Pricing with Contextual Elasticity and Heteroscedastic Valuation

Jianyu Xu [1]   **Yu-Xiang Wang** [2]

Abstract

We study an online contextual dynamic pricing problem, where customers decide whether to purchase a product based on its features and price. We introduce a novel approach to modeling a customer's expected demand by incorporating feature-based price elasticity, which can be equivalently represented as a valuation with heteroscedastic noise. To solve the problem, we propose a computationally efficient algorithm called "Pricing with Perturbation (PwP)", which enjoys an $O(\sqrt{dT \log T})$ regret while allowing arbitrary adversarial input context sequences. We also prove a matching lower bound at $\Omega(\sqrt{dT})$ to show the optimality regarding d and T (up to $\log T$ factors). Our results shed light on the relationship between contextual elasticity and heteroscedastic valuation, providing insights for effective and practical pricing strategies.

1. Introduction

Contextual pricing, a.k.a., Feature-based dynamic pricing, considers the problem of setting prices for a sequence of highly specialized or individualized products. With the growth of e-commerce and the increasing popularity of online retailers as well as customers, there has been a growing interest in this area (see, e.g., Amin et al., 2014; Qiang & Bayati, 2016; Javanmard & Nazerzadeh, 2019; Shah et al., 2019; Cohen et al., 2020; Xu & Wang, 2021; Bu et al., 2022).

Formulated as a learning problem, the seller has no prior knowledge of ideal prices but is expected to learn on the fly by exploring different prices and adjusting their pricing strategy after collecting every demand feedback from customers. Different from non-contextual dynamic pricing (Kleinberg & Leighton, 2003) where identical products are sold repeatedly, a contextual pricing agent is expected to

generalize from one product to another in order to successfully price a previously-unseen product. A formal problem setup is described below:

Contextual pricing. For $t = 1, 2, ..., T$:
1. A product occurs, described by a context $x_t \in \mathbb{R}^d$.
2. The seller (we) proposes a price $p_t \geq 0$.
3. The customer reveals a demand $0 \leq D_t \leq 1$.
4. The seller gets a reward $r_t = p_t \cdot D_t$.

Here T is the time horizon, and the (random) demand D_t is drawn from a distribution determined by context (or feature) x_t and price p_t. The sequence of contexts $\{x_t\}$ can be either independently and identically distributed (iids) or chosen arbitrarily by an adversary. The seller's goal is to minimize the cumulative *regret* against the sequence of optimal prices.

Existing works on contextual pricing usually assumes linearity on the demand, but they fall into two camps. On the one hand, the "linear demand" camp (Qiang & Bayati, 2016; Ban & Keskin, 2021; Bu et al., 2022) assumes the *demand* D_t as a (generalized) linear model. A typical model is $D_t = \lambda(\alpha p_t + x_t^T \beta) + \epsilon_t$. Here $\alpha < 0$ is a parameter closely related to the *price elasticity*. We will rigorously define a price elasticity in Appendix A.1 according to Parkin et al. (2002), where we also show that α is the *coefficient of elasticity*. Besides of α, other parameters like $\beta \in \mathbb{R}^d$ captures the base demand of products with feature x_t, ϵ_t is a zero-mean demand noise, and λ is a known monotonically increasing link function. With this model, we have a noisy observation on the expected demand, which is reasonable as the same product is offered many times in period t. On the other hand, the "linear valuation" camp (Cohen et al., 2020; Javanmard & Nazerzadeh, 2019; Xu & Wang, 2021) models a buyer's *valuation* y_t as linear and assumes a binary demand $D_t = \mathbb{1}[p_t \leq y_t]$. All three works listed above assume a *linear-and-noisy* model with $y_t = x_t^\top \theta^* + N_t$, where $\theta^* \in \mathbb{R}^d$ is an unknown linear parameter that captures common valuations and N_t is an idiosyncratic noise assumed to be iid.

Interestingly, the seemingly different modeling principles are closely connected to each other. In the "linear valuation" camp, notice that a customer's probability of "buying" a product equals $\mathbb{E}[D_t]$, which is further given by

---
[1]University of California, Santa Barbara [2]University of California, San Diego. Correspondence to: Jianyu Xu <xu_jy15@ucsb.edu>, Yu-Xiang Wang <yuxiangw@ucsd.edu>.

$$\mathbb{E}[D_t|p] = \mathbb{P}[y_t \geq p] := S(p - x_t^\top \theta^*),$$

where S is the survival function of N_t (i.e. $S(w) = 1 - \mathrm{CDF}(w)$ for $w \in \mathbb{R}$). This recovers a typical linear demand model by taking $\lambda(w) = S(-w)$ with $\alpha = -1$ and $\beta = \theta^*$. In other words, the distribution of N_t completely characterizes the demand function $\lambda(\cdot)$ and vice versa.

However, the "linear demand" camp is not satisfied with a fixed $\alpha = -1$, while the "linear valuation" camp are skeptical about an observable demand D_t even with zero-mean iid noise. One common limitation to both models is that neither captures how feature x_t affects the price elasticity.

Our model. To address this issue, we propose a natural model that unifies the perspectives of both groups. Also, we resolve the common limitation by modeling *heteroscedasticity*, where we assume that the elasticity coefficient α is *linearly* dependent on feature x_t. This contextual modeling originates from the fact that different products have different price elasticities (Anderson et al., 1997), and a linear elasticity is widely adopt in literature (Bijmolt et al., 2005). A detailed discussion will be provided in Section 6.

In specific, we assume:

$$D_t \sim \mathrm{Ber}(S(x_t^\top \eta^* \cdot p_t - x_t^\top \theta^*)), \tag{1}$$

which adopts a generalized linear demand model (GLM) and a Boolean-censored feedback simultaneously. From the perspective of valuation model, it is *equivalent* to assume

$$D_t = \mathbb{1}[p_t \leq y_t]$$
$$\text{where} \quad y_t = \frac{1}{x_t^\top \eta^*} \cdot (x_t^\top \theta^* + N_t) \tag{2}$$
$$\text{and} \quad \mathrm{CDF}_{N_t}(w) = 1 - S(w).$$

Although Eq. (1) seems more natural than Eq. (2), they are equivalent to each other (with reasonable assumptions on S). Notice that the random valuation y_t is *heteroscedastic*, which means its variance is not the same constant across a variety of x_t's. We provide a detailed interpretation of this linear fractional valuation model in appendix.

1.1. Contributions.

Our main results are twofold.

1. We propose a new demand model that assumes a feature-dependent price elasticity on every product. Equivalently, we model the heteroscedasticity on customers' valuations among different products. This model unifies the "linear demand" and "linear valuation" camps.

2. We propose a "Pricing with Perturbation (PwP)" algorithm that achieves $O(\sqrt{dT \log T})$ regret on this model, which is optimal up to $\log T$ factors. This regret upper bound holds for both i.i.d. and adversarial $\{x_t\}$ sequences.

1.2. Technical Novelty

To the best of our knowledge, we are the first to study a contextual pricing problem with heteroscedastic valuation and Boolean-censored feedback. Some existing works, including Javanmard & Nazerzadeh (2019); Miao et al. (2019); Ban & Keskin (2021); Wang et al. (2021a), focus on related topics and achieve theoretical guarantees. However, their methodologies are not applicable to our settings due to substantial obstacles, which we propose novel techniques to overcome.

Randomized surrogate regret. Xu & Wang (2021) solves the problem with $x_t^\top \eta^* = 1$, by taking the negative log-likelihood as a surrogate regret and running an optimization oracle that achieves a fast rate (i.e. an $O(\log T)$ regret). However, the log-likelihood is no longer a surrogate regret in our setting, since it is not "convex enough" and therefore cannot provide sufficient (Fisher) information. In this work, we overcome this challenge by constructing a *randomized* surrogate loss function, whose *expectation* is "strongly convex" enough to upper bound the regret.

OCO for adversarial inputs. Javanmard & Nazerzadeh (2019) and Ban & Keskin (2021) study the problem with unknown or heterogeneous noise variances (i.e. elasticity coefficients), but their techniques highly rely on the distribution of the feature distributions. As a result, their algorithm could be easily attacked by an adversarial $\{x_t\}$ series. In our work, we settle this issue by conducting an online convex optimization (OCO) scheme while updating parameters. Instead of estimating from the history that requires sufficient randomness in the inputs, our algorithm can still work well for adversarial inputs.

In addition, our algorithm has more advanced properties such as computational efficiency and information-theoretic optimality. For more highlights of our algorithm, please refer to Section 4.1.

2. Related Works

Here we present a review of the pertinent literature on contextual pricing and heteroscedasticity in machine learning, aiming to position our work within the context of related studies. For a detailed comparison on closely-related problem settings and regret bounds, please refer to Table 1. For more related works on non-contextual pricing, contextual pricing, contextual searching and contextual bandits, please refer to Wang et al. (2021b), Xu & Wang (2021), Krishnamurthy et al. (2021) and Zhou (2015) respectively.

Pricing with (Generalized) Linear Demand. As we mentioned in Section 1.2, there are a large number of recent works on contextual dynamic pricing problems, and we refer to Ban & Keskin (2021) as a detailed introduction. On the one hand, Qiang & Bayati (2016); Nambiar et al. (2019); Miao et al. (2019); Wang et al. (2021a); Ban & Keskin (2021); Bu et al. (2022) assume a (generalized) linear demand model with noise, i.e. $\mathbb{E}[D_t] = g(\alpha p_t - \beta^\top x_t)$. Among those papers, Miao et al. (2019) worksl with a fixed α while we assume α as context-dependent. Wang et al. (2021a) and Ban & Keskin (2021) are the closest to our problem setting, which consist of a generalized linear demand model and noisy observations. On the one hand, Ban & Keskin (2021) assumes independent add-on noises (while we allow binary martingale observations). With the development of a least-square estimator, they present an algorithm that achieves $\tilde{O}(s\sqrt{T})$ regret (with s being the sparsity factor). On the other hand, Wang et al. (2021a) further gets rid of the independence among noises and allow them to be idiosyncratic. They proposes a UCB-based algorithm with $\tilde{O}(d\sqrt{T})$ regret and another Thompson-Sampling-based algorithm with $\tilde{O}(d^{\frac{3}{2}}\sqrt{T})$ regret, both of which are sub-optimal in d. Moreover, all works mentioned above assume the context sequence $\{x_t\}$ to be i.i.d., whereas we consider it "too good to be true" and work towards an algorithm adaptive to adversarial input sequences.

Linear Valuation. Golrezaei et al. (2019); Shah et al. (2019); Goyal & Perivier (2021) adopts the linear valuation model $y_t = x_t^\top \theta^* + N_t$, which is a special case of our model as $x_t^\top \eta^* = 1$. However, existing works diverge on the assumption whether the seller has precise knowledge on the noise distribution. On the one hand, Cohen et al. (2020); Javanmard & Nazerzadeh (2019); Xu & Wang (2021) assume N_t's drawn from a **known** distribution. Specifically, both Javanmard & Nazerzadeh (2019) and Xu & Wang (2021) achieve an $O(d \log T)$ regret for stochastic and adversarial $\{x_t\}$ sequences, respectively. Javanmard & Nazerzadeh (2019) also studies the setting when $x_t^\top \eta^*$ is fixed but unknown (in our model) and achieves $O(d\sqrt{T})$ regret for stochastic $\{x_t\}$ sequences. In comparison, we achieve $O(\sqrt{dT \log T})$ on a more general problem and get rid of those assumptions. On the other hand, Fan et al. (2021); Luo et al. (2021); Xu & Wang (2022); Luo et al. (2022) study the problem with unspecified noise distributions. Fan et al. (2021) introduces an exploration-first algorithm with carefully-designed unbiased Method-of Moment estimators. Luo et al. (2021; 2022) proposes UCB-style algorithms with optimistic plug-in estimators, and Xu & Wang (2022) adopts an EXP-4 algorithm with hypothesis discretization. Their methodologies could potentially be extended to a generalized version of our problem setting, where the valuation noise (before scaling) N_t is drawn from an unknown distribution. However, it is worth noting that our problem setting

is not a special case of theirs: The expected valuation in their settings is still linear ($E[y_t] = x_t^\top \theta^*$), while our paper assumes $E[y_t] = \frac{x_t^\top \theta^*}{x_t^\top \eta^*}$ that is a linear fractional mapping. We have included more discussions in Appendix C (see "Algorithm and analysis for unknown link function $S(\cdot)$") regarding this aspect.

Nonparametric Valuation. There are various studies on nonparametric pricing problems. Specifically, Chen et al. (2019) studies the contextual pricing problem with non-parametric demand model and binary feedback, achieving a $\tilde{O}(T^{\frac{2+d}{4+d}})$ regret for d-dimension covariates. Ye & Jiang (2024) studies the non-contextual non-parametric pricing problem with β-th order smooth demand curve but without pre-knowledge on β, recovering the minimax rate of $\tilde{O}(T^{\frac{\beta+1}{2\beta+1}})$ given certain additional assumptions. perakis2023dynamic There are various studies on nonparametric pricing problems. Specifically, Chen et al. (2019) studies the contextual pricing problem with non-parametric demand model and binary feedback, achieving a $\tilde{O}(T^{\frac{2+d}{4+d}})$ regret for d-dimension covariates. Ye & Jiang (2024) studies the non-contextual non-parametric pricing problem with β-th order smooth demand curve but without pre-knowledge on β, recovering the minimax rate of $\tilde{O}(T^{\frac{\beta+1}{2\beta+1}})$ given certain additional assumptions. Perakis & Singhvi (2023) studies a non-contextual pricing problem with additive noises, where they improve the switching cost while also maintaining the optimal regret rate.

Heteroscedasticity. Since the valuation noise is scaled by a $\frac{1}{x_t^\top \eta^*}$ coefficient, the valuation is *heteroscedastic*, referring to a situation where the variance is not the same constant across all observations. Heteroscedasticity may lead to bias estimates or loss of sample information. There are several existing methods handling this problem, including weighted least squares method (Cunia, 1964), White's test (White, 1980) and Breusch-Pagan test (Breusch & Pagan, 1979). Furthermore, Anava & Mannor (2016) and Chaudhuri et al. (2017) study online learning problems with heteroscedastic variances and provide regret bounds. For a formal and detailed introduction, we refer the audience to the textbook of Kaufman (2013).

3. Problem Setup
3.1. Notations
To formulate the problem, we firstly introduce necessary notations and symbols used in the following sections. The sales session contains T rounds with T known to the seller in advance[1]. At each time $t = 1, 2, \ldots, T$, a product with

---

[1]Here we assume T known for simplicity. For unknown T, we may apply a "doubling epoch" trick as Javanmard & Nazerzadeh (2019) without affecting the regret rate.

	Known α	**Unknown fixed** α		**Heteroscedastic** $\alpha = x_t^\top \eta^*$	
Features	Stochastic & Adversarial	Stochastic	Adversarial	Stochastic	Adversarial
Upper Bound	$d \log T$ [XW21]	$d\sqrt{T}$ [JN19]	$? \Rightarrow \sqrt{dT}$ **This Work**	$s\sqrt{T}$ (independent noises) [BK21] $d\sqrt{T}$ (idiosyncratic noises) [WTL21]	$? \Rightarrow \sqrt{dT}$ **This Work**
Lower Bound	$d \log T$ [BR12]	$\sqrt{T}$ [JN19]		$\sqrt{T} \Rightarrow \sqrt{dT}$ **This Work**	

Table 1. Existing related literature and results on regret bounds, with $\tilde{O}(\cdot)$ dropped. Note that each adversarial setting covers the stochastic setting under the same assumptions. Here [XW21] stands for Xu & Wang (2021), [JN19] for Javanmard & Nazerzadeh (2019), [BR12] for Broder & Rusmevichientong (2012), [BK21] for Ban & Keskin (2021), and [WTL21] for Wang et al. (2021a).

feature $x_t \in \mathbb{R}^d$ occurs and we propose a price $p_t \geq 0$. Then the nature draws a demand $D_t \sim \text{Ber}(S(x_t^\top \eta^* \cdot p_t - x_t^\top \theta^*))$, where $\theta^*, \eta^* \in \mathbb{R}^d$ are fixed unknown linear parameters and the link function $S : \mathbb{R} \to [0, 1]$ is non-increasing. By the end of time t, we receive a reward $r_t = p_t \cdot D_t$.

Equivalently, this customer has a valuation $y_t = \frac{x_t^\top \theta^* + N_t}{x_t^\top \eta^*}$ with noise $N_t \in \mathbb{R}$, and then make a decision $\mathbb{1}_t = \mathbb{1}[p_t \leq y_t] = D_t$ after seeing the price p_t. Similarly, we receive a reward $r_t = p_t \cdot \mathbb{1}_t$. Assume $N_t \sim \mathbb{D}_F$ is independently and identically distributed (i.i.d.), with cumulative distribution function (CDF) $F = 1 - S$. Denote $s := S'$ and $f := F'$.

3.2. Definitions

Here we define some key quantities. Firstly, we define an expected reward function.

Definition 3.1 (expected reward function). Define

$$r(u, \beta, p) := \mathbb{E}[r_t | x_t^\top \theta^* = u, x_t^\top \eta^* = \beta, p_t = p] = p \cdot S(\beta \cdot p - u) \tag{3}$$

as the expected reward function.

Given this, we further define a greedy price function as the argmax of $r(u, \beta, p)$ over p.

Definition 3.2 (greedy price function). Define $J(u, \beta)$ as a greedy price function, i.e. the price that maximizes the expected reward given $u = x_t^\top \theta^*$ and $\beta = x_t^\top \eta^*$.

$$J(u, \beta) = \underset{p \in \mathbb{R}}{\operatorname{argmax}} \, r(u, \beta, p) = \underset{p \in \mathbb{R}}{\operatorname{argmax}} \, p \cdot S(\beta \cdot p - u) \tag{4}$$

Notice that

$$J(u, \beta) = \underset{p}{\operatorname{argmax}} \, p \cdot S(\beta p - u) = \frac{1}{\beta} \cdot \underset{\beta p}{\operatorname{argmax}} \, \beta p \cdot S(\beta p - u) = \frac{1}{\beta} J(u, 1). \tag{5}$$

According to Xu & Wang (2021, Section B.1), we have the following properties.

Lemma 3.3. *Denote* $\varphi(w) := -\frac{S(w)}{s(w)} - w = \frac{1 - F(w)}{f(w)} - w$, *and we have* $J(u, \beta) = \frac{u + \varphi^{-1}(u)}{\beta}$. *Also, for* $u \geq 0$ *and* $\beta > 0$, *we have* $\frac{\partial J(u, \beta)}{\partial u} \in (0, 1)$.

Then we define a negative log-likelihood function of parameter hypothesis (θ, η) given the results at time t.

Definition 3.4 (log-likelihood functions). Denote $\ell_t(\theta, \eta)$ as the negative log-likelihood at time t, and define $L_t(\theta, \eta)$ as their summations:

$$-\ell_t(\theta, \eta) = \mathbb{1}_t \cdot \log S(x_t \top \eta \cdot p_t - x_t^\top \theta)$$
$$+ (1 - \mathbb{1}_t) \cdot \log(1 - S(x_t^\top \eta \cdot p_t - x_t^\top \theta)).$$
$$L_t(\theta, \eta) = \sum_{\tau=1}^{t} \ell_t. \tag{6}$$

Finally, we define a round-t expected regret and a cumulative expected regret.

Definition 3.5 (regrets). Define $Reg_t(p_t) := r(x_t^\top \theta^*, x_t^\top \eta^*, J(x_t^\top \theta^*, x_t^\top \eta^*)) - r(x_t^\top \theta^*, x_t^\top \eta^*, p_t)$ as the expected regret at round t, conditioning on price p_t. Also, define the cumulative regret as follows

$$Regret = \sum_{t=1}^{T} Reg_t(p_t) \tag{7}$$

3.3. Assumptions

We establish three technical assumptions to make our analysis and presentation clearer. Firstly, we assume that all feature and parameter vectors are bounded within a unit ball in Euclidean norm. This assumption is without loss of generality as it only rescales the problem.

Assumption 3.6 (bounded feature and parameter spaces). Assume features $x_t \in \mathcal{H}_x$ and parameters $\theta \in \mathcal{H}_\theta, \eta \in \mathcal{H}_\eta$. Denote $U_p^d := \{x \in \mathbb{R}^d, \|x\|_p \leq 1\}$ as an L_p-norm unit ball in $\mathbb{R}^d$. Assume all $\mathcal{H}_x, \mathcal{H}_\theta, \mathcal{H}_\eta \in U_p^d$. Also, assume

$x^\top \theta > 0, \forall x \in \mathcal{H}_x, \theta \in \mathcal{H}_\theta$ and $x^\top \eta > C_\beta > 0, \forall x \in \mathcal{H}_x, \eta \in \mathcal{H}_\eta$ for some constant $C_\beta \in (0,1)$.

The positiveness of elasticity coefficient $x^\top \eta > 0$ comes from the *Law of Demand* (Gale, 1955; Hildenbrand, 1983), stating that the quantity purchased varies inversely with price. This is derived from the *Law of Diminishing Marginal Utilities* and has been widely accepted (Marshall, 2009). We will show the necessity of assuming an elasticity lower bound C_β in Appendix C. In specific, we claim that any algorithm will suffer a regret of $\Omega(\frac{1}{C_\beta})$. For the simplicity of notation, we denote $[\theta; \eta] := [\theta^\top, \eta^\top]^\top \in \mathbb{R}^{2d}$ as the combination of d-dimension column vectors θ and η. Since we know that $x_t^\top \theta \in [0,1]$ and $x_t^\top \eta \in [C_\beta, 1]$, we have $J(x_t^\top \theta, x_t^\top \eta) \in [J(0,1), J(1, C_\beta)]$. Later we will show that the price perturbation is no more than $\frac{J(0,1)}{10}$. Therefore, we may have the following assumption.

Assumption 3.7 (bounded prices). For any price p_t at each time $t = 1, 2, \ldots, T$, we require $p_t \in [c_1, c_2]$, where $c_1 = \frac{J(0,1)}{2}$ and $c_2 = 2J(1, C_\beta)$.

Similar to Javanmard & Nazerzadeh (2019), we also assume a log-concavity on the noise CDF.

Assumption 3.8 (log-concavity). Every D_t is independently sampled according to Eq. (1), with $S(\omega) \in [0,1]$ and $s(\omega) = S'(\omega) > 0, \forall \omega \in \mathbb{R}$. Equivalently, the valuation noise $N_t \sim \mathbb{D}_F$ is independently and identically distributed (i.i.d.), with CDF $F = 1 - S$. Assume that $S \in \mathbb{C}^2$, and S and $(1 - S)$ are strictly log-concave.

4. Main Results

To solve the contextual pricing problem with featurized elasticity, we propose our "Pricing with Perturbation (PwP)" algorithm. In the following, we firstly describe the algorithm and highlight its properties, then analyze (and bound) its cumulative regret, and finally prove a regret lower bound to show its optimality.

4.1. Algorithm

The pseudocode of PwP is displayed as Algorithm 1, which calls an ONS oracle (Algorithm 2).

At each time t, it inherits parameters θ_t and η_t from $(t-1)$ and takes in a context vector x_t. By trusting in θ_t and η_t, it calculates a greedy price $\hat{p}_t$ and outputs a perturbed version $p_t = \hat{p}_t + \Delta_t$. After seeing customer's decision $\mathbb{1}_t$, PwP calls an "Online Newton Step (ONS)" oracle (see Algorithm 2) to update the parameters as θ_{t+1} and η_{t+1} for future use.

Algorithm 1 Pricing with Perturbation (PwP)

1: **Input:** parameter spaces $\mathcal{H}_\theta, \mathcal{H}_\eta$, link function S, time horizon T, dimension d
2: **Initialization:** parameters $\theta_1 \in \mathcal{H}_\theta$, $\eta_1 \in \mathcal{H}_\eta$, price perturbation Δ, cumulative likelihood $L_0 = 0$, matrix $A_0 = \epsilon \cdot I_{2d}$ and parameter ϵ, γ
3: **for** $t = 1, 2, \ldots, T$ **do**
4:     Observe x_t;
5:     Calculate greedy price $\hat{p}_t = J(x_t^\top \theta_t, x_t^\top \eta_t)$
6:     Sample $\Delta_t = \Delta$ with $\Pr = 0.5$ and $\Delta_t = -\Delta$ with $\Pr = 0.5$;
7:     Propose price $p_t = \hat{p}_t + \Delta_t$;
8:     Receive the customer's decision $\mathbb{1}_t$;
9:     Construct negative log-likelihood $\ell_t(\theta, \eta)$ and $L_t(\theta, \eta)$ as eq. (6);
10:    Update parameters:

$$[\theta_{t+1}; \eta_{t+1}] \leftarrow ONS([\theta_t; \eta_t])$$

11: **end for**

Algorithm 2 Online Newton Step (ONS)

1: **Input: current parameter $[\theta_t, \eta_t]$, likelihood $\ell_t(\theta, \eta)$, matrix A_t, parameter γ, parameter spaces $\mathcal{H}_\theta$ and $\mathcal{H}_\eta$.**
2: Calculate $\nabla_t = \nabla \ell_t(\theta, \eta)$;
3: Rank-1 update: $A_t = A_{t-1} + \nabla_t \nabla_t^\top$;
4: Newton step: $[\hat{\theta}_{t+1}; \hat{\eta}_{t+1}] = [\hat{\theta}_t; \hat{\eta}_t] - \frac{1}{\gamma} A_t^{-1} \nabla_t$;
5: Projection: $[\theta_{t+1}; \eta_{t+1}] = \Pi_{\mathcal{H}_\theta \times \mathcal{H}_\eta}^{A_t}([\hat{\theta}_{t+1}; \hat{\eta}_{t+1}])$;

4.1.1. HIGHLIGHTS

We highlight the achievements of the PwP algorithm in the following three aspects.

In this pricing problem. As we mentioned in Section 1.2, the key to solving this contextual elasticity (or heteroscedastic valuation) pricing problem is to construct a surrogate loss function. Xu & Wang (2021) adopts negative log-likelihood in their setting, which does not work for ours since it is not "convex" enough. In our PwP algorithm, we overcome this challenge by introducing a perturbation Δ on the proposed greedy price. This idea originates from the observation that the *variance* of p_t contributes positively to the "convexity" of the expected log-likelihood, which helps "re-build" the upper-bound inequality.

In online optimization. PwP perturbs the greedy action (price) it should have taken. This idea is similar to a "Following the Perturbed Leader (FTPL)" algorithm (Hutter et al., 2005) that minimizes the summation of the empirical risk and a random loss function serving as a perturbation. However, this might lead to extra computational cost as the random perturbation is not necessarily smooth and therefore

hard to optimize. In this work, PwP introduces a possible way to overcome this obstacle: Instead of perturbing the objective function, we may directly perturb the action to explore its neighborhood. Our regret analysis and results indicate the optimality of this method and imply a potentially wide application.

In information theory. We show the following fact in the regret analysis of PwP: By adding Δ perturbation on p_t, we may lose $O(\Delta^2)$ in reward but will gain $O(\Delta^2) \cdot I$ in Fisher information (i.e. the expected Hessian of negative log-likelihood function) in return. By Cramer-Rao Bound, this leads to $O(\frac{1}{\Delta^2})$ estimation error. In this way, we quantify the information (observing from exploration) on the scale of reward, which shares the same idea with the Upper Confidence Bound (Lai & Robbins, 1985) method that always maximizes the summation of empirical reward and information-traded reward.

Besides, PwP is computationally efficient as it only calls the ONS oracle for once. As for the ONS oracle, it updates an $A_t^{-1} = (A_{t-1} + \nabla_t \nabla_t^\top)^{-1}$ at each time t, which is with $O(d^2)$ time complexity according to the following *Woodbury matrix identity*

$$(A + xx^\top)^{-1} = A^{-1} - \frac{1}{1 + x^\top A^{-1} x} A^{-1} x (A^{-1} x)^\top. \quad (8)$$

4.2. Regret Upper Bound

Now we analyze the regret of PwP and propose an upper bound up to constant coefficients.

Theorem 4.1. *Under Assumption 3.6, 3.7 and 3.8, by taking* $\Delta = \min\left\{\left(\frac{d \log T}{T}\right)^{\frac{1}{4}}, \frac{J(0,1)}{10}, \frac{1}{10}\right\}$, *the algorithm PwP guarantees an expected regret at* $O(\sqrt{dT \log T})$.

In the following, we prove Theorem 4.1 by stating a thread of key lemmas. We leave the detailed proof of those lemmas to Appendix A.

Proof. The proof overview can be displayed as the following roadmap of inequalities:

$$\mathbb{E}[Regret] = \sum_{t=1}^{T} Reg_t(p_t)$$

$$\leq \mathbb{E}\left[\sum_{t=1}^{T} O\left((x_t^\top(\theta_t - \theta^*))^2 + (x_t^\top(\eta_t - \eta^*))^2 + \Delta^2\right)\right]$$

$$\leq O\left(\frac{\sum_{t=1}^{T} \mathbb{E}\left[\ell_t(\theta_t, \eta_t) - \ell_t(\theta^*, \eta^*)\right]}{\Delta^2} + T \cdot \Delta^2\right)$$

$$\leq O\left(\frac{d \log T}{\Delta^2} + T \cdot \Delta^2\right) = O(\sqrt{dT \log T}).$$

$$(9)$$

Here the first inequality is by the smoothness of regret function (see Lemma 4.2), the second inequality is by a special "strong convexity" of $\ell_t(\theta, \eta)$ that contributes to the surrogate loss (see Lemma 4.3), the third inequality is by Online Newton Step (see Lemma 4.4), and the last equality is by the value of Δ. A rigorous version of Eq. (9) can be found in Appendix A.4.

We firstly show the smoothness of $Reg_t(p_t)$:

Lemma 4.2 (regret smoothness). *Denote* $p_t^* := J(x_t^\top \theta^*, x_t^\top \eta^*)$. *There exists constants* $C_r > 0$ *and* $C_J > 0$ *such that*

$$Reg_t(p_t) \leq C_r \cdot (p_t - p_t^*)^2$$
$$\leq C_r \cdot 2\left(C_J \cdot \left[(x_t^\top(\theta_t - \theta^*))^2 + (x_t^\top(\eta_t - \eta^*))^2\right] + \Delta^2\right). \quad (10)$$

While the first inequality of Eq. (10) is from the smoothness, and the second inequality is by the Lipschitzness of function $J(u, \beta)$. Please refer to Appendix A.2 for proof details. We then show the reason why the log-likelihood function can still be a surrogate loss with carefully randomized p_t.

Lemma 4.3 (surrogate expected regret). *There exists a constant* $C_l > 0$ *such that* $\forall \theta \in \mathcal{H}_\theta, \eta \in \mathcal{H}_\eta$, *we have*

$$\mathbb{E}[\ell_t(\theta, \eta) - \ell_t(\theta^*, \eta^*)|\theta_t, \eta_t]$$
$$\geq \frac{C_l \Delta^2}{10}[(\theta - \theta^*)^\top, (\eta - \eta^*)^\top]\begin{bmatrix} x_t x_t^\top & 0 \\ 0 & x_t x_t^\top \end{bmatrix}$$
$$\cdot \begin{bmatrix} \theta - \theta^* \\ \eta - \eta^* \end{bmatrix} \quad (11)$$
$$= \frac{C_l \cdot \Delta^2}{10}\left[(x_t^\top(\theta - \theta^*))^2 + (x_t^\top(\eta - \eta^*))^2\right].$$

This lemma is crucial. We show a proof sketch here and defer the detailed proof to Appendix A.3.

Proof sketch of Lemma 4.3. We show that there exist constants $C_l > 0, C_p > 0$ such that

1. $\nabla^2 \ell_t(\theta, \eta) \succeq C_l \cdot \begin{bmatrix} x_t x_t^\top & -p_t x_t x_t^\top \\ -p_t x_t x_t^\top & p_t^2 x_t x_t^\top \end{bmatrix}$, and

2. $\mathbb{E}\begin{bmatrix} x_t x_t^\top & -p_t x_t x_t^\top \\ -p_t x_t x_t^\top & p_t^2 x_t x_t^\top \end{bmatrix}|\theta_t, \eta_t$
$\succeq C_p \Delta^2 \begin{bmatrix} x_t x_t^\top & 0 \\ 0 & x_t x_t^\top \end{bmatrix}$.

The first property above relies on the exp-concavity of ℓ_t. Notice that the second property does not hold without the $\mathbb{E}$ notation, as the left hand side is a $(a - b)^2$ form while the right hand side is in a $(a^2 + b^2)$ form. In general, there

exist no constant $c > 0$ such that $(a - b)^2 \geq c(a^2 + b^2)$. However, due to the randomness of p_t, we have

$$\mathbb{E}[p_t^2 | \hat{p}_t] = \mathbb{E}[p_t | \hat{p}_t]^2 + \Delta^2. \tag{12}$$

In this way, the *conditional expectation* of the left hand side turns to $(a - b)^2 + \lambda \cdot b^2$ and we have

$$
\begin{aligned}
& (a - b)^2 + \lambda b^2 \\
= & \left(\frac{1}{\sqrt{1 + \frac{\lambda}{2}}} \cdot a - \sqrt{1 + \frac{\lambda}{2}} \cdot b \right)^2 + (1 - \frac{1}{1 + \frac{\lambda}{2}}) a^2 + \frac{\lambda}{2} b^2 \\
\geq & \frac{\frac{\lambda}{2}}{1 + \frac{\lambda}{2}} \cdot (a^2 + b^2).
\end{aligned}
$$

$$\tag{13}$$

Similarly, we upper bound $\begin{bmatrix} x_t x_t^\top & 0 \\ 0 & x_t x_t^\top \end{bmatrix}$ with $\mathbb{E}[\nabla^2 \ell_t(\theta, \eta) | \theta_t, \eta_t]$ up to a $C_p \cdot \Delta^2$ coefficient.

With those two properties above, along with a property of likelihood function that $\mathbb{E}[\nabla \ell_t(\theta^*, \eta^*)] = 0$, we can prove Lemma 4.3 by taking a Taylor expansion of ℓ_t at $[\theta^*; \eta^*]$. ∎

Finally, we cite a theorem from Hazan (2016) as our Lemma 4.4 that reveals the surrogate regret rate on negative log-likelihood functions.

Lemma 4.4. *With parameters* $G = \sup_{\theta \in \mathcal{H}_\theta, \eta \in \mathcal{H}_\eta} \|\nabla l_t(\theta, \eta)\|_2$, $D = \sup \|[\theta_1; \eta_1] - [\theta_2; \theta_2]\| \leq 2$, $\alpha = C_e$, $\gamma = \frac{1}{2} \min\{\frac{1}{4GD}, \alpha\}$ *and* $\epsilon = \frac{1}{\gamma^2 D^2}$ *and* $T > 4$, *Keep running Algorithm 2 for* $t = 1, 2, \ldots, T$ *guarantees:*

$$\sup_{\{x_t\}} \left\{ \sum_{t=1}^T \ell_t(\theta_t, \eta_t) - \min_{\theta \in \mathcal{H}_\theta, \eta \in \mathcal{H}_\eta} \sum_{t=1}^T \ell_t(\theta, \eta) \right\} \leq 5(\frac{1}{\alpha} + GD) \log T \tag{14}$$

With all these lemma above, we have proved every line of Eq. (9). ∎

4.3. Lower Bounds

We claim that PwP is near-optimal in information theory, by proposing a matching regret lower bound in Theorem 4.5. We present the proof with valuation model to match with existing results.

Theorem 4.5. *Consider the contextual pricing problem setting with Bernoulli demand model given in Eq. (1). With all assumptions in Section 3 hold, any pricing algorithm has to suffer a* $\Omega(\sqrt{dT})$ *worst-case expected regret for* $T \geq 2d^2(1 + \log d)$, *with* T *the time horizon and* d *the dimension of context.*

Proof Sketch. We defer the proof details to Appendix A.5. The main idea is to reduce d numbers of 1-dimension problems to this problem setting. In fact, we may consider the following problem setting:

1. Construct set $X = \{e_i := [0, \ldots, 0, 1, 0, \ldots, 0]^\top \in \mathbb{R}^d$ with only i^{th} place being $1, i = 1, 2, \ldots, d\}$.

2. Let $\theta^* = [\frac{u_1}{\sigma_1}, \frac{u_2}{\sigma_2}, \frac{u_3}{\sigma_3}, \ldots, \frac{u_d}{\sigma_d}]^\top, \eta^* = [\frac{1}{\sigma_1}, \frac{1}{\sigma_2}, \frac{1}{\sigma_3}, \ldots, \frac{1}{\sigma_d}]^\top$, and therefore we have $\frac{e_i^\top \theta^* + N_t}{e_i^\top \eta^*} = u_i + \sigma_i \cdot N_t$.

3. At each time $t = 1, 2, \ldots, T$, sample $x_t \sim X$ independently and uniformly at random.

In this way, we divide the whole time series T into d separated sub-problems, where the Sub-Problem i has a valuation model $y_t(i) = u_i + \sigma_i \cdot N_t$, for $i = 1, 2, \ldots, d$. Let $N_t \sim \mathcal{N}(0, 1), t = 1, 2, \ldots, T$, and therefore $y_t(i) \sim \mathcal{N}(u_i, \sigma_i^2)$ are independent Gaussian random variables. According to Hoeffding's Inequality, each Sub-Problem i has at least $\frac{T}{2d}$ time periods with high probability. According to Xu & Wang (2021, Theorem 12) (originated from Broder & Rusmevichientong (2012, Theorem 3.1)), the regret lower bound of each sub-problem is $\Omega(\sqrt{\frac{T}{2d}})$. Therefore, the total expected regret lower bound is $\Omega(d \cdot \sqrt{\frac{T}{2d}}) = \Omega(\sqrt{Td})$. ∎

5. Numerical Experiments

Here we conduct numerical experiments to validate the low-regret performance of our algorithm PwP. Since we are the first to study this heteroscadestic valuation model, we do not have a baseline algorithm working for exactly the same problem. However, we can modify the RMLP-2 algorithm in Javanmard & Nazerzadeh (2019) by only replacing their max-likelihood estimator (MLE) for θ^* with a new MLE for both θ^* and η^*. This modified RMLP-2 algorithm does not have a regret guarantee in our setting, but it may still serve as a baseline to compare with. n the following part, we will compare the cumulative regrets of our ONSPP algorithm with the (modified) RMLP-2 in the following two scenarios:

1. The linear-fractional valuation $y_t = \frac{x_t^\top \theta^* + N_t}{x_t^\top \eta^*}$.

2. A *fully-linear* valuation $y_t = x_t^\top \theta^* + x_t^\top \eta^* \cdot N_t$.

Notice that the second valuation model is also heteroscedastic, but the linear scalar only multiplies with the noise. We design the first experiment to show the regret performance of our algorithm, and the second experiment to imply the generality of our valuation model and our algorithm to model-misspecified settings. In order to show the regret dependence w.r.t. T, we plot all cumulative regret curves in log-log plots, where an α slope indicates an $O(T^\alpha)$ dependence. We repeat each experiment for 20 times.

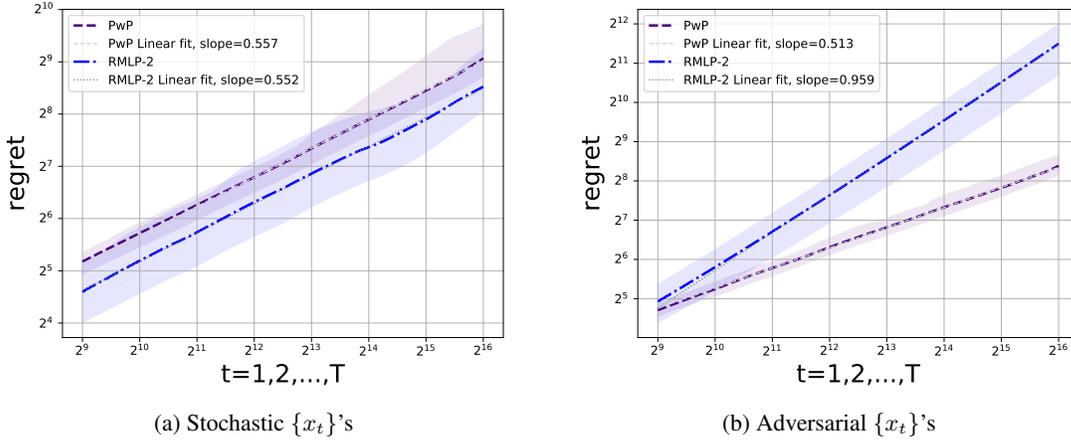(a) Stochastic $\{x_t\}$'s

(b) Adversarial $\{x_t\}$'s

Figure 1. The regret of PwP algorithm and a modified RMLP-2 algorithm on simulation data (generated according to Eq. (1)), plotted in log-log scales to indicate the regret dependence on T. Figure 1a and Figure 1b are for stochastic and adversarial $\{x_t\}$ sequences respectively. We also plot linear fits for those regret curves, where a slope-α line indicates an $O(T^\alpha)$ regret. The error bands are drawn with 0.95 coverage using Wald's test. From the figures, we know that PwP performs closely to its $O(\sqrt{T \log T})$ regret regardless of the types of input context sequences, whereas RMLP-2 fails in the attack of adversarial input.

5.1. Performances on well-assumed model

In this part, we test our algorithm ONSPP and the modified RMLP-2 on the correct demand model (as assumed in Eq. (1)) with both stochastic and adversarial $\{x_t\}$ sequences, respectively.

We test PwP and the modified RMLP-2 on the demand model assumed in Eq. (1) with both stochastic and adversarial $\{x_t\}$ sequences, respectively. Basically, we assume $T = 2^{16}$ $d = 2$, $N_t \sim \mathcal{N}(0, \sigma^2)$ with $\sigma = 0.5$, and we repeatedly run each algorithm for 20 times in each experiment setting. In order to show the regret dependence w.r.t. T, we plot all cumulative regret curves in log-log plots, where an α slope indicates an $O(T^\alpha)$ dependence.

Stochastic $\{x_t\}$. We implement and test PwP and RMLP-2 on stochastic $\{x_t\}$'s, where x_t are iid sampled from $\mathcal{N}(\mu_x, \Sigma_x)$ (for $\mu_x = [10, 10, \ldots, 10]^\top$ and some randomly sampled Σ_x) and then normalized s.t. $\|x_t\|_2 \leq 1$. The numerical results are shown in Figure 1a. Numerical results show that both algorithms achieve $\sim O(T^{0.56})$ regrets, which is close to the theoretic regret rate at $O(\sqrt{T \log T})$.

Adversarial $\{x_t\}$. Here we design an adversarial $\{x_t\}$ sequence to attack both algorithms. Since RMLP-2 divides the whole time horizon T into epochs with length $k = 1, 2, 3, \ldots$ sequentially and then does pure exploration at the beginning of each epoch, we may directly attack those pure-exploration rounds in the following way: (1) In each pure-exploration round (i.e. when $t = 1, 3, 6, \ldots, \frac{k(k+1)}{2}, \ldots$), let the context be $x_t = [1, 0]^\top$; (2) In any other round, let the context be $x_t = [0, 1]^\top$. In this way, the RMLP-2 algorithm

will never learn $\theta^*[2]$ and $\eta^*[2]$ since the inputs of pure-exploration rounds do not contain this information. Under this oblivious adversarial context sequence, we implement PwP and RMLP-2 and compare their performance. The results are shown in Figure 1b, indicating that PwP can still guarantee $O(T^{0.513})$ regret (close to $O(\sqrt{T \log T})$) while RMLP-2 runs into a linear regret.

As a high-level interpretation, the performance difference is because PwP adopts a "distributed" exploration at every time t while RMLP-2 makes it more "concentrated". Although both PwP and RMLP-2 take the same amount of exploration that optimally balance the reward loss and the information gain (and that is why they both perform well in stochastic inputs), randomly distributed exploration would save the algorithm from being "attacked" by oblivious adversary. In fact, this phenomenon is analog to ϵ-Greedy versus Exploration-first algorithms in multi-armed bandits. We will discuss more in Appendix C.

So far, we have presented the numerical results of running PwP and a modified RMLP-2 on the well-assumed demand model as Eq. (1) (or Eq. (2) equivalently). Besides of that, we also conduct experiments on a model-misspecification setting to show the robustness, where the true demand (or valuation) distribution is not the same as Eq. (1) or Eq. (2). The numerical results are presented in Appendix B.

6. Discussion

Here we discuss the motivation and the limitation of making Assumption 3.6. We leave the majority of discussion to Appendix C.

Assumption on linear elasticity. We assume that the price

elasticity $\alpha := x_t^\top \eta^*$ is linear on x_t. The reasons are three-fold. First, in accordance with the *parsimony principle*, we have opted for a linear model due to its simplicity and causality. This principle suggests that, all else being equal, the simpler model should be selected. Second, linear models provide a robust first-order approximation, are highly interpretable, and are computationally efficient for dynamic pricing where rapid decision-making is essential. Last but not least, the use of linear models to represent price elasticity is well-established in the literature, receiving empirical support from studies such as those by Bijmolt et al. (2005); Lijesen (2007) and Miller & Alberini (2016). While we recognize that price elasticity can exhibit complex behaviors, the linear approximation provides a strong foundation for initial analysis and has proven to be an effective tool in comparable research endeavors.

Necessity of lower-bounding $x_t^\top \eta^*$ from 0. As we state in Assumption 3.6, the price elasticity coefficient $x_t^\top \eta^*$ is lower bounded by a constant $C_\beta > 0$. On the one hand, this is necessary since we cannot have an upper bound on the optimal price without this assumption. On the other hand, according to Eq. (3), we know that $r(u, \beta, p) = r(u, 1, \beta \cdot p) \cdot \frac{1}{\beta}$, which indicates that the reward is rescaled by $\frac{1}{\beta}$. As a result, the regret should be proportional to $\frac{1}{C_\beta}$. Although a larger (i.e. closer to 0) elasticity would lead to a more *smooth* demand curve, this actually reduce the information we could gather from customers' feedback and slow down the learning process. We look forward to future researches getting rid of this assumption and achieve more adaptive regret rates. See more details in Appendix C.

Regret lower bounds for fixed unknown noise distributions. We claim a $\Omega(\sqrt{dT})$ regret lower bound in Theorem 4.5 with customers' demand model being Eq. (1). However, this result does not imply a $\Omega(\sqrt{dT})$ lower bound for customers' valuation being $y_t = x_t^\top \theta^* + N_t$ adopted by Javanmard & Nazerzadeh (2019); Cohen et al. (2020); Xu & Wang (2021). This is because our problem setting is more general than theirs, and our construction of $\Omega(\sqrt{dT})$ regret lower bounds are substantially beyond the scope of this specific subproblem. So far, the best existing regret lower bound for the linear noisy model ($y_t = x_t^\top \theta^* + N_t$) is still $\Omega(\sqrt{T})$. However, we conjecture that this should also be $\Omega(\sqrt{dT})$. The hardness of proving this lower bound comes from the fact that the noises are iid over time, which prevents a trivial separation into d independent sub-sequences as we do.

7. Conclusion

In summary, our work focuses on the problem of contextual pricing with highly differentiated products. We propose a contextual elasticity model that unifies the "linear demand" and "linear valuation" camps and captures the price effect and heteroscedasticity. To solve this problem, we develop an algorithm PwP, which utilizes Online Newton Step (ONS) on a surrogate loss function and proposes perturbed prices for exploration. Our analysis show that it guarantees a $O(\sqrt{dT \log T})$ regret even for adversarial context sequences. We also provide a matching $\Omega(\sqrt{dT})$ regret lower bound to show its optimality (up to $\log T$ factors). Besides, our numerical experiments also validate the regret bounds of PwP and its advantage over existing method. We hope this work would shed lights on the research of contextual pricing as well as online decision-making problems.

Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here. Please refer to Appendix C for a detailed discussion.

References

Amin, K., Rostamizadeh, A., and Syed, U. Repeated contextual auctions with strategic buyers. In *Advances in Neural Information Processing Systems (NIPS-14)*, pp. 622–630, 2014.

Anava, O. and Mannor, S. Heteroscedastic sequences: beyond gaussianity. In *International Conference on Machine Learning*, pp. 755–763. PMLR, 2016.

Anderson, P. L., McLellan, R. D., Overton, J. P., and Wolfram, G. L. Price elasticity of demand. *McKinac Center for Public Policy. Accessed October*, 13(2), 1997.

Baby, D., Xu, J., and Wang, Y.-X. Non-stationary contextual pricing with safety constraints. *Transactions on Machine Learning Research*, 2022.

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

Bijmolt, T. H., Van Heerde, H. J., and Pieters, R. G. New empirical generalizations on the determinants of price elasticity. *Journal of marketing research*, 42(2):141–156, 2005.

Breusch, T. S. and Pagan, A. R. A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the econometric society*, pp. 1287–1294, 1979.

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

Bu, J., Simchi-Levi, D., and Wang, C. Context-based dynamic pricing with partially linear demand model. In *Advances in Neural Information Processing Systems*, 2022.

Chaudhuri, K., Jain, P., and Natarajan, N. Active heteroscedastic regression. In *International Conference on Machine Learning*, pp. 694–702. PMLR, 2017.

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

Chen, X., Simchi-Levi, D., and Wang, Y. Utility fairness in contextual dynamic pricing with demand learning. *arXiv preprint arXiv:2311.16528*, 2023.

Chu, W., Li, L., Reyzin, L., and Schapire, R. Contextual bandits with linear payoff functions. In *International Conference on Artificial Intelligence and Statistics (AISTATS-11)*, pp. 208–214, 2011.

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

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

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

Cunia, T. Weighted least squares method and construction of volume tables. *Forest Science*, 10(2):180–191, 1964.

Fan, J., Guo, Y., and Yu, M. Policy optimization using semi-parametric models for dynamic pricing. *arXiv preprint arXiv:2109.06368*, 2021.

Gale, D. The law of supply and demand. *Mathematica scandinavica*, pp. 155–169, 1955.

Golrezaei, N., Jaillet, P., and Liang, J. C. N. Incentive-aware contextual pricing with non-parametric market noise. *arXiv preprint arXiv:1911.03508*, 2019.

Goyal, V. and Perivier, N. Dynamic pricing and assortment under a contextual mnl demand. *arXiv preprint arXiv:2110.10018*, 2021.

Hazan, E. Introduction to online convex optimization. *Foundations and Trends in Optimization*, 2(3-4):157–325, 2016.

Hildenbrand, W. On the" law of demand". *Econometrica: Journal of the Econometric Society*, pp. 997–1019, 1983.

Hutter, M., Poland, J., et al. Adaptive online prediction by following the perturbed leader. 2005.

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

Kaufman, R. L. *Heteroskedasticity in regression: Detection and correction*. Sage Publications, 2013.

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

Krishnamurthy, A., Lykouris, T., Podimata, C., and Schapire, R. Contextual search in the presence of irrational agents. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing (STOC-21)*, pp. 910–918, 2021.

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

Leme, R. P., Sivan, B., Teng, Y., and Worah, P. Learning to price against a moving target. In *International Conference on Machine Learning*, pp. 6223–6232. PMLR, 2021.

Lijesen, M. G. The real-time price elasticity of electricity. *Energy economics*, 29(2):249–258, 2007.

Luo, Y., Sun, W. W., et al. Distribution-free contextual dynamic pricing. *arXiv preprint arXiv:2109.07340*, 2021.

Luo, Y., Sun, W. W., and Liu, Y. Contextual dynamic pricing with unknown noise: Explore-then-ucb strategy and improved regrets. In *Advances in Neural Information Processing Systems*, 2022.

Marshall, A. *Principles of economics: unabridged eighth edition*. Cosimo, Inc., 2009.

Miao, S., Chen, X., Chao, X., Liu, J., and Zhang, Y. Context-based dynamic pricing with online clustering. *arXiv preprint arXiv:1902.06199*, 2019.

Miller, M. and Alberini, A. Sensitivity of price elasticity of demand to aggregation, unobserved heterogeneity, price trends, and price endogeneity: Evidence from us data. *Energy Policy*, 97:235–249, 2016.

Nambiar, M., Simchi-Levi, D., and Wang, H. Dynamic learning and pricing with model misspecification. *Management Science*, 65(11):4980–5000, 2019.

Parkin, M., Powell, M., and Matthews, K. *Economics*. Addison-Wesley, Harlow, 2002.

Perakis, G. and Singhvi, D. Dynamic pricing with unknown nonparametric demand and limited price changes. *Operations Research*, 2023.

Qiang, S. and Bayati, M. Dynamic pricing with demand covariates. *arXiv preprint arXiv:1604.07463*, 2016.

Shah, V., Johari, R., and Blanchet, J. Semi-parametric dynamic contextual pricing. *Advances in Neural Information Processing Systems*, 32, 2019.

Wang, H., Talluri, K., and Li, X. On dynamic pricing with covariates. *arXiv preprint arXiv:2112.13254*, 2021a.

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

White, H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: journal of the Econometric Society*, pp. 817–838, 1980.

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

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

Xu, J., Qiao, D., and Wang, Y.-X. Doubly fair dynamic pricing. In *International Conference on Artificial Intelligence and Statistics*, pp. 9941–9975. PMLR, 2023.

Ye, Z. and Jiang, H. Smoothness-adaptive dynamic pricing with nonparametric demand learning. In *International Conference on Artificial Intelligence and Statistics*, pp. 1675–1683. PMLR, 2024.

Zhou, L. A survey on contextual multi-armed bandits. *arXiv preprint arXiv:1508.03326*, 2015.

A. Definition and Proof Details

Here we show the proof details of the lemmas we have stated in Section 4.2. Before that, let us clarify some terminologies we mentioned in the main paper.

A.1. Definitions

Firstly, we rigorously define the concept of *price elasticity* occurring in Section 1.

Definition A.1 (Price Elasticity (Parkin et al., 2002))**.** Suppose $D(p)$ is a demand function of price p. Then the price elasticity E_d of demand is defined as

$$E_D := \frac{\Delta D(p)/D(p)}{\Delta p/p} = \frac{\partial D(p)}{\partial p} \cdot \frac{p}{D(p)}. \tag{15}$$

With this definition, along with our generalized linear demand model given in Eq. (1), the price elasticity for the expected demand $S(x_t^\top \eta^* \cdot p_t - x_t^\top \theta^*)$ is

$$
\begin{aligned}
E_D =& \frac{\partial S(x_t^\top \eta^* \cdot p_t - x_t^\top \theta^*)}{\partial p_t} \cdot \frac{p_t}{S(x_t^\top \eta^* \cdot p_t - x_t^\top \theta^*)} \\
=& x_t^\top \eta^* \cdot \frac{s(x_t^\top \eta^* \cdot p_t - x_t^\top \theta^*)}{S(x_t^\top \eta^* \cdot p_t - x_t^\top \theta^*)} \cdot p_t.
\end{aligned}
\tag{16}
$$

Here $s(\cdot) = S'(\cdot)$. Therefore, despite the effect of the link function and the price p_t, the price elasticity is proportional to the price coefficient $x_t^\top \eta^*$. This is why we call $x_t^\top \eta^*$ (or α in the general model $D(p) = \lambda(\alpha \cdot p + x_t^T \beta)$) the *elasticity coefficient* or *coefficient of elasticity* in Section 1.

A.2. Proof of Lemma 4.2

Proof. In order to prove Lemma 4.2, we show the following lemma that indicates the Lipschitzness of $J(u, \beta)$:

Lemma A.2 (Lipschitz of optimal price)**.** *There exists a constant $C_J > 0$ such that*

$$|J(u_1, \beta_1) - J(u_2, \beta_2)| \le C_J \cdot (|u_1 - u_2| + |\beta_1 - \beta_2|). \tag{17}$$

With this lemma, we get the second inequality of Eq. (10). We will prove this lemma later in this subsection. Now, we focus on the first inequality. Notice that

$$
\begin{aligned}
Reg_t(p_t) =& r(x_t^\top \theta^*, x_t^\top \eta^*, p_t^*) - r(x_t^\top \theta^*, x_t^\top \eta^*, p_t) \\
\le & -\frac{\partial r(u, \beta, p)}{\partial p}\Big|_{u=x_t^\top \theta^*, \beta=x_t^\top \eta^*, p=p_t^*}(p_t^* - p_t) \\
& -\frac{1}{2} \cdot \inf_{p \in [c_1, c_2], \beta \in [C_\beta, 1], u \in [0,1]} \frac{\partial^2 r(u, \beta, p)}{\partial p^2}\Big|_{u=x_t^\top \theta^*, \beta=x_t^\top \eta^*, p=p_t^*}(p_t^* - p_t)^2 \\
=& 0 + \frac{1}{2} \cdot \sup_{p \in [c_1, c_2], \beta \in [C_\beta, 1], u \in [0,1]} \{|2s(\beta \cdot p - u) \cdot \beta + p \cdot s'(\beta \cdot p - u) \cdot \beta^2|\}(p_t^* - p_t)^2.
\end{aligned}
\tag{18}
$$

Here the first line is by the definition of $Reg_t(p_t)$, the second line is by smoothness, the third line is by the optimality of p_t^*, and the last line is by calculus. Since $|2s(\beta \cdot p - u) \cdot \beta + p \cdot s'(\beta \cdot p - u) \cdot \beta^2|$ is continuous on $p \in [c_1, c_2], \beta \in [C_\beta, 1], u \in [0, 1]$, we denote this maximum as $2C_r$, which proves the first inequality of Eq. (10). ∎

Now we show the proof of Lemma A.2.

Proof of Lemma A.2. Since $J(u, \beta) = \frac{u + \varphi^{-1}(u)}{\beta}$ where $\varphi(w) = -\frac{S(w)}{s(w)} - w$. Notice that

$$\varphi'(w) = -\frac{d\frac{S(w)}{s(w)}}{dw} - 1 = \frac{d^2 \log(S(w))}{dw^2} \cdot \frac{S(w)^2}{s(w)^2} - 1 < -1 \tag{19}$$

since $S(w)$ is log-concave (as is assumed in Assumption 3.8). Given Eq. (19), we know that $\frac{d\varphi^{-1}(u)}{d(u)} = \frac{1}{\frac{d\varphi(w)}{dw}|_{w=\varphi^{-1}(u)}} \in$ $(-1, 0)$. Therefore, we have:

$$\frac{\partial J(u, \beta)}{\partial u} = \frac{1 + \frac{d\varphi^{-1}(u)}{du}}{\beta} \in (0, \frac{1}{C_\beta})$$

$$\frac{\partial J(u, \beta)}{\partial \beta} = \frac{\partial \frac{J(u,1)}{\beta}}{\partial \beta} = -\frac{J(u, 1)}{\beta^2} \in [-\frac{c_2}{C_\beta}, -c_1].$$

(20)

Therefore, we know that $J(u, \beta)$ is Lipschitz with respect to u and β respectively. Take $C_J = \max\{\frac{1}{C_\beta}, \frac{c_2}{C_\beta}\}$ and we get Eq. (17). ∎

A.3. Proof of Lemma 4.3

Proof. We firstly show the convexity (and exp-concavity) of $\ell_t(\theta, \eta)$ by the following lemma.

Lemma A.3 (exp-concavity). $\ell_t(\theta, \eta)$ *is convex and* C_e-*exp-concave with respect to* $[\theta; \eta]$, *where* $C_e > 0$ *is a constant dependent on* F *and* C_β. *Equivalently,* $\nabla^2 \ell_t(\theta, \eta) \succeq C_e \cdot \nabla \ell_t(\theta, \eta) \nabla \ell_t(\theta, \eta)^\top$. *Also, we have* $\nabla^2 \ell_t(\theta, \eta) \succeq C_l \cdot \begin{bmatrix} x_t x_t^\top & -p_t \cdot x_t x_t^\top \\ -p_t \cdot x_t x_t^\top & v_t^2 \cdot x_t x_t^\top \end{bmatrix}$ *for some constant* $C_l > 0$.

The proof of Lemma A.3 is mainly straightforward calculus, and we defer the proof to the end of this subsection. According to Lemma A.3, we have $\nabla^2 \ell_t(\theta, \eta) \succeq C_l \cdot \begin{bmatrix} x_t x_t^\top & -p_t \cdot x_t x_t^\top \\ -p_t \cdot x_t x_t^\top & p_t^2 \cdot x_t x_t^\top \end{bmatrix}$. Therefore, we know that

$$\ell_t(\theta, \eta) \geq \ell_t(\theta^*, \eta^*) + \nabla \ell_t(\theta^*, \eta^*)^\top \begin{bmatrix} \theta - \theta^* \\ \eta - \eta^* \end{bmatrix} + [(\theta - \theta^*)^\top, (\eta - \eta^*)^\top] C_l \begin{bmatrix} x_t x_t^\top & -p_t x_t x_t^\top \\ -p_t x_t x_t^\top & p_t^2 x_t x_t^\top \end{bmatrix} \begin{bmatrix} \theta - \theta^* \\ \eta - \eta^* \end{bmatrix}$$

(21)

According to the property of likelihood, we have $\mathbb{E}[\nabla \ell_t(\theta^*, \eta^*)|\theta_t, \eta_t] = 0$ for any θ_t and η_t. Combining this with Eq. (21), we get

$$\mathbb{E}[\ell_t(\theta, \eta) - \ell_t(\theta^*, \eta^*)|\theta_t, \eta_t] \geq C_l[(\theta - \theta^*)^\top, (\eta - \eta^*)^\top] \mathbb{E} \begin{bmatrix} x_t x_t^\top & -p_t x_t x_t^\top \\ -p_t x_t x_t^\top & p_t^2 x_t x_t^\top \end{bmatrix} |\theta_t, \eta_t \end{bmatrix} \begin{bmatrix} \theta - \theta^* \\ \eta - \eta^* \end{bmatrix}$$

(22)

Recall that $\hat{p}_t = J(x_t^\top \theta_t, x_t^\top \eta_t)$ and that $p_t = \hat{p}_t + \Delta_t$. Therefore, we know that the conditional expectations $\mathbb{E}[p_t|\theta_t, \eta_t] = \hat{p}_t$ and $\mathbb{E}[p_t^2|\theta_t, \eta_t] = \hat{p}_t^2 + \Delta^2$. Given this, we have

$$\mathbb{E} \begin{bmatrix} x_t x_t^\top & -p_t x_t x_t^\top \\ -p_t x_t x_t^\top & p_t^2 x_t x_t^\top \end{bmatrix} |\theta_t, \eta_t \end{bmatrix}$$

$$= \begin{bmatrix} x_t x_t^\top & -\hat{p}_t x_t x_t^\top \\ -\hat{p}_t x_t x_t^\top & (\hat{p}_t^2 + \Delta^2) x_t x_t^\top \end{bmatrix}$$

$$= \begin{bmatrix} x_t \\ -\hat{p}_t x_t \end{bmatrix} [x_t^\top, -\hat{p}_t x_t^\top] + \begin{bmatrix} 0 & 0 \\ 0 & \Delta^2 x_t x_t^\top \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{\sqrt{1+\frac{\Delta^2}{2}}} \cdot x_t \\ -\sqrt{1+\frac{\Delta^2}{2}} \hat{p}_t \cdot x_t \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{1+\frac{\Delta^2}{2}}} \cdot x_t^\top, -\sqrt{1+\frac{\Delta^2}{2}} \hat{p}_t \cdot x_t^\top \end{bmatrix} + \begin{bmatrix} (1 - \frac{1}{1+\frac{\Delta^2}{2}}) x_t x_t^\top & 0 \\ 0 & \frac{\Delta^2}{2} x_t x_t^\top \end{bmatrix}$$

(23)

Since $\Delta = \min \left\{ \left(\frac{d \log T}{T} \right)^{\frac{1}{4}}, \frac{J(0,1)}{10}, \frac{1}{10} \right\}$, we have $1 - \frac{1}{1+\frac{\Delta^2}{2}} = \frac{\frac{\Delta^2}{2}}{1+\frac{\Delta^2}{2}} \geq \frac{\Delta^2}{10}$. As a result, we have

$$\mathbb{E} \begin{bmatrix} x_t x_t^\top & -p_t x_t x_t^\top \\ -p_t x_t x_t^\top & p_t^2 x_t x_t^\top \end{bmatrix} |\theta_t, \eta_t \end{bmatrix}$$

$$\geq \frac{\Delta^2}{10} \cdot \begin{bmatrix} x_t x_t^\top & 0 \\ 0 & x_t x_t^\top \end{bmatrix}$$

(24)

This proves the lemma. ∎

Finally, we show the proof of Lemma A.3.

Proof of Lemma A.3. Recall that $\ell_t(\theta, \eta) = -\mathbb{1}_t \cdot \log(S(x_t^\top(p_t\eta - \theta))) - (1 - \mathbb{1}_t) \cdot \log(1 - S(x_t^\top(p_t\eta - \theta)))$. We first calculate the gradient and Hessian of $\ell_t(\theta, \eta)$ with respect to $[\theta; \eta]$. For notation simplicity, denote $w_t := x_t^\top(p_t\eta - \theta)$.

$$\nabla \ell_t = -\left(\mathbb{1}_t \cdot \frac{s(w_t)}{S(w_t)} - (1 - \mathbb{1}_t) \cdot \frac{s(w_t)}{1 - S(w_t)}\right) \cdot \begin{bmatrix} -x_t \\ p_t \cdot x_t \end{bmatrix} \tag{25}$$

$$\nabla^2 \ell_t = -\left(\mathbb{1}_t \cdot \frac{s'(w_t)S(w_t) - s(w_t)^2}{S(w_t)^2} + (1 - \mathbb{1}_t) \cdot \frac{-s'(w_t)(1 - S(w_t)) - s(w_t)^2}{(1 - S(w_t))^2}\right) \cdot \begin{bmatrix} -x_t \\ p_t \cdot x_t \end{bmatrix} \begin{bmatrix} -x_t^\top, & p_t \cdot x_t^\top \end{bmatrix}$$

$$= -\left(\mathbb{1}_t \cdot \frac{s'(w_t)S(w_t) - s(w_t)^2}{S(w_t)^2} + (1 - \mathbb{1}_t) \cdot \frac{-s'(w_t)(1 - S(w_t)) - s(w_t)^2}{(1 - S(w_t))^2}\right) \cdot \begin{bmatrix} x_t x_t^\top & -p_t \cdot x_t x_t^\top \\ -p_t \cdot x_t x_t^\top & p_t^2 \cdot x_t x_t^\top \end{bmatrix} \tag{26}$$

According to Assumption 3.8, we know that $S(w)$ and $(1 - S(w))$ are strictly log-concave, which indicates

$$\frac{d^2 \log(1 - S(w))}{dw^2} = \frac{-s'(w)(1 - S(w)) - s(w)^2}{(1 - S(w))^2} < 0$$
$$\frac{d^2 \log(S(w))}{dw^2} = \frac{s'(w)S(w) - s(w)^2}{S(w)^2} < 0, \forall w \in \mathbb{R}. \tag{27}$$

Since $w_t = p_t \cdot x_t^\top \eta - x_t^\top \theta$ where $p_t \in [c_1, c_2]$, we know that $w_t \in [-1, c_2]$. Since $\frac{d^2 \log(S(w))}{dw^2}$ and $\frac{d^2 \log(1 - S(w))}{dw^2}$ are continuous on $[-1, c_2]$, we know that

$$\mathbb{1}_t \cdot \frac{s'(w_t)S(w_t) - s(w_t)^2}{S(w_t)^2} + (1 - \mathbb{1}_t) \cdot \frac{-s'(w_t)(1 - S(w_t)) - s(w_t)^2}{(1 - S(w_t))^2}$$
$$\leq \sup_{w \in [-1, c_2]} \max\left\{\frac{s'(w_t)S(w_t) - s(w_t)^2}{S(w_t)^2}, \frac{-s'(w_t)(1 - S(w_t)) - s(w_t)^2}{(1 - S(w_t))^2}\right\} < 0. \tag{28}$$

Denote $C_l = -\sup_{w \in [-1, c_2]} \max\left\{\frac{s'(w_t)S(w_t) - s(w_t)^2}{S(w_t)^2}, \frac{-s'(w_t)(1 - S(w_t)) - s(w_t)^2}{(1 - S(w_t))^2}\right\} > 0$, and we know that

$$\nabla^2 \ell_t(\theta, \eta) \succeq C_l \cdot \begin{bmatrix} x_t x_t^\top & -p_t \cdot x_t x_t^\top \\ -p_t \cdot x_t x_t^\top & p_t^2 \cdot x_t x_t^\top \end{bmatrix}. \tag{29}$$

Similarly, we know that $\frac{s(w)}{S(w)}$ and $\frac{-s(w)}{1 - S(w)}$ are continuous on $[-1, c_2]$. Therefore, we may denote $C_G = \sup_{w \in [-1, c_2]} \max\left\{|\frac{s(w)}{S(w)}|, |\frac{-s(w)}{1 - S(w)}|\right\} > 0$ and get

$$\nabla \ell_t(\theta, \eta) \nabla \ell_t(\theta, \eta)^\top \preceq C_G^2 \cdot \begin{bmatrix} -x_t \\ p_t \cdot x_t \end{bmatrix} \begin{bmatrix} -x_t^\top, & p_t \cdot x_t^\top \end{bmatrix}. \tag{30}$$

Given all these above, we have

$$\nabla^2 \ell_t(\theta, \eta) \succeq C_l \cdot \begin{bmatrix} x_t x_t^\top & -p_t \cdot x_t x_t^\top \\ -p_t \cdot x_t x_t^\top & p_t^2 \cdot x_t x_t^\top \end{bmatrix}$$
$$= \frac{C_l}{C_G^2} \cdot C_G^2 \cdot \begin{bmatrix} -x_t \\ p_t \cdot x_t \end{bmatrix} \begin{bmatrix} -x_t^\top, & p_t \cdot x_t^\top \end{bmatrix} \tag{31}$$
$$\succeq \frac{C_l}{C_G^2} \cdot \nabla \ell_t(\theta, \eta) \nabla \ell_t(\theta, \eta)^\top.$$

Denote $C_e := \frac{C_l}{C_G^2}$ and we prove the lemma. ∎

A.4. Proof of Theorem 4.1

Proof. With all lemmas above, we have

$$
\begin{aligned}
\mathbb{E}[Regret] =& \mathbb{E}[\sum_{t=1}^{T} \mathbb{E}[Reg_t(p_t)|\theta_t, \eta_t]] \\
\leq& \mathbb{E}[\sum_{t=1}^{T} C_r \cdot 2 \cdot C_J \cdot \mathbb{E}[(x_t^\top(\theta_t - \theta^*))^2 + (x_T^\top(\eta_t - \eta^*))^2|\theta_t, \eta_t] + T \cdot C_r \cdot 2 \cdot \Delta^2] \\
\leq& \mathbb{E}[\sum_{t=1}^{T} 2C_r C_J \cdot \frac{10}{C_l \cdot \Delta^2} \cdot \mathbb{E}[\ell_t(\theta_t, \eta_t) - \ell_t(\theta^*, \eta^*)|\theta_t, \eta_t] + 2C_r T \Delta^2] \\
=& \frac{20 C_r C_J}{C_l \Delta^2} \mathbb{E}[\sum_{t=1}^{T} \ell_t(\theta_t, \eta_t) - \ell_t(\theta^*, \eta^*)] + 2C_r T \Delta^2 \\
=& O(\frac{d \log T}{\Delta^2} + \Delta^2 T) \\
=& O(\sqrt{dT \log T}).
\end{aligned}
\tag{32}
$$

Here the first line is by the law of total expectation, the second line is by Lemma 4.2, the third line is by Lemma 4.3, the fourth line is by equivalent transformation, the fifth line is by Lemma 4.4, and the sixth line is by the fact that $\Delta = \min\left\{\left(\frac{d \log T}{T}\right)^{\frac{1}{4}}, \frac{J(0,1)}{10}, \frac{1}{10}\right\}$. This holds the theorem. ∎

A.5. Proof of Theorem 4.5

Proof. Denote $\theta^* = [\theta_1, \theta_2, \ldots, \theta_d]^\top$ and $\eta^* = [\eta_1, \eta_2, \ldots, \eta_d]^\top$. We firstly construct a context set X as

$$
X = \{e_i, i = 1, 2, \ldots, d | e_i = [0, \ldots, 0, 1, 0, \ldots, 0]^\top \in \mathbb{R}^d, e_i[i] = 1, e_i[j] \neq 1, \forall j \neq i\}.
\tag{33}
$$

Then we sample each $x_t \sim_{i.i.d.} \mathbb{U}_X$, where $\mathbb{U}_X$ is a uniform distribution defined on each element of X (i.e. $\Pr[x_t = e_i] = \frac{1}{d}, \forall i \in [d], t \in [T]$). Denote $i_t := i$ if $x_t = e_i$. Now we decompose the indexes set $[T]$ of series $\{x_t\}_{t=1}^{T}$ into d subsets:

$$
S_i := \{t | x_t = e_i, t = 1, 2, \ldots, T\}, i = 1, 2, \ldots, d.
\tag{34}
$$

From the perspective of customers' valuations, we have $y_t = \frac{e_i^\top \theta^* + N_t}{e_i^\top \eta^*} = \frac{\theta_{i_t}}{\eta_{i_t}} + \frac{1}{\eta_{i_t}} \cdot N_t$ where N_t is an i.i.d. noise with known distribution. Let $N_t \sim_{i.i.d.} \mathcal{N}(0,1)$ as standard Gaussian noises. Therefore, each $i \in [d]$ determines a sub-problem (denoted as $\mathcal{P}_i$) that only happens when $t \in S_i$ and has a fixed valuation distribution, i.e. $y_t = \frac{\theta_i}{\eta_i} + \frac{1}{\eta_i} \cdot N_t \sim_{i.i.d.} \mathcal{N}(\frac{\theta_i}{\eta_i}, \frac{1}{\eta_i^2})$. For any $t \notin S_i$, neither y_t nor $\mathbb{1}_t$ is dependent on θ_i or η_i, which enables us to separately consider each $\mathcal{P}_i$. Denote $T_i := |S_i|$. In the following, we bound the least possible regret of this sub-problem as $\Omega(\sqrt{T_i})$.

Let $\theta_i = \frac{u_i}{\sigma_i}$ and $\eta_i = \frac{1}{\sigma_i}$ where $u_i, \sigma_i > 0$ are unknown parameters to be determined. Given this, customers' valuation distribution is $y_t \sim \mathcal{N}(u_i, \sigma_i^2)$ for $t \in S_i$. According to Theorem 12 of Xu & Wang (2021), let $u_i = \sqrt{\frac{\pi}{2}}$ and $\sigma_i \in \{1, 1 - T_i^{-\frac{1}{4}}\}$, and any algorithm has to suffer at least $\frac{1}{24000} \cdot \sqrt{T_i}$ regret.

Then we show that $T_i \geq \frac{T}{2d}$ with high probability. Notice that

$$
\begin{aligned}
\mathbb{E}[T_i] =& \mathbb{E}[\sum_{t=1}^{T} \mathbb{1}[x_t = e_i]] \\
=& \sum_{t=1}^{T} \Pr[x_t = e_i] \\
=& T \cdot \frac{1}{d} = \frac{T}{d}.
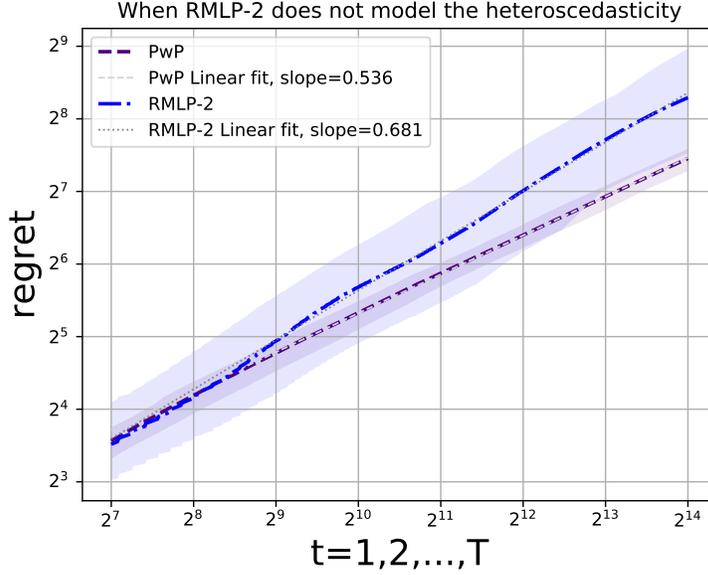\end{aligned}
\tag{35}
$$

Figure 2. Regrets of PwP versus the original homoscedastic RMLP-2 algorithm. In this log-log diagram, a $O(T^\alpha)$ regret curve is shown as a straight line with slope α. From the figure, we notice that PwP is optimal while RMLP-2 is sub-optimal, indicating the necessity of modeling homoscedasticity to achieve optimal regrets.

According to Hoeffding's Inequality, we have:

$$\Pr[|\sum_{t=1}^{T} \mathbb{1}[x_t = e_i] - \frac{T}{d}| \geq \frac{T}{2d}] \leq 2\exp\{-2\frac{T^2}{4d^2} \cdot \frac{1}{T}\} \tag{36}$$

$$\Rightarrow \quad \Pr[T_i \geq \frac{T}{2d}] \leq 2\exp\{-\frac{T}{2d^2}\}.$$

According to a union bound on the failure probability, with $Pr \geq 1 - 2d\exp\{-\frac{T}{2d^2}\}$, we have $T_i \geq \frac{T}{2d}, \forall i \in [d]$. Therefore, the expected regret satisfies

$$\begin{aligned}
\mathbb{E}[Regret] &= \mathbb{E}[\sum_{i=1}^{d} Regret(\mathcal{P}_i)] \\
&\geq \mathbb{E}[\sum_{i=1}^{d} \frac{1}{24000}\sqrt{T_i}] \\
&\geq \mathbb{E}[\sum_{i=1}^{d} \frac{1}{24000}\sqrt{\frac{T}{2d}} \cdot (1 - 2d\exp\{-\frac{T}{2d^2}\})] \\
&\geq \frac{1}{200000} \cdot \sqrt{Td}.
\end{aligned} \tag{37}$$

Here the last inequality comes from the assumption that $T \geq 2d^2(1 + \log d)$ and therefore $1 - 2d\exp\{-\frac{T}{2d^2}\} \geq 1 - \frac{2}{e} > \frac{1}{4}$. $\blacksquare$

B. More Experiments

B.1. Model Adaptivity

In this section, we show that it is necessary to model the heteroscedasticity. In specific, we compare PwP with the original RMLP-2 algorithm from Javanmard & Nazerzadeh (2019) that ignores heteroscedasticity in a heteroscedastic environment. We conduct both experiments for $T = 2^14$ rounds and repeat them for 10 epochs. The numerical results are displayed in
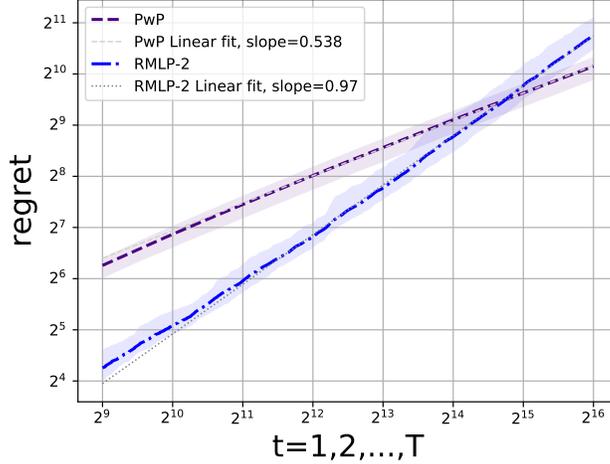
16

Figure 3. Regrets of misspecified PwP with expanded contexts, in comparison with a baseline RMLP-2 knowing the correct model. The results show that PwP still have a sub-linear regret in a certain period of time with context expansions, indicating that our linear demand model as Eq. (1) can be generalized to a linear valuation model as Eq. (38) in practice.

the lower figure, plotted in log-log diagrams. From the figure, we notice that the regret of RMLP-2 is much larger than PwP. Also, the slope of regrets of RMLP-2 is $0.681 >> 0.5$, indicating that it does not guarantee a $O(\sqrt{T})$ regret. In comparison, PwP still performs well as it achieves a $\sim O(T^{0.536})$ regret. This indicates that the algorithmic adaptivity of PwP to both homoscedastic and heteroscedastic environments is highly non-trivial, and a failure of capturing it would result in a substantial sub-optimality.

B.2. Model Misspecification

In Section 5, we compare the cumulative regrets of our PwP algorithm with the (modified) RMLP-2 on the *linear demand* model (as Eq. (1) or equivalently, the linear fractional valuation model as Eq. (2)). In this section, we consider a model-misspecific setting, where customer's true valuation is given by the following equation

$$y_t = x_t^\top \theta^* + x_t^\top \eta^* \cdot N_t \tag{38}$$

and the demand $D_t = \mathbb{1}_t = \mathbb{1}[p_t \leq y_t]$. As a result, Eq. (38) captures a *linear valuation* model with heteroscedastic valuation.

Now, we design an experiment to show the generalizability of both our PwP algorithm and our demand model as Eq. (1). In specific, we run the PwP algorithm that still models a customer's valuation as $\tilde{y}_t = \frac{x_t^\top \tilde{\theta}^* + \tilde{N}_t}{\tilde{x}_t^\top \tilde{\eta}^*}$, where $\tilde{x}_t \in \mathbb{R}^q$ is an *expanded* version of the original context x_t (i.e. $\tilde{x}_t = \pi(x_t)$ for some fixed expanding policy π) and $\tilde{\theta}^*, \tilde{\eta}^* \in \mathbb{R}^q$ are some fixed parameters[2]. Therefore, PwP is trying to learn those misspecified $\tilde{\theta}^*$ and $\tilde{\eta}^*$ although there does not exist such an underground truth.

We are curious whether the expansion of context (from x_t to $\tilde{x}_t$) would leverage the hardness of model misspecification. For $x = [x_1, x_2, \ldots, x_d]^\top$, denote $x^n := [x_1^n, x_2^n, \ldots, x_d^n]^\top$. Then for any context $x \in \mathbb{R}^d$, we specify each context-expanding policy as follows:

$$\begin{aligned} &\pi(x; x_0, \mathbf{a}) \\ &:= [x; (x - x_0)^{a_1}; (x - x_0)^{a_2}; \ldots; (x - x_0)^{a_m}]^\top \in \mathbb{R}^{(m+1)d}. \end{aligned} \tag{39}$$

The policy π in Eq. (39) is a polynomial expansion of x with index list $\mathbf{a} = [a_1, a_2, \ldots, a_m] \in \mathbb{Z}^m$, where $x_0 \in \mathbb{R}^d$ is a fixed start point of this expansion.

Now we consider the baseline to compare with. We claim that it is very challenging to solve the contextual pricing problem with customers' valuations being Eq. (38) with theoretic regret guarantees (although the $\Omega(\sqrt{T})$ lower bound given by

---

[2]We may assume $q \geq d$ without loss of generality.

Javanmard & Nazerzadeh (2019) still holds), and there are no existing algorithms targeting at this problem setting. However, there are still some straightforward algorithms that might approach it: For example, a max-likelihood estimate (MLE) of θ^* and η^*. In fact, we may still reuse the framework of RMLP-2 by replacing its MLE oracle according to the distribution given by Eq. (38). In the following, we will compare the performances of

1. PwP algorithm with the misspecified linear demand model as Eq. (1), with expanded context $\{x_t\}$'s, and

2. RMLP-2 algorithm on the correct linear valuation model as Eq. (38), with original context $\{x_t\}$'s.

We implement PwP and RMLP-2 on stochastic $\{x_t\}$ sequences (since RMLP-2 has already failed in the adversarial setting) and get numerical results shown as Figure 3. Here we choose $x_0 = [0.5, 0.5]^\top$ and $\mathbf{a} = [0, 1]$. For a model-misspecified online-learning algorithm, there generally exists an $O(\epsilon \cdot T)$ term in the regret rate, where ϵ is a parameter measuring the distance between the global optimal policy and the best *proper* policy (i.e. the best policy in the hypothesis set). However, our numerical results imply that PwP may still achieve a sub-linear regret within a certain time horizon T, whereas the baseline RMLP-2 that takes the correct model has a much worse regret. It is worth mentioning that PwP may still run into $\Omega(T)$ regret as T gets sufficiently large, due to model misspecification. These results imply that

1. Our linear demand model Eq. (1) can be generalized to a linear valuation model as Eq. (38) in practice.

2. Our PwP algorithm can still perform well in model-misspecification settings, and even better than a baseline MLE algorithm with a correct model in a certain period of time.

For the first phenomenon that our demand model can be generalized with context expansion tricks, we may understand it as a Taylor expansion (and we take a linear approximation) at $x_0 = [0.5, 0.5]^\top$. For the second phenomenon that PwP outperforms RMLP-2, it might be caused by the non-convexity of the log-likelihood function of the valuation model specified in Eq. (38). As a result, while RMLP-2 is solving a non-convex MLE and getting estimates far from the true parameters, PwP instead works on an online convex optimization problem within a larger space (which probably contain the underground truth) due to context expansions. Unfortunately, we do not have a rigorous analysis of those two phenomenons.

C. More Discussions
As supplementary to Section 6, here we discuss some potential extensions and impacts of our work with more details.

Assumption on lower-bounding elasticity as $C_\beta > 0$. Here we claim that the regret lower bound should have an $\Omega(\frac{1}{C_\beta})$ dependence on C_β. We prove this by contradiction. Without loss of generality, assume $C_\beta \in (0, 1)$. In specific, we construct a counter example to show it is impossible to have an $O(C_\beta^{-1+\alpha})$ regret for any $\alpha > 0$:

Firstly, let $\beta = C_\beta$. Suppose there exists an algorithm $\mathcal{A}$ that proposes a series of prices $\{p_t\}_{t=1}^T$ which achieve $O(C_\beta^{-1+\alpha})$ regret in any pricing problem instance under our assumptions.

Now, we consider another specific problem setting where $\beta = 1$ while all other quantities $\theta^*, \eta^*, \{x_t\}_{t=1}^T$ stay unchanged. Notice that the reward function has the following property:

$$r(u, \beta, p) = p \cdot S(\beta p - u) = \frac{1}{\beta} \cdot (\beta p) \cdot S(\beta p - u) = \frac{1}{\beta} \cdot r(u, 1, \beta p) \tag{40}$$

Therefore, we construct another algorithm $\mathcal{A}^*$ which proposes $C_\beta \cdot p_t$ at $t = 1, 2, \ldots, T$. According to the $O(C_\beta^{-1+\alpha})$ regret bound of $\mathcal{A}$, we know that $\mathcal{A}^*$ will suffer $C_\beta \cdot O(C_\beta^{-1+\alpha}) = O(C_\beta^\alpha)$ regret. Let $C_\beta \to 0^+$ and observe the regret of $\mathcal{A}^*$ on the latter problem setting (where $\beta = 1$). On the one hand, this is a fixed problem setting with information-theoretic lower regret bound at $\Omega(\log T)$. On the other hand, the regret will be bounded by $\lim_{C_\beta \to 0^+} O(C_\beta^\alpha) = 0$. They are contradictory to each other. Given this, we know that there does not exist such an $\alpha > 0$ such that there exists an algorithm that can achieve $O(C_\beta^{-1+\alpha})$. As a result, it is necessary to lower bound the elasticities by C_β from 0.

Adversarial attacks. Our PwP algorithm achieves near-optimal regret even for adversarial context sequences, while the baseline (modified) RMLP-2 algorithm fails in an oblivious adversary and suffer a linear regret. This is mainly caused by the fact that RMLP-2 takes a pure-exploration step at a *fixed* time series, i.e. $t = 1, 1 + 2, 1 + 2 + 3, \ldots, \frac{k(k+1)}{2}$. This issue might be leveraged by randomizing the position of pure-exploration steps: In each Epoch $k = 1, 2, \ldots$, it may firstly sample one out of all k rounds in this epoch uniformly at random, and then propose a totally random price at this specific round.

However, RMLP-2 still requires $\mathbb{E}[xx^\top] \succeq c \cdot I_d$ even with this trick.

Nonstationarity in Pricing   Although our PwP algorithm is applicable on heteroscedastic valuations, we still benchmark with an optimal fixed pricing policy that knows η^* and θ^* in advance. In reality, customers' valuations and elasticities might fluctuate according to the market environment, causing θ_t^* and η^* different over $t \in [T]$. Existing work such as Leme et al. (2021) and Baby et al. (2022) studies similar settings but assumes i.i.d. noises. It is worth to further investigate the setting when heteroscedasticity and nonstationarity occur simultaneously.

Algorithm and analysis for unknown link function $S(\cdot)$.   Unfortunately, our algorithm is unable to be generalized to the online contextual pricing problem with linear valuation and unknown noise distribution that has been studied by Fan et al. (2021). Indeed, the problem becomes substantially harder when the noise distribution is unknown to the agent. Existing works usually adopt bandits or bandit-like algorithms to tackle that problem. For example, Fan et al. (2021) approaches it with a combination of exploration-first and kernel method (or equivalently, local polynomial), and Luo et al. (2021) uses a UCB-styled algorithm. However, none of them close the regret gap even under the homoscedastic elasticity environment as they assumed, and the known lower bound is at least $\Omega(T^{\frac{2}{3}})$, or $\Omega(T^{\frac{m+1}{2m+1}})$ for smooth ones (Wang et al., 2021b). On the other hand, we study a parametric model, and it is not quite suitable for a bandit algorithm to achieve optimality in regret. In a nutshell, these two problems (known vs unknown noise distributions), although seem similar to each other, are indeed substantially different.

Comparing with Linear Bandits   We claim that no inclusion relationship exists between our setting and linear bandits (Chu et al., 2011), and neither setting's lower bound can imply the other. There are mainly two substantial differences. On the one hand, it is our expected *demand model* (instead of the expected linear *reward* for linear bandits) that falls in a *generalized-linear* model, with a link function bounded by $[0, 1]$. Consequently, the expected *reward* function in our model (i.e., the multiplication of price and demand) is neither linear nor generalized linear. On the other hand, while each action in linear bandits is represented by a vector of dimension d, our action, specifically the price p, is a scalar. This distinction further prevents us from adopting the dependence on d in the linear bandits lower bound to our model.

Linear demand model vs linear valuation model.   In this work, we adopt a generalized linear demand model with Boolean feedback, as assumed in Eq. (1). As we have stated in Appendix B, there exists a heteroscedastic linear valuation model as Eq. (38) that also captures a customer's behavior. However, this linear valuation model is actually harder to learn, as its log-likelihood function is non-convex. It is still an open problem to determine the minimax regret of an online contextual pricing problem with a valuation model like Eq. (38).

Determination of price perturbation.   In this work, we adopt a Rademacher-style perturbation with $\Delta = \frac{1}{\sqrt{T}}$, which optimally balances the exploration cost $T\Delta^2$ against the condition number $\frac{1}{\Delta^2}$ of the surrogate loss function. Notice that this requires a pre-knowledge on T, which may not be available in real-world scenarios. To address this, we apply a "doubling-epoch" trick as suggested by Javanmard & Nazerzadeh (2019). Specifically, we define each epoch k with a length of $\tau_k = 2^k$, and we treat each epoch as an independent pricing problem instance with $T = \tau_k$ and run our PwP algorithm accordingly. This approach allows our algorithm to adapt to an unknown T while preserving the $O(\sqrt{T})$ regret rate. Equivalently, we may set adaptive perturbations as $\Delta_t = (2^{\lfloor \log_2 t \rfloor})^{-\frac{1}{2}}$ in practice. It is reasonable to see a milder perturbation imposed on the greedy price as the time index t goes larger.

Ethic issues.   Since we study a dynamic pricing problem, we have to consider the social impacts that our methodologies and results could have. The major concern in pricing is *fairness*, which attracts increasing research interests in recent years (Cohen et al., 2021; 2022; Chen et al., 2023; Xu et al., 2023). In general, we did not enforce or quantify the fairness of our algorithm. In fact, we might not guarantee an individual fairness since PwP proposes random prices, which means even the same input x_t's would lead to different output prices. Despite the perturbations Δ_t we add to the prices, the pricing model (i.e. the parameters θ^* and η^*) is updating adaptively over time. This indicates that customers arriving later would have relatively fairer prices, since the model is evolving drastically at the beginning rounds and is converging to (local) optimal after a sufficiently long time period. We claim that our PwP algorithm is still fairer than the baseline RMLP-2 algorithm we compare with, since RMLP-2 takes pure explorations at some specific time. As a result, those customers who are given a totally random price would have a either much higher or much lower expected price than those occurring in exploitation rounds. However, it is still worth mentioning that RMLP-2 satisfies individual fairness within each pure-exploitation epoch, since it does not update parameters nor adding noises then.