Dynamic Pricing with Adversarial Covariates

Hanzhao Wang Kalyan Talluri Xiaocheng Li

Imperial College Business School, Imperial College London
{h.wang19, kalyan.talluri, xiaocheng.li}@imperial.ac.uk

Abstract

We revisit the dynamic pricing problem under the generalized linear demand model. Specifically, a seller can dynamically adjust the price of a product over a horizon of T time periods, and at each time period t, the demand of the product is jointly determined by the price and an observable covariate vector $x_t \in \mathbb{R}^d$ through an unknown generalized linear model. Existing works mainly study the case that the covariate vectors x_t's are independently and identically distributed (i.i.d.), while a few exceptions relax this assumption, but either sacrifice model generality or yield sub-optimal regret bounds. We develop a simple pricing algorithm using the principle of optimism in the face of uncertainty and derive an $O(d\sqrt{T}\log T)$ regret upper bound without assuming any statistical structure on the covariates x_t (which can be adversarially chosen). The regret upper bound matches the lower bound (even under the i.i.d. assumption) up to logarithmic factors, thus it shows the unnecessity of (i) the i.i.d. assumption and (ii) the regret's dependency on the (inverse) minimum eigenvalue of x_t's covariance matrix. Furthermore we discuss the condition under which a better regret is achievable and also the applicability of Thompson sampling algorithm. We hope our algorithms and analyses provide new tools for future works on this topic.

1 Introduction

Dynamic pricing, as a central topic in revenue management, has been studied extensively in the fields of operations research and computer science. Among the studies, a stream of works models the demand as a function of covariates or side information, and it gains its popularity from the increasing data availability and the emerging machine learning techniques. The stream is usually named after "dynamic pricing with covariates", "feature-based dynamic pricing", "personalized dynamic pricing", or "dynamic pricing with side information". The covariates (or side information) generally refer to a vector of features that characterize the pricing environment on both macro level (market) and micro level (customer). The problem of dynamic pricing with covariates usually considers a sequential setting where a seller sells a product over a horizon of T time periods. In each time period, upon the observation of covariates $x_t \in \mathbb{R}^d$, the seller decides a price and then observes the realized demand. The seller's objective is to maximize the revenue, and the demand is related to the covariates through an unknown function. Throughout the selling horizon, the seller needs to balance the learning of the demand function with the earning of the revenue, which exemplifies the classic exploration-exploitation tradeoff. The performance of a pricing policy or algorithm is measured by the gap between the expected cumulative revenue obtained by the policy and that of the optimal policy which knows the demand function a priori.

In this paper, we study an adversarial setting where the covariates x_t are adversarially generated and the demand depends on the covariates and the price through a generalized linear model. For the generalized linear model, it covers the two mainstream demand models – binary demand model and linear demand model as special cases (to be elaborated shortly). For the adversarial generation of the covariates, it does not relax the benchmark oracle considered in the dynamic pricing problem with i.i.d. covariates, but rather it aims to question the necessity of the i.i.d. assumption. Technically, such questioning removes the dependency of the inverse minimum eigenvalue $\lambda_{\min}^{-1}$ of the covariance matrix in the regret bound; practically, it is well motivated from the application context where the covariates for demand prediction usually exhibit non-i.i.d. patterns induced by trend, seasonality, promotions, etc. We will return to more discussions on this motivation in Section 1.3.

Paper	Regret Bound	Covariates	Price Coeff.	Demand Model (Key Asm.)
Qiang and Bayati (2016)	$O(d \log T)$	Martingale	Const.	Linear model with a known incumbent price
Cohen et al. (2020)	$O(d^2 \log T)$	Adversarial	$= 1$	Deterministic binary model
Cohen et al. (2020)	$\tilde{O}(d^{19/6} T^{\frac{2}{3}})$	Adversarial	$= 1$	Binary model with sub-Gaussian noise
Mao et al. (2018)	$\tilde{O}(T^{\frac{d}{d+1}})$	Adversarial	$= 1$	Deterministic binary model
Javanmard (2017)	$O(\sqrt{T})$	Adversarial	$= 1$	Binary model with log-concave noise
Javanmard and Nazerzadeh (2019)	$O(d \log T)$	i.i.d	$= 1$	Binary model with log-concave noise
Javanmard and Nazerzadeh (2019)	$\tilde{O}(\log d \sqrt{T})$	i.i.d	Const.	Binary model with log-concave noise
Luo et al. (2021)	$\tilde{O}(dT^{\frac{2}{3}})$	Adversarial	Const.	Binary model with log-concave unknown noise
Xu and Wang (2021)	$O(d \log T)$	Adversarial	$= 1$	Binary model with strictly log-concave noise
Ban and Keskin (2021)	$\tilde{O}(d\sqrt{T})$	i.i.d	Linear	Generalized linear model with sub-Gaussian noise
Ours	$O(d^2 \log^2 T)$	Adversarial	Const.	Generalized linear model with sub-Gaussian noise and known price coefficient
Ours	$\tilde{O}(d\sqrt{T})$	Adversarial	Linear	Generalized linear model with sub-Gaussian noise

Table 1: Summary of Existing Results: The notation $\tilde{O}(\cdot)$ omits the logarithmic factors, d is the dimension of the covariates and T is the horizon length. The column "Covariates" describes the assumption on the generation of the covariates. The column name "Price Coeff." is short for price coeffcient, i.e., the coefficient before the price for the corresponding demand model. The column "Demand Model" summarizes key assumptions for the corresponding paper.

Table 1 summarizes existing works on the problem of dynamic pricing with covariates with respect to regret bound, assumption, and demand model. For the regret bound, we focus on the dependency on the covariates dimension d and the horizon T. The assumption on the covariates consists of two types, i.i.d. and adversarial. Qiang and Bayati (2016) impose a Martingale type condition on the covariates which makes no essential difference to the i.i.d. assumption, because both are to ensure the minimum eigenvalue of the sample covariance matrix bounded away from zero. The price coefficient refers to the coefficient of the price in the demand model. The discussion of the price coefficient cannot be separated from that of the demand model. As mentioned earlier, the demand model refers to how the demand depends on the covariates and the price. There are mainly two models considered in the literature, linear and binary. Ban and Keskin (2021) consider the generalized linear demand model which umbrellas both models as special cases, extending the covariate-free case of the generalized linear demand model in (Broder and Rusmevichientong, 2012). Our paper considers this generalized linear demand model, but differs from Ban and Keskin (2021) in adopting adversarial covariates.

On a high-level, the existing literature on dynamic pricing with covariates can be grouped into two categories:

- Allow adversarial covariates but assume the knowledge of the price coefficient and the distribution of demand shock. In the following two subsections 1.1 and 1.2, we discuss a modeling issue associated with this assumption.
- Assume i.i.d. covariates but unknown price coefficient and unknown distribution of demand shock.

Our work is the first result along this stream that allows adversarial covariates and achieves optimal order of regret. In Subsection 1.3, we discuss the technical and practical advantages of removing the i.i.d. assumption.

1.1 Modeling issue of knowing both price coefficient and shock distribution

We first point out a common modeling issue for a stream of works on the binary demand model including (Mao et al., 2018; Javanmard, 2017; Cohen et al., 2020; Xu and Wang, 2021) and part of (Javanmard and Nazerzadeh, 2019). These works consider the *binary demand model* where the customer purchase behavior is described by a comparison between the customer's utility and the price. They corresponds to the rows with price coefficient $= 1$ in Table 1. Specifically, at time t, the binary demand

$$D_t = \begin{cases} 1, & x_t^\top \beta^* + \epsilon_t \geq p_t \\ 0, & x_t^\top \beta^* + \epsilon_t < p_t \end{cases} \tag{1}$$

where $x_t \in \mathbb{R}^d$ denotes the covariate vector, $\beta^* \in \mathbb{R}^d$ is the linear coefficient, p_t is the price, and ϵ_t models some unobserved utility shock. Under this model, $x_t^\top \beta^* + \epsilon_t$ represents the customer's utility where the term $x_t^\top \beta^*$ captures the part of the utility explained by the covariates x_t.

By assuming the knowledge of price coefficient and shock distribution, it means that (i) the coefficient of p_t is known to be 1 and (ii) ϵ_t's are i.i.d. and follow a known distribution. The above mentioned works study the problem under an online setting where the vector β^* is assumed unknown and the seller strives to learn β^* through observations over time. The generalized linear demand model that we study covers this setting as a special case; however, we point out that there are some modeling issues associated with this special case.

Claim 1. *The optimal pricing policy under the binary demand model* (1) *entails the knowledge of the full distribution of* ϵ_t.

The above claim is due to the heterogeneity of the covariates x_t. To see this, different x_t's may result in different optimal prices which aim to trade-off the purchase probability and the earned revenue under a realized purchase. Therefore, to identify the optimal price for all possible covariates x_t, one needs to know the full distribution of ϵ_t. If we assume the knowledge of the distribution of ϵ_t, then the dynamic pricing problem reduces purely to the learning of β^*, but this suffers from the following two issues.

Ill-posedness of the online setting. Under (1), the assumption on the knowledge of ϵ_t's distribution is usually justified by that it can be learned from history observations. Recall that Claim 1 tells the importance of the precise knowledge of the full distribution. On one hand, if one assumes that the true distribution can be learned precisely from the history data, why can not the coefficient β^* be learned at the same time? In fact, β^* and the distribution of ϵ_t can only be learned simultaneously. That is, any one of these two cannot be learned precisely without the other one learned precisely too. So, there is no point to consider an online setting where β^* is unknown while the distribution of ϵ_t is known. On the opposite end, if the distribution estimated from history observations suffers from an error against the true distribution, the online algorithm will inevitably suffer a linear regret due to the misspecifcation error that it assumes the estimated distribution to be true and only focuses on the learning of β^*.

Unfriendly to new data. Imagine at certain time, the seller has more covariates $y_t \in \mathbb{R}^m$ in addition to x_t. The additional covariates y_t can potentially improve the prediction of customer utility and thus reduce the variance of the demand shock ϵ_t. But as y_t's are new data, there is no history data to tell to what extent y_t's will change the distribution of ϵ_t. Meanwhile, the existing algorithms along this stream all rely on the knowledge of ϵ_t's distribution, so there is no way to start a new online learning procedure to learn the coefficients of y_t's and the new distribution of ϵ_t simultaneously. The same logic applies to the pricing of a newly launched product.

Based on these two points, we argue that for the literature along this line, the discussion of whether the covariates are i.i.d. or adversarially generated is meaningless if one does not first address the modeling issue of price coefficient $= 1$. A few works consider a deterministic binary demand model (for example, (Mao et al., 2018; Cohen et al., 2020)) where there is no ϵ_t term. We believe the deterministic model is even more restrictive in that it is equivalent to an assumption that the customer utility can be fully explained by the observed covariates and there is no additional randomness or unobservable covariates.

1.2 Remedy and more demand models

One way to address the above issue is to assume the distribution of ϵ_t belongs to a parametric family of distributions with some unknown parameter(s). It does not hurt to assume ϵ_t has zero mean because one can achieve this by letting the first dimension of the covariates x_t to be one. Then one can assume the distribution ϵ_t is zero-mean and scale-invariant with variance σ^{-2} for some unknown $\sigma > 0$.

Claim 2. *If one assumes for the model (1), ϵ_t's follow a zero-mean and scale-invariant distribution with variance σ^{-2} for some unknown $\sigma > 0$, then the model (1) is equivalent to the following binary demand model where the distribution of ϵ_t is known but the price coefficient σ is unknown,*

$$D_t = \begin{cases} 1, & x_t^\top \beta^* + \epsilon_t' \geq \sigma p_t \\ 0, & x_t^\top \beta^* + \epsilon_t' < \sigma p_t \end{cases} \tag{2}$$

where ϵ_t' follows some known distribution with mean zero and variance one.

Claim 2 can be easily justified by plugging in $\epsilon_t = \sigma \epsilon_t'$ and absorbing a factor of σ into the coefficient β^*. It implies that if we assume certain parameterized structure for ϵ_t, it is equivalent to assume a known distribution of ϵ_t but an unknown price coefficient. Recall that the previous assumption is on the knowledge of both the price coefficient and the shock distribution. Claim 2 tells that it is equivalent to relax either part, i.e.,

$$\text{known price coeff. } (=1) + \text{unknown variance} \quad \Leftrightarrow \quad \text{unknown price coeff.} + \text{known variance } (=1).$$

We note that it is unnecessary to assume both parts unknown because the thresholding conditions in (1) and (2) are scale-invariant.

From a modeling viewpoint, the unknown parameter σ leaves it open how informative the covariates are in explaining the customer utility. Among the literature, Javanmard and Nazerzadeh (2019) and Luo et al. (2021) consider the model (2). Specifically, Javanmard and Nazerzadeh (2019) focus on the high-dimensional setting and impose the i.i.d. assumption on the covariates; Luo et al. (2021) allow a non-parametric structure for the shock distribution (more general than (2)) and achieve an $\tilde{O}(dT^{2/3})$ regret. Xu and Wang (2021) also mention the model (2) and leave the achievability of $\sqrt{T}$ regret as an open question. The idea behind the generalized linear model pursues further along this stream and replaces the price coefficient σ in (2) with a linear function of the covariates x_t. Thus it can be viewed as a further generalization of the model (2). The rationale for such generalization is, the extent to which the covariates x_t explain the customer utility can be dependent not only on some unknown constant σ but also on x_t, and the concrete dependency is unknown. Given that we work on this generalized linear demand model, our result resolves the open question in (Xu and Wang, 2021).

Linear Demand Model. Another prevalent demand model is the *linear demand model*, where it assumes

$$D_t = x_t^\top \beta^* + x_t^\top \gamma^* \cdot p_t + \epsilon_t \tag{3}$$

where β^* and γ^* are fixed unknown parameter vectors. The demand shock ϵ_t's are mean-zero random variables, and it makes no essential difference to adapt the distribution of ϵ_t to the history up to time t. We remark that throughout the paper the specification of ϵ_t's distribution should depend on the context (demand model); for simplicity, we do not distinguish and use ϵ_t to represent the randomness in all the demand models.

Claim 3. *Under the linear demand model (3), if the demand shock ϵ_t is mean-zero and sub-Gaussian, the optimal pricing strategy is not affected by the specific distribution of ϵ_t.*

The above claim states that under the linear demand model, the distribution of ϵ_t makes no difference to the optimal pricing strategy as long as it is mean-zero and sub-Gaussian. This explains why papers on the linear demand model, unlike the case of the binary demand model, do not discuss the knowledge of the distribution ϵ_t.

The linear demand model has been extensively studied in operations research literature but mostly under the covariate-free setting, i.e., $x_t^\top \beta^* = a$ and $x_t^\top \gamma^* = b$ for two unknown constants a and b (See, for example den Boer and Zwart (2014); Keskin and Zeevi (2014, 2018)). These papers mainly focus on understanding why certainty-equivalent pricing policy fails and propose algorithms that ensure enough price dispersion so as to learn the parameters (a, b). A line of subsequent works are inspired by the analyses developed in these papers and consider the presence of the covariates, including (Qiang and Bayati, 2016; Javanmard and Nazerzadeh, 2019; Ban and Keskin, 2021; Bastani et al., 2021), and a common assumption made in these works is that the covariates are i.i.d. generated. Note that the generalized linear demand model covers the linear demand model as a special case. Our work contributes to this stream in relaxing the i.i.d. assumption and removing the regret's dependency on the inverse minimum eigenvalue $\lambda_{\min}^{-1}$ of the covariance matrix.

1.3 Practical motivation for adversarial covariates

In the following, we argue from three aspects why the i.i.d. covariates assumption may not be practical for demand prediction.

- Network/peer effect: Several studies show the peer effects are significant for a wide range of products' demand (Seiler et al., 2017; Goolsbee and Klenow, 2002; Bailey et al., 2019); (Nasr and Elshar, 2018) use Markov Chain to model such relationships between consecutive customers. In addition, Baardman et al. (2020) show that the consideration of the tendency among customers can improve the demand prediction accuracy with real data evidence from online retailers. Thus, the customers' features are more likely to exhibit short-term dependency.

- Seasonality and life cycle of product: Seasonality and life cycle patterns are often found in demand prediction practices (Neale and Willems, 2009). We can imagine that for a new geek product, for example, VR glasses, customers at the initial phase should be more familiar with this kind of product or be fun of this brand and after getting more feedback, there will be some risk-neutral customers. Further, for laptop customers, the proportional of students may increase during the back to school season.

- Competitors: The influence from competitors are usually considered by sellers when predicting demands Armstrong et al. (2005). For example, it is shown that Airbnb has significant influences on hotels (Zervas et al., 2017). However, the effect from competitors is usually complicated and hard to measure and the customer features are not identically distributed in this case.

Moreover, past demand data has been very useful in predicting the future demand (Ferreira et al., 2016; Ban and Keskin, 2021; Chen et al., 2021). For example, when we want to predict the demand on Day $t+1$, the customer demand on Day t, Day $t-1$, Day $t-6$ may turn out to be important features in the prediction model. The inclusion of such feature may not only violate the i.i.d. assumption, but also result in a collinearity between the features, and consequently a small minimum eigenvalue of the covariate matrix. The practical implication of our result is that the seller does not need to worry about such effect and can proceed with using the covariates even if they are adversarially generated.

1.4 Implications for generic dynamic pricing problems

The setting of dynamic pricing with covariates recovers that of dynamic pricing without covariate (or covariate-free) as a special case by letting $x_t = 1$ for all $t = 1, ..., T$. Thus our algorithms also apply to the covariate-free dynamic pricing problem. Early works on this topic (Broder and Rusmevichientong, 2012; Keskin and Zeevi, 2014; den Boer and Zwart, 2014) have extensively discussed the necessity of price exploration and devised algorithms that conduct "hard" exploration such as adding random price perturbations or setting random prices. Our results provide a generic analysis on how the "soft" exploration methods such as UCB and Thompson sampling (TS) can be applied to the dynamic pricing problem. In Section 3 and Section 4.1, we discuss the UCB pricing algorithm and the TS pricing algorithm, respectively.

Moreover, in Section 4.2, we identify the knowledge of the price coefficient as a simple sufficient condition for the achievability of $O(\log T)$ regret under the generalized linear demand model. The result generally explains why in Table 1, $O(\log T)$ regret is achieved for some papers while $\Omega(\sqrt{T})$ regret is inevitable for others. The knowledge of the price coefficient is comparable to the knowledge of one incumbent price in (Keskin and Zeevi, 2014; Qiang and Bayati, 2016) or the well-separatedness condition in (Broder and Rusmevichientong, 2012). Under such conditions, there is no need for price exploration, and the so-called certainty-equivalent policy will provide an optimal rate of regret.

1.5 Other related literature

Our algorithm can be viewed as the vanilla version of the upper confidence bound (UCB) algorithm for the dynamic pricing problem. The alternative algorithm of Thompson sampling has also been explored under the dynamic pricing problem (Ferreira et al., 2018; Bastani et al., 2021). Ferreira et al. (2018) consider a constrained version of the problem, also known as the price-based network revenue management problem, where there are finitely many allowable prices and a number of resource constraints. Bastani et al. (2021) considers the transfer learning of the demand model across different sources of data.

We note that Javanmard and Nazerzadeh (2019); Ban and Keskin (2021) also feature for a high-dimensional setting and perform a variable selection subroutine in their pricing algorithm via L_1 regularization. The adversarial covariates in our paper are cast under a low dimensional setting and cannot be dealt under the high-dimensional setting. To this end, we believe the i.i.d. structure and the minimum eigenvalue assumption in both papers are necessary for the high-dimensional setting. The usage of covariates is also considered for assortment problem in (Chen et al., 2021). For more on the history and origin of the dynamic pricing problem, especially the covariate-free case, we refer the readers to the review paper Den Boer (2015).

2 Model and Performance Measure

We consider the generalized linear demand model from Ban and Keskin (2021)

$$D_t = g(x_t^\top \beta^* + x_t^\top \gamma^* \cdot p_t) + \epsilon_t \quad \forall t = 1, 2..., T, \tag{4}$$

where $\beta^*, \gamma^* \in \mathbb{R}^d$ are true unknown parameters and $g(\cdot)$ is a known function, x_t is observable customer covariate vector and ϵ_t is the unobservable and idiosyncratic demand shock of the customer arriving in period t. We use $\mathcal{X}$ to denote the domain of x_t. Here, we mean by observable that the seller knows x_t before setting the price p_t. Denote $\theta^* := (\beta^*; \gamma^*)$ as the concatenated parameter vector.

We denote the expected revenue function under price p, parameter θ, and covariates x_t as

$$r(p; \theta, x) := p \cdot g(x_t^\top \beta + x_t^\top \gamma \cdot p),$$

and denote the optimal expected revenue function as

$$r^*(\theta, x) := \max_{p \geq 0} \; p \cdot g(x_t^\top \beta + x_t^\top \gamma \cdot p),$$

with the optimal pricing function

$$p^*(\theta, x) = \arg\max_{p \geq 0} \; p \cdot g(x_t^\top \beta + x_t^\top \gamma \cdot p).$$

Throughout the paper, we assume the one-dimensional price optimization problem $r^*(\theta, x)$ can be efficiently solved for any given parameter θ and covariates x.

Assumption 1 (Boundedness)**.** *We assume*

(a) There exists $\bar{\theta} > 0$ such that $\Theta = \{\theta \in \mathbb{R}^{2d} : \|\theta\|_2 \leq \bar{\theta}\}$ and $\theta^* \in \Theta$.

(b) The seller is allowed with a price range $[\underline{p}, \bar{p}]$ under all possible x and θ^*. We assume that $p^*(\theta, x)$ is in the interior of the feasible set $[\underline{p}, \bar{p}]$ for all $\theta \in \Theta$ and $x \in \mathcal{X}$.

(c) For all $x \in \mathcal{X}$ and $p \in [\underline{p}, \bar{p}]$, $\|(x, px)\|_2 \leq 1$.

The boundedness assumption in above is standard in the dynamic pricing literature and it is also well grounded in the practical application context. We note that if the raw covariates do not satisfy the last part, one can always perform some normalization for the covariates to meet the condition.

Assumption 2 (Properties of $g(\cdot)$)**.** *Assume $g(\cdot)$ is strictly increasing and differentiable, with bounded derivative over its domain. Specifically, there exist constants $\underline{g}, \bar{g} \in \mathbb{R}$ such that $0 < \underline{g} \leq g'(z) \leq \bar{g} < \infty$ for all $z = x^\top \beta + x^\top \gamma \cdot p$ where $\|x\|_2 \leq 1$, $\theta \in \Theta$, and $p \in [\underline{p}, \bar{p}]$.*

Assumption 3 (Demand Shock)**.** *Assume $\{\epsilon_t, t = 1, 2...\}$ is a $\bar{\sigma}^2$-sub-Gaussian martingale difference, i.e.,*

$$\mathbb{E}[\epsilon_t | \mathcal{H}_{t-1}] = 0 \quad \text{and} \quad \log\left(\mathbb{E}\left[\exp(s\epsilon_t) \middle| \mathcal{H}_{t-1}\right]\right) \leq \frac{\bar{\sigma}^2 s^2}{2}$$

for all $s \in \mathbb{R}$, where $\mathcal{H}_t := \sigma(p_1, ..., p_t, \epsilon_1, ..., \epsilon_t, x_1, ..., x_t, x_{t+1})$ and $\mathcal{H}_0 = \sigma(\emptyset, \Omega)$. Moreover, we assume $\bar{\sigma}^2$ is known a priori.

Assumption 2 and Assumption 3 concern the function g and the demand shock term, respectively. For Assumption 2, we note that the filtration definition includes the covariates at time $t+1$. This small change allows the demand shock ϵ_t to be dependent on the covariates x_t and hence gives us more modeling flexibility. For Assumption 3, it covers common distributions such as Normal and random variables with bounded support. The sub-Gaussian parameter $\bar{\sigma}^2$ is an upper bound proxy for the true variance of the random variable, which can be easily obtained for a bounded random variable. We remark that the demand shock ϵ_t should be understood under a specific context, and it may follow different classes of distributions, for example, under the models (2), (3) and (4).

In the following we show how the general model (4) recovers the binary demand model (2) and the linear demand model (3) as special cases.

Example 1 (Binary Demand Model)**.** *As noted in the previous section, the binary demand model (2) (also (1)) is considered by a number of papers as in Table 1 (denoted by binary model in the "Demand Model" column). We first restrict the first dimension of x_t to be always 1. To recover (2), we can then set the function $g(\cdot)$ in (4) to be the cumulative distribution function of $-\epsilon_t$, $\gamma^* = (-\sigma, 0, ..., 0)^\top$, and*

$$\epsilon_t = \begin{cases} 1 - g\left(x_t^\top \beta - \sigma p_t\right) & w.p.\ g\left(x_t^\top \beta - \sigma p_t\right), \\ -g\left(x_t^\top \beta - \sigma p_t\right) & w.p.\ 1 - g\left(x_t^\top \beta - \sigma p_t\right). \end{cases}$$

Thus it becomes the binary demand model (2). As a notational remark, we note ϵ_t in above should be understood under the model (4). The parameter σ is an unknown parameter that represents the price coefficient or describes the variance of the utility shock in (2); the sub-Gaussian parameter $\bar{\sigma}^2$ can be easily chosen as $1/4$ from the boundedness of ϵ_t.

Example 2 (Linear Demand Model)**.** *The linear demand model (3) can be easily recovered from (4) by letting the function $g(\cdot)$ be an identity function. As in Example 1, we restrict the first dimension of x_t to be always 1. Then, the model in (Qiang and Bayati, 2016) can be recovered by setting $\gamma^* = (b, 0, ..., 0)$; the covariate-free linear demand model in (den Boer and Zwart, 2014; Keskin and Zeevi, 2014) can be recovered by setting $\beta^* = (a, 0, ..., 0)$ and $\gamma^* = (b, 0, ..., 0)$.*

Performance Measure. Now, we define *regret* as the performance measure for the problem. Specifically,

$$\text{Reg}_T^\pi(x_1, ..., x_T) := \sum_{t=1}^{T} r^*(\theta^*, x_t) - \mathbb{E}\left[\sum_{t=1}^{T} r_t\right]$$

where θ^* denotes the true parameter vector, π denotes the policy/algorithm, r_t is the revenue obtained at time t under π, and the expectation is taken with respect to the demand shock ϵ_t's. The benchmark oracle (the first summation in above) is defined based on the optimal revenue function $r^*(\theta, x)$. It assumes the knowledge of the true θ^* but does not observe the realization of the demand shock ϵ_t when setting the price. In defining the regret, we allow the covariates x_t's to be adversarially generated. Therefore, no expectation is taken for x_t's in the regret definition, and we seek for a worst-case regret upper bound over all possible x_t's. For the case when x_t's are i.i.d., the regret definition involves one more layer of expectation on x_t's. Thus our regret bound is strictly stronger and can directly translate into a regret bound for the i.i.d. case.

3 UCB-Based Pricing

In this section, we introduce our first generic algorithm for the problem of dynamic pricing with adversarial covariates. Since we do not assume the knowledge of the distribution ϵ_t's, we adopt the quasi-maximum likelihood estimation (MLE) to learn the parameter θ. In the following, we first introduce some analytical results for the quasi-MLE problem and then present our algorithm.

3.1 Regularized Quasi-Maximum Likelihood Estimation

For the dynamic pricing problem, we define the *misspecified likelihood function* for the t-th observation as follows

$$l_t(\theta) := -\int_{D_t}^{g(z_t^\top \theta)} \frac{1}{h(u)}(u - D_t)\mathrm{d}u, \tag{5}$$

where $\theta = (\beta, \gamma)$ encapsulates the parameters, $z_t = (x_t, p_t x_t)$ is a column vector by concatenating the covariates, and $h(u) = g'(g^{-1}(u))$ for $u \in \mathbb{R}$. To make some intuitions for this function, we can first interpret the term $1/h(u)$ as a positive constant due to the monotonicity of the function g. Then the function l_t as a proxy for the true likelihood function aims to find a θ that minimizes the gap between $g(z_t^\top \theta)$ and D_t. This can be seen from that when one maximizes l_t over θ, it achieves its maximum when $g(z_t^\top \theta) = D_t$ (if possible). The likelihood function l_t takes the same form as the one in (Ban and Keskin, 2021), but the following analysis differs from the analysis therein and is more aligned with the analyses in the bandits literature (Abbasi-Yadkori et al., 2011; Filippi et al., 2010).

Based on l_t, we define the regularized quasi-likelihood estimator with parameter λ as:

$$\hat{\theta}_t := \arg\max_{\theta \in \Theta} -\lambda \underline{g} \|\theta\|_2^2 + \sum_{\tau=1}^{t} l_\tau(\theta), \tag{6}$$

where $\underline{g}$ denotes the lower bound of $g'(\cdot)$ as defined in Assumption 2. The estimator $\hat{\theta}_t$ will be used throughout the paper. The motivation for the regularization term is to overcome the singularity caused by the adversarial covariates and to ensure a curvature for the likelihood function.

Now we analyze the property of the estimator. The gradient and Hessian of l_t are

$$\nabla l_t(\theta) = \xi_t(\theta) z_t, \quad \nabla^2 l_t(\theta) = -\eta_t(\theta) z_t z_t^\top,$$

where $\xi_t(\theta) := D_t - g\left(z_t^\top \theta\right), \eta_t(\theta) := g'\left(z_t^\top \theta\right)$. The concise form of the gradient and Hessian justifies the choice of $h(u)$ in (5).

The following lemma states that under a non-anticipatory pricing policy/algorithm, the sequence of $\{\xi_t(\theta^*)\}_{t=1}^T$ is a martingale difference sequence adapted to history observations with zero-mean $\bar{\sigma}^2$-sub-Gaussian increments.

Lemma 1. *For $t = 1, ..., T$, we have*

$$\mathbb{E}\left[\xi_t(\theta^*) | \mathcal{H}_{t-1}\right] = 0.$$

In addition, $\xi_t(\theta^*)|\mathcal{H}_{t-1}$ is $\bar{\sigma}^2$-sub-Gaussian.

Proof. Note that

$$\mathbb{E}\left[\xi_t(\theta^*) | \mathcal{H}_{t-1}\right] = \mathbb{E}\left[\epsilon_t | \mathcal{H}_{t-1}\right] = 0.$$

Both the last part and the sub-Gaussianity come from Assumption 3. □

Let the (cumulative) score function

$$S_t := \sum_{\tau=1}^{t} \frac{\xi_\tau(\theta^*)}{\bar{\sigma}} z_\tau,$$

and define the (cumulative) design matrix

$$M_t := \lambda I_{2d} + \sum_{\tau=1}^{t} z_\tau z_\tau^\top$$

where I_{2d} is a $2d$-dimensional identity matrix.

The following theorem measures S_t's deviation in terms of a metric induced by M_t. It can be easily proved by an application of the martingale maximal inequality on the sequence of $\xi_t(\theta^*)$. Throughout the paper, we define the M-norm of a vector z,

$$\|z\|_M \coloneqq \sqrt{z^\top M z}$$

for a positive definite matrix M. We use $\det M$ to denote the determinant of a matrix M.

Theorem 1 (Theorem 20.4, Lattimore and Szepesvári (2020))**.** *For any regularization parameter $\lambda > 0$ and $\delta \in (0, 1)$,*

$$\mathbb{P}\left(\exists t \in \{1..., T\} : \|S_t\|^2_{M_t^{-1}} \geq 2\log\left(\frac{1}{\delta}\right) + \log\left(\frac{\det M_t}{\lambda^d}\right)\right) \leq \delta.$$

An implication of the theorem is to produce the following bound on the estimation error of $\hat{\theta}_t$. The proof is a combination of the curvature analysis of the objective function (6) and the above theorem. In fact, Proposition 1 is a key ingredient in removing the statistical assumption on the covariates x_t. Specifically, when we impose some i.i.d. assumption on x_t and assume the minimum eigenvalue of its covariance matrix is bounded away from zero, then we can upper bound the estimation error of θ^* in the Euclidean norm. When we do not impose such assumptions, the following proposition tells that we can still obtain an estimation error bound by measuring the distance according to the sampled design matrix M_t. As noted earlier, the technique is often seen in the linear bandits literature where we usually assume the features associated with the actions arrive in an adversarial manner (Abbasi-Yadkori et al., 2011; Filippi et al., 2010).

Proposition 1. *For any regularization parameter $\lambda > 0$, the following bound holds*

$$\mathbb{P}\left(\exists t \in \{1, ..., T\} : \left\|\hat{\theta}_t - \theta^*\right\|_{M_t} \geq 2\sqrt{\lambda}\bar{\theta} + \frac{2\bar{\sigma}}{\underline{g}}\sqrt{2\log\left(\frac{1}{\delta}\right) + \log\left(\frac{\det M_t}{\lambda^{2d}}\right)}\right) \leq \delta.$$

for any $\delta \in (0, 1)$.

The proof of Proposition 1 is deferred to the Appendix. It follows a standard analysis of the loss function. Intuitively, if we do Taylor expansion for the objective function (6), the first-order term will be characterized by Theorem 1, while the second-order term is controlled jointly by the matrix M_t and the properties of the function g given in Assumption 2.

We choose $\delta = \frac{1}{T}$ in Proposition 1 and define the function

$$\alpha(M) \coloneqq 2\sqrt{\lambda}\bar{\theta} + \frac{2\bar{\sigma}}{\underline{g}}\sqrt{2\log T + \log\left(\frac{\det M}{\lambda^d}\right)}.$$

Then we obtain the following corollary which gives us a small-probability confidence bound for the estimator $\hat{\theta}_t$. The first part of the corollary prescribes the confidence interval in our following algorithms, and the second part provides a uniform upper bound of $\alpha(M_t)$'s. Specifically, the first part is obtained by plugging $\delta = \frac{1}{T}$ into Proposition 1. And given that $\|x_t\|_2^2 \leq 1$ by assumption, we can apply Lemma 19.4 of Lattimore and Szepesvári (2020) (purely algebraic analysis with no stochasticity) and obtain the second part.

Corollary 1. *For all $\lambda > 0$,*

$$\mathbb{P}\left(\exists t \in \{1, ..., T\} : \left\|\hat{\theta}_t - \theta^*\right\|_{M_t} \geq \alpha(M_t)\right) \leq \frac{1}{T}.$$

Moreover, for all $t = 1, ..., T$,

$$\alpha(M_t) \leq \bar{\alpha} := 2\sqrt{\lambda}\bar{\theta} + \frac{2\bar{\sigma}}{\underline{g}}\sqrt{2\log(T) + 2d\log\left(\frac{2d\lambda + T}{2d\lambda}\right)}.$$

Now we complete the discussion of the quasi-MLE and we proceed to the algorithm and regret analysis in the following.

3.2 Algorithm and Regret Analysis

Algorithm 1 describes a UCB-based pricing algorithm. At each time t, the algorithm first constructs an estimator and the corresponding confidence bound based on the quasi-MLE introduced earlier. Then the algorithm finds the most "optimistic" parameter within the confidence bound and sets the price pretending this parameter to be the true. It is a standard exemplification of the idea of upper confidence bound (UCB), also known as the principle of optimism in the face of uncertainty.

Algorithm 1 UCB Pricing

Input: Regularization parameter λ.

for $t = 1, ..., T$ **do**

Compute the estimators $\hat{\theta}_{t-1}$ by (6) and its confidence interval

$$\Theta_t := \left\{\theta \in \Theta : \left\|\theta - \hat{\theta}_{t-1}\right\|_{M_{t-1}} \leq \alpha(M_{t-1})\right\}.$$

Observe feature x_t and choose the UCB parameter which maximizes the expected revenue:

$$\theta_t := (\beta_t, \gamma_t) = \arg\max_{\theta \in \Theta_t} \quad r^*(\theta, x_t) \tag{7}$$

Set the price by

$$p_t = p^*(\theta_t, x_t)$$

end for

While many existing dynamic pricing algorithms more or less utilize the special model structure for designing algorithms, Algorithm 1 is very simple yet generally applicable. In this light, we hope the algorithm and its analysis can work as a prototype for future study on this topic. Theorem 2 states the regret bound of Algorithm 1. For dimension d and horizon T, it meets the lower bound of the problem (See (Ban and Keskin, 2021)) up to logarithmic factors. Compared to the existing bounds (Qiang and Bayati, 2016; Javanmard and Nazerzadeh, 2019; Ban and Keskin, 2021), our bound does not involve the term $\lambda_{\min}^{-1}$ where $\lambda_{\min}$ represents the minimum eigenvalue of the covariance matrix of X_t. For other parameters like $\bar{p}$ and $\bar{\theta}$, we are unclear about whether their dependencies are optimal. They also appear in the existing regret bounds under the i.i.d. setting (Javanmard and Nazerzadeh, 2019; Ban and Keskin, 2021). Under the linear demand model, the parameter $\bar{\theta}$ is on the same magnitude with the demand D_t so it can be validly treated as a constant. But for binary linear demand model, whether the parameter $\bar{\theta}$ has some implicit dependency on d is contingent on the distribution of the utility shock in (2).

For Theorem 2, the proof idea is very intuitive. As the algorithm represents the confidence interval based on the matrix M_{t-1}, the current observation x_t will either induce a small single step regret or reduce the confidence interval significantly. Then we upper bound the regret of the algorithm to a summation sequence involving the covariates x_t's and matrices M_t's and then employ the elliptical potential lemma (as follows) to conclude the proof.

Lemma 2 (Elliptical Potential Lemma, (Lai and Wei, 1982)). *For any constant $\lambda \geq 1$ and sequence of $\{x_t\}_{t\geq 1}$ with $\|x_t\|_2 \leq 1$ for all $t \geq 1$ and $x_t \in \mathbb{R}^d$, define the sequence of covariance matrices:*

$$\Sigma_0 := \lambda I_d, \quad \Sigma_t := \lambda I_d + \sum_{\tau=1}^{t} x_\tau x_\tau^\top \quad \forall t \geq 1,$$

where I_d is the identity matrix with dimension d. Then for any $T \geq 1$, the following inequality holds

$$\sum_{t=1}^{T} \|x_t\|^2_{\Sigma_{t-1}^{-1}} \leq 2d \log\left(\frac{\lambda d + T}{\lambda d}\right).$$

Theorem 2. *Under Assumptions 1, 2 and 3, with any sequence $\{x_t\}_{t=1,\ldots,T}$, if we choose the regularization parameter $\lambda = 1$, the regret of Algorithm 1 is upper bounded by*

$$4\bar{p}\bar{g}\bar{\alpha}\sqrt{Td\log\left(\frac{2d+T}{2d}\right)} + \bar{p} = O(d\sqrt{T}\log T)$$

where $\bar{\alpha} = 2\bar{\theta} + \frac{2\bar{\sigma}}{\underline{g}}\sqrt{2\log T + 2d\log\left(\frac{2d+T}{2d}\right)}$ defined in Corollary 1 represents an upper bound for the confidence volume.

3.3 Discussion

Analysis-wise, the regret derivation of Algorithm 1 largely mimics the analysis of bandits problem with a generalized linear dependence (Filippi et al., 2010). The dynamic pricing with covariates problem differs in that (i) the action space becomes infinite and changes over time; (ii) there is a slight misalignment between the objective (reward) and the observation (demand). Specifically, the pricing decision at time t, if viewed as an action for the bandits problem, is a line segment prescribed jointly by the covariate x_t and the allowable price p_t. The analysis here draws a connection between the dynamic pricing and the bandits problem, and also provides an alternative route for the existing analyses on the dynamic pricing with covariates problem (See (Qiang and Bayati, 2016; Ban and Keskin, 2021; Zhu and Zheng, 2020) among others). As noted earlier, this new analysis relaxes the i.i.d. assumption and removes the dependency on the inverse minimum eigenvalue of the covariance matrix.

As for the computational aspect, the dynamic pricing problem has one extra layer of optimization problem (8) than a finite-arm bandits problem. Specifically, Algorithm 1 has two optimization problems to solve: (i) the quasi-MLE problem (6); (ii) the UCB optimization problem (8). In the following, we discuss these two problems separately.

Quasi-MLE problem (6)

In general, the quasi-MLE problem cannot be solved in closed-form. In practice, the problem (6) has to be solved by some optimization algorithm, and thus the solution output from the optimization algorithm might not be the exact optimal solution. We show that the pricing algorithm can also be adapted to the case of an approximate optimal solution. Algorithm 2 describes such an adaption where at each time t, an approximate solution $\check{\theta}_t$ is used to construct the confidence interval. The approximation gap Δ_t is formally defined as follows and it can be obtained by monitoring the dual optimization problem,

$$\Delta_t := \sum_{\tau=1}^{t} l_\tau(\hat{\theta}_t) - \lambda\underline{g}\|\hat{\theta}_t\|_2^2 - \sum_{\tau=1}^{t} l_\tau(\check{\theta}_t) + \lambda\underline{g}\|\check{\theta}_t\|_2^2.$$

To account for this approximation gap, the confidence bound needs to be enlarged accordingly such that it covers the true parameter θ^*.

Algorithm 2 UCB Pricing with Approximation

Input: Regularization parameter λ.
for $t = 1, ..., T$ **do**
Compute the estimators $\check{\theta}_{t-1}$ by (6) with approximation gap Δ_{t-1} and its confidence interval

$$\Theta_t := \left\{ \theta \in \Theta : \left\| \theta - \check{\theta}_{t-1} \right\|_{M_{t-1}} \leq \alpha(M_{t-1}) + \sqrt{\frac{2}{\underline{g}} \Delta_{t-1}} \right\}.$$

Observe feature x_t and choose the UCB parameter which maximizes the expected revenue:

$$\theta_t := (\beta_t, \gamma_t) = \underset{\theta \in \Theta_t}{\arg\max} \quad r^* (\theta, x_t) \tag{8}$$

Set the price by

$$p_t = p^* (\theta_t, x_t)$$

end for

Theorem 3. *Under Assumptions 1, 2 and 3, with any sequence $\{x_t\}_{t=1,...,T}$, if we choose the regularization parameter $\lambda = 1$, the regret of Algorithm 2 is upper bounded by*

$$4\bar{p}\bar{g} \left(\bar{\alpha} \sqrt{Td \log \left(\frac{2d + T}{2d} \right)} + \frac{\sqrt{2}}{2\sqrt{\underline{g}}} \sum_{t=1}^{T} \sqrt{\Delta_{t-1}} \mathbb{E} \left[\|z_t\|_{M_{t-1}^{-1}} \right] \right) + \bar{p}.$$

Moreover, if $\Delta_{t-1} \leq \bar{\Delta}$ for all $t = 1, ..., T$, the regret upper bound becomes

$$4\bar{p}\bar{g} \left(\bar{\alpha} + \sqrt{\frac{2\bar{\Delta}}{\underline{g}}} \right) \sqrt{Td \log \left(\frac{2d + T}{2d} \right)} + \bar{p}.$$

Theorem 3 provides a regret upper bound for Algorithm 2. Given that the objective function (6) is unnormalized, the uniform bound $\bar{\Delta}$ can be achieved through linear steps of gradient descent or stochastic gradient descent. This explains the application of online optimization techniques for the dynamic pricing problem in (Javanmard, 2017; Xu and Wang, 2021). The proof idea of Theorem 3 is almost identical to that of Theorem 2 by using the following lemma captures the distance between the approximate solution $\check{\theta}_t$ and the optimal solution $\hat{\theta}_t$. The lemma can be implied from the optimality condition.

Lemma 3. *Recall that $\hat{\theta}_t$ is the optimal solution to the optimization problem (6). For any $\theta \in \Theta$, we have*

$$\sum_{\tau=1}^{t} l_\tau(\hat{\theta}_t) - \lambda \underline{g} \|\hat{\theta}_t\|_2^2 - \sum_{\tau=1}^{t} l_\tau(\theta) + \lambda \underline{g} \|\theta\|_2^2 \geq \frac{1}{2} \underline{g} \|\hat{\theta}_t - \theta\|_{M_t}^2.$$

The lemma justifies the choice of the confidence bound in Algorithm 2 and it ensures that the confidence bound covers the true parameter with high probability.

UCB optimization problem (8)

We note that in general the optimization subroutine (8) may be non-convex and hard to solve. In fact, computation efficiency is a common issue for UCB algorithms. For example, the action selection step can be NP-hard for the linear bandit problem (Dani et al., 2008). The following two examples show that the subroutine could be solved efficiently under some special cases.

Example 3 (Covariate-free linear demand function)**.** *Consider the demand* $D_t = a + bp_t + \epsilon_t$*, where* $a > 0, b < 0$ *are unknown real-valued parameters. Then* (8) *becomes an optimization problem with concave objective function* $-\frac{a^2}{4b}$ *subject to a quadratic constraint.*

Example 4 (Binary demand observation with constant price coefficient)**.** *Consider the binary demand model* (2) *where the parameters* β *and* σ *are unknown. For each fixed value of* σ*, it is easy to verify that the revenue function is increasing with respect to* $x^\top \beta$*; hence the optimization problem* (8) *reduces to quadratically constrained linear program (QCLP). Then we can discretize the domain of* σ *into a number of possible candidates, and solve a QCLP for each candidate. Lastly, we pick the pair of* β *and* σ *that outputs the largest objective value.*

While these specific examples rely on the structure of the problem, one general solution to compute the UCB optimization problem is through Monte Carlo method. Note that for each time t, the confidence interval is an ellipsoid, so random sampling from the confidence interval can be done efficiently. Specifically, we randomly generated K samples from the uniform distribution over the confidence interval and we denote the samples as $\tilde{\theta}_k$'s. Then the optimization subroutine is solved by

$$\theta_t = \arg\max_{k=1,\dots,K} r^*(\tilde{\theta}_k, x_t).$$

In the numerical experiments, we will try out different values for K and examine the effect of sample size on the algorithm performance.

4 Towards More Efficient Computation and Better Regret

4.1 Thompson Sampling

Now we apply the method of Thompson sampling as a more efficient implementation of the algorithms in the previous section. Previous work (Ferreira et al., 2016) has discussed the possibility of applying Thompson sampling for revenue management where the authors studied on the covariate-free case and analyzed the constrained dynamic pricing/revenue management problem. Our result focuses on the handling of covariates and analyzes the problem under a frequentist rather than Bayesian setting.

In Algorithm 3, we first compute the quasi-MLE estimator $\hat{\theta}_{t-1}$ just as the previous algorithms. Then the algorithm samples a multivariate Gaussian vector ξ_t and uses ξ_t to generate a randomized estimator $\tilde{\theta}_{t-1}$. This sampling step can be viewed as an efficient substitute of the UCB optimization step (8) in Algorithm 1. Intuitively, the design matrix M_{t-1} represents the confidence bound for the current estimation of θ^* based on the past observations. In the algorithm, the matrix M_{t-1} twists the Gaussian vector ξ_t so as to encourage a random exploration in the direction that the current estimator is less confident. Then the algorithm pretends the sampled parameter $\tilde{\theta}_{t-1}$ as true and uses it to set the price p_t. The algorithm design and analysis largely follow the analysis of Thompson sampling for linear bandits by Abeille and Lazaric (2017).

Algorithm 3 Thompson Sampling with Covariates

Input: Regularization parameter λ.

for $t = 1, ..., T$ **do**

Compute the estimator $\hat{\theta}_{t-1}$ by (6) and observe covariates x_t.

Sample $\xi_t \sim \mathcal{N}(0, I_{2d})$ and compute the parameter

$$\tilde{\theta}_{t-1} := \hat{\theta}_{t-1} + \alpha(M_{t-1})M_{t-1}^{-1/2}\xi_t.$$

Set the price by

$$p_t = \arg\max_{p\in[\underline{p},\bar{p}]} \; r\left(p; \tilde{\theta}_{t-1}, x_t\right)$$

end for

Assumption 4 (Properties of $g(\cdot)$). *Let*

$$\tilde{\Theta} := \{\theta \in \mathbb{R}^{2d} : \|\theta - \tilde{\theta}\|_2 \le 2\bar{\alpha}\sqrt{d\log(4dT^2)} \text{ for some } \tilde{\theta} \in \Theta\}$$

where Θ is defined in Assumption 1. We assume $g(z)$ is strictly increasing, differentiable, convex and there exist constants $\underline{g}, \bar{g} \in \mathbb{R}$ such that $0 < \underline{g} \le g'(z) \le \bar{g} < \infty$ for all $z = x^\top\beta + x^\top\gamma \cdot p$ where $\|x\|_2 \le 1$, $\theta = (\beta, \gamma) \in \tilde{\Theta}$, and $p \in [\underline{p}, \bar{p}]$.

Assumption 4 is a stronger version of Assumption 2 on the properties of g. Specifically, the domain of parameter θ is enlarged from Θ to $\tilde{\Theta}$ so as to cover (with high probability) the randomized sampled parameters $\tilde{\theta}_t$'s in the algorithm. Assumption 4 basically requires that the properties of g in Assumption 2 hold on this larger domain of $\tilde{\Theta}$. We remark that the constants $\underline{g}$ and $\bar{g}$ are always one under the linear demand model; but for the binary demand model, these two constants may change after the domain is enlarged.

Theorem 4. *Under Assumption 1, 3, 4, with any sequence $\{x_t\}_{t=1,...,T}$, if we choose the regularization parameter $\lambda = 1$, the regret of Algorithm 3 can be bounded by*

$$8\left(8\sqrt{e\pi}+1\right)\bar{p}\bar{\alpha}d\bar{g}\sqrt{T\log(4dT^2)\log\left(\frac{2d+T}{2d}\right)} + 2\bar{p} = O\left((d\log T)^{\frac{3}{2}}\sqrt{T}\right).$$

Compared to Algorithm 1, there is an extra factor of $\sqrt{d}$ in the regret bound here. This extra factor is also inevitable for the existing analyses of Thompson sampling algorithms on the linear bandits problem (Agrawal and Goyal, 2013; Abeille and Lazaric, 2017). The proof of Theorem 4 largely follows the derivation in (Abeille and Lazaric, 2017). Some special care needs to be taken with respect to the price optimization step in the Algorithm 3. The price optimization restricts the optimal price to an interval of $[\underline{p}, \bar{p}]$, and thus some projection into this interval may sometimes be required for the unconstrained optimal price. The projection prevents a direct application of the gradient-based single step regret bound in (Abeille and Lazaric, 2017), but a similar argument can be made using the Lipschitzness of the function g and the boundedness of the covariates and the parameters. We defer the detailed proof to Section B.

While Algorithm 3 resolves the computational efficiency of the UCB optimization (8), we note that the Thompson Sampling algorithm is also compatible with the approximate quasi-MLE estimator. Algorithm 4 is parallel to Algorithm 2 in that an approximate solution to the quasi-MLE problem (6) is used. And the regret bound can be revised accordingly as follows. We remark that the Assumption 4 should be slightly changed in that the set $\tilde{\Theta}$ should be redefined by $\tilde{\Theta} := \{\theta \in \mathbb{R}^{2d} : \exists\tilde{\theta} \in \Theta, \|\theta - \tilde{\theta}\|_2 \le$

Algorithm 4 Thompson Sampling Pricing with Approximation

Input: Regularization parameter λ.

for $t = 1, ..., T$ **do**

Compute the estimator $\check{\theta}_{t-1}$ by (9) with approximation gap Δ_{t-1}, observe feature x_t.

Sample $\xi_t \sim \mathcal{N}(0, I_{2d})$ and compute the parameter

$$\tilde{\theta}_{t-1} := \check{\theta}_{t-1} + \left(\sqrt{\frac{2}{\underline{g}}\Delta_{t-1}} + \alpha(M_{t-1})\right) M_{t-1}^{-1/2}\eta_t.$$

Set the price by

$$p_t = \arg\max_{p\in[\underline{p},\bar{p}]} r\left(p; \tilde{\theta}_{t-1}, x_t\right)$$

end for

$2\sqrt{d\log(4dT^2)}(\bar{\alpha} + \bar{\Delta})\}$. This is because the approximate quasi-MLE solution may further enlarge the sampling region.

Theorem 5. *Under Assumption 1, 3, 4, with any sequence $\{x_t\}_{t=1,...,T}$, if we choose the regularization parameter $\lambda = 1$, then the regret of Algorithm 4 can be bounded by*

$$8\bar{g}\bar{p}\sqrt{d\log(4dT^2)}\left(8\sqrt{e\pi}+1\right)\left(\bar{\alpha}\sqrt{Td\log\left(\frac{2d+T}{2d}\right)} + \frac{\sqrt{2}}{2\sqrt{\underline{g}}}\sum_{t=1}^{T}\sqrt{\Delta_{t-1}}\|z_t\|_{M_{t-1}^{-1}}\right) + 2\bar{p}.$$

Moreover, if $\Delta_{t-1} \le \bar{\Delta}$ for all $t = 1, ..., T$, the regret bound becomes

$$8\bar{p}\bar{g}d\left(8\sqrt{e\pi}+1\right)\left(\bar{\alpha} + \sqrt{\frac{2\bar{\Delta}}{\underline{g}}}\right)\sqrt{T\log(4dT^2)\log\left(\frac{2d+T}{2d}\right)} + 2\bar{p}.$$

4.2 Better regret with known price coefficient

In this subsection, we consider the setting of known price coefficient, i.e., γ^* in (4) is known. From a modeling viewpoint, we remark that under the binary demand model (2), the assumption of known γ^* suffers from the same issue as we discussed in Section 1.1. Under the linear demand model with a constant price coefficient such as Example 2, the assumption may be potentially practical in that the price coefficient can be estimated through small perturbed pricing experiments (Nambiar et al., 2019). In general, the algorithm and analysis presented in the following are mainly for technical illustration purposes: (i) to identify a difference between the dynamic pricing problem and the bandits problem; (ii) to explain the achievability of $O(\log T)$ regret dependency in the literature.

Assumption 5 (Known γ^* and smoothness)**.** *Assume γ^* in (4) is known. In addition, assume there exists a constant C such that the optimal expected revenue function satisfies*

$$|r^*(\theta^*, x) - r\left(p^*(\theta, x); \theta^*, x\right)| \le C(x^\top\beta^* - x^\top\beta)^2,$$

for all $\theta, \theta^* \in \Theta$ with $\theta = (\beta; \gamma^*)$ and $\theta^* = (\beta^*; \gamma^*)$.

To interpret the condition in the assumption, we recall that $p^*(\theta, x)$ denotes the optimal price under the parameter θ. So the left-hand-side represents the revenue loss caused by using a wrong parameter θ for pricing, while the right-hand-side is quadratic in terms of the linear estimation error. One sufficient condition for the assumption is that $r(p; \theta^*, x)$ is continuously twice differentiable with respect to p for all possible θ^* and x, and $p^*(\theta, x)$ is Lipschitz in $x^\top\beta$. Essentially, this condition does not impose extra restriction upon Assumptions 1, 2, and 3; in other words, almost all the demand models that satisfy the previous assumptions also meet this condition under the knowledge of γ^*. For example, this condition can be met by the binary demand model (2) with a log-concave unknown noise (Javanmard and Nazerzadeh, 2019) and by the linear demand model (3). It is also analogous to the "well separation" condition in the covariate-free case (Broder and Rusmevichientong, 2012).

To proceed with the algorithm description, we first slightly revise the MLE estimator in Section 3.1 for known γ^*. Specifically, we redefine the misspecified likelihood function for the case of known γ^* as

$$\tilde{l}_t(\beta) := -\int_{D_t}^{g(x_t^\top \beta + x_t^\top \gamma^* \cdot p_t)} \frac{1}{h(u)}(u - D_t)\mathrm{d}u,$$

where the function $h(u)$ is the same as in Section 3.1. Then the estimator becomes

$$\hat{\beta}_t := \underset{\beta \in \Theta_\beta}{\arg\max} -\lambda \underline{g} \|\beta\|_2^2 + \sum_{\tau=1}^{t} \tilde{l}_\tau(\beta), \tag{9}$$

where Θ_β denotes the subspace $\{\beta : (\beta, \gamma^*) \in \Theta\}$. Compared to the previous case of unknown γ^*, the only change made here is to plug in the known value of γ^* and to restrict the attention to estimating the unknown β^*. Accordingly, we revise the definition of the (cumulative) design matrix as

$$\tilde{M}_t := \lambda I_d + \sum_{\tau=1}^{t} x_\tau x_\tau^\top$$

with I_d as an identity matrix of dimension d.

The following result is parallel to Corollary 1. We omit the proof as it is the same as the previous case of unknown γ^* except for some minor notation changes.

Proposition 2. *For all* $\lambda > 0$*,*

$$\mathbb{P}\left(\exists t \in \{1, ..., T\} : \left\|\hat{\beta}_t - \beta^*\right\|_{\tilde{M}_t} \geq \alpha(\tilde{M}_t)\right) \leq \frac{1}{T}.$$

Moreover, for all $t = 1, ..., T$*,*

$$\alpha(\tilde{M}_t) \leq \bar{\alpha}' := 2\sqrt{\lambda}\bar{\theta} + \frac{2\bar{\sigma}}{\underline{g}}\sqrt{2\log(T) + d\log\left(\frac{d\lambda + T}{d\lambda}\right)}.$$

Algorithm 5 describes a certainty-equivalent pricing policy. At each time step, the algorithm performs a regularized quasi-MLE to obtain the estimator $\hat{\beta}_{t-1}$. Then it assumes $\hat{\beta}_{t-1}$ to be the true parameter and finds the corresponding optimal price.

Algorithm 5 Certainty-Equivalent Pricing

Input: Regularization parameter λ.
for $t = 1, ..., T$ **do**
  Compute the estimator $\hat{\beta}_{t-1}$ by (9), observe feature x_t and set the price by

$$p_t = p^*\left(\hat{\theta}_{t-1}, x_t\right)$$

  where $\hat{\theta}_{t-1} = (\hat{\beta}_{t-1}; \gamma^*)$
end for

Theorem 6. *Under Assumptions 1, 2, 3, 5 and with any sequence* $\{x_t\}_{t=1,...,T}$*, if we choose the regu-*

larization parameter $\lambda = 1$, the regret of Algorithm 5 is upper bounded by

$$2C\bar{\alpha}'^2 d \log\left(\frac{d+T}{d}\right) + \bar{p} = O(d^2 \log^2 T)$$

where $\bar{\alpha}' = 2\bar{\theta} + \frac{2\bar{\sigma}}{\underline{g}}\sqrt{2\log T + d\log\left(\frac{d+T}{d}\right)}$ *is defined in Proposition 2 and C is defined in Assumption 5.*

Theorem 6 provides a regret upper bound for Algorithm 5. We remark that it is unnecessary for Algorithm 5 to compute the estimator at every time step. Javanmard and Nazerzadeh (2019) solve an L_1 regularized linear regression on geometric time intervals, and the scheme can also be applied to Algorithm 5 with the same order of regret bound. The frequent or infrequent estimation scheme makes no analytical difference and the choice mainly accounts for computation consideration. Xu and Wang (2021) study the special case of the binary demand model with unit price coefficient (i.e., known γ^*), and they derive the same order of regret bound as Theorem 6 under adversarial covariates using online Newton's method. The intuition is that the convergence rate of Newton's method is on the same order with the MLE estimator, so the corresponding output can be viewed as an approximate MLE estimator at each time step, and the approximation will not deteriorate the regret performance. This is aligned with Theorem 3 and Theorem 5 where an approximate quasi-MLE estimator is used.

In general, many existing the $o(\sqrt{T})$ regret bounds (Broder and Rusmevichientong, 2012; Javanmard, 2017; Javanmard and Nazerzadeh, 2019; Xu and Wang, 2021) fall into this paradigm of

$$\text{known price coefficient} \ + \ \text{certainty-equivalent policy}.$$

Intuitively, when the price coefficient is known, the price p_t will not interfere the learning of β^*. Thus there is no need to do price exploration like UCB or TS, and the regret purely reflects the cumulative learning rate of β^*. This disentanglement of pricing decisions from parameter estimation makes the setting of known γ^* analogous to the "full information" setting in online learning literature. In contrast, when the price coefficient is unknown, the pricing decisions will affect the learning rate of γ^*, thus the setting of unknown γ^* is more aligned with the "partial information" setting such as the bandits problem.

Interpreting the result under a linear demand model.

We use a linear demand model to further illustrate the contrast between $O(\sqrt{T})$ and $O(\log T)$ regret dependency. Consider the demand follows

$$D_t = a + bp_t + \epsilon_t$$

for some $a > 0$ and $b < 0$. At time t, the seller sets the price by $p_t = -\frac{\hat{a}_t}{2\hat{b}_t}$ with for some estimators $\hat{a}_t$ and $\hat{b}_t$ which could be from either Algorithm 1 (optimistic estimators) or Algorithm 5 (CE estimators). Then the single step regret can be expressed by a function of the true parameters and the estimators,

$$\text{Reg}_t = r^*((a,b)) - r(p_t;(a,b)) = -2b\left(\frac{\hat{a}_t}{2\hat{b}_t} - \frac{a}{2b}\right)^2 = -\frac{(a\hat{b}_t - \hat{a}_t b)^2}{2b\hat{b}_t^2} \tag{10}$$

- When the price coefficient is known, $\hat{b}_t = b$. The equality becomes

$$\text{Reg}_t = -\frac{1}{2b}(\hat{a}_t - a)^2.$$

- When the price coefficient is unknown, we cannot do more than a first-order Talyor expansion when we want to upper bound Reg_t by the estimation error, i.e.,

$$\text{Reg}_t \leq c(|\hat{a}_t - a| + |\hat{b}_t - b|) \tag{11}$$

for some $c > 0$.

For this example, whether the price coefficient b is known determines the space in which we view the right-hand-side of (10) as a function of $\hat{a}_t$ and $\hat{b}_t$. Intuitively, suppose that the estimation error is on the order of $\sqrt{1/t}$ (the intuition is precise when the covariates are i.i.d.). Then the right-hand-side will recover two different regret bounds under the two settings.

Generally, we remark that the first-order bound like (11) under a proper norm is always the first step for the analysis of UCB and TS algorithms, including our analysis for the dynamic pricing problem. For linear bandits problem, the LinUCB algorithm (Chu et al., 2011; Abbasi-Yadkori et al., 2011) can directly obtain this first-order bound and under TS algorithms, it can be obtained by some Bayesian arguments (Russo and Van Roy, 2014) or by maintaining a constant probability of choosing an optimistic action with anti-concentration sampling (Abeille and Lazaric, 2017).

5 Numerical Experiments and Conclusion

5.1 Numerical Experiments

We consider the linear demand model (3) for three groups of numerical experiments:

(a) Covariates x_t are i.i.d generated throughout all time periods.

(b) The horizon is split into two phases with equal length: In first phase, the first half of the covariates (dimension 1 to $d/2$) as a sub-vector are i.i.d generated over time while the second half (dimension $d/2+1$ to d) are all zero; the first half of the covariates (dimension 1 to $d/2$) are all zero while the second half (dimension $d/2+1$ to d) as a sub-vector are i.i.d generated over time.

(c) The horizon is split into size phases with equal length: In first three phases, at Phase $m = 1, 2, 3$, only x_t's m-th third covariates are non-zero and i.i.d. generated over time (just like (b) but with three groups of covariates). For Phase $m = 4, 5, 6$, it repeats the generation mechanism of Phase $m-3$, respectively.

Moreover, we set the allowable price range as $[0.1, 5]$. For each simulation trial, the parameter β^* is generated by a uniform distribution over $\frac{1}{\sqrt{d}}[1,2]^d$ and γ^* is generated by a uniform distribution over $-\frac{1}{\sqrt{d}}[0,1]^d$. The dimension d will be tested for different values as shown in the figures below. For experiments (a), (b) and (c), the covariates (if non-zero) is always generated i.i.d. from a uniform distribution over $\frac{1}{\sqrt{d}}[0,1]$. The normalizing factor $\frac{1}{\sqrt{d}}$ ensures the demand always stays on the same magnitude for different d.

For benchmark purpose, we adapt the covariate-free constrained iterated least square (CILS) algorithm (Keskin and Zeevi, 2014) for the covariate setting. Specifically, the price is set by

$$p_t = \begin{cases} \bar{p}_{t-1} + \text{sgn}(\delta_t)\kappa t^{-\frac{1}{4}}, & \text{if } |\delta_t| < \kappa t^{-\frac{1}{4}}, \\ p^*(\hat{\theta}_t), & \text{otherwise }, \end{cases}$$

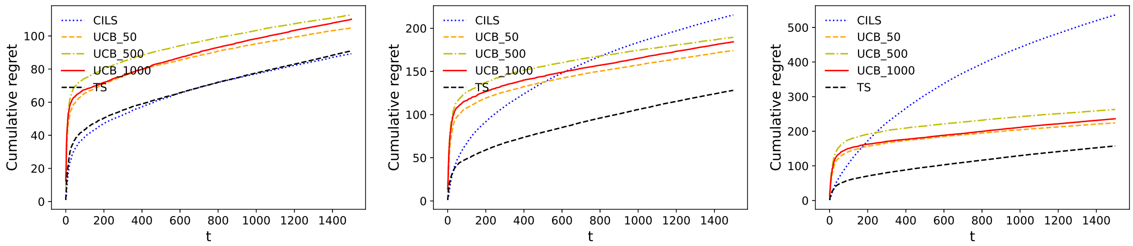
where $\hat{\theta}_t$ is the least square estimator for the unknown parameters, $\bar{p}_{t-1}$ is the average of the prices over the period 1 to $t-1$, and $\delta_t = p^*(\hat{\theta}_t) - \bar{p}_{t-1}$. The intuition is that if the tentative price $p^*(\hat{\theta}_t)$ stays too close to the history average, we will introduce a small perturbation as price experimentation to encourage the parameter learning. The parameter κ is a hyper-parameter and after a moderate tuning, we choose $\kappa = \frac{d}{10}$ in our experiments. For Algorithm 1, we also set the confidence set by

$$\Theta_t = \left\{ \theta \in \Theta : \left\| \hat{\theta}_{t-1} - \theta \right\|_{M_{t-1}}^2 \le \frac{d}{10} \right\}.$$

For Algorithm 3, we choose the sampled parameter by

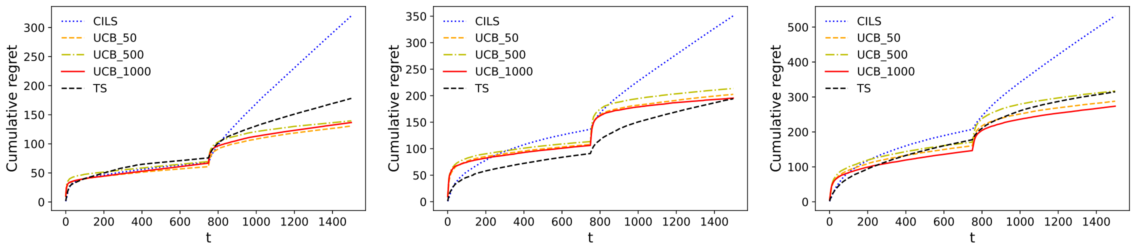
$$\tilde{\theta}_t = \hat{\theta}_t + \frac{\sqrt{d}}{25} M_t^{-1/2} \eta_t.$$

For both UCB and TS algorithms, we choose the regularization parameter $\lambda = 1$. We set the horizon $T = 1500$ and plot the cumulative gap between the online revenue and the optimal revenue. The curve is plotted based on an average over 100 simulation trials. We use covariate CILS to denote the benchmark algorithm.



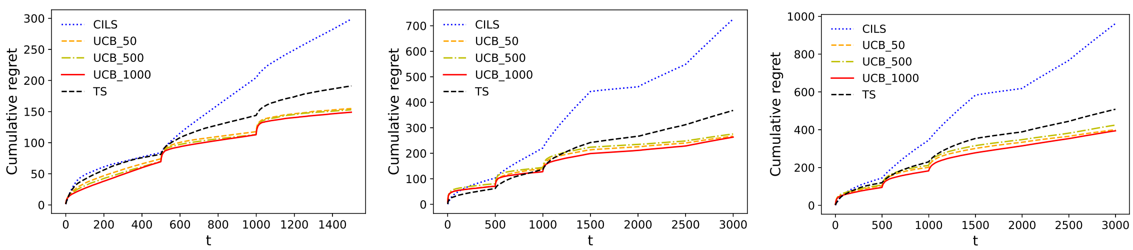
(a) $d = 6$ (b) $d = 12$ (c) $d = 18$

Figure 1: Experiment (a): i.i.d. covariates



(a) $d = 6$ (b) $d = 12$ (c) $d = 18$

Figure 2: Experiment (b): two phases with different distributions



(a) $d = 6$ (b) $d = 12$ (c) $d = 18$

Figure 3: Experiment (c): six phases with repeating patterns

Figure 1 shows Experiment (a) where the covariates are i.i.d. generated over time. We can see that all the algorithms exhibit sublinear curve and their performances are quite comparable to each other.

Figure 2 shows Experiment (b) where there are two phases with covariates from different distributions. We notice that the CILS algorithm performs well for the first phase but fails to learn the parameter during the second phase. Algorithm 1 performs stably well under all numbers of Monte Carlo samples. At the beginning of Phase 2, the curves of Algorithm 1 and Algorithm 3 grow quickly but then they all flatten. Figure 3 shows Experiment (c) where there are six phases and the later three phases repeat the distributions of the first three phases, respectively. We observe that both Algorithm 1 and Algorithm 3 have successfully learned all the parameters over the first three phases, so the corresponding curves do not grow significantly over the later three phases. Besides, we also note that the performance of Algorithm 1 is quite insensitive with respect to the number of samples K in the Monte Carlo method across all the experiments.

5.2 Discussion and Future Directions

In this paper, we consider the dynamic pricing problem under a generalized linear demand model where the demand is dependent on the price and covariates. We develop general-purposed algorithms and derive regret bounds under the settings of unknown price coefficient and known price coefficient. We conclude our discussion with a few possible future work directions.

- Demand model beyond Lipschitzness and unimodality.

  All the existing works discussed in this paper assume some Lipschitzness on the demand function. For the generalized linear demand model, this is exemplified in Assumption 2. Though the structure may be well justified in practice, the regret's dependency on the Lipschitz constant, as well as some other constants, is often overlooked in the regret analysis. A more complete characterization of the regret in terms of all the model parameters will bring more insights and it deserves more future efforts. den Boer and Keskin (2020) consider a piece-wise linear (discontinuous) demand function and the model therein might point to a direction for non-Lipschitzness. Besides, the consideration of a multimodal revenue function (Wang et al., 2021) in dynamic pricing under the adversarial covariates is not addressed in this work.

- Non-parametric distribution of ϵ_t under the binary demand model.

  In Section 1.1, we discuss the modeling issue of assuming the price coefficient to be 1. The binary demand model (2) partially resolves the problem but still imposes a parametric structure for the distribution. The question is what if the parametric distribution family is misspecified and what effect the misspecification will bring to the algorithm performance. Luo et al. (2021) consider a fully non-parametric distribution for ϵ_t and obtain a regret on the order of $T^{2/3}$. The performance deterioration might be the paid price for not knowing the parameterized structure of ϵ_t. It remains unclear whether the dependency on T can be further improved under a non-parametric distribution on ϵ_t.

- Online model selection.

  The existing works on dynamic pricing often assume the demand function belongs to a family of functions. The question is how to choose the proper function family and what if the true demand function lies outside the family. Besbes and Zeevi (2015) consider the covariate-free case and illustrate the robustness of a linear demand model. With the presence of covariates, Javanmard and Nazerzadeh (2019) and Ban and Keskin (2021) both consider the high-dimensional setting where the variable selection can be viewed as a model selection procedure. More generally, the question remains open whether the problem of dynamic pricing with covariates exhibits a similar formulation as the online model selection problem (Chatterji et al., 2020; Foster et al., 2019) where the true demand model may belong to a large family of models and the algorithm's regret is adaptively determined by the some complexity measure of the true demand model.

References

Abbasi-Yadkori, Yasin, Dávid Pál, Csaba Szepesvári. 2011. Improved algorithms for linear stochastic bandits. *Advances in neural information processing systems* **24** 2312–2320.

Abeille, Marc, Alessandro Lazaric. 2017. Linear thompson sampling revisited. *Artificial Intelligence and Statistics*. PMLR, 176–184.

Agrawal, Shipra, Navin Goyal. 2013. Thompson sampling for contextual bandits with linear payoffs. *International Conference on Machine Learning*. PMLR, 127–135.

Armstrong, Jon Scott, Kesten C Green, et al. 2005. Demand forecasting: evidence-based methods. Tech. rep., Citeseer.

Baardman, Lennart, Setareh Borjian Boroujeni, Tamar Cohen-Hillel, Kiran Panchamgam, Georgia Perakis. 2020. Detecting customer trends for optimal promotion targeting. *Manufacturing & Service Operations Management* .

Bailey, Michael, Drew M Johnston, Theresa Kuchler, Johannes Stroebel, Arlene Wong. 2019. Peer effects in product adoption. Tech. rep., National Bureau of Economic Research.

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

Bastani, Hamsa, David Simchi-Levi, Ruihao Zhu. 2021. Meta dynamic pricing: Transfer learning across experiments. *Management Science* .

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

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

Chatterji, Niladri, Vidya Muthukumar, Peter Bartlett. 2020. Osom: A simultaneously optimal algorithm for multi-armed and linear contextual bandits. *International Conference on Artificial Intelligence and Statistics*. PMLR, 1844–1854.

Chen, Xi, Zachary Owen, Clark Pixton, David Simchi-Levi. 2021. A statistical learning approach to personalization in revenue management. *Management Science* .

Chu, Wei, Lihong Li, Lev Reyzin, Robert Schapire. 2011. Contextual bandits with linear payoff functions. *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*. JMLR Workshop and Conference Proceedings, 208–214.

Cohen, Maxime C, Ilan Lobel, Renato Paes Leme. 2020. Feature-based dynamic pricing. *Management Science* **66**(11) 4921–4943.

Dani, Varsha, Thomas P Hayes, Sham M Kakade. 2008. Stochastic linear optimization under bandit feedback .

Den Boer, Arnoud 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, Arnoud V, N Bora Keskin. 2020. Discontinuous demand functions: estimation and pricing. *Management Science* **66**(10) 4516–4534.

den Boer, Arnoud V, Bert Zwart. 2014. Simultaneously learning and optimizing using controlled variance pricing. *Management science* **60**(3) 770–783.

Ferreira, Kris Johnson, Bin Hong Alex Lee, David Simchi-Levi. 2016. Analytics for an online retailer: Demand forecasting and price optimization. *Manufacturing & Service Operations Management* **18**(1) 69–88.

Ferreira, Kris Johnson, David Simchi-Levi, He Wang. 2018. Online network revenue management using thompson sampling. *Operations research* **66**(6) 1586–1602.

Filippi, Sarah, Olivier Cappe, Aurélien Garivier, Csaba Szepesvári. 2010. Parametric bandits: The generalized linear case. *NIPS*, vol. 23. 586–594.

Foster, Dylan J, Akshay Krishnamurthy, Haipeng Luo. 2019. Model selection for contextual bandits. *arXiv preprint arXiv:1906.00531* .

Goolsbee, Austan, Peter J Klenow. 2002. Evidence on learning and network externalities in the diffusion of home computers. *The Journal of Law and Economics* **45**(2) 317–343.

Javanmard, Adel. 2017. Perishability of data: dynamic pricing under varying-coefficient models. *The Journal of Machine Learning Research* **18**(1) 1714–1744.

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

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

Keskin, N Bora, Assaf Zeevi. 2018. On incomplete learning and certainty-equivalence control. *Operations Research* **66**(4) 1136–1167.

Lai, Tze Leung, Ching Zong Wei. 1982. Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *The Annals of Statistics* **10**(1) 154–166.

Lattimore, Tor, Csaba Szepesvári. 2020. *Bandit algorithms*. Cambridge University Press.

Luo, Yiyun, Will Wei Sun, Yufeng Liu. 2021. Distribution-free contextual dynamic pricing. *arXiv preprint arXiv:2109.07340* .

Mao, Jieming, Renato Paes Leme, Jon Schneider. 2018. Contextual pricing for lipschitz buyers. *NeurIPS*. 5648–5656.

Nambiar, Mila, David Simchi-Levi, He Wang. 2019. Dynamic learning and pricing with model misspecification. *Management Science* **65**(11) 4980–5000.

Nasr, Walid W, Ibrahim J Elshar. 2018. Continuous inventory control with stochastic and non-stationary markovian demand. *European Journal of Operational Research* **270**(1) 198–217.

Neale, John J, Sean P Willems. 2009. Managing inventory in supply chains with nonstationary demand. *Interfaces* **39**(5) 388–399.

Qiang, Sheng, Mohsen Bayati. 2016. Dynamic pricing with demand covariates. *Available at SSRN 2765257* .

Russo, Daniel, Benjamin Van Roy. 2014. Learning to optimize via posterior sampling. *Mathematics of Operations Research* **39**(4) 1221–1243.

Seiler, Stephan, Song Yao, Wenbo Wang. 2017. Does online word of mouth increase demand?(and how?) evidence from a natural experiment. *Marketing Science* **36**(6) 838–861.

Wang, Yining, Boxiao Chen, David Simchi-Levi. 2021. Multimodal dynamic pricing. *Management Science* .

Xu, Jianyu, Yu-xiang Wang. 2021. Logarithmic regret in feature-based dynamic pricing. *NeurIPS* .

Zervas, Georgios, Davide Proserpio, John W Byers. 2017. The rise of the sharing economy: Estimating the impact of airbnb on the hotel industry. *Journal of marketing research* **54**(5) 687–705.

Zhu, Feng, Zeyu Zheng. 2020. When demands evolve larger and noisier: Learning and earning in a growing environment. *International Conference on Machine Learning*. PMLR, 11629–11638.

A Proofs for Section 3

Proof of Proposition 1

Proof. We perform a second-order Taylor's expansion for the objective function of regularized quasi-MLE (6) around the true parameter θ^*. Let

$$Q_t(\theta) := \sum_{\tau=1}^{t} l_\tau(\theta).$$

We have

$$\begin{aligned} &Q_t(\theta^*) - \lambda \underline{g}\|\theta^*\|_2^2 - Q_t(\theta) + \lambda \underline{g}\|\theta\|_2^2 \\ =& -\left\langle \nabla Q_t(\theta^*) - \lambda \underline{g}\theta^*, \theta - \theta^* \right\rangle - \frac{1}{2}\left\langle \theta - \theta^*, \left(\nabla^2 Q_t(\theta') - \lambda \underline{g} I_{2d}\right)(\theta - \theta^*)\right\rangle \end{aligned} \tag{12}$$

for some θ' on the line segment between θ and θ^*.

By the optimality of $\hat{\theta}_t$,

$$Q_t(\theta^*) - \lambda \underline{g}\|\theta^*\|_2^2 \leq Q_t(\hat{\theta}_t) - \lambda \underline{g}\|\hat{\theta}_t\|_2^2.$$

Then from (12), we have

$$\left\langle \nabla Q_t(\theta^*) - \lambda \underline{g}\theta^*, \hat{\theta}_t - \theta^* \right\rangle + \frac{1}{2}\left\langle \hat{\theta}_t - \theta^*, \left(\nabla^2 Q_t(\tilde{\theta}') - \lambda \underline{g} I_{2d}\right)(\hat{\theta}_t - \theta^*)\right\rangle \geq 0, \tag{13}$$

for some $\tilde{\theta}'$ on the line segment between θ and θ^*.

From Assumption 2,

$$-\nabla^2 Q_t(\tilde{\theta}) = \sum_{\tau=1}^{t} g'\left(z_t^\top \tilde{\theta}\right) z_\tau z_\tau^\top \geq \underline{g} \cdot \sum_{\tau=1}^{t} z_\tau z_\tau^\top \quad \forall \tilde{\theta} \in \Theta.$$

Further, with Holder's inequality and (13),

$$
\begin{aligned}
\left\|\nabla Q_t(\theta^*) - \lambda \underline{g}\theta^*\right\|_{M_t^{-1}} \left\|\hat{\theta}_t - \theta^*\right\|_{M_t} \geq & \left\langle \nabla Q_t(\theta^*) - \lambda \underline{g}\theta^*, \hat{\theta}_t - \theta^* \right\rangle \\
\geq & \frac{1}{2} \left\langle \hat{\theta}_t - \theta^*, \left(-\nabla^2 Q_t(\theta') + \lambda \underline{g} I_{2d}\right)(\hat{\theta}_t - \theta^*) \right\rangle \\
\geq & \frac{1}{2}\underline{g} \left\langle \hat{\theta}_t - \theta^*, M_t(\hat{\theta}_t - \theta^*) \right\rangle \\
= & \frac{1}{2}\underline{g} \left\|\hat{\theta}_t - \theta^*\right\|_{M_t}^2
\end{aligned}
$$

almost surely. Consequently,

$$
\left\|\nabla Q_t(\theta^*) - \lambda \underline{g}\theta^*\right\|_{M_t^{-1}} \geq \frac{1}{2}\underline{g} \left\|\hat{\theta}_t - \theta^*\right\|_{M_t} \quad a.s. \tag{14}
$$

Recall that

$$
S_t = \sum_{\tau=1}^{t} \frac{\xi_\tau(\theta^*)}{\bar{\sigma}} z_\tau = \frac{1}{\bar{\sigma}} \nabla Q_t(\theta^*),
$$

which implies

$$
\begin{aligned}
\frac{1}{2\bar{\sigma}}\underline{g} \left\|\hat{\theta}_t - \theta^*\right\|_{M_t} \leq & \frac{1}{\bar{\sigma}} \left\|-\nabla Q_t(\theta^*) + \lambda \underline{g}\theta^*\right\|_{M_t^{-1}} \\
= & \frac{1}{\bar{\sigma}} \left\|-\bar{\sigma} S_t + \lambda \underline{g}\theta^*\right\|_{M_t^{-1}} \\
\leq & \|S_t\|_{M_t^{-1}} + \frac{\sqrt{\lambda}\underline{g}}{\bar{\sigma}} \sqrt{(\theta^*)^\top (\lambda M_t^{-1})\theta^*} \\
\leq & \|S_t\|_{M_t^{-1}} + \frac{\sqrt{\lambda}\underline{g}}{\bar{\sigma}} \|\theta^*\|_2.
\end{aligned}
$$

Here the first line comes from (14), the second line comes from the definition of S_t, the third lines comes from the norm inequality, and the last line is from the fact that $\lambda M_t^{-1} \leq I_{2d}$. Thus, we complete the proof from combining Theorem (1) with $\|\theta^*\|_2 \leq \bar{\theta}$. □

Proof of Theorem 2

Proof. We define the "good event" as $\mathcal{E} = \{\theta^* \in \Theta_t \text{ for } t = 1, ..., T\}$. From Corollary 1, we know

$$
\mathbb{P}(\mathcal{E}) \geq 1 - 1/T. \tag{15}
$$

At time t, under the event $\mathcal{E}$, the choice of θ_t in Algorithm 1 ensures

$$
r^*(\theta_t, x_t) \geq r^*(\theta^*, x_t). \tag{16}
$$

Thus, under the event $\mathcal{E}$, the single period regret can be bounded by

$$
\begin{aligned}
\mathrm{Reg}_t := r^*(\theta^*, x_t) - r(p_t; \theta^*, x_t) \leq & r^*(\theta_t, x_t) - r(p_t; \theta^*, x_t) \\
= & p_t \cdot \left(g\left(x_t^\top \beta_t + x_t^\top \gamma_t \cdot p_t\right) - g\left(x_t^\top \beta^* + x_t^\top \gamma^* \cdot p_t\right)\right) \\
\leq & \bar{p}\bar{g} \left|z_t^\top (\theta_t - \theta^*)\right| \\
\leq & \bar{p}\bar{g} \|z_t\|_{M_{t-1}^{-1}} \|\theta_t - \theta^*\|_{M_{t-1}} \\
\leq & \bar{p}\bar{g} \|z_t\|_{M_{t-1}^{-1}} \left(\|\theta_t - \hat{\theta}_{t-1}\|_{M_{t-1}} + \|\hat{\theta}_{t-1} - \theta^*\|_{M_{t-1}}\right) \\
\leq & 2\bar{p}\bar{g}\bar{\alpha} \|z_t\|_{M_{t-1}^{-1}}
\end{aligned}
$$

where the functions r^* and r are introduced in the Section 2. Here the first line is from (16), the third line is from Assumption 1, the fourth line is from Holder's inequality, and the last inequality is by Corollary 1 under the event $\mathcal{E}$.

Thus, the total expected regret (the expectation is with respect to the randomness of demand shocks) can be bounded by

$$\begin{aligned}
\mathrm{Reg}_T^{\pi_1}(x_1, ..., x_T) &= \sum_{t=1}^{T} \mathbb{E}\left[\mathrm{Reg}_t \cdot \mathbb{1}_{\mathcal{E}}\right] + \mathbb{E}\left[\mathrm{Reg}_t \cdot \mathbb{1}_{\mathcal{E}^c}\right] \\
&\leq 2\sum_{t=1}^{T} \bar{p}\bar{g}\bar{\alpha}\|z_t\|_{M_{t-1}^{-1}} + \mathbb{P}\left(\mathcal{E}^c\right) \cdot \bar{p}T \\
&\leq 2\bar{p}\bar{g}\bar{\alpha}\sqrt{T\sum_{t=1}^{T} \|z_t\|^2_{M_{t-1}^{-1}}} + \mathbb{P}\left(\mathcal{E}^c\right) \cdot \bar{p}T \\
&\leq 4\bar{p}\bar{g}\bar{\alpha}\sqrt{Td\log\left(\frac{2d\lambda + T}{2d\lambda}\right)} + \bar{p}
\end{aligned}$$

where π_1 denotes the pricing policy specified by Algorithm 1 and $\mathcal{E}^c$ denotes the complement of the event $\mathcal{E}$. The second inequality is by Holder's inequality and the last inequality is because (15) and Lemma 2. $\square$

Proof of Lemma 3

Proof. Recall that the regularized quasi-MLE at time t is defined as

$$Q_t(\theta) - \lambda \underline{g}\|\theta\|_2^2$$

with Hessian matrix

$$\sum_{\tau=1}^{t} -g'(z_\tau^\top \theta) z_\tau z_\tau^\top - \lambda \underline{g} I_{2d},$$

which is negative definite by the assumption that $g'(z_t^\top \theta) > 0$ (in the feasible domain of related parameters). Thus, the regularized quasi-MLE is concave in θ. We can then perform a second-order Taylor's expansion around the optimal solution $\hat{\theta}_t$ in Θ with any point $\theta \in \Theta$,

$$\begin{aligned}
&Q_t(\hat{\theta}_t) - \lambda \underline{g}\|\hat{\theta}_t\|_2^2 - Q_t(\theta) + \lambda \underline{g}\|\theta\|_2^2 \\
=& -\left\langle \nabla Q_t(\hat{\theta}_t) - \lambda \underline{g}\hat{\theta}_t, \theta - \hat{\theta}_t \right\rangle - \frac{1}{2}\left\langle \hat{\theta}_t - \theta, \left(\nabla^2 Q_t(\theta') - \lambda \underline{g} I_{2d}\right)(\hat{\theta}_t - \theta) \right\rangle \\
\geq& -\frac{1}{2}\left\langle \hat{\theta}_t - \theta, \left(\nabla^2 Q_t(\theta') - \lambda \underline{g} I_{2d}\right)(\hat{\theta}_t - \theta) \right\rangle \\
\geq& \frac{1}{2}\underline{g}\|\hat{\theta}_t - \theta\|^2_{M_t},
\end{aligned}$$

where $\theta' \in \Theta$ is a point between θ and $\hat{\theta}_t$, the first inequality is by the concavity and the optimality of $\hat{\theta}_t$ in a compact set Θ, and the second inequality is by $-\nabla^2 Q_t(\theta') \geq \underline{g}\sum_{\tau=1}^{t} z_\tau z_\tau^\top$. $\square$

B Proofs for Section 4

We first introduce some requirement for the sampling distribution.

Definition 1 (Sampling distribution (Abeille and Lazaric, 2017)). *A distribution $\mathcal{D}^{TS}$ is suitable for Thompson sampling if it is a multivariate distribution on $\mathbb{R}^{2d}$ absolutely continuous with respect to the Lebesgues measure which satisfies the following properties:*

- *(anti-concentration) there exists a positive probability q such that for any $u \in \mathbb{R}^{2d}$ with $\|u\|_2 = 1$,*

$$\mathbb{P}_{\eta \sim \mathcal{D}^{TS}}(u^\top \eta \geq 1) \geq q,$$

- *(concentration) there exist positive constants c, c' such that $\forall \delta \in (0, 1)$,*

$$\mathbb{P}_{\eta \sim \mathcal{D}^{TS}}\left(\|\eta\|_2 \leq \sqrt{2cd \log \frac{c'2d}{\delta}}\right) \geq 1 - \delta.$$

As shown in (Abeille and Lazaric, 2017), the Gaussian distribution $\xi \sim \mathcal{N}(0, I_{2d})$ that we use in Algorithm 3 satisfies the above definition with $c = c' = 2$ and $q = \frac{1}{4\sqrt{e\pi}}$. An alternative choice is $\eta \sim \mathcal{U}(0, \sqrt{2d})^{2d}$, uniform distribution on domain $[0, \sqrt{2d}]^{2d}$, and then $c = 1$, $c' = \frac{e}{2d}$ and $q = \frac{1}{16\sqrt{3\pi}}$.

When the sampling distribution of ξ in Algorithm 3 satisfies the Definition 1, it has the following property. Specifically, the following lemma tells that the sampled parameter $\tilde{\theta}_{t-1}$ will be "optimistic" with respect to some convex function f for a constant probability. By "optimistic", we mean the objective value under this parameter $\tilde{\theta}_{t-1}$ is no smaller than that under θ^*.

Lemma 4 (Lemma 3 of (Abeille and Lazaric, 2017)). *For the sampled parameter in Algorithm 3 $\tilde{\theta}_{t-1} = \hat{\theta}_{t-1} + \alpha(M_{t-1})M_{t-1}^{-1/2}\xi$ with $\xi \sim \mathcal{N}(0, I_{2d})$, then for any convex function $f(\theta)$ in θ and any $t \geq 1$,*

$$\mathbb{P}\left(f(\tilde{\theta}_{t-1}) \geq f(\theta^*) \middle| \mathcal{H}_{t-1}, \theta^* \in \Theta_{t-1}\right) \geq \frac{1}{8\sqrt{e\pi}}$$

where Θ_t is as defined in Algorithm 1.

Using this lemma, we proceed with the proof of Theorem 4.

Proof of Theorem 4

Proof. Let $\kappa(M) =: 2\sqrt{d \log(4dT^2)}\alpha(M)$. From Corollary 1, we know $\kappa(M) \leq 2\bar{\alpha}\sqrt{d \log(4dT^2)}$, which justifies the choice of $\tilde{\Theta}$ in Assumption 4. Compared to the definition of $\alpha(M)$ in the UCB case, the extra factor $2\sqrt{d \log(4dT^2)}$ in $\kappa(M)$ aims to account for the dispersion caused by the sampling in the Thompson sampling algorithm. Specifically, $\alpha(M_t)$ describes the volume of confidence bound for $\hat{\theta}_t$ in the UCB algorithm, while $\kappa(M_t)$ describes the volume (with high probability) of possible sampled parameter $\tilde{\theta}_t$. Define

$$\tilde{\Theta}_{t-1} \coloneqq \left\{\theta \in \mathbb{R}^{2d} : \|\theta - \hat{\theta}_{t-1}\|_{M_{t-1}} \leq \kappa(M_{t-1})\right\}.$$

Now, we define the good event for Thompson sampling as

$$\mathcal{E} = \{\theta^* \in \Theta_t, \tilde{\theta}_t \in \tilde{\Theta}_t \text{ for } t = 0, ..., T-1\}$$

where Θ_t is as defined in Algorithm 1.

Then by Definition 1 and Corollary 1, we have

$$\mathbb{P}(\mathcal{E}) \geq 1 - \frac{2}{T}.$$

Since the sampled parameter $\tilde{\theta}_t$ may be out of the original parameter set Θ, we revise the definition of optimal objective function by

$$\tilde{r}^*(\theta, x) := \max_{p \in [\underline{p}, \bar{p}]} r(p; \theta, x).$$

By Assumption 4, the function $g(\cdot)$ is convex, so $\tilde{r}^*(\theta, x)$ is also convex in θ by the preservation of convexity under linear transformation and maximization.

Now under event $\mathcal{E}$, the regret can be decomposed into

$$\begin{aligned}
\mathrm{Reg}_T^{\pi_3} \cdot \mathbb{1}_{\mathcal{E}} &= \sum_{t=1}^{T} \left(\tilde{r}^*(\theta^*, x_t) - r(p_t; \theta^*, x_t)\right) \cdot \mathbb{1}_{\mathcal{E}} \\
&= \sum_{t=1}^{T} \Big(\tilde{r}^*(\theta^*, x_t) - \tilde{r}^*(\tilde{\theta}_{t-1}, x_t) + \tilde{r}^*(\tilde{\theta}_{t-1}, x_t) - r(p_t; \theta^*, x_t)\Big) \cdot \mathbb{1}_{\mathcal{E}}
\end{aligned}$$

where π_3 denotes the pricing policy specified by Algorithm 3.

We denote

$$\begin{aligned}
\mathrm{Reg}_t^{(1)} &:= \Big(\tilde{r}^*(\theta^*, x_t) - \tilde{r}^*(\tilde{\theta}_{t-1}, x_t)\Big) \cdot \mathbb{1}_{\mathcal{E}}, \\
\mathrm{Reg}_t^{(2)} &:= \Big(\tilde{r}^*(\tilde{\theta}_{t-1}, x_t) - r(p_t; \theta^*, x_t)\Big) \cdot \mathbb{1}_{\mathcal{E}}.
\end{aligned}$$

With the same approach as the proof of Theorem 2, we have

$$\mathbb{E}\left[\sum_{t=1}^{T} \mathrm{Reg}_t^{(2)}\right] \le 8\bar{p}\bar{g}\bar{\alpha} d \sqrt{T \log(4dT^2) \log\left(\frac{2d+T}{2d}\right)}.$$

Now we focus on analyzing $\mathrm{Reg}_t^{(1)}$. Define

$$\Theta_t^{\mathrm{OPT}} = \{\theta \in \tilde{\Theta}_{t-1} : \tilde{r}^*(\theta, x_t) \ge \tilde{r}^*(\theta^*, x_t)\}.$$

The set contains parameters for which the corresponding optimal objective is larger than the optimal objective value under the true parameter. Then for any $\tilde{\theta} \in \Theta_t^{\mathrm{OPT}}$, we have

$$\begin{aligned}
\mathbb{E}\left[\mathrm{Reg}_t^{(1)} | \mathcal{H}_{t-1}\right] &\le \mathbb{E}\left[\Big(\tilde{r}^*(\tilde{\theta}; x_t) - \tilde{r}^*(\tilde{\theta}_{t-1}; x_t)\Big) \cdot \mathbb{1}_{\mathcal{E}} \middle| \mathcal{H}_{t-1}\right] \\
&= \mathbb{E}\left[\Big(\tilde{r}^*(\tilde{\theta}; x_t) - p_t \cdot g(x_t^\top \tilde{\beta}_{t-1} + x_t^\top \tilde{\gamma}_{t-1} \cdot p_t)\Big) \cdot \mathbb{1}_{\mathcal{E}} \middle| \mathcal{H}_{t-1}\right] \\
&\le \mathbb{E}\left[\tilde{p}^*(\tilde{\theta}, x_t) \cdot \Big(g(x_t^\top \tilde{\beta} + x_t^\top \tilde{\gamma} \cdot \tilde{p}^*(\tilde{\theta}, x_t)) - g(x_t^\top \tilde{\beta}_{t-1} + x_t^\top \tilde{\gamma}_{t-1} \cdot \tilde{p}^*(\tilde{\theta}, x_t))\Big) \cdot \mathbb{1}_{\mathcal{E}} \middle| \mathcal{H}_{t-1}\right] \\
&\le \bar{p}\bar{g}\mathbb{E}\left[\left|x_t^\top \tilde{\beta} + x_t^\top \tilde{\gamma} \cdot \tilde{p}^*(\tilde{\theta}, x_t) - (x_t^\top \tilde{\beta}_{t-1} + x_t^\top \tilde{\gamma}_{t-1} \cdot \tilde{p}^*(\tilde{\theta}, x_t))\right| \cdot \mathbb{1}_{\mathcal{E}} \middle| \mathcal{H}_{t-1}\right] \\
&= \bar{p}\bar{g}\mathbb{E}\left[\left|\tilde{z}_t(\tilde{\theta})^\top (\tilde{\theta} - \tilde{\theta}_{t-1})\right| \cdot \mathbb{1}_{\mathcal{E}} \middle| \mathcal{H}_{t-1}\right] \\
&\le 4\bar{p}\bar{g}\bar{\alpha}\sqrt{d \log(4dT^2)}\mathbb{E}\left[\|\tilde{z}_t(\tilde{\theta})\|_{M_{t-1}^{-1}} \cdot \mathbb{1}_{\mathcal{E}} \middle| \mathcal{H}_{t-1}\right],
\end{aligned} \tag{17}$$

where $\tilde{p}^*(\theta, x) = \arg\max_{p \in [\underline{p}, \bar{p}]} r(p; \theta, x)$ and $\tilde{z}_t(\theta) = (x_t, \tilde{p}^*(\theta, x_t) x_t)$. Here the third line is from the optimality of p_t, the fourth line is from Assumptions 1 and 4, and the last line comes from the definition of $\mathcal{E}$ together with Holder's inequality.

Let $\tilde{\theta}'_{t-1}$ be an independent copy of $\tilde{\theta}_{t-1}$ following the same distribution. Then we have

$$\begin{aligned}
\mathbb{E}\left[\mathrm{Reg}_t^{(1)}|\mathcal{H}_{t-1}\right] &\le 4\bar{p}\bar{g}\bar{\alpha}\sqrt{d\log(4dT^2)}\mathbb{E}\left[\|\tilde{z}_t(\tilde{\theta}'_{t-1})\|_{M_{t-1}^{-1}}\cdot\mathbb{1}_{\mathcal{E}}\middle|\tilde{\theta}'_{t-1}\in\Theta_t^{\mathrm{OPT}},\mathcal{H}_{t-1}\right]\\
&\le 4\bar{p}\bar{g}\bar{\alpha}\sqrt{d\log(4dT^2)}\mathbb{E}\left[\|\tilde{z}_t(\tilde{\theta}'_{t-1})\|_{M_{t-1}^{-1}}\cdot\mathbb{1}_{\mathcal{E}}\cdot\mathbb{1}_{\tilde{\theta}'_{t-1}\in\Theta_t^{\mathrm{OPT}}}\middle|\mathcal{H}_{t-1}\right]\cdot\mathbb{P}^{-1}\left(\tilde{\theta}'_{t-1}\in\Theta_t^{\mathrm{OPT}}|\mathcal{H}_{t-1},\theta^*\in\Theta_{t-1}\right)\\
&\le 4\bar{p}\bar{g}\bar{\alpha}\sqrt{d\log(4dT^2)}\mathbb{E}\left[\|\tilde{z}_t(\tilde{\theta}'_{t-1})\|_{M_{t-1}^{-1}}\middle|\mathcal{H}_{t-1}\right]\cdot\mathbb{P}^{-1}\left(\tilde{\theta}'_{t-1}\in\Theta_t^{\mathrm{OPT}}|\mathcal{H}_{t-1},\theta^*\in\Theta_{t-1}\right)\\
&\le 4\bar{p}\bar{g}\bar{\alpha}\sqrt{d\log(4dT^2)}\mathbb{E}\left[\|\tilde{z}_t(\tilde{\theta}_{t-1})\|_{M_{t-1}^{-1}}\middle|\mathcal{H}_{t-1}\right]\cdot 8\sqrt{e\pi}\\
&= 32\bar{p}\bar{\alpha}\bar{g}\sqrt{e\pi d\log(4dT^2)}\mathbb{E}\left[\|z_t\|_{M_{t-1}^{-1}}\middle|\mathcal{H}_{t-1}\right].
\end{aligned}$$

where $z_t = (x_t, p_t x_t)$ and p_t is the price used in the algorithm. Here the first line comes by replacing $\tilde{\theta}$ in (17) with a randomized parameter of $\tilde{\theta}'_{t-1}$ restricted to the set Θ_t^{OPT}. The second line comes from the property of conditional expectation. The third line removes the indicator functions. The fourth line applies Lemma 4 by setting the function f as $\tilde{r}^*(\cdot, x_t)$. The last line comes from the definition of $\tilde{z}_t(\theta)$ and z_t.

Finally, by using Lemma 2, we can conclude

$$\mathbb{E}\left[\sum_{t=1}^{T}\mathrm{Reg}_t^{(1)}\right] \le 64\bar{\alpha}\bar{g}\bar{p}\sqrt{e\pi d\log(4dT^2)}\sqrt{Td\log\left(\frac{2d+T}{2d}\right)}.$$

Combining the two parts of regret with the additional loss caused by the event $\mathcal{E}^C$ (bounded by $2\bar{p}$), we complete the proof. ☐

Proof of Theorem 6

Proof. Revise the definition of the "good event" as

$$\tilde{\mathcal{E}} := \left\{\left\|\hat{\beta}_t - \beta^*\right\|_{\tilde{M}_t} \le \alpha(\tilde{M}_t) \text{ for } t = 0, ..., T-1\right\},$$

From Proposition 2, we know

$$\mathbb{P}(\tilde{\mathcal{E}}) \ge 1 - 1/T.$$

Under the event $\tilde{\mathcal{E}}$, the single period regret can be bounded by

$$\begin{aligned}
\mathrm{Reg}_t &= r^*(\theta^*, x_t) - r(p_t; \theta^*, x_t)\\
&\le C\left|x_t^\top \beta^* - x_t^\top \hat{\beta}_{t-1}\right|^2\\
&\le C\|x_t\|^2_{\tilde{M}_{t-1}^{-1}}\|\hat{\beta}_{t-1} - \beta^*\|^2_{\tilde{M}_{t-1}}\\
&\le C\bar{\alpha}'^2\|x_t\|^2_{\tilde{M}_{t-1}^{-1}}.
\end{aligned}$$

Here the second line is from Assumption 5, the third line is from Holder's inequality, and the last inequality is by Proposition 2 under the event $\tilde{\mathcal{E}}$.

Thus, the total expected regret (the expectation is with respect to the randomness of demand shocks) can be bounded by

$$
\begin{aligned}
\mathrm{Reg}_T^{\pi_{\mathrm{CE}}}(x_1, ..., x_T) &= \sum_{t=1}^{T} \mathbb{E}\left[\mathrm{Reg}_t \cdot \mathbb{1}_{\tilde{\mathcal{E}}}\right] + \mathbb{E}\left[\mathrm{Reg}_t \cdot \mathbb{1}_{\tilde{\mathcal{E}}^c}\right] \\
&\leq \sum_{t=1}^{T} C\bar{\alpha}'^2 \|x_t\|^2_{\tilde{M}_{t-1}^{-1}} + \mathbb{P}\left(\tilde{\mathcal{E}}^c\right) \cdot \bar{p}T \\
&\leq 2C\bar{\alpha}'^2 d \log\left(\frac{d\lambda + T}{d\lambda}\right) + \bar{p}
\end{aligned}
$$

where π_{CE} denotes Algorithm 5 and $\tilde{\mathcal{E}}^c$ denotes the complement of the event $\tilde{\mathcal{E}}$. The last inequality is because of Lemma 2. □