \log T \cdot \frac{T-s}{\sqrt{s-t_0}} - O(\sqrt{T}) = \Omega\left(\frac{T-s}{\sqrt{s-t_0}}\right).$$

2. $1 < c \leq \log T$. In this case, we have

$$p(T-s)\Phi\left(-\frac{c}{\sigma}+\frac{(p-q_0)t_0}{\sigma\sqrt{s-t_0}}\right) \geq p(T-s)\Phi\left(-\frac{c}{\sigma}-\frac{|p-q_0|t_0}{\sigma\sqrt{s-t_0}}\right).$$

Let

$$x = \frac{c}{\sigma}+\frac{|p-q_0|t_0}{\sigma\sqrt{s-t_0}} > \frac{1}{\sigma}.$$

We have

$$\Phi(-x) \geq \frac{1}{x+1/x}\exp(-x^2/2) \text{ for all } x > 0.$$

Now since $1/x \leq \sigma$, and $c < \log T$, we have

$$\begin{aligned}
&\Phi\left(-\frac{c}{\sigma}-\frac{|p-q_0|t_0}{\sigma\sqrt{s-t_0}}\right)\\
\geq &\frac{1}{\log T/\sigma + |p-q_0|qt_0/\sigma+\sigma}\exp\left(-\frac{c^2}{2\sigma^2}-\frac{(p-q_0)^2t_0^2}{2\sigma^2(s-t_0)}-\frac{(q-p)|p-q_0|t_0}{\sigma^2}\right).
\end{aligned}$$

Therefore when T is large (recall $s \leq T^{1-\epsilon}$),

$$p(T-s)\Phi\left(-\frac{c}{\sigma}+\frac{(p-q_0)t_0}{\sigma\sqrt{s-t_0}}\right) \geq T^{1-\epsilon/4}\exp\left(-\frac{c^2}{2\sigma^2}-\frac{(p-q_0)^2t_0^2}{2\sigma^2(s-t_0)}-\frac{(q-p)|p-q_0|t_0}{\sigma^2}\right).$$

Now we also have when $p \leq q$,

$$\begin{aligned}&\exp\left(-\frac{2(q-p)(q_0-p)t_0}{\sigma^2}\right)\frac{2\sqrt{2}pT^{1-\epsilon/2}}{\sqrt{\pi}}\exp\left(-\frac{c^2}{2\sigma^2}-\frac{(p-q_0)^2t_0^2}{2\sigma^2(s-t_0)}-\frac{(q-p)(p-q_0)t_0}{\sigma^2}\right)\\ \leq&\exp\left(\frac{4q(q+q_0)t_0}{\sigma^2}\right)\frac{2\sqrt{2}p}{\sqrt{\pi}}T^{1-\epsilon/2}\exp\left(-\frac{c^2}{2\sigma^2}-\frac{(p-q_0)^2t_0^2}{2\sigma^2(s-t_0)}-\frac{(q-p)|p-q_0|t_0}{\sigma^2}\right).\end{aligned}$$

Thus when T is large,

$$\begin{aligned}&p(T-s)\Phi\left(-\frac{c}{\sigma}+\frac{(p-q_0)t_0}{\sigma\sqrt{s-t_0}}\right)\\ \geq&\exp\left(-\frac{2(q-p)(q_0-p)t_0}{\sigma^2}\right)\frac{2\sqrt{2}pT^{1-\epsilon/2}}{\sqrt{\pi}}\exp\left(-\frac{c^2}{2\sigma^2}-\frac{(p-q_0)^2t_0^2}{2\sigma^2(s-t_0)}-\frac{(q-p)(p-q_0)t_0}{\sigma^2}\right).\end{aligned}$$

Thus

$$S \geq (q-p)(T-s)-O(\sqrt{T}) \geq \frac{T-s}{\sqrt{s-t_0}}-O(\sqrt{T}) = \Omega\left(\frac{T-s}{\sqrt{s-t_0}}\right), \text{ since } s \leq T^{1-\epsilon}.$$

3. $c \leq 1$, $s > t_0+1$. In this case, we have

$$p(T-s)\Phi\left(-\frac{c}{\sigma}+\frac{(p-q_0)t_0}{\sigma\sqrt{s-t_0}}\right) \geq p(T-s)\Phi\left(-\frac{1}{\sigma}-\frac{|p-q_0|t_0}{\sigma}\right).$$

   Thus $S \geq p(T-s)\Phi(-1/\sigma-|p-q_0|t_0/\sigma)-O(\sqrt{T})-O(T^{1-\epsilon/2}) = \Omega(T)$.

4. $c \leq 1$, $s \leq t_0+1$, $p > q_0$. In this case,

$$p(T-s)\Phi\left(-\frac{c}{\sigma}+\frac{(p-q_0)t_0}{\sigma\sqrt{s-t_0}}\right) \geq q_0(T-s)\Phi\left(-\frac{1}{\sigma}\right).$$

We have $S \geq q_0(T-s)\Phi(-1/\sigma) - O(\sqrt{T}) - O(T^{1-\epsilon/2}) = \Omega(T)$.

5. $c \leq 1,\ s \leq t_0 + 1,\ p \leq q_0$,

$$\begin{aligned}
&p(T-s)\exp\left(-\frac{2qq_0t_0}{\sigma^2}\right)\left[1-\Phi\left(\frac{ps-q_0t_0+q(s-t_0)}{\sigma\sqrt{s-t_0}}\right)\right] \\
\geq &p(T-s)\exp\left(-\frac{2qq_0t_0}{\sigma^2}\right)\left[1-\Phi\left(\frac{(p+q)(s-t_0)}{\sigma\sqrt{s-t_0}}\right)\right] \\
\geq &\frac{q(T-s)}{2}\exp\left(-\frac{2qq_0t_0}{\sigma^2}\right)\left[1-\Phi\left(\frac{(q_0+q)}{\sigma}\right)\right],
\end{aligned}$$

where the last inequality is because $p \geq q/2$. We have

$$S \geq q(T-s)\exp\left(-2qq_0t_0/\sigma^2\right)\left[1-\Phi\left((q_0+q)/\sigma\right)\right]/2 - O(\sqrt{T}) - O(T^{1-\epsilon/2}) = \Omega(T).$$

Combiningd with the discussion for the case $p \geq q$, we have when $s < T^{1-\epsilon}$,

$$S = \Omega\left(\frac{T-s}{\sqrt{s-t_0}}\right).$$

□

Proof of Lemma 3

Proof. Denote $H_n = \{G_t | \mathrm{argmax}(G_t/t) \in [n, n+1)\} \cap H$. H_n is the set of trajectories that achieves maximum G_t/t at the interval $[n, n+1)$. $H = \sum_n H_n$. We bound the revenue of a trajectory in H by $\max_t TG_t/t$. Therefore for $G_t \in H_n$, we can bound it with $\max_{t\in[n,n+1)} TG_t/t$.

We first consider the probability that $G_t > qt + m,\ t \in [n, n+1)$. Let $y = G_n$. We have

$$\mathbb{P}(G_t > qt + m | G_n = y) = 2\Phi\left(-\frac{m + qn - y}{\sigma}\right).$$

The expected revenue one can get when $G_n = y$ is

$$
\begin{aligned}
&\mathbb{E}[\max \frac{TG_t}{t} | G_n = y] \\
\le& \frac{T}{n} \int_{\sqrt{n}\log T}^{\infty} (qn + q + m) \frac{2}{\sigma} \phi \left(- \frac{qn + m - y}{\sigma} \right) dm \\
=& \frac{T}{n} \int_{\sqrt{n}\log T}^{\infty} (qn + m - y) \frac{2}{\sigma} \phi \left(- \frac{qn + m - y}{\sigma} \right) dm + \frac{T(y-q)}{n} \int_{\sqrt{n}\log T}^{\infty} \frac{2}{\sigma} \phi \left(- \frac{qn + m - y}{\sigma} \right) dm \\
=& \frac{\sigma T}{n\sqrt{2\pi}} \exp(-(qn + \sqrt{n}\log T - y)^2/(2\sigma^2)) + \frac{2T(y-q)}{n} \Phi \left(- \frac{qn + \sqrt{n}\log T - y}{\sigma} \right).
\end{aligned}
$$

Now consider the distribution of G_n, we have

$$
\begin{aligned}
&\mathbb{E}_y[\mathbb{E}[\max \frac{TG_t}{t} | G_n = y]] \\
&\le \mathbb{P}(y \le qn + \sqrt{n}\log T/2) \mathbb{E}[\max TG_t/t | G_n = qn + \sqrt{n}\log T/2] \\
&+ \int_{y > qn + \sqrt{n}\log T/2} \mathbb{E}[\max TG_t/t | G_n = y] dF_y(y).
\end{aligned}
$$

For the first term, we have

$$
\begin{aligned}
&\mathbb{P}(y \le qn + \sqrt{n}\log T/2) \mathbb{E}[\max TG_t/t | G_n = qn + \sqrt{n}\log T/2] \\
\le& \frac{\sigma T}{n\sqrt{2\pi}} \exp(-(\sqrt{n}\log T/2)^2/(2\sigma^2)) + \frac{T(2qn + \sqrt{n}\log T - 2q)}{n} \Phi \left(- \frac{\sqrt{n}\log T/2}{\sigma} \right) \\
\le& \frac{\sigma T}{\sqrt{2\pi}} T^{-logT/(8\sigma^2)} + T(q + \log T - 2q/n) \frac{2\sigma}{\log T} \exp(-\log^2 T/(4\sigma^2)) \\
=& \frac{\sigma T}{\sqrt{2\pi}} T^{-logT/(8\sigma^2)} + T(q + \log T - 2q/n) \frac{2\sigma}{\log T} T^{-logT/(4\sigma^2)} \\
\le& T(3\sigma + 2\sigma q) T^{-logT/(8\sigma^2)}.
\end{aligned}
$$

For the second term, we have

$$
\begin{aligned}
&\int_{y>qn+\sqrt{n}\log T/2} \mathbb{E}[\max TG_t/t|G_n=y]dF_y(y) \\
\leq &\int_{y>qn+\sqrt{n}\log T/2} \frac{\sigma T+2T(y-q)}{n}\frac{1}{\sigma\sqrt{n-t_0}}\phi\left(\frac{y-q_0t_0-q(n-t_0)}{\sigma\sqrt{n-t_0}}\right)dy \\
= &\int_{y>qn+\sqrt{n}\log T/2} \frac{2T(y-q_0t_0-q(n-t_0))}{n}\frac{1}{\sigma\sqrt{n-t_0}}\phi\left(\frac{y-q_0t_0-q(n-t_0)}{\sigma\sqrt{n-t_0}}\right)dy \\
+ &\int_{y>qn+\sqrt{n}\log T/2} \frac{T\sigma+2T(q_0t_0+q(n-t_0-1))}{n}\frac{1}{\sigma\sqrt{n-t_0}}\phi\left(\frac{y-q_0t_0-q(n-t_0)}{\sigma\sqrt{n-t_0}}\right)dy \\
= &\frac{2T\sigma\sqrt{n-t_0}}{n}\exp\left(-\frac{(qn+\sqrt{n}\log T/2-q_0t_0-q(n-t_0))^2}{2\sigma^2(n-t_0)}\right) \\
+ &\frac{T\sigma+2T(q_0t_0+q(n-t_0-1))}{n}\Phi\left(-\frac{qn+\sqrt{n}\log T/2-q_0t_0-q(n-t_0)}{\sigma\sqrt{n-t_0}}\right) \\
\leq &\frac{2T\sigma\sqrt{n-t_0}}{n}\exp\left(-\frac{\log^2 T+4(q-q_0)t_0\log T/\sqrt{n}}{8\sigma^2}\right) \\
+ &\frac{T\sigma+2T(q_0t_0+q(n-t_0-1))}{n}\frac{\sigma\sqrt{n-t_0}}{(q-t_0)t_0+\sqrt{n}\log T/2}\exp\left(-\frac{\log^2 T+4(q-q_0)t_0\log T/\sqrt{n}}{8\sigma^2}\right)
\end{aligned}
$$

Combining the two terms together, we know that the expected revenue for trajectories at H_n is at the order of $O(T^{-logT/(8\sigma^2)})$. Sum $H_n$through $n=t_0,t_0+1,\cdots,T$. We have the expected revenue of trajectories at H has the same order, $O(T^{-logT/(8\sigma^2)})$. Notice that if a trajectory G_t satisfies $G_t \leq qt+\sqrt{n}\log T, \forall t\in[n,n+1)$, $n=t_0,t_0+1,\cdots,T$, then it must satisfy $G_t\leq qt+\sqrt{t}\log T, \forall t$. Hence we complete the proof. □

Proof of Lemma 4

Proof. From Lemma 3 we can only consider trajectories with $G_t \leq qt+\sqrt{t}\log T$. Denote $q_s = G_s/s$.

1. $p>q$ and $s<T^{1-\epsilon}$

From Lemma 1, we have when $p > q$,

$$p\mathbb{E}_{\tau_{p,s}}[\tau_{p,s} - s|G_s = y] \leq \frac{4p(q_s - p)s}{\sigma}\sqrt{T - s}.$$

Since we know $G_s < qs + s^{1/2}\log T$ and $s < T^{1-\epsilon}$

$$\begin{aligned} p\mathbb{E}_{\tau_{p,s}}[\tau_{p,s} - s|G_s = y] \leq & \frac{4ps^{1/2}\log T}{\sigma}\sqrt{T - s} \\ \leq & \frac{4(qs^{1/2}\log T + \log^2 T)}{\sigma}\sqrt{T - s} \\ = & O(T^{1-\epsilon/2}\log T). \end{aligned}$$

Hence we have

$$S = \Omega(T).$$

2. $q > p \geq q - s^{-1/2}\log^{-1} T$ and $s \leq T^{1-\epsilon}$

We have that

$$\begin{aligned} & p\mathbb{E}_{\tau_{p,s}}[\tau_{p,s} - s|G_s = y] \\ & \leq p\,(T - s)\left[1 - \exp\left(-\frac{2(q_s - p)(q - p)s}{\sigma^2}\right)\right] + \frac{4p(q_s - p)s}{\sigma}\exp\left(-\frac{2(q_s - p)(q - p)s}{\sigma^2}\right)\sqrt{T - s} \\ & \leq q\,(T - s)\left[1 - \exp\left(-\frac{4}{\sigma^2\log^2 T}\right)\right] \\ & + \frac{4q\sqrt{s}}{\sigma\log T}\sqrt{T - s} \\ & \leq q\,(T - s)\left[1 - \exp\left(-\frac{4}{\sigma^2\log T}\right)\right] + \frac{4qT^{1-\epsilon/2}}{\sigma\log T}. \end{aligned}$$

Hence

$$S = q(T - s) - q\,(T - s)\left[1 - \exp\left(-\frac{4}{\sigma^2\log T}\right)\right] - \frac{4qT^{1-\epsilon/2}}{\sigma\log T} = \Omega(T).$$

5. $p < q - s^{-1/2}\log^{-1} T$ and $s \leq T^{1-\epsilon}$

In this case

$$S \geq q(T-s) - (q - s^{-1/2}\log^{-1} T)(T-s) = \Omega(\frac{T-s}{\sqrt{s}\log T}).$$

Combining all the case together, we have $S = \Omega(\frac{T-s}{\sqrt{s}\log T})$.

□

Proof of Lemma 6

Proof. We consider the following cases.

- $p > q + \frac{1}{\log T}$, $s < T^{1-\epsilon}$

  Consider the distribution of G_t at $t = T^{1-\epsilon/2}$. If the stopping time $\tau < T^{1-\epsilon/2}$, the revenue is bounded by $pT^{1-\epsilon/2}$, if the process passes $T^{1-\epsilon/2}$, it is bounded by pT. We have

$$p\mathbb{E}[\tau_{p,s} - s] \leq pT^{1-\epsilon/2}\mathbb{P}(G_{T^{1-\epsilon/2}} < pT^{1-\epsilon/2}) + pT\mathbb{P}(G_{T^{1-\epsilon/2}} \geq pT^{1-\epsilon/2}).$$

  From Hoeffding's inequality we have

$$\begin{aligned}\mathbb{P}(G_{T^{1-\epsilon/2}} \geq pT^{1-\epsilon/2}) &\leq \exp\left(-\frac{2(T^{1-\epsilon/2}-s)^2[p-q+(p-q_0)s/(T^{1-\epsilon/2}-s)]^2}{(T^{1-\epsilon/2}-s)(\bar{q}-\underline{q})^2}\right) \\ &\leq \exp\left(-\frac{2(T^{1-\epsilon/2}-s)[1/\log T+(p-q_0)s/(T^{1-\epsilon/2}-s)]^2}{(\bar{q}-\underline{q})^2}\right) \\ &= O\left(\exp\left(-\frac{2T^{1-\epsilon/2}}{\log^2 T(\bar{q}-\underline{q})^2}\right)\right).\end{aligned}$$

  The probability of passing $T^{1-\epsilon}$ is very small, so the total revenue is bounded by

$O(T^{1-\epsilon/2})$. We have

$$S = q(T-s) - O(T^{1-\epsilon/2}) = \Omega(T).$$

- $q + \frac{1}{\log T} > p \geq q - \frac{1}{\sqrt{s}\log T}$, $s \leq T^{1-\epsilon}$

  We consider the distribution of G_t at $t = s\log^2 T$. With a similar argument in the last case, we can bound the revenue by

$$p\mathbb{E}[\tau_{p,s} - s] \leq p(s\log^2 T - s)\mathbb{P}(G_{s\log^2 T} < ps\log^2 T) + p(T-s)\mathbb{P}(G_{s\log^2 T} \geq ps\log^2 T).$$

To bound $\mathbb{P}(G_{s\log^2 T} \geq ps\log^2 T)$, we use Central Limit Theorem, []

$$\left|\mathbb{P}\left(\alpha \leq \frac{1}{\sqrt{n}\sigma}\sum_{i=1}^{n} X_i \leq \beta\right) - \int_{\alpha}^{\beta}\phi(t)dt\right| \leq \frac{C}{\sqrt{n}},$$

where C is a constant determined by the distribution of X_i. We have

$$\begin{aligned}
\mathbb{P}(G_{s\log^2 T} \geq ps\log^2 T) &= \mathbb{P}\left(\frac{1}{\sigma\sqrt{s\log^2 T - s}}\sum_{i=s+1}^{s\log^2 T} X_i \geq ps\log^2 T - q_s s\right) \\
&= \Phi\left(\frac{(q_s - p)s + (q-p)(s\log^2 T - s)}{\sigma\sqrt{s\log^2 T - s}}\right) + \frac{C}{\sqrt{s\log^2 T - s}} \\
&\leq \Phi\left(\frac{\frac{2\log T}{\sqrt{s}}s + \frac{1}{\sqrt{s}\log T}(s\log^2 T - s)}{\sigma\sqrt{s\log^2 T - s}}\right) + \frac{C}{\sqrt{s\log^2 T - s}} \\
&\leq \Phi\left(\frac{2\log T}{\sigma\sqrt{\log^2 T - 1}} + \frac{\sqrt{\log^2 T - 1}}{\sigma\log T}\right) + \frac{C}{\sqrt{s\log^2 T - s}}
\end{aligned}$$

where $q_s = G_s/s$ is the review level at s and C is a constant. The second first inequality uses the result from Lemma 5 that $q_s \leq q + \log T/\sqrt{s}$. As $T \to \infty$,

$$\frac{\sqrt{\log^2 T - 1}}{\sigma \log T} \to 1, \ \frac{C}{\log T} \to 0,$$

therefore we can find a large enough T' so that $\forall T > T'$,

$$\Phi\left(\frac{2\log T}{\sigma\sqrt{\log^2 T - 1}} + \frac{\sqrt{\log^2 T - 1}}{\sigma \log T}\right) + \frac{C}{\sqrt{s\log^2 T - s}} \leq \Phi\left(\frac{4}{\sigma}\right).$$

Now we bound the regret

$$\begin{aligned}
S &= q(T-s) - p\mathbb{E}[\tau_{p,s} - s] \\
&\geq q(T-s) - p(s\log^2 T - s) - p(T-s)\Phi(4/\sigma) \\
&\geq q(1-\Phi(4/\sigma))(T-s) - \Phi(4/\sigma)(T-s)/\log T - (q + \frac{1}{\log T})T^{1-\epsilon}\log^2 T \\
&= \Omega(T-s).
\end{aligned}$$

- $p < q - \frac{1}{\sqrt{s}\log T}$, $s \leq T^{1-\epsilon}$

  Since the price is small enough we can bound the regret directly,

$$\begin{aligned}
S &\geq q(T-s) - (q - \frac{1}{\sqrt{s}\log T})(T-s) \\
&= \frac{T-s}{\sqrt{s}\log T}.
\end{aligned}$$

Combining the above three cases we prove that for any price p, the regret is

$$S = \Omega\left(\frac{T-s}{\sqrt{s}\log T}\right).$$

□