Submitted to *Management Science*
manuscript (Please, provide the manuscript number!)

Context-Based Dynamic Pricing with Separable Demand Models

Jinzhi Bu
Department of Logistics and Maritime Studies, Faculty of Business, The Hong Kong Polytechnic University, Hung Hom, Kowloon, HongKong, jinzhi.bu@polyu.edu.hk

David Simchi-Levi
Institute for Data, Systems, and Society, Department of Civil and Environmental Engineering, and Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA 02139, dslevi@mit.edu

Chonghuan Wang
Center for Computational Science and Engineering and Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, chwang9@mit.edu

Motivated by the empirical evidence observed from the real-world dataset, this paper studies context-based dynamic pricing with separable demand models. Consider a seller selling a product over a finite horizon of T periods and facing an unknown expected demand function that admits a separable structure $f(p)+g(x)$, where $p\in\mathbb{R}$ and $x\in\mathbb{R}^d$ denote the product's price and features respectively. The seller does not know the exact expression of $f(p)$ or $g(x)$, but can dynamically adjust prices in each period based on the observed features and demands to learn their forms. The seller's objective is to maximize the T-period expected revenue. We systematically characterize the statistical complexity of the online learning problem under three configurations of demand models with different structures of $f(p)$ and $g(x)$. For each model, we design an efficient online learning algorithm with a provable regret upper bound. We also show that the upper bound is generally unimprovable by proving a matching regret lower bound in certain parameter regimes. Our results reveal fundamental differences in the optimal regret rates when $f(p)$ and $g(x)$ are endowed with different structures. The numerical results demonstrate that our learning algorithms are more effective than benchmark algorithms for all the three models, and also show the effects of the parameters associated with $f(p)$ and $g(x)$ on the algorithm's empirical regret.

Key words: separable model, dynamic pricing, contextual information, online learning

1. Introduction

The recent success of online retailers has provided an unprecedented data-rich environment for firms to take better advantage of the observable contextual information (or features, covariates) and thereby dynamically improve their pricing strategies. Examples of such contextual informa-

tion include products' characteristics, seasonality, festival information, economic indicators, etc. A fundamental question stemming from the presence of contextual information in dynamic pricing is how to select a predictive model to capture the dependency of demands on prices and contexts.

The model studied in this work is inspired by our collaboration with the industry partner (Li et al. 2023), one of the leading retailers of consumer electronics in Middle East. As a typical practice of this company, the prices for its products being sold, e.g., mobile phones, laptops, and electronics, are updated every shopping cycle, usually spanning two days. These updates are based on the past sales data and exogenous feature information. Our focus is on a specific product group of mobile phones for which we collect data from 2000 SKUs spanning the last three years. We note that approximately 83% of these SKUs have an average sales volume of less than 5 units per shopping cycle. The prevalence of the low-sales products limits the amount of information that can be distilled from the data of each single SKU, especially in terms of the price elasticity. Therefore, as pre-processing, we employ the traditional K-Means clustering technique to categorize products based on various historical attributes such as the discounted prices charged by the company, competitor prices, time since release, color, storage, brand, and sales volumes. The clustering allows us to assume that products within the same cluster share the same demand model.

After the above pre-processing step, we hope to identify a suitable predictive demand model for each cluster. One of the most widely used approaches is the non-parametric regression-tree-based model. We conducte experiments employing various standard techniques, including random forest regressor, LightGBM regressor (see, e.g., Ke et al. 2017) and Catboost regressor (see, e.g., Prokhorenkova et al. 2018). Nevertheless, even after clustering, the price elasticity derived by these methods tends to hover around zero, and the resulting price-demand curves for numerous SKUs exhibit near-flat trends. One potential pitfall of the non-parametric regression-tree-based model is that it may overlook different roles of the *endogenous* pricing decision and *exogenous* context information and simply regards them as homogeneous features in the predictive model. This underlines the necessity for differentiating the impacts of price and context in the demand model. More concretely, it drives us towards adopting a demand function with a separable structure, represented by $f(p)+g(x)$, where $p \in \mathbb{R}$ and $x \in \mathbb{R}^d$ denote the price and context respectively.

To implement the separable model $f(p)+g(x)$, we applied a piece-wise linear function to approximate $f(p)$ and Catboost regressor to estimate $g(x)$. We partition the dataset into a training set and a testing set, 80% and 20% chronologically ordered in the full dataset respectively. In industrial applications, gaining a deep understanding into the price elasticity is always emphasized. We then first compare the price-demand curve generated by our separable model against the non-parametric Catboost model. Figure 1 shows the result for one SKU in one testing period. Clearly seen from this figure, the separable model produces a price-demand curve that exhibits a declining trend (in

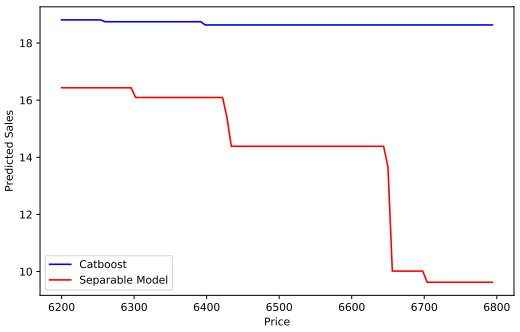


Figure 1 **Comparison of the price elasticity under Catboost model and separable model**

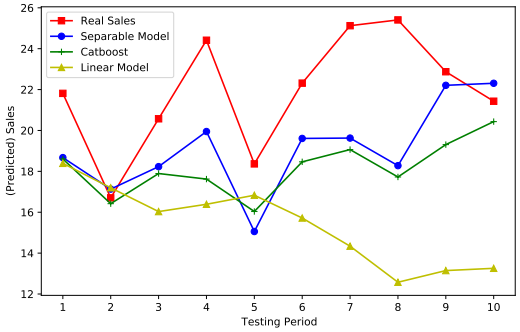


Figure 2 **Comparison of the predicted sales under three models versus the actual sales**

	Separable Model	Catboost Model	Linear Model
Avg. of squared errors	24.520	25.232	31.699
Std. of squared errors	225.192	238.185	269.583
Avg. of r^2 scores	0.644	0.642	0.544

Table 1 **Comparison of the prediction accuracy under three models for 678 SKUs**

the red curve), offering more insights into the price elasticity compared with the non-parametric Catboost model (in the blue curve). We also compare the predictive accuracy of this separable model with the parametric linear model and Catboost model in the testing dataset. For one representative cluster with 678 SKUs, Table 1 shows the average squared errors, the standard deviation of squared errors, and the average r^2 scores. The average r^2 scores are computed across the testing periods. That is, for every model in Table 1 and for each testing period, we calculate an r^2 score and then determine the overall average across all testing periods. We see that beyond an improved price elasticity curve, the separable model also outperforms the other two models slightly in all three metrics. In Figure 2, we also plot the predicted sales averaged over these 678 SKUs under the three models, alongside the actual average sales. While the average sales tend to be underestimated, the pattern under the separable model over the testing periods better aligns with that of the true one. More detailed empirical evidences can be found in Li et al. (2023).

Beyond the empirical evidence, the separable demand model also complements the existing studies on context-based dynamic pricing by preserving the advantages of two predominant modeling approaches in the literature. The first approach is to assume a parametric form to elucidate the relationship between demand and (price, context) pair. Among these models, the linear demand model, i.e., $bp + a^\top x$, is arguably the most fundamental one. It separates the effects of the price and context into an additive form, as our separable model does, but treats both effects in a linear manner. The parametric assumption usually brings much convenience to practical implementation, and encourages many efficient learning algorithms based on the well-developed regression theory. However, it may also lead to a serious issue of model mis-specification, as discussed in Nambiar

et al. (2019). The second approach is to employ a fully non-parametric model i.e., $d(p,x)$ for some unknown function $d(\cdot)$ without any specific parametric assumption. This model can be quite robust to different scenarios and application contexts. However, it fails to capitalize on the structural characteristics inherent to the problem itself. Compared to the parametric model, the separable model is still simple to interpret, but can alleviate the issue of model mis-specification. Compared to the non-parametric model, the separable model takes advantage of the separability structure that is validated from the real-world dataset, without sacrificing the robustness since $f(p)$ and $g(x)$ can be general functions.

1.1. Model and Research Questions

In this paper, we study a context-based dynamic pricing problem with separable demand models and online learning. Consider a seller selling one product over a horizon of T periods. At the beginning of each period t, the seller observes a context vector $x_t \in [0,1] \in \mathbb{R}^d$ encoding the product's characteristics and other exogenous information in period t, e.g., economic indicator, weather, competitors' prices. We assume that $\{x_t\}_{t\geq 1}$ are independently and identically distributed (i.i.d.) random variables (r.v.'s) drawn from some unknown distribution $\mathcal{P}$. The seller then chooses a price p_t from the feasible price range $[\underline{p}, \overline{p}]$ and customers observe the posted price p_t. The random demand D_t in period t is generated according to the following demand function: $D_t(p_t) = f(p_t) + g(x_t) + \varepsilon_t$, where $f(\cdot)$ and $g(\cdot)$ are unknown functions, and $\{\varepsilon_t\}_{t\geq 1}$ are i.i.d. mean-zero σ^2-sub-Gaussian r.v.'s, i.e., $\mathbb{E}[e^{\lambda\varepsilon_t}] \leq e^{\frac{\lambda^2\sigma^2}{2}}$ for any $\lambda \in \mathbb{R}$. The expected revenue under price p and context x_t is denoted as $r(p,x_t) := p(f(p) + g(x_t))$. The optimal price for period t is $p_t^* := \arg\max_{p\in[\underline{p},\overline{p}]} r(p,x_t)$. After observing the demand realization D_t in period t, the seller collects the revenue $p_t D_t$. An *admissible* pricing policy π is defined as a sequence of functions $\{\pi_t\}_{t\geq 1}$, where each $\pi_t(\cdot)$ maps the historical information observed up to the beginning of period t, denoted by vector $H_t = (x_1, p_1, D_1, \ldots, x_{t-1}, p_{t-1}, D_{t-1}, x_t)$, and possibly some external randomness to a feasible price in $[\underline{p}, \overline{p}]$. The performance of a pricing policy π is measured by *regret* $R^\pi(T)$, defined as the difference between the T-period expected revenue generated by the clairvoyant optimal policy and the pricing policy π. That is, $R^\pi(T) = \sum_{t=1}^T \mathbb{E}[r(p_t^*, x_t) - r(p_t, x_t)]$.

Intuitively, if no "smoothness" assumption is made on $f(p)$ and $g(x)$, the values of the two functions can change arbitrarily when p and x change, making the learning task virtually impossible. Thus, one needs appropriate structural assumptions on $f(p)$ and $g(x)$ so that the problem is learnable. As will be seen later, we systematically study three separable demand models under different structures of $f(p)$ and $g(x)$. Our central research question is: *what is the statistical complexity of the learning problem for each model considered?* To address this question, we first need to design an efficient online learning algorithm with a provable regret upper bound. One key challenge is

that although the underlying demand model has a separable structure, the seller can only observe the total demand, without knowing the exact contributions of $f(p)$ and $g(x)$. In order to maximize the expected revenue, the algorithm needs to identify both $f(p)$ and $g(x)$. Moreover, we also need to establish regret lower bounds for any admissible policies. As to this task, the key challenge lies in different structural properties of $f(p)$ and of $g(x)$ in each model. Accordingly, one needs to construct different hard instances satisfying the structure properties of $f(p)$ and of $g(x)$, and these instances should also adapt to some smoothness parameters for the function class of interest.

1.2. Main Results and Contributions

The main contributions of this paper lie in systematically characterizing and understanding the statistical complexity of the context-based dynamic pricing problem with separable demand models. Specifically, we consider three configurations of demand models under different structural properties of $f(p)$ and $g(x)$. Table 2 provides a summary for our main results and the comparison to the most relevant literature. Throughout this paper, the notations $\mathcal{O}(\cdot)$, $\Omega(\cdot)$ and $\Theta(\cdot)$ are used to hide constant factors, and $\widetilde{\mathcal{O}}(\cdot)$, $\widetilde{\Omega}(\cdot)$ and $\widetilde{\Theta}(\cdot)$ are used to hide both constant and logarithmic factors. We next discuss our results in more details.

Separable model with linear pricing effect (SMLPE). In this model, we assume $f(p)$ is linear and $g(x)$ is k_gth-order smooth with smoothness constant L. We develop a learning algorithm that melds the ideas of pricing with random shock borrowed from Nambiar et al. (2019) and contextual space binning and local polynomial approximation. For any $k_g > 0$, we prove a regret upper bound $\widetilde{\mathcal{O}}(\sqrt{T} \vee (L^2 T)^{\frac{d}{d+2k_g}})$ for this algorithm. For $0 < k_g \leq 1$, we establish a matching regret lower bound that must be incurred by any algorithm. This guarantees the rate-optimality of our algorithm in terms of the dependence on both L and T in the regime of $0 < k_g \leq 1$, which is exactly the class of commonly assumed Hölder continuous functions.

Separable model with linear contextual effect (SMLCE). In this model, we assume $f(p)$ is k_fth-order smooth with smoothness constant δ and $g(x)$ is linear. We devise a learning algorithm based on a local polynomial approximation borrowed from Wang et al. (2021b) and a new idea of optimism over OFU (Optimism in the Face of Uncertainty) for biased linear contextual bandit. For any $k_f > 0$, we prove that this algorithm attains a regret upper bound $\widetilde{\mathcal{O}}(\sqrt{T} \vee (\delta T^{k_f+1})^{\frac{1}{2k_f+1}})$, and also establish a matching lower bound $\Omega(\sqrt{T} \vee (\delta T^{k_f+1})^{\frac{1}{2k_f+1}})$. This guarantees the rate-optimality of our algorithm in terms of the dependence on both δ and T.

Separable model with non-parametric effects (SMNPE). In this model, we assume $f(p)$ is k_fth-order smooth with smoothness constant δ and $g(x)$ is k_gth-order smooth with smoothness constant L. Our algorithm hybrids the ideas of exploration-then-exploitation and local polynomial approximation. For any $k_g > 0$, we prove that the algorithm has the regret upper bound

$\widetilde{\mathcal{O}}(T^{\frac{3}{4}} \vee (L^d T^{d+k_g})^{\frac{1}{d+2k_g}})$. We also establish the regret lower bound $\Omega((L^d T^{d+k_g})^{\frac{1}{d+2k_g}})$ for $0 < k_g \leq 1$. Therefore, our algorithm almost attains the best-achievable regret rate in the regime of $0 < k_g \leq 1$.

Highlights of technical contributions. The technical contributions of this work are summarized in two aspects: (i) designing efficient learning algorithms with provable regret upper bounds, and (ii) constructing hard instances that adapt to different structural properties of $f(p)$ and $g(x)$ in the analysis of regret lower bounds. We give the highlights in the next two paragraphs.

Similar to many online learning problems, designing an efficient learning algorithm requires carefully balancing the *exploration-exploitation* trade-off. Morever, our problem faces an additional challenge – the seller can only observe the aggregate demand $D_t(p_t)$, but has no access to the exact contributions $f(p_t)$ and $g(x_t)$ make respectively to $D_t(p_t)$. Furthermore, the estimation errors of $f(p)$ and $g(x)$ are coupled, making it more difficult to analyze the regret upper bound. In SMLPE, we adopt the random shock pricing, an idea borrowed from Nambiar et al. (2019), which enables a direct estimation for $f(p)$ without knowing or accurately estimating $g(x)$. In SMNPE, in the exploration phase, we conduct randomized price experiments to directly estimate $f(p)$. Since SMLPE and SMNPE share the same structure of a high-order smooth function $g(x)$, they also apply the same technique of a local polynomial approximation. However, we can only obtain a biased observation of $g(x)$, because we only possess an estimator of $f(p)$ rather than its exact value. To address this challenge, we introduce a novel bridging optimization problem and a carefully designed clean event analysis to control the impact of the estimation error of $f(p)$ on the local high-order polynomial approximation of $g(x)$. In SMLCE, we take a different route than SMLPE and SMNPE by simultaneously estimating $f(p)$ and $g(x)$. This is achieved by leveraging the parametric property of $g(x)$. To chase the context-dependent optimal price in each period, we also propose a new idea of optimism over OFU for biased linear contextual bandit. This represents an additional layer of complexity compared with Wang et al. (2021b) where the authors study a dynamic pricing problem with a similar assumption on $f(p)$ but without contexts.

The analysis of the regret lower bound in each model involves constructing a series of hard problem instances so that any algorithm "performing well" in some problem instances would incur "high" regret in others. Although the existing studies, e.g., Rigollet and Zeevi (2010), Chen and Gallego (2021), Hu et al. (2022) and Wang et al. (2021b), also face a similar task, the construction of the hard instances in our setting is non-trivial due to two reasons: (i) the smoothness parameters k_g and k_f are not necessarily integer numbers, and (ii) in all three models, we hope to tightly characterize the impacts of constants L and δ in the regret bounds, which appears to be new in the dynamic pricing literature.

Table 2 Summary of the main results in this work and existing results in literature

Paper	Demand model	Regret upper bound	Regret lower bound	Key assumption
Dynamic pricing without context				
Besbes and Zeevi (2012)	$f(p)$	$\widetilde{\mathcal{O}}(T^{\frac{2k+1}{3k+1}})$	N/A	kth-order smoothness, concave revenue function
Keskin and Zeevi (2014)	$bp+a$	$\widetilde{\mathcal{O}}(\sqrt{T})$	$\Omega(\sqrt{T})$	Linear demands
Wang et al. (2021b)	$f(p)$	$\widetilde{\mathcal{O}}(T^{\frac{k+1}{2k+1}})$	$\Omega(T^{\frac{k+1}{2k+1}})$	kth-order smooth
Dynamic pricing with context				
Qiang and Bayati (2016)	$bp+a^\top x$	$\mathcal{O}(\log T)$	$\Omega(\log T)$	Known incumbent price
Nambiar et al. (2019)	$bp+g(x)$	$\mathcal{O}(\sqrt{T})$	$\Omega(\sqrt{T})$	The benchmark is the best linear model
Slivkins (2011)	$f(p,x)$	$\mathcal{O}(T^{\frac{d+2}{d+3}})$	$\Omega(T^{\frac{d+2}{d+3}})$	f: Lipschitz continuous
Chen and Gallego (2021)	$f(p,x)$	$\mathcal{O}((\log T)^2 T^{\frac{d+2}{d+4}})$	$\Omega(T^{\frac{d+2}{d+4}})$	f: Lipschitz continuous, smooth and locally concave revenue function
This work (Sec. 2 to 4)	$bp+g(x)$	$\widetilde{\mathcal{O}}(\sqrt{T}\vee(L^2T)^{\frac{d}{d+2k_g}})$	$\Omega(\sqrt{T}\vee(L^2T)^{\frac{d}{d+2k_g}})^\dagger$	g: k_gth-order smooth in $[0,1]^d$, f: k_fth-order smooth, where $k_g,k_f>0$
	$f(p)+a^\top x$	$\widetilde{\mathcal{O}}(\sqrt{T}\vee(\delta T^{k_f+1})^{\frac{1}{2k_f+1}})$	$\Omega(\sqrt{T}\vee(\delta T^{k_f+1})^{\frac{1}{2k_f+1}})$	
	$f(p)+g(x)$	$\widetilde{\mathcal{O}}(T^{\frac{3}{4}}\vee(L^dT^{d+k_g})^{\frac{1}{d+2k_g}})$	$\Omega((L^dT^{d+k_g})^{\frac{1}{d+2k_g}})^\dagger$	

† In this lower bound, we assume that $0<k_g\le 1$.

1.3. Literature Review

Li et al. (2023), together with their industrial practices, is the main motivation for us to consider the separable structure in revenue management. However, Li et al. (2023) mainly focus on the offline learning regime, where the data has been collected and the decision maker has no control over what kind of data they can see. Our paper lies in the regime of online learning, where the decision maker does not have any reliable historical data and need to learn on the fly. Thus, this paper is related to two streams of literature: dynamic pricing with demand learning, and bandits with contextual information. We next review the relevant works in each stream.

Dynamic pricing with demand learning. There is a vast literature on dynamic pricing with online demand learning, see two recent reviews, Chen et al. (2022) and Chen and Hu (2023). Earlier studies in this area consider the setting without contextual information. These works can be classified into two categories according to parametric models (see, e.g., Broder and Rusmevichientong 2012, Keskin and Zeevi 2014, Keskin et al. 2022a) and non-parametric models (see, e.g., Besbes and Zeevi 2009, Besbes and Zeevi 2012, Wang et al. 2014, Miao and Wang 2021, Wang et al. 2021b,

Li and Zheng 2023). We refer the reader to den Boer (2015) for a comprehensive review on this stream of literature.

Among the aforementioned studies, Wang et al. (2021b) is closely related to this work. Specifically, Wang et al. (2021b) consider a dynamic pricing problem without context, and assume the expected demand is $(k-1)$-times differentiable in the price, with the $(k-1)$-th derivative being Lipschtiz continuous. The authors design a learning algorithm by applying the OFU principle from linear contextual bandit, and prove that the optimal regret rate is $\widetilde{\Theta}(T^{\frac{k+1}{2k+1}})$. The algorithmic design and regret upper bound analysis of our SMLCE are inspired by their work, but there are two important differences. First, the optimal price is stationary in Wang et al. (2021b) in the absence of context, but can change over time in our contextual setting. Therefore, the multi-armed bandit protocol proposed in Wang et al. (2021b) to learn the fixed optimal price cannot be directly applied. To overcome this difficulty, we propose a novel idea of being more optimistic over OFU in price selection to chase the context-dependent optimal price. Second, the notion of k_fth-order smoothness in our work is a generalization to that in Wang et al. (2021b), where we allow k_f to be any real value while they assume k is an integer. Moreover, we study the effect of a smoothness parameter δ associated with $f(\cdot)$ on the regret bound, whereas Wang et al. (2021b) treat δ as a constant and ignores this effect. The general values of k_f and δ require a more careful selection of the parameters used by the algorithm, and a more sophisticated construction for the hard instances in the lower bound analysis.

There is a growing body of literature on context-based dynamic pricing with demand learning (e.g., Qiang and Bayati 2016, Javanmard and Nazerzadeh 2019, Miao et al. 2019, Cohen et al. 2020, Ban and Keskin 2021, Wang et al. 2021a, Keskin et al. 2022b). We refer to Ban and Keskin (2021) for a recent review, and next focus on the papers that are more related to ours. Qiang and Bayati (2016) consider the demand model $bp+a^\top x$ in an incumbent-price setting, i.e., the expected demand for some incumbent price is exactly known, and show that the greedy algorithm achieves the regret upper bound $\mathcal{O}(\log T)$. Nambiar et al. (2019) study the demand model $bp+g(x)$ and propose a random price shock (`RPS`) algorithm that generates randomized price shocks to estimate price elasticity. They prove that the regret of the `RPS` algorithm compared with the clairvoyant who knows the best *linear* approximation to $g(x)$ is in the order of $\mathcal{O}(\sqrt{T})$. In this paper, we borrow the idea of `RPS` to construct a learning algorithm for our SMLPE, but the regret we consider for the algorithm is benchmarked against the clairvoyant that knows the *true* demand function. Ban and Keskin (2021) consider *personalized* dynamic pricing where the price elasticity is affected by each individual customer's characteristics. Their demand model $g(\alpha^\top x+\beta^\top xp)$ does not admit a separable structure, since the price elasticity is context-dependent. By contrast, we focus on pricing for the whole customer population in different periods with dynamic features relevant to

products. Chen and Gallego (2021) consider a non-parametric demand model in context-based dynamic pricing. Assuming that the revenue function is Lipschitz continuous and locally concave, the authors prove that the optimal regret is $\widetilde{\Theta}(T^{\frac{d+2}{d+4}})$. Without leveraging separability structure, the algorithm in Chen and Gallego (2021) can lead to a sub-optimal regret in our setting. When $k_g = 1$, SMLPE becomes a special case of theirs, and the optimal regret in this case is improved to $\widetilde{\Theta}(\sqrt{T} \vee T^{\frac{d}{d+2}})$ due to the separable structure.

There is another stream of literature in context-based dynamic pricing, where the demand is formulated by a binary choice model (see, e.g., Javanmard and Nazerzadeh 2019, Mao et al. 2018, Cohen et al. 2020, Shah et al. 2019, Xu and Wang 2021, Fan et al. 2021, Luo et al. 2023). Adopting the different aggregate-demand modeling approach, our paper is very different from these works in both the algorithm development and regret analysis.

Bandits with contextual information. Our paper is also closely related to the literature studying bandits with contextual information. See Lattimore and Szepesvári (2018) and Slivkins (2019) for comprehensive reviews. The most studied model in contextual bandit is the linear model (see, e.g., Auer et al. 2002, Dani et al. 2008, Rusmevichientong and Tsitsiklis 2010, Chu et al. 2011, Abbasi-Yadkori et al. 2011), where the expected reward is a linear combination of contexts. The algorithms developed in these works are mostly built upon the celebrated idea of the OFU principle, which effectively balances the exploration-exploitation tradeoff. In SMLCE, we borrow the OFU idea to design a learning algorithm and apply the high-probability confidence bound constructed by Abbasi-Yadkori et al. (2011) to analyze the regret upper bound. Later studies extend the linear model to generalized linear model (see, e.g., Filippi et al. 2010, Li et al. 2012, Li et al. 2017). There are also a substantial amount of literature considering contextual bandits with non-parametric reward feedback under a general Hölder continuous assumption (see, e.g., Rigollet and Zeevi 2010, Perchet and Rigollet 2013, Hu et al. 2022). Among the studies on non-parametric bandits, Slivkins (2011) assuming a continuous action space is the most relevant to this work. For general Lipschitz reward function $f(x,p)$, where x is the context vector in $\mathbb{R}^d$ and p is the action vector in $\mathbb{R}^{d_p}$, Slivkins (2011) proves that the optimal regret is $\widetilde{\Theta}(T^{\frac{d+d_p+1}{d+d_p+2}})$. In particular, letting $d_p = 1$ leads to the optimal regret rate $\widetilde{\Theta}(T^{\frac{d+2}{d+3}})$ for a dynamic pricing problem. This rate is strictly higher than the optimal regret $\widetilde{\Theta}(T^{\frac{d+1}{d+2}})$ in our SMNPE with $k_g = 1$ and $d \geq 2$, which shows the benefit of the separable demand structure in reducing the statistical complexity.

1.4. Structure and Notations

Throughout this paper, we define $a \wedge b \triangleq \min\{a,b\}$ and $a \vee b \triangleq \max\{a,b\}$ for $a,b \in \mathbb{R}$, and use $[n]$ to denote the set $\{1,2,\cdots,n\}$ for any positive integer n. We use "context" and "feature" interchangeably. In Sec. 2 and Sec. 3, we study the semi-parametric model with linear pricing effect

and linear contextual effect respectively. In Sec. 4, we consider the non-parametric model. In Sec. 5, we provide further discussions on the comparison between the three models in Sec. 2, Sec. 3 and Sec. 4. A numerical study will be conducted in Sec. 6. We conclude this paper in Sec. 7.

2. Separable Model with Linear Pricing Effect (SMLPE)

In this section, we study the following separable demand model with linear pricing effect:

$$D_t(p) = bp + g(x_t) + \varepsilon_t, \qquad \forall p \in [\underline{p}, \overline{p}], \tag{1}$$

where b is an unknown parameter belonging to some known interval $[\underline{b}, \overline{b}] \subseteq (-\infty, 0)$. As a result, the revenue function $r(p, x_t) = p(bp + g(x_t))$ considered in this section is strongly concave in p. Following the literature, e.g., Ban and Keskin (2021), we assume that the optimal price $p_t^* = -\frac{g(x_t)}{2b} \in [\underline{p}, \overline{p}]$ for any $b \in [\underline{b}, \overline{b}]$ and $x_t \in [0,1]^d$. We now introduce our assumption on $g(\cdot)$. First, we define $\mathfrak{b}(k) := \sup\{i \in \mathbb{N} : i < k\}$ is the largest integer that is strictly less than k. For $\kappa = (\kappa_1, \kappa_2, \ldots, \kappa_d) \in \mathbb{N}^d$, we define $|\kappa| := \kappa_1 + \cdots + \kappa_d$, and $\partial^\kappa g = \partial_1^{\kappa_1} \partial_2^{\kappa_2} \cdots \partial_d^{\kappa_d} g := \frac{\partial^{|\kappa|} g}{\partial x_1^{\kappa_1} \partial x_2^{\kappa_2} \cdots \partial x_d^{\kappa_d}}$.

Assumption 1. *The function $g : [0,1]^d \to \mathbb{R}$ is k_gth-order smooth with constant $L > 0$, denoted by $g \in \mathcal{G}_d(k_g, L)$, if $g(\cdot)$ is $\mathfrak{b}(k_g)$-times differentiable on $[0,1]^d$ and for any $\kappa \in \mathbb{N}^d$ with $|\kappa| = \mathfrak{b}(k_g)$ and any $x_1, x_2 \in [0,1]^d$,*

$$|\partial^\kappa g(x_1) - \partial^\kappa g(x_2)| \le L \|x_1 - x_2\|^{k_g - \mathfrak{b}(k_g)}. \tag{2}$$

We refer to k_g and L as the smoothness degree and constant for $g(\cdot)$ respectively. Eq. (2) tells that $\mathfrak{b}(k_g)$th-order derivatives of $g(\cdot)$ are $(k_g - \mathfrak{b}(k_g))$-Hölder continuous. There are also some other almost equivalent definitions of smooth functions, see, e.g., Hu et al. (2022). In the above definition, k_g can be any non-negative real number and does not need to be an integer. When $k_g = 1$, $\mathcal{G}_d(k_g, L)$ is the class of Lipschitz continuous functions, which is often assumed in dynamic pricing literature (see, e.g., Chen and Gallego 2021). When $k_g < 1$, $\mathcal{G}_d(k_g, L)$ is the class of Hölder continuous functions. In this paper, we consider a general class of smooth functions to capture the effect of context. This assumption implies that the volumes of demands are similar if the contexts are similar. If this assumption fails, the historical sales data of one observed context is not informative for that of a new similar context, and thus learning is virtually impossible. An immediate property guaranteed by Assumption 1 is that there exists some constant $\bar{g} \ge 0$ such that $|g(x)| \le \bar{g}$ for any $x \in [0,1]^d$. Note that we give a special emphasis on the smoothness constant L for $g(\cdot)$, whose impact is usually overlooked in the regret analysis by the current literature. Nevertheless, as L affects how effectively a non-parametric function $g(\cdot)$ can be approximated by a polynomial function, we will study its impact on the regret bounds later.

2.1. Algorithm and Regret Upper Bound

In this subsection, we propose an Algorithm for SMLPE (`ASMLPE` for short) in Algorithm 1. As highlighted in Sec. 1.2, one key challenge in designing an algorithm for a separable demand model is the task of estimating both functions of $f(p)$ and $g(x)$ by using only the aggregate demand observations. For SMLPE, we combine two ideas: (i) pricing with random shock, and (ii) context space binning and local polynomial approximation, to address this challenge. Below we illustrate the details. For notation convenience, for any two vectors $x=(x(1),x(2),\ldots,x(d))\in[0,1]^d$ and $v=(v_1,v_2,\ldots,v_d)\in\mathbb{N}^d$, we denote $v!=\Pi_{i=1}^d v_i!$ and $x^v=\Pi_{i=1}^d(x(i))^{v_i}$. In Algorithm 1, we also adopt the notation $\phi_{\mathbf{M}_j}(x)=(1,\cdots,(x-x_j)^u,\cdots)$, where u takes all vectors in $\mathbb{N}^d$ satisfying $|u|\le\mathfrak{b}(k_g)$.

Pricing with random shock. Since $g(x)$ is unknown and may not be linear, a naive linear regression of d_t against (p_t,x_t) causes the price endogeneity effect and introduces bias to the estimation of price sensitivity b (see Nambiar et al. 2019). Even if $g(x)$ is a known linear function, directly applying the estimator to charge a myopic price suffers from a lack of exploration and can lead to insufficient learning. The technique we apply is called pricing with random shock, which is borrowed from Nambiar et al. (2019) and resolves the two drawbacks of linear regression simultaneously. Instead of regressing d_t against p_t, we estimate b by regressing d_t against the random shock Δ_t, which is an exogenous random variable with zero mean (see line 17 of Algorithm 1). Since

$$\frac{\sum_{s=1}^t\Delta_s d_s}{\sum_{s=1}^t\Delta_s^2}=\frac{\sum_{s=1}^t\Delta_s(bp_s+g(x_s)+\varepsilon_s)}{\sum_{s=1}^t\delta_s^2}=b+\frac{\sum_{s=1}^t\Delta_s(bp_s^g+g(x_s)+\varepsilon_s)}{\sum_{s=1}^t\delta_s^2},\tag{3}$$

$\frac{\sum_{s=1}^t\Delta_s d_s}{\sum_{s=1}^t\Delta_s^2}$ is an unbiased estimate of b. Note that this step estimates b even without any knowledge of $g(x)$, and therefore decouples the estimation error of $g(x)$ from that of b. Moreover, Eq. (3) takes advantages of all the data collected to estimate b regardless of where x_s lies, which leads to a high data utilization efficiency. Based on the estimate $\hat{b}_t$ and a local polynomial approximation for $g(x)$ to be explained later, the algorithm computes a greedy price p_t^g (see lines 11 and 12 of Algorithm 1) and charges a price p_t by adding the random shock Δ_t (see line 13 of Algorithm 1). A careful control of the magnitude of Δ_t balances the fundamental tradeoff between exploration and exploitation.

Context space binning and local polynomial approximation. Assumption 1 guarantees that $g(\cdot)$ cannot change dramatically in a local area. This leads to a natural idea of dividing the context space $[0,1]^d$ into different small bins and using a $\mathfrak{b}(k_g)$th order polynomial function to approximate $g(\cdot)$ in each bin. Specifically, we divide the context space $[0,1]^d$ into M^d equal-sized small bins. For each bin $\mathbf{M}_j$, we choose a fixed point $x_j\in\mathbf{M}_j$ and approximate $g(x)$ using its $\mathfrak{b}(k_g)$th order Taylor expansion at x_j, which is denoted by $P_{\mathbf{M}_j}(x)$ and defined as follows:

$$P_{\mathbf{M}_j}(x)=\sum_{i=0}^{\mathfrak{b}(k_g)}\sum_{\kappa\in\mathbb{N}^d:|\kappa|=i}\frac{\partial^\kappa g(x_j)}{\kappa!}(x-x_j)^\kappa.$$

To estimate the unknown coefficients of $P_{\mathbf{M}_j}(x)$ for each bin $\mathbf{M}_j$, we apply the linear regression to the residuals $\{d_s - p_s\hat{b}_t : 1 \le s \le t-1, x_s \in \mathbf{M}_j\}$ (see line 10 of Algorithm 1). However, there are two notable challenges. First, there exists an important tradeoff for the choice of M. If M is too large, the amount of data collected for each bin will be limited, discouraging the success of learning $g(\cdot)$. If M is too small, a larger approximation error of $g(\cdot)$ using a polynomial function $P_{\mathbf{M}_j}(x)$ will be incurred, leading to a poor pricing strategy. In fact, one can easily verify the following equation from Assumption 1:

$$\left|g(x) - P_{\mathbf{M}_j}(x)\right| = \left|\sum_{\kappa\in\mathbb{N}^d:|\kappa|=\mathfrak{b}(k_g)} \left(\frac{\partial^\kappa g(x_j')}{\kappa!}(x-x_j)^\kappa - \frac{\partial^\kappa g(x_j)}{\kappa!}(x-x_j)^\kappa\right)\right| = \mathcal{O}\left(\frac{L}{M^{k_g}}\right). \quad (4)$$

The choice of M is given in Theorem 1. Second, unlike the estimation process for b through random shock which is not affected by the quality of the estimate for $g(x)$, the estimation process for $g(x)$ from the local polynomial approximation and linear regression is subject to the estimation error for b. In particular, the residual $d_s - p_s\hat{b}_t$ is a biased observation of $g(x_s)$ since $\hat{b}_t \neq b$. The expectation bound of $(\hat{b}_t - b)^2$ established by Nambiar et al. (2019) does not suffice for our purpose, and we establish a more powerful high probability bound of $(\hat{b}_t - b)^2$ for every t. Building upon this, we introduce a novel bridging optimization problem aiming at managing the impact of the propagation of the estimation error of b to learning the function $g(x)$.

THEOREM 1. *Suppose Assumption 1 holds for demand function (1) and let Algorithm 1 run with* $M = \lceil (L^2T)^{\frac{1}{d+2k_g}} \rceil$. *Then the regret of Algorithm 1 is*

$$\widetilde{\mathcal{O}}\left(\sqrt{T} \vee (L^2T)^{\frac{d}{d+2k_g}}\right). \quad (5)$$

The upper bound in (5) consists of two parts $\widetilde{\mathcal{O}}(\sqrt{T})$ and $\widetilde{\mathcal{O}}((L^2T)^{\frac{d}{d+2k_g}})$. At a high level, $\widetilde{\mathcal{O}}(\sqrt{T})$ arises from the complexity of learning the price sensitivity b. For the simplest linear demand model without context, the squared estimation error of b under an asymptotically optimal policy exhibits a $t^{-1/2}$ order of magnitude (see, e.g., Keskin and Zeevi 2014). Therefore, a cumulative regret $\mathcal{O}(\sqrt{T})$ is in general unavoidable. The second term $\widetilde{\mathcal{O}}((L^2T)^{\frac{d}{d+2k_g}})$ captures the challenge of learning function $g(\cdot)$ within a high dimensional context space. This bound is increasing in d, consistent with the intuition that a higher dimension of context space leads to a more challenging task of learning $g(\cdot)$. It's also decreasing in k_g, aligned with the intuition that a larger value of k_g indicates a more accurate approximation of $g(\cdot)$ by a local polynomial function. The smoothness constant L also affects the order of the regret upper bound. If $L = \Theta(1)$ as assumed in the literature, the regret bound is always $\widetilde{\mathcal{O}}(T^{\frac{d}{d+2k_g}})$ except $d \le 2k_g$. However, if $L = \mathcal{O}(T^{\frac{k_g}{2d}-\frac{1}{4}})$, which can be very small when d is relatively large compared with k_g, the regret bound becomes $\widetilde{O}(\sqrt{T})$. In this case, the bottleneck becomes estimating b instead of $g(\cdot)$ because the latter can be very well approximated by a polynomial function. The proof of Theorem 1 is deferred to Appendix A.1.

Algorithm 1: Algorithm for Separable Model with Linear Pricing Effect (`ASMLPE`)

Input: price range $[\underline{p}, \overline{p}]$, bounds on the price coefficient $\underline{b}$ and $\overline{b}$, number of bins M, λ

Initialization:

Partition each dimension of the context $[0,1]$ into M segments of equal length, denoted as $\mathbf{M}_j$ for $j = 1, 2, \cdots, M^d$;

Select the center from every $\mathbf{M}_j$, denoted by $m_1, \cdots, m_{M^d}$;

Initialize $\mathcal{D}_{1,j} = \emptyset$ for each $j \in [M^d]$ and $\hat{b}_1 = \frac{\underline{b}+\overline{b}}{2}$;

Main Steps:

for $t = 1, 2, \cdots, T$ **do**

Set $\delta_t \leftarrow t^{-\frac{1}{4}}$;

Observe x_t and find $j \in [M^d]$ such that $x_t \in \mathbf{M}_j$;

$\hat{\theta}_{t,j} = (\lambda I + \sum_{(x_s,p_s,d_s)\in\mathcal{D}_{t,j}} \phi_{\mathbf{M}_j}(x_s)\phi_{\mathbf{M}_j}(x_s)^\top)^{-1} \sum_{(x_s,p_s,d_s)} (d_s - p_s\hat{b}_t)\phi_{\mathbf{M}_j}(x_s)$;

Set unconstrained greedy price: $p_t^u \leftarrow -\frac{\hat{\theta}_{t,j}^\top \phi_{\mathbf{M}_j}(x_t)}{2\hat{b}_t}$;

Project greedy price: $p_t^g \leftarrow \mathrm{Proj}(p_t^u, [\underline{p} + \delta_t, \overline{p} - \delta_t])$;

Generate an independent random variable $\Delta_t \leftarrow \delta_t$ w.p. $\frac{1}{2}$ and $\Delta_t \leftarrow -\delta_t$ w.p. $\frac{1}{2}$;

Set price $p_t \leftarrow p_t^g + \Delta_t$;

Observe realized demand d_t;

Update $\mathcal{D}_{t+1,j} \leftarrow \mathcal{D}_{t,j} \cup \{(x_t, p_t, d_t)\}$ and $\mathcal{D}_{t+1,i} \leftarrow \mathcal{D}_{t,i}$ for $i \neq j$;

Update $\hat{b}_{t+1} \leftarrow \mathrm{Proj}(\frac{\sum_{s=1}^t \Delta_s d_s}{\sum_{s=1}^t (\Delta_s)^2}, [\underline{b}, \overline{b}])$;

end for

2.2. Regret Lower Bound

We now establish a lower bound for SMLPE. We denote the regret of policy π as $R^\pi_{g,b,\mathcal{P},\mathcal{Q}}(T)$ under the demand function $bp + g(x) + \varepsilon$, where $g \in \mathcal{G}_d(k_g, L)$, and the distributions of x and ε are $\mathcal{P}$ and $\mathcal{Q}$ respectively. We use $\mathcal{E}(\sigma)$ to denote the class of σ^2-sub-Gaussian distributions.

THEOREM 2. *For $\mathcal{G}_d(k_g, L)$ with $0 < k_g \leq 1$ and $L > 0$, $[\underline{b}, \overline{b}] \subseteq (-\infty, 0)$ and $\sigma \geq 0$, there exists a constant $K_1 > 0$ independent of T and L, such that for any admissible policy π,*

$$\sup_{\substack{g\in\mathcal{G}_d(k_g,L),\\ b\in[\underline{b},\overline{b}],\mathcal{P},\mathcal{Q}\in\mathcal{E}(\sigma)}} R^\pi_{g,b,\mathcal{P},\mathcal{Q}}(T) \geq K_1 \cdot \left(\sqrt{T} \vee (L^2 T)^{\frac{d}{d+2k_g}}\right). \tag{6}$$

Theorem 2 shows that the regret upper bound achieved by Algorithm 1 in Theorem 1 is unimprovable in terms of the dependency on both T and L, for the case when $0 < k_g \leq 1$. To our knowledge, this is the first tight regret bound in the dynamic pricing literature in terms of the dependence of both T and L.

In Nambiar et al. (2019), a similar demand model is considered, but the performance of their learning algorithm is benchmarked with the clairvoyant optimal policy for the "best" linear demand function. Under such a setting, they show that the best achievable regret is $\Theta(\sqrt{T})$. By contrast, our regret notion is defined against the optimal policy endowed with knowledge of the *true* demand function. Consequently, our optimal regret encompasses an additional term $\widetilde{\Theta}((L^2T)^{\frac{d}{d+2k_g}})$, capturing the complexity for learning the function $g(\cdot)$. In Chen and Gallego (2021), the authors prove the optimal regret $\widetilde{\Theta}(T^{\frac{d+2}{d+4}})$ for general Lipschitz continuous demand functions (in both context and price) under a local concavity property. When $k_g = 1$ and $L = \Theta(1)$, the demand functions we consider here, represent a subset of their model. In this scenario, our optimal regret is $\widetilde{\Theta}(\sqrt{T} \vee T^{\frac{d}{d+2}})$, which is strictly lower than theirs. This reduction benefits from the separability structure assumed in our model. The improvement is more significant when d is small, but less significant as d increases. In terms of the algorithmic design, without utilizing the separability structure, their algorithm divides both the price space and context space into bins, and treats the learning problem in each small bin as an independent one. In our setting, we leverage all the historical data to estimate the price sensitivity to achieve information sharing among different bins in the context space.

We next sketch the proof of Theorem 2 and defer the detailed analysis to Appendix A.2. Note that the first lower bound $\Omega(\sqrt{T})$ is directly implied from the existing results (e.g., Theorem 1 in Keskin 2014) by choosing $g(x)$ to be a constant function. To show the second lower bound $\Omega((L^2T)^{\frac{d}{d+2k_g}})$, we borrow the idea from Rigollet and Zeevi (2010) and Chen and Gallego (2021) to construct a series of Hölder continuous functions that are "similar" to each other and difficult to distinguish. Additionally, we also need to construct the instance with the consideration of L. Specifically, we partition the context space $[0,1]^d$ into M^d equally sized bins, denoted as $\mathbf{M}_j$ for $j \in [M^d]$, by dividing each dimension into M intervals of equal length. We then construct a series of functions $\{g_{\mathbf{w}}(\cdot) : \mathbf{w} \in \{0,1\}^{M^d}\}$, each of which is indexed by a tuple $\mathbf{w} \in \{0,1\}^{M^d}$. The j-th coordinate of $\mathbf{w}$ determines the value of $g_{\mathbf{w}}(x)$ for $x \in \mathbf{M}_j$ as follows:

$$g_{\mathbf{w}}(x) = \begin{cases} |\underline{b}|\,(\underline{p}+\overline{p}) & \text{if } w_j = 0, \\ |\underline{b}|\,(\underline{p}+\overline{p}) + \frac{L}{2}(D(x,\partial\mathbf{M}_j))^{k_g} & \text{if } w_j = 1 \text{ and } D(x,\partial\mathbf{M}_j) \le \frac{1}{4^{1/k_g}M}, \\ |\underline{b}|\,(\underline{p}+\overline{p}) + \frac{L}{8M^{k_g}} & \text{if } w_j = 1 \text{ and } D(x,\partial\mathbf{M}_j) > \frac{1}{4^{1/k_g}M}, \end{cases} \tag{7}$$

where $\partial\mathbf{M}_j$ denotes the boundary of the bin $\mathbf{M}_j$, and $D(x,\mathbf{M}_j) := \inf\{\|x-y\| : y \in \partial\mathbf{M}_j\}$ denotes the Euclidean distance between x and $\partial\mathbf{M}_j$. Lemma 1 shows $g_{\mathbf{w}}(\cdot)$ is k_gth-order smooth if $0 < k_g \le 1$, whose proof will be given in Appendix A.3.

LEMMA 1. *For $0 < k_g \le 1$ and $L > 0$, let $M = \lceil (L^2T)^{\frac{1}{d+2k_g}} \rceil$. Then for each $\mathbf{w} \in \{0,1\}^{M^d}$, $g_{\mathbf{w}}(x)$ defined in Eq. (7) belongs to $\mathcal{G}_d(k_g, L)$.*

We then consider two demand functions $g_{(\mathbf{w}_{-j},w_j)}$ for $w_j=0,1$, where we use $(\mathbf{w}_{-j},w_j)$ to denote an index $\mathbf{w}\in\{0,1\}^{M^d}$ whose j-th coordinate is w_j and the other coordinates are $\mathbf{w}_{-j}$. When x falls into $\mathbf{M}_j$ with $w_j=0$, the optimal price is $\frac{p+\overline{p}}{2}$. For $w_j=1$, the optimal price is $\frac{p+\overline{p}}{2}+\frac{L}{16M^{k_g}|\underline{b}|}$. Thus, if the price charged by an algorithm in period t is greater than $\frac{p+\overline{p}}{2}+\frac{L}{32M^{k_g}|\underline{b}|}$, its gap with the optimal price under demand function $g_{(\mathbf{w}_{-j},0)}$ is greater than $\frac{L}{32M^{k_g}|\underline{b}|}$; and if the price charged by the algorithm is less than $\frac{p+\overline{p}}{2}+\frac{L}{32M^{k_g}|\underline{b}|}$, its gap with the optimal price under the other function $g_{(\mathbf{w}_{-j},1)}$ is still greater than $\frac{L}{32M^{k_g}|\underline{b}|}$. Bretagnolle–Huber inequality (see Bretagnolle and Huber 1979) guarantees that the minimal error of making one type of the mistakes depends on how well the algorithm can distinguish between the two demand functions. Then, by Kullback-Leibler (KL) divergence arguments, we guarantee that the minimal error can be lower bounded.

3. Separable Model with Linear Contextual Effect (SMLCE)

In this section, we study the following separable demand model with linear contextual effect:

$$D_t(p)=f(p)+a^\top x_t+\varepsilon_t,\qquad \forall p\in[\underline{p},\overline{p}], \tag{8}$$

where $a\in\mathbb{R}^d$ is an unknown vector capturing the linear contextual effect whose norm is bounded from above by a constant $\bar{a}>0$, i.e., $\|a\|\le\bar{a}$, and $f:[\underline{p},\overline{p}]\to\mathbb{R}$ is an unknown function capturing the pricing effect. In this section, we impose the k_fth-order smoothness assumption on $f(\cdot)$, which is similar to Assumption 1 under $d=1$. Recall that $\mathfrak{b}(k)=\sup\{i\in\mathbb{N}:i<k\}$ is the largest integer that is strictly less than k.

Assumption 2. *The function $f(\cdot):[\underline{p},\overline{p}]\to\mathbb{R}^+$ is k_fth-order smooth with constant $\delta>0$, denoted by $f\in\mathcal{F}(k_f,\delta)$, if $f(\cdot)$ is $\mathfrak{b}(k)$-times differentiable on $[\underline{p},\overline{p}]$ and for any $p,p'\in[\underline{p},\overline{p}]$,*

$$\left|f^{(\mathfrak{b}(k_f))}(p)-f^{(\mathfrak{b}(k_f))}(p')\right|\le\delta\left|p-p'\right|^{k_f-\mathfrak{b}(k_f)}. \tag{9}$$

Note that some commonly assumed smoothness conditions in the literature are covered by Assumption 2. For example, when $k_f=1$, $\mathcal{F}(k_f,\delta)$ is the class of Lipschitz continuous functions, as assumed in Besbes and Zeevi (2009) and Chen and Shi (2019). When $k=2$, $\mathcal{F}(k_f,\delta)$ includes all the functions with bounded second-order derivatives, as assumed in Wang et al. (2014), Besbes and Zeevi (2015), and Lei et al. (2014). When k_f is a general integer, Assumption 2 reduces to Assumption 1 in Wang et al. (2021b). We also emphasize that different from a variety of literature assuming concavity (see, e.g., Chen and Gallego 2021, Wang et al. 2014, Besbes and Zeevi 2015, Chen and Shi 2019), we do not make this assumption in this section. Again, we emphasize the dependency of this class of functions on the smoothness constant δ.

Algorithm 2: Algorithm for Separable Model with Linear Contextual Effect (`ASMLCE`)

Input: time horizon T, price range $[\underline{p}, \overline{p}]$, polynomial degree k_f, smoothness parameter C_0, context dimension d, upper bound $\bar{a}$, number of price segments N, error control terms Δ, noise variance proxy σ

Initialization:

Partition $[\underline{p}, \overline{p}]$ into N segments of equal length, denoted as $\mathbf{I}_j$ for $j = 1, 2, \cdots, N$;

Initialize for all $j \in [N]$: $\mathcal{D}_j = \emptyset$.

Main Steps:

for $t = 1, 2, \cdots, T$ **do**

Observe x_t;

for $i = 1, 2, \cdots, N$ **do**

$(\hat{\theta}_{t,i}, \hat{a}_{t,i}, V_{t,i}) = \texttt{RLC}\Big(\mathfrak{b}(k_f), d, \mathbf{I}_i, \mathcal{D}_i, \Delta, 1/T^2, C_0, \sigma, \bar{a}\Big)$;

$\gamma_{t,i} = \sigma\sqrt{(d + \mathfrak{b}(k_f)) \log\left(\frac{d+\mathfrak{b}(k_f)+|\mathcal{D}_i|}{d+\mathfrak{b}(k_f)}\right) + 4\log T} + \lambda^{\frac{1}{2}}(C_0^2 \mathfrak{b}(k_f) + \overline{a}^2)^{\frac{1}{2}} + \Delta\sqrt{|\mathcal{D}_i|}$;

$\hat{r}_{t,i} = \max_{p \in \mathbf{I}_i} p \times \Big(\langle \hat{\theta}_{t,i}, \varphi(p) \rangle + \langle \hat{a}_{t,i}, x_t \rangle + \gamma_{t,i}\sqrt{\phi(p, x_t)^\top V_{t,i}^{-1} \phi(p, x_t)} + \Delta\Big)$;

$\hat{p}_{t,i} = \arg\max_{p \in \mathbf{I}_i} p \times \Big(\langle \hat{\theta}_{t,i}, \varphi(p) \rangle + \langle \hat{a}_{t,i}, x_t \rangle + \gamma_{t,i}\sqrt{\phi(p, x_t)^\top V_{t,i}^{-1} \phi(p, x_t)} + \Delta\Big)$;

end for

Select $i_t = \arg\max_{i \le N} \hat{r}_{t,i}$ and charge $p_t = \hat{p}_{t,i_t}$;

Observe realized demand d_t;

Update $\mathcal{D}_{i_t} \leftarrow \mathcal{D}_{i_t} \cup \{(x_t, p_t, d_t)\}$;

end for

3.1. Algorithm and Regret Upper Bound

We construct an Algorithm for SMLCE (`ASMLCE` for short) in Algorithm 2. When $f(p)$ is k_fth-order smooth, we can naturally apply the idea of local polynomial approximation as what we did for $g(x)$ in Sec. 2. However, different from the context sequence $\{x_t : t \ge 1\}$ which is generated in an *i.i.d.* manner, the price sequence $\{p_t : t \ge 1\}$ is generated adaptively, posing more challenges in fitting $f(p)$ using a polynomial function. Moreover, the optimal price is context-dependent and changes over time. Therefore, the idea of embedding the biased linear contextual bandit into a multi-armed bandit protocol in Wang et al. (2021b) for dynamic pricing without context does not directly work. Algorithm 2 combines two ideas: (i) local polynomial approximation, and (ii) optimism over OFU for biased linear contextual bandit to address these challenges, which are illustrated below.

Local polynomial approximation. This idea is a one-dimensional-case implementation of the context space binning and local polynomial approximation in `ASMLPE`. Wang et al. (2021b) has also applied this idea to dynamic pricing without context. Specifically, we partition the price interval

Algorithm 3: Regression with Linear Context (`RLC`)

Input: polynomial degree k, context dimension d, domain $\mathbf{I}=[l,u]$, history $\mathcal{D}$, bias Δ, probability ϵ, smoothness parameter C, noise variance proxy σ, upper bound $\bar{a}$

Compute $\lambda=\frac{(u-l)^{2k}-1}{(u-l)^2-1}+d$;

Compute $V=\lambda I_{(k+d)\times(k+d)}+\sum_{(x,p,d)\in\mathcal{D}}\phi(p,x)\phi(p,x)^\top$, where $\phi(p,x)=(\varphi(p)^\top,x^\top)^\top$ and $\varphi(p)=(1,(p-l),\cdots,(p-l)^{k-1})^\top$;

Compute Ridge estimate $(\hat{\theta},\hat{a})=\arg\min\sum_{(x,p,d)\in\mathcal{D}}\left(d-\langle\theta,\varphi(p)\rangle-\langle a,x\rangle\right)^2+\lambda(\|\theta\|_2^2+\|a\|_2^2)$;

Output: $\hat{\theta},\hat{a},V$

$[\underline{p},\overline{p}]$ into N segments of equal size, denoted as $\mathbf{I}_1,\cdots,\mathbf{I}_N$ (see line 3 of Algorithm 2). For each price segment $\mathbf{I}_j:=[a_j,b_j)$, we use the following polynomial function of degree $\mathfrak{b}(k_f)$:

$$P_{\mathbf{I}_j}(p):=\sum_{i=0}^{\mathfrak{b}(k_f)}\frac{f^{(i)}(a_j)}{i!}(p-a_j)^i,$$

to locally approximate the true function $f(p)$. Similar as Eq. (4), the approximation error $|f(p)-P_{\mathbf{I}_j}(p)|$ is bounded by $\mathcal{O}(\delta/N^{k_f})$. Since the contextual effect is linear in this case, a polynomial function $P_{\mathbf{I}_j}(p)+a^\top x_t$ is used as a whole to approximate $f(p)+a^\top x_t$ in segment $\mathbf{I}_j$.

Optimism over OFU for biased linear contextual bandit. Applying the above local polynomial approximation, for each $t\geq 1$ and $p_t\in\mathbf{I}_j$, we rewrite $D_t(p)$ as follows:

$$D_t(p_t)=P_{\mathbf{I}_j}(p_t)+a^\top x_t+\beta_t=\theta_j^\top\varphi(p_t)+a^\top x_t+\beta_t, \tag{10}$$

where $\beta_t:=f(p_t)-P_{\mathbf{I}_j}(p_t)+\varepsilon_t$, $\varphi(p_t):=(1,(p_t-a_j),\cdots,(p_t-a_j)^{\mathfrak{b}(k_f)})$ and $\theta_j\in\mathbb{R}^{\mathfrak{b}(k_f)+1}$ whose i-th coordinate is $f^{(i)}(a_j)/i!$ for $0\leq i\leq\mathfrak{b}(k_f)$. Although the above model is quite similar to the linear contextual bandit in the literature (see, e.g., Abbasi-Yadkori et al. 2011, Chu et al. 2011), where the OFU idea is commonly adopted, there are two challenges in our problem. First, β_t in equation (10) contains a biased term $f(p_t)-P_{\mathbf{I}_j}(p_t)$ that is not mean-zero and depends on the pricing decision. We address this issue by borrowing the idea in Wang et al. (2021b) that adds an additional term Δ when implementing the OFU principle to compute an optimistic price in each segment. Second, as a unique challenge appearing in contextual dynamic pricing, the optimal price depends on the random context revealed in each period and changes over time. This is different from the setting in Wang et al. (2021b) where the optimal price remains a constant so that they can treat each price segment as an arm and the segment containing the *fixed* optimal price as the "best" arm. We overcome this by a new idea of the optimism over OFU, which proceeds as follows. We first implement the OFU principle in each price segment $\mathbf{I}_i$ and compute an optimistic price $\hat{p}_{t,i}$ and optimistic revenue $\hat{r}_{t,i}$ (see the for loop in lines 8 to 13 of Algorithm 2). Then we choose the

most optimistic price from the N candidate prices $\hat{p}_{t,1},\dots,\hat{p}_{t,N}$ that achieves the highest optimistic revenue (see line 14 of Algorithm 2). This idea helps to generate a price trajectory under which the algorithm's revenue in each period t is close to the true optimal revenue with high probability.

The following theorem presents the regret upper bound for Algorithm 2.

THEOREM 3. *Suppose Assumption 2 holds for demand function (8) and let Algorithm 2 run with* $N=\lceil \delta^{\frac{2}{2k_f+1}} T^{\frac{1}{2k_f+1}} \rceil +1$ *and* $\Delta=\delta(\overline{p}-\underline{p})^{k_f}/N^{k_f}$. *Then the regret of Algorithm 2 is*

$$\widetilde{\mathcal{O}}\left(d\left(\sqrt{T}\vee(\delta T^{k_f+1})^{\frac{1}{2k_f+1}}\right)\right).$$

The proof of Theorem 3 is deferred to Appendix B.1. We next discuss this result. First, when $\delta=\Theta(1)$, the regret upper bound is $\widetilde{\mathcal{O}}(dT^{(k_f+1)/(2k_f+1)})$, recovering Theorem 1 in Wang et al. (2021b) for the setting without contexts. This demonstrates that the additional linear contextual effect only brings a polynomial coefficient d to the regret bound. Similar to the impact of k_g in Theorem 5, this upper bound decreases in k_f. When $k_f=1$, $f(\cdot)$ is Lipschitz and the regret bound is $\widetilde{\mathcal{O}}(dT^{\frac{2}{3}})$, and when $k_f=\infty$, $f(\cdot)$ is infinitely differentiable and the regret bound becomes $\widetilde{\mathcal{O}}(d\sqrt{T})$. Second, when δ is not a constant, the regret bound behaves differently when δ belongs to different regimes. When $\delta=\Omega(T^{-\frac{1}{2}})$, the regret bound is $\widetilde{\mathcal{O}}(d((\delta T^{k_f+1})^{\frac{1}{2k_f+1}})$, and when δ decreases to $\mathcal{O}(T^{-\frac{1}{2}})$, the regret bound is always $\widetilde{\mathcal{O}}(d\sqrt{T})$. This shows that the smaller δ is, the lower the regret bound will be. The explanation is similar to the effect of L in Theorem 1. Third, in comparison to Theorem 1, the roles that dimension d plays are different in the regret bounds for SMLPE and SMLCE. In Theorem 1, d appears in the exponent of T due to the non-parametric assumption on $g(x)$, while in Theorem 3, d serves as a multiplicative factor in the regret bound due to the linear assumption on $g(x)$. This demonstrates that non-parametric contextual effect brings much more complexity to online learning than the parametric effect.

3.2. Regret Lower Bound

We now establish a regret lower bound for SMLCE. Similar to Theorem 2, we denote the regret of policy π by $R^{\pi}_{f,a,\mathcal{P},\mathcal{Q}}(T)$ under the demand function $f(p)+a^{\top}x+\varepsilon$, where $f\in\mathcal{F}(k_f,\delta)$, and the distributions of x and ε are $\mathcal{P}$ and $\mathcal{Q}$ respectively.

THEOREM 4. *For* $\mathcal{F}(k_f,\delta)$ *with* $k_f>0$ *and* $\delta>0$, $\bar{a}\geq 0$, *and* $\sigma\geq 0$, *there exists a constant* $K_2>0$ *independent of* T *and* δ, *such that for any admissible policy* π,

$$\sup_{\substack{f\in\mathcal{F}(k_f,\delta),\\ \|a\|\leq\bar{a},\mathcal{P},\mathcal{Q}\in\mathcal{E}(\sigma)}} R^{\pi}_{f,a,\mathcal{P},\mathcal{Q}}(T)\geq K_2\cdot\left(\sqrt{T}\vee\left(\delta T^{k_f+1}\right)^{\frac{1}{2k_f+1}}\right). \tag{11}$$

Theorem 4 shows that the regret upper bound achieved by Algorithm 2 in Theorem 3 is unimprovable in terms of its dependency on the learning horizon T and δ. Therefore, the optimal regret

rate for SMLCE is $\widetilde{\Theta}(\sqrt{T} \vee (\delta T^{k_f+1})^{\frac{1}{2k_f+1}})$. Note that the upper bound in Theorem 3 grows linearly in d, but the lower bound in Theorem 4 is independent of d. This is because in the construction of demand functions for the lower bound, we simple take $a = \mathbf{0}$. We leave the problems of analyzing the more complicated dimension-dependent lower bound as future research.

To prove Theorem 4, the first lower bound $\Omega(\sqrt{T})$ is directly implied from Theorem 1 in Keskin and Zeevi (2014) by letting $a = \mathbf{0}$ and $f(p) = \alpha + \beta p$ after appropriately choosing α and β. To show the second lower bound $\Omega((\delta T^{k_f+1})^{\frac{1}{2k_f+1}})$, we construct a series of demand functions that are in the class of $\mathcal{F}(k_f, \delta)$ and use the KL divergence arguments to bound the regret. Note that the smoothstep function constructed in Wang et al. (2021b) cannot be directly used here because in our problem k_f is not necessarily an integer. Besides, our analysis also requires the constructed demand functions to be dependent on δ such that the established lower bound achieves a tight dependency on δ, which is more complicated than that in Wang et al. (2021b).

We next describe how to construct the demand functions for our lower bound analysis. For simplicity, we assume $[\underline{p}, \overline{p}] = [1, 2]$. Similar to Hu et al. (2022), we introduce an infinitely differentiable function $u(x) = \exp(-\frac{1}{x(1-x)})\mathbf{1}\{x \in [0,1]\}$, and define function $S(x) := \left(\int_0^2 u(t)dt\right)^{-1} \int_{-\infty}^{x} u(t)dt$. Based on $S(x)$, we define $g_{k_f}(x)$ as follows:

$$g_{k_f}(x) = \frac{1}{Z_{k_f}}(S(x)\mathbf{1}\{x \le 1\} + S(2-x)\mathbf{1}\{x > 1\}), \quad \forall x \in [0,2], \tag{12}$$

where $Z_{k_f} > 0$ is a scaling parameter which guarantees that all the l-th derivatives of $g_{k_f}(x)$ are uniformly bounded on $[0,2]$ for $0 \le l \le \mathfrak{b}(k_f)$, and that $g_{k_f}^{(\mathfrak{b}(k_f))}(x)$ is $(k_f - \mathfrak{b}(k_f))$-Hölder continuous. To construct a series of demand functions, we partition the price range $[1,2]$ into J segments of equal length, denoted by $\mathbf{I}_1, \mathbf{I}_2, \ldots, \mathbf{I}_J$. For each $j = 0, 1, \ldots, J$, we let

$$f_j(p) := \begin{cases} \frac{\hat{\delta}}{2p} & \text{if } p \notin \mathbf{I}_j, \\ \frac{\frac{1}{2}\hat{\delta} + \eta \cdot g_{k_f}(2J(p-a_j))}{p} & \text{if } p \in \mathbf{I}_j, \end{cases} \tag{13}$$

where a_j is the left-end point of interval $\mathbf{I}_j$. We then establish the following lemma, whose proof is given in Appendix B.3.

Lemma 2. *Let* $\hat{\delta} = \delta / ((\sum_{i=0}^{\mathfrak{b}(k_f)} \frac{\mathfrak{b}(k_f)!}{i!}) \vee (\mathfrak{b}(k_f)+1)2^{\mathfrak{b}(k_f)-1})$, $J = \lceil 4(\mathfrak{b}(k_f)+1)2^{\mathfrak{b}(k_f)} \hat{\delta}^{\frac{2}{2k_f+1}} T^{\frac{1}{2k_f+1}} \rceil$ *and* $\eta = ((2\sigma) \wedge \frac{1}{2^{3k+1}}) \frac{1}{((\mathfrak{b}(k_f)+1)2^{\mathfrak{b}(k_f)})^{k_f}} \hat{\delta}^{\frac{1}{2k_f+1}} T^{-\frac{k_f}{2k_f+1}}$ *in Eq. (13). Then for each* $0 \le j \le J$, $f_j(p) \in \mathcal{F}(k_f, \delta)$.

Similar to Wang et al. (2021b), we establish the following inequalities: for each $j \in [J]$,

$$\left|\mathbb{E}_j^\pi[T_j] - \mathbb{E}_0^\pi[T_j]\right| \le \frac{1}{2} T \sqrt{\mathrm{KL}(\mathbb{P}_0^\pi(T_j) || \mathbb{P}_j^\pi(T_j))} \le \frac{1}{4\sigma} \sqrt{\mathbb{E}_0^\pi[T_j] T \eta}. \tag{14}$$

In Eq. (14), the first inequality is obtained by bounding $\left|\mathbb{E}_j^\pi[T_j] - \mathbb{E}_0^\pi[T_j]\right|$ via the total variation of $\mathbb{P}_0^\pi$ and $\mathbb{P}_j^\pi$ and applying Pinsker's inequality that relates the total variation of two probability measures with the KL divergence, and the second inequality is due to our construction of r_j. Letting $j^* := \arg\min_{1\le j\le J} \mathbb{E}_0^\pi[T_j]$, we must have $\mathbb{E}_0^\pi[T_{j^*}] \le \frac{T}{J}$. Since we set $\eta = \Theta((\delta T^{-k_f})^{\frac{1}{2k_f+1}})$ and $J = \Theta((\delta^2 T)^{\frac{1}{2k_f+1}})$, Eq. (14) guarantees that $\mathbb{E}_{j^*}^\pi[T_{j^*}] \le \frac{T}{2}$. This indicates that when the true revenue function is r_{j^*}, there are at least $\frac{T}{2}$ times when the selected prices do not fall into the "best" segment $\mathbf{I}_{j^*}$, leading to the regret loss $\Omega(T\eta) = \Omega((\delta T^{k_f+1})^{\frac{1}{2k_f+1}})$.

4. Separable Model with Non-Parametric Effects (SMNPE)

In this section, we study the general case, the fully non-parametric demand model with a separable structure. Specifically, the demand function is as follows:

$$D_t(p) = f(p) + g(x_t) + \varepsilon_t, \qquad \forall p \in [\underline{p}, \overline{p}]. \tag{15}$$

We state the assumptions made throughout this section.

ASSUMPTION 3. *$f(\cdot) \in \mathcal{F}(k_f, \delta)$ and $g(\cdot) \in \mathcal{G}_d(k_g, L)$.*

Without loss of generality, we also assume $\mathbb{E}[g(x)] = 0$ since if $\mathbb{E}[g(x)] \neq 0$, one can always absorb the constant $\mathbb{E}[g(x)]$ into $f(\cdot)$. SMLPE and SMLCE are two special cases of SMNPE considered here. When $k_f = 1$ and $k_g = 1$, SMNPE can be seen as a special case of Slivkins (2011) with an additional separable structure. Thus, one can expect our regret bound will not exceed theirs and will see how much benefit our separability structure will bring later in this section. Compared with Chen and Gallego (2021), we make the additional separability assumption, but do not assume local concavity of the revenue function as required in their paper.

4.1. Algorithm and Regret Upper Bound

In this section, we propose an Algorithm for SMNPE (`ASMNPE` for short) in Algorithm 4. The ideas of `ASMNPE` include: (i) context space binning and local polynomial approximation; (ii) pricing space discretization; and (iii) exploration then exploitation. The first idea is similar to that in Sec. 2 and thus omitted for brevity. We next illustrate ideas (ii) and (iii) in details.

Price space discretization. Since $f(\cdot)$ is non-parametric with an unknown structure, to learn this function, we adopt the idea of price space discretization to reduce our original pricing problem with a continuous action set to the one with a finite action set. Specifically, we divide $[\underline{p}, \overline{p}]$ into N segments of equal length, and take the midpoint of each segment to constitute a candidate price set $\bigcup_{i=1}^N \{\underline{p} + \frac{\overline{p}-\underline{p}}{N}(i - \frac{1}{2})\}$. Our algorithm then always selects the price from this candidate set. Again, a tradeoff exists when choosing parameter N to balance the discretization error and learning efficiency. The specific choice of N is given in Theorem 5.

Algorithm 4: Algorithm for Separable Model with Non-Parametric Effects (`ASMNPE`)

1 **Input:** time horizon T, price range $[\underline{p}, \overline{p}]$, context dimension d, number of bins for price N, parameter of bins for context M, exploration parameter n_0.
2 **Initialization:**
3 Define $P[i] = \underline{p} + \frac{\overline{p}-\underline{p}}{N}(i - \frac{1}{2})$ for each $1 \le i \le N$;
4 Initialize for all $1 \le i \le N$: $s_i = 0$ and $n_i = 0$;
5 Partition each dimension of context into M equal-length segments, denoted as $\mathbf{M}_j$, $j \in [M^d]$;
6 Initialize for all j where $j \le M^d$: $\mathcal{D}_{1,j} = \emptyset$.
7 **Main Steps:**
8 **for** $t = 1, 2, \cdots, n_0 N$ **do** `// Exploration phase`
9 Calculate $i = (t \mod N) + 1$ and charge price $p_t = P[i]$;
10 Observe realized demand d_t;
11 $s_i \leftarrow \frac{s_i \times n_i + d_t}{n_i + 1}$ and $n_i \leftarrow n_i + 1$;
12 **end for**
13 **for** $t = n_0 N + 1, n_0 N + 2, \cdots, T$ **do** `// Exploitation phase`
14 Observe x_t and find $j \in [M^d]$ such that $x_t \in \mathbf{M}_j$;
15 $\hat{\theta}_{t,j} = (\lambda I + \sum_{(x_s, p_s, d_s) \in \mathcal{D}_{t,j}} \phi_{\mathbf{M}_j}(x_s) \phi_{\mathbf{M}_j}(x_s)^\top)^{-1} \sum_{(x_s, p_s, d_s) \in \mathcal{D}_{t,j}} \sum_{i=1}^{N} (d_s - s_i) \mathbb{I}(p_s = P[i]) \phi_{\mathbf{M}_j}(x_s)$;
16 Select $m_t = \arg\max_{1 \le i \le N} P[i] \times (s_i + \hat{\theta}_{t,j}^\top \phi_{\mathbf{M}_j}(x_t))$ and charge $p_t = P[m_t]$;
17 Observe realized demand d_t;
18 Update $\mathcal{D}_{t+1,j} \leftarrow \mathcal{D}_{t,j} \cup \{(x_t, p_t, d_t)\}$ and $\mathcal{D}_{t+1,j'} \leftarrow \mathcal{D}_{t,j'}$ for $j' \ne j$;
19 **end for**

Exploration then exploitation. Our algorithm proceeds in an exploration-then-exploitation manner. In the exploration phase, each price $P[i]$ in the candidate set is charged for n_0 times, and $f(P[i])$ is estimated by taking the sample average of n_0 demand observations under price $P[i]$, denoted as s_i. The estimates $\{s_i\}_{1 \le i \le N}$ play two roles in the subsequent exploitation phase. First, for each period t in the exploitation phase, after observing the context x_t and the bin $\mathbf{M}_j$ it belongs to, we estimate $g(x_t)$ by computing the average of the residuals $\{d_k - s_k : (x_k, p_k, d_k) \in \mathcal{D}_{t,j}\}$, and these residuals are obtained only from the exploitation phase for periods when the context falls into the same bin $\mathbf{M}_j$. Second, we compute the proxy expected revenue for each candidate price $P[i]$ by replacing $f(P[i])$ with s_i and $g(x_t)$ with $\hat{\theta}_{t,j}^\top \phi_{\mathbf{M}_j}(x_t)$, and search for the best price in the candidate set that maximizes the proxy expected revenue (line 16). The exploration-then-exploitation method is a standard technique in bandit (see, e.g., Lattimore and Szepesvári 2018). It

also brings additional convenience in resolving our core challenge that only the aggregate demand can be observed. During the exploration phase, the prices are charged independently of all the contexts, which achieves fully randomized control and an estimation of $f(\cdot)$ without knowing $g(\cdot)$.

The following theorem shows the regret upper bound of Algorithm 4.

THEOREM 5. *Suppose Assumption 3 holds for demand function (15) and let Algorithm 4 run with* $N = \lceil T^{\frac{1}{4}} \rceil$, $n_0 = \lceil T^{\frac{1}{2}} \rceil$ *and* $M = \lceil (L^2 T)^{\frac{1}{d+2k_g}} \rceil$. *Then the regret of Algorithm 4 is*

$$\widetilde{\mathcal{O}}\left(T^{\frac{3}{4}} \vee \left(L^d T^{d+k_g} \right)^{\frac{1}{d+2k_g}} \right).$$

Note that the upper bound in Theorem 5 does not depend on the smoothness parameter k_f or δ of function $f(\cdot)$. In the next subsection, we will show that in almost all the cases, the upper bound in Theorem 5 is tight, meaning that even if $f(\cdot)$ becomes smoother, the regret bound cannot be further reduced and the complexity is determined by that of learning the function $g(\cdot)$.

As we have mentioned, when $k_g = 1$, our model reduces to a special case of Slivkins (2011) with an extra separability structure. For this case, our regret upper bound $\widetilde{\mathcal{O}}(T^{\frac{3}{4}} \vee T^{\frac{d+1}{d+2}})$ is no greater than the optimal regret rate $\widetilde{\Theta}(T^{\frac{d+2}{d+3}})$ in their paper, which is explained by our separability structure. When $d = 1$, our upper bound $\widetilde{\mathcal{O}}(T^{\frac{3}{4}})$ with separability structure is the same as that in Slivkins (2011) without separability structure. There are two possible reasons: our Algorithm 4 does not fully utilize the separability structure; or when contexts are univariate, the separability structure essentially cannot help to reduce the learning complexity.

4.2. Regret Lower Bound

We next establish a regret lower bound for SMNPE. For brevity, we denote the regret of policy π by $R^{\pi}_{f,g,\mathcal{P},\mathcal{Q}}(T)$ where $f \in \mathcal{F}(k_f, \delta)$, $g \in \mathcal{G}_d(k_g, L)$, and distribution of context x and random shock ε are $\mathcal{P}$ and $\mathcal{Q}$ respectively. Like before, $\mathcal{E}(\sigma)$ denotes the class of σ^2-sub-Gaussian distributions.

THEOREM 6. *For* $\mathcal{G}_d(k_g, L)$ *and* $\mathcal{F}(k_f, \delta)$ *with* $0 < k_g \le 1$, $L > 0$, $k_f > 0$ *and* $\delta > 0$, *there exists a constant* $K_3 > 0$ *independent of* T, *such that for any admissible policy* π,

$$\sup_{\substack{f \in \mathcal{F}(k_f,\delta), g \in \mathcal{G}_d(k_g,L), \\ \mathcal{P},\mathcal{Q} \in \mathcal{E}(\sigma)}} R^{\pi}_{f,g,\mathcal{P},\mathcal{Q}}(T) \ge K_3 \cdot \left(L^d T^{d+k_g} \right)^{\frac{1}{d+2k_g}}. \tag{16}$$

When $L = \Theta(1)$, the lower bound in Theorem 6 matches the upper bound in Theorem 5 as long as $d \ge 2$ or $k_g \le \frac{1}{2}$ and the optimal regret rate in this case equals $\widetilde{\Theta}(T^{\frac{d+k_g}{d+2k_g}})$. For the model studied in Section 2, when $L = \Theta(1)$ and $0 < k_g \le 1$, the optimal regret is $\widetilde{\Theta}(T^{\frac{d}{d+2k_g}})$, which is smaller than $\widetilde{\Theta}(T^{\frac{d+k_g}{d+2k_g}})$. This is due to the linear pricing effect and strong concavity of the revenue function assumed for the model of Section 2. For the model studied in Section 3, when $\delta = \Theta(1)$, the optimal regret is $\widetilde{\Theta}(T^{\frac{k_f+1}{2k_f+1}})$, which is also smaller than $\widetilde{\Theta}(T^{\frac{d+k_g}{d+2k_g}})$ when $k_f \ge 1$ and $k_g = 1$. This is because

the non-parametric form of $g(\cdot)$ in the model of this section usually governs the learning complexity. If $d=1$ and $k_g > \frac{1}{2}$, there is a gap between the upper bound $\widetilde{\mathcal{O}}(T^{\frac{3}{4}})$ in Theorem 5 compared with the lower bound $\Omega(T^{\frac{1+k_g}{1+2k_g}})$ in Theorem 6. Nevertheless, in the real-world business, the seller usually has more than one feature to utilize and $d=1$ is an extreme case that rarely happens in practice. It's also worth mentioning that when $L=\Omega(T^{\frac{k_g}{2d}-\frac{1}{4}})$, the lower bound in Theorem 6 also matches the upper bound in Theorem 5.

Now, we discuss the construction of the demand functions in the lower bound analysis and defer the complete proof of Theorem 6 to Appendix C.2. For simplicity, as before, we assume $[\underline{p},\overline{p}]=[1,2]$. Define $\hat{\delta}=\delta/((\sum_{i=0}^{\mathfrak{b}(k_f)}\frac{\mathfrak{b}(k_f)!}{i!})\vee(\mathfrak{b}(k_f)+1)2^{\mathfrak{b}(k_f)-1})$, $\eta=\Theta(T^{-\frac{k_g}{d+2k_g}})$, $J=\Theta((T^{\frac{k_g}{d+2k_g}})^{\frac{1}{k_f}})$, $M=\Theta((L^2T)^{\frac{1}{d+2k_g}})$. The exact definitions of η, J and M can be found in Appendix C.2. We can construct an eligible $f(p)$ as

$$f(p):=\begin{cases}\frac{\hat{\delta}}{2p} & \text{if } p\notin[\frac{3}{2}-\frac{1}{2J},\frac{3}{2}+\frac{1}{2J}],\\ \frac{\hat{\delta}+\eta\cdot g_{k_f}(2J(p-a_j))}{2p} & \text{if } p\in[\frac{3}{2}-\frac{1}{2J},\frac{3}{2}+\frac{1}{2J}],\end{cases} \tag{17}$$

where $g_k(\cdot)$ follows from the definition in Eq. (12). Lemma 2 guarantees that $f(p)\in\mathcal{F}^{k_f}([\underline{p},\overline{p}];\delta)$. To construct function $g(\cdot)$, similar to Sec. 2, we partition the context space $[0,1]^d$ into M^d equally sized bins, denoted as $\mathbf{M}_j$ for $j\in[M^d]$. For each $\mathbf{w}\in\{0,1\}^{M^d}$, we define $g_{\mathbf{w}}(x)$ for $x\in\mathbf{M}_j$:

$$g_{\mathbf{w}}(x)=\begin{cases}0 & \text{if } w_j=0,\\ \frac{L}{2}(D(x,\partial\mathbf{M}_j))^{k_g} & \text{if } w_j=1 \text{ and } D(x,\partial\mathbf{M}_j)\le(4\eta)^{\frac{1}{k_g}},\\ 4\eta & \text{if } w_j=1 \text{ and } D(x,\partial\mathbf{M}_j)>(4\eta)^{\frac{1}{k_g}},\end{cases} \tag{18}$$

where η should be chosen such that $(\frac{8\eta}{L})^{\frac{1}{k_g}}\le\frac{1}{2M}$, because otherwise, the third line of Eq. (18) is not well defined. From Lemma 1, we know that $g_{\mathbf{w}}(x)\in\mathcal{G}_d(k_g,L)$ for $k_g\le 1$.

Similar to Sec. 2, we consider two demand functions $g_{(\mathbf{w}_{-j},w_j)}$ for $w_j=0,1$. By our constructions, we can show that whenever x_t falls in $\mathbf{M}_j$, if $w_j=0$, the optimal price is $\frac{3}{2}$ and otherwise, the optimal price is 2. For any price that falls into $[\frac{3}{2}-\frac{1}{2J},\frac{3}{2}+\frac{1}{2J}]$, the regret under function $g_{(\mathbf{w}_{-j},1)}$ will be large; and for any price that does not belong to $[\frac{3}{2}-\frac{1}{2J},\frac{3}{2}+\frac{1}{2J}]$, the regret under $g_{(\mathbf{w}_{-j},0)}$ will not be negligible either. The remaining analysis is similar to that of Theorem 2 and we delay the details to Appendix C.2.

5. Discussions

In this section, we provide further discussions on the commonalities and differences between the three algorithms in Sections 2, 3 and 4 and their regret bounds. As we have previously discussed, the separable demand model with contexts presents two main challenges: balancing the exploration-exploitation trade-off in online learning and separately estimating $f(p)$ and $g(x)$ with only access

to the aggregate demand observations. In Table 3 below, we summarize how the three algorithms address these two challenges. In addition, one important aspect for understanding the performance of the proposed algorithm is the data utilization rate, or more specifically, how much historical data in each period are used to estimate $f(p)$ and $g(x)$. We also give a summary of this in Table 3. Since the first two models SMLPE and SMLCE are special cases of the third model SMNPE, we next compare the algorithms `ASMLPE` and `ASMLCE` for the first two models with `ASMNPE` for the third model respectively.

Table 3 Comparison of the three algorithms

	How to balance the exploration-exploitation trade-off?	How to separately estimate $f(p)$ and $g(x)$?	To what degree the historical data are utilized?
Sec. 2: $bp+g(x)$	Random shock	Random shock allows for estimating b without estimating $g(x)$	All the data are used to estimate b; The data in each bin are used to locally estimate $g(x)$
Sec. 3: $f(p)+a^\top x$	Optimism in the Face of Uncertainty	In each price segment, estimate $f(p)$ through polynomial approx. and $a^\top x$	The data in each price segment are used to locally estimate both $f(p)$ and $a^\top x$
Sec. 4: $f(p)+g(x)$	Explore-then-exploit	Separately estimate $f(p)$ in exploration phase and $g(x)$ in exploitation phase	Data in the exploration phase are used to estimate $f(p)$; Data in the exploitation phase are used to estimate $g(x)$;

Comparing SMLPE with SMNPE, both models assume the function $g(x)$ is non-parametric and k_gth-order smooth. Therefore, both `ASMLPE` and `ASMNPE` leverage the techniques of context space binning and local polynomial approximation to estimate $g(x)$ in each local bin. The key difference of the two algorithms arises from different assumptions made on the pricing effect. In SMLPE, the price is assumed to affect the demand linearly, which creates an opportunity for the algorithm to more effectively estimate function $f(p)$ and more sufficiently utilize the historical data. The random shock adopted by `ASMLPE` guarantees that the price sensitivity b can be estimated independently, preventing the estimation error of b from propagating due to poor estimation of $g(x)$ in the initial periods. Moreover, in each period t, the data collected from all historical periods $1, 2, \ldots, t-1$ can be utilized to estimate b, regardless of which bin each x_s falls into, leading to a high efficiency of data utilization. Unfortunately, without the linear price assumption, the idea of random shock does not work. For the model with non-parametric $f(p)$, `ASMNPE` adopts a common strategy in online learning called explore-then-exploit, by estimating $f(p)$ purely in the exploration phase and estimating $g(x)$ in the exploitation phase. In the exploration phase, `ASMNPE` conducts a randomized experiment on the discretized prices and uses the sample average approximation to estimate $f(p)$ under each discretized price p. The estimates for $f(p)$ remain unchanged in the exploitation phase, which may explain why an additional term $T^{3/4}$ appears in the regret upper bound.

Comparing SMLCE with SMNPE, both models assume the function $f(p)$ is non-parametric and k_fth-order smooth. In SMLCE, the linearity assumption of $g(x)$ greatly facilitates its estimation.

`ASMLCE` adopts the simple linear regression together with a polynomial approximation of $f(p)$ to jointly estimate $f(p)$ and a in each price segment. This is sufficient for achieving the optimal rate for learning $g(x)$. Moreover, the dimensionality d of the context space or the smoothness parameter of $g(\cdot)$ does not bring additional statistical complexity in the exponent of T. However, in SMNPE, one of the main challenges is to estimate the non-parametric function $g(x)$. In order to achieve the tight rate for learning the non-parametric function $g(x)$, `ASMNPE` trades the estimation accuracy of $f(p)$ for that of $g(x)$, by taking a more straightforward approach of splitting the exploitation phase from the exploration phase. It remains an open question how to improve the framework of `ASMNPE` and reduce the regret upper bound.

Finally, we compare the regret bounds for the three models. We mainly focus on the regime with $L=\Theta(1)$, $\delta=\Theta(1)$, $0<k_g\le 1$ and $d\ge 2$, where the optimal regrets for the three models are characterized by $\widetilde{\Theta}\Big(T^{\frac{d}{d+2k_g}}\Big)$, $\widetilde{\Theta}\Big(T^{\frac{k_f+1}{2k_f+1}}\Big)$, and $\widetilde{\Theta}\Big(T^{\frac{d+k_g}{d+2k_g}}\Big)$ respectively. Comparing SMLPE and SMLCE, we observe that their optimal regrets are both higher than $\widetilde{\Theta}(\sqrt{T})$. Intuitively, despite of the distinct demand structures, these two models include the fully parametric model $bp+a^\top x$ as a special case, which inevitably incurs $\widetilde{\Theta}(\sqrt{T})$ regret. Therefore, the complexity comes from tackling with the non-parametric component $g(x)$. When $d=1$, the optimal regret for SMLPE is characterized by $\widetilde{\Theta}\Big(T^{\frac{1}{1+2k_g}}\Big)$, and the role of k_g is different from that of k_f in $\widetilde{\Theta}\Big(T^{\frac{k_f+1}{2k_f+1}}\Big)$ for SMLCE. This distinction arises due to the inherent asymmetry between p and x. Specifically, p is an endogenous variable determined based on observed data, while x is an exogenous variable unaffected by historical data. Comparing SMLPE and SMNPE, the optimal regret of SMLPE is lower than that of SMNPE because of the concavity of the revenue function with respect to p in the former. Moreover, the expected revenue $p(f(p)+g(x))$ in SMNPE can even be a multi-modal function of p, which creates a substantial challenge for learning and optimizing. Comparing SMLCE and SMNPE, when $k_f\ge 1$, the optimal regret $\widetilde{\Theta}\Big(T^{\frac{k_f+1}{2k_f+1}}\Big)$ of SMLCE is consistently lower than the optimal regret $\widetilde{\Theta}\Big(T^{\frac{d+k_g}{d+2k_g}}\Big)$ of SMNPE noting that $\frac{k_f+1}{2k_f+1}\le\frac{2}{3}<\frac{d+k_g}{d+2k_g}$ due to $d\ge 2>k_g$. This reveals that when $f(p)$ becomes smoother, the primary complexity shifts from learning $f(p)$ to learning $g(x)$ when transiting from a linear function $a^\top x$ to a non-parametric function.

6. Numerical Study

In this section, we study the empirical performances of our algorithms. We measure the performance of a learning algorithm π by the relative regret defined as follows:

$$\frac{\sum_{t=1}^T \mathbb{E}[p_t^*(f(p_t^*)+g(x_t))-p_t(f(p_t)+g(x_t))]}{\sum_{t=1}^T \mathbb{E}[p_t^*(f(p_t^*)+g(x_t))]}\times 100\%,$$

where p_t^* is the optimal price under context x_t.

For each of the three models SMLPE, SMLCE and SMNPE, we compare the relative regret of our algorithms with the contextual zooming algorithm proposed by Slivkins (2011), which is designed for general Lipschitz bandits. For the first model SMLPE, we also evaluate the performance of `RPS` Algorithm proposed by Nambiar et al. (2019). As discussed, Nambiar et al. (2019) considers the problem with the same demand structure as our SMLPE. `RPS` Algorithm has been proven to converge to the best linear model asymptotically. However, when the true model is not linear, it may suffer from model mis-specification as will be shown later. For the second and third models, SMLCE and SMNPE, we also evaluate the performance of the linear greedy algorithm. It runs a linear regression in each period to get an estimate of the demand function by a linear function (see Algorithm 7 in Appendix E) and suggests a price that maximizes the proxy revenue function. For each instance tested in this section, we repeat the experiments for 50 independent runs, and compute the empirical relative regret by taking the average to approximate the true relative regret.

6.1. Numerical Results for SMLPE

In this subsection, we present the numerical results for SMLPE in two numerical settings below.

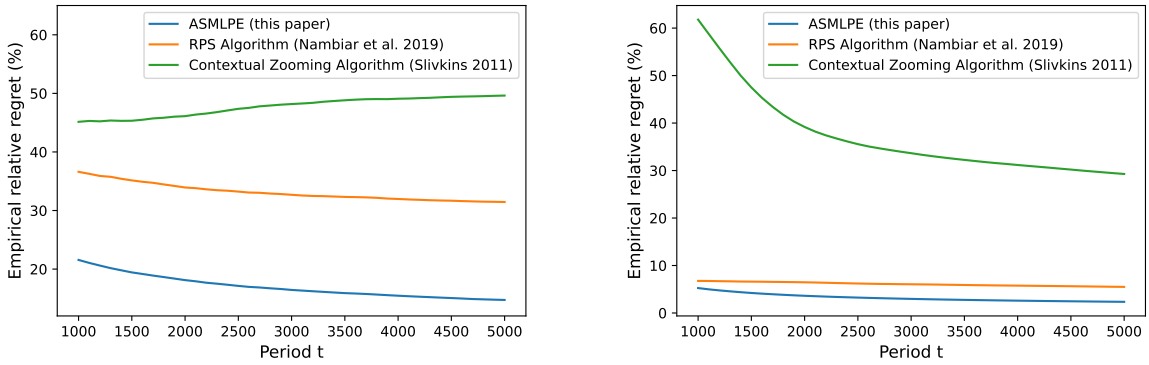
The first numerical setting. This setting comes from Nambiar et al. (2019) with a simple re-scaling of the domain of context. The demand is given by the following function:

$$D_t(p_t) = -0.9p_t + 1 + \frac{1}{2(2x_t + 0.03)} + \varepsilon_t, \tag{19}$$

where x_t is a one-dimensional random variable uniformly distributed over $[0,1]$ and the noise ε_t is normally distributed with mean 0 and standard deviation 0.1. The feasible price range is $[\underline{p}, \overline{p}] = [0.6, 9.81]$, and the range of b is $[\underline{b}, \overline{b}] = [-1.2, -0.5]$. Figure 3(a) shows that the relative regret of our `ASMLPE` keeps decreasing as t grows. By contrast, the `RPS` algorithm incurs an almost flat curve as t increases, indicating that it fails to converge to the true optimal pricing policy. The relative regret for contextual zooming can even increase as t increases. This may be caused by the reason that in (19), x_t has a certain probability near 0, so the derivative of $g(x)$ can be large, indicating that the Lipschitz constant L can be very large and identifying $g(\cdot)$ becomes more challenging. It's also worth noting that the contextual zooming algorithm in Slivkins (2011) is specifically designed for $(x_t, p_t) \in [0,1]^2$, and it is not straightforward to select proper hyper-parameters to adapt to the demand function in (19). In the second numerical setting, the assumptions in Slivkins (2011) are well satisfied.

The second numerical setting. In this setting, we let $f(p) = -p$ and $g(x)$ as follows:

$$g(x) := \begin{cases} \left(D(x, \partial([0,1]^d))\right)^{k_g} + 1, & \text{if } D(x, \partial([0,1]^d)) \le \frac{1}{4}, \\ (\frac{1}{4})^{k_g} - 1.4 \times \left(D(x, \partial([0,1]^d)) - \frac{1}{4}\right)^{k_g} + 1, & \text{otherwise,} \end{cases} \tag{20}$$



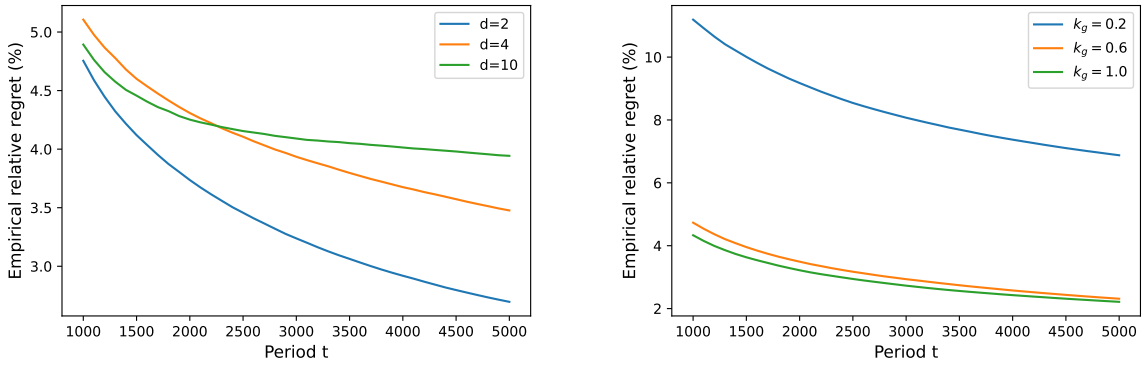
(a) Comparison of three algorithms under Eq. (19) (b) Comparison of three algorithms under Eq. (20)

Figure 3 Empirical relative regret of three algorithms under SMLPE

where $D(x, \partial([0,1]^d))$ is the Euclidean distance between x and the boundary of $[0,1]^d$. We assume that ϵ is a Gaussian random variable with mean zero and standard deviation 0.1, and x_t is uniformly distributed on $[0,1]^d$, and set $[\underline{b}, \overline{b}] = [-1.5, -0.5]$ and $[\underline{p}, \overline{p}] = [0.1, 1]$. In Figure 3(b), we compare the performances under our `ASMLPE` and the two benchmarks with $d = 2$ and $k_g = 1$. Since the assumptions in Slivkins (2011) are satisfied in this setting, the contextual zooming algorithm gradually converges. However, the convergence speed is relatively slow compared with our `ASMLPE`, which is the loss of not utilizing the separable structure. Although the RPS algorithm achieves a relative small regret, the curve is still quite flat.

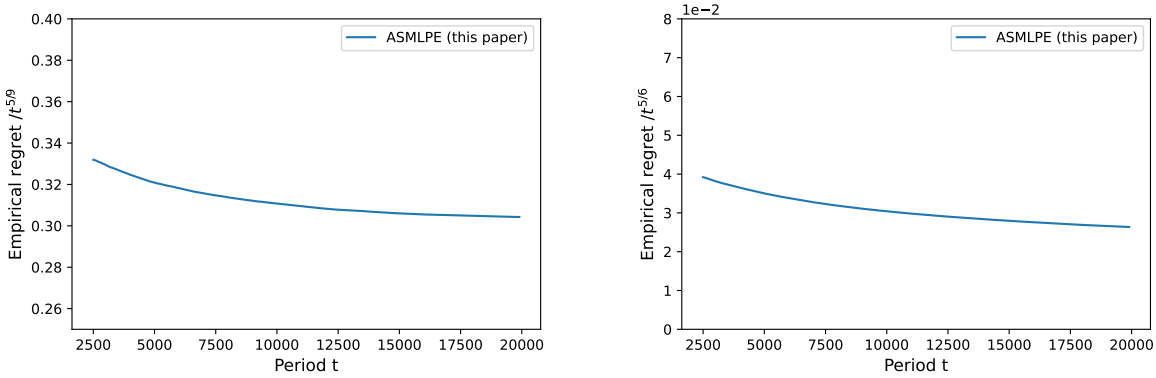
We then test the impacts of d and k_g on the empirical performance of `ASMLPE`. In Figure 4(a), we set $d = 2, 4, 10$ while fixing $k_g = 1$ in Eq. (20). As shown in the figure, when t is sufficiently large, the empirical relative regret increases with respect to d, which is consistent with our theoretical results. Although $d = 10$ slightly outperforms $d = 4$ when t is small, the decaying of $d = 4$ seems always faster than that of $d = 10$. In Figure 4(b), we set $k_g = 0.2, 0.6, 1$ while fixing $d = 1$ in Eq. (20). When k_g is relatively small, a little bit increment of k_g, i.e., from $k_g = 0.2$ to $k_g = 0.6$, will lead to a significant speed-up of the convergence. When k_g gets larger, e.g., from $k_g = 0.6$ to $k_g = 1$, further increasing its value can only bring a limited improvement.

We also test the rate of the empirical regret and compare it with the theoretical regret upper bound established in Theorem 1. We consider two instances under the demand function in Eq. (20): $d = 2, k_g = 0.8$ and $d = 10, k_g = 1$, whose theoretical regret upper bounds are $\widetilde{O}(T^{\frac{5}{9}})$ and $\widetilde{O}(T^{\frac{5}{6}})$, respectively. In Figure 5, we plot the scaled empirical regret, which is defined as the empirical regret divided by $t^{\frac{5}{9}}$ and $t^{\frac{5}{6}}$. We can see that the scaled regret gradually converges to a constant, which means that the empirical regret is aligned with the theoretical rate.



(a) Impact of dimension d (b) Impact of smoothness parameter k_g

Figure 4 Empirical relative regret of `ASMLPE` under SMLPE



(a) $d = 2, k_g = 0.8$ (b) $d = 10, k_g = 1$

Figure 5 Scaled empirical regret of `ASMLPE` under SMLPE

6.2. Numerical Results for SMLCE

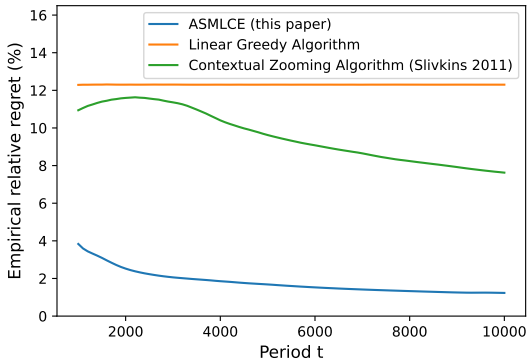
In this subsection, we present the numerical results for SMLCE in the following numerical setting:

$$D_t(p_t) = -\frac{4}{15}\delta p_t^{2.5} + 30 + \frac{1}{d}\mathbf{1}_d \cdot x_t + \varepsilon_t,$$

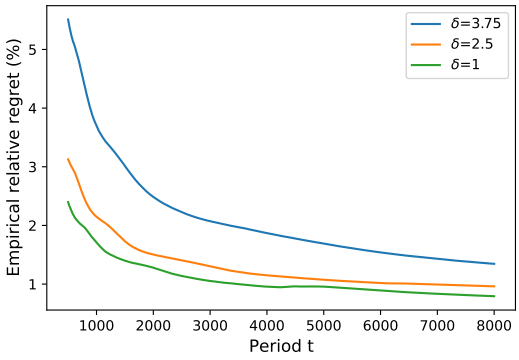
where $\mathbf{1}_d := (1, 1, \cdots, 1) \in \mathbb{R}^d$, x_t is uniformly distributed on $[0,1]^d$ and ε_t is i.i.d. zero-mean Gaussian with standard deviation 0.1. The optimal price under context x_t is $(\frac{15(30+\frac{1}{d}\mathbf{1}_d \cdot x_t)}{14\delta})^{1/2.5}$. By simple calculation, we know $k_f = 2.5$ in this case.

In Figure 6(a), we fix $\delta = 3.75$ and $[\underline{p}, \overline{p}] = [2, 3.5]$, and compare our `ASMLCE` with the contextual zooming algorithm and the linear greedy algorithm. The empirical relative regrets of both our `ASMLCE` and the contextual zooming algorithm decrease with respect to t. However, the latter converges slower than the former, which illustrates the benefit of utilizing the separable structure. The linear greedy algorithm has an almost flat empirical relative regret curve, indicating that this

policy fails to approach the optimal pricing policy. In Figure 6(b), we show the impact of δ by varying it in $\{3.75, 2.5, 1\}$. As δ becomes smaller, the empirical relative regret of `ASMLCE` decreases, which is consistent with our theoretical results. Similar to the effect of k_g in Figure 4(b), a small decrement of δ from 3.75 to 2.5 leads to a notable decrement in the regret, whereas the improvement is less significant by further decreasing the value of δ. In Figure 7, we plot the scaled empirical regret of `ASMLCE`, i.e., the empirical regret divided by $t^{7/12}$. We can see that the scaled regret experiences a very mild decrement when t increases from 2,000 to 10,000.



(a) Comparison of three algorithms



(b) Impact of parameter δ

Figure 6 Empirical relative regret under SMLCE

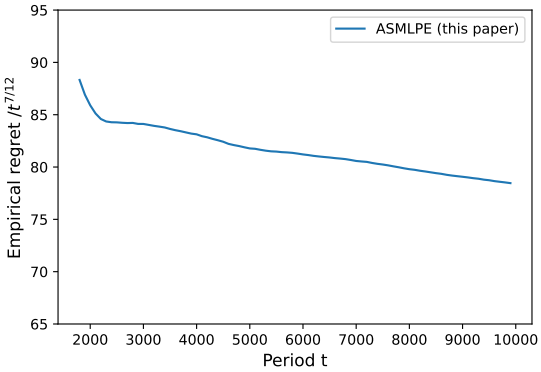


Figure 7 Scaled empirical regret of `ASMLPE` under SMLPE

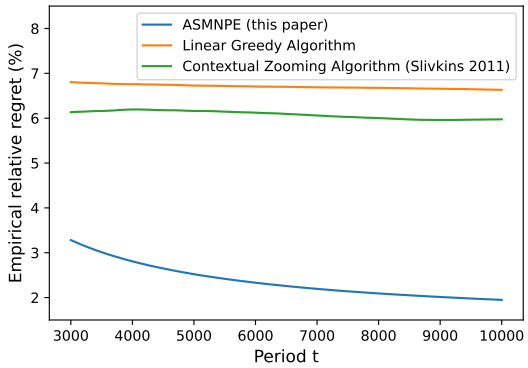


Figure 8 Empirical relative regret of three algorithms under SMNPE

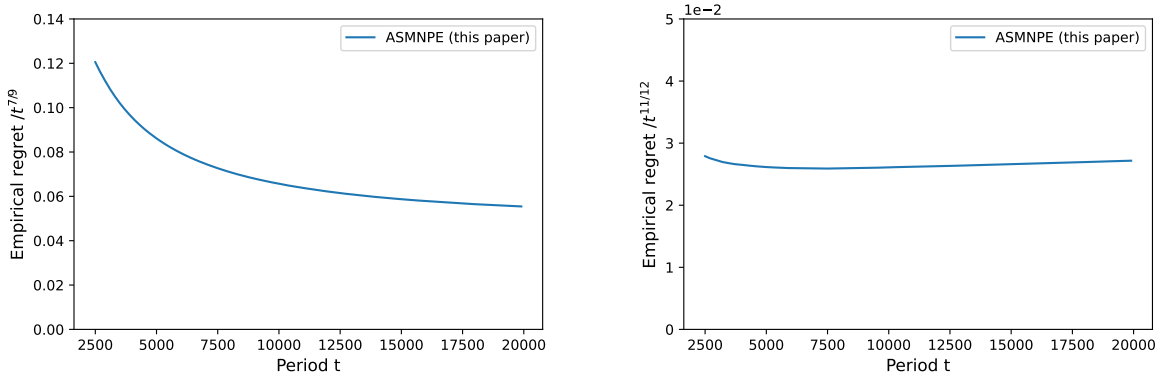
6.3. Numerical Results for SMNPE

In this subsection, we present the numerical results for the model SMNPE in the following numerical setting:

$$f(p) := \begin{cases} -p & \text{if } p \le 0.625, \\ -0.44 \times p - 0.35 & \text{otherwise,} \end{cases} \tag{21}$$

where $p \in [\underline{p}, \overline{p}] = [0.1, 1]$. By simple calculation, we can verify $k_f = \delta = 1$. In this subsection, we still apply $g(x)$ as Eq. (20) to test `ASMNPE`. We assume that x is uniformly distributed on $[0,1]^d$ and ε is a zero-mean Gaussian *r.v.* with standard deviation 0.1.

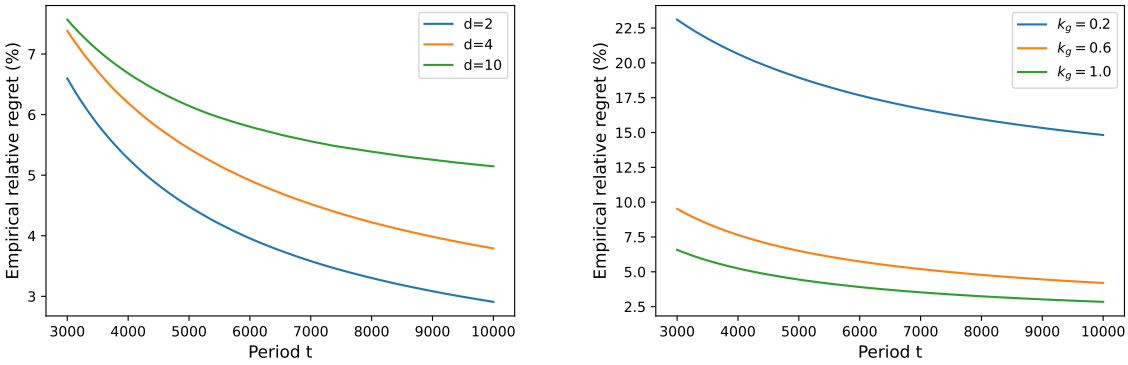
In Figure 8, we fix $k_g = 1$ and $d = 2$ and compare the performances of `ASMNPE`, the contextual zooming algorithm and the linear greedy policy As t increases, the relative regret of `ASMNPE` keeps decreasing, while remains almost a constant under the contextual zooming algorithm and the linear greedy algorithm. Similar as Figure 5, we test the order of the empirical regret in Figure 9. We still use the same two sets of parameters: $d = 2, k_g = 0.8$ and $d = 10, k_g = 1$ as in Figure 5, but the theoretical regret upper bounds become $\widetilde{\mathcal{O}}(T^{\frac{7}{9}})$ and $\widetilde{\mathcal{O}}(T^{\frac{11}{12}})$ respectively. As we can see from Figure 9, the scaled empirical regrets gradually converge to a constant, which verifies the theoretical results.



(a) $d = 2, k_g = 0.8$ (b) $d = 10, k_g = 1$

Figure 9 Scaled empirical regret of `ASMNPE` under SMNPE

We also study the impacts of d and k_g on the performance of `ASMNPE` in Figures 10(a) and 10(b) respectively. In Figure 10(a), we fix $k_g = 1$ and let $d = 2, 4, 10$. As shown in the figure, the lower dimension d is, the faster the algorithm `ASMNPE` converges. Figure 10(b) shows the impact of k_g on the performance of `ASMNPE`. Similar to Figure 4(b), when k_g is relatively small, increasing the value of k_g a little bit can lead to a significant improvement of the performance of `ASMNPE`.



(a) Impact of dimension d (b) Impact of smoothness parameter k_g

Figure 10 Empirical relative regret of `ASMNPE` under SMNPE

6.4. Model Misspecification and Robustness

In this subsection, we shed some light on the robustness of our algorithm in a model-misspecified setting. Specifically, we consider the following demand function for which the price and contextual effects are not separable:

$$D_t(p_t) = -(x_t + 0.5)p_t + g(x_t) + \varepsilon_t. \tag{22}$$

We assume $p_t \in [0.1, 1]$, x_t is a one-dimensional random variable uniformly distributed over $[0, 1]$, $g(x_t)$ follows the definition of Eq. (20) and the noise ε_t is normally distributed with mean 0 and standard deviation 0.1. We run our `ASMLPE` algorithm by letting $\underline{b} = -2$ and $\overline{b} = -0.3$, and plot its empirical relative regret in Figure 11. We can see that even under this misspecified setting, `ASMLPE` still exhibits a robust performance, with the empirical relative regret consistently below 4% when t exceeds 10, 000. As t increases to 50,000, the empirical relative regret does not seem to exhibit further reduction, which is unavoidable because of the model misspecification.

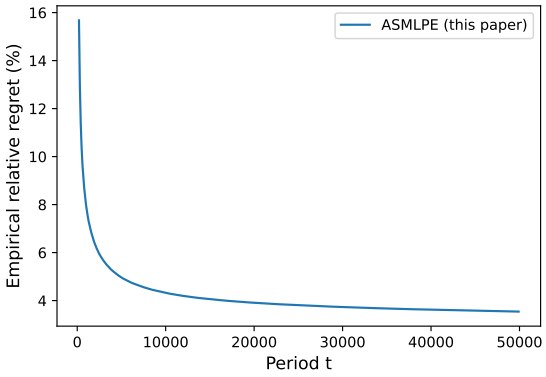


Figure 11 The empirical relative regret of `ASMLPE` under Eq. (22)

7. Concluding Remarks

In this work, we study context-based dynamic pricing with an unknown demand function. The demand model is assumed to have a separable structure between the pricing effect and the contextual effect. For each of the three models with different structural properties of the demand function, we propose an efficient learning algorithm with a provable regret upper bound. We also show that the upper bound is unimprovable by developing a matching regret lower bound in certain senses. One interesting direction for future research is to validate and further utilize such separable structures in other applications from operations management with high-dimensional contexts. Potential topics include context-based resource allocation, multi-product pricing, assortment optimization, and joint inventory control and pricing problems.

We also emphasize that the algorithms we develop in this paper assume prior knowledge of smoothness parameters k_g and k_f. In practice, such pre-knowledge cannot be available to the seller, especially when launching new products. Although perfect estimations of k_g and k_f are hardly possible, we can still infer their values or their ranges from similar products that have richer historical sales data. Various practical approaches can be applied for determining suitable k_g and k_f. Notably, among these approaches are criterion-based methods, such as Bayesian information criterion (see, e.g., Neath and Cavanaugh 2012) and Akaike information criterion (see, e.g., Cavanaugh and Neath 2019), both of which are widely embraced in the broader field of model selection. An alternate strategy is to successively fit the models in an increasing order and test the significance of regression coefficients at each step of model fitting. This iterative process continues until the t-test for the highest-order term is no longer statistically significant, which is called a forward selection procedure. On the other hand, we want to point out that Gur et al. (2022) highlights that devising contextual bandit algorithms capable of adapting to unknown smoothness of the reward function is a "nearly impossible" task. Conceptually, in our dynamic-pricing setting, pre-knowing k_g and k_f is also almost inevitable for any algorithms. Furthermore, recent advancements within the broader contextual bandit literature, notably within the context of similar smooth payoff functions studied by Hu et al. (2022), also underscore the need for knowing continuity parameters. This implies the inherent complexity of the problem when continuity parameters remain unknown.

Another crucial prior knowledge that we are assuming to be known is the separable structure itself. In our collaboration with the Middle Eastern retailer, we found that the separable model performed well on three years of historical data in a relatively heuristic way. Another possible approach to identify the separable structure is to treat separability as a hyperparameter and use empirical methods such as cross-validation to determine whether a separable model is preferable. However, the main challenge lies in designing a statistically rigorous test for separability, particularly how many samples are required to confidently conclude whether a function is separable,

particularly in the non-parametric and semi-parametric settings we consider. To the best of our knowledge, the literature does not yet offer a straightforward solution. We leave this as another important future work.

Acknowledgment: The authors are grateful to the department editor J. George Shanthikumar, the anonymous associate editor, and three anonymous referees for their constructive comments and suggestions that have helped to significantly improve both the content and exposition of this paper. The authors acknowledge the support from the MIT Data Science Laboratory. J. Bu acknowledges the support from the Research Grants Council of Hong Kong [Early Career Scheme Grant PolyU 25505322].

References

Abbasi-Yadkori Y, Pál D, Szepesvári C (2011) Improved algorithms for linear stochastic bandits. *Advances in Neural Information Processing Systems*, 2312–2320.

Auer P, Cesa-Bianchi N, Fischer P (2002) Finite-time analysis of the multiarmed bandit problem. *Machine learning* 47(2-3):235–256.

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

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

Besbes O, Zeevi A (2012) Blind network revenue management. *Operations research* 60(6):1537–1550.

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

Bretagnolle J, Huber C (1979) Estimation des densités: risque minimax. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 47(2):119–137.

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

Buja A, Hastie T, Tibshirani R (1989) Linear smoothers and additive models. *The Annals of Statistics* 453–510.

Cavanaugh JE, Neath AA (2019) The akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. *Wiley Interdisciplinary Reviews: Computational Statistics* 11(3):e1460.

Chen N, Gallego G (2021) Nonparametric pricing analytics with customer covariates. *Operations Research* 69(3):974–984.

Chen N, Hu M (2023) Data-driven revenue management: The interplay of data, model, and decisions. *Service Science* .

Chen X, Jasin S, Shi C (2022) *The Elements of Joint Learning and Optimization in Operations Management*, volume 18 (Springer Nature).

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

Chu W, Li L, Reyzin L, Schapire R (2011) Contextual bandits with linear payoff functions. *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, 208–214 (JMLR Workshop and Conference Proceedings).

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

Dani V, Hayes TP, Kakade SM (2008) Stochastic linear optimization under bandit feedback. *Proceedings of the 21st Conference on Learning Theory*.

den Boer AV (2015) Dynamic pricing and learning: historical origins, current research, and new directions. *Surveys in operations research and management science* 20(1):1–18.

Fan J, Guo Y, Yu M (2021) Policy optimization using semiparametric models for dynamic pricing. *Available at SSRN 3922825* .

Filippi S, Cappe O, Garivier A, Szepesvári C (2010) Parametric bandits: The generalized linear case. *Advances in Neural Information Processing Systems*, 586–594.

Gur Y, Momeni A, Wager S (2022) Smoothness-adaptive contextual bandits. *Operations Research* .

Hamilton JD (2020) *Time series analysis* (Princeton university press).

Hu Y, Kallus N, Mao X (2022) Smooth contextual bandits: Bridging the parametric and nondifferentiable regret regimes. *Operations Research* 70(6):3261–3281.

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

Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu TY (2017) Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems* 30.

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

Keskin NB (2014) Optimal dynamic pricing with demand model uncertainty: A squared-coefficient-of-variation rule for learning and earning. *Available at SSRN 2487364* .

Keskin NB, Li Y, Song JS (2022a) Data-driven dynamic pricing and ordering with perishable inventory in a changing environment. *Management Science* 68(3):1938–1958.

Keskin NB, Li Y, Sunar N (2022b) Data-driven clustering and feature-based retail electricity pricing with smart meters. *Available at SSRN 3686518* .

Lattimore T, Szepesvári C (2018) Bandit algorithms. *preprint* .

Lei YM, Jasin S, Sinha A (2014) Near-optimal bisection search for nonparametric dynamic pricing with inventory constraint. *Ross School of Business Paper* (1252).

Levin DA, Peres Y (2017) *Markov chains and mixing times*, volume 107 (American Mathematical Soc.).

Li L, Chu W, Langford J, Moon T, Wang X (2012) An unbiased offline evaluation of contextual bandit algorithms with generalized linear models. *Proceedings of the Workshop on On-line Trading of Exploration and Exploitation 2*, 19–36 (JMLR Workshop and Conference Proceedings).

Li L, Lu Y, Zhou D (2017) Provably optimal algorithms for generalized linear contextual bandits. *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, 2071–2080 (JMLR. org).

Li M, Simchi-Levi D, Tan R, Wang C, Wu MX (2023) Contextual offline demand learning and pricing with separable models. *Available at SSRN 4619018* .

Li X, Zheng Z (2023) Dynamic pricing with external information and inventory constraint. *Management Science* .

Luo Y, Sun WW, Liu Y (2023) Distribution-free contextual dynamic pricing. *Mathematics of Operations Research* .

Mao J, Leme R, Schneider J (2018) Contextual pricing for lipschitz buyers. *Advances in Neural Information Processing Systems* 31.

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

Miao S, Wang Y (2021) Network revenue management with nonparametric demand learning:\sqrt {T}-regret and polynomial dimension dependency. *Available at SSRN 3948140* .

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

Neath AA, Cavanaugh JE (2012) The bayesian information criterion: background, derivation, and applications. *Wiley Interdisciplinary Reviews: Computational Statistics* 4(2):199–203.

Perchet V, Rigollet P (2013) The multi-armed bandit problem with covariates. *The Annals of Statistics* 41(2):693–721.

Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A (2018) Catboost: unbiased boosting with categorical features. *Advances in neural information processing systems* 31.

Qiang S, Bayati M (2016) Dynamic pricing with demand covariates. *Available at SSRN 2765257* .

Rigollet P, Zeevi A (2010) Nonparametric bandits with covariates. *arXiv preprint arXiv:1003.1630* .

Rusmevichientong P, Tsitsiklis JN (2010) Linearly parameterized bandits. *Mathematics of Operations Research* 35(2):395–411.

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

Slivkins A (2011) Contextual bandits with similarity information. *Proceedings of the 24th annual Conference On Learning Theory*, 679–702 (JMLR Workshop and Conference Proceedings).

Slivkins A (2019) Introduction to multi-armed bandits. *arXiv preprint arXiv:1904.07272* .

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

Wang Y, Chen B, Simchi-Levi D (2021b) Multimodal dynamic pricing. *Management Science* 67(10):6136–6152.

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

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

Online Appendix for "Context-Based Dynamic Pricing with Separable Demand Models"

By Jinzhi Bu, David Simchi-Levi, and Chonghuan Wang

Appendix A. Proofs for Statements in Section 2

A.1. Proof for Theorem 1

For each $t \in [T]$, let j_t be the index of the bin that x_t belongs and note that j_t is also a random variable. The total regret can be decomposed as follows:

$$\begin{aligned}
\sum_{t=1}^{T} r_t &= \sum_{t=1}^{T} -\frac{g(x_t)}{2b}\left(b\cdot\left(-\frac{g(x_t)}{2b}\right)+g(x_t)\right) - p_t\left(bp_t+g(x_t)\right) \\
&= |b|\sum_{t=1}^{T}\left(-\frac{g(x_t)}{2b}-p_t\right)^2 \\
&\le 2|b|\sum_{t=1}^{T}\left(-\frac{g(x_t)}{2b}-p_t^0\right)^2+\left(p_t^0-p_t\right)^2 \\
&\le 2|b|\sum_{t=1}^{T}\left(-\frac{g(x_t)}{2b}-p_t^u\right)^2+\left(p_t^0-p_t\right)^2 \\
&\le 6|b|\sum_{t=1}^{T} \underbrace{\left(\frac{-g(x_t)+\theta_{j_t}^*\phi(x_t)}{2b}\right)^2}_{\text{regret from polynomial approximation of } g(\cdot)} + \underbrace{\left(\frac{\theta_{j_t}^*\phi(x_t)}{2b}-\frac{\theta_{j_t}^*\phi(x_t)}{2\hat{b}_t}\right)^2}_{\text{regret from estimation error of } b} \\
&\qquad + \underbrace{\left(\frac{\theta_{j_t}^*\phi(x_t)-\hat{\theta}_{t,j_t}\phi(x_t)}{2\hat{b}_t}\right)^2}_{\text{regret from regression}} + \underbrace{\left(p_t^0-p_t\right)^2}_{\text{regret from random shock}}, \qquad \text{(EC.1)}
\end{aligned}$$

where $p_t^0 := \mathrm{Proj}(p_t^u, [\underline{p}, \overline{p}])$, the first and the third inequalities follow from Cauchy-Schwarz inequality, and the second inequality is due to $-\frac{g(x_t)}{2b} \in [\underline{p}, \overline{p}]$.

The first term on the RHS of (EC.1) arises from approximating $g(\cdot)$ using a k-polynomial function in each local bin, and can be further upper bounded as follows by Eq. (4):

$$\left(\frac{-g(x_t)+\theta_j^*\phi(x_t)}{2b}\right)^2 \le \frac{\max_{x\in\mathbf{M}_{j_t}}(g(x)-\theta_j^*\phi(x))^2}{4\bar{b}^2} \le \frac{L^2 d^{k_g+\mathfrak{b}(k_g)}}{4\bar{b}^2 M^{2k_g}(\mathfrak{b}(k_g)!)^2} \qquad \text{(EC.2)}$$

The second term on the RHS of (EC.1) is due to the estimation error of price sensitivity b since

$$\left(\frac{\theta_j^*\phi(x_t)}{2b}-\frac{\theta_j^*\phi(x_t)}{2\hat{b}_t}\right)^2 \le \frac{\max_{x\in[0,1]^d}(g(x))^2}{4\bar{b}^4}(\hat{b}_t-b)^2. \qquad \text{(EC.3)}$$

When $t \ge 2$,

$$\mathbb{E}[(b-\hat{b}_t)^2] = \frac{1}{(\sum_{s=1}^{t} s^{-\frac{1}{2}})^2}\mathbb{E}\left[\left(\sum_{s=1}^{t}\Delta_s(bp_s^g+g(x_s)+\varepsilon_s)\right)^2\right]$$

$$
\begin{aligned}
&= \frac{1}{(\sum_{s=1}^{t} s^{-\frac{1}{2}})^2} \mathbb{E}\left[\sum_{s=1}^{t} \Delta_s^2 (bp_s^g + g(x_s) + \varepsilon_s)^2\right] \\
&\le \frac{3}{(\sum_{s=1}^{t} s^{-\frac{1}{2}})^2} \sum_{s=1}^{t} \mathbb{E}\left[\sum_{s=1}^{t} s^{-\frac{1}{2}} \left((bp_s^g)^2 + (g(x_s))^2 + \varepsilon_s^2\right)\right] \\
&\le \frac{3\left(\underline{b}^2 \overline{p}^2 + \max_{x\in[0,1]^d} (g(x))^2 + \sigma^2\right)}{\sqrt{t}},
\end{aligned} \tag{EC.4}
$$

where the first identity follows from Eq. (3), the second identity holds since $\delta_s = s^{-\frac{1}{2}}$ by its definition and when $s \neq k$, Δ_s is independent of $\Delta_k(bp_s^g + g(x_s) + \varepsilon_s)(bp_k^g + g(x_k) + \varepsilon_k)$, and from $\mathbb{E}[\Delta_s] = 0$, we have $\mathbb{E}[\Delta_s \Delta_k (bp_s^g + g(x_s) + \varepsilon_s)(bp_k^g + g(x_k) + \varepsilon_k)] = 0$, the first inequality follows from $\Delta_s \in \{-s^{-\frac{1}{4}}, s^{\frac{1}{4}}\}$ and Cauchy-Schwarz inequality, and the last inequality holds since when $t \ge 2$, $\sum_{s=1}^{t} s^{-\frac{1}{2}} \ge \int_1^{t+1} s^{-\frac{1}{s}} ds = 2(\sqrt{t+1} - 1) \ge \sqrt{t}$. Therefore, we can have

$$
\sum_{t=1}^{T} \mathbb{E}\left[\left(\frac{\theta_j^* \phi(x_t)}{2b} - \frac{\theta_j^* \phi(x_t)}{2\hat{b}_t}\right)^2\right] = \mathcal{O}(\sqrt{T}). \tag{EC.5}
$$

The third term on the RHS of (EC.1) represents the estimation error of $\hat{\theta}_{t,j_t}$ using the lasso regression. For simplicity, we denote $V_{t,j_t} = \lambda I + \sum_{(x_s,p_s,d_s)\in\mathcal{D}_{t,j_t}} \phi(x_i)\phi(x_i)^\top$. Note that

$$
\begin{aligned}
\hat{\theta}_{t,j_t} &= V_{t,j_t}^{-1} \sum_{(x_s,p_s,d_s)\in\mathcal{D}_{t,j_t}} (p_s b + g(x_s) + \varepsilon_s - p_s \hat{b}_t)\phi(x_s) \\
&= V_{t,j_t}^{-1} \sum_{(x_s,p_s,d_s)\in\mathcal{D}_{t,j_t}} (p_s b + (\theta_j^*)^\top \phi(x_s) + \Delta(s) + \varepsilon_s - p_s \hat{b}_t)\phi(x_s) \\
&= \theta_j^* - \lambda V_{t,j_t}^{-1} \theta_j^* I + V_{t,j_t}^{-1} \sum_{(x_s,p_s,d_s)\in\mathcal{D}_{t,j_t}} (\Delta(s) + \varepsilon_s)\phi(x_s) + (b - \hat{b}_t) V_{t,j_t}^{-1} \sum_{(x_s,p_s,d_s)\in\mathcal{D}_{t,j_t}} p_s \phi(x_s),
\end{aligned} \tag{EC.6}
$$

where $\Delta(s)$ is the approximation error of the best polynomial function at x_s. Note that by Assumption 1, we can guarantee that $|\Delta(s)| = \mathcal{O}(\frac{L}{M^{kg}}) := \Delta$. Then, based on Eq. (EC.6) and Cauchy-Schwarz inequality, we have for any x

$$
\begin{aligned}
\left|x^\top \hat{\theta}_{t,j_t} - x^\top \theta_j^*\right| \le \|x\|_{V_{t,j_t}^{-1}} &\left(\lambda^{\frac{1}{2}} \|\theta^*\|_2 + \left\|\sum_{(x_s,p_s,d_s)\in\mathcal{D}_{t,j_t}} \varepsilon_s \phi(x_s)\right\|_{V_{t,j_t}^{-1}}\right) \\
&+ \left|x^\top V_{t,j_t}^{-1} \sum_{(x_s,p_s,d_s)\in\mathcal{D}_{t,j_t}} \Delta(s)\phi(x_s)\right| + |b - \hat{b}_t| \left|x^\top V_{t,j_t}^{-1} \sum_{(x_s,p_s,d_s)\in\mathcal{D}_{t,j_t}} p_s \phi(x_s)\right|,
\end{aligned} \tag{EC.7}
$$

where we adopt the notation $\|a\|_V := \sqrt{a^\top V a}$ for some positive definite matrix V. The second term of Eq. (EC.7) can be further bounded as follows,

$$
\left|x^\top V_{t,j_t}^{-1} \sum_{(x_s,p_s,d_s)\in\mathcal{D}_{t,j_t}} \Delta(s)\phi(x_s)\right| \le \sqrt{\sum_{(x_s,p_s,d_s)\in\mathcal{D}_{t,j_t}} (x^\top V_{t,j_t}^{-1} \phi(x_s))^2 \sum_{(x_s,p_s,d_s)\in\mathcal{D}_{t,j_t}} \Delta(s)^2}
$$

$$\leq \sqrt{|\mathcal{D}_{t,j_t}|}\Delta\sqrt{\sum_{(x_s,p_s,d_s)\in\mathcal{D}_{t,j_t}}(x^\top V_{t,j_t}^{-1}\phi(x_s))^2}$$
$$= \sqrt{|\mathcal{D}_{t,j_t}|}\Delta\sqrt{x^\top V_{t,j_t}^{-1}\left(\sum_{(x_s,p_s,d_s)\in\mathcal{D}_{t,j_t}}\phi(x_s)\phi(x_s)^\top\right)V_{t,j_t}^{-1}x}$$
$$\leq \sqrt{|\mathcal{D}_{t,j_t}|}\Delta\,\|x\|_{V_{t,j_t}^{-1}}. \tag{EC.8}$$

Similarly, the third term of Eq. (EC.7) can be controlled by

$$|b-\hat{b}_t|\left|x^\top V_{t,j_t}^{-1}\sum_{(x_s,p_s,d_s)\in\mathcal{D}_{t,j_t}}p_s\phi(x_s)\right| \leq |b-\hat{b}_t|\sqrt{\sum_{(x_s,p_s,d_s)\in\mathcal{D}_{t,j_t}}(x^\top V_{t,j_t}^{-1}\phi(x_s))^2\sum_{(x_s,p_s,d_s)\in\mathcal{D}_{t,j_t}}p_s^2}$$
$$\leq |b-\hat{b}_t|\sqrt{|\mathcal{D}_{t,j_t}|}\overline{p}\,\|x\|_{V_{t,j_t}^{-1}}. \tag{EC.9}$$

By Theorem 1 in Abbasi-Yadkori et al. (2011), we have that for any $\epsilon>0$, with probability at least $1-\epsilon$, $\forall t\geq 0$ such that $x_t\in\mathbf{M}_{j_t}$

$$\left\|\sum_{(x_s,p_s,d_s)\in\mathcal{D}_{t,j_t}}\varepsilon_s\phi(x_s)\right\|_{V_{t,j_t}^{-1}} \leq \sigma\sqrt{2\log\left(\frac{\det(V_{t,j_t})^{1/2}\det(\lambda I)^{-1/2}}{\epsilon}\right)} := \gamma_{t,j_t}. \tag{EC.10}$$

Plugging Eqs. (EC.8), (EC.9), and (EC.10) into Eq. (EC.7), we can get

$$\left|x^\top\hat{\theta}_{t,j_t}-x^\top\theta_j^*\right| \leq \|x\|_{V_{t,j_t}^{-1}}\left(\lambda^{\frac{1}{2}}\|\theta^*\|_2+\gamma_{t,j_t}+\sqrt{|\mathcal{D}_{t,j_t}|}\Delta+\sqrt{|\mathcal{D}_{t,j_t}|}|b-\hat{b}_t|\overline{p}\right). \tag{EC.11}$$

Therefore, we have the following upper bound on the third term of (EC.1) with high probability:

$$\left(\frac{\theta_j^*\phi(x_t)-\hat{\theta}_{t,j_t}\phi(x_t)}{2\hat{b}_t}\right)^2 \leq \frac{1}{4\underline{b}^2}\left(\theta_j^*\phi(x_t)-\hat{\theta}_{t,j_t}\phi(x_t)\right)^2$$
$$\leq \frac{1}{\underline{b}^2}\|\phi(x_t)\|_{V_{t,j_t}^{-1}}^2\left(\lambda\|\theta^*\|_2^2+\gamma_{t,j_t}^2+|\mathcal{D}_{t,j_t}|\Delta^2+|\mathcal{D}_{t,j_t}||b-\hat{b}_t|^2\overline{p}^2\right).$$

Then, we have the following sum over t, with probability at least $1-M^d\epsilon$,

$$\sum_{t=1}^T\left(\frac{\theta_j^*\phi(x_t)-\hat{\theta}_{t,j_t}\phi(x_t)}{2\hat{b}_t}\right)^2 \leq \frac{1}{\underline{b}^2}\sum_{t=1}^T\|\phi(x_t)\|_{V_{t,j_t}^{-1}}^2\left(\lambda\|\theta^*\|_2^2+\gamma_{t,j_t}^2+|\mathcal{D}_{t,j_t}|\Delta^2+|\mathcal{D}_{t,j_t}||b-\hat{b}_t|^2\overline{p}^2\right)$$
$$= \frac{1}{\underline{b}^2}\sum_{j=1}^{M^d}\sum_{t=1}^T\|\phi(x_t)\|_{V_{t,j}^{-1}}^2\left(\lambda\|\theta^*\|_2^2+\gamma_{t,j}^2+|\mathcal{D}_{t,j}|\Delta^2\right)1_{x_t\in\mathbf{M}_j}$$
$$+\frac{1}{\underline{b}^2}\sum_{t=1}^T\|\phi(x_t)\|_{V_{t,j_t}^{-1}}^2|\mathcal{D}_{t,j_t}||b-\hat{b}_t|^2\overline{p}^2. \tag{EC.12}$$

The first term of Eq. (EC.12) can be bounded as follows,

$$\frac{1}{\underline{b}^2}\sum_{j=1}^{M^d}\sum_{t=1}^T\|\phi(x_t)\|_{V_{t,j}^{-1}}^2\left(\lambda\|\theta^*\|_2^2+\gamma_{t,j}^2+|\mathcal{D}_{t,j}|\Delta^2\right)1_{x_t\in\mathbf{M}_j}$$

$$
\begin{aligned}
&\leq \frac{1}{\underline{b}^2}\sum_{j=1}^{M^d}\left(\lambda\|\theta^*\|_2^2+\gamma_{T,j}^2+|\mathcal{D}_{T,j}|\Delta^2\right)\sum_{t=1}^{T}\|\phi(x_t)\|^2_{V_{t,j}^{-1}}1_{x_t\in\mathbf{M}_j}\\
&\leq \frac{1}{\underline{b}^2}\sum_{j=1}^{M^d}2\left(\lambda\|\theta^*\|_2^2+\gamma_{T,j}^2+|\mathcal{D}_{T,j}|\Delta^2\right)\log\frac{\det(V_{T,j})}{\det(\lambda I)}\\
&\leq \frac{1}{\underline{b}^2}\sum_{j=1}^{M^d}2\left(\lambda\|\theta^*\|_2^2+\gamma_{T,j}^2+|\mathcal{D}_{T,j}|\Delta^2\right)\log\frac{\det(V_{T,j})}{\det(\lambda I)}\\
&=\mathcal{O}(M^d\log^2 T+\sum_{j=1}^{M^d}|\mathcal{D}_{T,j}|\Delta^2\log T)\\
&=\mathcal{O}(M^d\log^2 T+T\Delta^2\log T),
\end{aligned}
\tag{EC.13}
$$

where the fourth inequality holds because $\det(V_{T,j})\leq(\lambda+T)^{d^{k_g+1}}$ for all j by determinant-trace inequality (see, e.g., Abbasi-Yadkori et al. 2011) and $\gamma_{T,j}\leq\sigma\sqrt{2d^{k_g+1}\log(\lambda+T)-2\log\epsilon}$, and the last equality holds because of the fact that $\sum_{j=1}^{M^d}|\mathcal{D}_{T,j}|=T$. Now we are going to bound the second term of Eq. (EC.12). We have the following claim on the bound of $(b-\hat{b}_t)^2$. For a fixed $t\in[T]$,

$$
\mathbb{P}\left((b-\hat{b}_t)^2\geq\epsilon\right)\leq 2\exp(-\epsilon c_{b1}t^{\frac{1}{2}}),
\tag{EC.14}
$$

where $c_{b1}=1/(4(\sigma^2+2b^2\overline{p}^2+2\overline{g}^2))$. The proof of Eq. (EC.14) is delayed to the last part of this proof. Eq. (EC.14) indicates that $\mathbb{P}\left((b-\hat{b}_t)^2\geq\frac{\log\frac{2}{\delta}}{c_{b_1}\sqrt{t}}\right)\leq\delta$. Define an good event $\mathcal{A}:=\{\forall t\in[T],(b-\hat{b}_t)^2\leq\frac{\log\frac{2}{\delta}}{c_{b_1}\sqrt{t}}\}$. By the union bound, we can claim that $\mathbb{P}(\mathcal{A})\geq 1-T\delta$. Under event $\mathcal{A}$, the second term of Eq. (EC.12) can be written as

$$
\frac{1}{\underline{b}^2}\sum_{t=1}^{T}\|\phi(x_t)\|^2_{V_{t,j_t}^{-1}}|\mathcal{D}_{t,j_t}||b-\hat{b}_t|^2\overline{p}^2\leq\frac{\overline{p}^2\log\frac{2}{\delta}}{c_{b_1}\underline{b}^2}\sum_{t=1}^{T}\frac{\|\phi(x_t)\|^2_{V_{t,j_t}^{-1}}||\mathcal{D}_{t,j_t}|}{\sqrt{t}}.
\tag{EC.15}
$$

What remains to be tackled is the RHS of Eq. (EC.15). The first important observation is the boundedness of $\sum_{t=1}^{s}\|\phi(x_t)\|^2_{V_{t,j_t}^{-1}}||\mathcal{D}_{t,j_t}|$ for any $s\in[T]$ and any realization of $j_1,\cdots,j_T$ as follows,

$$
\begin{aligned}
\sum_{t=1}^{s}\|\phi(x_t)\|^2_{V_{t,j_t}^{-1}}||\mathcal{D}_{t,j_t}|&=\sum_{j=1}^{M^d}\sum_{t=1}^{s}\|\phi(x_t)\|^2_{V_{t,j}^{-1}}||\mathcal{D}_{t,j}|1_{x_t\in\mathbf{M}_j}\\
&\leq\sum_{j=1}^{M^d}|\mathcal{D}_{s,j}|\sum_{t=1}^{s}\|\phi(x_t)\|^2_{V_{t,j}^{-1}}1_{x_t\in\mathbf{M}_j}\\
&\leq\sum_{j=1}^{M^d}2|\mathcal{D}_{s,j}|\log\frac{\det(V_{s,j})}{\det(\lambda I)}\\
&\leq 2d^{k_g+1}\log(\lambda+s)\sum_{j=1}^{M^d}|\mathcal{D}_{s,j}|\\
&\leq 2d^{k_g+1}\log(\lambda+T)s.
\end{aligned}
$$

Therefore, we can bound the RHS of Eq. (EC.15) by the optimal value of the following optimization problem,

$$\begin{aligned} \max_{v_1,\cdots,v_T} \quad & \sum_{t=1}^{T} \frac{v_t}{\sqrt{t}} \\ \text{s.t.} \quad & \sum_{t=1}^{s} v_t \le 2d^{k_g+1}\log(\lambda+T)s \qquad \forall s\in[T] \\ & v_t \ge 0 \qquad \forall t\in[T] \end{aligned} \tag{EC.16}$$

By the KKT conditions, the optimal solution to the optimization problem (EC.16) is $v_1=\cdots=v_T=2d^{k_g+1}\log(\lambda+T)$. Therefore, Eq. (EC.15) can be controlled by

$$\frac{1}{\overline{b}^2}\sum_{t=1}^{T}\|\phi(x_t)\|^2_{V_{t,j_t}^{-1}}|\mathcal{D}_{t,j_t}||b-\hat{b}_t|^2\overline{p}^2 \le \frac{2\overline{p}^2\log\frac{2}{\delta}}{c_{b_1}}d^{k_g+1}\log(\lambda+T)\sum_{t=1}^{T}\frac{1}{\sqrt{t}} \le 4\frac{\overline{p}^2\log\frac{2}{\delta}}{c_{b_1}}d^{k_g+1}\log(\lambda+T)\sqrt{T}. \tag{EC.17}$$

Putting Eqs. (EC.12) (EC.13) and (EC.17) together, by choosing $\delta=\epsilon=\frac{1}{T^2}$, we can transform the high probability bound into the bound of the expectation as follows.

$$\sum_{t=1}^{T}\mathbb{E}\left[\left(\frac{\theta_j^*\phi(x_t)-\hat{\theta}_{t,j_t}\phi(x_t)}{2\hat{b}_t}\right)^2\right] = \widetilde{\mathcal{O}}\left(M^d+\frac{L^2T}{M^{2k_g}}+\sqrt{T}\right) \tag{EC.18}$$

The last term on the RHS of (EC.1) comes from the regret of the random shock added to the greedy policy for exploration, and can be bounded by $\mathcal{O}(t^{-1/2})$:

$$\mathbb{E}[\left(p_t^0-p_t\right)^2] \le 2\mathbb{E}[(p_t^0-p_t^g)^2+(p_t^g-p_t)^2] \le 4\delta_t^2 = \frac{4}{\sqrt{t}}. \tag{EC.19}$$

Finally, putting Eqs. (EC.1), (EC.5), (EC.18) and (EC.19) together, we obtain the following upper bound on the total expected regret:

$$\sum_{t=1}^{T}\mathbb{E}[r_t] = \widetilde{\mathcal{O}}\left(\frac{L^2T}{M^{2k_g}}+\sqrt{T}+M^d\right). \tag{EC.20}$$

By setting $M=\lceil (L^2T)^{\frac{1}{d+2k_g}}\rceil$, we obtain the upper bound $\widetilde{\mathcal{O}}(\sqrt{T}+(L^2T)^{\frac{d}{d+2k_g}}) = \widetilde{\mathcal{O}}(\sqrt{T}\vee(L^2T)^{\frac{d}{d+2k_g}})$ in Theorem 1.

What remains to be proved is Eq. (EC.14). By the definition of $\hat{b}_{t+1}$, we have

$$\begin{aligned} \mathbb{P}\left((b-\hat{b}_{t+1})^2\ge\epsilon\right) &\le \mathbb{P}\left(\left(b-\frac{\sum_{s=1}^{t}\Delta_s d_s}{\sum_{s=1}^{t}(\Delta_s)^2}\right)^2\ge\epsilon\right) \\ &= \mathbb{P}\left(\left(\frac{\sum_{s=1}^{t}\Delta_s(bp_s^g+g(x_s)+\varepsilon_s)}{\sum_{s=1}^{t}(\Delta_s)^2}\right)^2\ge\epsilon\right) \\ &\le \mathbb{P}\left(\frac{\sum_{s=1}^{t}\Delta_s(bp_s^g+g(x_s)+\varepsilon_s)}{\sum_{s=1}^{t}(\Delta_s)^2}\ge\sqrt{\epsilon}\right)+\mathbb{P}\left(\frac{\sum_{s=1}^{t}\Delta_s(bp_s^g+g(x_s)+\varepsilon_s)}{\sum_{s=1}^{t}(\Delta_s)^2}\le-\sqrt{\epsilon}\right) \end{aligned} \tag{EC.21}$$

Note that $M_t^b := \sum_{s=1}^t \Delta_s(bp_s^g + g(x_s) + \varepsilon_s)$ is a martingale because

$$\mathbb{E}[\Delta_t(bp_t^g + g(x_t) + \varepsilon_t) \mid H_t] = (bp_t^g + g(x_t))\mathbb{E}[\Delta_t \mid H_t] + \mathbb{E}[\Delta_t\varepsilon_t \mid H_t] = 0.$$

Then, by Markov inequality,

$$\begin{aligned}
\mathbb{P}\left(\frac{\sum_{s=1}^t \Delta_s(bp_s^g + g(x_s) + \varepsilon_s)}{\sum_{s=1}^t (\Delta_s)^2} \geq \sqrt{\epsilon}\right) &\leq \frac{\mathbb{E}[e^{\eta \sum_{s=1}^t \Delta_s(bp_s^g + g(x_s) + \varepsilon_s)}]}{e^{\eta\sqrt{\epsilon}\sum_{s=1}^t \delta_s^2}} \\
&\leq \frac{\mathbb{E}[M_{t-1}^b \mathbb{E}[e^{\eta\Delta_t(bp_t^g + g(x_t) + \varepsilon_t)} \mid H_t]]}{e^{\eta\sqrt{\epsilon}\sum_{s=1}^t (\Delta_s)^2}} \\
&= \frac{\mathbb{E}[M_{t-1}^b \mathbb{E}[e^{\eta\Delta_t(bp_t^g + g(x_t))}\mathbb{E}[e^{\eta\Delta_t\varepsilon_t} \mid H_t, \Delta_t] \mid H_t]]}{e^{\eta\sqrt{\epsilon}\sum_{s=1}^t \delta_s^2}} \\
&\leq \frac{\mathbb{E}[M_{t-1}^b e^{\frac{\eta^2\delta_t^2\sigma^2}{2}} \mathbb{E}[e^{\eta\Delta_t(bp_t^g + g(x_t))} \mid H_t]]}{e^{\eta\sqrt{\epsilon}\sum_{s=1}^t \delta_s^2}} \\
&\leq \frac{\mathbb{E}\left[M_{t-1}^b e^{\frac{\eta^2\delta_t^2\sigma^2}{2}} e^{\frac{\eta^2\delta_t^2(bp_t^g + g(x_t))^2}{2}}\right]}{e^{\eta\sqrt{\epsilon}\sum_{s=1}^t \delta_s^2}} \\
&\leq \frac{\mathbb{E}\left[M_{t-1}^b e^{\frac{\eta^2\delta_t^2(\sigma^2 + 2b^2\overline{p}^2 + 2\overline{g}^2)}{2}}\right]}{e^{\eta\sqrt{\epsilon}\sum_{s=1}^t \delta_s^2}} \\
&\leq \frac{e^{\sum_{s=1}^t \frac{\eta^2\delta_s^2(\sigma^2 + 2b^2\overline{p}^2 + 2\overline{g}^2)}{2}}}{e^{\eta\sqrt{\epsilon}\sum_{s=1}^t \delta_s^2}} \\
&\leq \exp(-\epsilon c_{b1}(t+1)^{1/2}),
\end{aligned}$$

where $c_{b1} := \frac{1}{4(\sigma^2 + 2b^2\overline{p}^2 + 2\overline{g}^2)}$, the third inequality holds because Δ_t and ε_t are independent and ε_t is sub-Gaussian, the fourth inequality is due to Δ_t is sub-Gaussian with parameter δ^2, the fifth inequality adopts Cauthy-Schwarz inequality, and in the last inequality, we set $\eta = \frac{\sqrt{\epsilon}}{\sigma^2 + 2b^2\overline{p}^2 + 2\overline{g}}$ and use the fact that $\sum_{s=1}^t \delta_s^2 \geq \frac{1}{2}(t+1)^{1-2\alpha}$. Therefore, by symmetry, we can bound Eq. (EC.21) as

$$\mathbb{P}\left((b - \hat{b}_t)^2 \geq \epsilon\right) \leq 2\exp(-\epsilon c_{b1} t^{1/2}).$$

We finish the proof. Q.E.D.

A.2. Proof for Theorem 2

We fix the price elasticity to be $\underline{b}$, the distribution of contexts to be a uniform distribution on $\bigcup_{j=1}^{M^d}\{x \in \mathbf{M}_j : D(x, \partial\mathbf{M}_j) > \frac{1}{4^{1/k_g} M}\}$, and the distribution of random noise to be normal distribution. For any policy π, we establish the following lower bound on its worst-case regret by restricting to the functions $g_{\mathbf{w}}(\cdot)$ constructed in Eq. (7):

$$\sup_{g \in \mathcal{G}_d(k_g, L)} R_g^\pi(T) \geq \sup_{g \in \{g_{\mathbf{w}} : \mathbf{w} \in \{0,1\}^{M^d}\}} R_g^\pi(T)$$

$$
\begin{aligned}
&= |\underline{b}| \sup_{g\in\{g_{\mathbf{w}}:\mathbf{w}\in\{0,1\}^{M^d}\}} \sum_{t=1}^{T} \mathbb{E}_g^{\pi}\left[(p^*(x_t)-p_t)^2\right] \\
&\geq \frac{|\underline{b}|}{2^{M^d}} \sum_{\mathbf{w}\in\{0,1\}^{M^d}} \sum_{t=1}^{T} \mathbb{E}_{g_{\mathbf{w}}}^{\pi}\left[(p^*(x_t)-p_t)^2\right] \\
&= \frac{|\underline{b}|}{2^{M^d}} \sum_{\mathbf{w}\in\{0,1\}^{M^d}} \sum_{t=1}^{T}\sum_{j=1}^{M^d} \mathbb{E}_{g_{\mathbf{w}}}^{\pi}\left[(p^*(x_t)-p_t)^2\mathbb{I}_{\{x_t\in\mathbf{M}_j\}}\right] \\
&= \frac{|\underline{b}|}{2^{M^d}M^d} \sum_{j=1}^{M^d} \sum_{\mathbf{w}_{-j}\in\{0,1\}^{M^d-1}} \sum_{w_j\in\{0,1\}} \sum_{t=1}^{T} \mathbb{E}_{g_{(\mathbf{w}_{-j},w_j)}}^{\pi}\left[(p^*(x_t)-p_t)^2 \mid x_t\in\mathbf{M}_j\right],
\end{aligned}
\tag{EC.22}
$$

where in the second identity, we use $(\mathbf{w}_{-j}, w_j)$ to denote an index $\mathbf{w}$ whose j-th coordinate is w_j and the other coordinates are $\mathbf{w}_{-j}$, and the fact that $\mathbb{E}_{g_{\mathbf{w}}}^{\pi}[(p^*(x_t)-p_t)^2\mathbb{I}_{\{x_t\in\mathbf{M}_j\}}] = \mathbb{E}_{g_{\mathbf{w}}}^{\pi}[(p^*(x_t)-p_t)^2 \mid x_t\in\mathbf{M}_j]\mathcal{P}_u(x_t\in\mathbf{M}_j)$ and $\mathcal{P}_u(x_t\in\mathbf{M}_j)=\frac{1}{M^d}$ by our construction. Note that whenever x falls into $\mathbf{M}_j$, the optimal price $p^*(x)$ associated with any function $g_{\mathbf{w}}(\cdot)$ is

$$
p^*(x) = \begin{cases} \frac{\underline{p}+\overline{p}}{2} & w_j = 0 \\ \frac{\underline{p}+\overline{p}}{2} + \frac{L}{16M^{k_g}|\underline{b}|} & w_j = 1. \end{cases}
\tag{EC.23}
$$

We have the following lower bound on $\mathbb{E}_{g_{(\mathbf{w}_{-j},w_j)}}^{\pi}[(p^*(x_t)-p_t)^2 \mid x_t\in\mathbf{M}_j]$ in the RHS of (EC.22):

$$
\mathbb{E}_{g_{(\mathbf{w}_{-j},0)}}^{\pi}\left[(p^*(x_t)-p_t)^2\,\middle|\,x_t\in\mathbf{M}_j\right] \geq \left(\frac{L}{32M^{k_g}|\underline{b}|}\right)^2 \mathbb{P}_{g_{(\mathbf{w}_{-j},0)}}^{\pi,t}\left(\left\{p_t\geq\frac{\underline{p}+\overline{p}}{2}+\frac{L}{32M^{k_g}|\underline{b}|}\right\}\,\middle|\,x_t\in\mathbf{M}_j\right),
\tag{EC.24}
$$

$$
\mathbb{E}_{g_{(\mathbf{w}_{-j},1)}}^{\pi}\left[(p^*(x_t)-p_t)^2\,\middle|\,x_t\in\mathbf{M}_j\right] \geq \left(\frac{L}{32M^{k_g}|\underline{b}|}\right)^2 \mathbb{P}_{g_{(\mathbf{w}_{-j},1)}}^{\pi,t}\left(\left\{p_t\leq\frac{\underline{p}+\overline{p}}{2}+\frac{L}{32M^{k_g}|\underline{b}|}\right\}\,\middle|\,x_t\in\mathbf{M}_j\right),
\tag{EC.25}
$$

where $\mathbb{P}_{g_{(\mathbf{w}_{-j},w_j)}}^{\pi,t}$ denotes the probability measure for history under policy π and demand $g_{(\mathbf{w}_{-j},w_j)}$.

Bretagnolle–Huber inequality (see Bretagnolle and Huber 1979) guarantees that

$$
\begin{aligned}
&\mathbb{P}_{g_{(\mathbf{w}_{-j},0)}}^{\pi,t}\left(p_t\geq\frac{\underline{p}+\overline{p}}{2}+\frac{L}{32M^{k_g}|\underline{b}|}\,\middle|\,x_t\in\mathbf{M}_j\right) + \mathbb{P}_{g_{(\mathbf{w}_{-j},1)}}^{\pi,t}\left(p_t<\frac{\underline{p}+\overline{p}}{2}+\frac{L}{32M^{k_g}|\underline{b}|}\,\middle|\,x_t\in\mathbf{M}_j\right) \\
&\geq \frac{1}{2}\exp\left(-\mathrm{KL}\left(\mathbb{P}_{g_{(\mathbf{w}_{-j},0)}}^{\pi,t},\mathbb{P}_{g_{(\mathbf{w}_{-j},1)}}^{\pi,t}\right)\right).
\end{aligned}
\tag{EC.26}
$$

We next focus on analyzing the KL-divergence between the two probability measures $\mathbb{P}_{g_{(\mathbf{w}_{-j},0)}}^{\pi,t}(\cdot \mid x_t\in\mathbf{M}_j)$ and $\mathbb{P}_{g_{(\mathbf{w}_{-j},1)}}^{\pi,t}(\cdot \mid x_t\in\mathbf{M}_j)$. Noting the following identity

$$
\begin{aligned}
&\mathbb{P}_{g_{(\mathbf{w}_{-j},w_j)}}^{\pi,t}(X_1,P_1,d_1,\cdots,X_t,P_t \mid x_t\in\mathbf{M}_j) \\
&= \mathbb{P}_{g_{(\mathbf{w}_{-j},0)}}^{\pi,t-1}(X_1,P_1,d_1,\cdots,X_{t-1},P_{t-1},d_{t-1}) \times \pi(X_t,P_t \mid x_t\in\mathbf{M}_j, X_1,P_1,d_1,\cdots,X_{t-1},P_{t-1},d_{t-1}),
\end{aligned}
$$

and denoting $\pi(X_t, P_t \mid x_t \in \mathbf{M}_j, X_1, P_1, d_1, \cdots, X_{t-1}, P_{t-1}, d_{t-1})$ as π_t^j, we obtain

$$\begin{aligned}\mathrm{KL}\big(\mathbb{P}^{\pi,t}_{g_{(\mathbf{w}_{-j},0)}}(\cdot \mid x_t \in \mathbf{M}_j), \mathbb{P}^{\pi}_{g_{(\mathbf{w}_{-j},1)}}(\cdot \mid x_t \in \mathbf{M}_j)\big) &= \mathrm{KL}(\mathbb{P}^{\pi,t-1}_{g_{(\mathbf{w}_{-j},0)}} \times \pi_t^j, \mathbb{P}^{\pi,t-1}_{g_{(\mathbf{w}_{-j},1)}} \times \pi_t^j) \\ &= \mathrm{KL}(\mathbb{P}^{\pi,t-1}_{g_{(\mathbf{w}_{-j},0)}}, \mathbb{P}^{\pi,t-1}_{g_{(\mathbf{w}_{-j},1)}}) + \mathbb{E}^{\pi,t-1}_{g_{(\mathbf{w}_{-j},0)}}[\mathrm{KL}(\pi_t^j, \pi_t^j)] \\ &= \mathrm{KL}(\mathbb{P}^{\pi,t-1}_{g_{(\mathbf{w}_{-j},0)}}, \mathbb{P}^{\pi,t-1}_{g_{(\mathbf{w}_{-j},1)}}), \end{aligned} \tag{EC.27}$$

where the second identity follows from the chain rule of the KL divergence. Moreover, $\mathbb{P}^{\pi,t}_{g_{(\mathbf{w}_{-j},w_j)}}$ can be further decomposed as follows:

$$\begin{aligned}&\mathbb{P}^{\pi,t-1}_{g_{(\mathbf{w}_{-j},w_j)}}(X_1, P_1, d_1, \cdots, X_{t-1}, P_{t-1}, d_{t-1}) \\ &= \mathbb{P}^{\pi,t-2}_{g_{(\mathbf{w}_{-j},w_j)}}(X_1, P_1, d_1, \cdots, X_{t-2}, P_{t-2}, d_{t-2}) \times \mathcal{P}_u(X_{t-1}) \\ &\quad \times \pi(P_{t-1} \mid X_1, P_1, d_1, \cdots, X_{t-2}, P_{t-2}, d_{t-2}, X_{t-1}) \times \mu_{g_{(\mathbf{w}_{-j},w_j)}}(d_{t-1} | P_{t-1}, X_{t-1}). \end{aligned} \tag{EC.28}$$

Denoting $\pi(P_{t-1} \mid X_1, P_1, d_1, \cdots, X_{t-2}, P_{t-2}, d_{t-2}, X_{t-1})$ by $\pi_{\mathcal{F}_{t-2}, x_{t-1}}$, we get

$$\begin{aligned}&\mathrm{KL}(\mathbb{P}^{\pi,t-1}_{g_{(\mathbf{w}_{-j},0)}}, \mathbb{P}^{\pi,t-1}_{g_{(\mathbf{w}_{-j},1)}}) \\ &= \mathrm{KL}(\mathbb{P}^{\pi,t-2}_{g_{(\mathbf{w}_{-j},0)}}, \mathbb{P}^{\pi,t-2}_{g_{(\mathbf{w}_{-j},1)}}) + \mathbb{E}^{\pi,t-2}_{g_{\mathbf{w}_{-j},0}}[\mathrm{KL}(\mathcal{P}_u \times \pi_{\mathcal{F}_{t-2},x_{t-1}} \times \mu_{g_{(\mathbf{w}_{-j},0)}}, \mathcal{P}_u \times \pi_{\mathcal{F}_{t-2},x_{t-1}} \times \mu_{g_{(\mathbf{w}_{-j},1)}})] \\ &= \mathrm{KL}(\mathbb{P}^{\pi,t-2}_{g_{(\mathbf{w}_{-j},0)}}, \mathbb{P}^{\pi,t-2}_{g_{(\mathbf{w}_{-j},1)}}) + \mathbb{E}^{\pi,t-2}_{g_{(\mathbf{w}_{-j},0)}}[\mathrm{KL}(\mathcal{P}_u \times \pi_{\mathcal{F}_{t-2},x_{t-1}}, \mathcal{P}_u \times \pi_{\mathcal{F}_{t-2},x_{t-1}})] \\ &\quad + \mathbb{E}^{\pi,t-2}_{g_{(\mathbf{w}_{-j},0)}}[\mathbb{E}_{\mathcal{P}_u \times \pi_{\mathcal{F}_{t-2},x_{t-1}}}[\mathrm{KL}(\mu_{g_{(\mathbf{w}_{-j},0)}}, \mu_{g_{(\mathbf{w}_{-j},1)}})] \\ &= \mathrm{KL}(\mathbb{P}^{\pi,t-2}_{g_{(\mathbf{w}_{-j},0)}}, \mathbb{P}^{\pi,t-2}_{g_{(\mathbf{w}_{-j},1)}}) + \mathbb{E}^{\pi,t-2}_{g_{(\mathbf{w}_{-j},0)}}[\mathbb{E}_{\mathcal{P}_u \times \pi_{\mathcal{F}_{t-2},x_{t-1}}}[\mathrm{KL}(\mu_{g_{(\mathbf{w}_{-j},0)}}, \mu_{g_{(\mathbf{w}_{-j},1)}})], \end{aligned} \tag{EC.29}$$

where the first and second identities follow from the chain rule of KL divergence. Since we have assumed that ε follows a normal distribution with variance σ^2, the following equations hold:

$$\begin{aligned}\mathrm{KL}(\mu_{g_{(\mathbf{w}_{-j},0)}}(\cdot \mid p_{t-1}, x_{t-1}), \mu_{g_{(\mathbf{w}_{-j},1)}}(\cdot \mid p_{t-1}, x_{t-1})) &= \frac{1}{2\sigma^2}\left(\underline{b}p_{t-1} + g_{(\mathbf{w}_{-j},0)}(x_{t-1}) - \underline{b}p_{t-1} - g_{(\mathbf{w}_{-j},1)}(x_{t-1})\right)^2 \\ &= \frac{1}{2\sigma^2}\left(g_{(\mathbf{w}_{-j},0)}(x_{t-1}) - g_{(\mathbf{w}_{-j},1)}(x_{t-1})\right)^2 \\ &\le \frac{1}{2\sigma^2}\left(\frac{L}{8M^{k_g}}\right)^2 \mathbb{I}_{\{x_{t-1} \in \mathbf{M}_j\}}, \end{aligned}$$

where the last inequality holds because $g_{(\mathbf{w}_{-j},0)}$ and $g_{(\mathbf{w}_{-j},1)}$ only differ in $\mathbf{M}_j$. Plugging the above equation into Eq. (EC.29), we obtain

$$\begin{aligned}\mathrm{KL}(\mathbb{P}^{\pi,t}_{g_{(\mathbf{w}_{-j},0)}}, \mathbb{P}^{\pi,t}_{g_{(\mathbf{w}_{-j},1)}}) &= \mathrm{KL}(\mathbb{P}^{\pi,t-1}_{g_{(\mathbf{w}_{-j},0)}}, \mathbb{P}^{\pi,t-1}_{g_{(\mathbf{w}_{-j},1)}}) + \frac{1}{2\sigma^2 M^d}\left(\frac{L}{8M^{k_g}}\right)^2 \\ &= \frac{L^2}{128\sigma^2 M^{d+2k_g}} t. \end{aligned} \tag{EC.30}$$

where the first identity is holds since $\mathbb{E}_{\mathcal{P}_u \times \pi_{\mathcal{F}_{t-2},x_{t-1}}}[\mathbb{I}_{\{x_{t-1} \in \mathbf{M}_j\}}] = \frac{1}{M^d}$, and the second identity follows by repeatedly applying the first identity.

Combining Eqs. (EC.22), (EC.24), (EC.26), (EC.27) and (EC.30), we have

$$\sup_{g\in\mathcal{G}_d(k_g,L)} R_g^\pi(T) \geq \frac{|\underline{b}|}{2^{M^d}M^d}\sum_{j=1}^{M^d}\sum_{\mathbf{w}_{-j}\in\{0,1\}^{M^d-1}}\sum_{t=1}^{T}\frac{1}{2}\left(\frac{1}{16M^{k_g}|\underline{b}|}\right)^2\exp\left(-\frac{L^2}{128\sigma^2M^{d+2k_g}}t\right)$$
$$\geq \frac{|\underline{b}|}{4}T\left(\frac{L}{32M^{k_g}|\underline{b}|}\right)^2\exp\left(-\frac{L^2}{128\sigma^2M^{d+2k_g}}T\right).$$

Since $M=\lceil (L^2T)^{\frac{1}{d+2k_g}}\rceil$, we obtain the lower bound $\Omega((L^2T)^{\frac{d}{d+2k_g}})$ from the above inequality. Q.E.D.

A.3. Proof for Lemma 1

We prove that for any $x,y\in[0,1]^d$, $|g_{\mathbf{w}}(x)-g_{\mathbf{w}}(y)|\leq L||x-y||^{k_g}$ by considering two cases: x and y fall into the same bin in case 1, and x and y fall into different bins in case 2.

Case 1: $x,y\in\mathbf{M}_j$ for some $j\in[M^d]$. When x and y fall into the same bin $\mathbf{M}_j$, we divide the proof into four subcases.

Subcase 1.1: $w_j=0$, or $w_j=1$, $D(x,\partial\mathbf{M}_j)>\frac{1}{4^{1/k_g}M}$ and $D(y,\partial\mathbf{M}_j)>\frac{1}{4^{1/k_g}M}$. In this subcase, we have $g_{\mathbf{w}}(x)=g_{\mathbf{w}}(y)$, and the result is trivial.

Subcase 1.2: $w_j=1$, $D(x,\partial\mathbf{M}_j)\leq\frac{1}{4^{1/k_g}M}$ and $D(y,\partial\mathbf{M}_j)\leq\frac{1}{4^{1/k_g}M}$. Without loss of generality, we assume that $g_{\mathbf{w}}(x)\leq g_{\mathbf{w}}(y)$. Then we have the following equation:

$$\begin{aligned}\frac{L}{2}||x-y||^{k_g}+g_{\mathbf{w}}(x)&=\frac{L}{2}||x-y||^{k_g}+\frac{L}{2}(D(x,\partial\mathbf{M}_j))^{k_g}+|\underline{b}|(\underline{p}+\overline{p})\\&=\frac{L}{2}||x-y||^{k_g}+\frac{L}{2}\min_{z\in\partial\mathbf{M}_j}||z-x||^{k_g}+|\underline{b}|(\underline{p}+\overline{p})\\&=\frac{L}{2}\min_{z\in\partial\mathbf{M}_j}(||x-y||^{k_g}+||z-x||^{k_g})+|\underline{b}|(\underline{p}+\overline{p}).\end{aligned}\tag{EC.31}$$

If the following inequality holds: for $0<k_g\leq 1$,

$$||a+b||^{k_g}\leq||a||^{k_g}+||b||^{k_g},\ \forall a,b\in\mathbb{R}^d,\tag{EC.32}$$

then we have from (EC.31) that

$$\frac{L}{2}||x-y||^{k_g}+g_{\mathbf{w}}(x)\geq\frac{L}{2}\min_{z\in\partial\mathbf{M}_j}||z-y||^{k_g}+|\underline{b}|(\underline{p}+\overline{p})=\frac{L}{2}(D(y,\partial\mathbf{M}_j))^{k_g}+|\underline{b}|(\underline{p}+\overline{p})=g_{\mathbf{w}}(y),\tag{EC.33}$$

which then implies $|g_{\mathbf{w}}(x)-g_{\mathbf{w}}(y)|\leq\frac{L}{2}||x-y||^{k_g}$.

We now show (EC.32). Note that (EC.32) is simply the triangle inequality when $k_g=1$. When $0<k_g<1$, let $k_g{}'\in\mathbb{R}^+$ be such that $\frac{1}{k_g}+\frac{1}{k_g{}'}=1$. By applying Hölder's inequality, we have for any $a,b\in\mathbb{R}^+$,

$$(a+b)^{k_g}\leq\left((a^{k_g}+b^{k_g})^{\frac{1}{k_g}}(1^{k_g{}'}+1^{k_g{}'})^{\frac{1}{k_g{}'}}\right)^{k_g}=2^{k_g-1}(a^{k_g}+b^{k_g})<a^{k_g}+b^{k_g},\tag{EC.34}$$

Applying (EC.34) and the triangle inequality, we have for any $a, b \in \mathbb{R}^d$,

$$||a+b||^{k_g} \le (||a|| + ||b||)^{k_g} < ||a||^{k_g} + ||b||^{k_g},$$

which finishes the proof of (EC.32).

Subcase 1.3: $w_j = 1$, $D(x, \partial \mathbf{M}_j) \le \frac{1}{4^{1/k_g} M}$ and $D(y, \partial \mathbf{M}_j) > \frac{1}{4^{1/k_g} M}$. Let $\hat{\mathbf{M}}_j := \{z \in \mathbf{M}_j : D(z, \partial \mathbf{M}_j) > \frac{1}{4^{1/k_g} M}\}$. Since $\text{Proj}(x, \hat{\mathbf{M}}_j) \in \partial \hat{\mathbf{M}}_j$, then we have $g_{\mathbf{w}}(y) = g_{\mathbf{w}}(\text{Proj}(x, \hat{\mathbf{M}}_j)) = |\underline{b}|(\underline{p} + \overline{p}) + \frac{L}{8M^{k_g}}$ and

$$|g_{\mathbf{w}}(x) - g_{\mathbf{w}}(y)| = |g_{\mathbf{w}}(x) - g_{\mathbf{w}}(\text{Proj}(x, \hat{\mathbf{M}}_j))|. \tag{EC.35}$$

Since $\partial \hat{\mathbf{M}}_j = \{z \in \mathbf{M}_j : D(z, \partial \mathbf{M}_j) = \frac{1}{4^{1/k_g} M}\}$ and $\text{Proj}(x, \hat{\mathbf{M}}_j) \in \partial \hat{\mathbf{M}}_j$, it then follows that $D(\text{Proj}(x, \hat{\mathbf{M}}_j), \partial \mathbf{M}_j) = \frac{1}{4^{1/k_g} M}$. Since $D(x, \partial \mathbf{M}_j) \le \frac{1}{4^{1/k_g} M}$ from the assumption, by applying the result in subcase 1.2, we obtain

$$|g_{\mathbf{w}}(x) - g_{\mathbf{w}}(\text{Proj}(x, \hat{\mathbf{M}}_j))| \le \frac{L}{2} ||x - \text{Proj}(x, \hat{\mathbf{M}}_j)||^{k_g} = \frac{L}{2} \min_{z \in \hat{\mathbf{M}}_j} ||x - z||^{k_g} \le \frac{L}{2} ||x - y||^{k_g}, \tag{EC.36}$$

where the last inequality holds due to $y \in \hat{\mathbf{M}}_j$. Combining (EC.35) with (EC.36), we obtain $|g_{\mathbf{w}}(x) - g_{\mathbf{w}}(y)| \le \frac{L}{2} ||x - y||^{k_g}$.

Subcase 1.4: $w_j = 1$, $D(x, \partial \mathbf{M}_j) > \frac{1}{4^{1/k_g} M}$ and $D(y, \partial \mathbf{M}_j) \le \frac{1}{4^{1/k_g} M}$. The proof of this subcase is similar to subcase 1.3, and is omitted for brevity.

Case 2: $x \in \mathbf{M}_i$ and $y \in \mathbf{M}_j$ for $i \ne j$. When x and y fall into different bins, we divide the proof into five subcases.

Subcase 2.1: $w_i = w_j = 0$, or if $w_i = w_j = 1$, $D(x, \partial \mathbf{M}_i) > \frac{1}{4^{1/k_g} M}$ and $D(y, \partial \mathbf{M}_j) > \frac{1}{4^{1/k_g} M}$. In this subcase, we have $g_{\mathbf{w}}(x) = g_{\mathbf{w}}(y)$ and the result is trivial.

Subcase 2.2: $w_i = 0$ and $w_j = 1$. In this subcase, we have

$$\begin{aligned}
|g_{\mathbf{w}}(x) - g_{\mathbf{w}}(y)| &\le \frac{L}{2} (D(y, \partial \mathbf{M}_j))^{k_g} && \text{(EC.37)} \\
&\le \frac{L}{2} ||\text{Proj}(x, \partial \mathbf{M}_j) - y||^{k_g} && \text{(EC.38)} \\
&\le \frac{L}{2} ||\text{Proj}(x, \partial \mathbf{M}_j) - x||^{k_g} + \frac{L}{2} ||x - y||^{k_g} && \text{(EC.39)} \\
&= \frac{L}{2} ||\text{Proj}(x, \mathbf{M}_j) - x||^{k_g} + \frac{L}{2} ||x - y||^{k_g} && \text{(EC.40)} \\
&\le L ||x - y||^{k_g}. && \text{(EC.41)}
\end{aligned}$$

In the above equations, Eq. (EC.37) holds since under the assumption $w_i = 0$ and $w_j = 1$, if $D(y, \partial \mathbf{M}_j) \le \frac{1}{4^{1/k_g} M}$, $|g_{\mathbf{w}}(x) - g_{\mathbf{w}}(y)| = \frac{L}{2} (D(y, \partial \mathbf{M}_j))^{k_g}$, and if $D(y, \partial \mathbf{M}_j) > \frac{1}{4^{1/k_g} M}$, $|g_{\mathbf{w}}(x) - g_{\mathbf{w}}(y)| = \frac{L}{8M^{k_g}} \le \frac{L}{2} (D(y, \partial \mathbf{M}_j))^{k_g}$. Eq. (EC.38) follows from the definition of $D(y, \partial \mathbf{M}_j)$ and $\text{Proj}(x, \partial \mathbf{M}_j) \in \partial \mathbf{M}_j$. Eq. (EC.39) follows from (EC.32). Eq. (EC.40) holds since if $\text{Proj}(x, \mathbf{M}_j)$

is an interior point of $\mathbf{M}_j$, since $\mathbf{M}_j$ is a cubic, one can always construct a ball inside $\mathbf{M}_j$ with the center $\text{Proj}(x, \mathbf{M}_j)$, and the intersected point between the ball and the line connecting x and $\text{Proj}(x, \mathbf{M}_j)$ has a strictly shorter distance to x than $\text{Proj}(x, \mathbf{M}_j)$, leading to contradiction with the fact that $\text{Proj}(x, \mathbf{M}_j)$ is the closest point in the bin $\mathbf{M}_j$ to x. Thus, $\text{Proj}(x, \mathbf{M}_j)$ must be at the boundary $\partial\mathbf{M}_j$ and $\text{Proj}(x, \mathbf{M}_j) = \text{Proj}(x, \partial\mathbf{M}_j)$. Eq. (EC.41) follows from $y \in \mathbf{M}_j$ and the definition of $D(x, \partial\mathbf{M}_j)$.

Subcase 2.3: $w_i = 1$ and $w_j = 0$. The proof of this subcase is similar to subcase 2.2 and is omitted.

Subcase 2.4: $w_i = w_j = 1$, $D(x, \partial\mathbf{M}_i) \le \frac{1}{4^{1/k_g} M}$ and $D(y, \partial\mathbf{M}_j) \le \frac{1}{4^{1/k_g} M}$. Without loss of generality, we assume $g_{\mathbf{w}}(y) \ge g_{\mathbf{w}}(x)$. Then we have

$$\frac{L}{2}||x-y||^{k_g} + g_{\mathbf{w}}(x) = \frac{L}{2} \min_{z \in \partial\mathbf{M}_i} (||x-y||^{k_g} + ||z-x||^{k_g}) + |\underline{b}|(\underline{p} + \overline{p}) \ge \frac{L}{2} \min_{z \in \partial\mathbf{M}_i} ||z-y||^{k_g} + |\underline{b}|(\underline{p} + \overline{p}), \tag{EC.42}$$

where the inequality follows from (EC.32).

On the other hand, when K is sufficiently large, we have

$$\begin{aligned} g_{\mathbf{w}}(y) &= \frac{L}{2} \min_{z \in \partial\mathbf{M}_j} ||z-y||^{k_g} + |\underline{b}|(\underline{p} + \overline{p}) \\ &= \frac{L}{2} \min_{z \in \partial([-K,K]^d \setminus \text{int}(\mathbf{M}_j))} ||z-y||^{k_g} + |\underline{b}|(\underline{p} + \overline{p}) \qquad \text{(EC.43)} \\ &= \frac{L}{2} \min_{z \in [-K,K]^d \setminus \text{int}(\mathbf{M}_j)} ||z-y||^{k_g} + |\underline{b}|(\underline{p} + \overline{p}) \qquad \text{(EC.44)} \\ &\le \frac{L}{2} \min_{z \in \partial\mathbf{M}_i} ||z-y||^{k_g} + |\underline{b}|(\underline{p} + \overline{p}). \qquad \text{(EC.45)} \end{aligned}$$

In the above equations, Eq. (EC.43) holds since $\partial([-K,K]^d \setminus \text{int}(\mathbf{M}_j)) = \partial([-K,K]^d) \cup \partial\mathbf{M}_j$, and when K is sufficiently large, $D(y, \partial\mathbf{M}_j) < D(y, \partial([-K,K]^d))$ and $\text{Proj}(y, \partial(\mathbf{M}_j)) = \text{Proj}(y, \partial([-K,K]^d) \cup \partial\mathbf{M}_j)$. Eq. (EC.44) holds due to the same reason as (EC.40). Eq. (EC.45) holds since $\partial(\mathbf{M}_i) \subset [-K,K]^d$ and $\partial(\mathbf{M}_i) \cap \text{int}(\mathbf{M}_j) = \emptyset$ imply that $\partial(\mathbf{M}_i) \subset [-K,K]^d \setminus \text{int}(\mathbf{M}_j)$.

Combining Eqs. (EC.42) and (EC.45), we obtain $|g_{\mathbf{w}}(x) - g_{\mathbf{w}}(x)| \le \frac{L}{2}||x-y||^{k_g}$.

Subcase 2.5: $w_i = w_j = 1$, $D(x, \partial\mathbf{M}_j) \le \frac{1}{4^{1/k_g} M}$ and $D(y, \partial\mathbf{M}_j) > \frac{1}{4^{1/k_g} M}$. In this subcase, we have

$$\begin{aligned} |g_{\mathbf{w}}(x) - g_{\mathbf{w}}(y)| = \left| \frac{L}{2}(D(x, \partial\mathbf{M}_j))^{k_g} - \frac{L}{8M^{k_g}} \right| &= \frac{L}{8M^{k_g}} - \frac{L}{2}(D(x, \partial\mathbf{M}_j))^{k_g} \\ &< \frac{L}{2}(D(y, \partial\mathbf{M}_j))^{k_g} - \frac{L}{2}(D(x, \partial\mathbf{M}_j))^{k_g}. \end{aligned}$$

The remaining analysis is similar to subcase 2.4, and is omitted for brevity. Q.E.D.

Appendix B. Proofs for Statements in Section 3

B.1. Proof for Theorem 3

As preparations, we establish the following lemma showing that the true demand function $f(p) + a^\top x$ can be well-approximated by the linear function $\hat{\theta}_{t,j}^\top \varphi(p) + \hat{a}_{t,j}^\top x$ within price segment $\mathbf{I}_j$ after

running Algorithm 3. This result is quite standard and can be obtained easily by modifying the analysis in Abbasi-Yadkori et al. (2011), and thus we omit the proof.

LEMMA EC.1. *For each $j \in [N]$, with probability at least $1-\epsilon$, the following event holds: for any $t \in [T]$, $p \in \mathbf{I}$ and $x \in [0,1]^d$,*

$$\left| f(p) + a^\top x - \left(\hat{\theta}_{t,j}^\top \varphi(p) + \hat{a}_{t,j}^\top x \right) \right| \le \gamma_{t,j} \sqrt{\phi(p,x)^\top V_{t,j}^{-1} \phi(p,x)} + \Delta. \tag{EC.46}$$

Let $p_t^* := \arg\max_{p \in [\underline{p}, \overline{p}]} p(f(p) + a^\top x_t)$ be the optimal price for period t and $i_t^* \in [N]$ denote the index for the price segment p_t^* belongs to. Conditioning on the events guaranteed by Lemma EC.1 for each price segment $i \in [N]$, we have

$$\begin{aligned}
r_t &= p_t^* \left(f(p_t^*) + a^\top x_t \right) - p_t \left(f(p_t) + a^\top x_t \right) \\
&\le \max_{p \in \mathbf{I}_{i_t^*}} p \left(\langle \hat{\theta}_{t,i_t^*}, \varphi(p) \rangle + \langle \hat{a}_{t,i_t^*}, x_t \rangle + \gamma_{t,i_t^*} \sqrt{\phi(p,x_t)^\top V_{t,i_t^*}^{-1} \phi(p,x_t)} + \Delta \right) - p_t \left(f(p_t) + a^\top x_t \right) \\
&\le \max_{i \in [N]} \max_{p \in \mathbf{I}_i} p \left(\langle \hat{\theta}_{t,i}, \varphi(p) \rangle + \langle \hat{a}_{t,i}, x_t \rangle + \gamma_{t,i} \sqrt{\phi(p,x_t)^\top V_{t,i}^{-1} \phi(p,x_t)} + \Delta \right) - p_t \left(f(p_t) + a^\top x_t \right) \\
&= p_t \left(\langle \hat{\theta}_{t,i_t}, \varphi(p_t) \rangle + \langle \hat{a}_{t,i_t}, x_t \rangle + \gamma_{t,i_t} \sqrt{\phi(p_t,x_t)^\top V_{t,i_t}^{-1} \phi(p_t,x_t)} + \Delta \right) - p_t \left(f(p_t) + a^\top x_t \right) \\
&\le p_t \left| \langle \hat{\theta}_{t,i_t}, \varphi(p_t) \rangle + \langle \hat{a}_{t,i_t}, x_t \rangle - \left(f(p_t) + a^\top x_t \right) \right| + p_t \left(\gamma_{t,i_t} \sqrt{\phi(p_t,x_t)^\top V_{t,i_t}^{-1} \phi(p_t,x_t)} + \Delta \right) \\
&\le 2\overline{p} \left(\gamma_{t,i_t} \sqrt{\phi(p_t,x_t)^\top V_{t,i_t}^{-1} \phi(p_t,x_t)} + \Delta \right),
\end{aligned} \tag{EC.47}$$

where the first inequality follows from Eq. (EC.46) in Lemma EC.1 and $p_t^* \in \mathbf{I}_t^*$ by definition, the second equality is based on the design of our Algorithm 2 (line 13), and the last inequality again follows from Eq. (EC.46) in Lemma EC.1. We now need to bound $\sqrt{\phi(p_t,x_t)^\top V_{t,i_t}^{-1} \phi(p_t,x_t)}$.

Let $n_{T,j} := \sum_{t=1}^T \mathbb{I}_{p_t \in \mathbf{I}_j}$ be the number of times for which prices $p_1, p_2, \ldots, p_T$ selected by our algorithm fall into $\mathbf{I}_j$, and $\mathcal{A}$ denote the event that Eq. (EC.47) in Lemma EC.1 holds for any $t \in [T]$ and $i \in [N]$. When $\mathcal{A}$ holds, from Eq. (EC.47), the total regret can be bounded as follows:

$$\begin{aligned}
\sum_{t=1}^T r_t &\le 2\overline{p} \sum_{t=1}^T \gamma_{t,i_t} \left\| \phi(p_t,x_t) \right\|_{V_{t,i_t}^{-1}} + 2\overline{p}\Delta T \\
&\le 2\overline{p} \sum_{j=1}^N \gamma_{T,j} \sum_{t=1}^T \left\| \phi(p_t,x_t) \right\|_{V_{t,j}^{-1}} \mathbb{I}_{p_t \in \mathbf{I}_j} + 2\overline{p}\Delta T \\
&\le 2\overline{p} \sum_{j=1}^N \gamma_{T,j} \sqrt{n_{T,j}} \sqrt{\sum_{t=1}^T \left\| \phi(p_t,x_t) \right\|_{V_{t,j}^{-1}}^2 \mathbb{I}_{p_t \in \mathbf{I}_j}} + 2\overline{p}\Delta T \\
&\le 2\overline{p} \sum_{j=1}^N \gamma_{T,j} \sqrt{n_{T,j}} \sqrt{2(\mathfrak{b}(k)+d+1) \log \left(1 + \frac{n_{T,j}}{\mathfrak{b}(k)+d+1} \right)} + 2\overline{p}\Delta T,
\end{aligned} \tag{EC.48}$$

where the second inequality holds since $\{\gamma_{t,j}: 1\le t\le T\}$ is an increasing sequence for each $j\in[N]$, the third inequality follows from Cauchy-Schwarz inequality, and the last inequality holds due to $\sum_{t=1}^{T}\|\phi(p_t,x_t)\|^2_{V_{t,j}^{-1}}\mathbb{I}_{p_t\in\mathbf{I}_j}\le 2\log\det(V_{T,j})/\log(\lambda I)$ from the elliptical potential lemma (see, e.g., Lemma 11 in Abbasi-Yadkori et al. (2011)) and $\log\det(V_{T,j})\le(\mathfrak{b}(k)+d+1)\log(\lambda(1+\frac{n_{T,j}}{\mathfrak{b}(k)+d+1}))$.

Since $\epsilon=T^{-2}$, we then have

$$\begin{aligned}\gamma_{T,j}&=\sigma\sqrt{(\mathfrak{b}(k)+d+1)\log\left(\frac{\mathfrak{b}(k)+d+1+n_{T,j}}{\mathfrak{b}(k)+d+1}\right)-2\log\epsilon}+\lambda^{\frac12}(C_0^2(\mathfrak{b}(k)+1)+\overline{a}^2)^{\frac12}+\Delta\sqrt{n_{T,j}}\\&\le\sigma\sqrt{2(\mathfrak{b}(k)+d+1)\log(T+1)}+\lambda^{\frac12}(C_0^2(\mathfrak{b}(k)+1)+\overline{a}^2)^{\frac12}+\Delta\sqrt{n_{T,j}},\end{aligned}$$

and the first term in the RHS of (EC.48) is bounded by

$$\begin{aligned}&\sum_{j=1}^{N}\gamma_{T,j}\sqrt{n_{T,j}}\sqrt{2(\mathfrak{b}(k)+d+1)\log\left(1+\frac{n_{T,j}}{\mathfrak{b}(k)+d+1}\right)}\\&\le 2\sqrt{(\mathfrak{b}(k)+d+1)\log(T+1)}\left(\max\left\{\sigma\sqrt{(\mathfrak{b}(k)+d+1)\log(T+1)},\lambda^{\frac12}(C_0^2k+\overline{a}^2)^{\frac12}\right\}\cdot\sum_{j=1}^{N}\sqrt{n_{T,j}}+\Delta T\right)\\&\le 2\sqrt{(\mathfrak{b}(k)+d+1)\log(T+1)}\\&\qquad\cdot\left(\max\left\{\sigma\sqrt{(\mathfrak{b}(k)+d+1)\log(T+1)},\lambda^{\frac12}(C_0^2(\mathfrak{b}(k)+1)+\overline{a}^2)^{\frac12}\right\}\cdot\sqrt{\sum_{j=1}^{N}n_{T,j}}\cdot\sqrt{\sum_{j=1}^{N}1^2}+\Delta T\right)\\&=2\sqrt{(\mathfrak{b}(k)+d+1)\log(T+1)}\left(\max\left\{\sigma\sqrt{(\mathfrak{b}(k)+d+1)\log(T+1)},\lambda^{\frac12}(C_0^2(\mathfrak{b}(k)+1)+\overline{a}^2)^{\frac12}\right\}\cdot\sqrt{TN}+\Delta T\right).\end{aligned}$$

This, together with $\Delta=\Theta(\frac{\delta}{N^k})$ and (EC.48), implies

$$\sum_{t=1}^{T}r_t=\mathcal{O}\left(\sqrt{(\mathfrak{b}(k)+d+1)\log T}\right)\cdot\mathcal{O}\left(\sqrt{(\mathfrak{b}(k)+d+1)NT\log T}+\frac{\delta T}{N^k}\right).$$

To balance the two terms $\sqrt{(\mathfrak{b}(k)+d+1)NT\log T}$ and $\frac{\delta T}{N^k}$, we let $N=\lceil(T\delta^2)^{\frac{1}{2k+1}}\rceil+1$. When $\delta=\mathcal{O}(T^{-\frac12})$, $N=\Theta(1)$ and thus we obtain $\sqrt{(\mathfrak{b}(k)+d+1)NT}=\Theta(\sqrt{(\mathfrak{b}(k)+d+1)T})$ and $\frac{\delta T}{N^k}=\mathcal{O}(\sqrt{T})$. In this case, we get $\sum_{t=1}^{T}r_t=\mathcal{O}((\mathfrak{b}(k)+d+1)\sqrt{T}\log T)$. When $\delta=\Omega(T^{-\frac12})$, $N=\Theta((T\delta^2)^{\frac{1}{2k+1}})$ and we obtain $\sum_{t=1}^{T}r_t=\mathcal{O}((\mathfrak{b}(k)+d+1)\delta^{\frac{1}{2k+1}}T^{\frac{k+1}{2k+1}}\log T)$. Combining these two cases, we have $\sum_{t=1}^{T}r_t=\mathcal{O}((\mathfrak{b}(k)+d+1)((\delta T^{k+1})^{\frac{1}{2k+1}}\vee\sqrt{T})\log T)$.

Therefore, the total expected regret is upper bounded by

$$\begin{aligned}\sum_{t=1}^{T}\mathbb{E}[r_t]&=\sum_{t=1}^{T}\mathbb{E}[r_t|\mathcal{A}]\cdot\mathbb{P}(\mathcal{A})+\sum_{t=1}^{T}\mathbb{E}[r_t|\mathcal{A}^c]\cdot\mathbb{P}(\mathcal{A}^c)\\&=\widetilde{\mathcal{O}}\left((\mathfrak{b}(k)+d+1)\left((\delta T^{k+1})^{\frac{1}{2k+1}}\vee\sqrt{T}\right)\right)+\mathcal{O}\left(\frac{N}{T}\right),\end{aligned}\tag{EC.49}$$

where the second identity holds since from the union bound, $\mathbb{P}(\mathcal{A}^c)\le\frac{N}{\epsilon}=\frac{N}{T^2}$. Since $\mathcal{O}(\frac{N}{T})=\mathcal{O}(T^{-\frac{2k}{2k+1}}\delta^{\frac{2}{2k+1}})$ and $\delta=\mathcal{O}(1)$, the first term in the RHS of (EC.49) dominates. Q.E.D.

B.2. Proof for Theorem 4

As discussed in Sec. 3.2, it suffices to prove the lower bound $\Omega((\delta T^{k_f+1})^{\frac{1}{2k_f+1}})$. Note that we also only need to consider the case $\delta \geq T^{-\frac{1}{2}}$, since otherwise, $(\delta T^{k_f+1})^{\frac{1}{2k_f+1}} < \sqrt{T}$, and the desired lower bound in (11) becomes $\Omega(\sqrt{T})$, which is again obtained. Recall from Lemma 2 that $\hat{\delta} = \delta/((\sum_{i=0}^{\mathfrak{b}(k_f)} \frac{\mathfrak{b}(k_f)!}{i!}) \vee (\mathfrak{b}(k_f)+1)2^{\mathfrak{b}(k_f)-1})$, $J = \lceil 4(\mathfrak{b}(k_f)+1)2^{\mathfrak{b}(k_f)} \hat{\delta}^{\frac{2}{2k_f+1}} T^{\frac{1}{2k_f+1}} \rceil$ and $\eta = ((2\sigma) \wedge \frac{1}{2^{3k_f+1}}) \frac{1}{((\mathfrak{b}(k_f)+1)2^{\mathfrak{b}(k_f)})^{k_f}} \hat{\delta}^{\frac{1}{2k_f+1}} T^{-\frac{k_f}{2k_f+1}}$.

For each $j = 0, 1, 2 \ldots, J$ and $p \in [\underline{p}, \overline{p}]$, we define $r_j(p) = p \times f_j(p)$. Note that for each $j \in [J]$, the induced optimal price of $r_j(p)$ belongs to $\mathbf{I}_j$ and $r_j(p)$ differs from $r_0(p)$ only in $\mathbf{I}_j$, with the maximum difference characterized by parameter η. For any policy π, consider the random variable T_j denoting the number of times the prices selected by π fall into segment $\mathbf{I}_j$. We first claim the following inequality for any $1 \leq j \leq J$:

$$\left|\mathbb{E}_0^\pi[T_j] - \mathbb{E}_j^\pi[T_j]\right| \leq \frac{1}{4\sigma}\sqrt{\mathbb{E}_0^\pi[T_j]}T\eta, \tag{EC.50}$$

where $\mathbb{E}_0^\pi[\cdot]$ and $\mathbb{E}_j^\pi[\cdot]$ denote the expectation associated with the probability measure induced by policy π under demand model r_0 and r_j respectively. The proof of (EC.50) is deferred to the last part. Consider the index $j^* \in [J]$ that minimizes $\mathbb{E}_0^\pi[T_j]$ over $j \in [J]$. By Pigeonhole principle, we have $\mathbb{E}_0^\pi[T_{j^*}] \leq T/J$. From (EC.50), we further have

$$\mathbb{E}_{j^*}^\pi[T_{j^*}] \leq \frac{1}{4\sigma}\sqrt{\mathbb{E}_0^\pi[T_{j^*}]}T\eta + \mathbb{E}_0^\pi[T_{j^*}] \leq \frac{1}{4\sigma}\sqrt{\frac{T}{J}}T\eta + \frac{T}{J} \leq \frac{T}{2}, \tag{EC.51}$$

where the first inequality follows from (EC.50), the second inequality follows from the choice of j, and the last inequality holds since $\eta^2 \leq (2\sigma\hat{\delta}^{\frac{1}{2k_f+1}} T^{-\frac{k_f}{2k_f+1}})^2 = 4\sigma^2\hat{\delta}^{\frac{2}{2k_f+1}} T^{-\frac{2k_f}{2k_f+1}} \leq \sigma^2\frac{J}{T}$ implies $\frac{1}{4\sigma}\sqrt{\frac{T}{J}}T\eta \leq \frac{1}{4}T$ and $\hat{\delta} \geq T^{-\frac{1}{2}}$ implies $\frac{T}{J} = \frac{1}{4}\hat{\delta}^{-\frac{2}{2k_f+1}} T^{\frac{2k_f}{2k_f+1}} \leq \frac{1}{4}T$. Note that when the true demand function is $f_{j^*}(\cdot)$, in any period when policy π charges a price out of $\mathbf{I}_j$, a revenue loss η will be incurred by the definition of $r_j(\cdot)$. Hence, we have

$$\sup_{f \in \{f_1, f_2, \ldots, f_J\}} R_{f_j}^\pi(T, k_f, \delta) \geq R_{f_{j^*}}^\pi(T, k_f, \delta) \geq (T - \mathbb{E}_{j^*}^\pi[T_{j^*}])\eta \geq \frac{1}{2}T\eta = \Omega\left(\left(\delta T^{k_f+1}\right)^{\frac{1}{2k_f+1}}\right).$$

Finally, we complete the proof of Theorem 4 by proving (EC.50). For the sake of rigor, we define a probability space as follows. Let $\Omega = ([1,2] \times \mathbb{R})^T \times \{0, 1, 2, \ldots, T\}$ and $\mathcal{B}(\Omega)$ be the Borel algebra on Ω. For any $t \in [T]$, let P_t and D_t be measurable functions on $(\Omega, \mathcal{B}(\Omega))$ that map each $\omega = (p_1, d_1, p_2, d_2, \ldots, p_T, d_T) \in \Omega$ to p_t and d_t respectively. For any $j \in \{0\} \cup [J]$, let T_j be a measurable function on $(\Omega, \mathcal{B}(\Omega))$ that maps $\omega = (p_1, d_1, p_2, d_2, \ldots, p_T, d_T) \in \Omega$ to the cardinality of the set $\{1 \leq t \leq T : p_t \in \mathbf{I}_j\}$. We also define two functions $\mu_i^\pi : ([1,2] \times \mathbb{R})^T \to \mathbb{R}^+$ and $\nu_i^\pi : ([1,2] \times \mathbb{R})^T \times \{0, 1, 2, \ldots, T\} \to \mathbb{R}^+$ as follows:

$$\nu_j^\pi(p_1, d_1, p_2, d_2, \ldots, p_T, d_T) = \prod_{t=1}^T \left(\mu^\pi(p_t | p_1, d_1, \ldots, p_{t-1}, d_{t-1}) \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(d_t - f_j(p_t))^2}{2\sigma^2}}\right),$$
$$\mu_j^\pi(p_1, d_1, p_2, d_2, \ldots, p_T, d_T, t_j) = \nu_j^\pi(p_1, d_1, p_2, d_2, \ldots, p_T, d_T) \cdot \mathbf{1}_{\{t_j = |1 \leq t \leq T : p_t \in \mathbf{I}_j|\}},$$

where $\mu^\pi(p_t|p_1,d_1,\ldots,p_{t-1},d_{t-1})$ is the p.d.f. for p_t given $(p_1,d_1,\ldots,p_{t-1},d_{t-1})$. Let $\mathbb{P}_j^\pi(\cdot)$ be the following probability measure on $(\Omega,\mathcal{B}(\Omega))$: for any $B\in\mathcal{B}(\Omega)$, $\mathbb{P}_j^\pi(B)=\int_B \mu_j^\pi(w)dw$. Thus, $(\Omega,\mathcal{B}(\Omega),\mathbb{P}_j^\pi)$ constitute a probability space, and from the chain rule, $\nu_j^\pi(\cdot)$ and $\mu_j^\pi(\cdot)$ are the p.d.f. for $(P_1,D_1,P_2,D_2,\ldots,P_T,D_T)$ and $(P_1,D_1,P_2,D_2,\ldots,P_T,D_T,T_j)$ respectively. With a slight abuse of notation, we denote the distributions of T_j and $(P_1,D_1,P_2,D_2,\ldots,P_T,D_T)$ by $\mathbb{P}_i^\pi(T_j)$ and $\mathbb{P}_i^\pi(P_1,D_1,P_2,D_2,\ldots,P_T,D_T)$ respectively, and the conditional probability distribution of T_j given $(P_1,D_1,P_2,D_2,\ldots,P_T,D_T)$ by $\mathbb{P}_i^\pi(T_j|P_1,D_1,P_2,D_2,\ldots,P_T,D_T)$. For any given $0\le i\le J$ and $0\le j\le J$, $\mathbb{E}_i^\pi[T_j]$ is then the expectation of T_j under $\mathbb{P}_i^\pi$. Then we note that

$$\begin{aligned}\left|\mathbb{E}_0^\pi[T_j]-\mathbb{E}_j^\pi[T_j]\right| &\le \sum_{t=0}^{T} t\times\left|\mathbb{P}_0^\pi(t)-\mathbb{P}_j^\pi(t)\right| \le T\times\sum_{t=0}^{T}\left|\mathbb{P}_0^\pi(t)-\mathbb{P}_j^\pi(t)\right| \\ &= \frac{1}{2}T||\mathbb{P}_0^\pi(T_j)-\mathbb{P}_j^\pi(T_j)||_{\mathrm{TV}} \le \frac{1}{2}T\sqrt{\frac{1}{2}\mathrm{KL}(\mathbb{P}_0^\pi(T_j)||\mathbb{P}_j^\pi(T_j))},\end{aligned} \tag{EC.52}$$

where the first identity follows from the property of the total variation distance for discrete random variables, see, e.g., Proposition 4.2 in Levin and Peres (2017), and the last inequality follows from Pinsker's inequality.

To further bound $\mathrm{KL}(\mathbb{P}_0^\pi(T_j)||\mathbb{P}_j^\pi(T_j))$, we note that

$$\begin{aligned}\mathrm{KL}\Big(\mathbb{P}_0^\pi(T_j)\,||\,\mathbb{P}_j^\pi(T_j)\Big) &= \mathrm{KL}\Big(\mathbb{P}_0^\pi(P_1,D_1,P_2,D_2,\ldots,P_T,D_T,T_j)\,||\,\mathbb{P}_j^\pi(P_1,D_1,P_2,D_2,\ldots,P_T,D_T,T_j)\Big) \\ &\quad-\mathrm{KL}\Big(\mathbb{P}_0^\pi(P_1,D_1,P_2,D_2,\ldots,P_T,D_T|T_j)\,||\,\mathbb{P}_j^\pi(P_1,D_1,P_2,D_2,\ldots,p_T,D_T|T_j)\Big) \\ &\le \mathrm{KL}\Big(\mathbb{P}_0^\pi(P_1,D_1,P_2,D_2,\ldots,P_T,D_T,T_j)\,||\,\mathbb{P}_j^\pi(P_1,D_1,P_2,D_2,\ldots,P_T,D_T,T_j)\Big) \\ &= \mathbb{E}_0^\pi\left[\mathbb{E}_0^\pi\left[\log\frac{\mu_0^\pi(P_1,D_1,P_2,D_2,\ldots,P_T,D_T,T_j)}{\mu_j^\pi(P_1,D_1,P_2,D_2,\ldots,P_T,D_T,T_j)}\Big|(P_1,D_1,\ldots,P_T,D_T)\right]\right] \\ &= \mathbb{E}_0^\pi\left[\log\frac{\nu_0^\pi(P_1,D_1,P_2,D_2,\ldots,P_T,D_T)}{\nu_j^\pi(P_1,D_1,P_2,D_2,\ldots,P_T,D_T)}\right] \\ &= \frac{1}{2\sigma^2}\sum_{t=1}^{T}\mathbb{E}_0^\pi\left[(D_t-f_j(P_t))^2-(D_t-f_0(P_t))^2\right] \\ &= \frac{1}{2\sigma^2}\sum_{t=1}^{T}\mathbb{E}_0^\pi\left[(f_0(P_t)-f_j(P_t))^2\right] \\ &= \frac{1}{2\sigma^2}\sum_{t=1}^{T}\mathbb{E}_0^\pi\left[\mathbf{1}_{\{P_t\in\mathbf{I}_j\}}(f_0(P_t)-f_j(P_t))^2\right] \\ &\le \frac{1}{2\sigma^2}\mathbb{E}_0^\pi[T_j]\cdot\max_{p\in\mathbf{I}_j}(f_0(p)-f_j(p))^2 \\ &= \frac{1}{2\sigma^2}\mathbb{E}_0^\pi[T_j]\eta^2,\end{aligned} \tag{EC.53}$$

where the first identity follows from the chain rule for KL divergence, the first inequality follows from the fact that the KL divergence between any two probability distributions is non-negative, the

second identity holds due to the definition of KL divergence and the law of total expectation, the third identity holds since given $(P_1, D_1, P_2, D_2, \ldots, P_T, D_T)$, T_j takes the value $|\{1 \le t \le T : P_t \in \mathbf{I}_j\}|$ with probability one, and when $T_i = |\{1 \le t \le T : P_t \in \mathbf{I}_j\}|$, $\mu_i^\pi(P_1, D_2, P_2, D_2, \ldots, P_T, D_T, T_j) = \nu_i^\pi(P_1, D_2, P_2, D_2, \ldots, P_T, D_T)$, the fifth identity holds since $D_t = f_0(P_t) + \varepsilon_t$, $\mathbb{E}_0^\pi[\varepsilon_t] = 0$ and P_t is independent of ε_t, and the sixth identity holds since f_0 and f_j are only different in $\mathbf{I}_j$. Then (EC.50) is obtained by combining (EC.52) with (EC.53). Q.E.D.

B.3. Proof for Lemma 2

Recall that $\hat{\delta} = \delta/((\sum_{i=0}^{\mathfrak{b}(k_f)} \frac{\mathfrak{b}(k_f)!}{i!}) \vee (\mathfrak{b}(k_f)+1)2^{\mathfrak{b}(k_f)-1})$, $J = \lceil 4(\mathfrak{b}(k_f)+1)2^{\mathfrak{b}(k_f)} \hat{\delta}^{\frac{2}{2k_f+1}} T^{\frac{1}{2k_f+1}} \rceil$ and $\eta = ((2\sigma) \wedge \frac{1}{2^{3k_f+1}}) \frac{1}{((\mathfrak{b}(k_f)+1)2^{\mathfrak{b}(k_f)})^{k_f}} \hat{\delta}^{\frac{1}{2k_f+1}} T^{-\frac{k_f}{2k_f+1}}$. The result for $j = 0$ is trivial. When $1 \le j \le J$, from the properties (1) and (2) of $S(x)$, $g_{k_f}(x)$ is infinitely differentiable. Now we check the property of $f_j^{(\mathfrak{b}(k_f))}(p)$.

$$
\begin{aligned}
\left| f_j^{\mathfrak{b}(k_f)}(p_1) - f_j^{\mathfrak{b}(k_f)}(p_2) \right| &= \left| \left(\frac{r_j(p_1)}{p_1} \right)^{(\mathfrak{b}(k_f))} - \left(\frac{r_j(p_2)}{p_2} \right)^{(\mathfrak{b}(k_f))} \right| \\
&= \left| \sum_{i=0}^{\mathfrak{b}(k_f)} \binom{\mathfrak{b}(k_f)}{i} \left(\frac{1}{p_1} \right)^{(\mathfrak{b}(k_f)-i)} r_j^{(i)}(p_1) - \sum_{i=0}^{\mathfrak{b}(k_f)} \binom{\mathfrak{b}(k_f)}{i} \left(\frac{1}{p_2} \right)^{(\mathfrak{b}(k_f)-i)} r_j^{(i)}(p_2) \right| \\
&= \left| \sum_{i=0}^{\mathfrak{b}(k_f)} \binom{\mathfrak{b}(k_f)}{i} (-1)^{(\mathfrak{b}(k_f)-i)} (\mathfrak{b}(k_f)-i)! \left(\frac{r_j^{(i)}(p_1)}{p_1^{\mathfrak{b}(k_f)-i+1}} - \frac{r_j^{(i)}(p_2)}{p_2^{\mathfrak{b}(k_f)-i+1}} \right) \right| \\
&\le \left(\sum_{i=0}^{\mathfrak{b}(k_f)} \frac{\mathfrak{b}(k_f)!}{i!} \right) \max_{0 \le i \le \mathfrak{b}(k_f)} \left| \frac{r_j^{(i)}(p_1)}{p_1^{\mathfrak{b}(k_f)-i+1}} - \frac{r_j^{(i)}(p_2)}{p_2^{\mathfrak{b}(k_f)-i+1}} \right|, \qquad \text{(EC.54)}
\end{aligned}
$$

where the second identity follows the general Leibniz rule. Now we turn to the RHS of Eq. (EC.54). In the following, we discuss in three cases: (1) $p_1, p_2 \in \mathbf{I}_j$, (2) $p_1 \in \mathbf{I}_j$ and $p_2 \notin \mathbf{I}_j$, and (3) $p_1, p_2 \notin \mathbf{I}_j$.

Case 1: $p_1 \in \mathbf{I}_j$ and $p_2 \in \mathbf{I}_j$.

If $0 \le i \le \mathfrak{b}(k_f) - 1$, we have

$$
\begin{aligned}
\left| \frac{r_j^{(i)}(p_1)}{p_1^{\mathfrak{b}(k_f)-i+1}} - \frac{r_j^{(i)}(p_2)}{p_2^{\mathfrak{b}(k_f)-i+1}} \right| &\le \left| \frac{r_j^{(i)}(p_1)}{p_1^{\mathfrak{b}(k_f)-i+1}} - \frac{r_j^{(i)}(p_2)}{p_1^{\mathfrak{b}(k_f)-i+1}} \right| + \left| \frac{r_j^{(i)}(p_2)}{p_1^{\mathfrak{b}(k_f)-i+1}} - \frac{r_j^{(i)}(p_2)}{p_2^{\mathfrak{b}(k_f)-i+1}} \right| \\
&\le \left| r_j^{(i)}(p_1) - r_j^{(i)}(p_2) \right| + \max_{p \in [1,2]} \left| r_j^{(i)}(p) \right| \left| \frac{p_1^{\mathfrak{b}(k_f)-i+1} - p_2^{\mathfrak{b}(k_f)-i+1}}{p_1^{\mathfrak{b}(k_f)-i+1} p_2^{\mathfrak{b}(k_f)-i+1}} \right| \\
&\le \max_{p \in [1,2]} \left| r_j^{(i+1)}(p) \right| |p_1 - p_2| + \max_{p \in [1,2]} \left| r_j^{(i)}(p) \right| |p_1 - p_2| \left| \sum_{q=0}^{\mathfrak{b}(k_f)-i} p_1^q p_2^{\mathfrak{b}(k_f)-i-q} \right| \\
&\le \eta (2J)^{i+1} \max_{x \in [0,2]} \left| g_{k_f}^{(i+1)}(x) \right| |p_1 - p_2| \\
&\qquad + \eta (2J)^i \max_{x \in [0,2]} \left| g_{k_f}^{(i)}(x) \right| |p_1 - p_2| \, (\mathfrak{b}(k_f) - i + 1) 2^{\mathfrak{b}(k_f)-i} \\
&\le \left(\eta (2J)^{i+1} + (\mathfrak{b}(k_f) - i + 1) \eta (2J)^i 2^{\mathfrak{b}(k_f)-i} \right) |p_1 - p_2|
\end{aligned}
$$

$$\leq 2\eta(2J)^{i+1}|p_1-p_2|, \tag{EC.55}$$

where the third inequality follows from mean value theorem, the existence of the $(i+1)$-th derivatives, $p_1 \geq 1$, $p_2 \geq 1$ and the fact that $p_1^{\mathfrak{b}(k_f)-i} - p_2^{\mathfrak{b}(k_f)-i+1} = (p_1-p_2)(\sum_{q=0}^{\mathfrak{b}(k_f)-i+1} p_1^q p_2^{\mathfrak{b}(k_f)-i-q})$, the fourth inequality holds due to $p_1 \leq 2$ and $p_2 \leq 2$, the fifth inequality holds by our construction that $\max_{x\in[0,2]} \left|g_{k_f}^{(i)}(x)\right| \leq 1$, the last inequality follows from $(\mathfrak{b}(k_f)+1)2^{\mathfrak{b}(k_f)} \leq 2J$. Then, for $i = \mathfrak{b}(k_f)$,

$$\begin{aligned}
\left|\frac{r_j^{(\mathfrak{b}(k_f))}(p_1)}{p_1} - \frac{r_j^{(\mathfrak{b}(k_f))}(p_2)}{p_2}\right| &\leq \left|\frac{r_j^{(\mathfrak{b}(k_f))}(p_1)}{p_1} - \frac{r_j^{(\mathfrak{b}(k_f))}(p_2)}{p_1}\right| + \left|\frac{r_j^{(\mathfrak{b}(k_f))}(p_2)}{p_1} - \frac{r_j^{(\mathfrak{b}(k_f))}(p_2)}{p_2}\right| \\
&\leq \eta(2J)^{\mathfrak{b}(k_f)} \left|g^{(\mathfrak{b}(k_f))}(2J(p_1-a_j)) - g^{(\mathfrak{b}(k_f))}(2J(p_2-a_j))\right| \\
&\qquad + |p_1-p_2| \max_{p\in[1,2]} \left|r_j^{\mathfrak{b}(k_f)}(p)\right| \\
&\leq \eta(2J)^{\mathfrak{b}(k_f)} |2J(p_1-p_2))|^{k_f-\mathfrak{b}(k_f)} + \eta(2J)^{\mathfrak{b}(k_f)} |p_1-p_2| \\
&\leq (\eta(2J)^{k_f} + \eta(2J)^{\mathfrak{b}(k_f)}) |p_1-p_2|^{k_f-\mathfrak{b}(k_f)} \\
&\leq 2\eta(2J)^{k_f} |p_1-p_2|^{k_f-\mathfrak{b}(k_f)},
\end{aligned} \tag{EC.56}$$

where the second inequality is because $\left|\frac{r_j^{(\mathfrak{b}(k_f))}(p_2)}{p_1} - \frac{r_j^{(\mathfrak{b}(k_f))}(p_2)}{p_2}\right| \leq \left|\frac{1}{p_1} - \frac{1}{p_2}\right| \max_{p\in[1,2]} \left|r_j^{(\mathfrak{b}(k_f))}(p)\right| \leq |p_1-p_2| \max_{p\in[1,2]} \left|r_j^{(\mathfrak{b}(k_f))}(p)\right|$, the third inequality follows that $g^{(\mathfrak{b}(k_f))}(\cdot)$ is $(k_f - \mathfrak{b}(k_f))$-Hölder continuous, the fourth inequality holds because of $|p_1-p_2| \leq 1$ and $k_f - \mathfrak{b}(k_f) \leq 1$. Then, Eq. (EC.54) can be simplified

$$\begin{aligned}
\left|f_j^{\mathfrak{b}(k_f)}(p_1) - f_j^{\mathfrak{b}(k_f)}(p_2)\right| &\leq \left(\sum_{i=0}^{\mathfrak{b}(k_f)} \frac{\mathfrak{b}(k_f)!}{i!}\right) \max\left\{\max_{0\leq i\leq \mathfrak{b}(k_f)-1} 2\eta(2J)^{i+1}|p_1-p_2|, 2\eta(2J)^{k_f}|p_1-p_2|^{k_f-\mathfrak{b}(k_f)}\right\} \\
&\leq \left(\sum_{i=0}^{\mathfrak{b}(k_f)} \frac{\mathfrak{b}(k_f)!}{i!}\right) \max\left\{2\eta(2J)^{\mathfrak{b}(k_f)}|p_1-p_2|, 2\eta(2J)^{k_f}|p_1-p_2|^{k_f-\mathfrak{b}(k_f)}\right\} \\
&= \left(\sum_{i=0}^{\mathfrak{b}(k_f)} \frac{\mathfrak{b}(k_f)!}{i!}\right) \left(2\eta(2J)^{k_f}|p_1-p_2|^{k_f-\mathfrak{b}(k_f)}\right) \\
&\leq \delta |p_1-p_2|^{k_f-\mathfrak{b}(k_f)},
\end{aligned} \tag{EC.57}$$

where the last inequality holds due to $2\eta(2J)^{k_f} \leq 2\frac{1}{2^{3k_f+1}} \frac{1}{((\mathfrak{b}(k_f)+1)2^{\mathfrak{b}(k_f)})^{k_f}} \hat{\delta}^{\frac{1}{2k_f+1}} T^{-\frac{k_f}{2k_f+1}} (8(\mathfrak{b}(k_f)+1)2^{\mathfrak{b}(k_f)} \hat{\delta}^{\frac{2}{2k_f+1}} T^{\frac{1}{2k_f+1}})^{k_f} = \hat{\delta}$.

Case 2: $p_1 \in \mathbf{I}_j$ and $p_2 \notin \mathbf{I}_j$.

Note that Eq. (EC.55) still hold in this case. What we need to derive is the bound for $|\frac{r_j^{(\mathfrak{b}(k_f))}(p_1)}{p_1} - \frac{r_j^{(\mathfrak{b}(k_f))}(p_2)}{p_2}|$ (i.e., Eq. (EC.56) can not be directly applied here). Define $p_2' := \text{Proj}(p_2, \mathbf{I}_j)$. Note that

p_2' is either a_j or b_j, $r_j(a_j)=r_j(b_j)=\frac{1}{2}\hat{\delta}$ and $r_j^{(i)}(a_j)=r_j^{(i)}(b_j)=0$ for all $1\le i\le \mathfrak{b}(k_f)$. Thus, we can have

$$\left|\frac{r_j^{(\mathfrak{b}(k_f))}(p_1)}{p_1}-\frac{r_j^{(\mathfrak{b}(k_f))}(p_2)}{p_2}\right|=\left|\frac{r_j^{(\mathfrak{b}(k_f))}(p_1)}{p_1}-\frac{r_j^{(\mathfrak{b}(k_f))}(p_2')}{p_2'}\right|\le 2\eta(2J)^{k_f}|p_1-p_2'|^{k_f-\mathfrak{b}(k_f)}$$
$$\le 2\eta(2J)^{k_f}|p_1-p_2|^{k_f-\mathfrak{b}(k_f)}, \qquad \text{(EC.58)}$$

where the first equality holds due to $r_j^{(\mathfrak{b}(k_f))}(p_2)=r_j^{(\mathfrak{b}(k_f))}(p_2')=0$, the first inequality follows from Eq. (EC.56) because p_1 and p_2' are both in $\mathbf{I}_j$, and the second inequality holds due to the projection process. Together with Eqs. (EC.54), (EC.55) and (EC.58), by the same calculation of Eq. (EC.57), we can know $|f_j^{\mathfrak{b}(k_f)}(p_1)-f_j^{\mathfrak{b}(k_f)}(p_2)|\le\delta|p_1-p_2|^{k_f-\mathfrak{b}(k_f)}$.

Case 3: $p_1\notin\mathbf{I}_j$ and $p_2\notin\mathbf{I}_j$.

When p_1 and p_2 are not in $\mathbf{I}_j$, $r_j^{(i)}(p_1)=r_j^{(i)}(p_2)=0$, for all $1\le i\le\mathfrak{b}(k_f)$. From the first two lines of Eq. (EC.54), we can have

$$\begin{aligned}\left|f_j^{\mathfrak{b}(k_f)}(p_1)-f_j^{\mathfrak{b}(k_f)}(p_2)\right|&=\left|\frac{r_j(p_1)}{p_1^{\mathfrak{b}(k_f)+1}}-\frac{r_j(p_2)}{p_2^{\mathfrak{b}(k_f)+1}}\right|\\&\le\frac{1}{2}\hat{\delta}\left|p_1^{\mathfrak{b}(k_f)+1}-p_2^{\mathfrak{b}(k_f)+1}\right|\\&\le\frac{1}{2}\hat{\delta}|p_1-p_2|\left|\sum_{q=0}^{\mathfrak{b}(k_f)}p_1^q p_2^{\mathfrak{b}(k_f)-q}\right|\\&\le\frac{1}{2}\hat{\delta}(\mathfrak{b}(k_f)+1)2^{\mathfrak{b}(k_f)}|p_1-p_2|\\&\le\delta|p_1-p_2|^{k_f-\mathfrak{b}(k_f)},\end{aligned}$$

where the third inequality holds due to $p_1\le 2$ and $p_2\le 2$, and the last inequality follows from $\hat{\delta}\le\delta/((\mathfrak{b}(k_f)+1)2^{\mathfrak{b}(k_f)}-1)$.

Together with the above three cases, we draw the conclusion that for any $p_1,p_2\in[1,2]$, we have $|f_j^{\mathfrak{b}(k_f)}(p_1)-f_j^{\mathfrak{b}(k_f)}(p_2)|\le\delta|p_1-p_2|^{k_f-\mathfrak{b}(k_f)}$. We finish the proof. Q.E.D.

Appendix C. Proofs for Statements in Section 4

C.1. Proof for Theorem 5

For notation convenience, in this proof, we denote $p_t^*=\arg\max_{p\in[\underline{p},\overline{p}]}p(f(p)+g(x_t))$, $i_t=\arg\min_{0\le i\le N-1}|p_t^*-P[i]|$, $i_t^*=\max_{0\le i\le N-1}P[i]\times(f(P[i])+g(x_t))$ and $\hat{i}_t^*=\max_{0\le i\le N-1}P[i]\times(f(P[i])+(\theta_{j_t}^*)^\top\phi(x_t))$. Algorithm 4 operates in an exploration-then-exploitation manner. In the exploration phase, since $g(\cdot)$ and $f(\cdot)$ are bounded, the total regret is $\mathcal{O}(n_0N)$. We now turn to bound the regret in the exploitation phase. For each $t\ge[T]$, let j_t be the index of the bin that x_t falls into, which itself is a random variable. The regret can be decomposed as follows:

$$\sum_{t=n_0N}^{T}r_t=\sum_{t=n_0N}^{T}\max_{p\in[\underline{p},\overline{p}]}p(f(p)+g(x_t))-p_t(f(p_t)+g(x_t))$$

$$
\begin{aligned}
&= \sum_{t=n_0 N}^{T} \underbrace{\max_{p\in[\underline{p},\overline{p}]} p(f(p)+g(x_t)) - \max_{0\le i\le N-1} P[i]\times(f(P[i])+(\theta^*_{j_t})^\top\phi(x_t))}_{\text{regret from discretization error of price and context spaces}} \\
&+ \underbrace{\max_{0\le i\le N-1} P[i]\times(f(P[i])+(\theta^*_{j_t})^\top\phi(x_t)) - p_t(f(p_t)+(\theta^*_{j_t})^\top\phi(x_t))}_{\text{regret from failure of identifying the best discretized price}} \\
&+ \underbrace{p_t(f(p_t)+(\theta^*_{j_t})^\top\phi(x_t)) - p_t(f(p_t)+g(x_t))}_{\text{regret from discretization error of context space}}. \qquad \text{(EC.59)}
\end{aligned}
$$

The first term of Eq. (EC.59) comes from the discretization error of the price and context spaces. With Assumption 3, we establish the following inequalities:

$$
\begin{aligned}
&\max_{p\in[\underline{p},\overline{p}]} p(f(p)+g(x_t)) - \max_{0\le i\le N-1} P[i]\times(f(P[i])+(\theta^*_{j_t})^\top\phi(x_t)) \\
&= \max_{p\in[\underline{p},\overline{p}]} p(f(p)+g(x_t)) - \max_{0\le i\le N-1} P[i]\times(f(P[i])+g(x_t)) \\
&\quad + \max_{i\in[N]} P[i]\times(f(P[i])+g(x_t)) - \max_{0\le i\le N-1} P[i]\times(f(P[i])+(\theta^*_{j_t})^\top\phi(x_t)) \\
&\le p^*_t(f(p^*_t)+g(x_t)) - P[i_t]\times(f(P[i_t])+g(x_t)) \\
&\quad + P[i^*_t]\times(f(P[i^*_t])+g(x_t)) - P[i^*_t]\times(f(P[i^*_t])+(\theta^*_{j_t})^\top\phi(x_t)) \\
&= p^*_t f(p^*_t) - P[i_t]f(P[i_t]) + (p^*_t - P[i_t])g(x_t) + P[i^*_t](g(x_t) - (\theta^*_{j_t})^\top\phi(x_t)) \\
&\le p^*_t f(p^*_t) - P[i_t]f(P[i_t]) + \frac{\overline{p}-\underline{p}}{2N}\max_{x\in[0,1]^d}|g(x)| + \frac{\overline{p}Ld^{\frac{k_g+\mathfrak{b}(k_g)}{2}}}{M^{k_g}\mathfrak{b}(k_g)!}. \qquad \text{(EC.60)}
\end{aligned}
$$

In the first inequality of (EC.60), we apply the following inequalities:

$$
\begin{aligned}
&\max_{0\le i\le N-1} P[i]\times(f(P[i])+g(x_t)) \mid j_t \ge P[i_t]\times(f(P[i_t])+g(x_t)), \\
&\max_{0\le i\le N-1} P[i]\times(f(P[i])+(\theta^*_{j_t})^\top\phi(x_t)) \ge P[i^*_t]\times(f(P[i^*_t])+(\theta^*_{j_t})^\top\phi(x_t)).
\end{aligned}
$$

In the second inequality of (EC.60), we use the fact that p^*_t and $P[i_t]$ both belong to the i_t-th price segment from the definition of i_t, and since the length of each price segment is no more than $(\overline{p}-\underline{p})/N$ and $P[i_t]$ is the midpoint of the i_t-th price segment, the distance between p^*_t and $P[i_t]$ is no more than $(\overline{p}-\underline{p})/(2N)$.

The first term in the RHS of (EC.60) can be further upper bounded as follows:

$$
\begin{aligned}
p^*_t f(p^*_t) - P[i_t]f(P[i_t]) &= p^*_t f(p^*_t) - P[i_t]f(p^*_t) + P[i_t]f(p^*_t) - P[i_t]f(P[i_t]) \\
&\le \frac{\overline{p}-\underline{p}}{2N}\max_{p\in[\underline{p},\overline{p}]}|f(p)| + \overline{p}|f(p^*_t) - f(P[i_t])| \\
&\le \frac{\overline{p}-\underline{p}}{2N}\max_{p\in[\underline{p},\overline{p}]}|f(p)| + \overline{p}(C_0\vee\delta)|p^*_t - P[i_t]| \\
&\le \frac{\overline{p}-\underline{p}}{2N}\left(\max_{p\in[\underline{p},\overline{p}]}|f(p)| + \overline{p}(C_0\vee\delta)\right). \qquad \text{(EC.61)}
\end{aligned}
$$

In the second inequality of (EC.61), from our assumption on f, if $k_f = 1$, f is Lipschitz continuous with Lipschitz constant δ, and if $k_f > 1$, since $\sup_{p\in[\underline{p},\overline{p}]} |f'(p)| \le C_0$, from the mean value theorem, we have $|f(p) - f(p')| = |f'(\xi)(p-p')| \le C_0|p-p'|$. In both cases, $|f(p) - f(p')| \le (C_0 \vee \delta)|p-p'|$.

Combining Eqs. (EC.60) and (EC.61), we have the upper bound on the first term of Eq. (EC.59):

$$\begin{aligned}&\max_{p\in[\underline{p},\overline{p}]} p(f(p)+g(x_t)) - \max_{0\le i\le N-1} P[i]\times(f(P[i])+\mathbb{E}[g(x)|x\in\mathbf{M}_{j_t}])\\ &\le \frac{\overline{p}-\underline{p}}{2N}\left(\max_{p\in[\underline{p},\overline{p}]}|f(p)| + \overline{p}(C_0\vee\delta) + \max_{x\in[0,1]^d}|g(x)|\right) + \frac{\overline{p}Ld^{\frac{k_g+\mathfrak{b}(k_g)}{2}}}{M^{k_g}\mathfrak{b}(k_g)!}\end{aligned} \tag{EC.62}$$

The second term in Eq. (EC.59) captures the regret due to the failure of identifying the best discretized price. We have the following the upper bound with probability at least $1-(N+M)^d\epsilon$, whose proof is delayed to the last part of this proof:

$$\begin{aligned}&\sum_{t=n_0N}^{T} \max_{i\in[N]} P[i]\times(f(P[i])+(\theta^*_{j_t})^\top\phi(x_t)) - p_t(f(p_t)+(\theta^*_{j_t})^\top\phi(x_t))\\ &= \widetilde{\mathcal{O}}\left(\frac{T}{\sqrt{n_0}}\sqrt{\log\frac{1}{\epsilon}} + \sqrt{M^dT}\sqrt{\log\frac{1}{\epsilon}} + \Delta T\right).\end{aligned} \tag{EC.63}$$

The third term in Eq. (EC.59) is due to the discretization error of the context space and is easily shown to have the following upper bound from our assumption $g\in\mathcal{G}_d(k_g,L)$ and definition of $\mathbf{M}_{j_t}$:

$$\mathbb{E}\left[p_t(f(p_t)+(\theta^*_{j_t})^\top\phi(x_t)) - p_t(f(p_t)+g(x_t))\right] \le \frac{\overline{p}Ld^{\frac{k_g+\mathfrak{b}(k_g)}{2}}}{M^{k_g}\mathfrak{b}(k_g)!} \tag{EC.64}$$

Putting Eqs. (EC.59) to (EC.64) together, we can have

$$\begin{aligned}&\sum_{t=n_0N+1}^{T}\mathbb{E}[r_t]\\ &= \mathcal{O}\left(\frac{T}{N}\right) + \mathcal{O}\left(\frac{LT}{M^{k_g}}\right) + \mathcal{O}\left(\frac{T}{\sqrt{n_0}}\sqrt{\log\frac{1}{\epsilon}}\right) + \mathcal{O}\left(\sqrt{TM^d\log\frac{1}{\epsilon}}\right) + \mathcal{O}((N+M^d)T\epsilon).\end{aligned} \tag{EC.65}$$

After adding the regret in the first n_0N periods, we obtain the upper bound on the total expected regret:

$$\begin{aligned}\sum_{t=1}^{T}\mathbb{E}[r_t] &= \mathcal{O}(n_0N) + \mathcal{O}\left(\frac{T}{N}\right) + \mathcal{O}\left(\frac{LT}{M^{k_g}}\right) + \mathcal{O}\left(\frac{T}{\sqrt{n_0}}\sqrt{\log\frac{1}{\epsilon}}\right) + \mathcal{O}\left(\sqrt{TM^d\log\frac{1}{\epsilon}}\right) + \mathcal{O}((N+M^d)T\epsilon)\\ &= \mathcal{O}(T^{\frac{3}{4}}) + \mathcal{O}(L^{\frac{d}{d+2k_g}}T^{\frac{d+k_g}{d+2k_g}}) + \mathcal{O}(T^{\frac{3}{4}}) + \mathcal{O}(T^{\frac{3}{4}}\sqrt{\log T}) + \mathcal{O}(L^{\frac{d}{d+2k_g}}T^{\frac{d+k_g}{d+2k_g}}\sqrt{\log T}) + \mathcal{O}(T^{-(\frac{3}{4}\wedge\frac{2k_g}{d+2k_g})})\\ &= \widetilde{\mathcal{O}}(T^{\frac{3}{4}}\vee L^{\frac{d}{d+2k_g}}T^{\frac{d+k_g}{d+2k_g}}),\end{aligned}$$

where in the second identity, we let $N = \lceil T^{\frac{1}{4}}\rceil$, $n_0 = \lceil T^{\frac{1}{2}}\rceil$, $M = \lceil (L^2T)^{\frac{1}{d+2k_g}}\rceil$ and $\epsilon = T^{-2}$. This completes the proof of Theorem 5.

Proof of Eq. (EC.63). To bound the second term in Eq. (EC.59), we note that

$$\begin{aligned}
&\max_{0\le i\le N-1} P[i]\times(f(P[i])+(\theta^*_{j_t})^\top\phi(x_t))-p_t\times\left(f(p_t)+(\theta^*_{j_t})^\top\phi(x_t)\right)\\
&=\left(\max_{0\le i\le N-1} P[i]\times(f(P[i])+(\theta^*_{j_t})^\top\phi(x_t))-\max_{0\le i\le N-1} P[i]\times\left(s_i+(\hat\theta_{t,j_t})^\top\phi(x_t)\right)\right)\\
&\quad+\left(\max_{0\le i\le N-1} P[i]\times\left(s_i+(\hat\theta_{t,j_t})^\top\phi(x_t)\right)-p_t\times\left(f(p_t)+(\theta^*_{j_t})^\top\phi(x_t)\right)\right)\\
&\le\left(P[\hat i^*_t]\times(f(P[\hat i^*_t])+(\theta^*_{j_t})^\top\phi(x_t))-P[\hat i^*_t]\times\left(s_{\hat i^*_t}+(\hat\theta_{t,j_t})^\top\phi(x_t)\right)\right)\\
&\quad+\left(p_t\times\left(s_{m_t}+(\hat\theta_{t,j_t})^\top\phi(x_t)\right)-p_t\times\left(f(p_t)+(\theta^*_{j_t})^\top\phi(x_t)\right)\right)\\
&\le 2\overline{p}\max_{0\le i\le N-1}|f(P[i])-s_i|+2\overline{p}|(\hat\theta_{t,j_t})^\top\phi(x_t)-(\theta^*_{j_t})^\top\phi(x_t)|.
\end{aligned}\tag{EC.66}$$

To bound the first term $2\overline{p}\max_{0\le i\le N-1}|f(P[i])-s_i|$ in (EC.66), recall that n_0 is the number of times each price $P[i]$ is explored and s_i records the average demand observed under price $P[i]$ in the exploration phase. Denote $P[0]=P[N]$. For each $1\le i\le N$, we have

$$\begin{aligned}
f(P[i])-s_i&=f(P[i])-\frac{\sum_{k=0}^{n_0-1}d_{kN+i}}{n_0}\\
&=f(P[i])-\frac{\sum_{k=0}^{n_0-1}(f(P[i])+g(x_{kN+i})+\varepsilon_{kN+i})}{n_0}\\
&=-\frac{\sum_{k=0}^{n_0-1}(g(x_{i+kN})+\varepsilon_{i+kN})}{n_0}.
\end{aligned}\tag{EC.67}$$

Since g is bounded on $[0,1]^d$ and $\mathbb{E}[g(x)]=0$, from Hoffding's lemma, $g(x_{i+kN})$ is sub-Gaussian with variance proxy $\frac{1}{4}\max_{x_1,x_2\in[0,1]^d}(g(x_1)-g(x_2))^2$. Since $\{x_t:t\ge1\}$ and $\{\varepsilon_t:t\ge1\}$ are independent, we have $g(x_{i+kN})+\varepsilon_{i+kN}$ is sub-Gaussian with variance proxy $\sigma_1'^2:=\frac{1}{4}\max_{x_1,x_2\in[0,1]^d}(g(x_1)-g(x_2))^2+\sigma^2$. For any $\epsilon>0$, denote $\mathcal{A}:=\{\max_{0\le i\le N-1}|f(P[i])-s_i|\le\sigma_1'\sqrt{\frac{2}{n_0}\log\frac{2}{\epsilon}}\}$. From the union bound and Chernoff inequality, we have

$$\mathbb{P}(\mathcal{A})\ge1-\sum_{i=1}^{N}\mathbb{P}\left(|f(P[i])-s_i|>\sigma_1'\sqrt{\frac{2}{n_0}\log\frac{2}{\epsilon}}\right)\ge1-N\epsilon.\tag{EC.68}$$

To bound the second term $2\overline{p}|\hat a_{t,j_t}-\mathbb{E}[g(x)|x\in\mathbf{M}_{j_t}]|$ in (EC.66), similar to Eq. (EC.6), we can have

$$\begin{aligned}
\hat\theta_{t,j_t}&=V_{t,j_t}^{-1}\sum_{(x_s,p_s,d_s)\in\mathcal{D}_{t,j_t}}(f(p_s)+g(x_s)+\varepsilon_s-s_{m_s})\phi(x_s)\\
&=V_{t,j_t}^{-1}\sum_{(x_s,p_s,d_s)\in\mathcal{D}_{t,j_t}}(f(p_s)+(\theta^*_{j_t})^\top\phi(x_s)+\Delta(s)+\varepsilon_s-s_{m_s})\phi(x_s)\\
&=\theta^*_j-\lambda V_{t,j_t}^{-1}\theta^*_jI+V_{t,j_t}^{-1}\sum_{(x_s,p_s,d_s)\in\mathcal{D}_{t,j_t}}(\Delta(s)+\varepsilon_s)\phi(x_s)+V_{t,j_t}^{-1}\sum_{(x_s,p_s,d_s)\in\mathcal{D}_{t,j_t}}(f(p_s)-s_{m_s})\phi(x_s).
\end{aligned}\tag{EC.69}$$

Moreover, we have the following observation under the event $\mathcal{A}$,

$$\begin{aligned}\left|x^\top V_{t,j_t}^{-1}\sum_{(x_s,p_s,d_s)\in\mathcal{D}_{t,j_t}}(f(p_s)-s_{m_s})\phi(x_s)\right| &\le \sqrt{\sum_{(x_s,p_s,d_s)\in\mathcal{D}_{t,j_t}}(x^\top V_{t,j_t}^{-1}\phi(x_s))^2\sum_{(x_s,p_s,d_s)\in\mathcal{D}_{t,j_t}}\Delta(s)^2}\\ &\le \sqrt{\sum_{(x_s,p_s,d_s)\in\mathcal{D}_{t,j_t}}(f(p_s)-s_{m_s})^2}\sqrt{\sum_{(x_s,p_s,d_s)\in\mathcal{D}_{t,j_t}}(x^\top V_{t,j_t}^{-1}\phi(x_s))^2}\\ &= \sigma_1'\sqrt{\frac{|\mathcal{D}_{t,j_t}|}{n_0}\log\frac{2}{\epsilon}}\sqrt{x^\top V_{t,j_t}^{-1}\left(\sum_{(x_s,p_s,d_s)\in\mathcal{D}_{t,j_t}}\phi(x_s)\phi(x_s)^\top\right)V_{t,j_t}^{-1}x}\\ &\le \sigma_1'\sqrt{\frac{|\mathcal{D}_{t,j_t}|}{n_0}\log\frac{2}{\epsilon}}\|x\|_{V_{t,j_t}^{-1}}. \end{aligned}\tag{EC.70}$$

Then, following the procedure of Eqs. (EC.8), (EC.9), and (EC.10), we can get the following event of high probability which we denote as event $\mathcal{B}_j$, with a fixed j, for any $t\ge 1$ such that $x_t\in\mathbf{M}_j$

$$\left|x^\top\hat{\theta}_{t,j}-x^\top\theta_j^*\right|\le\|x\|_{V_{t,j}^{-1}}\left(\lambda^{\frac{1}{2}}\|\theta^*\|_2+\gamma_{t,j}+\sqrt{|\mathcal{D}_{t,j}|}\Delta+\sigma_1'\sqrt{\frac{|\mathcal{D}_{t,j}|}{n_0}\log\frac{2}{\epsilon}}\right).\tag{EC.71}$$

Moreover, we define $\mathcal{B}=\cap_{j\in[M^d]}\mathcal{B}_j$, and we can know $\mathbb{P}(\mathcal{A},\mathcal{B})\ge 1-(N+M^d)\epsilon$.

Therefore, we have the following inequality under events $\mathcal{A}$ and $\mathcal{B}$,

$$\begin{aligned}&\sum_{t=n_0N}^{T}\max_{0\le i\le N-1}P[i]\times(f(P[i])+(\theta_{j_t}^*)^\top\phi(x_t))-p_t\times\left(f(p_t)+(\theta_{j_t}^*)^\top\phi(x_t)\right)\\ &\le 2\overline{p}\sigma_1'T\sqrt{\frac{2}{n_0}\log\frac{2}{\epsilon}}+2\overline{p}\sum_{t=1}^{T}|(\hat{\theta}_{t,j_t})^\top\phi(x_t)-(\theta_{j_t}^*)^\top\phi(x_t)|\\ &\le 2\overline{p}\sigma_1'T\sqrt{\frac{2}{n_0}\log\frac{2}{\epsilon}}+2\overline{p}\sum_{t=1}^{T}\|\phi(x_t)\|_{V_{t,j_t}^{-1}}\left(\lambda^{\frac{1}{2}}\|\theta^*\|_2+\gamma_{t,j_t}+\sqrt{|\mathcal{D}_{t,j_t}|}\Delta+\sigma_1'\sqrt{\frac{|\mathcal{D}_{t,j_t}|}{n_0}\log\frac{2}{\epsilon}}\right)\\ &= 2\overline{p}\sigma_1'T\sqrt{\frac{2}{n_0}\log\frac{2}{\epsilon}}+2\overline{p}\sum_{t=1}^{T}\sum_{j=1}^{M^d}\|\phi(x_t)\|_{V_{t,j}^{-1}}\left(\lambda^{\frac{1}{2}}\|\theta^*\|_2+\gamma_{t,j}+\sqrt{|\mathcal{D}_{t,j}|}\Delta+\sigma_1'\sqrt{\frac{|\mathcal{D}_{t,j}|}{n_0}\log\frac{2}{\epsilon}}\right)1_{x_t\in\mathbf{M}_j}\\ &\le 2\overline{p}\sigma_1'T\sqrt{\frac{2}{n_0}\log\frac{2}{\epsilon}}+2\overline{p}\sum_{j=1}^{M^d}\left(\lambda^{\frac{1}{2}}\|\theta^*\|_2+\gamma_{T,j}+\sqrt{|\mathcal{D}_{T,j}|}\Delta+\sigma_1'\sqrt{\frac{|\mathcal{D}_{T,j}|}{n_0}\log\frac{2}{\epsilon}}\right)\sum_{t=1}^{T}\|\phi(x_t)\|_{V_{t,j}^{-1}}1_{x_t\in\mathbf{M}_j}.\end{aligned}$$

Note that $\sum_{t=1}^{T}\|\phi(x_t)\|_{V_{t,j}^{-1}}1_{x_t\in\mathbf{M}_j}\le\sqrt{|\mathcal{D}_{T,j}|\sum_{t=1}^{T}\|\phi(x_t)\|_{V_{t,j}^{-1}}^2 1_{x_t\in\mathbf{M}_j}}\le\sqrt{2|\mathcal{D}_{T,j}|\log\frac{\det(V_{T,j})}{\det(\lambda I)}}\le\sqrt{2|\mathcal{D}_{T,j}|\log(\lambda+T)}$, and thus,

$$\begin{aligned}&\sum_{t=n_0N}^{T}\max_{0\le i\le N-1}P[i]\times(f(P[i])+(\theta_{j_t}^*)^\top\phi(x_t))-p_t\times\left(f(p_t)+(\theta_{j_t}^*)^\top\phi(x_t)\right)\\ &\le 2\overline{p}\sigma_1'T\sqrt{\frac{2}{n_0}\log\frac{2}{\epsilon}}+2\overline{p}\sum_{j=1}^{M^d}\left(\lambda^{\frac{1}{2}}\|\theta^*\|_2+\gamma_{T,j}+\sqrt{|\mathcal{D}_{T,j}|}\Delta+\sigma_1'\sqrt{\frac{|\mathcal{D}_{T,j}|}{n_0}\log\frac{2}{\epsilon}}\right)\sqrt{2|\mathcal{D}_{T,j}|\log(\lambda+T)}\end{aligned}$$

$$\leq 2\overline{p}\sigma_1' T\sqrt{\frac{2}{n_0}\log\frac{2}{\epsilon}} + 2\overline{p}(\lambda^{\frac{1}{2}}\left\|\theta^*\right\|_2 + \gamma_{T,1})\sqrt{2M^d T\log(\lambda+T)}$$
$$+\sum_{j=1}^{M^d}\left(|\mathcal{D}_{T,j}|\Delta\sqrt{2\log(\lambda+T)} + \sigma_1'|\mathcal{D}_{T,j}|\sqrt{\frac{2\log(\lambda+T)}{n_0}\log\frac{2}{\epsilon}}\right)$$
$$=\widetilde{\mathcal{O}}\left(\frac{T}{\sqrt{n_0}}\sqrt{\log\frac{1}{\epsilon}} + \sqrt{\log\frac{1}{\epsilon}}\sqrt{M^d T} + \Delta T\right). \tag{EC.72}$$

Moreover, by a simple application of the union bound, we can have the $\mathbb{P}(\mathcal{A},\mathcal{B}) \geq 1-(N+M^d)\epsilon$. Thus, Eq. (EC.63) holds.

We complete the proof. Q.E.D.

C.2. Proof for Theorem 6

Recall that we have constructed a function $f(p) \in \mathcal{F}(k_f,\delta)$ and s series of functions $\{g_{\mathbf{w}} : \mathbf{w} \in \{0,1\}^{M^d}\}$. Also, we fix the distribution of contexts to be a uniform distribution on $\bigcup_{j=1}^{M^d}\{x \in \mathbf{M}_j : D(x,\partial\mathbf{M}_j) > (4\eta)^{\frac{1}{k_g}}\}$, and the distribution of contexts to be standard normal. For the ease of presentation, we omit the dependency of the regret and the expectation on these terms. Define $\hat{\delta} = \delta/((\sum_{i=0}^{\mathfrak{b}(k_f)}\frac{\mathfrak{b}(k_f)!}{i!}) \vee (\mathfrak{b}(k_f)+1)2^{\mathfrak{b}(k_f)-1})$, $\eta = 2^{-3k_f-2}L^{\frac{d}{d+2k_g}}\hat{\delta}T^{-\frac{k_g}{d+2k_g}}$, $J = 4\lceil(T^{\frac{k_g}{d+2k_g}})^{\frac{1}{k_f}}\rceil$, and $M = \lceil 2^{\frac{3k_f-1-k_g}{k_g}}\hat{\delta}^{-\frac{1}{k_g}}L^{\frac{2}{d+2k_g}}T^{\frac{1}{d+2k_g}}\rceil$. For convenience, we also define $\mathbf{I}_e := [\frac{3}{2}-\frac{1}{2J}, \frac{3}{2}+\frac{1}{2J}]$.

For any policy π, we establish the following lower bound on the worst-case regret by restricting the functions $f(\cdot)$ and $g_{\mathbf{w}}(\cdot)$ to what we have constructed:

$$\sup_{f\in\mathcal{F}(k_f,\delta),g\in G_d(k_g,L)} R_{f,g}^{\pi}(T) \geq \sup_{g\in\{g_{\mathbf{w}}:\mathbf{w}\in\{0,1\}^{M^d}\}} R_{f,g}^{\pi}(T)$$
$$\geq \frac{1}{2^{M^d}}\sum_{\mathbf{w}}\mathbb{E}_{f,g_{\mathbf{w}}}^{\pi}[\sum_{t=1}^{T}p^*(x_t)(f(p^*(x_t))+g_{\mathbf{w}}(x_t)) - p_t(f(p_t)+g_{\mathbf{w}}(x_t))]$$
$$= \frac{1}{2^{M^d}}\sum_{\mathbf{w}}\sum_{j=1}^{M^d}\mathbb{E}_{f,g_{\mathbf{w}}}^{\pi}\left[\left(\sum_{t=1}^{T}p^*(x_t)(f(p^*(x_t))+g_{\mathbf{w}}(x_t)) - p_t(f(p_t)+g_{\mathbf{w}}(x_t))\mathbb{I}_{x_t\in\mathbf{M}_j}\right)\right]$$
$$= \frac{1}{2^{M^d}M^d}\sum_{j=1}^{M^d}\sum_{\mathbf{w}_j\in\{0,1\}^{M^d-1}}\sum_{w_j\in\{0,1\}}\sum_{t=1}^{T}$$
$$\mathbb{E}_{f,g_{\mathbf{w}}}^{\pi}[p^*(x_t)(f(p^*(x_t))+g_{\mathbf{w}}(x_t)) - p_t(f(p_t)+g_{\mathbf{w}}(x_t)) \mid x_t\in\mathbf{M}_j]$$
$$= \frac{1}{2^{M^d}M^d}\sum_{j=1}^{M^d}\sum_{\mathbf{w}_j\in\{0,1\}^{M^d-1}}\sum_{t=1}^{T}$$
$$\mathbb{E}_{f,g_{(\mathbf{w}_{-j},0)}}^{\pi}[p^*(x_t)(f(p^*(x_t))+g_{\mathbf{w}}(x_t)) - p_t(f(p_t)+g_{\mathbf{w}}(x_t)) \mid x_t\in\mathbf{M}_j]$$
$$+\mathbb{E}_{f,g_{(\mathbf{w}_{-j},1)}}^{\pi}[p^*(x_t)(f(p^*(x_t))+g_{\mathbf{w}}(x_t)) - p_t(f(p_t)+g_{\mathbf{w}}(x_t)) \mid x_t\in\mathbf{M}_j]$$
$$= \frac{1}{2^{M^d}M^d}\sum_{j=1}^{M^d}\sum_{\mathbf{w}_j\in\{0,1\}^{M^d-1}}\sum_{t=1}^{T}\mathbb{E}_{f,g_{(\mathbf{w}_{-j},0)}}^{\pi}[p^*(x_t)f(p^*(x_t)) - p_t f(p_t) \mid x_t\in\mathbf{M}_j]$$

$$+\mathbb{E}^{\pi}_{f,g_{(\mathbf{w}_{-j},1)}}\left[p^*(x_t)(f(p^*(x_t))+g_{\mathbf{w}}(x_t))-p_t(f(p_t)+g_{\mathbf{w}}(x_t))\mid x_t\in\mathbf{M}_j\right].$$

By the definition of $f(\cdot)$, we can easily have

$$\mathbb{E}^{\pi}_{f,g_{(\mathbf{w}_{-j},0)}}\left[p^*(x_t)f(p^*(x_t))-p_tf(p_t)\mid x_t\in\mathbf{M}_j\right]\geq\frac{\eta}{Z_{k_f}}\mathbb{P}^{\pi,t}_{f,g_{(\mathbf{w}_{-j},0)}}(p_t\notin\mathbf{I}_e\mid x_t\in\mathbf{M}_j),\tag{EC.73}$$

where Z_{k_f} is the constant in the definition of the function $g_{k_f}(\cdot)$ in Eq. (12). Eq. (EC.73) can be interpreted as whenever p_t is not in $\mathbf{I}_e$, a regret of $\frac{\eta}{Z_{k_f}}$ will occur. As for $\mathbb{E}_{f,g_{(\mathbf{w}_{-j},1)}}\left[p^*(x_t)(f(p^*(x_t))+g_w(x_t))-p_t(f(p_t)+g_w(x_t))\mid x_t\in\mathbf{M}_j\right]$, when p_t falls into $\mathbf{I}_e$ and $x_t\in\mathbf{M}_j$, $p_t(f(p_t)+g_w(x_t))\leq\frac{1}{2}\hat{\delta}+\frac{\eta}{Z_k}+4\eta(\frac{3}{2}+\frac{1}{2J})\leq\frac{1}{2}\hat{\delta}+\frac{15}{2}\eta$. Besides, the value of $p^*(x_t)(f(p^*(x_t))+g(x_t))$ is $\frac{1}{2}\hat{\delta}+8\eta$ and $p^*(x_t)=2$. This means that when p_t falls into $\mathbf{I}_e$, a regret of at least $\frac{\eta}{2}$ will occur. Mathematically, we have

$$\mathbb{E}^{\pi}_{f,g_{(\mathbf{w}_{-j},1)}}\left[p^*(x_t)(f(p^*(x_t))+g_w(x_t))-p_t(f(p_t)+g_w(x_t))\mid x_t\in\mathbf{M}_j\right]\geq\frac{\eta}{2}\mathbb{P}^{\pi,t}_{f,g_{(\mathbf{w}_{-j},1)}}(p_t\in\mathbf{I}_e\mid x_t\in\mathbf{M}_j).\tag{EC.74}$$

The following route is similar to what we have derived in Appendix A.2. Again, we apply Bretagnolle–Huber inequality (see Bretagnolle and Huber 1979) to get: given $x_t\in\mathbf{M}_j$,

$$\mathbb{P}^{\pi,t}_{f,g_{(\mathbf{w}_{-j},0)}}(p_t\notin\mathbf{I}_e)+\mathbb{P}^{\pi,t}_{f,g_{(\mathbf{w}_{-j},1)}}(p_t\in\mathbf{I}_e)\geq\frac{1}{2}\exp\left(-\mathrm{KL}\left(\mathbb{P}^{\pi,t}_{f,g_{(\mathbf{w}_{-j},0)}},\mathbb{P}^{\pi,t}_{f,g_{(\mathbf{w}_{-j},1)}}\right)\right),\tag{EC.75}$$

where $\mathbb{P}^{\pi,t}_{f,g_{(\mathbf{w}_{-j},w_j)}}$ denotes the probability measure under policy π up to period t when the true demand function is made up of f and $g_{(\mathbf{w}_{-j},w_j)}$. With the chain rule of KL divergence, we have

$$\mathrm{KL}(\mathbb{P}^{\pi,t}_{f,g_{(\mathbf{w}_{-j},0)}},\mathbb{P}^{\pi,t}_{f,g_{(\mathbf{w}_{-j},1)}})\leq\frac{1}{2\sigma^2M^d}(4\eta)^2t.\tag{EC.76}$$

Hence, together with Eqs. (EC.75) and (EC.76), we can have

$$\sup_{f\in\mathcal{F}(k_f,\delta),g\in\mathcal{G}_d(k_g,L)}R^{\pi}_{\mathcal{P},f,g}(T)\geq\frac{\eta}{2(Z_{k_f}\vee 2)}T\exp\left(-\frac{1}{2\sigma^2M^d}(4\eta)^2T\right).\tag{EC.77}$$

By plugging the value of M and η, we can get the desired $\Omega(L^{\frac{d}{d+2k_g}}T^{\frac{d+k_g}{d+2k_g}})$ lower bound. Q.E.D.

Appendix D. Extensions to Separable Contextual Effect

In practice, a decision maker may have some prior knowledge about the relationship among features. For example, in time series analysis, a common operator is to decompose the demand into secular trends, seasonal variations, cyclical variations and irregular variations (see, e.g., Hamilton 2020), and consider these four components as independent ones. In this case, for example, the seasonal features and the cyclical features can be placed into two independent groups. This motivates us to consider the extension to separable contextual effect. Specifically, we express the context vector

$x \in \mathbb{R}^d$ as $[x(1), x(2), \cdots, x(n)]$, where each vector $x(i) \in \mathbb{R}^{d_i}$ represents a separate group for context. Moreover, we consider the following form of function $g(\cdot)$:

$$g(x) = g_0 + \sum_{i=1}^{n} g_i(x(i)), \tag{EC.78}$$

where g_0 is an unknown constant, and $\mathbb{E}[g_i(x(i))] = 0$ for all $i \in [n]$. In Eq. (EC.78), each group of features influences the expected demand independently, and changing one or several features in one group does not affect the other groups' contribution to the demand. When $d_i = 1$ for any $i \in [n]$ (i.e., each feature itself is a group), Eq. (EC.78) reduces to the well-known additive model in non-parametric regression literature (see, e.g., Buja et al. 1989).

Throughout this section, we make the following assumptions.

ASSUMPTION EC.1. *(a) For each $i \in [n]$, $g_i(\cdot) \in \mathcal{G}_{d_i}(k_{gi}, L_i)$ for some $0 < k_{gi} \le 1$;*
(b) $\{x(i)\}_{i=1}^n$ are jointly independent random vectors.

D.1. SMLPE with Separable Contextual Effect

In this subsection, we consider the extension of Sec. 2 and assume the following demand function:

$$D_t(p) = bp + g_0 + \sum_{i=1}^{n} g_i(x_t(i)) + \varepsilon_t, \qquad \forall p \in [\underline{p}, \overline{p}]. \tag{EC.79}$$

All the assumptions on the boundness of b, the distributions of context x and random noise ε remain the same as those in Sec. 2. We have the following theorem showing the optimal regret for the learning problem under model (EC.79).

THEOREM EC.1. *The optimal regret of the learning problem under model* (EC.79) *under Assumption EC.1 is*

$$\inf_{\pi} \sup_{\substack{g_i \in \mathcal{G}_{d_i}(k_{gi}, L_i) \forall i \in [n], \\ b \in [\underline{b}, \overline{b}], \mathcal{P}, \mathcal{Q} \in \mathcal{E}(\sigma)}} R^{\pi}_{g,b,\mathcal{P},\mathcal{D}}(T) = \widetilde{\Theta}\Big(\sqrt{T} \vee \max_{i \in [n]} (L_i^2 T)^{\frac{d_i}{d_i + 2k_{gi}}}\Big). \tag{EC.80}$$

Theorem EC.1 tells that the learning complexity is decided by the group of features that achieves $\max_{i \in [n]} (L_i^2 T)^{\frac{d_i}{d_i + 2k_{gi}}}$. From Theorems 1 and 2, we know that if $g(x)$ is k_{gi}th-order smooth with respect to each $x(i)$ and $L_i = \Theta(1)$ but does not have the separable structure, the optimal regret is $\widetilde{\Theta}(\sqrt{T} \vee T^{\frac{d}{d + 2\min_{i \in [n]} k_{gi}}})$, which is no less than the bound in Eq. (EC.80) since $d_i \le d$ and $k_{gi} \ge \min_{i \in [k]} k_{gi}$. To further see the benefit of the separable structure, if all $k_{gi}(\cdot)$'s share the same smoothness parameters k_{gi} and L_i for all $i \in [n]$, $\max_{i \in [n]} (L_i^2 T)^{\frac{d_i}{d_i + 2k_{gi}}}$ is decided by the largest d_i. Thus, if each d_i is relatively small compared with the total dimension d, the separable structure will greatly reduce the learning complexity. If all the L_i's and d_i's are the same, the bottleneck becomes the one with the smallest k_{gi}. The intuition is that the least smooth function is the hardest part of

the learning problem. Another special case is that when $L_i = \Theta(1)$, $d_i = 1$ and $k_{gi} \geq \frac{1}{2}$ for all $i \in [d]$, the demand function is completely decomposed into $n+1$ separate functions in $p, x(1), \ldots, x(n)$, and the optimal regret reduces to $\widetilde{\Theta}(\sqrt{T})$.

In Appendix D.1, we design Algorithm 5 for the demand model in (EC.79) and prove its regret upper bound in Theorem EC.3. The main ideas of Algorithm 5 are similar to Algorithm 1. The difference lies in that instead of partitioning $[0,1]^d$ into M^d bins, we separate the space $[0,1]^{d_i}$ of each context subgroup into $M_i^{d_i}$ bins. The intuition of why we get a better regret upper bound compared with SMLPE under the demand function in (EC.79) is that the total number of bins $\sum_{i=1}^n M_i^{d_i}$ can be greatly smaller than M^d since $d = \sum_{i=1}^n d_i$, meaning that the parameters we need to estimate in the local approximation can be significantly reduced.

The lower bound can be directly implied from Theorem 4 due to the following reason. When constructing $g(\cdot)$, we set $g_i(\cdot)$ for each subgroup $i \in [n]$ to 0, except the subgroup $i_{\max} := \max_{i\in[n]} (L_i^2 T)^{\frac{d_i}{d_i+2k_{gi}}}$. We consider a simpler scenario, where the decision maker is informed that all $g_i(\cdot)$'s are zero except $g_{i_{\max}}(\cdot)$, and thus the only function that need to be learned is $g_{i_{\max}}(\cdot)$. Then the demand model is reduced to $bp + g_{i_{\max}}(x(i_{\max})) + \varepsilon$, which is the exact model studied in Sec. 2. For this simpler scenario, we have obtained the desired lower bound from Theorem 4, which naturally serves as the lower bound for the original problem.

D.2. SMNPE with Separable Contextual Effect

In this subsection, we consider the extension of Sec. 4 and assume the following demand function:

$$D_t(p) = f(p) + \sum_{i=1}^n g_i(x_t(i)) + \varepsilon_t, \qquad \forall p \in [\underline{p}, \overline{p}]. \tag{EC.81}$$

Note that in Eq. (EC.81), compared with Eq. (EC.78), we absorb the constant g_0 into $f(p)$. The other assumptions on the price function $f(\cdot)$, the context distribution and the distribution of the random noise remain the same as those in Sec. 4. We have the following theorem in this subsection.

THEOREM EC.2. *If* $\max_{i\in[n]} \frac{d_i+k_{gi}}{d_i+2k_{gi}} \geq \frac{3}{4}$ *and* $L_i = \Theta(1)$*, the optimal regret of the learning problem under* (EC.81) *is*

$$\inf_{\pi} \sup_{\substack{g_i \in \mathcal{G}_{d_i}(k_{gi}, L_i) \forall i \in [k], \\ f \in \mathcal{F}(k_f, \delta), \mathcal{P}, \mathcal{Q} \in \mathcal{E}(\sigma)}} R^{\pi}_{g,f,\mathcal{P},\mathcal{D}}(T) = \widetilde{\Theta}\left(T^{\max_{i\in[n]} \frac{d_i+k_{gi}}{d_i+2k_{gi}}}\right). \tag{EC.82}$$

Comparing Theorem EC.2 with the regret bounds derived in Sec. 4, when $\max_{i\in[n]} \frac{d_i+k_{gi}}{d_i+2k_{gi}} \geq \frac{3}{4}$, the learning complexity is decided by the group of features that achieves $\max_{i\in[n]} \frac{d_i+k_{gi}}{d_i+2k_{gi}}$, smaller than $\max_{i\in[n]} \frac{d+k_{gi}}{d+2k_{gi}}$. When all the k_{gi}'s are the same, the complexity is decided by the largest dimension d_i. Similar to the observation with Sec. 7, when the largest dimension d_i is greatly smaller than the total dimension d, the learning complexity can be significantly reduced. When all the dimension d_i is the same, then the bottleneck of the learning problem is the smallest k_{gi}

since $\frac{d_i+k_{gi}}{d_i+2k_{gi}}$ increases when k_{gi} decreases. The only case that the condition $\max_{i\in[n]}\frac{d_i+k_{gi}}{d_i+2k_{gi}}\geq\frac{3}{4}$ does not hold is that $d_i=1$ and $k_{gi}>\frac{1}{2}$ for all $i\in[n]$, which requires that all the features' effects are separated from each other and all the functions g_i's are smooth enough. When there exists at least one group of features whose dimension is larger than 1, $\max_{i\in[n]}\frac{d_i+k_{gi}}{d_i+2k_{gi}}\geq\frac{3}{4}$ is satisfied.

Following similar ideas in Sec. 4.1, we design Algorithm 6 in Appendix D.2 and prove its regret upper bound $\widetilde{\mathcal{O}}(T^{\frac{3}{4}}\vee T^{\max_{i\in[n]}\frac{d_i+k_{gi}}{d_i+2k_{gi}}})$ in Theorem EC.4. To see the regret lower bound, similar to Sec. 7, we consider an easier case by setting all $g_j(\cdot)$'s to zero except the one with index $i_{\max}=\arg\max_{i\in[n]}\frac{d_i+k_{gi}}{d_i+2k_{gi}}$, for which we construct a similar function to that in the proof of Theorem 6. Knowing that $g_j(\cdot)=0$ for all $j\neq i_{\max}$, the decision maker faces the same learning problem as Sec. 4. From Theorem 6, we obtain the regret lower bound $\Omega(T^{\frac{d_{i_{\max}}+k_{gi_{\max}}}{d_{i_{\max}}+2k_{gi_{\max}}}})=\Omega(T^{\max_{i\in[n]}\frac{d_i+k_{gi}}{d_i+2k_{gi}}})$.

D.3. Algorithm and Regret Upper Bound for Appendix D.1

THEOREM EC.3. *Suppose Algorithm 5 runs with $M_i=\lceil(L_i^2T)^{\frac{1}{d_i+2k_{gi}}}\rceil$. The regret of Algorithm 5 is upper bounded by*

$$\widetilde{\mathcal{O}}\Big(\sqrt{T}\vee\max_{i\in[n]}(L_i^2T)^{\frac{d_i}{d_i+2k_{gi}}}\Big). \tag{EC.83}$$

Proof. For notation convenience, we denote $g_0+\sum_{i=1}^n g_i(x_t(i))$ as $\sum_{i=0}^n g_i(x_t(i))$. The total regret is upper bounded as follows:

$$\begin{aligned}
&\sum_{t=1}^T\mathbb{E}[r_t]\\
&=|b|\sum_{t=1}^T\mathbb{E}\left[\left(-\frac{\sum_{i=0}^n g_i(x_t(i))}{2b}-p_t\right)^2\right]\\
&\leq 2|b|\sum_{t=1}^T\mathbb{E}\left[\left(-\frac{\sum_{i=0}^n g_i(x_t(i))}{2b}-p_t^0\right)^2+\left(p_t^0-p_t\right)^2\right]\\
&\leq 2|b|\sum_{t=1}^T\mathbb{E}\left[\left(-\frac{\sum_{i=0}^n g_i(x_t(i))}{2b}-p_t^u\right)^2+\left(p_t^0-p_t\right)^2\right]\\
&=2|b|\sum_{t=1}^T\mathbb{E}\left[\left(-\frac{\sum_{i=0}^n g_i(x_t(i))}{2b}+\frac{\hat{g}_{0,t}+\sum_{i=1}^n\hat{a}_{t,i,\mathbf{j}_t(i)}}{2\hat{b}_t}\right)^2+\left(p_t^0-p_t\right)^2\right]\\
&\leq 6|b|\sum_{t=1}^T\mathbb{E}\Bigg[\left(\frac{-\sum_{i=0}^n g_i(x_t(i))+g_0+\sum_{i=1}^n\mathbb{E}[g_i(x(i))|x(i)\in\mathbf{M}^i_{\mathbf{j}_t(i)}]}{2b}\right)^2\\
&\qquad+\left(\frac{g_0+\sum_{i=1}^n\mathbb{E}[g_i(x(i))|x(i)\in\mathbf{M}^i_{\mathbf{j}_t(i)}]}{2b}-\frac{g_0+\sum_{i=1}^n\mathbb{E}[g_i(x(i))|x(i)\in\mathbf{M}^i_{\mathbf{j}_t(i)}]}{2\hat{b}_t}\right)^2\\
&\qquad+\left(\frac{g_0+\sum_{i=1}^n\mathbb{E}[g_i(x(i))|x(i)\in\mathbf{M}^i_{\mathbf{j}_t(i)}]-\hat{g}_{0,t}-\sum_{i=1}^n\hat{a}_{t,i,\mathbf{j}_t(i)}}{2\hat{b}_t}\right)^2+\left(p_t^0-p_t\right)^2\Bigg].
\end{aligned}\tag{EC.84}$$

Algorithm 5: Algorithm for SMLPE with Separable Context (`ASMLPE-SC`)

Input: price range $[\underline{p},\overline{p}]$, number of context groups n, bounds on the price coefficient $\underline{b}$ and $\overline{b}$, dimensions of each context group $d_1,\cdots,d_n$, parameters of bins $M_1,\cdots,M_n$.

Initialization:

for $i=1,2,\cdots,n$ **do**:

  Partition $[0,1]^{d_i}$ into $M_i^{d_i}$ cubes of equal size, denoted as $\mathbf{M}_j^i$ for $j=1,2,\cdots,M_i^{d_i}$;

  Initialize $\mathcal{D}_{i,j}=\emptyset$ for each $j\in[M_i^{d_i}]$;

end for

Initialize $\hat{b}_1=\frac{\underline{b}+\overline{b}}{2}$ and $\hat{g}_{0,1}=0$;

Main Steps:

for $t=1,2,\cdots,T$ **do**

  Set $\delta_t\leftarrow t^{-\frac{1}{4}}$;

  for $i=1,2,\cdots,n$ **do**:

    Observe $x_t(i)\in\mathbf{M}^i_{\mathbf{j}_t(i)}$ for some $\mathbf{j}_t(i)\in[M_i^{d_i}]$;

    If $\mathcal{D}_{t,i,\mathbf{j}_t(i)}=\emptyset$, set $\hat{a}_{t,i,\mathbf{j}_t(i)}\leftarrow 0$; otherwise, set $\hat{a}_{t,i,\mathbf{j}_t(i)}\leftarrow\frac{\sum_{(x_k,p_k,d_k)\in\mathcal{D}_{t,i,\mathbf{j}_t(i)}}(d_k-p_k\hat{b}_t-\hat{g}_{0,t})}{|\mathcal{D}_{t,i,\mathbf{j}_t(i)}|}$;

  end for

  Set unconstrained greedy price: $p_t^u\leftarrow-\frac{\hat{g}_{0,t}+\sum_{i=1}^n\hat{a}_{t,i,\mathbf{j}_t(i)}}{2\hat{b}_t}$;

  Project greedy price: $p_t^g\leftarrow\mathrm{Proj}(p_t^u,[\underline{p}+\delta_t,\overline{p}-\delta_t])$;

  Generate an independent random variable $\Delta_t\leftarrow\delta_t$ w.p. $\frac{1}{2}$ and $\Delta_t\leftarrow-\delta_t$ w.p. $\frac{1}{2}$;

  Set price $p_t\leftarrow p_t^g+\Delta_t$;

  Observe realized demand d_t;

  for $i=1,2,\cdots,n$ **do**:

    Update $\mathcal{D}_{t+1,i,\mathbf{j}_t(i)}\leftarrow\mathcal{D}_{t,i,\mathbf{j}_t(i)}\cup\{(x_t,p_t,d_t)\}$ and $\mathcal{D}_{t+1,i,j}\leftarrow\mathcal{D}_{t,i,j}$ for $j\neq\mathbf{j}_t(i)$;

  end for

  Update $\hat{b}_{t+1}\leftarrow\mathrm{Proj}(\frac{\sum_{s=1}^t\Delta_s d_s}{\sum_{s=1}^t(\Delta_s)^2},[\underline{b},\overline{b}])$;

  Set $\hat{g}_{0,t+1}\leftarrow\frac{\sum_{l=1}^t(d_l-\hat{b}_{t+1}p_l)}{t}$;

end for

Bound the first term of Eq. (EC.84). From Cauchy-Schwarz inequality, similar to Eq. (EC.2), we have

$$\left(\frac{-\sum_{i=1}^n g_i(x_t(i))+\sum_{i=1}^n\mathbb{E}[g_i(x(i))|x(i)\in\mathbf{M}^i_{\mathbf{j}_t(i)}]}{2b}\right)^2\le\frac{n}{4b^2}\sum_{i=1}^n\left(\mathbb{E}[g_i(x(i))|x(i)\in\mathbf{M}^i_{\mathbf{j}_t(i)}]-g_i(x_t(i))\right)^2$$
$$\le\frac{n}{4b^2}\sum_{i=1}^n\frac{L_i^2 d_i^{k_{gi}}}{M_i^{2k_{gi}}}. \tag{EC.85}$$

Bound the second term of Eq. (EC.84). For the second term on the RHS of Eq. (EC.84), we have

$$\left(\frac{g_0+\sum_{i=1}^n \mathbb{E}[g_i(x(i))|x(i)\in \mathbf{M}^i_{\mathbf{j}_t(i)}]}{2b} - \frac{g_0+\sum_{i=1}^n \mathbb{E}[g_i(x(i))|x(i)\in \mathbf{M}^i_{\mathbf{j}_t(i)}]}{2\hat{b}_t}\right)^2 \le \frac{\max_{x\in[0,1]^d}(g(x))^2}{4\underline{b}^4}(\hat{b}_t-b)^2.$$

Note that the following inequality continues to hold from the same arguments to Eq. (EC.4):

$$\mathbb{E}[(\hat{b}_t-b)^2] \le \frac{3(\underline{b}^2\overline{p}^2+\max_{x\in[0,1]^d}(g(x))^2+\sigma^2)}{\sqrt{t}} := c_{e,1}/\sqrt{t}. \tag{EC.86}$$

Thus, the second term is upper bounded by $\mathcal{O}(t^{-\frac{1}{2}})$.

Bound the third term of Eq. (EC.84). To bound the third term on the RHS of Eq. (EC.84), we first apply Cauchy-Schwarz inequality and obtain

$$\begin{aligned}&\left(\frac{g_0+\sum_{i=1}^n \mathbb{E}[g_i(x(i))|x(i)\in \mathbf{M}^i_{\mathbf{j}_t(i)}]-\hat{g}_{0,t}-\sum_{i=1}^n \hat{a}_{t,i,\mathbf{j}_t(i)}}{2\hat{b}_t}\right)^2\\ &\le \frac{n+1}{4\underline{b}^2}\left((g_0-\hat{g}_{0,t})^2+\sum_{i=1}^n(\mathbb{E}[g_i(x(i))|x(i)\in \mathbf{M}^i_{\mathbf{j}_t(i)}]-\hat{a}_{t,i,\mathbf{j}_t(i)})^2\right),\end{aligned} \tag{EC.87}$$

Note that for each $t\ge 2$, the first term $(g_0-\hat{g}_{0,t})^2$ in the RHS of Eq. (EC.87) is upper bounded as follows:

$$\begin{aligned}\mathbb{E}\left[(g_0-\hat{g}_{0,t})^2\right] &= \mathbb{E}\left[\left(g_0-\frac{\sum_{l=1}^{t-1}(d_l-\hat{b}_t p_l)}{t-1}\right)^2\right]\\ &= \mathbb{E}\left[\left(g_0-\frac{\sum_{l=1}^{t-1}(bp_l+g_0+\sum_{i=1}^n g_i(x_l(i))+\varepsilon_l-\hat{b}_t p_l)}{t-1}\right)^2\right]\\ &\le 3\mathbb{E}\left[\left(\frac{\sum_{l=1}^{t-1}(bp_l-\hat{b}_t p_l)}{t-1}\right)^2+\left(\frac{\sum_{l=1}^{t-1}(g(x_l)-g_0)}{t-1}\right)^2+\left(\frac{\sum_{l=1}^{t-1}\varepsilon_l}{t-1}\right)^2\right]\\ &\le 3\left[\overline{p}^2\mathbb{E}\left[(\hat{b}_t-b)^2\right]+\frac{\max_{x,y\in[0,1]^d}(g(x)-g(y))^2}{4(t-1)}+\frac{\sigma^2}{t-1}\right]\\ &\le 3\left[\overline{p}^2\mathbb{E}\left[(\hat{b}_t-b)^2\right]+\frac{n\sum_{i=1}^n L_i^2 d_i^{k_{gi}}}{4(t-1)}+\frac{\sigma^2}{t-1}\right],\end{aligned} \tag{EC.88}$$

where in the first inequality, we use the Cauchy-Schwarz inequality and $\sum_{i=1}^n g_i(x_l(i))=g(x_l)-g_0$, and the second inequality follows from Hoeffding lemma and the fact that $\{g(x_l)-g_0: 1\le l\le t-1\}$ are i.i.d. bounded r.v.'s with zero mean.

For the second term $\sum_{i=1}^n\left(\mathbb{E}[g_i(x(i))|x(i)\in \mathbf{M}^i_{\mathbf{j}_t(i)}]-\hat{a}_{t,i,\mathbf{j}_t(i)}\right)^2$ in the RHS of Eq. (EC.87), since the analysis is similar for $i=1,2,\ldots,n$, we next focus on the case of $i=1$. Denote the time indices

when the first group of context (i.e., $x(i)$) falls into $\mathbf{M}^1_{\mathbf{j}_t(1)}$ by $1 \le s_1 < s_2 < \cdots < s_{|\mathcal{D}_{t,1,\mathbf{j}_t(1)}|} \le t-1$. If $\mathcal{D}_{t,1,\mathbf{j}_t(1)} \neq \emptyset$, we have

$$
\begin{aligned}
&\mathbb{E}\left[\left(\mathbb{E}[g_1(x(1))|x(1)\in\mathbf{M}^1_{\mathbf{j}_t(1)}] - \hat{a}_{t,1,\mathbf{j}_t(1)}\right)^2\right] \\
&= \mathbb{E}\left[\left(\mathbb{E}[g_1(x(1))|x(1)\in\mathbf{M}^1_{\mathbf{j}_t(1)}] - \frac{\sum_{(x_k,p_k,d_k)\in\mathcal{D}_{t,1,\mathbf{j}_t(1)}}(d_k - p_k\hat{b}_t - \hat{g}_{0,t})}{\left|\mathcal{D}_{t,1,\mathbf{j}_t(1)}\right|}\right)^2\right] \\
&= \mathbb{E}\left[\left(\mathbb{E}[g_1(x(1))|x(1)\in\mathbf{M}^1_{\mathbf{j}_t(1)}] - \frac{\sum_{(x_k,p_k,d_k)\in\mathcal{D}_{t,1,\mathbf{j}_t(1)}}(p_k b + \sum_{i=0}^k g_i(x_k(i)) + \varepsilon_k - p_k\hat{b}_t - \hat{g}_{0,t})}{\left|\mathcal{D}_{t,1,\mathbf{j}_t(1)}\right|}\right)^2\right] \\
&\le 5\mathbb{E}\Bigg[\left(\mathbb{E}[g_1(x(1))|x(1)\in\mathbf{M}^1_{\mathbf{j}_t(1)}] - \frac{\sum_{(x_k,p_k,d_k)\in\mathcal{D}_{t,1,\mathbf{j}_t(1)}} g_1(x_k(1))}{|\mathcal{D}_{t,1,\mathbf{j}_t(1)}|}\right)^2 + \overline{p}^2(b-\hat{b}_t)^2 + (g_0-\hat{g}_{0,t})^2 \\
&\qquad + \left(\frac{\sum_{(x_k,p_k,d_k)\in\mathcal{D}_{t,1,\mathbf{j}_t(1)}}\sum_{i=2}^n g_i(x_k(i))}{|\mathcal{D}_{t,1,\mathbf{j}_t(1)}|}\right)^2 + \left(\frac{\sum_{(x_k,p_k,d_k)\in\mathcal{D}_{t,1,\mathbf{j}_t(1)}}\varepsilon_k}{|\mathcal{D}_{t,1,\mathbf{j}_t(1)}|}\right)^2\Bigg]. \qquad \text{(EC.89)}
\end{aligned}
$$

The first term on the RHS of Eq. (EC.89) can be bounded as follows:

$$
\begin{aligned}
&\mathbb{E}\left[\left(\mathbb{E}[g_1(x(1))|x(1)\in\mathbf{M}^1_{\mathbf{j}_t(1)}] - \frac{\sum_{(x_k,p_k,d_k)\in\mathcal{D}_{t,1,\mathbf{j}_t(1)}} g_1(x_k(1))}{|\mathcal{D}_{t,1,\mathbf{j}_t(1)}|}\right)^2\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\left(\mathbb{E}[g_1(x(1))|x(1)\in\mathbf{M}^1_{\mathbf{j}_t(1)}] - \frac{\sum_{(x_k,p_k,d_k)\in\mathcal{D}_{t,1,\mathbf{j}_t(1)}} g_1(x_k(1))}{|\mathcal{D}_{t,1,\mathbf{j}_t(1)}|}\right)^2 \middle| s_1,\cdots,s_{|\mathcal{D}_{t,1,\mathbf{j}_t(1)}|}, |\mathcal{D}_{t,1,\mathbf{j}_t(1)}|, \mathbf{j}_t(1)\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\left(\mathbb{E}[g_1(x(1))|x(1)\in\mathbf{M}^1_{\mathbf{j}_t(1)}] - \frac{\sum_{l\in[|\mathcal{D}_{t,1,\mathbf{j}_t(1)}|]} g_1(x_{s_l}(1))}{|\mathcal{D}_{t,1,\mathbf{j}_t(1)}|}\right)^2 \middle| s_1,\cdots,s_{|\mathcal{D}_{t,1,\mathbf{j}_t(1)}|}, |\mathcal{D}_{t,1,\mathbf{j}_t(1)}|, \mathbf{j}_t(1)\right]\right] \\
&\le \mathbb{E}\left[\mathbb{E}\left[\frac{\left(\max_{x,y\in\mathbf{M}^1_{\mathbf{j}_t(1)}}(g_1(x)-g_1(y))\right)^2}{4}\frac{1}{|\mathcal{D}_{t,1,\mathbf{j}_t(1)}|}\right] \middle| |\mathcal{D}_{t,1,\mathbf{j}_t(1)}|, \mathbf{j}_t(1)\right] \\
&\le \frac{\left(\max_{x,y\in[0,1]^{d_1}}(g_1(x)-g_1(y))\right)^2}{4}\mathbb{E}\left[\frac{1}{|\mathcal{D}_{t,1,\mathbf{j}_t(1)}|}\right] \\
&\le \frac{L_1^2 d_1^{k_{g1}}}{4}\mathbb{E}\left[\frac{1}{|\mathcal{D}_{t,1,\mathbf{j}_t(1)}|}\right], \qquad \text{(EC.90)}
\end{aligned}
$$

where the first equality holds due to the law of tower property, and the last inequality holds since

$$
\begin{aligned}
\max_{x,y\in[0,1]^{d_1}}(g_1(x(1))-g_1(y(1))) &\le \max_{x_1,y_1\in[0,1]^d_1}(g_1(x_1)-g_1(y_1)) + \max_{x,y\in[0,1]^{d-d_1}}\Big(\sum_{i=2}^n g_i(x(i)) - \sum_{i=2}^n g_i(y(i))\Big) \\
&= \max_{x,y\in[0,1]^d}(g(x)-g(y)).
\end{aligned}
$$

The second and the third terms of Eq. (EC.89) have already been bounded in (EC.86) and (EC.88) respectively. We now work on the fourth term.

$$
\mathbb{E}\left[\left(\frac{\sum_{(x_k,p_k,d_k)\in\mathcal{D}_{t,1,\mathbf{j}_t(1)}}\sum_{i=2}^n g_i(x_k(i))}{|\mathcal{D}_{t,1,\mathbf{j}_t(1)}|}\right)^2\right]
$$

$$
\begin{aligned}
&= \mathbb{E}\left[\mathbb{E}\left[\left(\frac{\sum_{l\in[|\mathcal{D}_{t,1,\mathbf{j}_t(1)}|]}\sum_{i=2}^n g_i(x_{s_l}(i))}{|\mathcal{D}_{t,1,\mathbf{j}_t(1)}|}\right)^2 \middle| s_1,\cdots,s_{|\mathcal{D}_{t,1,\mathbf{j}_t(1)}|}, |\mathcal{D}_{t,1,\mathbf{j}_t(1)}|, \mathbf{j}_t(1)\right]\right] \\
&\le \mathbb{E}\left[\mathbb{E}\left[\frac{1}{4|\mathcal{D}_{t,1,\mathbf{j}_t(1)}|}\left(\max_{x,y\in[0,1]^{d-d_1}}\left(\sum_{i=2}^n g_i(x(i)) - \sum_{i=2}^n g_i(y(i))\right)\right)^2 \middle| |\mathcal{D}_{t,1,\mathbf{j}_t(1)}|, \mathbf{j}_t(1)\right]\right] \\
&\le \mathbb{E}\left[\frac{n\sum_{i=1}^n L_i^2 d_i^{k_{gi}}}{4|\mathcal{D}_{t,1,\mathbf{j}_t(1)}|}\right], \qquad \text{(EC.91)}
\end{aligned}
$$

where the equation holds due to the tower property, the first inequality follows from the facts that $x(i)$ is independent from $x(1)$ for all $i \neq 1$. Plugging Eqs. (EC.90) and (EC.91) into Eq. (EC.89), we can get

$$
\begin{aligned}
&\mathbb{E}\left[\left(\mathbb{E}[g_1(x(1))|x(1)\in\mathbf{M}^1_{\mathbf{j}_t(1)}] - \hat{a}_{t,1,\mathbf{j}_t(1)}\right)^2 1_{\{|\mathcal{D}_{t,1,\mathbf{j}_t(1)}|\ge 1\}}\right] \\
&\le 5\mathbb{E}\left[\left(\frac{\frac{1}{2}n\sum_{i=1}^n L_i^2 d_i^{k_{gi}} + \sigma^2}{|\mathcal{D}_{t,1,\mathbf{j}_t(1)}|} + 4\frac{\overline{p}^2 c_{e,1}}{\sqrt{t}} + \frac{\frac{3}{4}n\sum_{i=1}^n L_i^2 d_i^{k_{gi}} + 3\sigma^2}{t}\right) 1_{\{|\mathcal{D}_{t,1,\mathbf{j}_t(1)}|\ge 1\}}\right] \\
&\le 5\mathbb{E}\left[\left(\frac{\frac{5}{4}n\sum_{i=1}^n L_i^2 d_i^{k_{gi}} + 4\sigma^2}{|\mathcal{D}_{t,1,\mathbf{j}_t(1)}|} + 4\frac{\overline{p}^2 c_{e,1}}{\sqrt{t}}\right) 1_{\{|\mathcal{D}_{t,1,\mathbf{j}_t(1)}|\ge 1\}}\right],
\end{aligned}
$$

where in the first inequality we discard the indicator function, and the second inequality follows from $t > |\mathcal{D}_{t,1,\mathbf{j}_t(1)}|$. For convenience, we define $\max_{x,y\in[0,1]^d}(g(x)-g(y))^2 \le n\sum_{i=1}^n L_i^2 d_i^{k_{gi}} := c_{e,2}$. If $\mathcal{D}_{t,1,\mathbf{j}_t(1)} = \emptyset$, $\hat{a}_{t,1,\mathbf{j}_t(1)} = 0$ and we have

$$
\mathbb{E}\left[\left(\mathbb{E}[g_1(x(1))|x(1)\in\mathbf{M}^1_{\mathbf{j}_t(1)}] - \hat{a}_{t,1,\mathbf{j}_t(1)}\right)^2 1_{\{|\mathcal{D}_{t,1,\mathbf{j}_t(1)}|=0\}}\right] \le \max_{x\in[0,1]^{d_1}}(g_1(x))^2\mathbb{E}\left[1_{\{|\mathcal{D}_{t,1,\mathbf{j}_t(1)}|=0\}}\right].
$$

Therefore, the expectation of Eq. (EC.87) is bounded as

$$
\begin{aligned}
&\mathbb{E}\left[\left(\frac{g_0 + \sum_{i=1}^n \mathbb{E}[g_i(x(i))|x(i)\in\mathbf{M}^i_{\mathbf{j}_t(i)}] - \hat{g}_{0,t} - \sum_{i=1}^n \hat{a}_{t,i,\mathbf{j}_t(i)}}{2\hat{b}_t}\right)^2\right] \\
&\le \frac{n+1}{4\underline{b}^2}\mathbb{E}\left[\frac{3\overline{p}^2 c_{e,1}}{\sqrt{t}} + \frac{\frac{3}{4}c_{e,2}+3\sigma^2}{t} + 5\sum_{i=1}^n\left(\frac{\frac{5}{4}c_{e,2}+4\sigma^2}{|\mathcal{D}_{t,i,\mathbf{j}_t(i)}|} + \frac{4\overline{p}^2 c_{e,1}}{\sqrt{t}}\right)1_{\{|\mathcal{D}_{t,i,\mathbf{j}_t(i)}|\ge 1\}} + \max_{x\in[0,1]^{d_i}}(g_i(x))^2 1_{\{|\mathcal{D}_{t,i,\mathbf{j}_t(i)}|=0\}}\right] \\
&\le \frac{n+1}{4\underline{b}^2}\mathbb{E}\left[\frac{(20n+3)\overline{p}^2 c_{e,1}}{\sqrt{t}} + \frac{\frac{3}{4}c_{e,2}+3\sigma^2}{t} + \sum_{i=1}^n\frac{\frac{25}{4}c_{e,2}+20\sigma^2}{|\mathcal{D}_{t,i,\mathbf{j}_t(i)}|}1_{\{|\mathcal{D}_{t,i,\mathbf{j}_t(i)}|\ge 1\}} + \max_{x\in[0,1]^{d_i}}(g_i(x))^2 1_{\{|\mathcal{D}_{t,i,\mathbf{j}_t(i)}|=0\}}\right].
\end{aligned}
$$

Bound the fourth term of Eq. (EC.84). The fourth term is exactly the same as Eq. (EC.19) and $\mathbb{E}[(p_t^0 - p_t)^2|\mathbf{j}_1,\mathbf{j}_2,\ldots,\mathbf{j}_t] \le \frac{4}{\sqrt{t}}$.

Bound total regret. Based on what we have derived above, we now bound the total regret.

$$
\begin{aligned}
&\sum_{t=1}^T \mathbb{E}[r_t] \\
&\le 6|b|\sum_{t=1}^T \mathbb{E}\left[\frac{n}{4b^2}\sum_{i=1}^n\frac{L_i^2 d_i^{k_{gi}}}{M_i^{2k_{gi}}} + \frac{\max_{x\in[0,1]^d}(g(x))^2}{4\underline{b}^4}\frac{c_{e,1}}{\sqrt{t}} + \frac{n+1}{4\underline{b}^2}\left(\frac{(20n+3)\overline{p}^2 c_{e,1}}{\sqrt{t}} + \frac{\frac{3}{4}c_{e,2}+3\sigma^2}{t} + \right.\right.
\end{aligned}
$$

$$\sum_{i=1}^{n}\frac{\frac{25}{4}c_{e,2}+20\sigma^2}{|\mathcal{D}_{t,i,\mathbf{j}_t(i)}|}1_{\{|\mathcal{D}_{t,i,\mathbf{j}_t(i)}|\geq 1\}}+\max_{x\in[0,1]^{d_i}}(g_i(x))^2 1_{\{|\mathcal{D}_{t,i,\mathbf{j}_t(i)}|=0\}}\Bigg)\Bigg]$$

$$\leq 6|b|\left(\frac{n}{4\underline{b}^2}\sum_{i=1}^{n}\frac{L_i^2 d_i^{k_{gi}}}{M_i^{2k_{gi}}}T\right)+\left(\frac{\max_{x\in[0,1]^d}(g(x))^2 c_{e,1}}{4\underline{b}^4}+\frac{(n+1)(20n+3)\overline{p}^2 c_{e,1}}{4\underline{b}^2}\right)(2\sqrt{T}-1)+\frac{(n+1)(\frac{3}{4}c_{e,2}+3\sigma^2)}{4\underline{b}^2}\cdot$$

$$(\log T+1)+\sum_{i=1}^{n}\mathbb{E}\left[\left(\frac{25}{4}c_{e,2}+20\sigma^2\right)\sum_{t=1}^{T}\frac{1}{|\mathcal{D}_{t,i,\mathbf{j}_t(i)}|}1_{\{|\mathcal{D}_{t,i,\mathbf{j}_t(i)}|\geq 1\}}+\max_{x\in[0,1]^{d_i}}(g_i(x))^2 1_{\{|\mathcal{D}_{t,i,\mathbf{j}_t(i)}|=0\}}\right]$$

$$=\sum_{i=1}^{n}\mathcal{O}\left(\frac{L_i^2 T}{M_i^{2k_{gi}}}\right)+\mathcal{O}(\sqrt{T})+\mathcal{O}(\log T)+\sum_{i=1}^{n}\widetilde{\mathcal{O}}\left(M_i^{d_i}\right).$$

The equation holds due to $\sum_{t=1}^{T}1_{\{|\mathcal{D}_{t,i,\mathbf{j}_t(i)}|=0\}}\leq M_i^{d_i}$. After taking $M_i=\lceil (L_i^2T)^{\frac{1}{d_i+2k_{gi}}}\rceil$, we obtain the regret upper bound $\widetilde{\mathcal{O}}(\sqrt{T}+\sum_{i=1}^{n}(L^2T)^{\frac{d_i}{d_i+2k_{gi}}})$. Q.E.D.

D.4. Algorithm and Regret Upper Bound for Appendix D.2

THEOREM EC.4. *Suppose for all $i\in[n]$, $L_i=\Theta(1)$, $k_{gi}\in(0,1]$, and Algorithm 6 runs with $N=\lceil T^{\frac{1}{4}}\rceil$, $n_0=\lceil T^{\frac{1}{2}}\rceil$ and $M_i=\lceil T^{\frac{1}{d_i+2k_{gi}}}\rceil$. The regret of Algorithm 6 is upper bounded by*

$$\widetilde{\mathcal{O}}\left(T^{\frac{3}{4}}\vee T^{\max_{i\in[n]}\frac{d_i+k_{gi}}{d_i+2k_{gi}}}\right).$$

Proof. Note that the exploration phase of Algorithm 6 is exactly the same as Algorithm 4. Therefore, the analysis on the estimation accuracy of s_i and the regret in this phase remain the same. For the analysis in the exploration phase, the main ideas are similar, and for completeness, we provide all the necessary details. Similar to the proof of Theorem 5, for notation convenience, we denote $p_t^*=\arg\max_{p\in[\underline{p},\overline{p}]}p(f(p)+\sum_{j=1}^{n}g_j(x_t(j)))$, $i_t=\arg\min_{0\leq i\leq N-1}|p_t^*-P[i]|$, $i_t^*=\max_{0\leq i\leq N-1}P[i]\times(f(P[i])+\sum_{j=1}^{n}g_j(x_t(j)))$ and $\hat{i}_t^*=\max_{0\leq i\leq N-1}P[i]\times(f(P[i])+\sum_{j=1}^{n}\mathbb{E}[g_j(x(j))|x(j)\in\mathbf{M}^j_{\mathbf{j}_t(j)}])$.

In the exploration phase, since $g_i(\cdot)$ and $f(\cdot)$ are all bounded, the total regret is $\mathcal{O}(n_0N)$. Similar to Eq. (EC.59), we decompose the regret as follows:

$$\mathbb{E}\left[\sum_{t=n_0N+1}^{T}r_t\right]$$

$$=\sum_{t=n_0N+1}^{T}\mathbb{E}\left[\max_{p\in[\underline{p},\overline{p}]}p\big(f(p)+\sum_{j=1}^{n}g_j(x_t(j))\big)-p_t\big(f(p_t)+\sum_{j=1}^{n}g_j(x_t(j))\big)\right]$$

$$=\sum_{t=n_0N+1}^{T}\mathbb{E}\Bigg[\max_{p\in[\underline{p},\overline{p}]}p\left(f(p)+\sum_{j=1}^{n}g_j(x_t(j))\right)-\max_{0\leq i\leq N-1}P[i]\times\left(f(P[i])+\sum_{i=0}^{n}\mathbb{E}[g_j(x(j))|x(j)\in\mathbf{M}^j_{\mathbf{j}_t(j)}]\right)$$

$$+\max_{0\leq j\leq N-1}P[i]\times\left(f(P[i])+\sum_{j=1}^{n}\mathbb{E}[g_j(x(j))|x(j)\in\mathbf{M}^j_{\mathbf{j}_t(j)}]\right)-p_t\left(f(p_t)+\sum_{j=1}^{n}\mathbb{E}[g_j(x(j))|x(j)\in\mathbf{M}^j_{\mathbf{j}_t(j)}]\right)$$

$$+p_t\left(f(p_t)+\sum_{j=1}^{n}\mathbb{E}[g_j(x(j))|x(j)\in\mathbf{M}^j_{\mathbf{j}_t(j)}]\right)-p_t\left(f(p_t)+\sum_{j=1}^{n}g_j(x_t(j))\right)\Bigg]. \tag{EC.92}$$

Algorithm 6: Algorithm for SMNPE with Separable Context (`ASMNPE-SC`)

Input: time horizon T, price range $[\underline{p},\overline{p}]$, context dimension d, number of context groups n, numbers of discretized prices N, exploration parameter n_0, dimensions of each context group $d_1,\cdots,d_n$, parameters of bins $M_1,\cdots,M_n$.

Initialization:

Define $P[i]=\underline{p}+\frac{\overline{p}-\underline{p}}{N}(i+\frac{1}{2})$ for each $0\le i\le N-1$;

Initialize for all $0\le i\le N-1$: $s_i=0$ and $n_i=0$;

for $j=1,2,\cdots,n$ **do**:

  Partition $[0,1]^{d_j}$ into $M_j^{d_j}$ cubes of equal size, denoted as $\mathbf{M}_{j'}^j$ for $j'=1,2,\cdots,M_j^{d_j}$;

  Initialize $\mathcal{D}_{n_0N+1,j,j'}=\emptyset$ for each $j'\in[M_j^{d_j}]$;

end for

Main Steps:

for $t=1,2,\cdots,n_0N$ **do** `// Exploration phase`

  Calculate $i=(t \mod N)$ and charge price $p_t=P[i]$;

  Observe realized demand d_t;

  $s_i\leftarrow\frac{s_i\times n_i+d_t}{n_i+1}$ and $n_i\leftarrow n_i+1$;

end for

for $t=n_0N+1,\cdots,T$ **do** `// Exploitation phase`

  for $j=1,2,\cdots,n$ **do**:

    Observe $x_t(j)\in\mathbf{M}_{\mathbf{j}_t(j)}^j$ for some $\mathbf{j}_t(j)\in[M_j^{d_j}]$;

    If $\mathcal{D}_{t,j,\mathbf{j}_t(i)}=\emptyset$, $\hat{a}_{t,j,\mathbf{j}_t(j)}\leftarrow 0$; otherwise, $\hat{a}_{t,j,\mathbf{j}_t(j)}\leftarrow\frac{\sum_{(x_k,p_k,d_k)\in\mathcal{D}_{t,j,\mathbf{j}_t(j)}}\sum_{i=0}^{N-1}(d_k-s_i)\mathbb{I}(p_k=P[i])}{\left|\mathcal{D}_{t,j,\mathbf{j}_t(j)}\right|}$;

  end for

  Select $m_t=\arg\max_{0\le i\le N-1}P[i]\times(s_i+\sum_{j=1}^n\hat{a}_{t,j,\mathbf{j}_t(j)})$ and charge $p_t=P[m_t]$;

  Observe realized demand d_t;

  for $j=1,2,\cdots,n$ **do**:

    Update $\mathcal{D}_{t+1,j,\mathbf{j}_t(j)}\leftarrow\mathcal{D}_{t,j,\mathbf{j}_t(j)}\cup\{(x_t,p_t,d_t)\}$ and $\mathcal{D}_{t+1,j,j'}\leftarrow\mathcal{D}_{t,j,j'}$ for $j'\neq\mathbf{j}_t(j)$;

  end for

end for

Bound the first term in Eq. (EC.92). Following Eq. (EC.60), we have

$$\mathbb{E}\left[\max_{p\in[\underline{p},\overline{p}]}p\left(f(p)+\sum_{j=1}^n g_j(x_t(j))\right)-\max_{0\le i\le N-1}P[i]\times\left(f(P[i])+\sum_{j=1}^n\mathbb{E}[g_j(x(j))|x(j)\in\mathbf{M}_{\mathbf{j}_t(j)}^j]\right)\right]$$
$$\le\mathbb{E}\left[p_t^*\left(f(p_t^*)+\sum_{j=1}^n g_j(x_t(j))\right)-P[i_t]\times\left(f(P[i_t])+\sum_{j=1}^n g_j(x_t(j))\right)\right]$$

$$
\begin{aligned}
&+\mathbb{E}\left[P[i_t^*]\times\left(f(P[i_t^*])+\sum_{j=1}^n g_j(x_t(j))\right)-P[i_t^*]\times\left(f(P[i_t^*])+\sum_{j=1}^n \mathbb{E}[g_j(x(j))|x(j)\in\mathbf{M}^j_{\mathbf{j}_t(j)}]\right)\right]\\
&=\mathbb{E}\Big[p_t^* f(p_t^*)-P[i_t]f(P[i_t])\Big]+\mathbb{E}\Big[(p_t^*-P[i_t])g(x_t)\Big]+\mathbb{E}\left[P[i_t^*]\left(\sum_{j=1}^n g_j(x_t(j))-\sum_{j=1}^n \mathbb{E}[g_j(x(j))|x(j)\in\mathbf{M}^j_{\mathbf{j}_t(j)}]\right)\right]\\
&\le\mathbb{E}\Big[p_t^* f(p_t^*)-P[i_t]f(P[i_t])\Big]+\frac{\overline{p}-\underline{p}}{2N}\max_{x\in[0,1]^d}|g(x)|+\sum_{j=1}^n\frac{\overline{p}L_j d_j^{\frac{k_{gj}}{2}}}{M_j^{k_{gj}}}\\
&\le\frac{\overline{p}-\underline{p}}{2N}\left(\max_{p\in[\underline{p},\overline{p}]}|f(p)|+\overline{p}(C_0\vee\delta)+\max_{x\in[0,1]^d}|g(x)|\right)+\sum_{j=1}^n\frac{\overline{p}L_j d_j^{\frac{k_{gj}}{2}}}{M_j^{k_{gj}}},
\end{aligned}
\tag{EC.93}
$$

where the last inequality holds due to Eq. (EC.61).

Bound the second term in Eq. (EC.92). We note that

$$
\begin{aligned}
&\max_{0\le i\le N-1}P[i]\times\left(f(P[i])+\sum_{j=1}^n\mathbb{E}[g_j(x(j))|x(j)\in\mathbf{M}^j_{\mathbf{j}_t(j)}]\right)-p_t\left(f(p_t)+\sum_{j=1}^n\mathbb{E}[g_j(x(j))|x(j)\in\mathbf{M}^j_{\mathbf{j}_t(j)}]\right)\\
&=\left(\max_{0\le i\le N-1}P[i]\times\left(f(P[i])+\sum_{j=1}^n\mathbb{E}[g_j(x(j))|x(j)\in\mathbf{M}^j_{\mathbf{j}_t(j)}]\right)-\max_{0\le 1\le N-1}P[i]\times\left(s_i+\sum_{j=1}^n\hat{a}_{t,j,\mathbf{j}_t(j)}\right)\right)\\
&\quad+\left(\max_{0\le 1\le N-1}P[i]\times\left(s_i+\sum_{j=1}^n\hat{a}_{t,j,\mathbf{j}_t(j)}\right)-p_t\left(f(p_t)+\sum_{j=1}^n\mathbb{E}[g_j(x(j))|x(j)\in\mathbf{M}^j_{\mathbf{j}_t(i)}]\right)\right)\\
&\le 2\overline{p}\max_{0\le i\le N-1}|f(P[i])-s_i|+2\overline{p}\sum_{j=1}^n\left|\hat{a}_{t,j,\mathbf{j}_t(j)}-\mathbb{E}[g_j(x(j))|x(j)\in\mathbf{M}^j_{\mathbf{j}_t(j)}]\right|.
\end{aligned}
\tag{EC.94}
$$

To bound the second term in Eq. (EC.94), we focus on the case of $j=1$ and the analysis for $2\le j\le n$ can be obtained similarly. If $|\mathcal{D}_{t,1,\mathbf{j}_t(1)}|\ge 1$, by the definition of $\hat{a}_{t,1,\mathbf{j}_t(1)}$, we have

$$
\begin{aligned}
&\left|\hat{a}_{t,1,\mathbf{j}_t(1)}-\mathbb{E}[g_1(x(1))|x(1)\in\mathbf{M}^1_{\mathbf{j}_t(1)}]\right|\\
&\le\max_{0\le i\le N-1}|f(P[i])-s_i|+\left|\frac{\sum_{(x_k,p_k,d_k)\in\mathcal{D}_{t,1,\mathbf{j}_t(1)}}(\sum_{j=1}^n g_j(x_k(j))-\mathbb{E}[g_1(x(1))|x(1)\in\mathbf{M}^1_{\mathbf{j}_t(1)}]+\varepsilon_k)}{|\mathcal{D}_{t,1,\mathbf{j}_t(1)}|}\right|\\
&\le\max_{0\le i\le N-1}|f(P[i])-s_i|+\left|\frac{\sum_{(x_k,p_k,d_k)\in\mathcal{D}_{t,1,\mathbf{j}_t(1)}}(g_1(x_k(1))-\mathbb{E}[g_1(x(1))|x(1)\in\mathbf{M}^1_{\mathbf{j}_t(1)}])}{|\mathcal{D}_{t,1,\mathbf{j}_t(1)}|}\right|\\
&\quad+\left|\frac{\sum_{(x_k,p_k,d_k)\in\mathcal{D}_{t,1,\mathbf{j}_t(1)}}\sum_{j=2}^n g_j(x_k(j))+\varepsilon_k}{|\mathcal{D}_{t,1,\mathbf{j}_t(1)}|}\right|.
\end{aligned}
\tag{EC.95}
$$

For the second term in the RHS of Eq. (EC.95), we bound its expectation as follows:

$$
\begin{aligned}
&\mathbb{E}\left[\left|\frac{\sum_{(x_k,p_k,d_k)\in\mathcal{D}_{t,1,\mathbf{j}_t(1)}}(g_1(x_k(1))-\mathbb{E}[g_1(x(1))|x(1)\in\mathbf{M}^1_{\mathbf{j}_t(1)}])}{|\mathcal{D}_{t,1,\mathbf{j}_t(1)}|}\right|\right]\\
&=\mathbb{E}\left[\mathbb{E}\left[\left|\frac{\sum_{(x_k,p_k,d_k)\in\mathcal{D}_{t,1,\mathbf{j}_t(1)}}(g_1(x_k(1))-\mathbb{E}[g_1(x(1))|x(1)\in\mathbf{M}^1_{\mathbf{j}_t(1)}])}{|\mathcal{D}_{t,1,\mathbf{j}_t(1)}|}\right|\Big|\mathbf{j}_{n_0N+1}(1),\cdots,\mathbf{j}_t(1)\right]\right]
\end{aligned}
$$

$$
\begin{aligned}
&\le \mathbb{E}\left[\int_0^\infty 2\exp\left(-\frac{8|\mathcal{D}_{t,1,\mathbf{j}_t(1)}|\epsilon^2}{\max_{x_1,x_2\in[0,1]^{d_1}}(g_1(x_1)-g_1(x_2))^2}\right)\mathrm{d}\epsilon\right]\\
&= \mathbb{E}\left[\frac{\sqrt{\pi}\max_{x_1,x_2\in[0,1]^{d_1}}|g_1(x_1)-g_1(x_2)|}{2\sqrt{2|\mathcal{D}_{t,1,\mathbf{j}_t(1)}|}}\right]\\
&\le \mathbb{E}\left[\sqrt{\frac{\pi\sigma_1'^2}{2|\mathcal{D}_{t,1,\mathbf{j}t(1)}|}}\right]. \qquad \text{(EC.96)}
\end{aligned}
$$

In the above equations, the first inequality holds since conditional on $\mathbf{j}_{n_0N+1}(1),\cdots,\mathbf{j}_t(1)$, $\{g_1(x_k(1))-\mathbb{E}[g_1(x(1))|x(1)\in\mathbf{M}^1_{\mathbf{j}_t(1)}]:(x_k,p_k,d_k)\in\mathcal{D}_{t,1,\mathbf{j}_t(1)}\}$ are i.i.d. sub-Gaussian r.v.'s with variance proxy $\frac{\max_{x_1,x_2\in[0,1]^{d_1}}(g_1(x_1)-g_1(x_2))^2}{4}$, and from Hoeffding's inequality, we have

$$
\begin{aligned}
&\mathbb{P}\left(\left|\frac{\sum_{(x_k,p_k,d_k)\in\mathcal{D}_{t,1,\mathbf{j}_t(1)}}(g_1(x_k(1))-\mathbb{E}[g_1(x(1))|x(1)\in\mathbf{M}^1_{\mathbf{j}_t(1)}])}{|\mathcal{D}_{t,1,\mathbf{j}_t(1)}|}\right|>\epsilon\middle|\mathbf{j}_{n_0N+1}(1),\cdots,\mathbf{j}_t(1)\right)\\
&\le 2\exp\left(-\frac{8|\mathcal{D}_{t,1,\mathbf{j}_t(1)}|\epsilon^2}{\max_{x_1,x_2\in[0,1]^{d_1}}(g_1(x_1)-g_1(x_2))^2}\right).
\end{aligned}
$$

The second inequality of (EC.96) holds due to $\max_{x_1,x_2\in[0,1]^{d_1}}|g_1(x_1)-g_1(x_2)|\le\max_{x_1,x_2\in[0,1]^d}|g(x_1)-g(x_2)|\le 2\sigma_1'$ (recall the definition of $\sigma_1'^2=\frac{1}{4}\max_{x_1,x_2\in[0,1]^d}(g(x_1)-g(x_2))^2+\sigma^2$ from Appendix C.1.).

For the third term of Eq. (EC.95), we bound its expectation as follows:

$$
\begin{aligned}
&\mathbb{E}\left[\left|\frac{\sum_{(x_k,p_k,d_k)\in\mathcal{D}_{t,1,\mathbf{j}_t(1)}}\sum_{j=2}^n g_j(x_k(j))+\varepsilon_k}{|\mathcal{D}_{t,1,\mathbf{j}_t(1)}|}\right|\right]\\
&= \mathbb{E}\left[\mathbb{E}\left[\left|\frac{\sum_{(x_k,p_k,d_k)\in\mathcal{D}_{t,1,\mathbf{j}_t(1)}}\sum_{j=2}^n g_j(x_k(j))+\varepsilon_k}{|\mathcal{D}_{t,1,\mathbf{j}_t(1)}|}\right|\middle|\mathbf{j}_{n_0N+1}(1),\cdots,\mathbf{j}_t(1)\right]\right]\\
&\le \mathbb{E}\left[\int_0^\infty 2\exp\left(-\frac{2|\mathcal{D}_{t,1,\mathbf{j}t(1)}|\mu^2}{\sigma_1'^2}\right)\mathrm{d}\mu\right]\\
&= \mathbb{E}\left[\sqrt{\frac{\pi\sigma_1'^2}{2|\mathcal{D}_{t,1,\mathbf{j}t(1)}|}}\right]. \qquad \text{(EC.97)}
\end{aligned}
$$

The inequality in Eq. (EC.97) holds because $x(j)$ is independent of $x(1)$ for all $j\neq 1$, and thus given $\mathbf{j}_{n_0N+1}(1),\cdots,\mathbf{j}_t(1)$, the distributions of $x(j)$ for $j\neq 1$ are not influenced, and $\sum_{j=2}^n g_j(x_k(j))+\varepsilon_k$ is zero-mean sub-Gaussian random variable with variance proxy $\frac{\max_{x_1,x_2\in[0,1]^d}(\sum_{j=2}^n g_j(x_1(j))-\sum_{j=2}^n g_j(x_2(j)))^2}{4}+\sigma^2\le\sigma_1'^2$.

Therefore, plugging in Eqs. (EC.96) and (EC.97) into (EC.95), the expectation of the second term of Eq. (EC.94) can be bounded as

$$
2\overline{p}\sum_{j=1}^n\mathbb{E}\left[\left|\hat{a}_{t,j,\mathbf{j}_t(j)}-\mathbb{E}[g_j(x(j))|x(j)\in\mathbf{M}^j_{\mathbf{j}_t(i)}]\right|\right]
$$

$$
\begin{aligned}
&= 2\overline{p}\sum_{j=1}^{n}\mathbb{E}\left[\left|\hat{a}_{t,j,\mathbf{j}_t(j)} - \mathbb{E}[g_j(x(j))|x(j)\in\mathbf{M}^j_{\mathbf{j}_t(i)}]\right| 1_{|\mathcal{D}_{t,1,\mathbf{j}t(1)}|\geq 1} + \left|\hat{a}_{t,j,\mathbf{j}_t(j)} - \mathbb{E}[g_j(x(j))|x(j)\in\mathbf{M}^j_{\mathbf{j}_t(i)}]\right| 1_{|\mathcal{D}_{t,1,\mathbf{j}t(1)}|=0}\right]\\
&\leq 2\overline{p}\sum_{j=1}^{n}\mathbb{E}\left[\left(\max_{0\leq i\leq N-1}|f(P[i])-s_i| + 2\sqrt{\frac{\pi\sigma_1'^2}{2|\mathcal{D}_{t,j,\mathbf{j}t(j)}|}}\right) 1_{|\mathcal{D}_{t,j,\mathbf{j}t(j)}|\geq 1} + \max_{x\in[0,1]^{d_j}}|g_j(x)| 1_{|\mathcal{D}_{t,j,\mathbf{j}t(j)}|=0}\right]\\
&\leq 2\overline{p}n\mathbb{E}\left[\max_{0\leq i\leq N-1}|f(P[i])-s_i|\right] + 2\overline{p}\sum_{j=1}^{n}\mathbb{E}\left[2\sqrt{\frac{\pi\sigma_1'^2}{2|\mathcal{D}_{t,j,\mathbf{j}t(j)}|}} 1_{|\mathcal{D}_{t,j,\mathbf{j}t(j)}|\geq 1}\right]\\
&\qquad + 2\overline{p}\max_{j\in[n],x\in[0,1]^{d_j}}|g_j(x)|\sum_{j=1}^{n}\mathbb{E}[1_{|\mathcal{D}_{t,j,\mathbf{j}t(j)}|=0}]. \qquad \text{(EC.98)}
\end{aligned}
$$

It only remains to bound $\mathbb{E}\left[\max_{0\leq i\leq N-1}|f(P[i])-s_i|\right]$. Similar to Appendix C.1, for any $\epsilon>0$, denote $\mathcal{A}:=\{\max_{0\leq i\leq N-1}|f(P[i])-s_i|\leq\sigma_1'\sqrt{\frac{2}{n_0}\log\frac{2}{\epsilon}}\}$. Eq. (EC.68) still holds here, and then we have $\mathbb{P}(\mathcal{A})\geq 1-N\epsilon$. Hence,

$$
\begin{aligned}
\mathbb{E}\left[\max_{0\leq i\leq N-1}|f(P[i])-s_i|\right] &= \mathbb{E}\left[\max_{0\leq i\leq N-1}|f(P[i])-s_i|\Big|\mathcal{A}\right]\mathbb{P}(\mathcal{A}) + \mathcal{O}(1)\mathbb{P}(\mathcal{A}^c)\\
&\leq \sigma_1'\sqrt{\frac{2}{n_0}\log\frac{2}{\epsilon}} + \mathcal{O}(N\epsilon). \qquad \text{(EC.99)}
\end{aligned}
$$

Bound the third term in Eq. (EC.92). For this term, we have

$$
p_t\left(f(p_t)+\sum_{j=1}^{n}\mathbb{E}[g_j(x(j))|x(j)\in\mathbf{M}^j_{\mathbf{j}_t(j)}]\right) - p_t\left(f(p_t)+\sum_{j=1}^{n}g_j(x_t(j)\right) \leq \overline{p}\sum_{j=1}^{n}\frac{\overline{p}L_j d_j^{\frac{k_{gj}}{2}}}{M_j^{k_{gj}}}. \qquad \text{(EC.100)}
$$

Bound the total regret. Combining Eqs. (EC.92), (EC.93), (EC.98), (EC.100) and (EC.99), we have

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{t=1}^{T}r_t\right] = \mathbb{E}\left[\sum_{t=1}^{n_0N}r_t\right] + \mathbb{E}\left[\sum_{t=n_0N+1}^{T}r_t\right]\\
&\leq \mathcal{O}(n_0N)\frac{\overline{p}-\underline{p}}{2}\left(\max_{p\in[\underline{p},\overline{p}]}|f(p)| + \overline{p}(C_0\vee\delta) + \max_{x\in[0,1]^d}|g(x)|\right)\frac{T-n_0N}{N} + 2\sum_{j=1}^{n}\frac{\overline{p}L_j d_j^{\frac{k_{gj}}{2}}}{M_j^{k_{gj}}}(T-n_0N)\\
&\qquad + 2\overline{p}n\sigma_1'\sqrt{\frac{2}{n_0}\log\frac{2}{\epsilon}}(T-n_0N) + 2\overline{p}\sum_{j=1}^{n}\sum_{t=n_0N+1}^{T}\mathbb{E}\left[2\sqrt{\frac{\pi\sigma_1'^2}{2|\mathcal{D}_{t,j,\mathbf{j}t(j)}|}} 1_{|\mathcal{D}_{t,j,\mathbf{j}t(j)}|\geq 1}\right] + \mathcal{O}(TN\epsilon)\\
&\qquad + 2\overline{p}\max_{j\in[n],x\in[0,1]^{d_j}}|g_j(x)|\sum_{j=1}^{n}\sum_{t=n_0N+1}^{T}\mathbb{E}[1_{|\mathcal{D}_{t,j,\mathbf{j}t(j)}|=0}]\\
&= \mathcal{O}(n_0N) + \mathcal{O}\left(\frac{T}{N}\right) + \sum_{j=1}^{n}\mathcal{O}\left(\frac{T}{M_j^{k_{gj}}}\right) + \mathcal{O}\left(\frac{T}{\sqrt{n_0}}\sqrt{\log\frac{2}{\epsilon}}\right) + \sum_{j=1}^{n}\mathcal{O}\left(\sqrt{TM_j^{d_j}}\right) + \sum_{j=1}^{n}\mathcal{O}\left(M_j^{d_j}\right) + \mathcal{O}(TN\epsilon)\\
&= \widetilde{\mathcal{O}}(T^{\frac{3}{4}}) + \sum_{j=1}^{n}\widetilde{\mathcal{O}}\left(T^{\frac{d_j+k_{gj}}{d_j+2k_{gj}}}\right). \qquad \text{(EC.101)}
\end{aligned}
$$

The first identity of Eq. (EC.101) follows from that

$$\sum_{t=n_0N+1}^{T}\frac{1}{\sqrt{|D_{t,j,\mathbf{j}_t(j)}|}}1_{|\mathcal{D}_{t,j,\mathbf{j}t(j)}|\geq 1}=\sum_{s=1}^{M_j^{d_j}}\sum_{t=n_0N+1}^{T}\frac{1}{\sqrt{|D_{t,j,s}|}}1_{\{|\mathcal{D}_{t,j,s}|\geq 1,\mathbf{j}_t(j)=s\}}$$
$$=\sum_{s=1}^{M_j^{d_j}}\sum_{s=1}^{|\mathcal{D}_{t,j,s}|}\frac{1}{\sqrt{s}}\leq 2\sum_{s=1}^{M_j^{d_j}}\sqrt{|D_{t,j,s}|}\leq 2\sqrt{TM_j^{d_j}},$$

where the last inequality follows Cauchy-Schwarz inequality and $\sum_{s=1}^{M_j^{d_j}}|D_{t,j,s}|\leq T$, and

$$\sum_{t=n_0N+1}^{T}1_{\{|\mathcal{D}_{t,j,\mathbf{j}t(j)}|=0\}}=\sum_{s=1}^{M_j^{d_j}}\sum_{t=n_0N+1}^{T}1_{\{|\mathcal{D}_{t,j,s}|=0,\mathbf{j}t(j)=s\}}\leq M_j^{d_j}.$$

In the second identity of Eq. (EC.101), we let $N=\lceil T^{\frac{1}{4}}\rceil$, $n_0=\lceil T^{\frac{1}{2}}\rceil$, and $M_i=\lceil T^{\frac{1}{d_i+2k_{gj}}}\rceil$ and $\epsilon=T^{-2}$. This completes the proof of Theorem EC.4. Q.E.D.

Appendix E. Linear Greedy Algorithm in Section 6

Algorithm 7: Linear Greedy Algorithm

Input: price range $[\underline{p},\overline{p}]$, bounds on the price coefficient $\underline{b}$ and $\overline{b}$

Initialization:

Initialize $\hat{a}_1=0$, $\hat{b}_1=\frac{\underline{b}+\overline{b}}{2}$, $\hat{c}_1=0$;

Main Steps:

for $t=1,2,\cdots,T$ **do**

  Observe x_t;

  Set unconstrained greedy price: $p_t^u\leftarrow-\frac{\hat{a}_t+\hat{c}_t^\top x_t}{2\hat{b}_t}$;

  Project greedy price: $p_t^g\leftarrow\text{Proj}(p_t^u,[\underline{p},\overline{p}])$;

  Set price $p_t\leftarrow p_t^g$;

  Observe realized demand d_t;

  Update $(\hat{a}_{t+1},\hat{b}_{t+1},\hat{c}_{t+1})\leftarrow\arg\min_{\alpha,\beta\in[\underline{b},\overline{b}],\gamma}\sum_{s=1}^{T}(d_s-\alpha-\beta p_s-\gamma^\top x_s)^2$;

end for