Fairness-aware Online Price Discrimination with Nonparametric Demand Models

Xi Chen*
Leonard N. Stern School of Business, New York University, xc13@stern.nyu.edu

Xuan Zhang
University of Illinois at Urbana-Champaign, xuan6@illinois.edu

Yuan Zhou
BIMSA, timzhouyuan@gmail.com

Price discrimination, which refers to the strategy of setting different prices for different customer groups, has been widely used in online retailing. Although it helps boost the collected revenue for online retailers, it might create serious concern in fairness, which even violates the regulation and law. This paper studies the problem of dynamic discriminatory pricing under fairness constraints. In particular, we consider a finite selling horizon of length $T$ for a single product with two groups of customers. Each group of customers has its unknown demand function that needs to be learned. For each selling period, the seller determines the price for each group and observes their purchase behavior. While existing literature mainly focuses on maximizing revenue, ensuring fairness among different customers has not been fully explored in the dynamic pricing literature. In this work, we adopt the fairness notion from (Cohen et al. 2021a). For price fairness, we propose an optimal dynamic pricing policy in terms of regret, which enforces the strict price fairness constraint. In contrast to the standard $\sqrt{T}$-type regret in online learning, we show that the optimal regret in our case is $\tilde{O}(T^{4/5})$. We further extend our algorithm to a more general notion of fairness, which includes demand fairness as a special case. To handle this general class, we propose a soft fairness constraint and develop the dynamic pricing policy that achieves $\tilde{O}(T^{4/5})$ regret.

Key words: Dynamic pricing, Demand Learning, Fairness, Nonparametric Demands

1. Introduction

Data-driving algorithms have been widely applied to automated decision-making in operations management, such as personalized pricing, online recommendation. Traditional operational decisions mainly seek ”globally optimal” decisions. However, such decisions could be unfair to a certain population segment (e.g., a demographic group or protected class). The issue of fairness is particularly critical in e-commerce. Indeed, the increasing prominence of e-commerce has given retailers unprecedented power to understand customers as individuals and to implement discriminatory pricing strategies that could be unfair to a specific customer group. As pointed by Wallheimer

* Author names listed in alphabetical order.
For example, the study in Pandey & Caliskan (2021) analyzes more than 100 million ride-hailing observations in Chicago. It shows that higher fare prices appear in neighborhoods with larger “non-white populations, higher poverty levels, younger residents, and higher education levels.” In the auto loan market, “the Black and Hispanic applicant’s loan approval rates are 1.5 percentage points lower, even controlling for creditworthiness” (Butler et al. 2021). Amazon once charged the customers who discussed DVDs at the website DVDTalk.com more than 40% than other customers for buying DVDs (Streitfeld 2000). As a consequence, Amazon publicly apologized and made refunds to 6,896 customers. In addition to customers’ backfiring, many regulations have been made to ensure fairness in various industries, e.g., regulations by Consumer Financial Protection Bureau (CFPB) for charges in financial services and by Equal Employment Opportunity Commission (EEOC) for employment screening. Violation of these regulations in the decision-making cycle could lead to severe consequences. Despite the importance of fairness in decision-making, the study of fairness-aware in dynamic pricing is still somewhat limited.

This paper studies the problem of dynamic pricing with nonparametric demand functions under fairness constraints. There is a wide range of fairness notations from the online learning community. However, most of these notions do not fit the pricing applications (see the related work in Section 2 for more discussions). Instead, we adopt the fairness notation from a recent paper on static pricing in the operations management literature (Cohen et al. 2021a). To highlight our main idea, we consider the simplest setup of monopoly selling over \( T \) periods to two customer groups without any inventory constraints. Although the two-group setting might be too simple in practice, it serves as the foundation in studying group fairness. Each group of customers has its underlying demand function in price, which is unknown to the seller. At each period \( t = 1, \ldots, T \), the seller offers a single product to each group \( i \in \{1, 2\} \) with the price \( p_1^{(t)} \) and \( p_2^{(t)} \), and observes the realized demands from each group. Most existing dynamic pricing literature only focuses on learning the demand function to maximize revenue over time. This work enforces the fairness constraints into this dynamic pricing problem with demand learning. We first consider a natural notion of price fairness introduced by Cohen et al. (2021a). Later, we will extend to a more general fairness notion, which includes demand fairness as a special case. According to Cohen et al. (2021a), price fairness has been widely adopted by regulations such as those from Federal Deposit Insurance Corporation (FDIC). More specifically, let \( p_1^\star \) and \( p_2^\star \) denote the optimal price for each customer group without any fairness constraint. Price fairness requires that for all time periods \( t \), we have

\[
|p_1^{(t)} - p_2^{(t)}| \leq \lambda |p_1^\star - p_2^\star|.
\]
The parameter $\lambda \in (0, 1)$ controls the fairness level, which should be pre-defined by the seller to meet a specific internal or external regulation. The smaller the value of $\lambda$, the more strict fairness constraint the seller needs to achieve. We note that a concurrent work by Cohen et al. (2021b) studies a dynamic pricing under the absolute price fairness with parametric demand function, which restricts the price difference to be upper bounded by a fixed constant. We believe the fairness definition in (1) could be practically more favorable in some scenarios as it controls the relative gap between the offered prices and the gap between the optimal prices. If the gap between the optimal prices is large (e.g., two customer groups are very heterogeneous), it is reasonable to allow more difference in price between the two groups. More discussions on the difference from (Cohen et al. 2021b) will be provided in Section 2. For the ease of presentation, we refer to such a constraint in Eq. (1) as the “hard fairness constraint”, in contrast to the “soft fairness constraint” introduced later, which only requires the pricing policy to approximately satisfy the fairness constraint. The main goal of this paper is to develop an efficient dynamic pricing policy that ensures the fairness constraint in Eq. (1) over the entire selling horizon and to quantify the revenue gap between the dynamic pricing and static pricing under the fairness constraint.

For the price fairness constraint in Eq. (1), we developed a dynamic pricing algorithm that achieves the regret at the order of $\tilde{O}(T^{4/5})$, where $\tilde{O}(\cdot)$ hides logarithmic factors in $T$. The cumulative regret is defined as the revenue gap between our pricing policy and the static clairvoyant prices under the fairness constraint. This regret bound is fundamentally different from the $\sqrt{T}$-type bound in ordinary dynamic pricing (Broder & Rusmevichientong 2012, Wang et al. 2014) due to the intrinsic difficulty raised by the fairness constraint. Please see Section 2 on related works for more discussions and comparisons.

Our pricing policy contains three stages. The first stage tries to estimate the optimal prices $p_1^*$ and $p_2^*$ without fairness constraints. By leveraging the common assumption that the revenue is a strongly concave function in demand (Jasin 2014), we developed a tri-section search algorithm to obtain the estimates $\hat{p}_1^*$ and $\hat{p}_2^*$. The estimates $\hat{p}_1^*$ and $\hat{p}_2^*$ from the first stage enable us to construct an approximate fairness constraint. Based on $\hat{p}_1^*$ and $\hat{p}_2^*$, the second stage uses a discretization technique to estimate optimal prices $\hat{p}_1^*$ and $\hat{p}_2^*$ of the static pricing problem under the fairness constraint. For the rest of the time periods, we will offer the prices $\hat{p}_1^*$ and $\hat{p}_2^*$ to two groups of customers. The policy is easy to implement as it is essentially an explore-exploit scheme. We further establish the information-theoretical lower bound $\Omega(T^{4/5})$ to show that the regret of the policy is optimal up to a logarithmic factor.

The second part of the paper extends the price fairness to a general fairness measure. Consider a fairness measure $M_i(p)$ for each group $i \in \{1, 2\}$, where $M_i$ is a Lipschitz continuous function in price. For example, the case $M_i(p) = p$ reduces the price fairness. We could also consider demand
fairness by defining $M_i(p) = d_i(p)$, where $d_i$ is the expected demand for the $i$-th customer group. Now the fairness constraint can be naturally extended to $|M_1(p_1) - M_2(p_2)| < \lambda |M_1(p_1^\#) - M_2(p_2^\#)|$. However, in practice, it is impossible to enforce such a hard constraint over all time periods since the fairness measure $M_i$ is unknown to the seller. For example, in demand fairness, the demand function $d_i$ needs to be learned via the interactions between the seller and customers. To this end, we propose the “soft fairness constraint”, which adds the following penalty term to the regret minimization problem,

$$
\gamma \max \left( \left| M_1(p_1^{(t)}) - M_2(p_2^{(t)}) \right| - \lambda |M_1(p_1^\#) - M_2(p_2^\#)|, 0 \right), \tag{2}
$$

where the parameter $\gamma$ balances between the regret minimization and the fairness constraint. When $\gamma = 0$, there would be no fairness constraint; while when $\gamma = \infty$, it is equivalent to “hard fairness constraint”. Under the general fairness measure $M_i$ and the soft fairness constraint in Eq. (2), we develop a dynamic pricing policy, which achieves the penalized regret at the order of $\tilde{O}(T^{4/5})$ for $\gamma \leq O(1)$.

With the problem setup, the main technical contributions of this paper are summarized as follows.

(a) On the algorithm side, we design a two-stage exploration procedure (corresponding to the first two stages of the algorithm) to learn the fairness-aware optimal prices. While the tri-section and discretization techniques have been used in pricing literature (see, e.g., (Lei et al. 2014) and (Wang et al. 2014) respectively), we adapt them to the new fairness constraint and make sure that the constraint is not violated even during the exploration stages. The efficiency of the new exploration procedure becomes worse (due to the fairness constraint) compared to the ordinary pricing problem, and the balance between exploration and exploitation also changes. We find this new optimal tradeoff between exploration and exploitation, which leads to the $\tilde{O}(T^{4/5})$ regret.

To establish this regret bound, we establish a key structural result between the price gap of the constrained optimal solution and that of the unconstrained optimal solution (see Lemma 4 in Section 4.2). At a higher level, the lemma states that the optimal fairness-aware pricing strategy should utilize the “discrimination margin” allowed by the fairness constraint as much as possible. This result could shed light on other fairness-aware problems.

\footnote{As discussed before, an extremely large $\gamma$ enforces the hard fairness constraint which is almost impossible to be met due to the unknown $M_i$. Therefore, we have to assume an upper bound on $\gamma$ to control the penalized regret. In our result, $\gamma$ can be as large as $O(1)$ to achieve the desired regret bound. This is a mild constraint since the maximum possible profit per selling period is also $O(1)$, and it makes sense to assume that the penalty imposed due to the fairness violation is comparable to the profit.}
(b) On the lower bound side, we show (somewhat surprisingly) that the compromise of the two-stage exploration is necessary and our $\widetilde{O}(T^{4/5})$ regret is indeed optimal. The lower bound construction is technically quite non-trivial. In particular, we construct pairs of hard instances where 1) the demand functions are similar to their counterparts in the pair, and 2) the unconstrained optimal prices for each demand function are quite different, which leads to different fairness constraints. By contrasting these two properties, we are eventually able to derive that $\Omega(T^{4/5})$ regret has to be paid in order to properly learn the fairness constraint and the optimal fairness-aware prices. An additional layer of challenge in our lower bound proof is that our constructed hard instances has to satisfy the standard assumptions in pricing literature (such as demands inversely correlated with prices and the law of diminishing returns) in order to become a real pricing problem instance, where in contrast the usual online learning lower bounds (such as bandits) do not have such requirements. We also note that in the existing lower bounds in dynamic pricing (Besbes & Zeevi 2009, Broder & Rusmevichientong 2012, Wang et al. 2014), it suffices to analyze the linear demand functions, which are relatively simple and automatically satisfy the assumptions. In our work, however, we have to construct more complicated demand functions and it requires substantially more technical efforts to make these functions also satisfy the desired assumptions. We believe that the technical tools developed in this paper considerably enrich the lower bound techniques in dynamic pricing literature.

The rest of the paper is organized as follows. In Section 2, we review the relevant literature in dynamic pricing and fairness-aware online learning. We formally introduce the problem setup in Section 3. Section 4 provides the dynamic pricing policy under price fairness and establishes the regret upper bound. The matching lower bound is provided in Section 5. In Section 6, we extend the price fairness to the general fairness measure and develop the corresponding dynamic pricing algorithm. The numerical simulation study will be provided in Section 7, followed by the conclusion in Section 8. Some technical proofs are relegated to the supplementary material.

2. Related Works

There are two lines of relevant research: one on dynamic pricing and the other on fairness machine learning. This section briefly reviews related research from both lines.

**Dynamic Pricing.** Due to the increasing popularity of online retailing, dynamic pricing has become an active research area in the past decade. Please refer to (Bitran & Caldentey 2003, Elmaghraby & Keskinocak 2003, den Boer 2015) for comprehensive surveys. We will only focus on the single-product pricing problem. The seminal work by Gallego & Van Ryzin (1994) laid out
the foundation of dynamic pricing. Earlier work in dynamic pricing often assumes that demand information is known to the retailer a priori. However, in modern retailing industries, such as fast fashion, the underlying demand function cannot be easily estimated from historical data. This motivates a body of research on dynamic pricing with demand learning (see, e.g., Araman & Caldentey (2009), Besbes & Zeevi (2009), Farias & Van Roy (2010), Broder & Rusmevichientong (2012), Harrison et al. (2012), den Boer & Zwart (2013), Keskin & Zeevi (2014), Wang et al. (2014), Lei et al. (2014), Chen et al. (2015), Miao et al. (2019), Wang et al. (2021) and references therein).

Along this line of research, Besbes & Zeevi (2009) first proposed a separate explore-exploit policy, which lead to sub-optimal regret of $\tilde{O}(T^{3/4})$ for nonparametric demands and $\tilde{O}(T^{2/3})$ for parametric demands. Wang et al. (2014) improved this result by developing joint exploration and exploitation policies that achieve the optimal regret of $\tilde{O}(T^{1/2})$. Lei et al. (2014) further improved the result by removing the logarithmic factor in $T$. For more practical considerations, den Boer & Zwart (2013) proposed a controlled variance pricing policy and Keskin & Zeevi (2014) proposed a semi-myopic pricing policy for a class of parametric demand functions. Broder & Rusmevichientong (2012) established the lower bound of $\Omega(\sqrt{T})$ for the general dynamic pricing setting and proposed a $O(\log T)$-regret policy when demand functions satisfy a “well-separated” condition. In addition, several works proposed Bayesian policies for dynamic pricing (Farias & Van Roy 2010, Harrison et al. 2012).

As compared to the obtained regret bounds in existing dynamic pricing literature, the fairness constraint would completely change the order of the regret. Our results show that with the fairness constraint, the optimal regret becomes $\tilde{O}(T^{4/5})$. In our algorithm, the first stage of the pure exploration phase, i.e., learning the difference $|M_1(p_1^*) - M_2(p_2^*)|$ in the fairness constraint, is the most regret-producing step. The popular learning-while-doing techniques (such as the Upper Confidence Bound and Thompson Sampling algorithms) in many existing dynamic pricing and online learning paper seem not helpful in our problem to further reduce the regret. Intuitively, this is due to the fundamental difference between exploring the fairness constraint and exploiting the (near-)optimal fairness-aware pricing strategy. Such an intuition has been rigorously justified by our lower bound theorem, showing that our explore-exploit algorithm cannot be further improved in terms of the minimax regret. The separation between our regret bound and the usual $\sqrt{T}$-type regret in dynamic pricing literature also illustrates the intrinsic difficulty from the information-theoretical perspective raised by the fairness constraint.

There are many interesting extensions of single product dynamic pricing, such as network revenue management (see, e.g., Gallego & Van Ryzin (1997), Ferreira et al. (2018), Chen & Shi (2019) and references therein), dynamic pricing in a changing environment (Besbes et al. 2015, Keskin & Zeevi 2016), infrequent price update with a limited number of price changes (Cheung et al. 2017),
and personalized pricing with potentially high-dimensional covariates (Nambiar et al. 2019, Ban & Keskin 2021, Lobel et al. 2018, Chen & Gallego 2021, Javanmard & Nazerzadeh 2019, Chen et al. 2021b,a). It would be interesting future directions to study how to ensure fairness under these more general dynamic pricing setups.

**Fairness.** The topic of fairness has been extensively studied in economics and recently attracts a lot of attention from the machine learning community. There is a wide range of different definitions of “fairness”, and many of them are originated from economics literature and are relevant to causal inference (e.g., the popular “predictive parity” definition (Kasy & Abebe 2021)). Due to space limitations, we will omit detailed discussions on these definitions and only highlight a few relevant to the online learning setting. The interested readers might refer to the book (Barocas et al. 2019) and the survey (Hutchinson & Mitchell 2019) for a comprehensive review of different notions of fairness.

One classical notion of fairness is the “individual fairness” introduced by Dwork et al. (2012). Considering an action $a$ that maps a context $x \in X$ to a real number, the individual fairness is essentially the Lipschitz continuity of the action $a$, i.e., $|a(x_1) - a(x_2)| \leq \lambda \cdot d(x_1, x_2)$, where $d(\cdot, \cdot)$ is a certain distance metric. The notion of “individual fairness” can be extended to the “fairness-across-time” and “fairness-in-hindsight” for a sequence of decisions (Gupta & Kamble 2019), which requires that actions cannot be changed too fast over time nor too much over different contextual information. However, we believe that individual fairness might not suit the pricing problem since society is more interested in protecting different customer groups. In contrast, the fairness notation adopted in our paper can be viewed as a kind of group fairness. Other types of group-fairness have been adopted in different types of operations problems, e.g., group bandit models (Baek & Farias 2021), online bipartite matching (Ma et al. 2021), and fair allocation (Cai et al. 2021). For multi-armed bandit models, one notion of fairness introduced by Liu et al. (2017) is the “smooth fairness”, which requires that for two arms with similar reward distributions, the choice probabilities have to be similar. Another popular notion of fairness in bandit literature is defined as follows: if the arm/choice $A$ has higher expected utility than another arm $B$, the arm $A$ should have a higher chance to be pulled (Joseph et al. 2016). In the auto-loan example, it means a more-qualified applicant should consistently get a higher chance of approval. A similar notion of fairness based on $Q$ functions has been adopted by Jabbari et al. (2017) in reinforcement learning. However, these notions of fairness are not designed to protect different customer groups. In pricing applications (e.g., auto-loan example), these fairness definitions may not capture the requirement of regulations.

In recent years, fairness has been incorporated into a wide range of operations problems. For example, Chen & Wang (2018) investigated the fairness of service priority and price discount in
a shared last-mile transportation system. Bateni et al. (2016) studied fair resource allocation in
the ads market. It adopts weighted proportional fairness proposed by Bertsimas et al. (2011).
Balseiro et al. (2021) introduced the regularized online allocation problem and proposed a primal-
dual approach to handle the max-min fairness regularizer. Kandasamy et al. (2020) studied online
demand estimation and allocation under the max-min fairness with applications to cloud comput-
ing. In the pricing application considered in this paper, research has been devoted to game-theoretic
models with fairness constraint in duopoly markets (see, e.g., Li & Jain (2016) and references
therein). In contrast, we consider a monopoly market and thus do not adopt game-theoretical
modeling. For the static pricing problem with known demand functions, Kallus & Zhou (2021) for-
mulated a multi-objective optimization problem, which takes the price parity and long-run welfare
into consideration. Cohen et al. (2021a) proposed new fairness notions designed for the pricing
problem (e.g., fairness in price and demand) and investigated the impact of these types of fair-
ness on social welfare. As the group fairness in (Cohen et al. 2021a) captures practical regulation
requirements, our work is built on these fairness notions and extends them to dynamic pricing with
nonparametric demand learning. We also note that the work by Cohen et al. (2021a) has already
studied the impact of the fairness constraint in the static problem, which is measured by the rev-
ue gap between the static pricing problem with and without fairness constraint. For example,
Proposition 2 in (Cohen et al. 2021a) characterizes the revenue loss as a function of λ. Therefore,
we will focus on the revenue gap between dynamic and static problems both under the fairness
constraint. We also note that Cohen et al. (2021a) extends the fairness notion to K groups, i.e., for
any 1 ≤ i < j ≤ K, it requires |M_i(p_i(t)) − M_j(p_j(t))| ≤ λ max_{1 ≤ k < l ≤ K} |M_k(p_k^#) − M_l(p_l^#)|. As a future
direction, it would be interesting to extend our policy to handle multiple customer groups.

A very recent work by Cohen et al. (2021b) studies the learning-while-doing problem for dynamic
pricing under fairness. The key difference is that they defined the fairness constraint as an absolute
upper bound of the price gap between different groups (i.e., |p_1 − p_2| ≤ C for some fixed constant C),
while we consider a relative price gap in (1). In addition, they studied a simple parametric demand
model in the form of generalized linear model in price. In contrast, we allow a fully nonparametric
demand model without any parametric assumption of d(·). Under the absolute price gap with
parametric demand models, the work by Cohen et al. (2021b) is able to achieve a √T-type regret.
On the other hand, the relative fairness requires us to learn the optimal prices p_1^# and p_2^#
for each individual group under nonparametric demand functions, which leads to a larger regret of
O(T^{4/5}). The optimality of the O(T^{4/5}) regret also indicates that our problem is fundamentally more difficult
than the one in (Cohen et al. 2021b).

Finally, we assume the protected group information is available to the seller. Recent work by
Kallus et al. (2021) studied how to assess fairness when the protected group membership is not
observed in the data. It would also be an interesting direction to extend our work to the setting with hidden group information.

3. Problem Formulation

We consider a dynamic discriminatory pricing problem with fairness constraints. Suppose that there are $T$ selling periods in total, and two groups of customers (labeled by 1 and 2). At each selling period $t = 1, 2, \ldots, T$, the seller offers a single product, with a marginal cost $c \geq 0$, to two groups of customers. The seller also decides a price $p_i(t) \in [p_l, p_u]$ for each group $i \in \{1, 2\}$, where $[p_l, p_u]$ is the known feasible price range. We assume that each group $i \in \{1, 2\}$ of customers has its own demand function $d_i(\cdot) : [p_l, p_u] \rightarrow [0, 1]$, where $d_i(p)$ is the expected demand from group $i$ when offered price $p$. The demand functions $\{d_i(\cdot)\}$ are unknown to the seller beforehand.

When offering the product to each group of customers, we denote the realized demand from group $i$ by $D_i(t) \in [0, 1]$ (up to normalization), which is essentially a random variable satisfying $E[D_i(t) \mid p_i(t), F_{t-1}] = d_i(p_i)$ and $F_{t-1}$ is the natural filtration up to selling period $t-1$. For example, when $D_i(t)$ follows a Bernoulli distribution with mean $d_i(p_i)$, the binary value of $D_i(t)$ represents whether the customer group $i$ makes a purchase (i.e., $D_i(t) = 1$) or not (i.e., $D_i(t) = 0$). By observing $D_i(t)$, the seller earns the profit $\sum_{i \in \{1, 2\}} (p_i(t) - c)D_i(t)$ at the $t$-th selling period.

If the seller has known the demand functions $\{d_i(\cdot)\}$ beforehand, and is not subject to any fairness constraint, her optimal prices for two groups are the following unconstrained clairvoyant solutions:

$$p_i^* = \arg \max_{p \in [p_l, p_u]} R_i(p) := (p - c)d_i(p), \quad \forall i \in \{1, 2\},$$

(3)

where $R_i(p)$ is the expected single-period revenue function for group $i$. Following the classical pricing literature (Gallego & Van Ryzin 1997), under the one-to-one correspondence between the price and demand and other regularity conditions, we could also express the price as a function of demand for each group (i.e., $p_i(d)$ for $i \in \{1, 2\}$). This enables us to define the so-called revenue-demand function, which expresses the revenue as a function of demand instead of price: $R_i(d) := (p_i(d) - c)d$. For commonly used demand functions (e.g., linear, exponential, power, and logit), the revenue-demand function is concave in the demand. The concavity assumption is widely assumed in pricing literature (see, e.g., Jasin (2014)) and is critical in designing our policy. We also note that the revenue function is not concave in price for many examples (e.g., logit demand).

Now, we are ready to formally introduce the fairness constraint. Let $M_i(p)$ denote a fairness measure of interest for group $i$ at price $p$, where $M_i$ can be any Lipschitz function. When $M_i(p) = p$, it reduces to the price fairness in (Cohen et al. 2021a). When $M_i(p) = d_i(p)$, it corresponds to
the demand fairness in (Cohen et al. 2021a). For any given fairness measure, the hard fairness constraint requires that

\[ |M_1(p_1^{(t)}) - M_2(p_2^{(t)})| \leq \lambda |M_1(p_1^t) - M_2(p_2^t)|, \quad \forall t \in \{1, 2, 3, \ldots, T\}, \tag{4} \]

where \( \lambda > 0 \) is the parameter for the fairness level that is selected by the seller to meet internal goals or satisfy regulatory requirements. The smaller \( \lambda \) is, the more strict fairness constraint the seller has to meet. We also note that the parameter \( \lambda \) is not a tuning-parameter of the algorithm. Instead, this is the fundamental fairness level that should be pre-determined by the seller either based on a certain internal/external regulation or on how much revenue the seller is willing to sacrifice (see Eq. (6)).

With the fairness measure in place, let \( \{p_1^*, p_2^*\} \) denote the fairness-aware clairvoyant solution, i.e., the optimal solution to the following static optimization problem,

\[
\max_{p_1, p_2 \in [p, \bar{p}]} R_1(p_1) + R_2(p_2), \tag{5}
\]

subject to \(|M_1(p_1) - M_2(p_2)| \leq \lambda |M_1(p_1^*) - M_2(p_2^*)|\).

For the ease of notation, we omit the dependency of \( \{p_1^*, p_2^*\} \) on \( \lambda \). We also note that the work by Cohen et al. (2021a) quantifies the tradeoff between the strictness of the fairness constraint and the overall revenue in the static problem. In particular, it shows that for linear demands and price or demand fairness,

\[
R_1(p_1^*) + R_2(p_2^*) - (R_1(p_1^*) + R_2(p_2^*)) = O((1 - \lambda)^2). \tag{6}
\]

In other words, the revenue loss due to imposing a stronger fairness constraint (as \( \lambda \) decreases to zero) grows at the rate of \( O((1 - \lambda)^2) \). As this tradeoff has already been explored in (Cohen et al. 2021a), the main goal of our paper is on developing an online policy that can guarantee the fairness through the entire time horizon.

In the learning-while-doing setting where the seller does not know the demand beforehand, she has to learn demand functions during selling periods, and maximize her total revenue, while in the meantime, obeying the fairness constraint. Equivalently, the seller would like to minimize the regret, which is the difference between her expected total revenue and the fairness-aware clairvoyant solution:

\[
\text{Reg}_T := \mathbb{E} \sum_{t=1}^{T} \left[ R_1(p_1^t) + R_2(p_2^t) - R_1(p_1^*) - R_2(p_2^*) \right], \tag{7}
\]

where \( p_1^* \) and \( p_2^* \) are the fairness-aware clairvoyant solution defined in Eq. (5).

In this paper, we will first focus on price fairness (i.e., \( M_i(p) = p \)) and establish matching regret upper and lower bounds for price-fairness-aware dynamic pricing algorithms. We will then
extend our algorithm to general fairness measure function \(\{M_i(p)\}\). However, in practical scenarios where \(M_i(p)\) is not accessible to the seller beforehand, only the noisy observation \(M_i^{(t)}\) with 
\[
\mathbb{E}[M_i^{(t)} | p_i^{(t)}, \mathcal{F}_{t-1}] = M_i(p_i^{(t)}) \quad (\text{for } i \in \{1, 2\})
\]
is revealed after the seller’s pricing decisions during selling period \(t\). A natural example is the demand fairness, where \(M_i(p) = d_i(p)\), and the seller could only observe a noisy demand realization at the offered price. In this case, it is impossible for the seller to satisfy the hard constraint in Eq. (4) at first a few selling periods (as there is a limited number of observations of \(M_i\) available). To this end, we propose the “soft fairness constraint” and add the soft fairness constraint as a penalty term to the regret minimization problem. In particular, the penalized regret incurred at time \(t\) takes the following form,
\[
\left[R_1(p_1^t) + R_2(p_2^t) - R_1(p_1^{(t)}) - R_2(p_2^{(t)})\right] + \gamma \max \left(\left|\left|M_1(p_1^{(t)}) - M_2(p_2^{(t)})\right| - \lambda \left|\left|M_1(p_1^t) - M_2(p_2^t)\right|, 0\right\right\right),
\tag{8}
\]
where the first term is the standard regret, the second term is the penalty term for violating the fairness constraint, and \(\gamma\) is a pre-defined parameter to balance between the regret and the fairness constraint. Subsequently, for general fairness measure, the seller aims to minimize the following cumulative penalized regret:
\[
\text{Reg}_T^{\text{soft}} := \mathbb{E} \sum_{t=1}^{T} \left\{ \left[R_1(p_1^t) + R_2(p_2^t) - R_1(p_1^{(t)}) - R_2(p_2^{(t)})\right] + \gamma \max \left(\left|\left|M_1(p_1^{(t)}) - M_2(p_2^{(t)})\right| - \lambda \left|\left|M_1(p_1^t) - M_2(p_2^t)\right|, 0\right\right\right) \right\}.
\]
Throughout the paper, we will make the following standard assumptions on demand functions and fairness measure:

**Assumption 1.** (a) The demand-price functions are monotonically decreasing and injective Lipschitz, i.e., there exists a constant \(K \geq 1\) such that for each group \(i \in \{1, 2\}\), it holds that
\[
\frac{1}{K} |p - p'| \leq |d_i(p) - d_i(p')| \leq K |p - p'|, \quad \forall p, p' \in [\underline{p}, \overline{p}].
\]
(b) The revenue-demand functions are strongly concave, i.e., there exists a constant \(C\) such that for each group \(i \in \{1, 2\}\), it holds that
\[
R_i(\tau d + (1 - \tau)d') \geq \tau R_i(d) + (1 - \tau)R_i(d') + \frac{1}{2} C \tau(1 - \tau)(d - d')^2, \quad \forall d, d', \tau \in [0, 1].
\]
(c) The fairness measures are Lipschitz, i.e., there exists a constant \(K'\) such that for each group \(i \in \{1, 2\}\), it holds that:
\[
|M_i(p) - M_i(p')| \leq K' |p - p'|, \quad \forall p, p' \in [\underline{p}, \overline{p}].
\]
Algorithm 1: Fairness-aware Dynamic Pricing with Demand Learning

1. For each group $i \in \{1, 2\}$, run $\text{EXPLOREUNCONSTRAINEDOPT}$ (Algorithm 2) separately with the input $z = i$, and obtain the estimate of the optimal price without fairness constraint $\hat{p}^i_*$.

2. Given $\hat{p}^1_*$ and $\hat{p}^2_*$, run $\text{EXPLOREUNCONSTRAINEDOPT}$ (Algorithm 3), and estimate the optimal price under the fairness constraint $(\hat{p}^1_*, \hat{p}^2_*)$.

3. For each of the remaining selling periods, offer the price $\hat{p}^i_*$ to each customer group $i \in \{1, 2\}$.

(d) There exits a constant $\bar{M} \geq 1$ such that the noisy observation $M^{(t)}_i \in [0, \bar{M}]$ for every selling period $t$ and customer group $i \in \{1, 2\}$.

Assumptions 1(a) and 1(b) are rather standard assumptions made for the demand-price and revenue-demand functions in pricing literature (see, e.g., Wang et al. (2014) and references therein). On the other hand, the fairness measure functions $M_i(p)$ are first studied in the context of dynamic pricing. Assumptions 1(c) and 1(d) assert necessary and mild regularity conditions on these functions and their noisy realizations.

In the rest of this paper, we will first investigate the optimal regret rate that can be achieved under the setting of price fairness. Once we obtain a clear understanding about price fairness, we will proceed to study more general fairness settings.

4. Dynamic Pricing Policy under Price Fairness

Starting with the price fairness (i.e., $M_i(p) = p$), we develop the fairness-aware pricing algorithm and establish its theoretical property in terms of regret.

Our algorithm (see Algorithm 1) runs in the explore-and-exploit scheme. In the exploration phase, the algorithm contains two stages. The first stage separately estimates the optimal prices for two groups without any fairness constraint. Using the estimates as input, the second stage learns the (approximately) optimal prices under the fairness constraint. Then the algorithm enters the exploitation phase, and uses the learned prices for each group to optimize the overall revenue. The algorithm is presented in Algorithm 1. Note that the algorithm will terminate whenever the time horizon $T$ is reached (which may happen during the exploration stage).

Now we describe two subroutines used in exploration phase: $\text{EXPLOREUNCONSTRAINEDOPT}$ (and $\text{EXPLORECONSTRAINEDOPT}$). We note that according to our theoretical results in Theorems 1 and 2, $\text{EXPLOREUNCONSTRAINEDOPT}$ and $\text{EXPLORECONSTRAINEDOPT}$ will only run in $\tilde{O}(T^{4/5})$ and $\tilde{O}(T^{3/5})$ time periods, respectively. Therefore, the estimated prices $\hat{p}^1_*$ and $\hat{p}^2_*$ will be offered for the most time periods in the entire selling horizon of length $T$. 
Algorithm 2: EXPLOREUNCONSTRAINEDOPT

**Input**: the customer group index \( z \in \{1, 2\} \)

**Output**: the estimated unconstrained optimal price \( \hat{p}^z \) for group \( z \)

1. \( p_L \leftarrow p, \ p_R \leftarrow \bar{p}, \ r \leftarrow 0; \)
2. while \(|p_L - p_R| > 4T^{-1/5}\) do
   3. \( r \leftarrow r + 1; \)
   4. \( p_{m1} \leftarrow \frac{2}{3}p_L + \frac{1}{3}p_R, \ p_{m2} \leftarrow \frac{1}{3}p_L + \frac{2}{3}p_R; \)
   5. Offer price \( p_{m1} \) to both customer groups for \( \frac{25K^4T^2}{C^2}T^{4/5} \ln T \) selling periods and denote the average demand from customer group \( z \) by \( \hat{d}_{m_1} \);
   6. Offer price \( p_{m2} \) to both customer groups for \( \frac{25K^4T^2}{C^2}T^{4/5} \ln T \) selling periods and denote the average demand from customer group \( z \) by \( \hat{d}_{m_2} \);
   7. if \( \hat{d}_{m_1}(p_{m1} - c) > \hat{d}_{m_2}(p_{m2} - c) \) then \( p_R \leftarrow p_{m2}; \) else \( p_L \leftarrow p_{m1}; \)
8. return \( \hat{p}^z = \frac{1}{2}(p_L + p_R); \)

The EXPLOREUNCONSTRAINEDOPT subroutine. Algorithm 2 takes the group index \( z \in \{1, 2\} \) as input, and estimates the unconstrained clairvoyant solution \( \hat{p}^z \) for group \( z \) (i.e., without the fairness constraint).

Algorithm 2 runs in a trisection fashion. The algorithm keeps an interval \([p_L, p_R]\) and shrinks the interval by a factor of \( 2/3 \) during each iteration while keeping the estimation target \( p^z \) within the interval. The iterations are indexed by the integer \( r \), and during each iteration, the two trisection prices \( p_{m_1} \) and \( p_{m_2} \) are selected. For either trisection price, both customer groups are offered the price (so that the price fairness is always satisfied) for a carefully chosen number of selling periods (as in Lines 5 and 6, where \( K \) and \( C \) are defined in Assumption 1), and the estimated demand from group \( z \) is calculated. Note that in practice, one may not have the access to the exact value \( K \) and \( C \), and the algorithm may use a large enough estimate for \( K \) and \( 1/C \), or use \( \text{poly log} T \) and \( \frac{1}{\text{poly log} T} \) instead. In the latter case, the theoretical analysis will work through for sufficiently large \( T \) and the regret remains at the same order up to \( \text{poly log} T \) factors. Finally, in Line 7 of the algorithm, we construct the new (shorter) interval based on estimated demands corresponding to the trisection prices.

Concretely, for Algorithm 2, we prove the following upper bounds on the number of selling periods used by the algorithm and its estimation error.

**Theorem 1.** For any input \( z \in \{1, 2\} \), Algorithm 2 uses at most \( O\left(\frac{K^4T^2}{C^2}T^{4/5} \log T \log(pT)\right) \) selling periods and satisfies the fairness constraint during each period. Let \( \hat{p}^z \) be the output of the procedure. With probability \((1 - O(T^{-2}\log(pT)))\), it holds that \(|\hat{p}^z - p^z| \leq 4T^{-\frac{1}{2}}\). Here, only universal constants are hidden in the \( O(\cdot) \) notations.
For the ease of presentation, the proof of Theorem 1 will be provided in later in Section 4.1.

The ExploreConstrainedOPT subroutine. Suppose we have run ExploreUnconstrainedOPT for each \( z \in \{1, 2\} \) and obtained both \( \hat{p}^\#_1 \) and \( \hat{p}^\#_2 \). ExploreConstrainedOPT estimates the constrained (i.e., fairness-aware) clairvoyant solution for both groups. The pseudo-code of the procedure is presented in Algorithm 3. In this procedure, we assume without loss of generality that \( \hat{p}^\#_1 \leq \hat{p}^\#_2 \) since otherwise we can always switch the labels of the two customer groups.

To investigate the property of this algorithm, we establish a key relation between the gap of the constrained optimal prices and that of the unconstrained optimal prices (see Lemma 4 in Section 4.2). In particular, Lemma 4 will show that the optimal offline clairvoyant fairness-aware pricing solution would fully exploit the fairness constraint so that Eq. (5) becomes tight, i.e., \( |p^*_1 - p^*_2| = \lambda |\hat{p}^\#_1 - \hat{p}^\#_2| \). This key relationship is proved by a monotonicity argument for the optimal total revenue as a function of the discrimination level (measured by the ratio between the price gap and that of the unconstrained optimal solution).

Using this key relationship, Algorithm 3 first sets \( \xi \) so that \( 2\xi \) is a lower estimate of the unconstrained optimal price gap (i.e., \( \lambda |\hat{p}^\#_1 - \hat{p}^\#_2| \)) and \( 2\lambda \xi \) is a lower estimate of the constrained optimal price gap (i.e., \( |\hat{p}^\#_1 - \hat{p}^\#_2| \)). The algorithm then tests the mean price \( (p^*_1 + p^*_2)/2 \) using the discretization technique. More specifically, the algorithm identifies a grid of possible mean prices \( \{\ell_1, \ell_2, \ldots, \ell_J\} \).

For each price checkpoint \( \ell_j \), the algorithm would try \( \ell_j - \frac{\lambda \xi}{2} \) and \( \ell_j + \frac{\lambda \xi}{2} \) as the fairness-aware prices for the two customer groups (so that the price gap \( \lambda \xi \) is bounded by \( \lambda |\hat{p}^\#_1 - \hat{p}^\#_2| \)), and estimate the corresponding demands and revenue. The algorithm finally reports the optimal prices among these price checkpoints based on the estimated revenue.

Formally, we state the following guarantee for Algorithm 3, and its proof will be relegated to in Section 4.2.

**Theorem 2.** Suppose that \( |\hat{p}^\#_1 - p^\#_1| \leq 4T^{-1/5} \) and \( |\hat{p}^\#_2 - p^\#_2| \leq 4T^{-1/5} \). Algorithm 3 uses at most \( O(pT^{3/5}\ln T) \) selling periods and satisfies the price fairness constraint during each selling period. With probability \( (1 - O(pT^{-2})) \), the procedure returns a pair of price \( (\hat{p}^*_1, \hat{p}^*_2) \) such that \( |\hat{p}^*_1 - \hat{p}^*_2| \leq \lambda |\hat{p}^\#_1 - \hat{p}^\#_2| \) and

\[
R_1(p^*_1) + R_2(p^*_2) - R_1(\hat{p}^*_1) - R_2(\hat{p}^*_2) \leq O(KpT^{-1/5}).
\]

Here, only universal constants are hidden in the \( O(\cdot) \) notations.

Based on Theorems 1 and 2, we are ready to state the regret bound of our main algorithm.

**Theorem 3.** With probability \( (1 - O(T^{-1})) \), Algorithm 1 satisfies the fairness constraint and its regret is at most \( O(T^{4/5} \log^2 T) \). Here, the \( O(\cdot) \) notation only hides the polynomial dependence on \( p \), \( K \) and \( 1/C \).
Algorithm 3: ExploreConstrainedOPT

Input: the estimated unconstrained optimal prices $\hat{p}_1^♯$ and $\hat{p}_2^♯$, assuming that $\hat{p}_1^♯ \leq \hat{p}_2^♯$ (without loss of generality)

Output: the estimated constrained optimal prices $\hat{p}_1^∗$ and $\hat{p}_2^∗$

1. $\xi \leftarrow \max\{|\hat{p}_1^♯ - \hat{p}_2^♯| - 8T^{-1/5}, 0\}$;
2. $J \leftarrow \lceil (\bar{p} - p)T^{1/2} \rceil$ and create $J$ price checkpoints $\ell_1, \ell_2, \ldots, \ell_J$ where $\ell_j \leftarrow \bar{p} + \frac{j}{J}(\bar{p} - p)$;
3. for each $\ell_j$ do
   4. Repeat the following offerings for $6T^{2/5} \ln T$ selling periods: offer price $p_1(j) \leftarrow \max\{p, \ell_j - \frac{\lambda \xi}{2}\}$ to customer group 1 and price $p_2(j) \leftarrow \min\{p, \ell_j + \frac{\lambda \xi}{2}\}$ to customer group 2;
5. Denote the average demand from customer group $i \in \{1, 2\}$ by $\hat{d}_i(j)$;
6. $\hat{R}(j) \leftarrow \hat{d}_1(j)(p_1(j) - c) + \hat{d}_2(j)(p_2(j) - c)$;
7. $j^* \leftarrow \arg \max_{j \in \{1, 2, \ldots, J\}} \{\hat{R}(j)\}$;
8. return $\hat{p}_1^* \leftarrow p_1(j^*)$ and $\hat{p}_2^* \leftarrow p_2(j^*)$;

Proof of Theorem 3. The proof will be carried out conditioned on the desired events of both Theorem 1 and Theorem 2, which happens with probability at least $(1 - O(T^{-1}))$. We can first easily verify that the first 2 steps of Algorithm 1 satisfy the fairness constraint; and the 3rd step also satisfies the fairness constraint since $|\hat{p}_1^♯ - \hat{p}_2^♯| \leq \lambda|p_1^♯ - p_2^♯|$ by Theorem 2. We then turn to bound the regret of the algorithm. Since the first two steps use at most $O(T^{1/2} \log^2 T)$ time periods, they incur at most $O(T^{1/2} \log^2 T)$ regret. By the desired event of Theorem 2, the regret incurred by the third step is at most $T \times O(T^{-1/2}) = O(T^{3/4})$.

4.1. Proof of Theorem 1 for ExploreUnconstrainedOPT

In this subsection, we establish the theoretical guarantee for ExploreUnconstrainedOPT in Theorem 1.

First, the following lemma upper bounds the number of time periods used by the algorithm.

Lemma 1. Each invocation of Algorithm 2 spends at most $O\left(\frac{\kappa^4f^2}{C^2} T^{3/2} \log T \log(pT)\right)$ selling periods, where only a universal constant is hidden in the $O(\cdot)$ notation.

Proof of Lemma 1. It is easy to verify that the length of the trisection interval $p_R - p_L$ shrinks by a factor of $2/3$ after each iteration, and therefore there are at most $\log_{3/2}((\bar{p} - p)T^{1/5}) = O(\log(pT))$ iterations. Also note that within each iteration, the algorithm uses at most $O\left(\frac{\kappa^4f^2}{C^2} T^{4/5} \log T\right)$ selling periods. The lemma then follows.
We then turn to upper bound the estimation error of the algorithm. For each iteration $r$, we define the following event

$$A_r := \{p^*_r \in [p_L, p_R] \text{ at the end of iteration } r\}.$$ 

Let $r^*$ be the last iteration. We note that 1) $A_0$ always holds, 2) $A_{r^*}$, if holds, would imply the desired estimation error bound ($|\hat{p}^*_r - p^*_r| \leq T^{-1/5}$). Therefore, to prove the desired error bound in Theorem 1, we first prove the following lemma.

**Lemma 2.** For each $r \in \{1, 2, \ldots, r^*\}$, we have that

$$\Pr[A_r | A_{r-1}] \geq 1 - 4T^{-2}.$$ 

**Proof of Lemma 2.** Given the event $A_{r-1}$, we focus on iteration $r$. During this iteration, by Azuma’s inequality, we first have that for each trisection point $i \in \{1, 2\}$, it holds that

$$\Pr \left[ |\hat{d}_{m_i} - d_z(p_{m_i})| \leq \frac{16C}{40K^2}T^{-2/5} \right]$$

$$\geq 1 - 2 \exp \left( - \frac{16C}{40K^2}T^{-2/5} \right)^2 \cdot \frac{5K^4p^2}{C^2}T^{4/5} \ln T = 1 - 2T^{-2}.$$ 

The rest of the proof will be conditioned on that

$$\forall i \in \{1, 2\}, |\hat{d}_{m_i} - d_z(p_{m_i})| \leq \frac{16C}{40K^2}T^{-2/5},$$ 

(9) which happens with probability at least $(1 - 4T^{-2})$ by a union bound.

To establish $A_r$, let $p_L$ and $p_R$ be the values taken at the beginning of iteration $r$, and we discuss the following three cases.

**Case 1:** $p^*_r \in [p_{m_1}, p_{m_2}]$. $A_r$ automatically holds in this case.

**Case 2:** $p^*_r \in [p_L, p_{m_1})$. In this case, by Line 7 of the algorithm, to establish $A_r$, we need to show that $\hat{d}_{m_1}(p_{m_1} - c) > \hat{d}_{m_2}(p_{m_2} - c)$. By Item (a) of Assumption 1, we have that

$$|d_z(p_{m_1}) - d_z(p_{m_2})| \geq \frac{1}{R} |p_{m_1} - p_{m_2}| \geq \frac{4T^{-1/5}}{3K}.$$ 

(10)

Also, by Item (b) of Assumption 1, when $d_z(p^*_r) > d_z(p_{m_1}) > d_z(p_{m_2})$, we have that

$$d_z(p_{m_1})(p_{m_1} - c) - d_z(p_{m_2})(p_{m_2} - c) = R_z(d_z(p_{m_1})) - R_z(d_z(p_{m_2})) \geq \frac{C}{2} (d_z(p_{m_1}) - d_z(p_{m_2}))^2$$

$$\geq \frac{16CT^{-2/5}}{18K^2},$$ 

(11)

where in the last inequality we applied Eq. (10). Together with Eq. (9), we have that

$$\hat{d}_{m_1}(p_{m_1} - c) - \hat{d}_{m_2}(p_{m_2} - c) \geq \frac{16CT^{-2/5}}{18K^2} - 2 \times \frac{16C}{40K^2T^{-2/5}} \times \frac{1}{p} > 0.$$ 

Therefore, $A_r$ holds in this case.

**Case 3:** $p^*_r \in (p_{m_2}, p_R]$. This case can be similarly handled as Case 2 by symmetry.

Combining the 3 cases above, the lemma is proved. □
Finally, since \( r^* \leq O(\log(pT)) \), we have that \( \mathcal{A}_{r^*} \) holds with probability at least \( 1 - O(T^{-2}\log(pT)) \). Together with Lemma 1, we prove Theorem 1.

### 4.2. Proof of Theorem 2 for EXPLOREUNCONSTRAINEDOPT

First, the following lemma upper bounds the number of time periods used by the algorithm.

**Lemma 3.** Algorithm 3 uses at most \( O(\overline{\pi}T^{3/5}\ln T) \) selling periods, where only an universal constant is hidden in the \( O(\cdot) \) notation.

**Proof.** For each price checkpoint \( \ell_j \), the algorithm uses at most \( 6T^{2/5}\ln T \) selling periods. Since there are \( J = \lceil (\overline{\pi} - \underline{\pi})T^{1/5} \rceil \) selling price checkpoints, the total selling periods used by the algorithm is at most \( O(\overline{\pi}T^{3/5}\ln T) \). \( \square \)

We next turn to prove the (near-)optimality of the estimated prices \( \hat{p}_1 \) and \( \hat{p}_2 \). To this end, we first establish the following key relation between the price gap of the constrained optimal solution and that of the unconstrained optimal solution.

**Lemma 4.** \( p_1^* - p_2^* = \lambda(p_1^* - p_2^*) \).

**Proof.** In this proof we assume without loss of generality that \( p_1^* \leq p_2^* \) as the other case can be similarly handled by symmetry.

Since \( R_1(d) \) is a unimodal function and \( d_1(p) \) is a monotonically decreasing function, we have that \( R_1(p) = R_1(d_1(p)) \) is a unimodal function. Similarly, \( R_2(p) = R_2(d_2(p)) \) is also a unimodal function. Under the price fairness constraint, we have that

\[
(p_1^*, p_2^*) = \arg\max_{(p_1, p_2) : |p_1 - p_2| \leq \lambda |p_1^* - p_2^*|} \{ R_1(p_1) + R_2(p_2) \}. \tag{12}
\]

We first claim that \( p_1^* \leq p_2^* \), since otherwise (when \( p_1^* > p_2^* \)), the objective value of the feasible solution \( (p_1, p_2) = (p_2^*, p_2^*) \) is \( R_1(p_2^*) + R_2(p_2^*) > R_1(p_1^*) + R_2(p_2^*) \geq R_1(p_1^*) + R_2(p_2^*) \) (where the first inequality is due to the unimodality of \( R_1(p) \)), contradicting to the optimality of \( (p_1^*, p_2^*) \). We also claim that \( p_1^* \geq p_2^* \), since otherwise (when \( p_1^* < p_2^* \), we also have that \( p_2^* + p_1^* - p_1^* \leq p_1^* + \lambda(p_2^* - p_1^*) \leq p_2^* \) and \( p_2^* + p_2^* - p_1^* > p_2^* \)), the objective value of the feasible solution \( (p_1, p_2) = (p_1^*, p_2^* + p_1^* - p_2^*) \) is \( R_1(p_1^*) + R_2(p_2^* + p_1^* - p_2^*) \geq R_1(p_1^*) + R_2(p_2^* + p_1^* - p_2^*) > R_1(p_1^*) + R_2(p_2^*) \) (where the second inequality is due to the unimodality of \( R_2(p) \)), also contradicting to the optimality of \( (p_1^*, p_2^*) \). To summarize, we have shown that \( p_1^* \in [p_1^*, p_2^*] \).

Since \( R_1(p_1) \) is monotonically decreasing when \( p_1 \in [p_1^*, p_2^*] \), by Eq. (12), we have that

\[
p_1^* = \arg\max_{p_1 \in [p_1^*, p_2^*]} \{ R_1(p_1) \} = \max\{p_1^*, p_2^* + \lambda(p_1^* - p_2^*)\}. \tag{13}
\]

Here, we use \([a \pm b]\) to denote the interval \([a - b, a + b]\) for any \( a \in \mathbb{R} \) and \( b \geq 0 \).
In a similar way, we can also work with \( p^*_2 \) and show that

\[
p^*_2 = \min\{p^*_2, p^*_1 - \lambda(p^*_1 - p^*_2)\}. \tag{14}
\]

Combining Eq. (13) and Eq. (14), we conclude that \( p^*_1 - p^*_2 = \lambda(p^*_1 - p^*_2) \) and the lemma is proved. \( \square \)

The following lemma provide bounds for the \( \xi \) parameter which is used in the algorithm to control the price gaps between the two customer groups.

**Lemma 5.** Suppose that \( |\tilde{p}_1^2 - \tilde{p}_1^1| \leq 4T^{-1/5} \) and \( |\tilde{p}_2^2 - \tilde{p}_2^1| \leq 4T^{-1/5} \), we have that \( \lambda \xi \leq \lambda|p^*_1 - p^*_2| \) and \( \lambda \xi \geq \max\{0, \lambda|p^*_1 - p^*_2| - 16T^{-1/5}\} \).

**Proof.** We first have that

\[
\lambda \xi = \lambda \max\{0, |\tilde{p}_1^2 - \tilde{p}_2^2| - 8T^{-1/5}\} \leq \lambda \max\{0, |p^*_1 - p^*_2| + 8T^{-1/5} - 8T^{-1/5}\} = \lambda |p^*_1 - p^*_2|.
\]

We also have that

\[
\lambda \xi = \lambda \max\{0, |\tilde{p}_1^2 - \tilde{p}_2^2| - 8T^{-1/5}\} \\
\geq \lambda \max\{0, |p^*_1 - p^*_2| - 8T^{-1/5} - 8T^{-1/5}\} \geq \max\{0, \lambda |p^*_1 - p^*_2| - 16T^{-1/5}\}. \quad \square
\]

The following lemma shows that our discretization scheme always guarantees that there is a price check point to approximate the constrained optimal prices.

**Lemma 6.** There exists \( \tilde{j} \in \{1, 2, \ldots, J\} \) such that both \( p_1(\tilde{j}), p_2(\tilde{j}) \in [\bar{p}, \ubar{p}] \) and \( |p_1(\tilde{j}) - p_1^*| \leq 9T^{-1/5}, |p_2(\tilde{j}) - p_2^*| \leq 9T^{-1/5}\).

**Proof.** Consider \( \tilde{j} = \arg \min_j |\ell_j - (p_1^* + p_2^*)/2| \), we have that \( |\ell_j - (p_1^* + p_2^*)/2| \leq T^{-1/5} \). Now, by Lemma 4 and Lemma 5, we have that \( |\ell_j - \frac{\lambda \xi}{2} - p_1^*| \leq 9T^{-1/5} \) and \( |\ell_j + \frac{\lambda \xi}{2} - p_2^*| \leq 9T^{-1/5} \). Therefore, we also have that \( |p_1(\tilde{j}) - p_1^*| \leq 9T^{-1/5} \) and \( |p_2(\tilde{j}) - p_2^*| \leq 9T^{-1/5} \). \( \square \)

We now prove the following lemma for the (near-)optimality of the estimated constrained prices.

**Lemma 7.** With probability \( (1 - 4(\bar{p} - \ubar{p})T^{-2}) \), we have that \( R_1(\tilde{p}_1^1) + R_2(\tilde{p}_2^2) \geq R_1(p_1^*) + R_2(p_2^*) - (4 + 18K)\bar{p}T^{-1/5} \).

**Proof.** By Azuma’s inequality, for each \( j \in \{1, 2, \ldots, J\} \), with probability \( 1 - 4T^{-3} \), it holds that

\[
|\tilde{d}_1(j) - d_1(p_1(j))| \leq T^{-1/5} \quad \text{and} \quad |\tilde{d}_2(j) - d_2(p_2(j))| \leq T^{-1/5}. \tag{15}
\]

Therefore, by a union bound, Eq. (15) holds for each \( j \in \{1, 2, \ldots, J\} \) with probability at least \( 1 - 4JT^{-3} \geq 1 - 4(\bar{p} - \ubar{p})T^{-2} \).Conditioned on this event, we have that

\[
\forall j \in \{1, 2, \ldots, J\} : \quad |\tilde{R}(j) - (R_1(p_1(j)) + R_2(p_2(j))| \leq 2\bar{p}T^{-1/5}. \tag{16}
\]
With Eq. (16), and let $\tilde{j}$ be the index designated by Lemma 6, we have that

$$R_1(\hat{p}_1) + R_2(\hat{p}_2) = R_1(p_1(\hat{j})) + R_2(p_2(\hat{j})) \geq \hat{R}(\hat{j}) - 2\tilde{p}T^{-1/5}$$

$$\geq \hat{R}(\tilde{j}) - 2\tilde{p}T^{-1/5} \geq R_1(p_1(\tilde{j})) + R_2(p_2(\tilde{j})) - 4\tilde{p}T^{-1/5}. \tag{17}$$

By Lemma 6 and Item (a) of Assumption 1, we have that

$$R_1(p_1(\tilde{j})) + R_2(p_2(\tilde{j})) \geq R_1(p_1^*) + R_2(p_2^*) - 2 \times 9T^{-1/5} \times \tilde{p}K. \tag{18}$$

Combining Eq. (17) and Eq. (18), we prove the lemma. □

We are now ready to prove Theorem 2. Note that the sample complexity is upper bounded due to Lemma 3. So long as $|\hat{p}_1^t - p_1^t| \leq T^{-1/5}$ and $|\hat{p}_2^t - p_2^t| \leq T^{-1/5}$, the price fairness is always satisfied due to the first inequality shown in Lemma 5. Finally, the (near-)optimality of the estimated prices $\hat{p}_1^t$ and $\hat{p}_2^t$ is guaranteed by Lemma 7.

5. Lower Bound

When $\lambda \in (\epsilon, 1 - \epsilon)$ where $\epsilon > 0$ is a positive constant, we will show that the expected regret of a fairness-aware algorithm is at least $\Omega(T^{4/5})$. Formally, we prove the following lower bound theorem.

**Theorem 4.** Suppose that $\pi$ is an online pricing algorithm that satisfies the price fairness constraint with probability at least 0.9 for any problem instance. Then for any $\lambda \in (\epsilon, 1 - \epsilon)$ and $T \geq \epsilon^{-C_{LB}}$ (where $C_{LB} > 0$ is a universal constant), there exists a pricing instance such that the expected regret of $\pi$ is at least $\frac{1}{160}\epsilon^2T^{4/5}$.

To prove such a lower bound, we need to construct hard instances for any fairness-aware algorithm. We first set $p = 1$, $\tilde{p} = 2$, and $c = 0$. For any two expected demand rate functions $d, d' : [p, \tilde{p}] \rightarrow [0, 1]$, we define a problem instance $I(d, d')$ as follows: at each time step, when offered a price $p \in [p, \tilde{p}]$, the stochastic demand from group 1 follows the Bernoulli distribution $\text{Ber}(d(p))$, and the stochastic demand from group 2 follows $\text{Ber}(d'(p))$.

**Construction of the hard instances.** We now construct two problem instances $I = I(d_1, d_3)$ and $I' = I(d_2, d_3)$, where $d_i(p) = R_i(p)/p$ for $i \in \{1, 2, 3\}$, and we define $R_i$’s as follows.

$$R_1(p) = \frac{1}{4} - \frac{1}{A}(p - 1 - \frac{\sqrt{h}}{4})^2, \quad p \in [1, 2],$$

$$R_2(p) = \begin{cases} 
\frac{1}{4} - \frac{1}{2A}(p - 1 + \frac{\sqrt{h}}{4})^2, & p \in [1, 1 + \frac{5\sqrt{h}}{4}], \\
\frac{1}{4} - \frac{3}{2A}(p - 1 - \frac{3\sqrt{h}}{4})^2 - \frac{3h}{4A}, & p \in [1 + \frac{5\sqrt{h}}{4}, 1 + \frac{7\sqrt{h}}{4}], \\
\frac{1}{4} - \frac{1}{A}(p - 1 - \frac{\sqrt{h}}{4})^2, & p \in [1 + \frac{7\sqrt{h}}{4}, 2], 
\end{cases}$$

$$R_3(p) = \frac{1}{8} - \frac{1}{A}(p - 2)^2, \quad p \in [1, 2].$$
Here, $A \geq 1$ is a large enough universal constant and $h \geq 0$ depends on $T$, both of which will be chosen later.

We verify the following properties of the constructed demand and profit rate functions.

**Lemma 8.** When $h \in (0,0.01)$ and $20 \leq A \leq 30$, the following statements hold.

(a) $d_i(p) \in [1/20,1/4]$ for all $i \in \{1,2,3\}$ and $p \in [1,2]$.
(b) $d_i(p)$ and $R_i(p)$ are continuously differentiable functions for all $i \in \{1,2,3\}$ and $p \in [1,2]$.
(c) For each $p \in [1,2]$, $i \in \{1,2,3\}$, $\partial d_i/\partial p < -\frac{1}{4h} < 0$, and $R_i$ is strongly concave as a function of $d_i$.
(d) For each $p \in [1,1 + \sqrt[3]{3h}]$, it holds that $|d_i(p) - d_2(p)| \leq \frac{h}{4A}$.
(e) For each $p \in [1,1 + \sqrt[3]{3h}]$, it holds that $D_{KL}(\text{Ber}(d_1(p))\|\text{Ber}(d_2(p))) \leq 5h^2/3A^2$.
(f) For any demand rate function $d(p)$ defined on $p \in [1,2]$, let $p^\ddagger(d) = \arg \max_{p \in [1,2]} \{p \cdot d(p)\}$ be the unconstrained clairvoyant solution; we have that $p^\ddagger(d_1) = 1 + \sqrt[3]{\frac{A}{h}}$, $p^\ddagger(d_2) = 1$, and $p^\ddagger(d_3) = 2$.

In the above lemma, Items (a)-(c) show that the constructed functions are real demand functions satisfying the standard assumptions in literature (also listed in Assumption 1); Items (d)-(e) show that the first two demand functions ($d_1$ and $d_2$) are very similar to each other and therefore it requires relatively more observations from noisy demands to differentiate them; Item (f) simply asserts the unconstrained optimal price for each demand function, which will be used later in our lower bound proof.

The proof of Lemma 8 and all other proofs in the rest of this section can be found in the supplementary material.

For any problem instance $\mathcal{J} \in \{\mathcal{I},\mathcal{I}'\}$, and any demand function $d$ that is employed by a customer group in $\mathcal{J}$, we denote by $p^\ddagger(d;\mathcal{J})$ the price for the customer group in the optimal fairness-aware clairvoyant solution to $\mathcal{J}$. We first compute the optimal fairness-aware solutions to both of our constructed problem instances as follows.

**Lemma 9.** Suppose that $h \leq \epsilon^2/40$, we have the following equalities.

$$p^\ddagger(d_1;\mathcal{I}) = \frac{1}{2}(p^\ddagger(d_1) + p^\ddagger(d_3)) - \frac{\lambda}{2}(p^\ddagger(d_3) - p^\ddagger(d_1)),$$

$$p^\ddagger(d_3;\mathcal{I}) = \frac{1}{2}(p^\ddagger(d_1) + p^\ddagger(d_3)) + \frac{\lambda}{2}(p^\ddagger(d_3) - p^\ddagger(d_1)),$$

$$p^\ddagger(d_2;\mathcal{I}') = \frac{1}{2}(p^\ddagger(d_1) + p^\ddagger(d_3)) - \frac{\lambda}{2}(p^\ddagger(d_3) - p^\ddagger(d_2)),$$

$$p^\ddagger(d_3;\mathcal{I}') = \frac{1}{2}(p^\ddagger(d_1) + p^\ddagger(d_3)) + \frac{\lambda}{2}(p^\ddagger(d_3) - p^\ddagger(d_2)).$$

**The price of a cheap first-group price.** By Lemma 9, we see that when $h \leq \epsilon^2/40$, we have that both $p^\ddagger(d_1;\mathcal{I})$ and $p^\ddagger(d_2;\mathcal{I}')$ are greater than $1 + \frac{\sqrt[3]{3h}}{4}$. For any pricing strategy $(p,p')$, we say
it is cheap for the first group if \( p \leq 1 + \frac{7\sqrt{h}}{4} \). The following lemma lower bounds the regret of a fairness-aware pricing strategy when it is cheap for the first group (and therefore deviates from the optimal solution).

**Lemma 10.** Suppose that \( h \leq \epsilon^4/400 \). For any fairness-aware pricing strategy \((p_1, p_3)\) for the problem instance \( I = I(d_1, d_3) \), if \( p_1 \in [1, 1 + \frac{7\sqrt{h}}{4}] \), we have that

\[
[R_1(p^*(d_1; I)) + R_3(p^*(d_3; I))] - [R_1(p_1) + R_3(p_3)] \geq \frac{\epsilon^2}{4A}.
\]

Similarly, for any fairness-aware pricing strategy \((p_2, p_3)\) for the problem instance \( I' = I(d_2, d_3) \), if \( p_2 \in [1, 1 + \frac{7\sqrt{h}}{4}] \), we have that

\[
[R_2(p^*(d_2; I')) + R_3(p^*(d_3; I'))] - [R_2(p_2) + R_3(p_3)] \geq \frac{\epsilon^2}{4A}.
\]

**The price of identifying the wrong instance.** If a pricing strategy misidentifies the underlying instance \( I' \) by \( I \) and satisfies the fairness condition of \( I \), we show in the following lemma that the significant regret would occur when we apply such a pricing strategy to \( I' \).

**Lemma 11.** Suppose that \( h \leq \epsilon^2/40 \) and \((p_2, p_3)\) is a pricing strategy that satisfies the fairness condition of \( I \), i.e.,

\[
|p_2 - p_3| \leq \lambda |p^*(d_3) - p^*(d_1)|.
\]

Then we have that

\[
[R_2(p^*(d_2; I')) + R_3(p^*(d_3; I'))] - [R_2(p_2) + R_3(p_3)] \geq \frac{\epsilon \lambda \sqrt{h}}{4A}.
\]

**The necessity of cheap first-group prices to separate the two instances apart.** We now show that one has to offer cheap first-group prices to separate \( I \) from \( I' \). This is intuitively true because the only difference between \( I \) and \( I' \) is the demand of the first group when the price is less than \( 1 + \frac{7\sqrt{h}}{4} \). Formally, for any online pricing policy algorithm \( \pi \) and any problem instance \( J \in \{I, I'\} \), let \( P_{J,\pi} \) be the probability measure induced by running \( \pi \) in \( J \) for \( T \) time periods. For each time period \( t \in \{1, 2, \ldots, T\} \), let \( p^{(t)}(d; J, \pi) \) denote the price offered by \( \pi \) to the customer group with demand function \( d \) in the problem instance \( J \). The following lemma upper bounds the KL-divergence between \( P_{I,\pi} \) and \( P_{I',\pi} \) (note that the upper bound relates to the expected number of cheap first-group prices).

**Lemma 12.** For any \( \pi \), it holds that

\[
D_{KL}(P_{I,\pi} \| P_{I',\pi}) \leq \sum_{t=1}^{T} \Pr_{P_{I,\pi}} \left[ p^{(t)}(d_1; I, \pi) \in \left[ 1, 1 + \frac{7\sqrt{h}}{4} \right] \right] \cdot \frac{4h^2}{A^2},
\]

where \( \Pr_{P}[\cdot] \) denotes the probability under the probability measure \( P \).
Now we have all the technical tools prepared. To prove our main lower bound theorem, note that any pricing algorithm has to offer enough amount of cheap first-group prices in order to learn whether the underline instance is $I$ or $I'$ (otherwise, misidentifying the instances would lead to a large regret). On the other hand, the learning process itself also incurs regret. In the following proof, we rigorously lower bound any possible tradeoff between these two types of regret and show the desired $\Omega(\frac{T^4}{4^{\frac{4}{5}}})$ bound.

**Proof of Theorem 4.** We set $h = T^{-2/5}$ and $A = 10$, and discuss the following two cases.

*Case 1:* $\sum_{t=1}^{T} \Pr_{P,I,\pi}[p^{(t)}(d_1; I, \pi) \in [1, 1 + \frac{7\sqrt{h}}{4}]] \geq \frac{A}{400h^2}$. Now, invoking Lemma 10, we have that the expected regret incurred by $\pi$ for instance $I$ is at least

$$\frac{A^2}{400h^2} \cdot \frac{\epsilon^2}{4A} = \frac{\epsilon^2 A}{1600h^2} \cdot \frac{\epsilon^2 T^{4/5}}{160}.$$

*Case 2:* $\sum_{t=1}^{T} \Pr_{P,I,\pi}[p^{(t)}(d_1; I, \pi) \in [1, 1 + \frac{7\sqrt{h}}{4}]] < \frac{A}{400h^2}$. Let $E$ be the event that $\pi$ satisfies the fairness constraints for instance $I$. We have that

$$\Pr_{P,I,\pi}[E] \geq 0.9. \quad (19)$$

Invoking Lemma 12 and Pinsker’s inequality (Lemma EC.1), we have that

$$\left| \Pr_{P,I,\pi}[E] - \Pr_{P,I',\pi}[E] \right| \leq \sqrt{\frac{1}{2} \cdot \frac{A^2}{400h^2} \cdot \frac{4h^2}{A^2}} \leq 0.1. \quad (20)$$

Combining Eq. (19) and Eq. (20), we have that

$$\Pr_{P,I',\pi}[E] \geq 0.8.$$

When $E$ happens for instance $I'$, by Lemma 11, we have that the regret incurred by the pricing strategy at each time step is at least $\frac{\epsilon \sqrt{h}}{5A}$. Therefore, the expected regret of $\pi$ for instance $I'$ is at least

$$\Pr_{P,I',\pi}[E] \cdot T \cdot \frac{\epsilon \sqrt{h}}{4A} \geq 0.8 \cdot T \cdot \frac{\epsilon^2 \sqrt{h}}{4A} \geq \frac{1}{5} \epsilon^2 T^{4/5}.$$

Combining the two cases, we prove the theorem. □

6. Extension to General Fairness Measure

In this section, we extend our fairness-aware dynamic pricing algorithm in Section 4 to general fairness measure $\{M_i(p)\}$ with soft constraints. We will present a policy and prove that its regret can also be controlled by the order of $\widetilde{O}(T^{4/5})$.

Our policy is presented in Algorithm 4. Similar to the algorithm for price fairness, Algorithm 4 also works in the explore-and-exploit manner, where the first two stages are the explore phases. The first exploration stage, the EXPLOREUNCONSTRAINEDOPT subroutine, is exactly Algorithm 2 introduced in Section 4, which serves to estimate the unconstrained optimal prices $p^*_1$ and $p^*_2$. Below we describe the new subroutine EXPLORECONSTRAINEDOPTGENERAL used in the second step.
Suppose we have already run \( \text{ExploreUnconstrainedOPT} \) (Algorithm 2) separately with the input \( z = i \), and obtain the estimation of the optimal price without fairness constraint \( \hat{p}_i^1 \).

Given \( \hat{p}_1^1 \) and \( \hat{p}_2^1 \), run \( \text{ExploreConstrainedOPTGeneral} \) (Algorithm 5), and obtain \( (\hat{p}_1^*, \hat{p}_2^*) \).

For each of the remaining selling periods, offer \( \hat{p}_i^* \) to the customer group \( i \).

---

**Algorithm 4: Fairness-aware Dynamic Pricing for General Fairness Measure**

1. For each group \( i \in \{1, 2\} \), run \( \text{ExploreUnconstrainedOPT} \) (Algorithm 2) separately with the input \( z = i \), and obtain the estimation of the optimal price without fairness constraint \( \hat{p}_i^1 \).
2. Given \( \hat{p}_1^1 \) and \( \hat{p}_2^1 \), run \( \text{ExploreConstrainedOPTGeneral} \) (Algorithm 5), and obtain \( (\hat{p}_1^*, \hat{p}_2^*) \).
3. For each of the remaining selling periods, offer \( \hat{p}_i^* \) to the customer group \( i \).

---

**Algorithm 5: ExploreConstrainedOPTGeneral**

**Input**: the estimated unconstrained optimal prices \( \hat{p}_1^2 \) and \( \hat{p}_2^2 \), assume that \( \hat{p}_1^2 \leq \hat{p}_2^2 \) (without loss of generality)

**Output**: the estimated constrained optimal prices \( \hat{p}_1^2 \) and \( \hat{p}_2^2 \)

1. \( \xi \leftarrow \max\{\lvert \hat{p}_1^2 - \hat{p}_2^2 \rvert, 0 \} \);
2. \( J \leftarrow \lceil (n - p)T^{1/2} \rceil \) and create \( J \) price checkpoints \( \ell_1, \ell_2, \ldots, \ell_J \) where \( \ell_j \leftarrow p + \frac{j}{J}(n - p) \);
3. For each \( \ell_j \) do
4. Repeat the following offering for \( 6T^{2/5} \ln T \) selling periods: offer price \( \ell_j \) to both of the customer groups;
5. For each customer group \( i \in \{1, 2\} \), denote the average demand from the customer group \( \hat{d}_i(\ell_j) \), and the average of the observed fairness measurement value by \( \hat{M}_i(\ell_j) \);
6. Let \( \hat{R}_i(\ell_j) \leftarrow \hat{d}_i(\ell_j) \cdot (\ell_j - c) \), for each \( i \in \{1, 2\} \);
7. For each \( i \in \{1, 2\} \), round up \( \hat{p}_j^2 \) to the nearest price checkpoint, namely \( \ell_{t_{i}} \);
8. For all pairs \( j_1, j_2 \in \{1, 2, \ldots, J\} \), let
   \[ \hat{G}(\ell_{j_1}, \ell_{j_2}) \leftarrow \hat{R}_1(\ell_{j_1}) + \hat{R}_2(\ell_{j_2}) - \gamma \max \left( \lvert \hat{M}_1(\ell_{j_1}) - \hat{M}_2(\ell_{j_2}) \rvert - \lambda \right) \]
9. Let \( (j_1^*, j_2^*) \leftarrow \arg \max_{j_1, j_2 \in \{1, 2, \ldots, J\}} \{ \hat{G}(\ell_{j_1}, \ell_{j_2}) \} \);
10. return \( (\hat{p}_1^*, \hat{p}_2^*) \leftarrow (\ell_{j_1^*}, \ell_{j_2^*}) \);

The \( \text{ExploreConstrainedOPTGeneral} \) subroutine. Suppose we have already run \( \text{ExploreUnconstrainedOPT} \) and obtained both \( \hat{p}_1^1 \) and \( \hat{p}_2^1 \). The \( \text{ExploreConstrainedOPTGeneral} \) estimates the clairvoyant solution with the soft fairness constraint for both groups. The pseudo-code of this procedure is presented in Algorithm 5.

Similar to the \( \text{ExploreConstrainedOPT} \) (Algorithm 3) in Section 4, Algorithm 5 also adopts the discretization technique. The key differences are that: (1) we also need to calculate the estimation of the fairness measure functions \( M_i(\cdot) \) at each price checkpoint \( \ell_j \); and (2) with soft fairness constraints, we are allowed to consider every pair of prices to the two customer groups; however, we
need to deduct the fairness penalty term from the estimated revenue from each pair of discretized prices at Line 8 of the algorithm.

Formally, we state the following guarantee for Algorithm 5 and its proof will be provided in Section 6.1.

**Theorem 5.** Suppose that $|\hat{p}_1^x - p_1^x| < 4T^{-\frac{1}{2}}$ and $|\hat{p}_2^x - p_2^x| < 4T^{-\frac{1}{2}}$. Also assume that $\gamma \leq O(1)$. Algorithm 5 uses at most $O(T^{\frac{1}{2}} \ln T)$ selling periods in total, and with probability at least $(1 - O(T^{-1}))$, the procedure returns a pair of prices $(\hat{p}_1, \hat{p}_2)$ such that

$$[R_1(p_1^* + R_2(p_2^*) - R_1(\hat{p}_1) - R_2(\hat{p}_2))]$$

$$+ \gamma \max(|M_1(\hat{p}_1) - M_2(\hat{p}_2)| - \lambda |M_1(p_1^x) - M_2(p_2^x)|, 0) \leq O(T^{-\frac{1}{2}}). \quad (21)$$

Here the $O(\cdot)$ notation hides the polynomial dependence on $\bar{p}$, $\gamma$, $K$, $K'$, $M$ and $C$.

Note that the Left-Hand-Side of Eq. (21) is the penalized regret incurred by a single selling period when the offered prices are $\hat{p}_1$ and $\hat{p}_2$.

Combining Theorem 1 and Theorem 5, we are ready to state the regret bound of the algorithm.

**Theorem 6.** Assume that $\gamma \leq O(1)$. With probability $(1 - O(T^{-1}))$, the cumulative penalized regret of Algorithm 4 is at most $\text{Reg}^\text{soft} \leq O(T^{1/5} \log^2 T)$. Here the $O(\cdot)$ notation hides the polynomial dependence on $\bar{p}$, $\gamma$, $K$, $K'$, $M$, and $C$.

**Proof.** The proof will be carried out conditioned on the desired events of both Theorem 1 and Theorem 5, which happens with probability at least $(1 - O(T^{-1}))$. Since the first two steps use at most $O(T^{\frac{1}{2}} \log^2 T)$ selling periods, they incur at most $O(T^{\frac{1}{2}} \log^2 T) \times O(\bar{p} + M) = O(T^{\frac{1}{2}} \log^3 T)$ penalized regret. By the desired event of Theorem 5, the penalized regret incurred by the third step is at most $T \times O(T^{-\frac{1}{2}}) = O(T^{\frac{1}{2}})$. □

### 6.1. Proof of Theorem 5 for ExploreConstrainedOPTGeneral

First, the following lemma upper bounds the number of the selling periods used by the algorithm:

**Lemma 13.** Algorithm 5 uses at most $O(\bar{p}T^{\frac{3}{2}} \log T)$ selling periods, where only an universal constant is hidden in $O(\cdot)$ notation.

**Proof.** For each price checkpoint $\ell_j$ the algorithm uses at most $6T^{\frac{2}{3}} \ln T$ selling periods. Since there are $J = \lfloor (\bar{p} - \bar{p})T^{\frac{1}{2}} \rfloor$ selling price checkpoints, the total number of selling periods used by the algorithm is at most $O(\bar{p}T^{\frac{3}{2}} \log T)$. □

We then turn to upper bound the penalized regret incurred by the estimated prices $\hat{p}_1$ and $\hat{p}_2$. Define

$$G(p_1, p_2) := R_1(p_1) + R_2(p_2) - \gamma \max(|M_1(p_1^x) - M_2(p_2)| - \lambda |M_1(p_1^x) - M_2(p_2)|, 0).$$
Note that $G(p_1^*, p_2^*) = R_1(p_1^*) + R_2(p_2^*)$ and therefore the Left-Hand-Side of Eq. (21) equals to $G(p_1^*, p_2^*) - G(\hat{p}_1^*, \hat{p}_2^*)$. To upper bound this quantity, and noting that both $\hat{p}_1^*$ and $\hat{p}_2^*$ are selected from the discretized price checkpoints $\{\ell_j\}_{j \in \{1, 2, \ldots, J\}}$, we first prove the following lemma which shows that it suffices to choose the prices from the discretized price checkpoints. In other words, Lemma 14 upper bounds the regret due to the discretization method.

**Lemma 14.** $\max_{j_1, j_2 \in \{1, 2, \ldots, J\}} \{G(\ell_{j_1}, \ell_{j_2})\} \geq G(p_1^*, p_2^*) - 2(\bar{p}K + \gamma K') \cdot T^{-\frac{1}{2}}$.

**Proof.** For each customer group $i \in \{1, 2\}$, we find the nearest price checkpoint, namely $\ell_{i1}^*$ to the optimal fairness-aware price $p_i^*$. Note that we always have that $|\ell_{i1}^* - p_i^*| \leq T^{-\frac{1}{2}}$.

By item (a) and (c) of Assumption 1, we have that

$$
\left|G(p_1^*, p_2^*) - G(\ell_{i1}^*, \ell_{i2}^*)\right|
\leq |R_1(p_1^*) - R_1(\ell_{i1}^*)| + |R_2(p_2^*) - R_2(\ell_{i2}^*)| + \gamma|M_1(p_1^*) - M_1(\ell_{i1}^*)| + \gamma|M_2(p_2^*) - M_2(\ell_{i2}^*)|
\leq 2(\bar{p}K + \gamma K')T^{-\frac{1}{2}}.
$$

Note that $\max_{j_1, j_2 \in \{1, 2, \ldots, J\}} \{G(\ell_{j_1}, \ell_{j_2})\} \geq G(\ell_{i1}^*, \ell_{i2}^*)$, and we prove the lemma. □

The following lemma uniformly upper bounds the estimation error for $G$ at all pairs of price checkpoints.

**Lemma 15.** Suppose that $|\hat{p}_i^* - p_i^*| \leq 4T^{-\frac{1}{2}}$ holds for each $i \in \{1, 2\}$. With probability at least $(1 - 12(\bar{p} - \underline{p})T^{-3})$, we have that

$$
\left|\tilde{G}(\ell_{j_1}, \ell_{j_2}) - G(\ell_{j_1}, \ell_{j_2})\right|
\leq 2(\bar{p} + \gamma M + 2\gamma M(\bar{M} + 5K')) T^{-\frac{1}{2}}
$$
holds for all $j_1, j_2 \in \{1, 2, \ldots, J\}$.

**Proof.** For each $i \in \{1, 2\}$, since $|\hat{p}_i^* - p_i^*| \leq 4T^{-\frac{1}{2}}$ and $|\ell_{i1}^* - \hat{p}_i^*| \leq T^{-\frac{1}{2}}$ (due to the rounding operation at Line 7), we have that $|p_i^* - \ell_{i1}^*| \leq 5T^{-\frac{1}{2}}$. By item (c) of Assumption 1, we have that

$$
|M_i(\ell_{i1}) - M_i(p_i^*)| \leq 5K'T^{-\frac{1}{2}}. \tag{22}
$$

For each price checkpoint $\ell_j$ and each customer group $i \in \{1, 2\}$, by Azuma’s inequality, with probability at least $(1 - 2T^{-3})$, we have that

$$
\left|\hat{d}_i(\ell_j) - d_i(\ell_j)\right| \leq T^{-\frac{1}{2}}. \tag{23}
$$

Therefore, by a union bound, Eq. (23) holds for all $j \in \{1, 2, \ldots, J\}$ and all $i \in \{1, 2\}$ with probability at least $1 - 4(\bar{p} - \underline{p})T^{-2}$. Conditioned on this event, we have that

$$
\left|\hat{R}_i(\ell_j) - R_i(\ell_j)\right| \leq \bar{p}T^{-\frac{1}{2}}, \quad \forall j \in \{1, 2, \ldots, J\}, i \in \{1, 2\}. \tag{24}
$$
Similarly, for each price checkpoint $\ell_j$ and each customer group $i \in \{1, 2\}$, by Azuma’s inequality, with probability at least $(1 - 2T^{-3})$,
\[
\left| \tilde{M}_i(\ell_i) - M_i(\ell_i) \right| \leq \mathcal{M}T^{-\frac{1}{4}}.
\] (25)

By a union bound, Eq. (25) holds for all $j \in \{1, 2, \ldots, J\}$ and all $i \in \{1, 2\}$ with probability at least $1 - 4(\bar{p} - \underline{p})T^{-2}$. Conditioned on this event, we have that
\[
\left| \tilde{M}_1(\ell_{t_1}) - M_1(\ell_{t_1}) \right| \leq \mathcal{M}T^{-\frac{1}{4}}, \quad \text{and} \quad \left| \tilde{M}_2(\ell_{t_2}) - M_2(\ell_{t_2}) \right| \leq \mathcal{M}T^{-\frac{1}{4}}.
\]
Together with Eq. (22), we have that
\[
\left| \tilde{M}_1(\ell_{t_1}) - \tilde{M}_2(\ell_{t_2}) \right| - \left| M_1(p_1^2) - M_2(p_2^2) \right| \leq (2\mathcal{M} + 10K')T^{-\frac{1}{4}}.
\] (26)

Now, combining Eq. (24) and Eq. (26), and by the definition of $G(\cdot, \cdot)$, for any $j_1, j_2 \in \{1, 2, \ldots, J\}$, we have that
\[
\begin{align*}
&\left| \tilde{G}(\ell_{j_1}, \ell_{j_2}) - G(\ell_{j_1}, \ell_{j_2}) \right| \\
&\leq \left| \tilde{R}_1(\ell_{j_1}) - R_1(\ell_{j_1}) \right| + \left| \tilde{R}_2(\ell_{j_2}) - R_2(\ell_{j_2}) \right| + \gamma \left( \left| \tilde{M}_1(\ell_{j_1}) - M_1(\ell_{j_1}) \right| + \left| \tilde{M}_2(\ell_{j_2}) - M_2(\ell_{j_2}) \right| \right) \\
&\quad + \gamma \lambda \left( \left| \tilde{M}_1(\ell_{t_1}) - \tilde{M}_2(\ell_{t_2}) \right| - \left| M_1(p_1^2) - M_2(p_2^2) \right| \right) \\
&\leq (2\bar{p} + 2\gamma \mathcal{M} + 2\gamma \lambda (2\mathcal{M} + 10K'))T^{-\frac{1}{4}}.
\end{align*}
\]

Finally, collecting the failure probabilities, we prove the lemma. □

Combining Lemma 14 and Lemma 15, we are able to prove Theorem 5.

**Proof of Theorem 5.** Conditioned on that the desired event of Lemma 15 (which happens with probability at least $1 - 12(\bar{p} - \underline{p})T^{-3} \geq 1 - O(T^{-1})$), we have that
\[
G(\tilde{p}_1^*, \tilde{p}_2^*) \geq \tilde{G}(\tilde{p}_1^*, \tilde{p}_2^*) - 2(p + \gamma \mathcal{M} + 2\gamma \lambda (\mathcal{M} + 5K'))T^{-\frac{1}{4}}
\]
\[
= \max_{j_1, j_2 \in \{1, 2, \ldots, J\}} \{ G(\ell_{j_1}, \ell_{j_2}) \} - 2(p + \gamma \mathcal{M} + 2\gamma \lambda (\mathcal{M} + 5K'))T^{-\frac{1}{4}}
\]
\[
\geq \max_{j_1, j_2 \in \{1, 2, \ldots, J\}} \{ G(\ell_{j_1}, \ell_{j_2}) \} - 4(\bar{p} + \gamma \mathcal{M} + 2\gamma \lambda (\mathcal{M} + 5K'))T^{-\frac{1}{4}}
\]
\[
\geq G(p_1^*, p_2^*) - 2(\bar{p}K + \gamma K')T^{-\frac{1}{4}} - 4(\bar{p} + \gamma \mathcal{M} + 2\gamma \lambda (\mathcal{M} + 5K'))T^{-\frac{1}{4}}.
\]

Here, the first two inequalities are due to the desired event of Lemma 15, the equality is by Line 9 of the algorithm, and the last inequality is due to Lemma 14.

Observing that the Left-Hand-Side of Eq. (21) equals to $G(p_1^*, p_2^*) - G(\tilde{p}_1^*, \tilde{p}_2^*)$, we prove the theorem. □
7. Numerical Study

In this section, we provide experimental results to demonstrate Algorithm 1 for price fairness and Algorithm 4 for demand fairness. For simplicity, we refer to Algorithm 1 as FDP-DL (Fairness-aware Dynamic Pricing with Demand Learning) and Algorithm 4 as FDP-GFM (Fairness-aware Dynamic Pricing - Generalized Fairness Measure).

In the experiment, we assume that the mean demand of each group $i$ takes the following form:

$$d_1(p_1) = \frac{1}{2} \exp(1 - p) \quad \text{and} \quad d_2(p_2) = \frac{1}{2} \exp\left(\frac{1-p}{2}\right).$$

The realized demand at each time period $t$, $D_t^i$, follows the Bernoulli distribution with the mean $d_i(p_t^i)$. We further set the price range to be $[p, \overline{p}] = [1, 2]$.

For the ease of illustration, we assume the cost $c = 0$. We vary the key fairness parameter $\lambda$ from 0.2, 0.5 to 0.8 (a larger $\lambda$ indicates more relaxed fairness requirement), and vary the selling periods $T$ from 100,000 to 1,000,000. For each different parameter setting, we would repeat the experiments for 100 times and report the average performance in terms of the cumulative regret.

For the illustration purpose, we consider two methods in the dynamic pricing literature that handles nonparametric demand functions: (1) a tri-section search algorithm adapted from (Lei et al. 2014), and (2) a nonparametric Dynamic Pricing Algorithm (DPA) adapted from (Wang et al. 2014). Both baseline algorithms try to learn the optimal price by shrinking the price interval. The key difference between the tri-section search and DPA is the number of difference prices to be tested at each learning period: the tri-section search will only test two prices while DPA will test $\text{poly}(T)$ prices at each learning period.

As previous dynamic pricing algorithms with nonparametric demand learning do not take fairness into consideration, it is hard to make a direct comparison. For the illustration purpose, we simply assume that the benchmark algorithms provide the same price to both customer groups at each time period. This is perhaps the most intuitive way to guarantee the fairness for benchmark algorithms. Note that under the single-customer-group setting, both baseline algorithms provide the almost optimal regret bound $\tilde{O}(\sqrt{T})$ up to a poly-logarithmic factor. On the other hand, a single-price-at-a-time algorithm would have a theoretical regret lower bound of $\Omega(T)$ . Indeed, in the worst-case scenario, always offering the same price might not well satisfy at least one customer group.

In Figure 1, we show the regrets of our algorithm and the benchmark algorithms. We use log-log figures to better show the relationship between the regret and the total time period. For better illustration, we also perform linear fitting using the sample data we get from the experiments. As one can see, the slope of the linear line for our algorithm is very close to 0.8 = 4/5, while the slopes of the baseline algorithms are close to 1. The slope of our algorithm verifies our theoretical result in Theorem 3.
Another interesting observation is that the benefit of our algorithm, comparing to baseline algorithms, becomes more significant when $\lambda$ becomes larger. This is because when $\lambda$ is smaller, the benefits of distinguishing the best prices of two customer groups also get smaller. Indeed, the baseline algorithm can be viewed as a special case when $\lambda = 0$.

**General Case**

In the experiment for the general fairness measure, we consider the demand fairness (i.e., $M_i(p^{(t)}_i) = D_i(p^{(t)}_i)$) in Algorithm 4 (FDP-GFM). Similar to the experiment setup of the Algorithm 1, we set $\lambda = 0.2, 0.5,$ and $0.8$ and let the maximum selling periods $T$ vary from $100,000$ to $1,000,000$. Furthermore, recall the penalized regret in (8). We set the parameter $\gamma = 1$ to balance the penalty and the original objective. We test each setting for 100 times and report the average performance.

The result is shown in Figure 2. As one can see, the results are also quite similar to the previous one: the slope of the fitted line of the Algorithm 4 (FDP-GFM) is close to $0.8 = 4/5$, which
Figure 2  The performance on regret for Algorithm 4. Here the x-axis the log of the total number periods $T$ and the y-axis the log of the cumulative regret. We consider three values of fairness-ware parameter $\lambda = 0.2$, 0.5 and 0.8.

matches the regret bound of $\tilde{O}(T^{4/5})$. Similarly, the baseline algorithms perform much worse and the corresponding slopes are close to 1.

8. Conclusions

This paper extends the static pricing under fairness constraints from Cohen et al. (2021a) to the dynamic discriminatory pricing setting. We propose fairness-aware pricing policies that achieve $O(T^{4/5})$ regret and establish its optimality. There are several future directions. First, although we only study the setting of two protected customer groups, this setting can be easily generalized to $K \geq 2$ groups using the fairness constraint from (Cohen et al. 2021a): $|M_i(p_i) - M_j(p_j)| \leq \lambda \max_{1 \leq i' < j' \leq K} |M_{i'}(p_i) - M_{j'}(p_j)|$ for all $1 \leq i < j \leq K$ pairs. Second, as this paper focuses on the fairness constraint, we omit operational constraints, such as the inventory constraint. It would be interesting to explore the dynamic discriminatory pricing under the inventory constraint. Finally, with the advance of technology in decision-making, fairness has become a primary ethical con-
cern, especially in the e-commerce domain. We would like to explore more fairness-aware revenue management problems.

Acknowledgement

Xi Chen and Yuan Zhou would like to thank the support from JPMorgan Faculty Research Awards. We also thank helpful discussions from Ivan Brugere, Jiahao Chen, and Sameena Shah.

References

Araman, V. F., & Caldentey, R. (2009). Dynamic pricing for nonperishable products with demand learning. *Operations Research, 57*(5), 1169–1188.

Baek, J., & Farias, V. F. (2021). Fair exploration via axiomatic bargaining. *arXiv preprint arXiv:2106.02553*.

Balseiro, S. R., Lu, H., & Mirrokni, V. (2021). Regularized online allocation problems: Fairness and beyond. *arXiv preprint arXiv:2007.00514v3*.

Ban, G.-Y., & Keskin, N. B. (2021). Personalized dynamic pricing with machine learning: High-dimensional features and heterogeneous elasticity. *Management Science, 67*(9), 5549–5568.

Barocas, S., Hardt, M., & Narayanan, A. (2019). *Fairness and Machine Learning*. fairmlbook.org. [http://www.fairmlbook.org](http://www.fairmlbook.org).

Bateni, M. H., Chen, Y., Ciocan, D. F., & Mirrokni, V. (2016). Fair resource allocation in a volatile marketplace. In *Proceedings of the 2016 ACM Conference on Economics and Computation*.

Bertsimas, D., Farias, V. F., & Trichakis, N. (2011). The price of fairness. *Operations Research, 59*(1), 17–31.

Besbes, O., Gur, Y., & Zeevi, A. (2015). Non-stationary stochastic optimization. *Operations Research, 63*(5), 1227–1244.

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

Bitran, G., & Caldentey, R. (2003). An overview of pricing models for revenue management. *Manufacturing & Service Operations Management, 5*(3), 203–229.

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

Butler, A. W., Mayer, E. J., & Weston, J. (2021). Racial discrimination in the auto loan market. *SSRN preprint SSRN: 3301009*.

Cai, W., Gaebler, J., Garg, N., & Goel, S. (2021). Fair allocation through selective information acquisition. In *AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society*.

Chen, N., & Gallego, G. (2021). Nonparametric pricing analytics with customer covariates. *Operations Research, 69*(3), 974–984.
Chen, Q., Jasin, S., & Duenyas, I. (2015). Real-time dynamic pricing with minimal and flexible price adjustment. *Management Science, 62*(8), 2437–2455.

Chen, X., Owen, Z., Pixton, C., & Simchi-Levi, D. (2021a). A statistical learning approach to personalization in revenue management. *Management Science (to appear)*.

Chen, X., Simchi-Levi, D., & Wang, Y. (2021b). Privacy-preserving dynamic personalized pricing with demand learning. *Management Science (to appear)*.

Chen, Y., & Shi, C. (2019). Network revenue management with online inverse batch gradient descent method. *Available at SSRN 3331939*.

Chen, Y., & Wang, H. (2018). Why are fairness concerns so important? lessons from a shared last-mile transportation system. *Available at SSRN 3168324*.

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

Cohen, M. C., Elmachtoub, A. N., & Lei, X. (2021a). Price discrimination with fairness constraints. *Management Science (forthcoming)*.

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

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

den Boer, A. V., & Zwart, B. (2013). Simultaneously learning and optimizing using controlled variance pricing. *Management Science, 60*(3), 770–783.

Dwork, C., Hardt, M., Pitassi, T., Reingold, O., & Zemel, R. (2012). Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*.

Elmaghraby, W., & Keskinocak, P. (2003). Dynamic pricing in the presence of inventory considerations: Research overview, current practices, and future directions. *Management Science, 49*(10), 1287–1309.

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

Ferreira, K. J., Simchi-Levi, D., & Wang, H. (2018). Online network revenue management using thompson sampling. *Operations Research, 66*(6), 1586–1602.

Gallego, G., & Van Ryzin, G. (1994). Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Management Science, 40*(8), 999–1020.

Gallego, G., & Van Ryzin, G. (1997). A multiproduct dynamic pricing problem and its applications to network yield management. *Operations Research, 45*(1), 24–41.

Gupta, S., & Kamble, V. (2019). Individual fairness in hindsight. In *Proceedings of the 2019 ACM Conference on Economics and Computation*. 
Harrison, J. M., Keskin, N. B., & Zeevi, A. (2012). Bayesian dynamic pricing policies: Learning and earning under a binary prior distribution. *Management Science, 58*(3), 570–586.

Hutchinson, B., & Mitchell, M. (2019). 50 years of test (un)fairness: Lessons for machine learning.

Jabbari, S., Joseph, M., Kearns, M., Morgenstern, J., & Roth, A. (2017). Fairness in reinforcement learning. In *Proceedings of the 34th International Conference on Machine Learning*.

Jasin, S. (2014). Reoptimization and self-adjusting price control for network revenue management. *62*(5), 1168–1178.

Javanmard, A., & Nazerzadeh, H. (2019). Dynamic pricing in high-dimensions. *Journal of Machine Learning Research, 20*(9), 1–49.

Joseph, M., Kearns, M., Morgenstern, J. H., & Roth, A. (2016). Fairness in learning: Classic and contextual bandits. In *Advances in Neural Information Processing Systems*.

Kallus, N., Mao, X., & Zhou, A. (2021). Assessing algorithmic fairness with unobserved protected class using data combination. *Management Science (to appear)*.

Kallus, N., & Zhou, A. (2021). Fairness, welfare, and equity in personalized pricing. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*.

Kandasamy, K., Sela, G.-E., Gonzalez, J. E., Jordan, M. I., & Stoica, I. (2020). Online learning demands in max-min fairness. *arXiv preprint arXiv:2012.08648v1*.

Kasy, M., & Abebe, R. (2021). Fairness, equality, and power in algorithmic decision-making.

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

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

Lei, Y. M., Jasin, S., & Sinha, A. (2014). Near-optimal bisection search for nonparametric dynamic pricing with inventory constraint. *Available at SSRN 2509425*.

Li, K. J., & Jain, S. (2016). Behavior-based pricing: An analysis of the impact of peer induced fairness. *Management Science, 62*(9), 2705–2721.

Liu, Y., Radanovic, G., Dimitrakakis, C., Mandal, D., & Parkes, D. C. (2017). Calibrated fairness in bandits. *arXiv preprint arXiv:1707.01875*.

Lobel, I., Leme, R. P., & Vladu, A. (2018). Multidimensional binary search for contextual decision-making. *Operations Research, 66*(5), 1346–1361.

Ma, W., Xu, P., & Xu, Y. (2021). Group-level fairness maximization in online bipartite matching. *arXiv preprint, arXiv:2011.13908v3*.

Miao, S., Chen, X., Chao, X., Liu, J., & Zhang, Y. (2019). Context-based dynamic pricing with online clustering. *arXiv preprint arXiv:1902.06199*.
Nambiar, M., Simchi-Levi, D., & Wang, H. (2019). Dynamic learning and price optimization with endogeneity effect. *Management Science, 65*(11), 4951–5448.

Pandey, A., & Caliskan, A. (2021). Disparate impact of artificial intelligence bias in ridehailing economy’s price discrimination algorithms. In *Proceedings of the AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society*.

Streitfeld, D. (2000). On the web, price tags blur. https://www.washingtonpost.com/archive/politics/2000/09/27/on-the-web-price-tags-blur/14daea51-3a64-488f-8e6b-c1a3654773da/.

Wallheimer, B. (2018). Are you ready for personalized pricing. https://review.chicagobooth.edu/marketing/2018/article/are-you-ready-personalized-pricing.

Wang, Y., Chen, X., Chang, X., & Ge, D. (2021). Uncertainty quantification for demand prediction in contextual dynamic pricing. *Production and Operations Management, 30*(6), 1703–1717.

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

EC.1. Proofs of Technical Lemmas for the Lower Bound in Section 5

EC.1.1. Pinsker’s Inequality

**Lemma EC.1.** If $P$ and $Q$ are two probability distributions on a measurable space $(X, \Sigma)$, then for any event $A \in \Sigma$, it holds that

$$|P(A) - Q(A)| \leq \sqrt{\frac{1}{2} \text{KL}(P||Q)},$$

where

$$\text{KL}(P||Q) = \int_X \left( \ln \frac{dP}{dQ} \right) dP$$

is the Kullback–Leibler divergence.

EC.1.2. Proof of Lemma 8

For readers’ convenience, we restate the claims of Lemma 8 and present the proofs immediately after each claim.

(a) $d_i(p) \in [1/20, 1/4]$ for all $i \in \{1, 2, 3\}$ and $p \in [1, 2]$.

**Proof.** When $A \geq 10$ and $h \in (0, 1)$, one can easily calculate that that $d_i(p) \in [1/10, 1/4]$ when $p \in [1, 2]$. $\square$

(b) $d_i(p)$ and $R_i(p)$ are continuously differentiable functions for all $i \in \{1, 2, 3\}$ and $p \in [1, 2]$.

**Proof.** Note that $R_i(p) = p \cdot d_i(p)$, thus we only need to prove $R_i(p)$ is a continuously differentiable function for each $i \in 1, 2, 3$ and $p \in [1, 2]$. Note that this is obviously true for $R_1(p)$ and $R_3(p)$. Thus we only need to prove $R_2(p)$ is a continuously differentiable function on $[1, 2]$.

To prove $R_2(p)$ is continuously differentiable at $p = 1 + \frac{5\sqrt{h}}{4}$, we only need to prove $R_2((1 + \frac{5\sqrt{h}}{4})_+)$ and $\partial_- R_2(1 + \frac{5\sqrt{h}}{4}) = \partial_+ R_2(1 + \frac{5\sqrt{h}}{4})$.

Recall that

$$R_2(p) = \begin{cases} 
\frac{1}{4} - \frac{1}{2A} (p - 1 + \frac{\sqrt{h}}{4})^2, & p \in [1, 1 + \frac{5\sqrt{h}}{4}) \\
\frac{1}{4} - \frac{3}{2A} (p - 1 - \frac{3\sqrt{h}}{4})^2 - \frac{3h}{4A}, & p \in [1 + \frac{5\sqrt{h}}{4}, 1 + \frac{7\sqrt{h}}{4}) \\
\frac{1}{4} - \frac{1}{A} (p - 1 - \frac{\sqrt{h}}{4})^2, & p \in [1 + \frac{7\sqrt{h}}{4}, 2] 
\end{cases}$$

Let $G_1(p) = \frac{1}{4} - \frac{1}{2A} (p - 1 + \frac{\sqrt{h}}{4})^2$, $G_2(p) = \frac{1}{4} - \frac{3}{2A} (p - 1 - \frac{3\sqrt{h}}{4})^2 - \frac{3h}{4A}$, $G_3(p) = \frac{1}{4} - \frac{1}{A} (p - 1 - \frac{\sqrt{h}}{4})^2$, this means we only need to prove $G_1(p) = G_2(p)$ and $G_1'(p) = G_2'(p)$ when $p = 1 + \frac{5\sqrt{h}}{4}$, and prove that $G_2(p) = G_3(p)$ and $G_2'(p) = G_3'(p)$ when $p = 1 + \frac{7\sqrt{h}}{4}$. 

When \( p = 1 + \frac{5\sqrt{h}}{4}, \)

\[
G_1(p) - G_2(p) = \left[ \frac{1}{4} - \frac{1}{2A}(p - 1 + \frac{\sqrt{h}}{4})^2 \right] - \left[ \frac{1}{4} - \frac{3}{2A}(p - 1 - \frac{3\sqrt{h}}{4})^2 - \frac{3h}{4A} \right] = 0,
\]

\[
G'_1(p) - G'_2(p) = \frac{4 - 4p - \sqrt{h}}{4A} - \frac{9\sqrt{h} + 12 - 12p}{4A} = 0.
\]

Similarly, when \( p = 1 + \frac{7\sqrt{h}}{4}, \)

\[
G_2(p) - G_3(p) = \left[ \frac{1}{4} - \frac{3}{2A}(p - 1 - \frac{3\sqrt{h}}{4})^2 - \frac{3h}{4A} \right] - \left[ \frac{1}{4} - \frac{1}{A}(p - 1 - \frac{\sqrt{h}}{4})^2 \right] = 0,
\]

\[
G'_2(p) - G'_3(p) = \frac{9\sqrt{h} + 12 - 12p}{4A} - \frac{8 + 2\sqrt{h} - 8p}{4A} = 0.
\]

This means \( R_2(p) \) is a twice differentiable function in \( p \in [1, 2] \). □

(c) For each \( p \in [1, 2], \, i \in \{ 1, 2, 3 \}, \, \partial d_i / \partial p < -\frac{1}{40} < 0, \) and \( R_i \) is strongly concave as a function of \( d_i \).

**Proof.** Since \( 0 < h \leq 0.01 \) and \( 20 \leq A \leq 30 \), we have that

\[
\frac{\partial d_1}{\partial p} = \frac{-4A + h + 8\sqrt{h} - 16p^2 + 16}{16Ap^2} \leq -\frac{1}{16} + \frac{0.81}{320} \leq -\frac{1}{30} < 0;\]

\[
\frac{\partial d_2}{\partial p} = \begin{cases} 
-\frac{8A + h - 8\sqrt{h} - 16p^2 + 16}{32Ap^2} \leq -\frac{1}{16} + \frac{0.01}{640} \leq -\frac{1}{30}, & p \in [1, 1 + \frac{5\sqrt{h}}{4}] \\
-\frac{8A + 51h + 72\sqrt{h} - 48p^2 + 48}{32Ap^2} \leq -\frac{1}{16} + \frac{7.71}{640} \leq -\frac{1}{30}, & p \in [1 + \frac{5\sqrt{h}}{4}, 1 + \frac{7\sqrt{h}}{4}) \\
-\frac{4A + h + 8\sqrt{h} - 16p^2 + 16}{16Ap^2} \leq -\frac{1}{16} + \frac{0.81}{320} \leq -\frac{1}{30}, & p \in [1 + \frac{7\sqrt{h}}{4}, 2) 
\end{cases}
\]

\[
\frac{\partial d_3}{\partial p} = \frac{-A - 8p^2 + 32}{8Ap^2} < -\frac{\frac{1}{8} + \frac{24}{8A}}{p^2} < -\frac{1}{40}.
\]

Similarly we have

\[
\frac{\partial d_1}{\partial p} = \frac{-4A + h + 8\sqrt{h} - 16p^2 + 16}{16Ap^2} \geq -\frac{1}{4} - \frac{1}{10} = -\frac{7}{20};\]

\[
\frac{\partial d_2}{\partial p} = \begin{cases} 
-\frac{8A + h - 8\sqrt{h} - 16p^2 + 16}{32Ap^2} \geq -\frac{1}{4} - \frac{1}{20} = -\frac{3}{10}, & p \in [1, 1 + \frac{5\sqrt{h}}{4}] \\
-\frac{8A + 51h + 72\sqrt{h} - 48p^2 + 48}{32Ap^2} \geq -\frac{1}{4} - \frac{3}{20} = -\frac{2}{5}, & p \in [1 + \frac{5\sqrt{h}}{4}, 1 + \frac{7\sqrt{h}}{4}) \\
-\frac{4A + h + 8\sqrt{h} - 16p^2 + 16}{16Ap^2} \geq -\frac{1}{4} - \frac{1}{10} = -\frac{7}{20}, & p \in [1 + \frac{7\sqrt{h}}{4}, 2) 
\end{cases}
\]

\[
\frac{\partial d_3}{\partial p} = \frac{-A - 8p^2 + 32}{4Ap^2} \geq -\frac{1}{8} - \frac{1}{10} = -\frac{9}{40}.
\]

Thus, for each \( p \in [1, 2], \, i \in \{ 1, 2, 3 \}, \)

\[
-\frac{2}{5} \leq \frac{\partial d}{\partial p} < -\frac{1}{40} < 0.
\]
To prove that \( R_i \) is strongly concave as a function of \( d_i \) (for each \( i \in \{1, 2, 3\} \)), we only need to show that the second-order semi-derivatives are upper bounded by a negative constant (which will be \(-\frac{1}{12}\) in the following proof).

Note that whenever \( R_i \) is twice-differentiable with respect to \( d_i \), we have that

\[
\frac{\partial^2 R_i}{(\partial d_i)^2} = \frac{\partial}{\partial p} \left( \frac{\partial R_i}{\partial d_i} \right) \left[ \frac{\partial p}{\partial d_i} \right].
\]

For \( i = 1 \), since \( R_1 \) is continuously twice-differentiable with respect to \( d_1 \), we only need to check that (while noticing that \( 20 \leq A \leq 30 \) and \( 0 < h < 0.01 \))

\[
\frac{\partial}{\partial p} \frac{\partial R_1}{\partial d_1} = \frac{\partial}{\partial p} \left[ \left( \frac{\partial R_1}{\partial p} \right) \left( \frac{\partial p}{\partial d_1} \right) \right] = \frac{\partial}{\partial p} \left[ \left( \frac{\sqrt{h} - 4p + 4}{2A} \right) \left( -\frac{16A^2}{4A + h + 8\sqrt{h} - 16p^2 + 16} \right) \right] = \frac{16p(\sqrt{h} - 4p + 4)^2(\sqrt{h} + 2p + 4) - 4A(\sqrt{h} - 6p + 4))}{(4A - h - 8\sqrt{h} + 16p^2 - 16)^2} > 0
\]

\[
> \frac{16 \cdot (0 - 4 \cdot 20 \cdot (0.1 - 6 + 4))}{(4 \cdot 30 + 16 \cdot 4 - 16)^2} > \frac{1}{20}.
\]

And therefore,

\[
\frac{\partial^2 R_1}{(\partial d_1)^2} = \frac{\partial}{\partial p} \left( \frac{\partial R_1}{\partial d_1} \right) \left[ \frac{\partial p}{\partial d_1} \right] \leq \frac{1}{20} \left( -\frac{5}{2} \right) \leq -\frac{1}{8}.
\]

For \( i = 3 \), we have that

\[
\frac{\partial}{\partial p} \frac{\partial R_3}{\partial d_3} = \frac{\partial}{\partial p} \left[ \left( \frac{\partial R_3}{\partial p} \right) \left( \frac{\partial p}{\partial d_3} \right) \right] = \frac{\partial}{\partial p} \left[ \left( \frac{4 - 2p}{A} \right) \left( -\frac{8A^2}{A - 8p^2 + 32} \right) \right] = \frac{16p(A(3p - 4) + 8(p + 4)(p - 2)^2)}{(A + 8p^2 - 32)^2} \geq \frac{8}{45},
\]

where the last inequality can be verified for \( A \in [20, 30] \). Therefore,

\[
\frac{\partial^2 R_3}{\partial (d_3)^2} = \frac{\partial}{\partial p} \left( \frac{\partial R_3}{\partial d_3} \right) \left[ \frac{\partial p}{\partial d_3} \right] \leq \frac{8}{45} \left( -\frac{5}{2} \right) \leq -\frac{4}{9}.
\]

Finally, for \( i = 2 \), we calculate the second-order derivatives for every interval where \( R_i \) admits continuously second-order derivative with respect to \( d_i \).
When $p \in [1, 1 + \frac{5\sqrt{h}}{4})$, we have that

$$\frac{\partial}{\partial p} \frac{\partial R_2}{\partial d_2} = \frac{\partial}{\partial p} \left[ \left( \frac{\partial R_2}{\partial p} \right) \left( \frac{\partial R_2}{\partial d_2} \right) \right]$$

$$= \frac{\partial}{\partial p} \left[ \left( 4 - 4p - \sqrt{h} \right) \left( \frac{32Ap^2}{4A} \right) \right]$$

$$= \frac{16p((\sqrt{h} + 4p - 4)(2p + 2) - \sqrt{h}) + 8A(\sqrt{h} + 6p - 4))}{(8A - h + 8\sqrt{h} + 16p^2 - 16)^2} > 0$$

$$> \frac{16 \cdot (0 + 8 \cdot 20 \cdot 2)}{(8 \cdot 30 + 16 \cdot 4 - 16)^2}$$

$$> \frac{1}{30}.$$ 

When $p \in (1 + \frac{5\sqrt{h}}{4}, 1 + \frac{7\sqrt{h}}{4})$, we have that

$$\frac{\partial}{\partial p} \frac{\partial R_2}{\partial d_2} = \frac{\partial}{\partial p} \left[ \left( \frac{\partial R_2}{\partial p} \right) \left( \frac{\partial R_2}{\partial d_2} \right) \right]$$

$$= \frac{\partial}{\partial p} \left[ \left( \frac{9\sqrt{h} - 12p + 12}{4A} \right) \left( \frac{32Ap^2}{-8A + 51h + 72\sqrt{h} - 48p^2 + 48} \right) \right]$$

$$= \frac{48p((6p - 3\sqrt{h} - 4)(8A - 51h - 82\sqrt{h} - 48) + 96p^3)}{(8A - 51h - 72\sqrt{h} + 48p^2 - 48)^2} > 0$$

$$> \frac{48 \cdot (2 \cdot (8 \cdot 20 - 0.51 - 0.82 - 48) + 96)}{(8 \cdot 30 + 48 \cdot 4 - 48)^2}$$

$$> \frac{1}{30}.$$ 

When $p \in (1 + \frac{7\sqrt{h}}{4}, 2]$, $R_2(p) = R_1(p)$, which means that $\frac{\partial}{\partial p} \frac{\partial R_2}{\partial d_2} > \frac{1}{20} > \frac{1}{30}$. Therefore, for each $p \in [1, 1 + \frac{5\sqrt{h}}{4}) \cup (1 + \frac{5\sqrt{h}}{4}, 1 + \frac{7\sqrt{h}}{4}) \cup (1 + \frac{7\sqrt{h}}{4}, 2]$, we have that

$$\frac{\partial^2 R_2}{\partial d_2^2} = \frac{\partial}{\partial p} \left[ \frac{\partial R_2}{\partial d_2} \right] \left( \frac{\partial p}{\partial d_2} \right) \leq \frac{1}{30} \left( -\frac{5}{2} \right) \leq -\frac{1}{12}.$$ 

This also means that the second-order semi-derivatives of $R_2$ with respect to $d_2$ at $p \in \{1 + \frac{5\sqrt{h}}{4}, 1 + \frac{7\sqrt{h}}{4}\}$ are also upper bounded by $-\frac{1}{12}$. □

(d) For each $p \in [1, 1 + \frac{7\sqrt{h}}{4}]$, it holds that $|d_1(p) - d_2(p)| \leq \frac{h}{4A}$.

**Proof.** Since

$$|d_2(p) - d_1(p)| = \frac{|R_2(p) - R_1(p)|}{p} \leq |R_2(p) - R_1(p)|,$$

we only need to show that $|R_2(p) - R_1(p)| \leq \frac{h}{4A}$, which can be verified by discussing the two cases that $p \in [1, 1 + \frac{5\sqrt{h}}{4})$ and $p \in [1 + \frac{5\sqrt{h}}{4}, 1 + \frac{7\sqrt{h}}{4}]$. □

(e) For each $p \in [1, 1 + \frac{7\sqrt{h}}{4}]$, it holds that $D_{KL}(Ber(d_1(p))||Ber(d_2(p))) \leq 5h^2 / 3A^2$. 

Proof. From item (a) we know that \(d_1(p) \in [\frac{1}{20}, \frac{1}{3}]\) and \(d_2(p) \in [\frac{1}{20}, \frac{1}{3}]\). Thus,

\[
D_{\text{KL}}(\text{Ber}(d_1(p))|\text{Ber}(d_2(p)))
= d_1(p) \log \frac{d_1(p)}{d_2(p)} + (1 - d_1(p)) \log \frac{1 - d_1(p)}{1 - d_2(p)}
= d_1(p) \log \left(1 + \frac{d_1(p) - d_2(p)}{d_2(p)}\right) + (1 - d_1(p)) \log \left(1 + \frac{d_2(p) - d_1(p)}{1 - d_2(p)}\right)
\leq (d_1(p) - d_2(p)) \left(\frac{d_1(p) - d_2(p)}{d_2(p)}\right) + (1 - d_1(p)) \log \left(1 + \frac{d_2(p) - d_1(p)}{1 - d_2(p)}\right)
= (d_1(p) - d_2(p))^2.
\]

From item (d), \(|d_1(p) - d_2(p)| \leq \frac{h}{4A}\), thus

\[
D_{\text{KL}}(\text{Ber}(d_1(p))|\text{Ber}(d_2(p))) \leq \frac{h^2}{16A^2} \leq \frac{5h^2}{3A^2}.
\]

\(\square\)

(f) For any demand rate function \(d(p)\) defined on \(p \in [1, 2]\), let \(p^\star(d) = \arg \max_{p \in [1, 2]} \{p \cdot d(p)\}\) be the unconstrained clairvoyant solution; we have that \(p^\star(d_1) = 1 + \frac{\sqrt{h}}{4}, p^\star(d_2) = 1\), and \(p^\star(d_3) = 2\).

Proof. Since \(R_1(p)\) and \(R_3(p)\) are quadratic functions, one can easily show that \(p^\star(d_1) = 1 + \frac{\sqrt{h}}{4}\) and \(p^\star(d_3) = 2\). It is also straightforward to verify that \(R_2(p)\) is monotonically decreasing when \(p \in [1, 2]\). Thus \(p^\star(d_2) = 1\). \(\square\)

**EC.1.3. Proof of Lemma 9**

**Proof of Lemma 9.** We first compute \(p^\star(d_1; \mathcal{I})\) and \(p^\star(d_3; \mathcal{I})\). By the unimodality of \(R_1(\cdot)\) and \(R_3(\cdot)\), one can easily verify that \(p^\star(d_1) \leq p^\star(d_1; \mathcal{I}) \leq p^\star(d_3; \mathcal{I}) \leq p^\star(d_3)\) and

\[
|p^\star(d_1; \mathcal{I}) - p^\star(d_3; \mathcal{I})| = \lambda |p^\star(d_1) - p^\star(d_3)|.
\]

Let \(\Delta := |p^\star(d_3) - p^\star(d_1)|\). We have that

\[
p^\star(d_1; \mathcal{I}) = \arg \max_{p \in [p^\star(d_1), p^\star(d_3) - \lambda \Delta]} \{R_1(p) + R_3(p + \lambda \Delta)\}
= \arg \max_{p \in [p^\star(d_1), p^\star(d_3) - \lambda \Delta]} \left\{\frac{3}{8} - \frac{1}{A} \left( (p - 1 - \frac{\sqrt{h}}{4})^2 + (p + \lambda \Delta - 2)^2 \right) \right\}
= \arg \max_{p \in [p^\star(d_1), p^\star(d_3) - \lambda \Delta]} \left\{\frac{3}{8} - \frac{1}{A} \left( (p - p^\star(d_1))^2 + (p + \lambda \Delta - p^\star(d_3))^2 \right) \right\}
= \frac{1 + \lambda}{2} p^\star(d_1) + \frac{1 - \lambda}{2} p^\star(d_3),
\]

where the third inequality uses Item (f) of Lemma 8. We then immediately get

\[
p^\star(d_3; \mathcal{I}) = p^\star(d_1; \mathcal{I}) + \lambda \Delta = \frac{1 - \lambda}{2} p^\star(d_1) + \frac{1 + \lambda}{2} p^\star(d_3).
\]
We then proceed to compute $p^*(d_2;\mathcal{I}')$ and $p^*(d_3;\mathcal{I}')$. Similarly, we have that $p^*(d_2) \leq p^*(d_2;\mathcal{I}') \leq p^*(d_3;\mathcal{I}') \leq p^*(d_3)$ and

$$|p^*(d_2;\mathcal{I}') - p^*(d_3;\mathcal{I}')| = \lambda|p^*(d_2) - p^*(d_3)|.$$  

Let $\Delta' := |p^*(d_3) - p^*(d_2)| = 1$, and we have that 

$$p^*(d_2;\mathcal{I}') = \arg \max_{p \in [p^*(d_2);p^*(d_3) - \lambda \Delta'] \cup \mathbb{R}} \{R_2(p) + R_3(p + \lambda \Delta')\}.$$  

Since $R_2(p) \leq \frac{1}{4}$ and $R_3(p)$ is monotonically increasing when $p \in [1,2]$, we have that

$$\max_{p \in [1,1 + \frac{7\sqrt{h}}{2}]} \{R_2(p) + R_3(p + \lambda \Delta')\} \leq \frac{1}{4} + R_3\left(1 + \frac{\sqrt{\frac{7h}{4} + \lambda \Delta'}}{2}\right) = \frac{3}{8} - \frac{1}{A}\left(1 + \frac{\sqrt{h}}{2} + \lambda - 2\right)^2.$$  

(EC.1)

When $p = 1 + \frac{7\sqrt{h}}{2}$, note that 

$$[R_2(p) + R_3(p + \lambda \Delta')]\big|_{p = 1 + \frac{7\sqrt{h}}{2}} = R_2\left(1 + \frac{7\sqrt{\frac{7h}{4} + \lambda \Delta'}}{2}\right) + R_3\left(1 + \frac{\sqrt{\frac{7h}{4} + \lambda \Delta'}}{2}\right) = \frac{3}{8} - \frac{1}{A}\left(1 + \frac{\sqrt{h}}{2} + \lambda - 2\right)^2.$$  

(EC.2)

Combining Eq. (EC.1) and Eq. (EC.2), we have that 

$$[R_2(p) + R_3(p + \lambda \Delta')]\big|_{p = 1 + \frac{7\sqrt{h}}{2}} - \max_{p \in [1,1 + \frac{7\sqrt{h}}{2}]} \{R_2(p) + R_3(p + \lambda \Delta')\} \geq \frac{1}{A}\left(1 + \frac{\sqrt{\frac{7h}{4} + \lambda - 2}}{4} - \frac{13\sqrt{\frac{7h}{4}}}{4} - \frac{1}{A}\left(1 + \frac{7\sqrt{\frac{7h}{4} + \lambda - 2}}{2}\right)^2\right) = \frac{\sqrt{h}}{4A}\left(14 - 14\lambda - 79\sqrt{\frac{7h}{4}}\right) \geq 0,$$

where the last inequality is because $\lambda \leq 1 - \epsilon$ and $h \leq \frac{e^2}{90}$. Note that $h \leq \frac{e^2}{90}$ also implies that $1 + \frac{7\sqrt{h}}{2} \leq p^*(d_3) - \lambda \Delta'$ is a valid candidate for $p^*(d_2;\mathcal{I}')$. Therefore, we have that $p^*(d_2;\mathcal{I}') \geq 1 + \frac{7\sqrt{h}}{4}$, and 

$$p^*(d_2;\mathcal{I}') = \arg \max_{p \in [1 + \frac{7\sqrt{h}}{4},p^*(d_3) - \lambda \Delta']} \{R_2(p) + R_3(p + \lambda \Delta')\} = \arg \max_{p \in [1 + \frac{7\sqrt{h}}{4},p^*(d_3) - \lambda \Delta']} \{R_1(p) + R_3(p + \lambda \Delta')\} = \arg \max_{p \in [1 + \frac{7\sqrt{h}}{4},p^*(d_3) - \lambda \Delta']} \left\{\frac{1}{2} - \frac{1}{A}\left((p - p^*(d_1))^2 + (p + \lambda \Delta' - p^*(d_3))^2\right)\right\} = \frac{1}{2}(p^*(d_1) + p^*(d_3)) - \frac{\lambda}{2}(p^*(d_3) - p^*(d_2)).$$

Finally, we have that 

$$p^*(d_3;\mathcal{I}') = p^*(d_2;\mathcal{I}') + \lambda \Delta' = \frac{1}{2}(p^*(d_1) + p^*(d_3)) + \frac{\lambda}{2}(p^*(d_3) - p^*(d_2)).$$
EC.1.4. Proof of Lemma 10

Proof of Lemma 10. Similar to the proof of Lemma 9, we define \( \Delta := |p^3(d_3) - p^3(d_1)| \leq 1 \). By Lemma 9, we have that

\[
R_1(p^*(d_1, I)) = \frac{1}{4} - \frac{1}{A} \left( \frac{1 - \lambda \Delta}{2} \right)^2, \quad R_3(p^*(d_3, I)) = \frac{1}{8} - \frac{1}{A} \left( \frac{1 - \lambda \Delta}{2} \right)^2. \tag{EC.1}
\]

When \( p_1 \in [1, 1 + \frac{7\sqrt{h}}{4}] \), we have that \( p_3 \in [1, 1 + \frac{7\sqrt{h}}{4} + \lambda \Delta] \). Using the monotonicity of \( R_3(\cdot) \) and the fact that \( R_1(p_1) \leq 1/4 \), we have that

\[
R_1(p^*(d_1, I)) + R_2(p^*(d_3, I)) - R_1(p_1) - R_3(p_3) \geq \frac{3}{8} - \frac{2}{A} \left( \frac{1 - \lambda \Delta}{2} \right)^2 - \frac{1}{4} - R_3 \left( 1 + \frac{7\sqrt{h}}{4} + \lambda \Delta \right) = \frac{1}{A} \left( 1 - \frac{7\sqrt{h}}{4} - \lambda \Delta \right)^2 - \frac{\Delta^2}{2A} (1 - \lambda)^2. \tag{EC.3}
\]

Note that

\[
\left( 1 - \frac{7\sqrt{h}}{4} - \lambda \Delta \right)^2 - \frac{\Delta^2}{2} (1 - \lambda)^2 = (1 - \lambda \Delta)^2 - \frac{1}{2} (\Delta - \lambda \Delta)^2 - \frac{7\sqrt{h}}{2} (1 - \lambda \Delta)^2 + \frac{49h}{16}
\]

\[
\geq \frac{1}{2} (1 - \lambda \Delta)^2 - \frac{7\sqrt{h}}{2} (1 - \lambda \Delta)^2 + \frac{49h}{16}
\]

\[
\geq \frac{1}{2} \epsilon^2 - \frac{7\sqrt{h}}{2} + \frac{49h}{16} \geq \frac{\epsilon^2}{4}, \tag{EC.4}
\]

where in the first two inequalities, we used \( \lambda \in (0, 1 - \epsilon) \) and \( \Delta \in (0, 1) \), in the last inequality, we used that \( h \leq \epsilon^4/400 \). Combining Eq. (EC.3) and Eq. (EC.4), we prove that

\[
R_1(p^*(d_1, I)) + R_2(p^*(d_3, I)) - R_1(p_1) - R_3(p_3) \geq \frac{\epsilon^2}{4A}.
\]

The second part of the lemma can be proved in the same way. \( \square \)

EC.1.5. Proof of Lemma 11

Proof of Lemma 11. Similar to the proof of Lemma 9, we define \( \Delta := |p^3(d_3) - p^3(d_1)| \) and \( \Delta' := |p^3(d_3) - p^3(d_2)| \).

Our assumption that \((p_2, p_3)\) satisfies the fairness condition of \(I\) is equivalent to that \(|p_3 - p_2| \leq \lambda \Delta\). By the unimodality of \( R_2(\cdot) \) and \( R_3(\cdot) \), we have that

\[
\max_{p_2, p_3 \mid |p_2 - p_3| \leq \lambda \Delta} \{ R_2(p_2) + R_3(p_3) \} = \max_{p \in [p^3(d_2), p^3(d_3) - \lambda \Delta]} \{ R_2(p) + R_3(p + \lambda \Delta) \}. \tag{EC.5}
\]

Similar to Eq. (EC.1) and Eq. (EC.2) in the proof of Lemma 9, we have that

\[
|R_2(p) + R_3(p + \lambda \Delta)| \bigg|_{p_1 = 1 + \frac{7\sqrt{h}}{4}} - \max_{p \in [1, 1 + \frac{7\sqrt{h}}{4}]} \{ R_2(p) + R_3(p + \lambda \Delta) \}
\]

\[
\geq \frac{1}{A} \left( \left( 1 + \frac{7\sqrt{h}}{4} + \lambda \Delta - 2 \right)^2 - \left( \frac{13\sqrt{h}}{4} \right)^2 - \left( 1 + \frac{7\sqrt{h}}{2} + \lambda \Delta - 2 \right)^2 \right)
\]

\[
= \frac{\sqrt{h}}{4A} \left( 14 - 14 \lambda \Delta - 79\sqrt{h} \right) \geq 0,
\]
where the last inequality is because $\lambda \Delta \leq 1 - \epsilon$ and $h \leq \frac{\epsilon^2}{40}$. Note that $h \leq \frac{\epsilon^2}{40}$ also implies that $1 + \frac{7\sqrt{h}}{2} \leq p^\ast(d_3) - \lambda \Delta$ is a valid solution to the RHS of Eq. (EC.5). Together with Eq. (EC.5), we have that

$$
\max_{p_2,p_3:|p_2-p_3| \leq \lambda \Delta} \{R_2(p_2) + R_3(p_3)\} = \max_{p \in [1, \frac{7\sqrt{h}}{2}, p^\ast(d_3) - \lambda \Delta]} \{R_2(p) + R_3(p + \lambda \Delta)\}
$$

$$
= \max_{p \in [1, \frac{7\sqrt{h}}{2}, p^\ast(d_3) - \lambda \Delta]} \left\{ \frac{3}{8} - \frac{1}{A} \left( (p - p^\ast(d_1))^2 + (p + \lambda \Delta - p^\ast(d_3))^2 \right) \right\}
$$

$$
= \frac{3}{8} - \frac{1}{2A} (p^\ast(d_3) - p^\ast(d_1) - \lambda \Delta)^2. \tag{EC.6}
$$

By Lemma 9, we compute that

$$
R_2(p^\ast(d_2;\mathcal{T}')) + R_3(p^\ast(d_3;\mathcal{T}')) = R_2\left( \frac{1}{2} (p^\ast(d_1) + p^\ast(d_3)) + \frac{\lambda \Delta'}{2} \right) + R_3\left( \frac{1}{2} (p^\ast(d_1) + p^\ast(d_3) - \frac{\lambda \Delta'}{2} \right)
$$

$$
= \frac{3}{8} - \frac{1}{2A} (p^\ast(d_3) - p^\ast(d_1) - \lambda \Delta')^2. \tag{EC.7}
$$

Combining Eq. (EC.6) and Eq. (EC.7), we conclude that

$$
R_2(p^\ast(d_2;\mathcal{T}')) + R_3(p^\ast(d_3;\mathcal{T}')) - [R_2(p_2) + R_3(p_3)]
$$

$$
\geq \frac{1}{2A} (p^\ast(d_3) - p^\ast(d_1) - \lambda \Delta)^2 - \frac{1}{2A} (p^\ast(d_3) - p^\ast(d_1) - \lambda \Delta')^2
$$

$$
= \frac{\lambda}{2A} (2\Delta - \lambda(\Delta + \Delta'))(\Delta' - \Delta) \geq \frac{\lambda}{2A} \cdot 2(1 - \lambda) \cdot \frac{\sqrt{h}}{4} \geq \epsilon \frac{\lambda \sqrt{h}}{4}. \tag{EC.8}
$$

**EC.1.6. Proof of Lemma 12**

Proof of Lemma 12. Let $\mathcal{P}^{\leq t}_{\mathcal{J},\pi}$ be the probability measure over the first $t$ time periods of $\mathcal{P}_{\mathcal{J},\pi}$ (for $\mathcal{J} \in \{\mathcal{I}, \mathcal{T}\}$). Observe that $\mathcal{P}^{\leq t}_{\mathcal{I},\pi}$ is exactly $\mathcal{P}_{\mathcal{I},\pi}$. Therefore, to prove the lemma, it suffices to prove that, for any $t \in \{1, 2, 3, \ldots, T\}$,

$$
D_{KL}(\mathcal{P}^{\leq t}_{\mathcal{I},\pi} \| \mathcal{P}^{\leq t}_{\mathcal{T},\pi}) \leq D_{KL}(\mathcal{P}^{\leq t-1}_{\mathcal{I},\pi} \| \mathcal{P}^{\leq t-1}_{\mathcal{T},\pi}) + \text{Pr}_{\mathcal{P}_{\mathcal{T},\pi}} \left[ p^{(i)}(d_1;\mathcal{T},\pi) \in \left[ 1, 1 + \frac{7\sqrt{h}}{4} \right] \right] \cdot \frac{4h^2}{A^2}. \tag{EC.8}
$$

Let $H_i$ be the pricing and demand records for both customer groups during time periods $1, 2, \ldots, t$, and we slightly abuse the notation by also denoting by $\mathcal{P}^{\leq t}_{\mathcal{J},\pi}(H_i)$ the corresponding probability density for $\mathcal{P}^{\leq t}_{\mathcal{J},\pi}$. Note that

$$
D_{KL}(\mathcal{P}^{\leq t}_{\mathcal{I},\pi} \| \mathcal{P}^{\leq t}_{\mathcal{T},\pi}) = \mathbb{E}_{H_t \sim \mathcal{P}^{\leq t}_{\mathcal{I},\pi}} \left[ \ln \frac{\mathcal{P}^{\leq t}_{\mathcal{I},\pi}(H_t)}{\mathcal{P}^{\leq t}_{\mathcal{T},\pi}(H_t)} \right]
$$

$$
= \mathbb{E}_{(H_{t-1}, p_{1}^{(i)}, d_{1}^{(i)}, p_{2}^{(i)}, d_{2}^{(i)}) \sim \mathcal{P}^{\leq t}_{\mathcal{I},\pi}} \left[ \ln \frac{\mathcal{P}^{\leq t-1}_{\mathcal{I},\pi}(H_{t-1})}{\mathcal{P}^{\leq t-1}_{\mathcal{T},\pi}(H_{t-1})} + \ln \frac{\mathcal{P}^{\leq t-1}_{\mathcal{I},\pi}(H_{t-1})}{\mathcal{P}^{\leq t-1}_{\mathcal{T},\pi}(H_{t-1})} + \ln \frac{\mathcal{D}_{KL}(\text{Ber}(d_1^{(i)}) \| \text{Ber}(d_2^{(i)})))}{\mathcal{D}_{KL}(\text{Ber}(d_1^{(i)}) \| \text{Ber}(d_2^{(i)})))} \right]. \tag{EC.9}
$$

$$
= \mathbb{E}_{H_{t-1} \sim \mathcal{P}^{\leq t-1}_{\mathcal{I},\pi}} \left[ \ln \mathcal{P}^{\leq t-1}_{\mathcal{I},\pi}(H_{t-1}) \right] + \mathbb{E}_{p_{1}^{(i)} \sim \mathcal{P}^{\leq t}_{\mathcal{I},\pi}} \left[ D_{KL}(\text{Ber}(d_1^{(i)}) \| \text{Ber}(d_2^{(i)}))) \right]. \tag{EC.10}
$$
In Eq. (EC.9), $p_i^{(t)}$ and $D_i^{(t)}$ respectively denote the price for and the demand from the customer group $i$, and we use $\text{Ber}(\cdot|\mu)$ to denote the probability mass of the Bernoulli distribution with parameter $\mu$. In Eq. (EC.9), we only consider the KL-divergence between the customer group 1 because the demand rates and distributions of the customer group 2 are the same for $I$ and $I'$.

Note that the first term in Eq. (EC.10) is exactly $D_{KL}(P_{I,\pi}^{\leq t-1} || P_{I',\pi}^{\leq t-1})$, and we upper bound second term in Eq. (EC.10) by

$$E_{p_i^{(t)} \sim P_{I,\pi}^{\leq t}} \left[ D_{KL}(\text{Ber}(d_1(p_i^{(t)})) \parallel \text{Ber}(d_2(p_i^{(t)}))) \right]$$

$$\leq \Pr_{p_i^{(t)} \sim P_{I,\pi}^{\leq t}} \left[ p_i^{(t)} \in \left[ 1, 1 + \frac{7\sqrt{h}}{4} \right] \right] \cdot \sup_{p \in \left[ 1, 1 + \frac{7\sqrt{h}}{4} \right]} \left\{ D_{KL}(\text{Ber}(d_1(p)) \parallel \text{Ber}(d_2(p))) \right\}$$

$$\leq \Pr_{p_i^{(t)} \sim P_{I,\pi}^{\leq t}} \left[ p_i^{(t)} \in \left[ 1, 1 + \frac{7\sqrt{h}}{4} \right] \right] \cdot \frac{4h^2}{A^2}, \quad \text{(EC.11)}$$

where the first inequality is because $d_1(p) = d_2(p)$ for all $p \in \left[ 1, 1 + \frac{7\sqrt{h}}{4}, 2 \right]$, and the second inequality is due to Item (e) of Lemma 8. Combining Eq. (EC.10) and Eq. (EC.11), we prove Eq. (EC.8), and therefore prove the lemma. □