Doubly Fair Dynamic Pricing

Jianyu Xu **Dan Qiao** **Yu-Xiang Wang**

Computer Science Department,
University of California, Santa Barbara

Abstract

We study the problem of online dynamic pricing with two types of fairness constraints: a *procedural fairness* which requires the *proposed* prices to be equal in expectation among different groups, and a *substantive fairness* which requires the *accepted* prices to be equal in expectation among different groups. A policy that is simultaneously procedural and substantive fair is referred to as *doubly fair*. We show that a doubly fair policy must be random to have higher revenue than the best trivial policy that assigns the same price to different groups. In a two-group setting, we propose an online learning algorithm for the 2-group pricing problems that achieves $\tilde{O}(\sqrt{T})$ regret, zero procedural unfairness and $\tilde{O}(\sqrt{T})$ substantive unfairness over T rounds of learning. We also prove two lower bounds showing that these results on regret and unfairness are both information-theoretically optimal up to iterated logarithmic factors. To the best of our knowledge, this is the first dynamic pricing algorithm that learns to price while satisfying two fairness constraints at the same time.

1 Introduction

Pricing problems have been studied since Cournot (1897). In a classical pricing problem setting such as Kleinberg and Leighton (2003); Broder and Rusmevichientong (2012); Besbes and Zeevi (2015), the seller (referred as "we") sells identical products in the following scheme.

Online pricing. For $t = 1, 2, \ldots, T$:
1. The customer valuates the product as y_t.
2. The seller proposes a price v_t concurrently without knowing y_t.
3. The customer makes a decision $\mathbf{1}_t = \mathbf{1}(v_t \leq y_t)$.
4. The seller receives a reward (revenue) $r_t = v_t \cdot \mathbf{1}_t$.

Here T is the time horizon known to the seller in advance[1],

---

[1] Here we assume T known for simplicity of notations. In fact,

---

and y_t's are drawn from a fixed distribution independently. The goal is to approach an optimal price that maximizes the expected revenue-price function. In order to make this, we should learn gradually from the binary feedback and improve our knowledge on customers' valuation distribution (or so-called "demands" (Kleinberg and Leighton, 2003)).

In recent years, with the development of price discrimination and personalized pricing strategies, *fairness* issues on pricing arose social and academic concerns (Kaufmann et al., 1991; Chapuis, 2012; Richards et al., 2016; Eyster et al., 2021). Customers are usually not satisfied with price discrimination, which can lead to reduced willingness to purchase and a damaged reputation for the seller. In the online pricing problem defined above, when we are selling identical items to customers from different *groups* (e.g., divided by gender, race, age, etc.), it can be unfair to offer different optimal price to each group: Optimal prices in different groups are not necessarily the same, and unfairness occurs if different customers are provided or buying the same item with different prices. Inspired by the concept of *procedural and substantive unconscionability* (Elfin, 1988), we define a *procedural unfairness* measuring the difference of *proposed prices* between the two groups, and a *substantive unfairness* measuring the difference of *accepted prices* between the two groups. Given these notions, our goal is to approach the optimal pricing policy that maximizes the expected total revenue with no procedural and substantive unfairness.

The concept of procedural fairness has been well established in Cohen et al. (2022) as "price fairness", while the concept of the substantive fairness is new to this paper. In fact, both procedural and substantive fairness have significant impacts on customers' experience and social justice. For instance, these notions help prevent the following two scenarios:

- Perspective buyers who are women found that they are offered consistently higher average price than men for the same product.

---

if T is unknown, then we may apply a "doubling epoch" trick as Javanmard and Nazerzadeh (2019) and the regret bounds are the same.

- Women who have bought the product found that they paid a higher average price than men who have bought the product.

Therefore, a good pricing strategy has to satisfy both procedural and substantive fairness.

However, these constraints are very hard to satisfy even with full knowledge on customers' demands. If we want to fulfill those two sorts of fairness perfectly by proposing deterministic prices for different groups, the only thing we can do is to trivially set the same price in all groups and to maximize the weighted average revenue function by adjusting this uniformly fixed price with existing methods such as Kleinberg and Leighton (2003). Consider the following example:

Example 1. Customers form two disjoint groups, where 30% customers are in Group 1 and the rest 70% are in Group 2. For each price in $\{\$0.625, \$0.7, \$1\}$, customers in two groups have different acceptance rates:

Acceptance Rate	$0.625	$0.7	$1
G_1 (30%)	3/5	1/2	1/2
G_2 (70%)	4/5	4/5	1/2

The figure below shows the expected revenue functions of prices in each group, where the red dashed line is their weighted average by population.
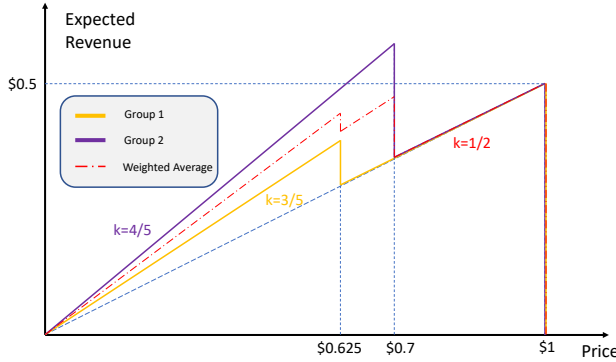
In Example 1, the only way to guarantee both fairness constraints is to propose the same price for both groups, and the optimal price is *$1* whose expected revenue is $\$0.5$ as is shown in the figure.

However, if we instead propose a *random price distribution* to each group and inspect those fairness notions *in expectation*, then there may exist price distributions that satisfy both of the fairness constraints and achieve higher expected revenue than any fixed-price strategy. Here a price distribution is the distribution over the prices for customers, and the exact price for each customer is *sampled* from this distribution *independently*. This random price sampling process can be implemented by marketing campaigns or promotions

such as random discounts or randomly-distributed coupons. Again, we consider Example 1 and the following random policy:

- For customers from G_1, propose *$0.625* with probability $\frac{20}{29}$ and *$1* with probability $\frac{9}{29}$.

- For customers from G_2, propose *$0.7* with probability $\frac{25}{29}$ and *$1* with probability $\frac{4}{29}$.

Under this policy, the expected proposed price and the expected accepted price in both groups are $\$\frac{43}{58}$ and $\$\frac{8}{11}$ respectively. Furthermore, the expected revenue is $\$\frac{74}{145} > \$0.5$, which means that this random policy performs better than the best fixed-price policy. It is worth mentioning that this is exactly the optimal doubly-fair random policy in this specific setting, but the proof of its optimality is highly nontrivial (and we put it in Appendix B.3 as part of the proof of Theorem 9).

In this work, we consider a two-group setting and we denote a *policy* as the tuple of two price distributions over the two groups respectively. Therefore, we can formally define the optimal policy as follows:

$$
\begin{aligned}
\pi_* = \operatorname*{argmax}_{\pi=(\pi^1,\pi^2)} \; q \cdot & \mathbb{E}_{v_t^1 \sim \pi^1, y_t^1 \sim \mathbb{D}^1}[v_t^1 \cdot \mathbb{1}(v_t^1 \leq y_t^1)] \\
+ (1-q) \cdot & \mathbb{E}_{v_t^2 \sim \pi^2, y_t^2 \sim \mathbb{D}^2}[v_t^2 \cdot \mathbb{1}(v_t^2 \leq y_t^2)] \\
s.t. \quad & \mathbb{E}_{\pi^1}[v_t^1] = \mathbb{E}_{\pi^2}[v_t^2] \\
& \mathbb{E}_{\pi^1,\mathbb{D}^1}[v_t^1|\mathbb{1}(v_t^1 \leq y_t^1) = 1] \\
& = \mathbb{E}_{\pi^2,\mathbb{D}^2}[v_t^2|\mathbb{1}(v_t^2 \leq y_t^2) = 1]
\end{aligned}
\tag{1}
$$

Here $\pi^1, v_t^1, y_t^1, \mathbb{D}^1$ and $\pi^2, v_t^2, y_t^2, \mathbb{D}^2$ are the proposed price distributions, proposed prices, customer's valuations and valuation distributions of Group 1 and Group 2 respectively, and q is the share (proportion) that Group 1 takes. From (1), the optimal policy under the in-expectation fairness constraints should be random in general[2]. However, even we know the exact $\mathbb{D}^1$ and $\mathbb{D}^2$, it is still a very hard problem to get π_*: Both sides of the second constraint in (1) are conditional expectations (i.e., fractions of expected revenue over expected acceptance rate) and is thus not convex (and also not quasiconvex). To make it more realistic (and also harder), the seller actually has no direct access to customers' demands $\mathbb{D}^1$ and $\mathbb{D}^2$ at the beginning. Therefore, in this work we consider a T-round *online* learning and pricing setting, where we could learn these demands from those *Boolean-censored* feedback (i.e., customers' decisions) and improve our pricing policy to approach π_* in (1).

In order to measure the performance of a specific policy, we define a *regret* metric that equals the expected revenue dif-

---

[2]Notice that a fixed-price policy can also be considered as "random".

ference between this policy and the optimal policy. We also quantify the procedural and substantive unfairness that are equal to the absolute difference of expected proposed/accepted prices in two groups. We will establish a more detailed problem setting in Section 3.

Summary of Results Our contributions are threefold:

- We design an algorithm, FPA, that achieves an $O(\sqrt{T}d^{\frac{3}{2}}\log\frac{d\log T}{\epsilon})$ cumulative regret with 0 procedural unfairness and $O(\sqrt{T}d^{\frac{3}{2}}\log\frac{d\log T}{\epsilon})$ substantive unfairness, with probability at least $(1-\epsilon)$. Here d is the total number of prices allowed to be chosen from. These results indicate that our FPA is asymptotically no-regret and fair as T gets large.

- We show that the regret of FPA is optimal with respect to T, as it matches $\Omega(\sqrt{T})$ regret lower bound up to $\log\log T$ factors.

- We show that the unfairness of FPA is also optimal with respect to T up to $\log\log T$ factors, as it has no procedural unfairness and its substantive unfairness matches the $\Omega(\sqrt{T})$ lower bound for any algorithm achieving an optimal $O(\sqrt{T})$ regret.

To the best of our knowledge, we are the first to study a pricing problem with multiple fairness constraints, where the optimal pricing policy is necessary to be random. We also develop an algorithm that is able to approach the best random pricing policy with high probability and at the least cost of both revenue and fairness.

Technical Novelty. Our algorithm is a "conservative policy-elimination"-based strategy that runs in epochs with doubling batch sizes as in Auer et al. (2002a). We cannot directly apply the action-elimination algorithm for multi-armed bandits as in Cesa-Bianchi et al. (2013), because the policy space is an infinite set and we cannot afford to try each one out. The fairness constraints further complicate things. Our solution is to work out just a few representative policies that are "good-and-exploratory", which can be used to evaluate the revenue and fairness of all other policies, then eliminate those that are unfair or have suboptimal revenue. Since we do not have direct access to the demand function, the estimated fairness constraints are changing over epochs due to estimation error. Therefore, it is non-trivial to keep the target optimal policy inside our "good policy set" during iterations. We settle this issue by setting the criteria of a "good policy" conservatively.

Our lower bound is new too and it involves techniques that could be of independent interest to the machine learning theory community. Notice that it is possible to have a perfectly fair algorithm by trivially proposing the same fixed price for both groups. It is highly non-trivial to show the unfairness lower bound within the family of regret-optimal algorithms. We present our result in Section 5.3 by establishing two similar problem settings that any algorithm cannot distinguish them efficiently and showing that a mismatch would cause a compatible amount of regret and substantive unfairness.

2   Related Works

Here we discuss some literature closely related to this work. Please refer to Appendix A for a broader discussion.

Dynamic Pricing Single product dynamic pricing problem has been well-studied through Kleinberg and Leighton (2003); Besbes and Zeevi (2009); Wang et al. (2014); Chen et al. (2019); Wang et al. (2021). The crux is to learn and approach the optimal of a revenue curve from Boolean-censored feedback. In specific, Kleinberg and Leighton (2003) proves $\Theta(\log\log T), \Theta(\sqrt{T})$ and $\Theta(T^{\frac{2}{3}})$ minimax regret bounds under noise-free, infinitely smooth and stochastic/adversarial valuation assumptions, sequentially. Wang et al. (2021) further shows a $\Theta(T^{\frac{K+1}{2K+1}})$ minimax regret bound for K^{th}-smooth revenue functions. In all these works, the decision space is continuous. In our problem setting, we require the prices to be chosen from a fixed set of d prices, and show a bandit-style $\Omega(\sqrt{dT})$ regret lower bound similar to Auer et al. (2002b).

Fairness in Machine Learning Fairness is a long-existing topic that has been extensively studied. In the machine learning community, fairness is defined from mainly two perspectives: the *group fairness* and the *individual fairness*. In a classification problem, for instance, (Dwork et al., 2012) defines these two notions as follows: (1) A group fairness requires different groups to have identical result distributions in statistics, which further includes the concepts of "demographic parity" (predictions independent to group attributes) and "equalized odds" (predictions independent to group attributes *conditioning on* the true labels). In Agarwal et al. (2018), these group fairness are reduced to linear constraints. The two fairness definitions we make in this work, the *procedural fairness* and the *substantive fairness*, belong to group fairness. (2) An individual fairness (Hardt et al., 2016) requires the difference of predictions on two individuals to be upper bounded by a distance metric of their intrinsic features. The notion "time fairness" is often considered as individual fairness as well. We provide a more detailed discussion on the line of work that address fairness concerns or stochastic constraints with online learning techniques in Appendix A.

Fairness in Pricing Recently there are many works contributing to pricing fairness problems (Kaufmann et al., 1991; Frey and Pommerehne, 1993; Chapuis, 2012; Richards et al., 2016; Priester et al., 2020; Eyster et al., 2021; Yang et al., 2022). As is stated in Cohen et al. (2022), in a pricing problem with fairness concerns, the concept of fairness in existing works is modeled either as a utility or

budget that trades-off the revenue or as a hard constraint that prevent us from taking the best action directly. Cohen et al. (2022) chooses the second model and defines four different types of fairness in pricing: price fairness, demand fairness, surplus fairness and no-purchase valuation fairness, each of which indicates the difference of prices, the acceptance rate, the surplus (i.e., (valuation − price) if bought and 0 otherwise) and the average valuation of not-purchasing customers in two groups is bounded, sequentially. They show that it is impossible to achieve any pair of different fairness notions simultaneously (with deterministic prices). In fact, this can be satisfied if they allow random pricing policies. Maestre et al. (2018) indeed builds their fairness definition upon random prices by introducing a "Jain's Index", which indicates the homogeneity of price distributions among different groups (i.e., our procedural fairness notion). They develop a reinforcement-learning-based algorithm to provide homogeneous prices, with no theoretic guarantees.

Cohen et al. (2021) and Chen et al. (2021) study the online-learning-fashion pricing problem as we do. Cohen et al. (2021) considers both group (price) fairness and individual (time) fairness, and their algorithm FaPU solves this problem with sublinear regret while guaranteeing fairness. They further study the pricing problem with demand fairness that are unknown and needs learning. In this setting, they propose another FaPD algorithm that achieves the optimal $\tilde{O}(\sqrt{T})$ regret and guarantees the demand fairness "almost surely", i.e., upper bounded by $\delta \cdot T$ as a budget. Chen et al. (2021) considers two different sorts of fairness constraints: (1) Price fairness constraints (as in Cohen et al. (2022)) are enforced; (2)Price fairness constraints are generally defined (and maybe not accessible), where they adopt "soft fairness constraints" by adding the fairness violation to the regret with certain weights. In both cases, they achieve $\tilde{O}(T^{\frac{4}{5}})$ regrets. These learning-based fairness requirements are quite similar to our problem setting, but in our setting the fairness constraints are non-convex (while theirs are linear) and are also optimized to corresponding information-theoretic lower bounds without undermining the optimal regret.

3 Problem Setup

In this section, we describe the problem setting of online pricing, introduce new fairness definitions and set the goal of our algorithm design.

Problem Description. We start with the online pricing process. The whole selling session involves customers from two groups (G_1 and G_2) and lasts for T rounds. Prices are only allowed to be chosen from a *known* and *fixed* set of d prices: $\mathbf{V} = \{v_1, v_2, \ldots, v_d\}$, where $0 < v_1 < v_2 < \ldots < v_d \leq 1$. Denote $\Delta^d = \{x \in \mathbb{R}^d_+, \|x\|_1 = 1\}$ as the probabilistic simplex. At each time $t = 1, 2, \ldots, T$, we propose a pricing policy $\pi = (\pi^1, \pi^2)$ consisting of two probabilistic distributions $\pi^1, \pi^2 \in \Delta^d$ over all d prices. A customer then arrives with an observable group attribution

G_e ($e \in \{1, 2\}$), and we propose a price by sampling a v_t^e from $\mathbf{V}$ according to distribution π^e. At the same time, the customer generates a valuation y_t^e *in secret*, where y_t^e is sampled independently and identically from some fixed unknown distribution $\mathbb{D}_e$. Afterward, we observe a feedback $\mathbf{1}_t^e = \mathbf{1}(v_t^e \leq y_t^e)$ and receive a reward(revenue) $r_t^e = \mathbf{1}_t^e \cdot v_t^e$.

Key Quantities. Here we define a few quantities and functions that is necessary to formulate the problem. Denote $\mathbf{v} := [v_1, v_2, \ldots, v_d]^\top$, $[d] := \{1, 2, \ldots, d\}$ and $\mathbb{1} := [1, 1, \ldots, 1]^\top \in \mathbb{R}^d$ for simplicity. Denote $F_e(i) := \Pr_{\mathbb{D}_e}[y_t^e \geq v_i], e = 1, 2, i \in [d]$ as the probability of price v_i being accepted in G_e. Since $v_i < v_j$ for $i < j$, we know that $F_e(1) \geq F_e(2) \geq \ldots \geq F_e(d)$. Notice that all $F_e(i)$'s are *unknown* to us. Define a matrix $F_e := \mathrm{diag}(F_e(1), F_e(2), \ldots, F_e(d))$.

As a result, for a customer from G_e ($e = 1, 2$), we know that

- The expected proposed price is $\mathbf{v}^\top \pi^e$.

- The expected reward(revenue) is $\mathbf{v}^\top F_e \pi^e$.

- The expected acceptance rate is $\mathbb{1}^\top F_e \pi^e$.

- The expected accepted price is $\frac{\mathbf{v}^\top F_e \pi^e}{\mathbb{1}^\top F_e \pi^e}$.

Denote the proportion of G_1 in all potential customers as q ($0 < q < 1$) which is fixed and *known* to us, and we assume that every customer is chosen from all potential customers uniformly at random. As a consequence, we can define the expected revenue of a policy π.

Definition 2 (Expected Revenue). *For any pricing policy $\pi = (\pi^1, \pi^2) \in \Pi$, define its expected revenue (given F_1 and F_2) as the weighted average of the expected rewards of G_1 and G_2.*

$$R(\pi; F_1, F_2) := \Pr[\textit{Customer is from } G_1] \cdot \mathbb{E}[r_t^1]$$
$$+ \Pr[\textit{Customer is from } G_2] \cdot \mathbb{E}[r_t^2] \quad (2)$$
$$= q \cdot \mathbf{v}^\top F_1 \pi^1 + (1-q) \cdot \mathbf{v}^\top F_2 \pi^2$$

Also, we can define the two different unfairness notions based on these results above.

Definition 3 (Procedural Unfairness). *For any pricing policy $\pi \in \Pi$, define its procedural unfairness as the absolute difference between the expected* proposed *prices of two groups.*

$$U(\pi) := |\mathbf{v}^\top \pi^1 - \mathbf{v}^\top \pi^2| = |\mathbf{v}^\top (\pi^1 - \pi^2)|. \quad (3)$$

Procedural unfairness is totally tractable as we have full access to $\mathbf{v}^\top$ and π. Therefore, we can define a policy family $\Pi := \{\pi = (\pi^1, \pi^2), U(\pi) = 0\}$ that contains all policies with no procedural unfairness. Now we define a substantive unfairness as another metric.

Definition 4 (Substantive Unfairness). *For any pricing policy $\pi \in \Pi$, define its substantive unfairness as the difference between the expected* accepted *prices of two groups.*

$$
\begin{aligned}
S(\pi; F_1, F_2) :=& \big| \mathbb{E}[v^1 | v^1 \sim \pi^1, v^1 \text{ being accepted}] \\
& - \mathbb{E}[v^2 | v^2 \sim \pi^2, v^2 \text{ being accepted}] \big| \\
=& \left| \frac{\mathbf{v}^\top F_1 \pi^1}{\mathbb{1}^\top F_1 \pi^1} - \frac{\mathbf{v}^\top F_2 \pi^2}{\mathbb{1}^\top F_2 \pi^2} \right|.
\end{aligned}
\tag{4}
$$

Substantive unfairness is not as tractable as procedural unfairness, as we have no direct access to the true F_1 and F_2. Ideally, the optimal policy that we want to achieve is:

$$
\begin{aligned}
\pi_* =& \underset{\pi = (\pi^1, \pi^2) \in \Pi}{\operatorname{argmax}} \ R(\pi; F_1, F_2) \\
s.t. & \quad U(\pi) = 0, \quad S(\pi; F_1, F_2) = 0.
\end{aligned}
\tag{5}
$$

The feasibility of this problem is trivial: policies such as $\pi^1 = \pi^2 = [0, \ldots, 0, 1, 0, \ldots, 0]^\top$ (i.e., proposing the same fixed price despite the customer's group attribution) are always feasible. However, this problem is in general highly non-convex and non-quasi-convex. Finally, we define a *(cumulative) regret* that measure the performance of any policy π:

Definition 5 (Regret). *For any algorithm $\mathcal{A}$, define its cumulative regret as follows:*

$$
\begin{aligned}
Reg_T(\mathcal{A}) :=& \sum_{t=1}^{T} Reg(\pi_t; F_1, F_2) \\
:=& \sum_{t=1}^{T} R(\pi_*; F_1, F_2) - R(\pi_t; F_1, F_2).
\end{aligned}
\tag{6}
$$

Here π_t is the policy proposed by $\mathcal{A}$ at time t.

Notice that we define the per-round regret by comparing the performance of π_t with the optimal policy π_* under constraints. Therefore, $Reg(\pi_t; F_1, F_2)$ is possible to be negative if $\pi_t \in \Pi$ but $S(\pi_t; F_1, F_2) > 0$. Similarly, we define a *cumulative substantive unfairness* as $S_T(\mathcal{A}) := \sum_{t=1}^{T} S(\pi_t; F_1, F_2)$.

Goal of Algorithm Design Our ultimate goal is to approach π_* in the performance. In the online pricing problem setting we adopt, however, we cannot guarantee $S(\pi_t; F_1, F_2) = 0$ for all π_t we propose at time $t = 1, 2, \ldots, T$ since we do not know F_1 and F_2 in advance. Instead, we may suffer a gradually vanishing unfairness as we learn F_1 and F_2 better. Therefore, our goal in this work is to design an algorithm that guarantees an optimal regret while suffering 0 cumulative procedural unfairness and the least cumulative substantive unfairness.

Technical Assumptions. Here we make some mild assumptions that help our analysis.

Assumption 1 (Least Probability of Acceptance). *There exists a fixed constant $F_{\min} > 0$ such that $F_e(d) \geq F_{\min}, e = 1, 2$.*

Assumption 1 not only ensures the definition of expected accepted price to be sound (by ruling out these unacceptable prices), but also implies $S(\pi, F_1, F_2)$ to be Lipschitz. Besides, we can always achieve this by reducing v_d. We will provide a detailed discussion in Section 6.

Assumption 2 (Number of Possible Prices). *We treat d, the number of prices, as an amount independent from T. Also, we assume $d = O(T^{\frac{1}{3}})$.*

Assumption 2 is a necessary condition of applying $\Omega(\sqrt{dT})$ regret lower bound, and we will explain more in Appendix B.2. As we are more curious about how the fairness constraints affect the interactive pricing process over time, this assumption separates the dependence on d and helps to show the optimality of our algorithm w.r.t. T.

4 Algorithm

In this section, we propose our Fairly Pricing Algorithm (FPA) in Algorithm 1 and then discuss the techniques we develop and apply to achieve the "no-regret" and "no-unfairness" goal.

4.1 Algorithm Components

Algorithm 1 takes the following inputs: time horizon T, price set $\mathbf{V}$, error probability ϵ, a universal constant L as the coefficient of the performance-fairness tradeoff on constraint relaxations (see Lemma 14), and q as the proportion that G_1 takes. We also adopt the following techniques and components that contribute to its no-regret and no-unfairness performance.

4.1.1 Before Epochs.

In this stage, we keep proposing the highest price v_d for $\tau_0 = O(\log T)$ rounds to estimate(lower-bound) the least accepting probability $F_{\min}$.

Before Epochs:

Initialize counter $M_{0,e} = 0$ and $N_{0,e} = 0$ for $e = 1, 2$.
for $t = 1, 2, \ldots, \tau_0 = 2 \log T \log \frac{16}{\epsilon}$ **do**
    Denote the customer's group index as $e_t \in \{1, 2\}$.
    Set $M_{0,e_t} += 1$.
    Propose the highest price v_d.
    If accepted, set $N_{0,e_t} += 1$.
end for
Output $\hat{F}_{\min} = \min\{\frac{N_{0,1}}{2M_{0,1}}, \frac{N_{0,2}}{2M_{0,2}}\}$.

4.1.2 Doubling Epochs.

Despite the "before epochs" stage, we divide the whole time space into epochs $k = 1, 2, \ldots$, where each epoch k has a length $\tau_k = O(\sqrt{T} \cdot 2^k)$ that doubles the length of epoch $(k-1)$. Intuitively, a longer epoch can improve the estimates and help the algorithm select better policies, which would in

---

Algorithm 1 Fairly Pricing Algorithm (FPA)

1: **Input:** Time horizon T, prices set $\mathbf{V}$, error probability ϵ, universal constant L, proportion q.
2: **Before Epochs:** Keep proposing the highest price v_d and estimate the lowest accepting probability as $\hat{F}_{\min}$. (See Section 4.1.1.)
3: **Initialization:** Candidate policy set $\Pi_1 = \Pi := \{\pi = (\pi^1, \pi^2), U(\pi) = 0\}$ and price index set $I_0^1 = I_0^2 = [d]$.
4: **for** Epoch $k = 1, 2, \ldots$ **do**
5:     Set epoch length τ_k, reward uncertainty $\delta_{k,r}$ and unfairness uncertainty $\delta_{k,s}$.
6:     **Select good-and-exploratory policies** from Π_k and form a set A_k. (See Section 4.1.3.)
7:     **Estimate acceptance probabilities** $F_e(i)$ as $\tilde{F}_{k,e}(i)$ for $e = 1, 2$ and $i = 1, 2, \ldots, d$. (See Section 4.1.4.)
8:     Let $\hat{F}_{k,e} = \operatorname{diag}(\bar{F}_{k,e}(1), \bar{F}_{k,e}(2), \ldots, \bar{F}_{k,e}(d)), e = 1, 2$.
9:     **Get empirical optimal policy** $\hat{\pi}_{k,*}$: Solve (7) with Algorithm 2

$$\hat{\pi}_{k,*} = \operatorname*{argmax}_{\pi \in \Pi_k} R(\pi, \hat{F}_{k,1}, \hat{F}_{k,2}), \text{ s.t. } S(\pi, \hat{F}_{k,1}, \hat{F}_{k,2}) \leq \delta_{k,s}. \tag{7}$$

10:     **Update the policy set** Π_k: Solve (8) with Algorithm 3

$$\Pi_{k+1} = \big\{\pi : \pi \in \Pi_k, S(\pi, \hat{F}_{k,1}, \hat{F}_{k,2}) \leq \delta_{k,s}, R(\pi, \hat{F}_{k,1}, \hat{F}_{k,2}) \geq R(\hat{\pi}_{k,*}, \hat{F}_{k,1}, \hat{F}_{k,2}) - \delta_{k,r} - L \cdot \delta_{k,s}\big\}. \tag{8}$$

11: **end for**

---

return reduce the regret in the next epoch with even longer time horizon.

4.1.3 Good-and-Exploratory Policies.

A profitable pricing policy might not be suitable of running in consideration of exploration, which is important for estimating $F_1(i)$ and $F_2(i)$ that facilitates the policy elimination. We resolve this issue by keeping a set of *good-and-exploratory* policies: After eliminating sub-optimal policies at the end of previous epoch, for each price v_i in group G_e we find out a policy in the remaining policies that maximizes the probability of proposing v_i in G_e at the beginning of current epoch. The larger this probability is, the more times v_i can be chosen in G_e, which would lead to a better estimate of $F_e(i)$. Here we give up to estimate the acceptance probability of those v_i with $\leq \frac{1}{\sqrt{T}}$ to be chosen by the optimal policy π_*, as it would not affect the elimination process and the performance substantially.

Select good-and-exploratory policies:

Initialize $A_k = \emptyset$, $I_k^1 = I_{k-1}^1$ and $I_k^2 = I_{k-1}^2$.
for Group $e = 1, 2$ and for price index $i \in I_{k-1}^e$, **do**
    {Pick up policy maximizing each probability:}
    Get $\tilde{\pi}_{k,i,e} = \operatorname{argmax}_{\pi \in \Pi_k} \pi^e(i)$.
    if $\tilde{\pi}_{k,i,e}^e(i) \geq \frac{1}{\sqrt{T}}$ **then**
        Let $A_k = A_k \cup \{\tilde{\pi}_{k,i,e}\}$
    else
        Remove i from I_k^e.
    end if
end for
Output A_k.

4.1.4 Probability Estimates.

Within epoch k, we run a set of "good-and-exploratory policies" with equal shares of τ_k and then update the estimates of F_1 and F_2. This will in return help the algorithm update the set of "good-and-exploratory policies" as the next epoch starts.

Estimate acceptance probabilities:

Initialize $M_{k,e}(i) = N_{k,e}(i) = 0, \forall i \in [d], e = 1, 2$.
for each policy $\pi \in A_k$, **do**
    Run π for a batch of $\frac{\tau_k}{|A_k|}$ rounds.
    For each time a price v_i is *proposed* in G_e, set $M_{k,e}(i) += 1$.
    For each time a price v_i is *accepted* in G_e, set $N_{k,e}(i) += 1$.
end for
For $e = 1, 2$, set $\bar{F}_{k,e}(i) = \max\{\frac{N_{k,e}(i)}{M_{k,e}(i)}, \hat{F}_{\min}\}$ for $i \in I_k^e$, and $\bar{F}_{k,e}(i) = \hat{F}_{\min}$ otherwise.
Output vectors $\bar{F}_{k,e}$ for $e = 1, 2$.

Policy Eliminations. At the end of each epoch k, we update the candidate policy set by eliminating those substantially sub-optimal policies: Firstly, we select an empirical optimal policy $\hat{\pi}_{k,*}$ that maximizes $R(\pi, \hat{F}_{k,1}, \hat{F}_{k,2})$ while guaranteeing $S(\pi, \hat{F}_{k,1}, \hat{F}_{k,2}) \leq \delta_{k,s}$. After that, we eliminate those policies that meet one of the following two criteria:

- Large unfairness:

$$S(\pi, \hat{F}_{k,1}, \hat{F}_{k,2}) > \delta_{k,s},$$

- Large regret:

$$R(\pi, \hat{F}_{k,1}, \hat{F}_{K,2}) < R(\hat{\pi}_{k,*}, \hat{F}_{k,1}, \hat{F}_{k,2}) - \delta_{k,r} - L \cdot \delta_{k,s}.$$

Here we adopt two subtractors on the regret criteria: $\delta_{k,r}$ for the estimation error in $R(\pi)$ caused by $\hat{F}_{k,e}$, and $L \cdot \delta_{k,s}$ for the possible increase of optimal reward by allowing $S(\pi) \leq \delta_{k,s}$ instead of $S(\pi) = 0$. In this way, we can always ensure the optimal policy π_* (i.e., the solution of (5)) to remain and also guarantee the other remaining policies perform similarly to π_*.

4.2   Computational Cost

Here we show that our FPA algorithm is efficient in computation.

On the one hand, FPA is *oracle-efficient* due to the doubling-epoch design, as we only run each oracle and update each parameter for $O(\log T)$ times. On the other hand, these arg-max oracles can also be implemented in a time-efficient way, although both (7) and (8) are highly non-convex on the constraints. In fact, a sufficient condition of solving (7) is to solve the following constrained optimization problem:

$$
\begin{aligned}
(\hat{\pi}_{k,*}, w^*) = \underset{\pi \in \Pi_k, w \in [0, \frac{1}{F_{\min}}]}{\arg\max} & R(\pi, \hat{F}_{k,1}, \hat{F}_{k,2}), \\
s.t. \quad & v^\top F_1 \pi^1 = w \cdot \mathbb{1}^\top F_1 \pi^1, \\
& v^\top F_2 \pi^2 \geq (w - \delta_{k,s}) \cdot \mathbb{1}^\top F_2 \pi^2, \\
& v^\top F_2 \pi^2 \leq (w + \delta_{k,s}) \mathbb{1}^\top F_2 \pi^2.
\end{aligned}
\tag{9}
$$

Since $w \in [0, \frac{1}{F_{\min}}]$ is a scalar, the optimization problem in (9) is a linear programming for any fixed w. Therefore, we may solve this problem by conducting a linear search over a series of $\{w_\ell\} \subset [0, \frac{1}{\hat{F}_{\min}}]$ (since $\hat{F}_{\min} \leq F_{\min}$) and solve the linear programming at each fixed $w = w_\ell$, which can be implemented as Algorithm 2.

Notice that we choose the linear searching step $\epsilon = \frac{\delta_{k,s}}{2}$ due to the Lipschitzness of the objective function and all constraints. In this way, the discretization error is upper bounded by $O(\delta_{k,s})$, which is in the same order of algorithmic regret w.r.t. T and d (as we will show in the proof of Theorem 6). Since we only solve the linear programming problem for $O(\frac{1}{\delta_{k,s}})$ times (which is a polynomial of T and d) and a linear programming problem can be solved in polynomial time, we know that Algorithm 2 is time-efficient. Similarly, we have the following Algorithm 3 that solves (8) efficiently as well.

5   Regret and Unfairness Analysis

In this section, we analyze the regret and unfairness of our FPA algorithm. We first present an $\tilde{O}(\sqrt{T}d^{\frac{3}{2}})$ regret upper bound along with an $\tilde{O}(\sqrt{T}d^{\frac{3}{2}})$ unfairness upper

---

Algorithm 2 Oracle: Empirical Optimal

1: **Input:** Matrix $\hat{F}_{k,1}, \hat{F}_{k,2}$, policy set Π_k, parameter $\hat{F}_{\min}$, allowed estimation error $\delta_{k,s}$, step length $\epsilon = \frac{\delta_{k,s}}{2}$.
2: **Initialization:** Pick a $\hat{\pi}_{k,*} \in \Pi_k$ arbitrarily.
3: **for** $\ell = 0, 1, 2, \ldots$ **do**
4:    Let $w_\ell = \ell \cdot \epsilon$
5:    **if** $w_\ell > \frac{1}{\hat{F}_{\min}}$ **then**
6:       Break.
7:    **end if**
8:    Solve the following linear program and get $\hat{\pi}_{k,*}$.

$$
\begin{aligned}
\hat{\pi}_{k,\ell,*} = \underset{\pi \in \Pi_k}{\arg\max} \ & R(\pi, \hat{F}_{k,1}, \hat{F}_{k,2}), \\
s.t. \quad & v^\top F_1 \pi^1 = w_\ell \cdot \mathbb{1}^\top F_1 \pi^1, \\
& v^\top F_2 \pi^2 \geq (w_\ell - \delta_{k,s}) \cdot \mathbb{1}^\top F_2 \pi^2, \\
& v^\top F_2 \pi^2 \leq (w_\ell + \delta_{k,s}) \mathbb{1}^\top F_2 \pi^2.
\end{aligned}
$$

9:    {Compare with best existing solution.}
10:   **if** $R(\hat{\pi}_{k,\ell,*}, \hat{F}_{k,1}, \hat{F}_{k,2}) > R(\hat{\pi}_{k,*}, \hat{F}_{k,1}, \hat{F}_{k,2})$ **then**
11:      Substitute $\hat{\pi}_{k,*} \leftarrow \hat{\pi}_{k,\ell,*}$.
12:   **end if**
13: **end for**
14: Return $\hat{\pi}_{k,*}$.

Algorithm 3 Oracle: Policy Elimination

1: **Input:** Matrix $\hat{F}_{k,1}, \hat{F}_{k,2}$, policy set Π_k, parameter $\hat{F}_{\min}$, $\delta_{k,s}$, $\delta_{k,r}$, constant L, policy $\hat{\pi}_{k,*}$ step length $\epsilon = \frac{\delta_{k,s}}{2}$.
2: **Initialization:** Feasible set $\Pi_{k+1} = \emptyset$.
3: **for** $\ell = 0, 1, 2, \ldots$ **do**
4:    Let $w_\ell = \ell \cdot \epsilon$
5:    **if** $w_\ell > \frac{1}{\hat{F}_{\min}}$ **then**
6:       Break.
7:    **end if**
8:    Solve the following linear program and get a feasible set $\Pi_{k+1,w}$.

$$
\begin{cases}
v^\top F_1 \pi^1 = w_\ell \cdot \mathbb{1}^\top F_1 \pi^1, \\
v^\top F_2 \pi^2 \geq (w_\ell - \delta_{k,s}) \cdot \mathbb{1}^\top F_2 \pi^2, \\
v^\top F_2 \pi^2 \leq (w_\ell + \delta_{k,s}) \mathbb{1}^\top F_2 \pi^2 \\
R(\pi, \hat{F}_{k,1}, \hat{F}_{k,2}) \geq R(\hat{\pi}_{k,*}, \hat{F}_{k,1}, \hat{F}_{k,2}) - \delta_{k,r} - L \cdot \delta_{k,s}.
\end{cases}
$$

9:    Update $\Pi_{k+1} \leftarrow \Pi_{k+1} \cup \Pi_{k+1,w}$
10: **end for**
11: Return Π_{k+1}.

bound. Then we show both of them are optimal (w.r.t. T) up to $\log \log T$ factors by presenting matching lower bounds.

5.1 Regret Upper Bound

First of all, we propose the following theorem as the main results for our Algorithm 1 (FPA).

Theorem 6 (Regret and Unfairness). *FPA guarantees an $O(\sqrt{T}d^{\frac{3}{2}} \log \frac{d \log T}{\epsilon})$ regret with no procedural unfairness and an $O(\sqrt{T}d^{\frac{3}{2}} \log \frac{d \log T}{\epsilon})$ substantive unfairness with probability $1 - \epsilon$.*

Proof sketch. We prove this theorem by induction w.r.t. epoch index k. Firstly, we start with the induction assumption that $\pi_* \in \Pi_k$, which naturally holds as $k = 1$. Meanwhile, we show a high-probability bound on the estimation error of each $F_e(i)$ for epoch k, according to concentration inequalities. With this, we derive the estimation error bound of $R(\pi, F_1, F_2)$ and $S(\pi, F_1, F_2)$ for each policy $\pi \in \Pi_k$ in epoch k, given that we are always taking a policy that maximizes the probability of proposing a specific price. After that, we bound the regret and unfairness of each policy remaining in Π_{k+1}, and therefore bound the regret and unfairness of epoch $(k + 1)$ with high probability. Finally, we show that the optimal fair policy π_* (defined in (5)) is also in Π_{k+1}, which matches the induction assumption for Epoch $(k + 1)$. By adding up these performance over epochs, we get the cumulative regret and unfairness respectively. Please refer to Appendix B.1 for a detailed proof. ∎

Remark 7. *Our algorithm guarantees $O(\sqrt{T} \log \log T)$ regret and unfairness simultaneously, whose average-over-time match the generic estimation error of $O(\frac{1}{\sqrt{T}})$. It implies that these fairness constraints do not bring informational obstacles to the learning process. In fact, these upper bounds are tight up to $O(\log \log T)$ factors, which are shown in Theorem 8 and Theorem 9.*

5.2 Regret Lower Bound

Here we show the regret lower bound of the pricing problem.

Theorem 8 (Regret lower bound). *Assume $d \leq T^{\frac{1}{3}}$. Given the online two-group fair pricing problem and the regret definition as (6), any algorithm would at least suffer an $\Omega(\sqrt{dT})$ regret.*

We may prove Theorem 8 by a reduction to online pricing problem with no fairness constraints: Given a problem setting where the two groups are identical, i.e. $F_1(i) = F_2(i), \forall i \in [d]$, and let $q = 0.5$. Notice that any policy satisfying $\pi^1 = \pi^2$ is procedurally and substantively fair, and the optimal policy is to keep proposing the best fixed price. Therefore, this can be reduced to an online identical-product (i.e., non-contextual) pricing problem, and we present a bandit-style lower bound proof in Appendix B.2 inspired by Auer et al. (2002b).

5.3 Unfairness Lower Bound with Optimal Revenue

Here we show that any optimal algorithm has to suffer an $\Omega(\sqrt{T})$ substantive unfairness.

Theorem 9 (Substantive Unfairness Lower Bound). *For any constant C_x, there exists constants $C_u > 0$ such that any algorithm with an $C_x \cdot T^{\frac{1}{2}}$ cumulative regret and zero procedural unfairness has to suffer an $C_u \cdot T^{\frac{1}{2}}$ substantive unfairness.*

It is worth mentioning that this result is different from ordinary lower bounds on the regret, as it also requires the algorithm to be optimal. In general, we propose 2 different problem settings, and we show the following four facts:

- No algorithm can perform well (i.e. low regret and low substantive unfairness) in both settings.

- Any algorithm cannot efficiently distinguish between the two settings.

- Failing in distinguishing between them would lead to a large substantive unfairness.

- Avoid distinguishing between them would suffer even larger regret or substantive unfairness.

In order to prove these, we make use of Example 1 presented in Section 1. One of the settings is exactly Example 1, and the other one is identical to it except these 0.5 acceptance rates (i.e., \$0.7 for G_1 and \$1 for both groups) are now $(0.5 - \zeta)$ in both groups. We get close-form solutions to both problem settings. We further show that the two settings are indistinguishable in information theory, and we either fail or avoid distinguishing them at least for $\Omega(T)$ rounds. Also, we show that the (reward or fairness) loss of avoiding distinguishing is more than $\Omega(\frac{1}{\sqrt{T}})$, and that the fairness loss of failing to distinguish is $\Omega(\frac{1}{\sqrt{T}})$. As we have a tight budget on the cumulative regret, we have to suffer a $\Omega(T \cdot \frac{1}{\sqrt{T}}) = \Omega(\sqrt{T})$ unfairness lower bound as a trade-off. Please refer to Appendix B.3 for more details.

6 Discussion

Here we discuss some open issues and potential extensions of this work. For more discussions on settings, techniques and social impacts, please refer to Appendix C.

Improvement on Technical Assumptions. In this work, we assume the existence of a lower bound $F_{\min} > 0$ of the acceptance rate of all prices for both groups. This assumption is stronger than our expectation, as the seller would not know the highest price that customers would accept. We assume this for two reasons: (1) Without assuming $F_e(i) > 0$, the substantive unfairness function might be undefined. For instance, if a pricing policy is completely unacceptable in G_1 (with *no* accepted prices) but is acceptable in G_2, then

is it a fair policy? (2) With a constantly large probability of acceptance, we can estimate every $\hat{F}_e(i)$ and bound it away from 0 and therefore leads to the Lipschitzness of $S(\pi, \hat{F}_1, \hat{F}_2)$. However, there might exist an algorithm that works for $F_e(i) > 0$ generally and maintains these optimalities as well, which is an open problem to the future.

Feelings of Fairness in FPA. In algorithm 1, notice that we run each $\tilde{\pi} \in A_k$ for a continuous batch of $\frac{\tau_k}{|A_k|} = \Omega(\sqrt{T} \cdot 2^k)$, which is long enough for customers to experience the fairness: Customers would figure out their average seeing and buying prices by comparing these amounts with customers from the other group. On the contrary, if we run an adaptive algorithm that changes the running policy at every time, the customers would feel like they are treated by different policies without knowing that all policies are quite fair (since now each customer has no comparison).

Relaxation on Substantive Fairness. In this work, our algorithm approaches the optimal policy (as the solution of (5)) through an online learning framework. This ensures an asymptotic fairness as $T \rightarrow +\infty$, but we still cannot guarantee perfect any-time fairness precisely (i.e., $S(\pi_t) = 0, \forall t$). Therefore, it is more practical to consider the following inequality-constraint optimization problem:

$$\pi_{\delta,*} = \operatorname*{argmax}_{\pi=(\pi^1,\pi^2)\in\Pi} R(\pi; F_1, F_2)$$
$$s.t. \quad U(\pi) = 0, \quad S(\pi; F_1, F_2) \leq \delta. \tag{10}$$

Comparing (5) with (10), we know that $R(\pi_*) \leq R(\pi_{\delta,*})$. According to Lemma 14, we further know that $R(\pi_*) \geq R(\pi_{\delta,*}) - L \cdot \delta$. Naturally, the substantive unfairness definition is now $\max\{0, S(\pi; F_1, F_2) - \delta\}$. If we still consider this problem under the framework of online learning, then two questions arose naturally: What are the optimal regret rate and (substantive) unfairness rate like? And how can we achieve them simultaneously? From our results in this work, we only know that (1) If $\delta = 0$, then both rates are $\Theta(\sqrt{T})$, and (2) if $\delta \geq 1$, then the optimal regret is $\Theta(\sqrt{T})$ and the optimal unfairness is 0 (as it is reduced to the unconstrained pricing problem). In fact, for $\delta = O(\sqrt{1/T})$, we may still achieve $O(\sqrt{T})$ regret and unfairness, but it is not clear if they are always optimal. For $\delta > \sqrt{1/T}$, we conjecture that the optimal regret is still $\Theta(\sqrt{T})$ and the optimal unfairness could be $\Theta(1/(\sqrt{T}\delta))$.

Trade-offs between Procedural and Substantive Fairness. We conjecture that it is not likely to trade-off substantive fairness with procedural fairness, as the lower bound on substantive fairness comes from the indistinguishability between two similar settings. To set unfair prices intentionally would not substantially speed-up the learning process and make the two environments more distinguishable.

Optimal Policy on the Continuous Space. In this work, we restrict our price choices in a fixed price set $\mathbf{V} =$ $\{v_1, v_2, \ldots, v_d\}$ and aims at the optimal distributions on these v_i's. However, if we are allowed to propose any price within $[0, 1]$, then the optimal policy could be a tuple of two *continuous* distributions that outperforms any policy restricted on $\mathbf{V}$. Even if we know that customers' valuations are all from $\mathbf{V}$, the optimal policy is not necessarily located inside $\mathbf{V}$ due to the fairness constraints. This optimization problem is even harder than (5), and the online-learning scheme further increases its hardness. Existing methods such as continuous distribution discretization (Xu and Wang, 2022) might work, but would definitely lead to an exponential time complexity.

From Two Groups to Multi Groups Our problem setting assumes that there are two groups of customers in total. We choose to study a two-group setting to simplify the presentation. In practice, however, it is very common that customers are coming from many groups with different valuations even on the same product. In fact, we believe it straightforward to extend our techniques and results to G-group settings, as long as we determine a metric of multi-group unfairness. For instance, if we choose to define the multi-group unfairness as the summation of pairwise unfairness of all $O(G^2)$ pairs of groups, we may adjust our algorithm by lengthening each epoch by $G/2$ times and keeping everything the same as in this paper. In this way, the upper regret bound would be $\tilde{O}(G^3\sqrt{T}d^{2/3})$, which is $O(G^3)$ times larger than the existing regret bound. Therefore, it is still optimal w.r.t. T up to iterative-log factors. However, this notion of pairwise unfairness is somewhat controversial in multi-group settings, and we will provide more discussions regarding this generalization on appendix C.1.

7 Conclusion

In this work, we study the online pricing problem with fairness constraints. Specifically, we introduce two fairness notions, a *procedural fairness* and a *substantive fairness*, which respectively ensure the equality of proposed and accepted prices between two different groups. To satisfy these two constraints simultaneously, we adopt *random* pricing policies and establish the objective function and rewards in expectation. To solve this problem with unknown demands, we develop a policy-elimination-based algorithm FPA that achieves an $\tilde{O}(\sqrt{T})$ regret with zero procedural unfairness and within an $\tilde{O}(\sqrt{T})$ substantive unfairness. We show that our algorithm is optimal in both regret and unfairness up to $\log \log T$ factors, by proving an $\Omega(\sqrt{T})$ regret lower bound and an $\Omega(\sqrt{T})$ unfairness lower bound for any optimal algorithm with $O(\sqrt{T})$ regret.

Acknowledgment

Science.

References

Agarwal, A., Beygelzimer, A., Dudík, M., Langford, J., and Wallach, H. (2018). A reductions approach to fair classification. In *International Conference on Machine Learning*, pages 60–69. PMLR.

Auer, P., Cesa-Bianchi, N., and Fischer, P. (2002a). Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2):235–256.

Auer, P., Cesa-Bianchi, N., Freund, Y., and Schapire, R. E. (2002b). The nonstochastic multiarmed bandit problem. *SIAM Journal on Computing*, 32(1):48–77.

Barocas, S., Hardt, M., and Narayanan, A. (2017). Fairness in machine learning. *Nips tutorial*, 1:2.

Bechavod, Y., Jung, C., and Wu, S. Z. (2020). Metric-free individual fairness in online learning. *Advances in neural information processing systems*, 33:11214–11225.

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

Besbes, O. and Zeevi, A. (2015). On the (surprising) sufficiency of linear models for dynamic pricing with demand learning. *Management Science*, 61(4):723–739.

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

Cesa-Bianchi, N., Dekel, O., and Shamir, O. (2013). Online learning with switching costs and other adaptive adversaries. *Advances in Neural Information Processing Systems*, 26.

Chapuis, J. M. (2012). Price fairness versus pricing fairness. *Revenue & Yield Management eJournal*, page 12.

Chen, Q., Jasin, S., and Duenyas, I. (2019). Nonparametric self-adjusting control for joint learning and optimization of multiproduct pricing with finite resource capacity. *Mathematics of Operations Research*, 44(2):601–631.

Chen, X., Zhang, X., and Zhou, Y. (2021). Fairness-aware online price discrimination with nonparametric demand models. *arXiv preprint arXiv:2111.08221*.

Chowdhury, S. R. and Zhou, X. (2022). Differentially private regret minimization in episodic markov decision processes. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 6375–6383.

Cohen, M. C., Elmachtoub, A. N., and Lei, X. (2022). Price discrimination with fairness constraints. *Management Science*.

Cohen, M. C., Lobel, I., and Paes Leme, R. (2020). Feature-based dynamic pricing. *Management Science*, 66(11):4921–4943.

Cohen, M. C., Miao, S., and Wang, Y. (2021). Dynamic pricing with fairness constraints. *Available at SSRN 3930622*.

Cournot, A. A. (1897). *Researches into the Mathematical Principles of the Theory of Wealth*. Macmillan.

Delage, E. and Iancu, D. A. (2015). Robust multistage decision making. In *The operations research revolution*, pages 20–46. Informs.

Dwork, C., Hardt, M., Pitassi, T., Reingold, O., and Zemel, R. (2012). Fairness through awareness. In *Proceedings of the 3rd innovations in theoretical computer science conference*, pages 214–226.

Elfin, R. (1988). The future use of unconscionability and impracticability as contract doctrines. *Mercer L. Rev.*, 40:937.

Eyster, E., Madarász, K., and Michaillat, P. (2021). Pricing under fairness concerns. *Journal of the European Economic Association*, 19(3):1853–1898.

Fan, J., Guo, Y., and Yu, M. (2021). Policy optimization using semiparametric models for dynamic pricing. *arXiv preprint arXiv:2109.06368*.

Frey, B. S. and Pommerehne, W. W. (1993). On the fairness of pricing—an empirical survey among the general population. *Journal of Economic Behavior & Organization*, 20(3):295–307.

Garcelon, E., Perchet, V., Pike-Burke, C., and Pirotta, M. (2021). Local differential privacy for regret minimization in reinforcement learning. *Advances in Neural Information Processing Systems*, 34:10561–10573.

Gupta, S. and Kamble, V. (2021). Individual fairness in hindsight. *J. Mach. Learn. Res.*, 22(144):1–35.

Hardt, M., Price, E., and Srebro, N. (2016). Equality of opportunity in supervised learning. *Advances in neural information processing systems*, 29.

Javanmard, A. and Nazerzadeh, H. (2019). Dynamic pricing in high-dimensions. *The Journal of Machine Learning Research*, 20(1):315–363.

Joseph, M., Kearns, M., Morgenstern, J. H., and Roth, A. (2016). Fairness in learning: Classic and contextual bandits. *Advances in neural information processing systems*, 29.

Kaufmann, P. J., Ortmeyer, G., and Smith, N. C. (1991). Fairness in consumer pricing. *Journal of Consumer Policy*, 14(2):117–140.

Kleinberg, R. and Leighton, T. (2003). The value of knowing a demand curve: Bounds on regret for online posted-price auctions. In *IEEE Symposium on Foundations of Computer Science (FOCS-03)*, pages 594–605. IEEE.

Lai, T. L. and Robbins, H. (1985). Asymptotically efficient adaptive allocation rules. *Advances in applied mathematics*, 6(1):4–22.

Leme, R. P. and Schneider, J. (2018). Contextual search via intrinsic volumes. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS-18)*, pages 268–282. IEEE.

Liu, A., Leme, R. P., and Schneider, J. (2021). Optimal contextual pricing and extensions. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA-21)*, pages 1059–1078. SIAM.

Maestre, R., Duque, J., Rubio, A., and Arévalo, J. (2018). Reinforcement learning for fair dynamic pricing. In *Proceedings of SAI Intelligent Systems Conference*, pages 120–135. Springer.

Priester, A., Robbert, T., and Roth, S. (2020). A special price just for you: Effects of personalized dynamic pricing on consumer fairness perceptions. *Journal of Revenue and Pricing Management*, 19(2):99–112.

Qiao, D. and Wang, Y.-X. (2022a). Near-optimal differentially private reinforcement learning. *arXiv preprint arXiv:2212.04680*.

Qiao, D. and Wang, Y.-X. (2022b). Offline reinforcement learning with differential privacy. *arXiv preprint arXiv:2206.00810*.

Qiao, D., Yin, M., Min, M., and Wang, Y.-X. (2022). Sample-efficient reinforcement learning with loglog (T) switching cost. *arXiv preprint arXiv:2202.06385*.

Richards, T. J., Liaukonyte, J., and Streletskaya, N. A. (2016). Personalized pricing and price fairness. *International Journal of Industrial Organization*, 44:138–153.

Tran, C., Fioretto, F., Van Hentenryck, P., and Yao, Z. (2021). Decision making with differential privacy under a fairness lens. In *IJCAI*, pages 560–566.

Vietri, G., Balle, B., Krishnamurthy, A., and Wu, S. (2020). Private reinforcement learning with pac and regret guarantees. In *International Conference on Machine Learning*, pages 9754–9764. PMLR.

Wang, Y., Chen, B., and Simchi-Levi, D. (2021). Multimodal dynamic pricing. *Management Science*.

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

Xu, J. and Wang, Y.-X. (2021). Logarithmic regret in feature-based dynamic pricing. *Advances in Neural Information Processing Systems*, 34.

Xu, J. and Wang, Y.-X. (2022). Towards agnostic feature-based dynamic pricing: Linear policies vs linear valuation with unknown noise. *International Conference on Artificial Intelligence and Statistics (AISTATS)*.

Yang, Z., Fu, X., Gao, P., and Chen, Y.-J. (2022). Fairness regulation of prices in competitive markets. *Available at SSRN*.

Yu, H., Neely, M., and Wei, X. (2017). Online convex optimization with stochastic constraints. *Advances in Neural Information Processing Systems*, 30.

Zhao, F., Qiao, D., Redberg, R., Agrawal, D., Abbadi, A. E., and Wang, Y.-X. (2022a). Differentially private linear sketches: Efficient implementations and applications. *arXiv preprint arXiv:2205.09873*.

Zhao, X., Li, L., and Wang, Y.-X. (2022b). Provably confidential language modelling. *arXiv preprint arXiv:2205.01863*.

APPENDIX

A   More Related Works

This part serves as a complement to Section 2.

Dynamic Pricing   Single product dynamic pricing problem has been well-studied through Kleinberg and Leighton (2003); Besbes and Zeevi (2009); Wang et al. (2014); Chen et al. (2019); Wang et al. (2021). The crux is to learn and approach the optimal of a revenue curve (in continuous price space) from Boolean-censored feedback. Kleinberg and Leighton (2003) studies this problem under three settings: noise-free, infinitely smooth and stochastic/adversarial valuation assumptions, and proves $\Theta(\log \log T), \Theta(\sqrt{T})$ and $\Theta(T^{\frac{2}{3}})$ minimax regret bounds, sequentially. Wang et al. (2021) further proves a $\Theta(T^{\frac{K+1}{2K+1}})$ minimax regret bound for K^{th}-smooth revenue functions. The key to these problem is to parametrize the revenue curve and estimate these parameters to bound the error. Therefore, a smoothness assumption would reduce the number of local parameters necessary for trading-off the information loss caused by randomness.

There are recently a variety of works on feature-based dynamic pricing problems (Cohen et al., 2020; Leme and Schneider, 2018; Javanmard and Nazerzadeh, 2019; Xu and Wang, 2021; Liu et al., 2021; Fan et al., 2021; Xu and Wang, 2022) where the sellers are asked to sell different products and set prices according to each of their features. All of these literatures listed above adopt a linear feature, i.e., customer's (expected) valuations are linearly dependent on the feature with fixed parameters. In specific, there are a few different problem settings:

1. When customer's valuations are deterministic, Cohen et al. (2020) proposes a binary-search-based algorithm and achieves a $O(\log T)$ regret, which is later improved by Leme and Schneider (2018) with a better $O(d \log \log T)$ regret that matches the information-theoretic lower bound even for the single product pricing problem.

2. When customers' valuations are linear and noisy. For the setting where the noise distribution is known to the seller, Javanmard and Nazerzadeh (2019) and Xu and Wang (2021) achieves the optimal $O(d \log T)$ regret in stochastic and adversarial settings respectively. Javanmard and Nazerzadeh (2019) further achieves $O(\sqrt{dT})$ regret for unknown but parametric noise distributions, which is also optimal according to Xu and Wang (2021). For totally unknown noise, Xu and Wang (2022) shows a $O(T^{\frac{3}{4}})$ regret in the non-continuous case while Fan et al. (2021) proves a $O(T^{\frac{2K+1}{4K-1}})$ for K^{th}-smooth noises, and both of them are not likely to be optimal (where the current lower bounds are $\Omega(T^{\frac{2}{3}})$ and $\Omega(T^{\frac{K+1}{2K+1}})$ respectively).

Fairness in Machine Learning   Ethic-related problems have attracted more and more concerns in contemporary studies of decision-making, including privacy (Qiao and Wang, 2022b; Vietri et al., 2020; Garcelon et al., 2021; Chowdhury and Zhou, 2022; Qiao and Wang, 2022a; Zhao et al., 2022a; Tran et al., 2021; Zhao et al., 2022b), fairness (Dwork et al., 2012; Hardt et al., 2016; Cohen et al., 2022) and robustness (Delage and Iancu, 2015). Fairness is a long-existing topic that has been extensively studied. In machine learning society, fairness is defined from different perspectives. On the one hand, the concept of *group fairness* requires different groups to receive identical treatment in statistics. In a classification problem, for instance, there are mainly two different types of group fairness: (1) A "demographic parity" (Dwork et al., 2012) that requires the outcome of a classifier to be statistically independent to the group information, and (2) an "equalized odds" (including "equal opportunity" as a relaxation) (Hardt et al., 2016) that requires the prediction of a classifier to be *conditionally* independent to the group information given the true label. In Agarwal et al. (2018), these probabilistic constraints are further modified as linear constraints, and therefore the fair classification problem is reduced to a cost-sensitive classification problem. It is worth mentioning that Agarwal et al. (2018) *allows* an ϵ_k-unfairness due to the learning error and *assumes $\epsilon_k = O(n^{-\alpha})$* with some $\alpha \leq \frac{1}{2}$, while we *quantify* the learning-caused unfairness and *upper and lower bound* the cumulative unfairness without pre-assuming its scale.

On the other hand, (Dwork et al., 2012) also proposes the concept of "individual fairness" (or "Lipschitz property") where the difference of treatments toward two individuals should be upper bounded by a distance metric of their intrinsic features, i.e., $D(\mu_x, \mu_y) \leq d(x, y)$ where x, y are features and μ_x, μ_y are the distributions of actions onto x and y respectively. The notion "time fairness" is often considered as individual fairness as well. For a more inclusive review on different definitions of fairness in machine learning, please refer to Barocas et al. (2017).

Fairness in Online Learning   Besides existing works on general machine learning fairness, there are some works that study online-learning or bandit problems. This is similar to our setting as we adopt an online pricing process. Among these works, Joseph et al. (2016) studies multi-armed and contextual bandits with fairness constraints. Their non-contextual setting is related to our works as our pricing problem can also be treated as a bandit. Their definition of δ-fairness is defined as comparisons among probabilities of taking actions, which is similar to our definition on procedural fairness. However, their fairness definitions are defined from the perspective of arms (i.e. actions): better actions worth larger probability to take. In comparison, our fairness definitions are more on the results: different groups share the same expected prices. Bechavod et al. (2020) studies an online learning (in specific, an online classification) problem with unknown and non-parametric constraints on individual fairness at each round. They develop an adaptive algorithm that guarantees an $O(\sqrt{T})$ regret as well as an $O(\sqrt{T})$ cumulative fairness loss. However, their problem settings are quite different from ours. Primarily, they assume individual fairness as a constraint, while our fairness definitions are indeed group fairness. Also, their online classification problem is different from our online pricing problem as they have full access to the regret function while we even do not have full-information reward (i.e., which is Boolean-censored). Similarly, their fairness loss is accessible although the unfair pairs of (τ_1, τ_2) are not fully accessible, while in our settings we do not know the $S(\pi; F_1, F_2)$ function at all. Besides, we have to satisfy two constraints at one time and one of them (the substantive fairness) is highly non-convex. Gupta and Kamble (2021) studies an online learning problem with two different sorts of individual fairness constraints over time: a "fairness-across-time" (FT) and a "fairness-in-hindsight" (FH). They show that it is necessary to have a linear regret under FT constraints, and they also propose a CAFE algorithm that achieves an optimal regret under FH constraints.

Despite the specific properties of fairness constraints, we may also consider the framework of constraint online learning. Yu et al. (2017) studies an online convex optimization (OCO) problem with stochastic constraints, which might be applicable to online fair learning. However, their problem settings and methodologies are largely different from ours: Firstly, their constraints are assumed convex while our substantive fairness constraint (i.e., the $S(\pi; F_1, F_2)$ function) is highly non-convex. Also, they have a direct access to the realized objective function $f^t(x_t)$ at each time while our pricing problem only has a Boolean-censored feedback. More importantly, Yu et al. (2017) assumes the availability of unbiased samples on constraint-related variables. In specific, their constraints are roughly $g_k(x) < 0$, and by the end of each period t they receive an unbiased sample of $g_k(x_t)$ for the x_t they have taken. On the contrary, we do not have any unbiased sample of $S(\pi, F_1, F_2)$ at each time, since there is only one customer from one of the two groups. Therefore, we cannot make use of their results in our problem setting.

Fairness in Pricing   There are some works also studying the fairness problem in dynamic pricing besides of the works we discuss in Section 2. For example, Richards et al. (2016) discusses some fairness issues regarding personalized pricing from the perspective of econometrics. Eyster et al. (2021) studies a phenomenon where customers would mistakenly attribute the cost increases to a time unfairness, and they propose methods to release customer's feeling of such unfairness by adjusting prices correspondingly. Chapuis (2012) looked into two fairness concerns called *price fairness* and *pricing fairness*, which indicates the *distributional* and *procedural* fairness of the pricing process respectively, from the **seller**'s perspective. In fact,

their price fairness is more likely to be our *procedural fairness* definition although it is not in their paper. This is because that we are considering the fairness from **customers**' perspective, where their observations on prices serve as a procedure of their decision process and their decision on whether or not to buy is actually indicating the fairness of results. There are more interesting works as is listed in Section 2, and we refer the readers to Chen et al. (2021) where there is a more comprehensive review on pricing and fairness.

B Proof Details

B.1 Proof of Theorem 6

Proof. First of all, we specify the parameters initialized in Algorithm 1: Let $C_q = 3\max\{\frac{1}{q}, \frac{1}{1-q}\}$ and $c_t = \max\{3, \sqrt{\frac{3}{\hat{F}_{\min}}}\}$. For $k = 1, 2, \ldots$, let $\tau_k = \frac{28C_q}{3} \cdot d\sqrt{T}\log(\frac{16d\log T}{\epsilon}) \cdot 2^k$, $\delta_{k,r} = 4c_t \log\frac{16d\log T}{\epsilon}d^{\frac{3}{2}}\sqrt{\frac{C_q}{\tau_k}}$, $\delta_{k,s} = \frac{32c_t}{(\hat{F}_{\min})^2}\log\frac{16d\log T}{\epsilon}d^{\frac{3}{2}}\sqrt{\frac{C_q}{\tau_k}}$. Denote $\bar{F}_{0,1} := \frac{N_{0,1}}{M_{0,1}}$ and $\bar{F}_{0,2} := \frac{N_{0,2}}{M_{0,2}}$, where $M_{0,e}$ and $N_{0,e}$ are counters defined in Section 4.1.1. Also denote $F_{0,1} := F_1(d)$ and $F_{0,2} := F_2(d)$ for simplicity. Now we prove that $\hat{F}_{\min} \leq F_{\min}$ with high probability. Recall than $C_q = 3\max\{\frac{1}{q}, \frac{1}{1-q}\}$. Recall that $\tau_0 = 2\log T\log\frac{16}{\epsilon}$. According to Hoeffding's Inequality, we have:

$$\Pr[|\bar{F}_{0,1} - F_{0,1}| \geq \frac{F_{0,1}}{2}] \leq 2\exp\{-2(\frac{F_{0,1}}{2})^2 \cdot \frac{1}{C_q}\tau_0\}$$

$$\Leftrightarrow \quad \Pr[\frac{F_{0,1}}{2} \leq \frac{3F_{0,1}}{2}] \geq 1 - 2\exp\{-(F_{0,1})^2\frac{1}{C_q}\log T\log\frac{16}{\epsilon}\} \tag{11}$$

$$\geq 1 - \frac{\epsilon}{8}.$$

Here the last inequality comes from Assumption 1 that $F_{0,1} \geq F_{\min} > 0$ and therefore we have $(F_{\min})^2\frac{1}{C_q}\log T \geq 1$ with large T. Therefore, we have $\frac{F_{0,1}}{2} \leq \bar{F}_{0,1} \leq \frac{3F_{0,1}}{2}$ with probability at least $1 - \frac{\epsilon}{8}$. Similarly, we have $\frac{F_{0,2}}{2}\bar{F}_{0,2} \leq \frac{3F_{0,2}}{2}$ with probability at least $1 - \frac{\epsilon}{8}$. Therefore, with $\Pr \geq 1 - \frac{\epsilon}{4}$, we have $\hat{F}_{\min} = \frac{1}{2}\min\{\bar{F}_{0,1}, \bar{F}_{0,2}\} \leq \min\{\frac{3F_{0,2}}{4}, \frac{3F_{0,2}}{4}\} = \frac{3}{4}F_{\min} < F_{\min}$.

We define some notations that are helpful to our proof. In epoch k, recall that we have:

$$\tilde{\pi}_{k,i,e} = \underset{\pi \in \Pi_k}{\operatorname{argmax}} \pi^e(i). \tag{12}$$

For simplicity, for every $i \in I_k^e$, denote $\rho_{k,e}(i) := \tilde{\pi}_{k,i,e}^e(i)$ that is the largest probability of choosing price v_i in G_e among all policies in Π_k. For those $i \notin I_k^e$, we find out the largest k' such that $i \in I_{k'}^e$ and let $\rho_{k,e}(i) := \tilde{\pi}_{k',i,e}^e(i)$. According to the fact that $\Pi = \Pi_1 \supset \Pi_2 \supset \ldots \supset \Pi_k \supset \Pi_{k+1} \supset \ldots$, we have

$$\pi^e(i) \leq \rho_{k,e}(i), \forall \pi \in \Pi_k; e = 1, 2; i \in [d]. \tag{13}$$

Next, we prove the following lemmas together by induction over epoch index $k = 1, 2, \ldots$. We firstly state that

Lemma 10. *Recall the optimal policy π_* defined in (5). Before Epoch k, we have $\pi_* \in \Pi_k$ with high probability (the failure probability will be totally bounded at the end of this proof).*

which is natural at $k = 1$ as $\Pi_1 = \Pi$. Now, suppose Lemma 10 holds for $\leq k$, then we have:

Lemma 11 (Number of Choosing v_i in G_e)**.** *For $M_{k,e}(i)$ and $N_{k,e}(i)$ defined in Algorithm 1, for any $e = 1, 2; i \in I_k^e$, with $\Pr \geq 1 - \frac{\epsilon}{2\log T}$ we have:*

$$\frac{\rho_{k,e}(i) \cdot \tau_k}{4d \cdot C_q} \leq M_{k,e}(i),$$

$$|N_{k,e}(i) - M_{k,e}(i) \cdot F_e(i)| \leq c_t \cdot \sqrt{F_e(i) \cdot M_{k,e}(i)} \cdot \log\frac{16d\log T}{\epsilon}. \tag{14}$$

Here $c_t = \max\{3, \sqrt{\frac{3}{\hat{F}_{\min}}}\}$.

Proof of Lemma 11. For any $i \in I_k^e$, there exists a policy $\tilde{\pi}_{k,i,e}$ running in Epoch k for at least $\frac{\tau_k}{|A_k|}$ rounds, and. Therefore, we have $\mathbb{E}[M_{k,e}(i)] \geq \rho_{k,e}(i) \cdot \frac{\tau_k}{|A_k|} \cdot \min\{q, 1-q\} \geq \rho_{k,e}(i) \cdot \frac{\tau_k}{|A_k| \cdot C_q}$. According to Bernstein's Inequality, for any $e = 1, 2; i \in I_k^e$ we have:

$$
\begin{aligned}
&\Pr[|M_{k,e}(i) - \mathbb{E}[M_{k,e}(i)]| \leq \frac{\mathbb{E}[M_{k,e}(i)]}{2}] \\
&\geq 1 - 2\exp\{-\frac{\frac{1}{2}(\frac{\mathbb{E}[M_{k,e}(i)]}{2})^2}{\sum_{t=1}^{\frac{\tau_k}{|A_k|}} \rho_{k,e}(i) \cdot \frac{1}{C_q}(1 - \rho_{k,e}(i) \cdot \frac{1}{C_q}) + \frac{1}{3} \cdot 1 \cdot \frac{\mathbb{E}[M_{k,e}(i)]}{2}}\} \\
&\geq 1 - 2\exp\{-\frac{\frac{1}{8}(\mathbb{E}[M_{k,e}(i)])^2}{\rho_{k,e}(i) \cdot \frac{\tau_k}{|A_k|C_q} + \frac{1}{6} \cdot \mathbb{E}[M_{k,e}(i)]}\} \\
&\geq 1 - 2\exp\{-\frac{\frac{1}{8}(\mathbb{E}[M_{k,e}(i)])^2}{\mathbb{E}[M_{k,e}(i)] + \frac{1}{6} \cdot \mathbb{E}[M_{k,e}(i)]}\} \\
&= 1 - 2\exp\{-\frac{1}{8 \cdot \frac{7}{6}}\mathbb{E}[M_{k,e}(i)]\} \\
&\geq 1 - 2\exp\{-\frac{3}{28}\rho_{k,e}(i) \cdot \frac{\tau_k}{|A_k| \cdot C_q}\} \\
&= 1 - 2\exp\{-\frac{3}{28}\rho_{k,e}(i) \cdot \frac{\frac{28}{3} \cdot d\sqrt{T}\log(\frac{16d\log T}{\epsilon}) \cdot 2^k}{2d \cdot C_q}\} \\
&= 1 - 2\exp\{-\rho_{k,e}(i)\sqrt{T} \cdot \log(\frac{16d\log T}{\epsilon}) \cdot \frac{d \cdot 2^k}{2d}\} \\
&\geq 1 - 2\exp\{-\log(\frac{16d\log T}{\epsilon})\} \\
&= 1 - 2 \cdot \frac{\epsilon}{16d\log T} \\
&= 1 - \frac{\epsilon}{8d\log T}.
\end{aligned}
\tag{15}
$$

Here the second line is because that

$$
\sum_{t=1}^{\tau_k} \mathbb{E}[(\mathbf{1}(\text{choosing } v_i \text{ at time } t))^2] \leq \sum_{t=1}^{\frac{\tau_k}{|A_k|}} \mathbb{E}[(\mathbf{1}(\text{ running } \tilde{\pi}_{k,i,e} \text{ and choosing } v_i \text{ at time } t))]
$$

, the third line is for $1 - \rho_{k,e}(i) \cdot \frac{1}{C_q} \leq 1$, the fourth and sixth line are from $\mathbb{E}[M_{k,e}(i)] \geq \frac{\rho_{k,e}(i) \cdot \tau_k}{|A_k| \cdot C_q}$, the seventh line is by plugging in $\tau_k = \frac{28C_q}{3} \cdot d\sqrt{T}\log(\frac{16d\log T}{\epsilon}) \cdot 2^k$, the eighth line is equivalent transformation and the ninth line is for $\rho_{k,e}(i) \geq \frac{1}{\sqrt{T}}$ according to Line 9 of Algorithm 1. As a result, with probability at least $1 - \frac{\epsilon}{8d\log T}$, we have

$$
M_{k,e}(i) \geq \frac{\mathbb{E}[M_{k,e}(i)]}{2} \geq \frac{\rho_{k,e}(i) \cdot \tau_k}{|A_k| \cdot C_q} \geq \frac{\rho_{k,e}(i) \cdot \tau_k}{4dC_q}.
\tag{16}
$$

Now, we analyze $N_{k,e}(i)$ for $i \in I_k^e$. Again, from Line 15 of Algorithm 1 we know that $N_{k,e}(i) = \sum_{t=1}^{M_{k,e}(i)} \mathbb{1}(v_t \text{ is accepted in } G_e)$. Therefore, we apply Bernstein's Inequality and get:

$$
\begin{aligned}
&\Pr[|N_{k,e}(i) - M_{k,e}(i) \cdot F_e(i)| \geq c_t \cdot \sqrt{M_{k,e}(i) \cdot F_e(i)} \log \frac{16d\log T}{\epsilon}] \\
&\leq 2\exp\{-\frac{\frac{1}{2}c_t^2 \cdot M_{k,e}(i)F_e(i)(\log \frac{16d\log T}{\epsilon})^2}{M_{k,e}(i)F_e(i)\log \frac{16d\log T}{\epsilon}(1 - F_e(i)) + \frac{1}{3} \cdot (c_t \cdot \sqrt{M_{k,e}(i) \cdot F_e(i)} \log \frac{16d\log T}{\epsilon})}\} \\
&\leq 2\exp\{-\frac{\frac{1}{2}c_t^2 \log \frac{16d\log T}{\epsilon}}{1 + \frac{c_t}{3}}\} \\
&\leq \frac{\epsilon}{8\log T}.
\end{aligned}
$$

Here the last line is by $c_t = \max\{3, \sqrt{\frac{3}{\hat{F}_{\min}}}\} \geq 3$ and therefore $\frac{\frac{1}{2}c_t^2}{1+\frac{c_t}{3}} \geq 1$. As a result, for $e = 1, 2; i \in I_k^e$, with $\Pr \geq 1 - \frac{\epsilon}{8 \log T}$ we have

$$|N_{k,e}(i) - M_{k,e}(i) \cdot F_e(i)| \leq c_t \cdot \sqrt{M_{k,e}(i)F_e(i)} \log \frac{16d \log T}{\epsilon}.$$

That is to say,

$$\begin{aligned}
|\bar{F}_{k,e}(i) - F_e(i)| &= |\max\{\frac{N_{k,e}(i)}{M_{k,e}(i)}, \hat{F}_{\min}\} - F_e(i)| \\
&\leq |\frac{N_{k,e}(i)}{M_{k,e}(i)}| \\
&\leq c_t \cdot \sqrt{\frac{F_e(i)}{M_{k,e}(i)}} \log \frac{16d \log T}{\epsilon} \\
&\leq c_t \log \frac{16d \log T}{\epsilon} \sqrt{F_e(i)} \cdot \sqrt{\frac{4dC_q}{\rho_{k,e}(i)\tau_k}}.
\end{aligned}$$

Here the first line is by definition of $\bar{F}_{k,e}$, the second line is because $\hat{F}_{\min} \leq F_{\min} \leq F_e(i)$, the third line is by the inequality above and the last line is by (16). Therefore, with $\Pr \geq 1 - \frac{\epsilon}{2 \log T}$, (14) holds for $e = 1, 2$ and for $\forall i \in I_k^e$. ∎

Given Lemma 11, we have the following corollary directly:

Corollary 12 (Estimation Error of $\bar{F}_{k,e}(i)$). *Assume that Lemma 11 holds. For $\bar{F}_{k,e}(i) = \max\{\frac{N_{k,e}(i)}{M_{k,e}(i)}, \hat{F}_{\min}\}$ defined in Algorithm 1, for any $e = 1, 2; i \in I_k^e$, we have:*

$$|\bar{F}_{k,e}(i) - F_e(i)| \leq c_t \cdot \log \frac{16d \log T}{\epsilon} \sqrt{\frac{4dF_e(i)C_q}{\rho_{k,e}(i)\tau_k}}. \tag{17}$$

For simplicity, denote $R(\pi) := R(\pi, F_1, F_2)$, $S(\pi) := S(\pi, F_1, F_2)$, $\hat{R}_k(\pi) := R(\pi, \bar{F}_{k,1}, \bar{F}_{k,2})$ and $\hat{S}_k(\pi) := S(\pi, \bar{F}_{k,1}, \bar{F}_{k,2})$. Based on Corollary 12, we can bound the estimation error of $\hat{R}_k(\pi)$ and $\hat{S}_k(\pi)$ by the following lemma:

Lemma 13 (Estimation Error of R and S Functions). *Given Lemma 11, we have:*

$$\begin{aligned}
|R(\pi) - \hat{R}_k(\pi)| &\leq \frac{\delta_{k,r}}{2}, \\
|S(\pi) - \hat{S}_k(\pi)| &\leq \frac{\delta_{k,s}}{2}.
\end{aligned} \tag{18}$$

Here $\delta_{k,r} = 4c_t \log \frac{16d \log T}{\epsilon} d^{\frac{3}{2}} \sqrt{\frac{C_q}{\tau_k}}$ and $\delta_{k,s} = \frac{32c_t}{\hat{F}_{\min}^2} \log \frac{16d \log T}{\epsilon} d^{\frac{3}{2}} \sqrt{\frac{1}{\tau_k}}$ as is defined in Theorem 6.

Proof of Lemma 13. First of all, we show that for any $e = 1, 2; i = 1, 2, \ldots, d$ and for any $\pi \in \Pi_k$,

$$|\bar{F}_{k,e}(i) - F_e(i)| \cdot \pi^e(i) \leq c_t \cdot \log \frac{16d \log T}{\epsilon} \sqrt{\frac{4dC_q}{\tau_k}}. \tag{19}$$

In fact, when $i \in I_k^e$, according to Lemma 11 we have

$$\begin{aligned}
|\bar{F}_{k,e}(i) - F_e(i)| \cdot \pi^e(i) &\leq |\bar{F}_{k,e}(i) - F_e(i)| \cdot \rho_{k,e}(i) \\
&\leq c_t \cdot \log \frac{16d \log T}{\epsilon} \sqrt{\frac{4dF_e(i)C_q}{\rho_{k,e}(i)\tau_k}} \cdot \rho_{k,e}(i) \\
&\leq c_t \cdot \log \frac{16d \log T}{\epsilon} \sqrt{\frac{4dC_q \cdot (\rho_{k,e}(i))}{\tau_k}} \\
&\leq c_t \cdot \log \frac{16d \log T}{\epsilon} \sqrt{\frac{4dC_q}{\tau_k}}.
\end{aligned} \tag{20}$$

When $i \notin I_k^e$, we know that $\rho_{k,e}(i) \le \frac{1}{\sqrt{T}}$ and thus $\pi^e(i) \le \frac{1}{\sqrt{T}}, \forall \pi \in \Pi_k$ according to (13). Also, since $\pi_* \in \Pi_k$ by induction, we know that $\pi_*^e \le \rho_{k,e}(i) \le \frac{1}{\sqrt{T}}$. Therefore, we have

$$
\begin{aligned}
|\bar{F}_{k,e}(i) - F_e(i)| \cdot \pi^e(i) &\le |\bar{F}_{k,e}(i) - F_e(i)| \cdot \rho_{k,e}(i) \\
&\le |\bar{F}_{k,e}(i)\pi^e(i) - F_e(i)| \cdot \frac{1}{\sqrt{T}} \\
&\le 1 \cdot \frac{1}{\sqrt{T}} \\
&\le c_t \cdot \log \frac{16d \log T}{\epsilon} \sqrt{\frac{4dC_q}{\tau_k}}.
\end{aligned}
\tag{21}
$$

Here we assume that $\log \frac{16d \log T}{\epsilon} > 1$ without losing of generality (i.e., T is sufficiently large and ϵ can be arbitrarily close to zero), and the last inequality comes from $c_t \ge 3 > 1$ and $4d \ge 1$ and $\tau_k \le T$. Combining (20) and (21), we know that (19) holds for all $e = 1, 2; i \in [d]$. Remember that $R(\pi) = q \cdot \sum_{i=1}^{d} F_1(i)\pi^1(i) + (1-q) \cdot \sum_{j=1}^{d} F_2(j)\pi^2(j)$ and that $\hat{R}_k(\pi) = q \cdot \sum_{i=1}^{d} \bar{F}_{k,1}(i)\pi^1(i) + (1-q) \cdot \sum_{j=1}^{d} \hat{F}_{k,2}(j)\pi^2(j)$. Therefore, we may bound the error between $\hat{R}_k(\pi)$ and $R(\pi)$. For $\forall \pi \in \Pi_k$, we have

$$
\begin{aligned}
|\hat{R}_k(\pi) - R(\pi)| &\le q \cdot \sum_{i=1}^{d} |\bar{F}_{k,1}(i) - F_1(i)| \cdot \pi^1(i) + (1-q) \cdot \sum_{j=1}^{d} |\bar{F}_{k,2}(j) - F_2(j)| \cdot \pi^2(i) \\
&\le q \cdot \sum_{i=1}^{d} c_t \cdot \log \frac{16d \log T}{\epsilon} \sqrt{\frac{4dC_q}{\tau_k}} + (1-q) \cdot \sum_{j=1}^{d} c_t \cdot \log \frac{16d \log T}{\epsilon} \sqrt{\frac{4dC_q}{\tau_k}} \\
&= c_t \cdot \log \frac{16d \log T}{\epsilon} \sqrt{\frac{4dC_q}{\tau_k}} \cdot d \\
&= \frac{\delta_{k,r}}{2}.
\end{aligned}
\tag{22}
$$

Similarly, for the error between $\hat{S}_k(\pi)$ and $S(\pi)$ as $\pi \in \Pi_k$, we have:

$$
\begin{aligned}
|\hat{S}_k(\pi) - S(\pi)| =& \|\frac{\mathbf{v}^\top \hat{F}_{k,1}\pi^1}{\mathbb{1}^\top \hat{F}_{k,1}\pi^1} - \frac{\mathbf{v}^\top \hat{F}_{k,2}\pi^2}{\mathbb{1}^\top \hat{F}_{k,2}\pi^2}| - |\frac{\mathbf{v}^\top F_1\pi^1}{\mathbb{1}^\top F_1\pi^1} - \frac{\mathbf{v}^\top F_2\pi^2}{\mathbb{1}^\top F_2\pi^2}\| \\
\leq& |\frac{\mathbf{v}^\top \hat{F}_{k,1}\pi^1}{\mathbb{1}^\top \hat{F}_{k,1}\pi^1} - \frac{\mathbf{v}^\top \hat{F}_{k,2}\pi^2}{\mathbb{1}^\top \hat{F}_{k,2}\pi^2} - (\frac{\mathbf{v}^\top F_1\pi^1}{\mathbb{1}^\top F_1\pi^1} - \frac{\mathbf{v}^\top F_2\pi^2}{\mathbb{1}^\top F_2\pi^2})| \\
\leq& |\frac{\mathbf{v}^\top \hat{F}_{k,1}\pi^1}{\mathbb{1}^\top \hat{F}_{k,1}\pi^1} - \frac{\mathbf{v}^\top F_1\pi^1}{\mathbb{1}^\top F_1\pi^1}| + |\frac{\mathbf{v}^\top \hat{F}_{k,2}\pi^2}{\mathbb{1}^\top \hat{F}_{k,2}\pi^2} - \frac{\mathbf{v}^\top F_2\pi^2}{\mathbb{1}^\top F_2\pi^2}| \\
=& \sum_{e=1}^2 |\frac{\mathbf{v}^\top \hat{F}_{k,e}\pi^e}{\mathbb{1}^\top \hat{F}_{k,e}\pi^e} - \frac{\mathbf{v}^\top F_e\pi^e}{\mathbb{1}^\top F_e\pi^e}| \\
=& \sum_{e=1}^2 |\frac{(\mathbf{v}^\top \hat{F}_{k,e}\pi^e)(\mathbb{1}^\top F_e\pi^e) - (\mathbf{v}^\top F_e\pi^e)(\mathbb{1}^\top \hat{F}_{k,e}\pi^e)}{(\mathbb{1}^\top \hat{F}_{k,e}\pi^e)(\mathbb{1}^\top F_e\pi^e)}| \\
=& \sum_{e=1}^2 \frac{|(\pi^e)^\top (\hat{F}_{k,e} - F_e)\mathbf{v} \cdot (\mathbb{1}^\top F_e\pi^e) + (\mathbf{v}^\top F_e\pi^e)\mathbb{1}^\top (F_e - \hat{F}_{k,e})\pi^e|}{|(\mathbb{1}^\top \hat{F}_{k,e}\pi^e)| \cdot |(\mathbb{1}^\top F_e\pi^e)|} \\
\leq& \sum_{e=1}^2 \frac{(\mathbb{1}^\top F_e\pi^e) \cdot \sum_{i=1}^d \pi^e(i)(\bar{F}_{k,e}(i) - F_e(i))v_i + (\mathbf{v}^\top F_e\pi^e) \cdot \sum_{j=1}^d 1 \cdot (F_e(j) - \hat{F}_{k,e}(j)) \cdot \pi^e(j)}{|\hat{F}_{\min}| \cdot |\hat{F}_{\min}|} \\
\leq& \sum_{e=1}^2 \frac{1 \cdot \sum_{i=1}^d \pi^e(i)|\bar{F}_{k,e}(i) - F_e(i)| \cdot 1 + 1 \cdot \sum_{j=1}^d 1 \cdot |F_e(j) - \hat{F}_{k,e}(j)|\pi^e(j)}{(\hat{F}_{\min})^2} \\
\leq& \frac{1}{(\hat{F}_{\min})^2} \sum_{e=1}^2 (\sum_{i=1}^d c_t \cdot \log \frac{16d\log T}{\epsilon} \sqrt{\frac{4dC_q}{\tau_k}} + \sum_{j=1}^d c_t \cdot \log \frac{16d\log T}{\epsilon} \sqrt{\frac{4dC_q}{\tau_k}}) \\
=& \frac{1}{(\hat{F}_{\min})^2} \cdot 2d \cdot c_t \cdot \log \frac{16d\log T}{\epsilon} \sqrt{\frac{4dC_q}{\tau_k}} \\
\leq& \frac{\delta_{k,s}}{2}.
\end{aligned}
$$

$$(23)$$

$\blacksquare$

Since we have $\hat{S}_k(\pi) \leq \delta_{k,s}, \forall \pi \in \Pi_{k+1}$ by definition in Algorithm 1, we know that for any policy $\pi \in \Pi_{k+1}$,

$$
\begin{aligned}
S(\pi) =& \hat{S}_k(\pi) + (S(\pi) - \hat{S}_k(\pi)) \\
\leq& \hat{S}_k(\pi) + |S(\pi) - \hat{S}_k(\pi)| \\
\leq& \delta_{k,s} + \frac{\delta_{k,s}}{2} \\
\leq& 2\delta_{k,s}.
\end{aligned}
$$

$$(24)$$

Therefore, any policy remaining in Π_{k+1} suffers at most $2\delta_{k,s}$ unfairness. Now let us bound the regret of any policy in Π_{k+1}. Here we firstly propose a lemma.

Lemma 14 (Small Relaxation Gain). *Recall that π_* is the solution to (5). Define a $\pi_{\delta,*}$ as follows.*

$$
\begin{aligned}
\pi_{\delta,*} =& \operatorname*{argmax}_{\pi \in \Pi} R(\pi) \\
& s.t. \qquad S(\pi) \leq \delta.
\end{aligned}
$$

$$(25)$$

Then there exists a constant $L \in \mathbb{R}_+$ such that $R(\pi_{\delta,*}) - R(\pi_*) \leq \frac{L}{2} \cdot \delta$.

We leave the proof of Lemma 14 to the end of this section. Given Lemma 14 and the previous Lemma 13, we have:

$$
\begin{aligned}
\hat{R}(\hat{\pi}_{k,*}) - \hat{R}(\pi_*) =& \hat{R}(\hat{\pi}_{k,*}) - R(\hat{\pi}_{k,*}) + R(\hat{\pi}_{k,*}) - R(\pi_{2\delta_{k,s},*}) + R(\pi_{2\delta_{k,s}}, *) - R(\pi_*) + R(\pi_*) - \hat{R}(\pi_*) \\
\leq& |\hat{R}(\hat{\pi}_{k,*}) - R(\hat{\pi}_{k,*})| + (R(\hat{\pi}_{k,*}) - R(\pi_{2\delta_{k,s},*})) + (R(\pi_{2\delta_{k,s}}, *) - R(\pi_*)) + |R(\pi_*) - \hat{R}(\pi_*)| \\
\leq& \frac{\delta_{k,r}}{2} + 0 + \frac{L}{2} \cdot 2\delta_{k,s} + \frac{\delta_{k,r}}{2} \\
=& \delta_{k,r} + L \cdot \delta_{k,s}
\end{aligned}
\tag{26}
$$

By definition of Π_{k+1} at (8), we know that $\pi_* \in \Pi_{k+1}$, which holds Lemma 10 at $k+1$ and therefore completes the induction. As a result, all Lemma 10, Lemma 11, Lemma 13 and Lemma 14 holds for all $k = 1, 2, \ldots$. As a result, we may calculate the total regret and substantive unfairness as follows.

For the regret, we may divide the whole time horizon T into three stages:

1. Stage 0: Before epochs where we propose v_d for $\tau_0 = 2 \log T \log \frac{16}{\epsilon}$ rounds in either G_1 or G_2. The regret for this stage is $O(\log T \log \frac{1}{\epsilon})$.

2. Stage 1: Epoch 1 where we try every price for $2 \cdot \frac{\tau_1}{2d}$ rounds in either G_1 or G_2. The regret for this stage is $O(\tau_1) = O(d\sqrt{T} \log \frac{\log T}{\epsilon})$.

3. Stage 2: Epoch $k = 2, 3, \ldots$. In each epoch k, every policy π we run satisfies $\pi \in \Pi_k$. Therefore, for any π running in Epoch $k = 2, 3, \ldots$, we have

$$
\begin{aligned}
R(\pi_*) - R(\pi) =& (R(\pi_*) - \hat{R}_{k-1}(\pi_*)) + (\hat{R}_{k-1}(\pi_*) - \hat{R}_{k-1}(\pi)) + (\hat{R}_{k-1}(\pi) - R(\pi)) \\
\leq& \frac{\delta_{k-1,r}}{2} + (\delta_{k-1,r} + L \cdot \delta_{k-1,s}) + \frac{\delta_{k-1,r}}{2} \\
=& 2\delta_{k-1,r} + L \cdot \delta_{k-1,s}.
\end{aligned}
\tag{27}
$$

The second line is by definition of Π_k for $k \geq 2$ and by Lemma 13. Suppose there are K epochs in total, and then we know that:

$$
T \geq \sum_{k=1}^{K} \tau_k = \frac{28C_q}{3} \cdot d\sqrt{T} \cdot \log \frac{16d \log T}{\epsilon} \cdot \sum_{k=1}^{K} 2^k.
$$

Solve the equaltion above and we get $K = O(\log \frac{\sqrt{T}}{d \log \frac{d \log T}{\epsilon}})$ and $K \leq \frac{1}{2} \log T$. Therefore, the total regret of Stage 2 is

$$
Reg = O(\sum_{k=2}^{K} \tau_k \cdot (2\delta_{k-1,r} + L \cdot \delta_{k-1,s})) = O(\sqrt{T} \cdot d^{\frac{3}{2}} \log \frac{d \log T}{\epsilon})
\tag{28}
$$

Add the regret of all three stages above, we get that the total regret is $O(\sqrt{T} \cdot d^{\frac{3}{2}} \log \frac{d \log T}{\epsilon})$.

For the unfairness, we derive it similarly in three stages:

1. Stage 0: Before epochs where we propose v_d for $\tau_0 = 2 \log T \log \frac{16}{\epsilon}$ rounds in either G_1 or G_2. The unfairness for this stage is 0 as we always propose the same price to both groups.

2. Stage 1: Epoch 1 where we try every price for $2 \cdot \frac{\tau_1}{2d}$ rounds in either G_1 or G_2. The regret for this stage is 0 as well.

3. Stage 2: Epoch $k = 2, 3, \ldots$. In each epoch k, every policy π we run satisfies $\pi \in \Pi_k$. Therefore, for any π running in Epoch $k = 2, 3, \ldots$, we have

$$
\begin{aligned}
S(\pi) =& \hat{S}_{k-1}(\pi) + (S(\pi) - \hat{S}_{k-1}(\pi)) \\
\leq& \delta_{k-1,s} + \frac{\delta_{k-1,s}}{2} \\
\leq& \frac{3\delta_{k-1,s}}{2}.
\end{aligned}
\tag{29}
$$

Here the last line is by definition of Π_k for $k \geq 2$ and by Lemma 13. Therefore, the total unfairness of Stage 2 is

$$
Unf \leq \sum_{k=2}^{K} \tau_k \cdot \frac{3\delta_{k-1,s}}{2} = O(\sqrt{T} \cdot d^{\frac{3}{2}} \log \frac{d \log T}{\epsilon}).
\tag{30}
$$

Therefore, the total substantive unfairness of all three stages is $O(\sqrt{T} \cdot d^{\frac{3}{2}} \log \frac{d \log T}{\epsilon})$ as well.

Finally, we count the probability of failure of all stages. For Stage 0, the failure probability is $\Pr_0 \leq \frac{\epsilon}{4}$. For each epoch, the failure probability is $\Pr_k \leq \frac{\epsilon}{2 \log T}$. Since there are $K \leq \frac{\log T}{2}$ epochs, the total failure probability is $\Pr_{failure} \leq \Pr_0 + K \cdot \Pr_k \leq \frac{\epsilon}{4} + K \cdot \frac{\epsilon}{2 \log T} \leq \frac{\epsilon}{2} < \epsilon$. That is to say, Theorem 6 holds with probability at least $\Pr \geq 1 - \epsilon$. ∎

At the end of this subsection, we prove Lemma 14 as we promised above.

Proof of Lemma 14. Denote any policy $\pi \in \Pi$ as $\pi = (\pi^1, \pi^2)$. For the simplicity of notation, we denote the following functions:

(a) Define $R_1(\pi^1) = \mathbf{v}^\top F_1 \pi^1$;

(b) Define $R_2(\pi^2) = \mathbf{v}^\top F_2 \pi^2$;

(c) Define $S_1(\pi^1) = \frac{\mathbf{v}^\top F_1 \pi^1}{\mathbb{1}^\top F_1 \pi^1}$;

(d) Define $S_2(\pi^2) - \frac{\mathbf{v}^\top F_2 \pi^2}{\mathbb{1}^\top F_2 \pi^2}$.

For $\pi_{\delta,*}$ defined in (10), denote $V_s := S_1(\pi_{\delta,*}^1)$ and $z = S_2(\pi_{\delta,*}^2) - V_s$. Therefore, we know that $V_s \in [v_1, 1]$ (recalling that $v_1 > 0$) and $z \in [-\delta, \delta]$. According to the optimality of $\pi_{\delta,*}$, we have:

$$\pi_{\delta,*} = \underset{\pi \in \Pi, V_s \in [v_1, 1], z \in [-\delta, \delta]}{\text{argmax}} q R_1(\pi^1) + (1-q) R_2(\pi^2)$$
$$\text{s.t.} \quad S_1(\pi^1) = V_s \tag{31}$$
$$S_2(\pi^2) = V_s + z$$

. Consider the constraint $S_2(\pi^2) - V_s \in [-\delta, \delta]$, we can derive the following relaxation:

$$S_2(\pi^2) - V_s \in [-\delta, \delta]$$
$$\Leftrightarrow \quad -\delta \leq \frac{\mathbf{v}^\top F_2 \pi^2}{\mathbb{1}^\top F_2 \pi^2} - V_s \leq \delta \tag{32}$$
$$\Rightarrow \quad -\delta(\mathbb{1}^\top F_2 \pi^2) \leq \mathbf{v}^\top F_2 \pi^2 - V_s \cdot \mathbb{1}^\top F_2 \pi^2 \leq \delta(\mathbb{1}^\top F_2 \pi^2)$$
$$\Rightarrow \quad -\delta \leq \mathbf{v}^\top F_2 \pi^2 - V_s \cdot \mathbb{1}^\top F_2 \pi^2 \leq \delta.$$

This is because $\mathbb{1}^\top F_2 \pi^2 \in [F_{\min}, 1] \subset (0, 1]$. Therefore, we may define $\theta_\delta = (\theta_\delta^1, \theta_\delta^2) \in \Pi$ such that

$$\theta_\delta := \underset{\theta \in \Pi, r, w \in [v_1 \cdot F_{\min}, 1]}{\text{argmax}} q R_1(\theta^1) + (1-q) R_2(\theta^2)$$
$$\text{s.t.} \quad \mathbf{v}^\top F_1 \theta^1 = w$$
$$\cdot \mathbb{1}^\top F_1 \theta^1 = \frac{w}{V_s} \tag{33}$$
$$\mathbf{v}^\top F_2 \theta^2 = r$$
$$-\delta \leq v_1 \cdot \mathbb{1}^\top F_2 \theta^2 - \frac{r \cdot v_1}{V_s} \leq \delta,$$

for any $\theta \geq 0$. Here we make use of the fact that $V_s \in [v_1, 1]$. Notice that $[v_1 \cdot F_{\min}, 1]$ contains all possible r's due to the fact that $F_e(i) > F_{\min}$ and $v_i \geq v_1$ for any $i \in [d]$ and $e \in \{1, 2\}$, then we have $R_2(\theta_\delta^2) \geq R_2(\pi_{\delta,*}^2)$ as a relaxation of conditions, which means that $R(\theta_\delta) \geq R(\pi_{\delta,*})$. Consider another policy π_{start}:

$$\pi_{start} := \underset{\pi \in \Pi}{\text{argmax}} \, q R_1(\pi^1) + (1-q) R_2(\pi^2)$$
$$\text{s.t.} \quad S_1(\pi^1) = V_s \tag{34}$$
$$S_2(\pi^2) = V_s.$$

Therefore, we know that when $\delta = 0$, we have $\theta_0 = \pi_{start}$ exactly. Also, since π_* can also be defined as follows:

$$
\begin{aligned}
\pi_* = \underset{\pi \in \Pi, v_s \in [v_1, 1]}{\operatorname{argmax}} \quad & q R_1(\pi^1) + (1-q) R_2(\pi^2) \\
s.t. \quad & S_1(\pi^1) = v_s \\
& S_2(\pi^2) = v_s.
\end{aligned}
\tag{35}
$$

According to the optimality of π_* over all $v_s \in [v_1, 1]$ while π_{start} is restricted on a specific V_s, we have $R(\pi_*) \geq R(\pi_{start}) = R(\theta_0)$. Recall that we also have $R(\theta_\delta) \geq R(\pi_{\delta,*})$. Therefore, as long as we show that there exists a constant L such that $R(\theta_\delta) - R(\theta_0) \leq \frac{L}{2} \cdot \delta$, then it is sufficient to show that $R(\pi_{\delta,*}) - R(\pi_*) \leq \frac{L}{2} \cdot \delta$.

Denote

$$
\tilde{\theta}_\delta = [(\theta_\delta^1)^\top, w, \frac{w \cdot V_1}{V_s}, (\theta_\delta^2)^\top, r, \frac{r \cdot v_1}{V_s}]^\top \in \mathbb{R}^{2d+4}.
\tag{36}
$$

Of course $\|\tilde{\theta}_\delta\|_1 \leq 1 + w + \frac{w}{V_s} + 1 + r + \frac{r}{V_s} \leq 4 + 2\frac{2}{v_1}$. Denote the domain of $\tilde{\theta}\delta$ as $\mathcal{D}(\delta)$. Therefore, we know that for any $\theta \in \mathcal{D}(\delta)$, we have

$$
\begin{aligned}
\theta &\succeq 0 \\
[\mathbb{1}_d^\top, 0, 0, \dots, 0]\theta &= 1 \\
[0, \dots, 0, 0, 0, \mathbb{1}_d^\top, 0, 0]\theta &= 1 \\
[0, \dots, 0, 1, 0, \dots, 0]\theta &\leq 1 \text{ (for 1 in the } (d+1)^{th} \text{ place)} \\
[0, \dots, 0, 1, 0, \dots, 0]\theta &\leq 1 \text{ (for 1 in the } (d+2)^{th} \text{ place)} \\
[0, \dots, 0, 1, 0]\theta &\leq 1 \\
[0, \dots, 0, 0, 1]\theta &\leq 1
\end{aligned}
\tag{37}
$$

Denote $\tilde{\mathcal{D}}(\delta)$ as the space of all θ satisfying (37), and we know that $\tilde{\mathcal{D}}(\delta) \supseteq \mathcal{D}(\delta)$ and $\tilde{\mathcal{D}}(\delta)$ is a bounded, close and convex set with only linear boundaries. Also, denote the following fixed parameters:

$$
\begin{aligned}
\mathbf{a} &:= [q \cdot (\mathbf{v}^\top F_1), 0, 0, (1-q) \cdot (\mathbf{v}^\top F_2), 0, 0] \in \mathbb{R}^{2d+4} \\
\mathbf{b}_1 &:= [\mathbf{v}^\top F_1, -1, 0, 0, \dots, 0] \in \mathbb{R}^{2d+4} \\
\mathbf{b}_2 &:= [v_1 \mathbb{1}^\top F_1, 0, -1, 0, 0, \dots, 0] \in \mathbb{R}^{2d+4} \\
\mathbf{g} &:= [0, \dots, 0, 0, 0, \mathbf{v}^\top F_2, -1, 0] \in \mathbb{R}^{2d+4} \\
\mathbf{d} &:= [0, \dots, 0, 0, 0, v_1 \cdot \mathbb{1}^\top F_2, 0, -1] \in \mathbb{R}^{2d+4}.
\end{aligned}
\tag{38}
$$

Again, these parameters are all constants under the same problem setting. Given these parameters, for the definition of θ_δ in (33), we may transform that definition into the following one equivalently:

$$
\begin{aligned}
\tilde{\theta}_\delta := \underset{\theta \in \tilde{\mathcal{D}}(\delta)}{\operatorname{argmax}} \quad & \mathbf{a}^\top \theta \\
s.t. \quad & \mathbf{b}_1^\top \theta = 0 \\
& \mathbf{b}_2^\top \theta = 0 \\
& \mathbf{g}^\top \theta = 0 \\
& \mathbf{d}^\top \theta \in [-\delta, \delta].
\end{aligned}
\tag{39}
$$

Since $\tilde{\mathcal{D}}(\delta) \supseteq \mathcal{D}(\delta)$, we know that $\mathbf{a}^\top \tilde{\theta}_\delta \geq R(\theta_\delta)$. Denote

$$
\tilde{\mathcal{D}}_{abg}(\delta) := \{\theta | \theta \in \tilde{\mathcal{D}}(\delta), \mathbf{b}_1^\top \theta = 0, \mathbf{b}_2^\top \theta = 0, \mathbf{g}^\top \theta = 0\},
$$

and we know that $\tilde{\mathcal{D}}_{abg}(\delta)$ is also a bounded close and convex set with only linear boundaries. Therefore, (39) is equivalent to the following definition:

$$
\begin{aligned}
\tilde{\theta}_\delta := \underset{\theta \in \tilde{\mathcal{D}}_{abg}(\delta)}{\operatorname{argmax}} \quad & \mathbf{a}^\top \theta \\
s.t. \quad & \mathbf{d}^\top \theta \in [-\delta, \delta].
\end{aligned}
\tag{40}
$$

Now we present the following lemma.

Lemma 15 (Bounded Shifting). *Given any space $Q \subset \mathbb{R}^n$ that is bounded, close and convex with only linear boundaries, consider the following subset $Q_0 := \{x \in Q, d^\top x = 0\} \neq \emptyset$. Then there exists a constant C_L such that for any $z \in \mathbb{R}$, $Q_z := \{x \in Q, d^\top x = z\}$ and any $\theta_z \in Q_z$, there always exists a $\theta_0 \in Q_0$ such that $\|\theta_z - \theta_0\|_2 \leq C_L \cdot |z|$.*

Proof of Lemma 15. Without loss of generality, we assume that $z > 0$. Denote $Q^+ = Q \cap \{x : d^\top x \geq 0\}$. Because Q is bounded, close and convex with only linear boundaries, the number of vertex of Q^+ must be finite. The vertex set of Q^+ can be decomposed as $V = V_0 + V_1$, where V_0 denotes the vertex such that $d^\top x = 0$ while V_1 denotes the vertex such that $d^\top x > 0$. In addition, $Q_0 = Q^+ \cap \{x : d^\top x = 0\}$ is the cross section while we define $B = \{x : d^\top x = 0\}$.

For each point $x \in Q_0$, we define β_x to be $\min\{$The intersection angle between B and $\overrightarrow{xv}, \ v \in V_1\}$. Due to the fact that β_x is continuous upon x, $\beta_x > 0$ and the domain Q_0 is bounded and close, there exists a $\beta_{\min} > 0$ such that $\beta_x \geq \beta_{\min}, \forall x \in Q_0$. Then we construct a corresponding cone $Cone_x$ for each $x \in Q_0$ such that $Cone_x = \{v : d^\top v \geq 0$, and the intersection angle between B and $\overrightarrow{xv} \geq \beta_{\min}\}$.

Since Q^+ is bounded, close and convex with only linear boundaries, for any point $\theta_z \in Q_z$, there exits $v_1, v_2, \cdots, v_k$ and $a_1, a_2, \cdots, a_k$ such that $v_i \in V, a_i \geq 0, \forall i \in [k]$ and $\sum_{i=1}^k a_i = 1$ and it holds that $\theta_z = \sum_{i=1}^k a_i v_i$. Then according to our construction of the cones, for each selected vertex v_i, there exists a cone $Cone_{t_i}$ such that $t_i \in Q_0$ and $v_i \in Cone_{t_i}$. We claim that $Cone_{\sum_{i=1}^k a_i t_i} = \{\sum_{i=1}^k a_i f_i : f_i \in Cone_{t_i}\}$. Therefore, it holds that $\theta_z = \sum_{i=1}^k a_i v_i \in Cone_{\sum_{i=1}^k a_i t_i}$. Consider this $\theta_0 = \sum_{i=1}^k a_i t_i$, because Q_0 is convex, we have $\theta_0 \in Q_0$. In addition, $\|\theta_0 - \theta_z\|_2 \leq \frac{|z|}{\|d\|_2 \cdot \sin(\beta_{\min})}$, which means by choosing $C_L = \frac{1}{\|d\|_2 \cdot \sin(\beta_{\min})}$, the proof is complete. ∎

Denote $z_\delta := \tilde{\theta}_\delta$ and we know that $|z_\delta| \leq \delta$. In order to apply Lemma 15, we have to ensure that $\tilde{\mathcal{D}}_{abg}(0) \neq \emptyset$. In fact, notice that $\tilde{\theta}_0 \in \tilde{\mathcal{D}}_{abg} \cap \{d^\top = 0\}$. With Lemma 15, there exists a $\hat{\theta}_0 \in \tilde{\mathcal{D}}_{abg}(0)$ such that $\|\tilde{\theta}_\delta - \hat{\theta}_0\|_2 \leq \frac{L}{2}|z| \leq \frac{L}{2}\delta$. As a result, we have:

$$
\begin{aligned}
\mathbf{a}^\top \tilde{\theta}_\delta - \mathbf{a}^\top \hat{\theta}_0 &\leq \|a\|_2 \cdot \|\tilde{\theta}_\delta - \hat{\theta}_0\|_2 \\
&\leq \|a\|_1 \cdot C_L \cdot \delta \\
&\leq (q \cdot \mathbf{v}^\top F_1 \mathbb{1} + (1-q) \cdot \mathbf{v}^\top F_2 \mathbb{1}) \cdot C_L \cdot \delta \\
&:= C_a \cdot C_L \cdot \delta.
\end{aligned}
\tag{41}
$$

By definition of $\tilde{\theta}_\delta$, we know that $\tilde{\theta}_0$ maximizes $\mathbf{a}^\top \theta$ in $\tilde{\mathcal{D}}_{abg}(0)$, which means that $\mathbf{a}^\top \tilde{\theta}_0 \geq \mathbf{a}^\top \hat{\theta}_0$. As a result, we have:

$$
\begin{aligned}
R(\theta_\delta) - R(\theta_0) &= \mathbf{a}^\top \tilde{\theta}_\delta - \mathbf{a}^\top \tilde{\theta}_0 \\
&\leq \mathbf{a}^\top \tilde{\theta}_\delta - \mathbf{a}^\top \hat{\theta}_0 \\
&\leq C_a \cdot C_L \cdot \delta.
\end{aligned}
\tag{42}
$$

Therefore, we have $R(\pi_{\delta,*}) - R(\pi_*) \leq R(\theta_\delta) - R(\pi_{start}) = R(\theta_\delta) - R(\theta_0) \leq C_a \cdot C_L \cdot \delta$. Let $L := 2 \cdot C_a \cdot C_L$ and this holds the lemma. ∎

B.2   Proof of Theorem 8

As is stated in Section 5.2, we may reduce this fair pricing problem to an ordinary online pricing problem with no fairness constraints. Therefore, we only need to prove the following theorem.

Theorem 16 (Regret Lower Bound). *Consider the online pricing problem with T rounds and d fixed prices in $[0, c]$ for $3 \leq d \leq T^{1/3}$ and some constant $c > 0$. Then any algorithm has to suffer at least $\Omega(\sqrt{dT})$ regret.*

Here we mainly adopt the proof roadmap of Kleinberg and Leighton (2003).

Proof. We let $c = 12$ without losing generality. Let $\epsilon = \sqrt{\frac{d}{T}}, l = 1, a_0 = 4l, a_i = (1 + \frac{\epsilon}{l})^i \cdot a_0, i = 1, 2, \ldots, d$, then we have: $4l = a_0 < a_1 < a_2 < \ldots < a_{d-1} < a_d < 12l$.

Define some distributions on the prices $\{a_i\}_{i=1}^d$:

- $\mathbb{P}_0$, with acceptance rates of each price: $P_0 = [\frac{l}{a_1}, \frac{l}{a_2}, \ldots, \frac{l}{a_{d-1}}, \frac{l}{a_d}]^T$, where $P_0(i) = \Pr[y \geq a_i] = \Pr[y > a_{i-1}] = \frac{l}{a_i} < \frac{1}{4}$.

- $\mathbb{P}_j$, with acceptance rates of each price: $P_j = [\frac{l}{a_1}, \frac{l}{a_2}, \ldots, \frac{l}{a_{j-1}}, \frac{l+\epsilon}{a_j}, \frac{l}{a_{j+1}}, \ldots, \frac{l}{a_{d-1}}, \frac{l}{a_d}]^T$, where $P_j(i) = \frac{l}{a_i} + \frac{\epsilon}{a_i} \cdot \mathbb{1}(i = j) \leq \frac{1}{4}$.

In the following part, we propose and prove the following lemma:

Lemma 17. *For any algorithm S, $\exists j \in \{1, 2, \ldots, d\}$, such that $Reg_{\mathbb{P}_j}(S) = \Omega(\sqrt{Td})$.*

Proof. Suppose f is a function: $\{0, 1\}^T \to [0, M]$. Denote $\mathbf{r} = [\mathbb{1}_1, \mathbb{1}_2, \ldots, \mathbb{1}_T]^\top$ as a vector containing the customer's decisions in sequence. Then for any $j = 1, 2, \ldots, d$ we have:

$$
\begin{aligned}
&\mathbb{E}_{\mathbb{P}_j}[f(\mathbf{r})] - \mathbb{E}_{\mathbb{P}_0}[f(\mathbf{r})] \\
&= \sum_{\mathbf{r}} f(\mathbf{r}) \cdot (\mathbb{P}_j[\mathbf{r}] - \mathbb{P}_0[\mathbf{r}]) \\
&\leq \sum_{\mathbf{r}:\mathbb{P}_j[\mathbf{r}] \geq \mathbb{P}_0[\mathbf{r}]} f(\mathbf{r})(\mathbb{P}_j[\mathbf{r}] - \mathbb{P}_0[\mathbf{r}]) \\
&\leq M \cdot \sum_{\mathbf{r}:\mathbb{P}_j[\mathbf{r}] \geq \mathbb{P}_0[\mathbf{r}]} f(\mathbf{r})(\mathbb{P}_j[\mathbf{r}] - \mathbb{P}_0[\mathbf{r}]) \\
&= \frac{M}{2} \|\mathbb{P}_j - \mathbb{P}_0\|_1.
\end{aligned}
\tag{43}
$$

Here we cite a lemma from *Cover and Thomas, Elements of Information theory*, Lemma 11.6.1.

Lemma 18.

$$
KL(\mathbb{P}_1 \| \mathbb{P}_2) \geq \frac{1}{2 \ln 2} \|\mathbb{P}_1 - \mathbb{P}_2\|_1^2.
$$

Since

$$
\begin{aligned}
&KL(\mathbb{P}_0(\mathbf{r}) \| \mathbb{P}_j(\mathbf{r})) \\
&= \sum_{t=1}^{T} KL(\mathbb{P}_0[r_t|\mathbf{r}_{t-1}] \| \mathbb{P}_j[r_t|\mathbf{r}_{t-1}]) \\
&= \sum_{t=1}^{t} \mathbb{P}_0(i_t \neq j) \cdot 0 + \mathbb{P}_0(i_t = j) \cdot KL(\frac{l}{a_j} \| \frac{l}{a_j} + \frac{\epsilon}{a_j}).
\end{aligned}
\tag{44}
$$

The first equality comes from the chain rule of decomposing a KL-divergence, and the second equality is because $\mathbb{1}_t$ satisfies a Bernoulli distribution $B(1, \frac{l}{a_{i_t}} + \frac{\epsilon}{a_{i_t}} \cdot \mathbb{1}(i_t = j))$. Now we propose another lemma:

Lemma 19. *If $\frac{1}{12} \leq p \leq \frac{1}{4}$, then we have: $KL(p, p + \epsilon) \leq 12\epsilon^2$ for sufficiently small ϵ.*

According to this Lemma 19, we have:

$$
\begin{aligned}
KL(\mathbb{P}_0(\mathbf{r}) \| \mathbb{P}_j(\mathbf{r})) &\leq \sum_{t=1}^{T} \mathbb{P}_0(i_t = j) \cdot 12\epsilon^2 \\
&\leq \sum_{t=1}^{T} \mathbb{P}_0(i_t = j) \cdot \frac{12}{16l^2} \epsilon^2.
\end{aligned}
\tag{45}
$$

Therefore, we have:

$$
\begin{aligned}
&\mathbb{E}_{\mathbb{P}_j}[f(\mathbf{r})] - \mathbb{E}_{\mathbb{P}_0}[f(\mathbf{r})] \\
&\leq \frac{M}{2} \frac{2 \ln 2}{\cdot} \sqrt{KL(\mathbb{P}_0(\mathbf{r}) \| \mathbb{P}_j(\mathbf{r}))} \\
&\leq \frac{\sqrt{6 \ln 2} M}{4} \cdot (\sqrt{\sum_{t=1}^{T} \mathbb{P}_0(i_t = j)}) \cdot \epsilon.
\end{aligned}
\tag{46}
$$

Denote $N_j := \sum_{t=1}^{T} \mathbb{1}(i_t = j)$, and hence:

$$\mathbb{E}_{\mathbb{P}_j}[f(\mathbf{r})] - \mathbb{E}_{\mathbb{P}_j}[f(\mathbf{r})] \leq \frac{\sqrt{6\ln 2}}{4} M(\sqrt{\mathbb{E}_{\mathbb{P}_0}[N_j]}) \cdot \epsilon. \tag{47}$$

Now let $f(\mathbf{r}) = N_j$, i.e., let function f simulate the algorithm which make choices of i_t's from historical results of $\{\mathbb{1}_1, \mathbb{1}_2, \ldots, \int_{t-1}\}$ (It is straightforward that $\{i_1, i_2, \ldots, i_{t-1}\}$ are also historical results crucial for deciding i_t. However, for a deterministic algorithm, it can generate $i_1, i_2, \ldots, i_{t-1}$ directly from $\emptyset, \{\mathbb{1}_1\}, \{\mathbb{1}_1, \mathbb{1}_2\}, \ldots, \{\mathbb{1}_1, \mathbb{1}_2, \ldots, \mathbb{1}_{t-2}\}$.) Now, $0 \leq f(\mathbf{r}) \leq T$, which indicates that $M = T$.. Then it turns out that

$$
\begin{aligned}
\mathbb{E}_{\mathbb{P}_j}[N_j] - \mathbb{E}_{\mathbb{P}_0}[N_j] &\leq \frac{\sqrt{6\ln 2}}{4} \cdot T \cdot \epsilon \cdot \sqrt{\mathbb{E}_{\mathbb{P}_0}[N_j]} \\
\Rightarrow \frac{1}{d}\sum_{j=1}^{d} \mathbb{E}_{\mathbb{P}_j}[N_j] &\leq \frac{1}{d}\sum_{j=1}^{d}(\mathbb{E}_{\mathbb{P}_0}[N_j] + \frac{\sqrt{6\ln 2}}{4} \cdot T \cdot \epsilon \cdot \sqrt{\mathbb{E}_{\mathbb{P}_0}[N_j]}) \\
&= \frac{T}{d} + \frac{\sqrt{6\ln 2}}{4} \cdot \frac{T}{d} \cdot \epsilon \cdot \sum_{j=1}^{d} \sqrt{\mathbb{E}_{\mathbb{P}_\rangle}[N_j]} \\
&\leq \frac{T}{d} + \frac{\sqrt{6\ln 2}}{4} \cdot \epsilon \cdot \frac{T}{d} \cdot \sqrt{Td} \\
&= \frac{T}{d} + \frac{\sqrt{6\ln 2}}{4} \cdot \sqrt{\frac{d}{T}} \cdot \frac{T}{d} \cdot \sqrt{Td} \\
&\leq \frac{T}{3} + 0.525 \cdot T \\
&\leq 0.9T
\end{aligned}
\tag{48}
$$

Here the second line is an average over all $j = 1, 2, \ldots, d$ of the first line, the third line uses the fact that $\sum_{j=1}^{d} \mathbb{E}[N_j] = T$, the fourth line applies a Cauchy-Schwarz Inequality that $Td = (\sum_{j=1}^{d} \mathbb{E}_{\mathbb{P}_0}[N_j])(d \cdot 1) \geq (\sum_{j=1}^{d} \sqrt{\mathbb{E}_{\mathbb{P}_0}[N_j]})^2$, the fifth line plugs in the values that $\epsilon = \sqrt{\frac{d}{T}}$, the sixth line uses the fact that $\ln 2 < 0.7$, and the last line holds for sufficient large T.

From Equation 48, we know that $\exists j \in \{1, 2, \ldots, d\}$ such that $\mathbb{E}_{\mathbb{P}_j}[N_j] \leq 0.9T$. As a result, we have:

$$
\begin{aligned}
Reg_{\mathbb{P}_j}(S) &\geq (1 - 0.9)T(\frac{l + \epsilon}{a_j} \cdot a_j - \frac{l}{a_{i_t}} \cdot a_{i_t}), \forall i_t \neq j \\
&= 0.1T(l + \epsilon - l) \\
&= 0.1T\epsilon \\
&= 0.1\sqrt{Td}.
\end{aligned}
\tag{49}
$$

Therefore, the $\Omega(\sqrt{Td})$ regret bound holds. ∎

∎

B.3 Proof of Theorem 9

Proof. Prior to our technical analysis, we briefly introduce the roadmap of proving the unfairness lower bound.

  (i) We construct two different but very similar problem settings: one is exactly Example 1, the other is identical to Example 1 except all probabilities of 0.5 are now changed into $(0.5 - \zeta)$, where $\zeta = C \cdot T^{-\frac{1}{2}+\eta}$ for some super small constant $C \geq 0$ and some small $\eta \geq 0$. In the following, we may call them the "Problem 0" (or P_0) and the "Problem ζ" (or P_ζ) sequentially.

 (ii) We derive the close-form solutions to both Problem 0 and Problem ζ, where we also parameterize the reward function with the expected proposed price V_r and the proposed accepted price V_s. Of course Problem ζ is more general and we may get the solutions to Problem 0 by simply let $\zeta = 0$.

(iii) We show that there does not exist any policy π that satisfies both of the following conditions simultaneously:

- π is within $C_0 \cdot T^{-\frac{1}{2}+\eta}$-suboptimal (w.r.t. regret) and within $C_0 \cdot T^{-\frac{1}{2}+\eta}$-unfair (w.r.t. fairness) in $P0$.

- π is within $C_0 \cdot T^{-\frac{1}{2}+\eta}$-suboptimal (w.r.t. regret) and within $C_0 \cdot T^{-\frac{1}{2}+\eta}$-unfair (w.r.t. fairness) in $P\zeta$.

(iv) We show that any algorithm have to distinguish P_0

According to the roadmap above, we firstly construct the following example as the problem setting for lower bound proof.

Example 20. *Customers form 2 disjoint groups: Group 1 takes 30% proportion of customers, and Group 2 takes the rest 70%. In specific,*

- *In Group 1, 40% customers valuate the item as \$0, $(10\% + \zeta)$ valuate customers it as \$0.625, and $(50\% - \zeta)$ customers valuate it as \$1.*

- *In Group 2, 20% customers valuate it as \$0, $(30\% + \zeta)$ customers valuate it as \$0.7, and $(50\% - \zeta)$ customers valuate it as \$1.*

Here $\zeta = C \cdot T^{-\frac{1}{2}+\eta}$ is a small amount, where $0 \le C \le 10^{-10}$. In other words, we have $\mathbf{v}^\top = [\frac{5}{8}, \frac{7}{10}, 1], F_1 = \mathrm{diag}\{0.6, 0.5 - \zeta, 0.5 - \zeta\}, F_2 = \mathrm{diag}\{0.8, 0.8, 0.5 - \zeta\}$ and our policy $\pi = (\pi^1, \pi^2)$ where $\pi^1, \pi^2 \in \Delta^3$. Our goal is to approach the following optimal policy

$$
\begin{aligned}
\pi_{\zeta,*} = &\underset{\pi=(\pi^1,\pi^2)\in\Pi}{\mathrm{argmax}} \quad R(\pi; F_1, F_2) \\
&s.t. \quad U(\pi) = 0 \\
&\qquad S(\pi; F_1, F_2) = 0.
\end{aligned}
\tag{50}
$$

For any policy π feasible to the constraints in (51), denote its expected accepted price as V_s (identical in both groups) and its expected proposed price as V_r (identical in both groups as well). Notice that $V_r \ge V_s \ge \frac{5}{8}$, we define $\alpha = V_r - V_s$ as their difference, and therefore we know that $\alpha \ge 0$. Again, we denote $R(\pi, F_1, F_2)$ as $R(\pi)$ without causing misunderstandings. Here we propose the following lemma regarding Example 20.

Lemma 21 (Close-form solution to Example 20). *For the problem setting defined in Example 20, we have:*

$$
\begin{aligned}
\pi^1_{\zeta,*} =&[\frac{20 - 40\zeta}{29 - 10\zeta}, 0, \frac{9 + 30\zeta}{29 - 10\zeta}]^\top, \\
\pi^2_{\zeta,*} =&[0, \frac{25 - 50\zeta}{29 - 10\zeta}, \frac{4 + 40\zeta}{29 - 10\zeta}]^\top, \forall \zeta \in [0, 10^{-10}].
\end{aligned}
\tag{51}
$$

Besides, for any feasible policy π and its corresponding V_s and α, we have:

$$
R(\pi) = \frac{71 - 30\zeta}{100} \cdot V_s + \frac{(100 - 60\zeta) - (142 - 60\zeta)V_s}{(8V_s - 5)(1 - V_s)25} \cdot V_s \cdot \alpha.
\tag{52}
$$

Proof of Lemma 21. For any feasible policy $\pi = (\pi^1, \pi^2)$, it has to satisfy the following equations for $e = 1, 2$:

$$
\begin{cases}
\mathbb{1}^\top \pi^e = 1 \\
\mathbf{v}^\top \pi^e = V_s + \alpha \\
\dfrac{\mathbf{v}^\top F_e \pi^e}{\mathbb{1}^\top F_e \pi^e} = V_s.
\end{cases}
$$

This is equivalent to the following linear equations system

$$
\begin{cases}
\mathbb{1}^\top \pi^e = 1 \\
\mathbf{v}^\top \pi^e = V_s + \alpha \\
(\mathbf{v} - V_s \cdot \mathbb{1})^\top F_e \pi^e = 0.
\end{cases}
\tag{53}
$$

This is further equivalent to $A_1 \pi^1 = [1, V_s + \alpha, 0]^\top$ and $A_2 \pi^2 = [1, V_s + \alpha, 0]^\top$ where

$$A_1(V_s, \zeta) = \begin{bmatrix} 1 & 1 & 1 \\ \frac{5}{8} & \frac{7}{10} & 1 \\ (\frac{5}{8} - V_s) \cdot \frac{3}{5} & (\frac{7}{10} - V_s) \cdot (\frac{1}{2} - \zeta) & (1 - V_s) \cdot (\frac{1}{2} - \zeta) \end{bmatrix}. \tag{54}$$

and

$$A_2(V_s, \zeta) = \begin{bmatrix} 1 & 1 & 1 \\ \frac{5}{8} & \frac{7}{10} & 1 \\ (\frac{5}{8} - V_s) \cdot \frac{4}{5} & (\frac{7}{10} - V_s) \cdot \frac{4}{5} & (1 - V_s) \cdot (\frac{1}{2} - \zeta) \end{bmatrix}. \tag{55}$$

Here we may omit the parameters (V_s, ζ) without misunderstanding. For $V_s = \frac{5}{8}$, the only possible policy is to propose the lowest price $\frac{5}{8}$ for both groups, and the expected reward is $0.3 \times \frac{5}{8} \times \frac{3}{5} + 0.7 \times \frac{5}{8} \times \frac{4}{5} = 0.4625 < 0.5 - \zeta$. Therefore, it is suboptimal as its expected reward is less than that of a deterministic policy keep proposing 1 as a price (whose reward is $0.5 - \zeta$). In the following, we only consider the case when $V_s > \frac{5}{8}$. Solve these linear equation systems and get

$$\begin{aligned} \pi^1 &= A_1^{-1} [1, V_s + \alpha, 0]^\top \\ &= \frac{1}{3(8V_s - 5)(1 + 10\zeta)} \\ &\begin{bmatrix} 120\alpha(1 - 2\zeta) \\ -((1 + 10\zeta)8V_s + 10(1 - 8\zeta)) \cdot \alpha - (8(1 + 10\zeta)V_s^2 - 13(1 + 10\zeta)V_s + (1 + 10\zeta)5) \\ 10(8(1 + 10\zeta)V_s - 2(1 + 28\zeta)) \cdot \alpha + (1 + 10\zeta)(8V_s - 5)(10V_s - 7) \end{bmatrix} \end{aligned} \tag{56}$$

and

$$\begin{aligned} \pi^2 &= A_2^{-1} [1, V_s + \alpha, 0]^\top \\ &= \frac{1}{3(1 - V_s)(3 + 10\zeta)} \\ &\begin{bmatrix} 4(((3 + 10\zeta)10V_s - (6 + 100\zeta))\alpha - (3 + 10\zeta)(10V_s - 7)(V_s - 1)) \\ (-5) \cdot (((3 + 10\zeta)8V_s - 80\zeta)\alpha + (3 + 10\zeta)(8V_s - 5)(V_s - 1)) \\ 24\alpha \end{bmatrix}. \end{aligned} \tag{57}$$

On the one hand, we can get the explicit form of $R(\pi)$ w.r.t. V_s and α:

$$\begin{aligned} R(\pi) &= q \cdot \mathbf{v}^\top F_1 \pi^1 + (1 - q) \cdot \mathbf{v}^\top F_2 \pi^2 \\ &= \frac{71 - 30\zeta}{100} V_s + \frac{(100 - 60\zeta) - (142 - 60\zeta)V_s}{(8V_s - 5)(1 - V_s)25} \cdot V_s \alpha. \end{aligned} \tag{58}$$

On the other hand, we have a few constraints to be applied. Since π is a probabilistic distribution, we have $\pi^e(i) \geq 0, e = 1, 2; i = 1, 2, 3$, which lead to

$$\begin{cases} 120\alpha(1 - 2\zeta) \geq 0 \\ -((1 + 10\zeta)8V_s + 10(1 - 8\zeta)) \cdot \alpha - (8(1 + 10\zeta)V_s^2 - 13(1 + 10\zeta)V_s + (1 + 10\zeta)5) \geq 0 \\ 10(8(1 + 10\zeta)V_s - 2(1 + 28\zeta)) \cdot \alpha + (1 + 10\zeta)(8V_s - 5)(10V_s - 7) \geq 0 \\ 4(((3 + 10\zeta)10V_s - (6 + 100\zeta))\alpha - (3 + 10\zeta)(10V_s - 7)(V_s - 1)) \geq 0 \\ (-5) \cdot (((3 + 10\zeta)8V_s - 80\zeta)\alpha + (3 + 10\zeta)(8V_s - 5)(V_s - 1)) \geq 0 \\ 24\alpha \geq 0. \end{cases} \tag{59}$$

From (59), we may derive the following upper and lower bounds for α.

(a) The first line and the last line of (59) is naturally satisfied.

(b) From the second line, we have

$$\alpha \leq \frac{(1 + 10\zeta)(8V_s - 5)(1 - V_s)}{(1 + 10\zeta)8V_s + 10(1 - 8\zeta)} := B_1. \tag{60}$$

(c) From the third line, we have

$$\alpha \geq \frac{(1 + 10\zeta)(8 \cdot V_s - 5)(7 - 10V_s)}{(1 + 10\zeta)8 \cdot V_s - 2(1 + 28\zeta)} \cdot \frac{1}{10} := B_2. \tag{61}$$

(d) From the fourth line, we have

$$\alpha \geq \frac{(3+10\zeta)(10V_s-7)(1-V_s)}{(3+10\zeta)10V_s-(6+100\zeta)} := B_3. \tag{62}$$

(e) From the fifth line, we have

$$\alpha \leq \frac{(3+10\zeta)(8V_s-5)(1-V_s)}{(3+10\zeta)8\cdot V_s-80\zeta} := B_4. \tag{63}$$

We get four constraints on α as above, where (60) and (63) are upper bounds, and (61) and (62) are lower bounds. Compare B_1 with B_4, we notice that

$$\frac{B_1}{B_4} = \frac{\frac{80\zeta}{3+10\zeta}}{8V_s+10\cdot\frac{1-8\zeta}{1+10\zeta}} \leq 1. \tag{64}$$

Therefore, (60) is tighter than (63). For the comparison between B_2 and B_3, we notice that $B_2 < 0 < B_3$ when $V_s > \frac{7}{10}$ and $B_2 \geq 0 \geq B_3$ when $V_s \leq \frac{7}{10}$.

In the following part, we derive the optimal policy by cases.

(a) When $\frac{5}{8} < V_s \leq \frac{50-30\zeta}{71-30\zeta}$, we have

$$
\begin{aligned}
R(\pi) =& \frac{71-30\zeta}{100}V_s + \frac{(100-60\zeta)-(142-60\zeta)V_s}{(8V_s-5)(1-V_s)25}\cdot V_s\alpha \\
\leq& \frac{71-30\zeta}{100}V_s + \frac{(100-60\zeta)-(142-60\zeta)V_s}{(8V_s-5)(1-V_s)25}\cdot V_s\cdot B_1 \\
=& \frac{71-30\zeta}{100}V_s + \frac{(100-60\zeta)-(142-60\zeta)V_s}{(8V_s-5)(1-V_s)25}\cdot V_s\cdot\frac{(1+10\zeta)(8V_s-5)(1-V_s)}{(1+10\zeta)8V_s+10(1-8\zeta)} \\
=& \frac{71-30\zeta}{100}V_S + \frac{100-142V_s-60\zeta(1-V_s)}{25(8V_s+10\cdot\frac{1-8\zeta}{1+10\zeta})}\cdot V_s \\
=& \frac{71-30\zeta}{100}V_S + \frac{100-142V_s-60\zeta(1-V_s)}{25(8V_s+10)-\frac{450\zeta}{1+10\zeta}}\cdot V_s \\
<& \frac{71-30\zeta}{100}V_S + \frac{100-142V_s-60\zeta(1-V_s)+\frac{450\zeta}{1+10\zeta}}{25(8V_s+10)}\cdot V_s \\
<& \frac{71}{100}V_s + \frac{100.1-142V_s}{25(8V_s+10)}\cdot V_s \\
=& \frac{71(8\cdot V_s+10)+(100.1-142\cdot V_s)\cdot 4}{100(8\cdot V_s+10)}\cdot V_s \\
=& \frac{11104\cdot V_s}{1000(8\cdot V_s+10)} \\
\leq& \frac{11104}{8000} - \frac{\frac{5}{4}\times 11104}{1000(8\times\frac{50}{71}+10)} \\
\leq& 0.50019
\end{aligned}
\tag{65}
$$

Here the first inequality (line 2) is by $(100-60\zeta)-(142-60\zeta)V_s \geq 0$ as $V_s \leq \frac{50-30\zeta}{71-30\zeta}$ and by $\alpha \leq B_1$, the second inequality (line 6) is by the fact that V_s's coefficient is within $(0,1)$, the third inequality (line 7) is by $\zeta \leq 10^{-10} < \frac{1}{4500}$, the fourth inequality (line 10) is by $V_s \leq \frac{50}{71}$ and the last inequality (line 11) is by numerical computations. We will later show that 0.50019 is not optimal.

(b) When $V_s > \frac{50-30\zeta}{71-30\zeta}$, we know that $V_s > \frac{7}{10}$ as $\zeta < \frac{1}{30}$. Therefore, we know $B_2 < 0 < B_3$ and we have

$$\alpha \geq B_3 = \frac{(3+10\zeta)(10V_s-7)(1-V_s)}{(3+10\zeta)10V_s-(6+100\zeta)}.$$

As a result, we have

$$
\begin{aligned}
R(\pi) &= \frac{71 - 30\zeta}{100} V_s - \frac{(142 - 60\zeta)V_s - (100 - 60\zeta)}{(8V_s - 5)(1 - V_s)25} \cdot V_s \alpha \\
&\leq \frac{71 - 30\zeta}{100} V_s - \frac{(142 - 60\zeta)V_s - (100 - 60\zeta)}{(8V_s - 5)(1 - V_s)25} \cdot V_s \cdot B_3 \\
&\leq \frac{71 - 30\zeta}{100} V_s - \frac{(142 - 60\zeta)V_s - (100 - 60\zeta)}{(8V_s - 5)(1 - V_s)25} \cdot V_s \cdot \frac{(3 + 10\zeta)(10V_s - 7)(1 - V_s)}{(3 + 10\zeta)10V_s - (6 + 100\zeta)}
\end{aligned}
\tag{66}
$$

Also, we can derive an upper bound for V_s as $B_3 \leq \alpha \leq B_1$.

$$
\begin{aligned}
\frac{(3 + 10\zeta)(10V_s - 7)(1 - V_s)}{(3 + 10\zeta)10V_s - (6 + 100)} &\leq \frac{(1 + 10\zeta)(8 \cdot V_s - 5)(1 - V_s)}{(1 + 10\zeta)8 \cdot V_s + 10(1 - 8\zeta)} \\
\Leftrightarrow \qquad V_s &\leq \frac{8 + 10\zeta}{11 + 10\zeta}.
\end{aligned}
\tag{67}
$$

Therefore, we may solve the maximal of $R(\pi)$ on $\frac{50 - 30\zeta}{71 - 30\zeta} \leq V_s \leq \frac{8 + 10\zeta}{11 + 10\zeta}$ by combining (66).

$$
\begin{aligned}
\frac{\partial R(\pi)}{\partial V_s} &= \frac{3(-3135 + 9870V_s - 7491V_s^2) + 3\zeta(-46430 + 145660V_s - 114638V_s^2)}{20(-5 + 8V_s)^2(-3 - 50\zeta + 15V_s + 50\zeta V_s)^2} \\
&\quad + \frac{+3\zeta^2(97900 - 315800V_s + 251740V_s^2) + 3\zeta^3(15000 - 30000V_s + 15000V_s^2)}{20(-5 + 8V_s)^2(-3 - 50\zeta + 15V_s + 50\zeta V_s)^2} \\
&= \frac{-22473(V_s - \frac{1645 - 6\sqrt{2685}}{2497})(V_s - \frac{1645 + 6\sqrt{2685}}{2497}) + 3(-46430 + 145660V_s - 114638V_s^2)\zeta + o(\zeta^2)}{20(-5 + 8V_s)^2(-3 - 50\zeta + 15V_s + 50\zeta V_s)^2} \\
&\gtrsim \frac{-22473(V_s - 0.5343)(v_s - 0.7833)}{20(-5 + 8V_s)^2(-3 - 50\zeta + 15V_s + 50\zeta V_s)^2} \\
&> 0.
\end{aligned}
\tag{68}
$$

Here the "$\gtrsim$" inequality is because the coefficient of ζ in any monomial above is within $\pm 10^6$, which indicates that any monomial containing ζ is within ± 0.0001. The last line is because $\frac{7}{10} \leq \frac{50 - 30\zeta}{71 - 30\zeta} \leq V_s \leq \frac{8 + 10\zeta}{11 + 10\zeta} \leq \frac{3}{4}$ and therefore $(V_s - 0.5343)(v_s - 0.7833) < 0$. As a result, we know that $R(\pi)$ is monotonically increasing as V_s increases within the range above. Therefore, we have:

$$
R(\pi) \leq R(\pi)|_{\alpha = B_3} \leq R(\pi)|_{\alpha = B_3 \text{ and } V_s = \frac{8 + 10\zeta}{11 + 10\zeta}} = \frac{37(1 - 2\zeta)(4 + 5\zeta)}{10(29 - 10\zeta)}
\tag{69}
$$

By plugging in $V_s = \frac{8 + 10\zeta}{11 + 10\zeta}$ into $\alpha = B_3$ and the close-form feasible solutions of π^1 and π^2 (i.e., (56) and (57)), we may get:

$$
\begin{aligned}
\alpha = B_3 &= \frac{3(1 + 10\zeta)(3 + 10\zeta)}{2(29 - 10\zeta)(11 + 10\zeta)} \\
\pi^1 &= [\frac{20 - 40\zeta}{29 - 10\zeta}, 0, \frac{9 + 30\zeta}{29 - 10\zeta}]^\top \\
\pi^2 &= [0, \frac{25 - 50\zeta}{29 - 10\zeta}, \frac{4 + 40\zeta}{29 - 10\zeta}]^\top.
\end{aligned}
\tag{70}
$$

Pushing back (70) to (50), we verify that $R(\pi)_{\max} = \frac{37(1 - 2\zeta)(4 + 5\zeta)}{10(29 - 10\zeta)}$ and therefore all inequalities in (69) hold as equalities.

Notice that $\frac{37(1 - 2\zeta)(4 + 5\zeta)}{10(29 - 10\zeta)} > 0.50019$, and therefore the optimal policy $\pi_{\zeta,*}$ is what we derive in (70). This holds the lemma. ∎

With Lemma 21, we know that $\pi_{\zeta,*}^1 = [\frac{20 - 40\zeta}{29 - 10\zeta}, 0, \frac{9 + 30\zeta}{29 - 10\zeta}]^\top$ and $\pi_{\zeta,*}^2 = [0, \frac{25 - 50\zeta}{29 - 10\zeta}, \frac{4 + 40\zeta}{29 - 10\zeta}]^\top$. We denote $V_{s,\zeta}^* := \frac{8 + 10\zeta}{3 + 10\zeta}$ and $\alpha_\zeta^* = \frac{3(1 + 10\zeta)(3 + 10\zeta)}{2(29 - 10\zeta)(11 + 10\zeta)}$ for future use. We also know that the optimal policy for Example 1 (i.e., $\zeta = 0$) is exactly what we proposed, i.e., $\pi_*^1 = [\frac{20}{29}, 0, \frac{9}{29}]^\top$ and $\pi_*^2 = [0, \frac{25}{29}, \frac{4}{29}]^\top$.

Let us go back to the two problems: P_0 defined in Example 1 and P_ζ defined in Example 20, where we consider the following four conditions:

- π is within $C_0 \cdot T^{-\frac{1}{2}+\eta}$-suboptimal (w.r.t. regret) in P_0 (denoted as Condition A).

- π is within $C_0 \cdot T^{-\frac{1}{2}+\eta}$-suboptimal (w.r.t. regret) in P_ζ (denoted as Condition B).

- π is within $C_0 \cdot T^{-\frac{1}{2}+\eta}$-unfair (w.r.t. fairness) in P_0 (denoted as Condition C).

- π is within $C_0 \cdot T^{-\frac{1}{2}+\eta}$-unfair (w.r.t. fairness) in P_ζ (denoted as Condition D).

According to our proof roadmap, we then prove the following lemma:

Lemma 22 (No policy fitting in P_0 and P_ζ). *There exist constants $C_0 > 0$ such that there does not exist any policy $\pi \in \Pi$ that satisfies all of Condition $ABCD$ (denoting $A \wedge B \wedge C \wedge D$) simultaneously.*

Corollary 23. *The space of Π can be divided as the following 3 subspaces:*

1. *Policies satisfying Condition AC (denoted as Space AC).*

2. *Policies satisfying Condition BD (denoted as Space BD).*

3. *Policies satisfying Condition (denoted as Outer Spaces) $\bar{A}\bar{B} \vee \bar{C}\bar{D} \vee \bar{A}\bar{D} \vee \bar{B}\bar{C}$. and these three subspaces are pairwise disjoint.*

Proof of Lemma 22. Let $C_1 = \frac{C}{W}$ and $C_2 = \frac{C}{W \cdot L}$ where $L > 0$ is a constant from Lemma 14 and $W \geq 10$ to be specified later. Let $C_0 = \min\{C_1, C_2\}$, and we prove the lemma by contradiction. Suppose there exists a policy π satisfies the four conditions above, and then we denote the expected accepted prices in G_1 and G_2 in Problem ζ are $V_{s,\zeta}$ and $V_{s,\zeta} + \beta_\zeta$ sequentially, where $\beta_\zeta \in [0, C_2 T^{-\frac{1}{2}+\eta}]$. Here we assume $\beta \geq 0$ without losing generality as we will not use the specific property of G_1 versus G_2. Also, we denote α_ζ as the difference between the expected proposed price in both groups (denoted as $V_{r,\zeta}$) and $V_{s,\zeta}$.

Now, consider a corresponding policy:

$$\check{\pi} := \begin{cases} G_1 : \mathbb{E}[\text{accepted price}] = V_{s,\zeta}, \mathbb{E}[\text{proposed price}] = V_{s,\zeta} + \alpha_\zeta \\ G_2 : \mathbb{E}[\text{accepted price}] = V_{s,\zeta}, \mathbb{E}[\text{proposed price}] = V_{s,\zeta} + \alpha_\zeta \end{cases} \tag{71}$$

According to (67), we know that $\frac{5}{8} \leq V_{s,\zeta} \leq V_{s,\zeta}^* = \frac{8+10\zeta}{11+10\zeta}$ and $R(\check{\pi}) \leq R(\pi_{\zeta,*})$. Therefore, we have:

$$\begin{aligned} R(\pi) &\leq R(\check{\pi}) + L \cdot \beta_\zeta \\ &\leq R(\pi_{\zeta,*}) - \min_{V_s \in [\frac{5}{8}, \frac{8+10\zeta}{11+10\zeta}]} \frac{\partial R(\pi)}{\partial V_s} \cdot (V_{s,\zeta}^* - V_{s,\zeta}) + L \cdot \beta_\zeta \\ &\leq R(\pi_{\zeta,*}) - \frac{1}{4} \cdot (V_{s,\zeta}^* - V_{s,\zeta}) + L \cdot \beta_\zeta. \end{aligned} \tag{72}$$

Here the first line comes from Lemma 14, the second line comes from the fact that $f(x_1) - f(x_2) \geq \min_x(f'(x))(x_1 - x_2)$ for $x_1 \geq x_2$ and $f'(x) > 0$, and the third line comes from the fact that $\frac{\partial R(\pi)}{\partial V_s} \geq \frac{1}{4}$ for $V_s \in [0.625, 0.728]$ as $0.728 > \frac{8+10\zeta}{11+10\zeta}$ for $\zeta \leq 10^{-10}$. Also, since π satisfies a low-regret condition, we have

$$R(\pi_{\zeta,*}) - R(\pi) \leq C_1 \cdot T^{-\frac{1}{2}+\eta}.$$

Combining with (72), we have

$$\begin{aligned} \frac{1}{4}(V_{s,\zeta}^* - V_{s,\zeta}) - L \cdot \beta_\zeta &\leq C_1 T^{-\frac{1}{2}+\eta} \\ \Rightarrow \quad (V_{s,\zeta}^* - V_{s,\zeta}) &\leq C_1 T^{-\frac{1}{2}+\eta} + L \cdot \beta_\zeta \\ &\leq (C_1 + LC_2) T^{-\frac{1}{2}+\eta}. \end{aligned} \tag{73}$$

Notice that this is suitable for any $\zeta \in [0, 10^{-10}]$, we may have the same result for both $\zeta = CT^{-\frac{1}{2}+\eta}$ and for $\zeta = 0$. We denote $\zeta_0 = 0$ and $\zeta_1 = CT^{-\frac{1}{2}+\eta}$ where $C = 10^{-10}$. Therefore, we have:

$$(V_{s,\zeta_0}^* - V_{s,\zeta_0}) \leq (C_1 + LC_2)T^{-\frac{1}{2}+\eta},$$
$$(V_{s,\zeta_1}^* - V_{s,\zeta_1}) \leq (C_1 + LC_2)T^{-\frac{1}{2}+\eta}. \tag{74}$$

Now let us bound $(\alpha_\zeta^* - \alpha_\zeta)$ for both ζ_0 and ζ_1. From (60) and (62), we have

$$B_3|_{V_s=V_{s,\zeta}} \leq \alpha_\zeta \leq B_1|_{V_s=V_{s,\zeta}}$$
$$B_3|_{V_s=V_{s,\zeta}^*} = \alpha_\zeta^* = B_1|_{V_s=V_{s,\zeta}^*}$$
$$\Rightarrow \quad \min_{V_s \in [0.7, 0.75]} \{\frac{\partial B_3}{\partial V_s}, \frac{\partial B_1}{\partial V_s}\} \cdot (V_{s,\zeta}^* - V_{s,\zeta}) \leq (\alpha_\zeta^* - \alpha_\zeta) \leq \max_{V_S \in [0.7, 0.75]} \{\frac{\partial B_3}{\partial V_s}, \frac{\partial B_1}{\partial V_s}\} (V_{s,\zeta}^* - V_{s,\zeta}) \tag{75}$$
$$\Rightarrow \quad 0 \leq 0.05(V_{s,\zeta}^* - V_{s,\zeta}) \leq (\alpha_\zeta^* - \alpha_\zeta) \leq 0.6(V_{s,\zeta}^* - V_{s,\zeta}).$$

Therefore, we have:

$$0 \leq V_{r,\zeta}^* - V_{r,\zeta} = V_{s,\zeta}^* + \alpha_\zeta^* - (V_{s,\zeta} + \alpha_\zeta) \leq (1 + 0.6)(V_{s,\zeta}^* - V_{s,\zeta}) \leq \frac{8}{5} \cdot (C_1 + LC_2)T^{-\frac{1}{2}+\eta}.$$

Therefore, we know that

$$V_{r,\zeta_0}^* \geq V_{r,\zeta_0} \geq V_{r,\zeta_0}^* - \frac{8}{5} \cdot (C_1 + LC_2)T^{-\frac{1}{2}+\eta},$$
$$V_{r,\zeta_1}^* \geq V_{r,\zeta_1} \geq V_{r,\zeta_1}^* - \frac{8}{5} \cdot (C_1 + LC_2)T^{-\frac{1}{2}+\eta}. \tag{76}$$
$$\Rightarrow \quad |V_{r,\zeta_1} - V_{r,\zeta_0}| \geq |V_{r,\zeta_0}^* - V_{r,\zeta_1}^*| - (C_1 + LC_2)T^{-\frac{1}{2}+\eta}.$$

HOWEVER, we have $V_{r,\zeta_1} = V_{r,\zeta_0}$ since they are the expected proposed price of the same pricing policy π in P_0 and P_ζ where the prices sets are all the same! Therefore, we have $|V_{r,\zeta_0}^* - V_{r,\zeta_1}^*| - (C_1 + LC_2)T^{-\frac{1}{2}+\eta} \leq 0$. Since $V_{r,\zeta_0}^* = \frac{43}{58}$ and $V_{r,\zeta_1} = \frac{43+10\zeta_1}{58-20\zeta_1}$, we have $|V_{r,\zeta_0}^* - V_{r,\zeta_1}^*| = \frac{360\zeta}{29(29-10\zeta)} \geq \frac{C}{3} \cdot T^{-\frac{1}{2}+\eta}$. Since $C_1 = \frac{C}{W} \leq 10^{-11}$ and $C_2 = \frac{C}{W \cdot L} \leq \frac{1}{L} \times 10^{-11}$, we know that $|V_{r,\zeta_0}^* - V_{r,\zeta_1}^*| > (C_1 + LC_2)T^{-\frac{1}{2}+\eta}$, which contradicts to the inequality we derived. Therefore, the lemma is proved by contradiction. ∎

In the following, we set $\zeta = \zeta_1 = CT^{-\frac{1}{2}+\eta}$ where $C = 10^{-10}$ as is defined in the proof of Lemma 22. Now let us go back to the main stream of proving Theorem 9. We also make it by contradiction. For any given C_x, without loss of generality, we may assume that $C_u \leq C_x$ to be specified later. Define $x = \frac{C_x}{\log T}$, and therefore $C_x T^{\frac{1}{2}} = T^{\frac{1}{2}-x}$. We will make use of Example 1 and Example 20, and let $\eta > x$ to be specified later. Therefore, we have $C_u \cdot T^{\frac{1}{2}} \leq C_x \cdot T^{\frac{1}{2}} = T^{\frac{1}{2}-x}$, which means that the contradiction is a **sufficient condition** to the following result: Suppose there exists an $x > 0$ and an algorithm such that it can always achieve $O(T^{\frac{1}{2}+x})$ regret with zero procedural unfairness and $O(T^{\frac{1}{2}-x})$ substantive unfairness. According to Corollary 23, we know that any policy $\pi \in \Pi$ are in exact one of those three spaces. In our problem setting, denote the policy we take at time $t = 1, 2, \ldots, T$ as π_t. Now we show that: among all policies $\{\pi_t\}_{t=1}^T$ we have taken, there are at most $O(T^{1-\eta+x})$ policies in all T policies having been played belonging to the Outer Space defined in Corollary 23. In fact, for any policy π in the Outer Space, we have:

(i) When $\pi \in \bar{A}\bar{B}$, the policy π will definitely suffer a regret $C_0 \cdot T^{-\frac{1}{2}+\eta}$, no matter which the problem setting is (i.e., P_0 or P_ζ). In order to guarantee $O(T^{\frac{1}{2}+x})$ regret, there are at most $N_1 = O(T^{1-\eta+x}) = o(T)$ rounds to play a policy in $\bar{A}\bar{B}$.

(ii) When $\pi \in \bar{C}\bar{D}$, the policy π will definitely suffer a substantive unfairness $C_0 \cdot T^{-\frac{1}{2}+\eta}$, no matter which the problem setting is (i.e., P_0 or P_ζ). In order to guarantee $O(T^{\frac{1}{2}-x})$ regret, there are at most $N_2 = O(T^{1-\eta-x}) = o(T)$ rounds to play a policy in $\bar{C}\bar{D}$.

(iii) When $\pi \in \bar{A}\bar{D} \vee \bar{B}\bar{C}$, in either P_0 or P_ζ it suffers something (that could be either $C_0 \cdot T^{-\frac{1}{2}+\eta}$ regret or $C_0 \cdot T^{-\frac{1}{2}+\eta}$ unfairness). As we have to guarantee $O(T^{\frac{1}{2}+x})$ regret and $O(T^{\frac{1}{2}-x})$ substantive unfairness, there are still at most $N_3 = O(\max\{T^{1-\eta+x}, T^{1-\eta-x}\}) = O(T^{1-\eta+x}) = o(T)$.

Therefore, the number of rounds when we select and play a policy from Space AC or Space BD is at least $T - o(T) \geq \frac{T}{2}$. Notice that if a policy in AC, then it performs well in P_0 but not necessarily in P_ζ. Similarly, if a policy in BD, then it performs well in P_ζ but not necessarily in P_0. Therefore, two questions emerges:

- How do policies in AC perform in P_ζ? and How do policies in BD perform in P_0? Specifically, we only care about the substantive fairness.

- How can we distinguish between P_ζ and P_0?

For distinguishablity, denote $F_1(\zeta) = \text{diag}\{0.6, 0.5 - \zeta, 0.5 - \zeta\}$ and $F_2(\zeta) = \text{diag}\{0.8, 0.8, 0.5 - \zeta\}$. Denote $S_0(\pi) := S(\pi, F_1(0), F_2(0))|_{\zeta=0}$ and $S_\zeta(\pi) := S(\pi, F_1(\zeta), F_2(\zeta))$. In the following, we propose two lemmas that help us prove. The first lemma, Lemma 24, shows that failing to distinguish would lead to large substantive unfairness, which answers the first question above.

Lemma 24. *There exists a constant C_{ac} such that: for any policy $\pi \in AC$, we have $S_\zeta(\pi) > C_{ac} \cdot T^{-\frac{1}{2}+\eta}$. There also exists a constant C_{bd} such that: for any policy $\pi \in BD$, we have $S_0(\pi) > C_{bd} \cdot T^{-\frac{1}{2}+\eta}$.*

Proof of Lemma 24 . We firstly prove the first half of this lemma, and then demonstrate the second half (which can be proved in exact the same way.)

First of all, we have the close-form solution to both P_0 and P_ζ in (51). Therefore, we have

$$S_\zeta(\pi_{0,*}) = \frac{12\zeta(1 - 2\zeta)}{(11 - 6\zeta)(11 - 10\zeta)}. \tag{77}$$

Now, consider any policy $\pi \in AC$. Similar to the Proof of Lemma 22, we define its accepted prices in G_1 and G_2 are $V_{s,0}$ and $V_{s,0} + \beta$ where $\beta \in [0, C_2 T^{-frac12+\eta}]$. We also denote the expected proposed price in both group as $V_{r,0} = V_{s,0} + \alpha_0$. Also, define a corresponding policy $\check{\pi}$:

$$\check{\pi} := \begin{cases} G_1 : \mathbb{E}[\text{accepted price}] = V_{s,0}, \mathbb{E}[\text{proposed price}] = V_{s,0} + \alpha_0 \\ G_2 : \mathbb{E}[\text{accepted price}] = V_{s,0}, \mathbb{E}[\text{proposed price}] = V_{s,0} + \alpha_0. \end{cases} \tag{78}$$

Notice that $\pi^1 = (A_1(V_{s,0}, 0))^{-1}[1, V_{r,0}, 0]^\top$ and $\pi^2 = (A_2(V_{s,0}, 0))^{-1}[1, V_{r,0}, \beta \cdot \mathbb{1}^\top F_2 \pi^2]^\top$. In comparison, we have $\check{\pi}^1 = (A_1(V_{s,0}, 0))^{-1}[1, V_{r,0}, 0]^\top$ and $\check{\pi}^2 = (A_2(V_{s,0}, 0))^{-1}[1, V_{r,0}, 0]^\top$. Therefore, we have:

$$\begin{aligned} \pi^1 &= \check{\pi}^1 \\ \|\pi^2 - \check{\pi}^2\|_1 &= \|(A_2(V_{s,0}, 0))^{-1}([1, V_{r,0}, \beta \cdot \mathbb{1}^\top F_2 \pi^2]^\top - [1, V_{r,0}, 0]^\top)\|_1 \\ &\leq \|(A_2(V_{s,0}, 0))^{-1}[0, 0, \beta]\|_1 \\ &= \|(A_2(V_{s,0}, 0))^{-1}_{[:,3]}\|_1 \cdot \beta \\ &\leq \frac{100\beta}{9(7V_s - 5)}. \end{aligned} \tag{79}$$

Also, since $F_{\min} \leq \mathbb{1}^\top F_2 \pi^2 \leq 1$ and $\|\mathbf{v}^\top F_2\|_1 \leq d$ always hold, we know that $\|\frac{\partial S_\zeta(\pi)}{\partial \pi^2}\| \leq \frac{d}{F_{\min}} \cdot \|\pi^2\|_1 = \frac{d}{F_{\min}}$. Since $V^*_{s,0} \approx \frac{8}{11}$ and all $V_{s,0}$ we consider are around it (According to (74)), we may assume that $(7V_s - 5) > \frac{1}{2} \cdot (\frac{8}{11} - \frac{5}{7}) > \frac{1}{200}$. Therefore, we have:

$$\begin{aligned} |S_\zeta(\check{\pi}) - S_\zeta(\pi)| &\leq \frac{d}{F_{\min}} \|\check{\pi}^2 - \pi^2\|_2 \\ &\leq \frac{100d\beta}{9(7 \cdot V_s - 5)F_{\min}} \\ &\leq \frac{100dC_2}{9(7 \cdot V_s - 5)F_{\min}} \cdot T^{-\frac{1}{2}+\eta} \\ &\leq \frac{3000dC_2}{F_{\min}} \cdot T^{-\frac{1}{2}+\eta}. \end{aligned} \tag{80}$$

Also, according to the proof of Lemma 22, we know that $|V_{s,0} - V_{s,0}^*| \leq (C_1 + LC_2)T^{-\frac{1}{2}+\eta}$ and $|\alpha_0^* - \alpha_0| \leq 0.6(C_1 + LC_2)T^{-\frac{1}{2}+\eta}$ (as $\zeta = 0$). Plugging in (56) and (57), we have:

$$
\begin{aligned}
\|\pi_{0,*}^1 - \check{\pi}^1\|_1 \leq & 50 \cdot ((120 + 8 + 10 + 10 \times (8 + 2))|\alpha_0^* - \alpha_0| + (13 + 106)|V_{s,0} - V_{s,0}^*|) \\
\leq & 1309(C_1 + LC_2)T^{-\frac{1}{2}+\eta} \\
\|\pi_{0,*}^2 - \check{\pi}^2\|_1 \leq & \frac{1}{3 \times 0.2 \times 3}((4 \times 36 + 120 + 24)|\alpha_0^* - \alpha_0| + (120 + 51 + 120 + 39)|V_{s,0} - V_{s,0}^*|) \\
\leq & 350(C_1 + LC_2)T^{-\frac{1}{2}+\eta}
\end{aligned}
\tag{81}
$$

Therefore, we have:

$$
\begin{aligned}
|S_\zeta(\pi_{0,*}) - S_\zeta(\check{\pi})| \leq & \frac{d}{F_{\min}}\|\pi_{0,*} - \check{\pi}\|_2 \\
= & \frac{d}{F_{\min}}(\|\pi_{0,*}^1 - \check{\pi}^1\|_2 + \|\pi_{0,*}^2 - \check{\pi}^2\|_2) \\
\leq & \frac{d}{F_{\min}}(\|\pi_{0,*}^1 - \check{\pi}^1\|_1 + \|\pi_{0,*}^2 - \check{\pi}^2\|_1) \\
\leq & \frac{d}{F_{\min}} \cdot (1309(C_1 + LC_2)T^{-\frac{1}{2}+\eta} + 350(C_1 + LC_2)T^{-\frac{1}{2}+\eta}) \\
\leq & \frac{d}{F_{\min}}2000(C_1 + LC_2)T^{-\frac{1}{2}+\eta}.
\end{aligned}
\tag{82}
$$

Recall that $C_1 = \frac{C}{W}$ and $C_2 = \frac{C}{W...L}$. Now, we let $W = 10^6 \frac{d}{F_{\min}}$. Therefore, we have:

$$
\begin{aligned}
S_\zeta(\pi) = & S_\zeta(\pi) - S_\zeta(\check{\pi}) + S_\zeta(\check{\pi}) - S_\zeta(\pi_{0,*}) + S_\zeta(\pi_{0,*}) \\
\geq & S_\zeta(\pi_{0,*}) - |S_\zeta(\pi) - S_\zeta(\check{\pi})| - |S_\zeta(\check{\pi}) - S_\zeta(\pi_{0,*})| \\
\geq & \frac{12\zeta(1 - 2\zeta)}{(11 - 2\zeta)(11 - 6\zeta)} - \frac{3000dC_2}{F_{\min}} \cdot T^{-\frac{1}{2}+\eta} - \frac{d}{F_{\min}}2000(C_1 + LC_2)T^{-\frac{1}{2}+\eta} \\
\geq & \frac{1}{20}\zeta - \frac{5000d(C_1 + LC_2)}{F_{\min}} \cdot T^{-\frac{1}{2}+\eta} \\
= & \frac{1}{20}C \cdot T^{-\frac{1}{2}+\eta} - \frac{5000dC}{F_{\min} \cdot W} \cdot T^{-\frac{1}{2}+\eta} \\
\geq & \frac{1}{20}C \cdot T^{-\frac{1}{2}+\eta} - \frac{1}{200}C \cdot T^{-\frac{1}{2}+\eta} \\
\geq & \frac{1}{30}C \cdot T^{-\frac{1}{2}+\eta}.
\end{aligned}
\tag{83}
$$

Let $C_{ac} = \frac{1}{30} \cdot C$ and this lemma holds. $\blacksquare$

Define $\mathbb{P}_{P_0}$ and $\mathbb{P}_{P_\zeta}$ as the probabilistic distribution of customer's feedback at each round. In order to increase the information for distinguishing between two problem settings, we assume that a customer would always tell us whether or not she accept the price \$1, at each time $t = 1, 2, \ldots, T$. Therefore, both $\mathbb{P}_{P_0}$ and $\mathbb{P}_{P_\zeta}$ are binomial distributions $B(T, 0.5)$ and $B(T, 0.5 - \zeta)$. Here we present another lemma, the Lemma 25, that indicates the hardness of distinguishing the two settings.

Lemma 25. *Consider the $N \geq \frac{T}{2}$ rounds when we play a policy in $AC \vee BD$. For any algorithm ϕ, denote $\phi_t = 1$ if $\pi_t \in AC$ and $\phi_t = 0$ if $\pi_t \in BD$. Then we have:*

$$
\max\{\mathbb{E}_{P_0}[\sum_{t=1}^N \phi_t], \mathbb{E}_{P_\zeta}[\sum_{t=1}^N (1 - \phi_t)]\} \geq \frac{1}{8}T \cdot \exp(-T^{2\eta}).
\tag{84}
$$

Proof of Lemma 25 . In fact, we have:

$$
\begin{aligned}
\max\{\mathbb{E}_{P_0}[\sum_{t=1}^{N}\phi_t], \mathbb{E}_{P_\zeta}[\sum_{t=1}^{N}(1-\phi_t)]\} &\geq \frac{\mathbb{E}_{P_0}[\sum_{t=1}^{N}\phi_t] + \mathbb{E}_{P_\zeta}[\sum_{t=1}^{N}(1-\phi_t)]}{2} \\
&= N \cdot \frac{\mathbb{P}_{P_0}[\phi_t == 1] + \mathbb{P}_{P_\zeta}[\phi_t == 0]}{2} \\
&\geq \frac{T}{4} \cdot (\mathbb{P}_{P_0}[\phi_t == 1] + \mathbb{P}_{P_\zeta}[\phi_t == 0]) \\
&\geq \frac{T}{8} \cdot \exp(-N \cdot KL(\mathbb{P}_{P_0}||\mathbb{P}_{P_\zeta})) \\
&\geq \frac{T}{8} \cdot \exp(-N \cdot KL(Ber(0.5)||Ber(0.5-\zeta))) \\
&\geq \frac{T}{8} \cdot \exp(-N \cdot 12\zeta^2) \\
&= \frac{T}{8} \cdot \exp(-N \cdot 12(C \cdot T^{-\frac{1}{2}+\eta})^2) \\
&\geq \frac{T}{8} \cdot \exp(-12C^2 T^{2\eta}) \\
&\geq \frac{T}{8} \cdot \exp(-T^{2\eta}).
\end{aligned}
\tag{85}
$$

Here the first line is for $\max \geq$ average, the second is by definition of ϕ_t, the third line is for $N \geq \frac{T}{2}$, the fourth line is from Fano's Inequality that $\mathbb{P}_0[\phi == 1] + \mathbb{P}_1[\phi == 0] \geq \frac{1}{2} \cdot \exp\{-N \cdot KL(\mathbb{P}_0||\mathbb{P}_1)\}$ for any distributions $\mathbb{P}_0$ and $\mathbb{P}_1$, the fifth line is by definition of P_0 and P_ζ that they are only different in the customers' feedback satisfying $Ber(0.5)$ and $\Pr = 0.5 - \zeta$ for some actions, respectively, the sixth line is from Lemma 19, the seventh line is for $\zeta = C \cdot T^{-\frac{1}{2}+\eta}$, the eighth line is for $N \leq T$, and the last line is for $12C^2 \leq 1$. ∎

With the two lemma above, we know that

- For any algorithm ϕ, we either run at least $\frac{T}{8} \cdot \exp(-T^{2\eta})$ rounds with some $\pi_t \in AC$ when the problem setting is P_ζ, or run at least $\frac{T}{8} \cdot \exp(-T^{2\eta})$ rounds with some $\pi_t \in BD$ when the problem setting is P_0, according to Lemma 25.

- For each round we mismatching the problem setting, we will suffer a $\min\{C_{ac}, C_{bd}\} \cdot T^{-\frac{1}{2}+\eta}$ unfairness, according to Lemma 24.

Given these two facts, denote $C_{\min} := \frac{1}{8}\min\{C_{ac}, C_{bd}\}$ and we at least have $C_{\min}T \cdot \exp(-T^{2\eta}) \cdot T^{-\frac{1}{2}+\eta}$ unfairness. For $x = \frac{C_x}{\log T}$ with any constant C_x, we let $\eta = \frac{3x}{2} = \frac{3C_x}{2\log T}$ and therefore $\eta > x$. As a result, we have

$$
\begin{aligned}
C_{\min}T \cdot \exp(-T^{2\eta}) \cdot T^{-\frac{1}{2}+\eta} &= C_{\min}\exp(-T^{2\eta} + \eta\log T)T^{\frac{1}{2}} \\
&= C_{\min}\exp(-T^{2\cdot\frac{C_x}{\log T}} + \frac{3C_x}{2\log T} \cdot \log T)T^{\frac{1}{2}} \\
&= C_{\min}\exp(-\exp(2 \cdot \frac{C_x}{\log T} \cdot \log T) + \frac{3C_x}{2})T^{\frac{1}{2}} \\
&= C_{\min}\exp(-\exp(2C_x) + \frac{3C_x}{2})T^{\frac{1}{2}}
\end{aligned}
\tag{86}
$$

Let $C_u = \frac{C_{\min}\exp(-\exp(2C_x)+\frac{3C_x}{2})}{2}$, and then the result of the equation above contradicts with the suppose that the unfairness does not exceed $C_u \cdot T^{\frac{1}{2}}$. Therefore, we have proved the theorem.

∎

C More Discussion

Here we discuss more on the problem settings we assumed, the techniques we used and the social impacts our algorithm might have, as a complement to Section 6.

C.1 Potential Generalizations of Current Problem Setting.

Currently we make a few technical assumptions that qualify the applications of our algorithm. In fact, these assumptions are mild and can be released by some tricks: On the one hand, we can always meet the requirement of Assumption 1 by reducing v_d. By running a binary-search algorithm for the highest acceptable price (with constant acceptance probability), we can find the feasible v_d within $O(\log T \log T)$ rounds (where $\log T$ for binary search and another $\log T$ for the concentration of a constant-expectation random variable, as we did in estimating $F_{\min}$). Since $O(\log T \log T)$ is much smaller than $O(\sqrt{T})$ as the optimal regret and unfairness, this would not harm the regret and unfairness substantially. On the other hand, we assume the prices to be chosen from a fixed and finite price set $\mathbf{V}$, which not only restricted our action but might lead to suboptimality from the perspective of a larger scope. In fact, if we allow the prices to be selected in the whole $[0, 1]$ range, a pricing policy can be a tuple of two continuous distributions over $[0, 1]$. To solve this problem, we may parametrize the distribution and learn the best parameters. We may also discretize the price space into small grids, i.e. prices are $\mathbf{V} = \{\gamma, 2\gamma, \ldots, (d-1)\gamma, d\gamma = 1\}$, where $\gamma = T^{-\alpha}$ with some constant α and d$= T^{\alpha}$ as a consequence. It is intrinsically a specific way of parametrization. According to the "half-Lipschitz" nature of pricing problem as well as our Lemma 14 along with the Lipschiness of $S(\pi; F_1, F_2)$, we know that the per-round discretization error would be upper bounded by $O(T^{-\alpha})$. Let the cumulative discretization error $O(T^{\frac{1}{-\alpha}})$ balances the cumulative regret (or substantive unfairness), i.e., $O(d^{\frac{3}{2}}\sqrt{T}) = 0(T^{\frac{1}{2}+\frac{3\alpha}{2}})$, we can achieve an upper bound on both the regret at $O(T^{\frac{4}{5}})$ by letting $\alpha = T^{\frac{1}{5}}$. However, this is not necessarily optimal as our original results in theorem 6 only match the lower bounds w.r.t. T but not to d. Therefore, it would also be an interesting problem to see the minimax regret/unfairness dependent w.r.t. d.

Besides of the assumptions we have made, there are other notions regarding our problem setup that can be generalized. Firstly, we may generalize our problem setting from two groups to multiple $G \geq 3$ groups. Again, the feasible set is not empty as we can always propose the same fixed price to all groups. However, there is not a directly generalization of the fairness definition, which we defined as the difference of the expectation of certain amount between two groups. We might defined it as "pairwise unfairness" by comparing the same difference among each pair of groups and adding them up, as we mentioned in section 6. However, this notion might not be a proper way to quantify the unfairness people suffer: Consider the case where the expected proposed prices in $(G-1)$ groups are very high and the last one is very low, and compare this case with another case where the expected proposed prices in 1 groups are very high and in the other $(G-1)$ groups are very low. The unfairness in these two cases should be definitely different, as the first seems more acceptable (i.e., being nice to only the minority versus being nice to only the majority). However, their "pairwise unfairness" are exactly the same in these two cases. Therefore, a better notion of procedural/substantive unfairness should be established for multiple $G \geq 3$ groups before we design and run fair pricing algorithms.

Secondly, we may generalize the modeling on customers from i.i.d. to strategic. For example, what if a customer tries multiple times until getting the lowest price of the distribution for this group. This is an adaptive adversary and therefore very hard to deal with even in the simplest decision-making process such as bandits.

Thirdly, we may also include more fairness concern. Currently we are considering the two types of fairness, but we define the cumulative fairness based as the summation of expected per-round unfairness. This definition does not take into consideration the changing of policies. For example, if we propose a fair policy at each round, but the policies over time changes drastically, then it is hard for the customers to feel or experience such a fairness. In our algorithm design of FPA, we always play the same policy for at least $\frac{\tau_k}{2d} = \Omega(\sqrt{T})$ rounds as a batch until the policy changes. This is a long enough time period for customers to experience fairness since at least a $Omega(\sqrt{T})$ number of customers from both groups would come and buy items under the same policy according to the Law of Large Numbers. However, this would still cause a feeling of unfairness for the two customers who are arriving almost simultaneously but the policy is just changed after the first customer buy or decide not to buy. Therefore, there exists necessity for us to consider the time/individual fairness under this online pricing problem scheme.

C.2 Potential Generalization of Techniques

Here we discuss a little bit more on the probable extension of the techniques we developed in our algorithm design and analysis.

A Good-and-exploratory Policy Set Our algorithm FPA maintains and updates a "good-and-exploratory" policy A_k in each epoch. Each policy in this set performs close to the optimal policy in both regret and unfairness reductions. A similar idea in reinforcement learning related research exists in Qiao et al. (2022) where they select policies that visit each (horizon, state, action) tuple most sufficiently while ensuring that the policy is low in regret. In fact, if we imagine an "exploratory" policy as the one that would elevate the most "information" (i.e., that would reduce the most uncertainty), then the "good-and-exploratory" policy-selection process is equivalent to an "Upper Confidence Bound" method Lai and Robbins

(1985) where we always pull the arm with the highest upper confidence bound in a multi-armed bandit. The only difference is that: for traditional exploration-and-exploitation balancing algorithm, we only need to improve our estimation on the parameters of these optimal or near-optimal policies. However, in our problem setting, we have to guarantee a uniform error bound, i.e., we have to improve our estimation on all parameters instead of only those optimal-related ones. This is because that we have to improve the estimation on constraints as well as on the revenue function. In our algorithm design, we handle this problem by keeping eliminating a feasible policy set, which in turn releases the algorithm from estimating those unnecessary parameters. In a nutshell, our methods can be applied broadly in online-decision-making problems.

Unfairness Lower Bound Proof on Optimal Algorithms   The main idea of our proof of the $\Omega(\sqrt{T})$ unfairness lower bound on any algorithm with $O(\sqrt{T})$ optimal regret is to construct a trade-off on unfairness and regret between two adjacent problem settings. We first bound the "bad policies" away from each problem setting, to avoid those policies that are super fair in both setting but performs poor in both setting as well. Then we show that policies with small-enough regret and unfairness on one setting should suffer a large regret on the other. Finally we end the proof by showing that we will definitely make $\tilde{\Theta}(T)$ times of mistakes in expectation, according to information theory. We believe that this scheme can be used in proving a variety of trading-off lower bounds.

C.3   Social Impacts

In this work, we develop methods to prompt the procedural and substantive fairness of customers from all groups. We believe that our techniques and results would enhance the unity of people with different gender, race, age, cultural backgrounds, and so on. However, it is definitely correct that we have to *treat differently* to different group of people. In order to ensure the fairness from customers' perspective, the seller is required to behave unfairly. Of course we could partly get rid of this issue by leaving the generating process of a random price to the nature, i.e., we let each customer draw a coupon from a box randomly. However, this only means that the seller's pricing process is fair but not leads to a fair result, as customers' coupon varies a lot from person to person. This turn out to be the exact issue named as "pricing and price fairness" proposed in Chapuis (2012) regarding the fairness of a seller's behavior. Maybe in the future we could develop an algorithm that is not only profitable but also ensures the fairness from both the seller and the customers' perspective, which could be a truly "doubly fair" dynamic pricing.