# Learning to sell a focal-ancillary combination

Hanzhao Wang Xiaocheng Li Kalyan Talluri

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

**Abstract**

A number of products are sold in the following sequence: first, an essential focal product is presented, and if purchased, discretionary ancillary products are subsequently offered. Examples include airline flights followed by seat selection or insurance, hotel rooms followed by spa treatments, and electronics such as tablets paired with accessories like electronic pencils. The firm has to decide on a sale format—whether to sell them in sequence unbundled, or together as a bundle—and how to price them. Since the discretionary ancillary is considered by the customer only after the purchase of the focal product, the sale strategy chosen by the firm creates an information and learning dependency between the products: for instance, offering only a bundle would preclude learning customers' valuation for the focal and ancillary products individually. In this paper, we study learning strategies for such focal and ancillary item combinations under the following scenarios: (a) pure unbundling to all customers, (b) a personalized mechanism, where, depending on some observed features of the customers, the two products are presented and priced as a bundle or in sequence, (c) the firm can change the selling strategy only once over the entire time horizon. We design pricing algorithms for all three scenarios assuming a demand distribution which depends on a linear function of customer features, and show that their regrets are upper bounded by $O\big(d^2\sqrt{T}\log T\big)$, where $d$ denotes the dimension of the customer features and $T$ represents the time horizon. Furthermore, we conduct numerical experiments with both simulated and real-world data to demonstrate the effectiveness of the proposed algorithms.

**Keywords.** Bundle pricing, Airline ancillary pricing, Policy learning, Dynamic pricing

## 1 Introduction

When consumers purchase an essential focal product or service, firms often offer additional discretionary items that complement or enhance the focal item. Typically, the focal item constitutes a necessary or essential purchase, whereas ancillary offerings are discretionary add-ons presented at a subsequent purchase stage. These ancillary items are not critical to the primary use of the focal item; thus, consumers' initial purchase decisions for the focal product can be independent of subsequent discretionary offerings. Consider the following illustrative examples:

- Retail: Purchasing a tablet or computer (essential focal item) as opposed to accessories such as electronic pencils or keyboards (discretionary ancillary items).
- Airline: Buying a flight ticket or a bundled flight with luggage (essential focal items) versus selecting seat upgrades, travel insurance, or in-flight food service (discretionary ancillary items).
- Hotel: Reserving a hotel room (essential focal item) compared to opting for additional services like spa treatments or parking (discretionary ancillary items).

These examples underscore the significant role ancillary revenue plays in increasing the seller's revenue in many industries. Thus, studying the sale mechanisms for such a focal-ancillary combination is of great interest.

It should be clear that the sale of a focal-ancillary combination has some idiosyncracies; specifically, there is a certain order and dependence on how they are sold. The firm can choose to sell them either bundled together or in a sequential manner with two distinct prices, where the consumer is offered the ancillary only after purchasing the focal product.

If the firm has a good estimate of the parameters of the demand model, it may be able to decide on bundled or unbundled offers up-front, but if it has to learn the parameters, then the choice of the sale mechanism would influence the learning of the parameters. For instance, if the firm were to follow a strict bundling strategy from the beginning, it may learn the valuations for the bundle as a whole, but will be unable to determine if an unbundling strategy is better as it does not learn the valuations for the focal and ancillary products separately. Even if the firm were to follow an unbundling strategy, the price of the focal item would influence both the number and types of samples it can collect on the ancillary product, and hence the rate at which it can learn the ancillary product valuations. Thus, the exploration-exploitation trade-off is more complicated than in the existing literature on dynamic pricing with learning: both the exploration and the exploitation are dependent on each step of a sequence of decisions (i.e., selling mechanism decision, focal item's pricing, and ancillary item's pricing) and significantly more challenging to analyze.

In this paper, we study learning strategies for such focal and ancillary item combinations under three distinct scenarios assuming a demand whose distribution depends on a linear function of customer features. First, a *pure unbundling* setting, which requires the seller to determine the prices for both the focal and ancillary products independently; Second, a *personalized mechanism* where the seller has the flexibility to decide the selling mechanism and the pricing based on the customer's covariates. Finally, a *one-switch* scenario where the firm is restricted to a uniform selling strategy for all customers but has the option to make a single strategic switch. Such a scenario can be due to strategic considerations, such as the impracticality of frequently altering selling strategies; instead, it exclusively offers bundled selections to all customers. These scenarios reflect different levels of flexibility that a seller might have in setting prices and choosing sales strategies in real-life contexts.

For each scenario, we design learning and pricing algorithms (and the switching time for the one-switch case) with regret upper bounded by $O(d^2\sqrt{T}\log T)$, where $d$ is the dimension of customer features and $T$ is the time horizon. The algorithms have to be carefully designed to explore the unknown demand functions because of the sequential structure of the purchase process. We further numerically validate these algorithms on both simulated and real-world datasets. The contributions of the paper are as follows:

1. For the pure unbundling case, we give an algorithm that sets prices for the focal and ancillary separately while learning the parameters of the demand model. The key challenge in designing this algorithm is to handle the "uninformative covariates" with respect to the ancillary item. Specifically, due to the sequential structure of purchases, we cannot obtain information on ancillary valuations if the customer does not purchase the focal product. This is in contrast to a single-item dynamic pricing with learning where the item is shown to every customer. To counter this, we develop a *Lower Confidence Bound* (LCB) pricing strategy to encourage the purchase of the focal item and show, with no distributional assumptions on the covariates of the arriving customers, a regret of $O\left(d^2\sqrt{T}\log T\right)$. Moreover, under an i.i.d. assumption on the distribution of the covariates, we improve the bound to $O\left(d^2\log^2(T)/q^*\right)$ where $q^*$ represents the expected purchase probability of the focal product under optimal pricing with the true parameters.

2. When the mechanism can be personalized to each customer based on their observed covariates (i.e., offering either a bundle or the focal-ancillary sequence), the fact that the firm can make an ancillary sale only if it sells the focal item again creates difficulties in learning and setting a personalized strategy. We extend the LCB pricing strategy to this personalized setting using ideas from *Information Directed Sampling (IDS)* due to [44]. Our algorithm achieves an $O\left(d^2\sqrt{T}\log T\right)$ regret again without making any distributional assumptions on the covariates.

3. For the one-switch case, our algorithm employs a "safe-switching" condition, which ensures that the switch to the bundling strategy only occurs when it is "safe" enough, avoiding the risk of switching to a bundling strategy that might be suboptimal. The algorithm achieves a regret of $O\left(d^2\log T\sqrt{\frac{T}{q^*}}\right)$ under an i.i.d assumption on the distribution of the covariates.

4. We conduct numerical experiments based on both a simulated dataset as well as a real-world dataset from an airline company. We compare our algorithms with benchmark algorithms and derive managerial insights from the analysis of the real-world airline dataset, particularly concerning the decision to bundle or not.

The rest of the paper is organized as follows: In §2 we survey the relevant literature, in §3 we set out the model and the notations, and in the three subsequent sections analyze our learning algorithms for the three different scenarios—the pure unbundling case in §4, the personalized mechanism in §5 and the one-switch case in §6. In §7, we present our numerical experiments, including analyses of the results on simulated data and real-world data from the airline industry. Finally, we conclude the paper with potential future directions in §8.

# 2 Literature Review

In this section, we survey the literature relevant to our paper.

## 2.1 Pricing of Bundles

The economic motivation for bundling and its use for price discrimination of a heterogenous set of consumers is well studied in both theoretical and empirical works [49, 2, 41, 25, 38, 16].

As we mentioned earlier, the sale of focal and ancillary products is distinct because of the asymmetric dependence of the latter on the former. Add-on pricing and the economics of lock-in, hold-up and giveaways have been examined in the Industrial Organization literature from various angles [46, 19, 21]. The analysis of Gomes and Tirole [24] shows the monopolist's tendency to extract rents from the ancillary product is tempered by the probability of missed sales.

In the Revenue Management literature, the major focus is on pricing of either the bundled products [54, 37, 1, 14] or pricing for sequentially selling the focal and ancillary [18, 53, 3, 9], or data-driven methods for bundling design [20], or bundling design with disposal [40]. A few works study other mechanisms: assortment optimization at checkout [15], loot box design [12] and robust optimization for pricing [11]. As far as we know, we are the first paper to consider learning for the joint selling strategy and pricing of dependent products.

## 2.2 Bandits and Dynamic Pricing with Learning

Our techniques and regret frameworks come from the learning literature and we collect some relevant references here.

**Cascading Bandits.** In the cascading bandits model, the agent offers $K$ items to the user, the user examines the list from top to bottom and chooses the first attractive item, a learning variant of the cascade model where the agent needs to learn the $K$ most attractive items. This bandit setting was first introduced by Kveton et al. [32] and then extended in Li et al. [36], Kveton et al. [33]. There are important differences between our model and the cascading bandits model: (i) In cascading bandits, the user examines a product only when they reject all previous ones, but for us, the ancillary item is considered only when they buy the focal item; (ii) The reward function in cascading bandits is the probability that the customer will accept at least one product, but in our model the objective is to maximize revenue; (iii) The decision variable in cascading bandits is the set of recommended items and their positions, while the decision variables for us are the prices as well as whether to bundle.

**Meta Bandits.** Another related bandits problem is meta bandits [4, 10, 31], where several tasks arrive in sequence. The unknown parameters in each task are sampled from the same (meta) prior distribution and thus the (meta) posterior distribution for the later tasks can be updated based on the observations from previous tasks. Hong et al. [26] further extend the above setting to the generalized hierarchical Bayesian bandits which allow more complicated arrival patterns of the tasks and Bastani et al. [7] consider the corresponding meta-pricing problem. Because the information/estimations from our sub-tasks are not shared (e.g., the observed demands of the focal item cannot estimate the demand function of the ancillary), deploying a meta-learning algorithm with a single-item pricing algorithm is not applicable to our problem.

**Dynamic Pricing with Learning.** In this stream on learning for feature-based dynamic pricing [17, 28, 6], the authors usually consider a single (focal) product. However, our demand model is for both the focal and ancillary items, in addition to a more complicated sale mechanism. Chen et al. [13] consider the revenue maximization problem with unknown parameters and a cascading model for customer behavior. However, the decision variables in this work are the rankings of offered items, and the customer is assumed to purchase at most one item. Another work, Song and Li [48], is based on a similar customer purchase process model as ours and also focuses on optimal dynamic pricing. However, their focus is on a situation with finite inventory, full information (the demand functions are known), and no customer covariates; thus the dynamic pricing is driven by the state of inventory, while the dynamic pricing in our work is to trade-off learning the demand and earning by choosing the optimal selling strategy and the optimal pricing.

One recent work by Simchi-Levi et al. [47] also studies the online pricing problem for focal and ancillary items with unknown demands; However, the key differences are as follows:

- In our model, we allow the seller to choose both the selling strategy (bundling or unbundling) and the prices. The model by Simchi-Levi et al. [47] restricts the seller to an unbundling strategy, while allowing customers who have already purchased the focal item to buy only the ancillary item.
- Our model allows personalized selling strategies and pricing by leveraging customer covariates (which can be non-i.i.d. across time) through a parametric demand model. Their model assumes a universal non-parametric demand model for all customers, resulting in identical optimal prices.
- Our model allows prices to be set in continuous intervals. Their model assumes the prices are chosen from finite sets.
- Our model focuses on one focal item and one ancillary item. Their model considers multiple products with a size constraint on the number of discounted ancillary products.

# 3 Model

A seller has two products to sell, a focal product and an ancillary product, over a horizon of $T$ time periods. The focal product is the main revenue driver, while the ancillary product is a closely related product that boosts the total earnings of the seller. The distinguishing feature of this combination is that the ancillary product can be purchased/used only after the focal product has been purchased. From a revenue management perspective, there are two decisions to make (for the seller): (i) whether to *bundle* the two items, and (ii) how to *price*.

If they are bundled, the seller will post one single price for the bundle, and the customer will decide to purchase either the bundle (both of the products) or nothing. If they are sold unbundled, the two products have their own individual prices, and the customer will first consider the focal product. If it is purchased, then the customer will consider the ancillary product.

We denote the chosen selling strategy in each period $t = 1, ..., T$ as $A_t \in \{b, u\}$, where $b$ denotes bundling and $u$ denotes unbundling. We represent the prices under unbundling as $[p_f, p_a]_u$ (subscript $u$) where $p_f, p_a$ are the focal and ancillary product prices, and bundling as $p_b$ (subscript $b$). The prices may vary over different time periods, indexed by a subscript $t$.

We assume $p_f, p_a, p_b \in [\underline{p}, \bar{p}]$, although our analysis can be easily generalized to the case when the feasible sets are different for the three prices. We assume there are no capacity constraints and both products have zero marginal costs, so that revenue equals profit.

Customer context is a $d$-dimensional feature vector $x_t \in \mathcal{X} \subseteq \mathbb{R}^d$ that describes the selling environment and the customer features in time period $t$, where $\mathcal{X}$ is the set of possible contexts. We assume customers do not form rational expectations and are not strategic on their time of arrival. In contrast to the literature on feature-based dynamic pricing [28, 6], we also consider the case without assumptions on the statistical structure of the $x_t$'s.

Denote the demand vector of a customer at a certain period as $\left([d_f, d_a]_u, d_b\right)$, where $d_f, d_a, d_b \in \{0, 1\}$, with 1 indicating a sale and 0, no sale. When the seller chooses unbundling, $d_b = 0$, and, in addition, if $d_f = 0$, then $d_a = 0$. Likewise, $d_f = d_a = 0$ when the seller chooses bundling.

The demand model for a single time period for an observation of $x$ and prices $[p_f, p_a]_u, p_b$ is as follows: When the selling strategy is unbundling, the utility of the arriving customer for the focal product is $x^\top \theta_f^* + \epsilon_f$ and that for the ancillary product $x^\top \theta_a^* + \epsilon_a$ where the random variables $\epsilon_f$ and $\epsilon_a$ capture utility unexplained by the feature vector $x$.

For focal and ancillary products, there are two fixed unknown coefficient vectors $\theta_f^*, \theta_a^*$ from a bounded set $\Theta \subset \mathbb{R}^d$ that we assume does not change over time. Without loss of generality, we also assume the price coefficient to be 1 by rescaling the utility accordingly. Thus the demand in the unbundled setting in a single period is

$$[d_f, d_a]_u = \begin{cases} [1,1], & \text{if } x^\top \theta_f^* + \epsilon_f \geq p_f \text{ and } x^\top \theta_a^* + \epsilon_a \geq p_a, \\ [1,0], & \text{if } x^\top \theta_f^* + \epsilon_f \geq p_f \text{ and } x^\top \theta_a^* + \epsilon_a < p_a, \\ [0,0], & \text{if } x^\top \theta_f^* + \epsilon_f < p_f, \end{cases}$$

(and the bundling demand $d_b = 0$ given the selling strategy is unbundling). We assume $\epsilon_f$ and $\epsilon_a$ are independent.

This model of consumer decision-making is standard in the economics literature (see for instance §5 of Gomes and Tirole [24]) and assumes the firm does not reveal the ancillary price before the focal item purchase (economists refer to this as "drip" pricing) and we assume the consumer does not form rational expectations of the ancillary price; i.e., the consideration set is not affected by the combined price.

Alternatively, if the selling strategy is bundling, the utility of the arriving customer for the bundle

is modeled by $x^\top \theta_b^* + \epsilon_b$ where $\theta_b^* := \theta_f^* + \theta_a^*$ denotes the coefficient vector for the product bundle, and $\epsilon_b$ is the utility shock for the bundle. Here we assume $\epsilon_b = \epsilon_f + \epsilon_a$ (the results in this paper continue to hold if $\epsilon_b$ takes a different distribution than $\epsilon_f + \epsilon_a$). We assume that there are no complementary interactions between the two products, meaning that customers treat them independently and add their values together. To illustrate, consider the purchase of a flight ticket (the focal product) and luggage (the ancillary product). The utility derived from the luggage depends on whether the customer needs to carry many items and is independent of the flight ticket itself (conditional on the customer features). This assumption of additive utility is commonly made in both the Economics literature [2, 38, 25] and the Operations Management literature [37, 14].

In other words, we are making the following assumption:

**Assumption 1** (Utility)**.** *We assume that the utility of the bundle is equal to the sum of the utilities of the focal and ancillary product and $\epsilon_f$ and $\epsilon_a$ are independent.*

The demand under the bundling sale mechanism hence is

$$d_b = \begin{cases} 1, & \text{if } x^\top \theta_b^* + \epsilon_b \geq p_b, \\ 0, & \text{otherwise} \end{cases}$$

with $d_f = d_a = 0$ (as the selling strategy is bundling).

Let $F_{\epsilon_f}$ and $F_{\epsilon_a}$ denote the cumulative distribution function of $\epsilon_f$ and $\epsilon_a$. Let $F_{\epsilon_b}$ denote the cumulative distribution function of $\epsilon_b$ (which as we mentioned needs not to be the convolution of $F_{\epsilon_f}$ and $F_{\epsilon_a}$). Then the probability model under the two mechanisms for the realized demand with feature vector $x$ is as follows. For the unbundling case:

$$[d_f, d_a]_u = \begin{cases} [1,1], & \text{w.p. } \left(1 - F_{\epsilon_f}(p_f - x^\top \theta_f^*)\right)\left(1 - F_{\epsilon_a}(p_a - x^\top \theta_a^*)\right), \\ [1,0], & \text{w.p. } \left(1 - F_{\epsilon_f}(p_f - x^\top \theta_f^*)\right) F_{\epsilon_a}(p_a - x^\top \theta_a^*), \\ [0,0], & \text{w.p. } F_{\epsilon_f}(p_f - x^\top \theta_f^*). \end{cases}$$

and for the bundling case:

$$d_b = \begin{cases} 1, & \text{w.p. } 1 - F_{\epsilon_b}(p_b - x^\top \theta_b^*), \\ 0, & \text{w.p. } F_{\epsilon_b}(p_b - x^\top \theta_b^*). \end{cases}$$

We now turn to the multi-period setting. First, we consider the random utility shock as idiosyncratic noises independent over time. For simplicity, we also assume they are identically distributed according to the law of $F_{\epsilon_f}, F_{\epsilon_a}$ and $F_{\epsilon_b}$. We use $(\epsilon_{t,f}, \epsilon_{t,a}, \epsilon_{t,b})$ to denote the utility shock at time $t$. Accordingly, we use $[p_{t,f}, p_{t,a}]_u, p_{t,b}$ and $[d_{t,f}, d_{t,a}]_u, d_{t,b}$ to denote the price and realized demand at time $t$. In this way, all the above model formulations can be cast into a multi-period form by replacing the price, demand, feature, and utility shock with their time-varying counterparts. The prices and the selling strategy (to bundle or not) are decisions to make. The seller collects observations of customer demands to learn the unknown fixed coefficient vector $(\theta_f^*, \theta_a^*)$ so as to better understand the demand mechanism and optimize the prices and the selling strategy.

## 3.1 Assumptions

We summarize our boundedness structure on the parameter $\theta$ ($= \theta_f, \theta_a$ or $\theta_b$), the feature vector $x$, and the prices by the following assumption.

**Assumption 2** (Boundedness)**.** *We assume*

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

*(b) For any covariate vector $x \in \mathcal{X}$, $\|x\|_2 \leq 1$.*

*(c) For focal, ancillary, and bundle products, the prices are restricted to a range $[\underline{p}, \bar{p}]$ under all possible $x$ and $\theta$. Furthermore, $\forall \theta \in \Theta, x \in \mathcal{X}$, the optimal prices to maximize expected revenue fall in the range $[\underline{p}, \bar{p}]$.*

Next, the following assumption describes the structure we impose on the random utility shocks $\epsilon_f, \epsilon_a$ and $\epsilon_b$. This type of assumption, together with the distribution knowledge, is often treated as part of the customer choice model and has been standard in the feature-based dynamic pricing literature [27, 28, 6, 55]; it is satisfied by several common distribution families, including the normal, uniform, logistic, and exponential distributions. The emphasis of this line of literature, as in our paper, is on the interplay of parameter learning and pricing decision-making; the assumption on the underlying distribution ensures a moderate speed of learning.

**Assumption 3** (Distribution)**.** *We assume the random utility shock distributions $F_{\epsilon_f}, F_{\epsilon_a}, F_{\epsilon_b}$ are continuously differentiable functions with density function $f_{\epsilon_f}, f_{\epsilon_a}, f_{\epsilon_b}$. Moreover, we assume that the $F_{(\cdot)}(v)$'s are strictly increasing in the interval $\left[\underline{p} - \bar{\theta}, \bar{p} + \bar{\theta}\right]$, and that both $F_{(\cdot)}(v)$ and $1 - F_{(\cdot)}(v)$ are log-concave in $v$ (we note that many common distributions satisfy log-concavity), and there exist constants $B$ and $B'$ such that*

$$B = \max_{v \in \left[\underline{p} - \bar{\theta}, \bar{p} + \bar{\theta}\right]} \max\left\{f_{\epsilon_f}(v), f_{\epsilon_a}(v), f_{\epsilon_b}(v)\right\},$$

*and*

$$B' = \max_{v \in \left[\underline{p} - \bar{\theta}, \bar{p} + \bar{\theta}\right]} \max\left\{|f'_{\epsilon_f}(v)|, |f'_{\epsilon_a}(v)|, |f'_{\epsilon_b}(v)|\right\}.$$

Based on the distribution functions, the following two quantities will be critical in determining the learning rate of the problem:

$$\nu = \inf_{|v| \leq \bar{p} + \bar{\theta}} \left\{ \min_{i \in \{f,a,b\}} \{-\log'' F_{\epsilon_i}(v), -\log''(1 - F_{\epsilon_i}(v))\} \right\},$$
$$\mu = \sup_{|v| \leq \bar{p} + \bar{\theta}} \left\{ \max_{i \in \{f,a,b\}} \{-\log' F_{\epsilon_i}(v), -\log'(1 - F_{\epsilon_i}(v))\} \right\}.$$

Under Assumption 2 and 3, we know that $\nu$ and $\mu$ are well-defined. As we will see in the later sections, they govern the strong convexity and smoothness of the log-likelihood function.

We will use the capital letter $X$ for random variables, and $x$ for realized values. For a positive definite matrix $\Sigma$, we define the $\Sigma$-norm of a vector $x$ as,

$$\|x\|_\Sigma := \sqrt{x^\top \Sigma x}.$$

## 3.2 Selling Strategy with Known Parameters

In this section, we collect results on the optimal pricing and expected demand for a single period when the true parameters are known. For brevity, we denote the deterministic part of the product valuation (given a feature $x$) as $v_f = x^\top \theta_f$, $v_a = x^\top \theta_a$, and $v_b = x^\top \theta_b = v_f + v_a$.

The seller's expected single-period revenue under the bundling strategy is given by

$$r_b(p_b; v_b) := p_b \left(1 - F_{\epsilon_b}(p_b - v_b)\right).$$

and under the unbundling strategy,

$$r_u(p_f, p_a; v_f, v_a) \coloneqq (p_f + p_a(1 - F_{\epsilon_a}(p_a - v_a))) \left(1 - F_{\epsilon_f}(p_f - v_f)\right).$$

For a cumulative distribution function $F_{\epsilon_i}(v)$, $i \in \{f, a, b\}$, define

$$g_i(v) \coloneqq v + \phi_i^{-1}(-v),$$

where $\phi_i(v) = v - (1 - F_{\epsilon_i}(v))/f_{\epsilon_i}(v)$ is the virtual valuation function ($\phi_i^{-1}$ is well-defined under our assumptions). The revenue-maximizing pricing rule can be defined using the functions $g_i(\cdot)$, $i \in \{f, a, b\}$ as follows.

**Lemma 1.** *[Optimal pricing for single item; Javanmard and Nazerzadeh [28]] The revenue-maximizing price is given by*

$$p_b^*(v_b) = g_b(v_b)$$

*for the bundling product and*

$$p_a^*(v_a) = g_a(v_a)$$

*for the ancillary product.*

The focal product's optimal price, however, has to be derived differently as it depends on the expected revenue of the ancillary product $r_a$. In the following, we write the optimal price for the focal product as a function of $v_f$ and $r_a$.

**Lemma 2.** *[Optimal pricing for focal item] The revenue-maximizing price for the focal product when $r_a$ is the expected revenue of the ancillary product, is given by*

$$p_f^*(v_f, r_a) = g_f(v_f + r_a) - r_a.$$

*Proof.* Proof of Lemma 2.

Given $v_f$ and $r_a$, we can write the expected revenue for unbundling as below:

$$(p_f + r_a)(1 - F_{\epsilon_f}(p_f - v_f)).$$

Thus, by the first order condition, the optimal $p_f^*(v_f, r_a)$ should satisfy

$$p_f^* = \frac{1 - F_{\epsilon_f}(p_f^* - v_f)}{f_{\epsilon_f}(p_f^* - v_f)} - r_a.$$

Rearranging it we get

$$\phi_f(p_f^* - v_f) + v_f + r_a = 0.$$

Thus with the definition of $g_f(\cdot)$, the optimal pricing function is

$$p_f^*(v_f, r_a) = g_f(v_f + r_a) - r_a.$$

When the seller prices the products according to the optimal functions in Lemma 1 and 2, we can define the optimal expected revenue as a function of the underlying valuation $v$, as $r_i(v) \coloneqq g_i(v)\left(1 - F_{\epsilon_i}(g_i(v) - v)\right), i \in \{f, a, b\}$.

Thus the optimal revenue function under the bundling strategy is

$$r_b^*(v_b) = r_b(v_b)$$

while the optimal revenue function for the unbundling strategy is

$$r_u^*(v_f, v_a) = r_f(v_f + r_a(v_a)).$$

The optimal selling strategy, for the single-period case, with known parameters, hence is simply given as $b$ if $r_b^*(v_b) > r_u^*(v_f, v_a)$ and $u$ otherwise.

### 3.3 Performance Measure

We consider an online policy $\pi$ that does not know the parameters $\theta_f^*, \theta_a^*, \theta_b^*$ and has to learn them. We denote $R^\pi$ as the (cumulative) revenue from $\pi$ and measure its performance by the *regret*:

$$\text{Regret}^\pi := r^* - \mathbb{E}[R^\pi]$$

where $r^*$ refers to the optimal expected revenue when we know the parameters and can optimize both the selling strategy (when available) as well as the price(s) for each arriving customer. We remark that both the revenue functions are defined for an arbitrary fixed sequence of the covariates $x_t$, and thus the expectation is over the utility shocks $\epsilon_f, \epsilon_a$ and $\epsilon_b$ and, potentially, randomness in the policy $\pi$.

It is important to define $r^*$ appropriately in different settings to align with the specific constraints of each setting and avoid vacuous bounds. For instance, in some pure unbundling pricing cases, where $\pi$ can only apply some pricing policy with the unbundling selling strategy, the bundling strategy (with the optimal pricing) can be more profitable. In these cases, using the bundling strategy with its optimal pricing as the benchmark would unfairly skew the analysis and result in a regret linear in $T$ for any unbundling-based pricing policy, including the optimal one (with the unbundling strategy). Therefore, we leave the definition of regret in an open form here and will specify it later under the context of the three settings.

We develop algorithms and derive regret upper bounds to evaluate the performance. In presenting the upper bound, we treat parameters such as $B, B', \nu, \mu, \bar{\theta}$ (pertaining to the known utility shock distributions) as constants and emphasize the dependence of regret on parameters like $d$, $T$, and other related quantities.

## 4 Pure Unbundling

We first study learning under a *pure unbundling* setting where only the unbundling strategy is available. This is a common selling setting in the airline and hotel industry. Our analysis and discussion of the pure unbundling setting aim to highlight technical and operational differences between two-product sequential pricing and single-product pricing. Also, it prepares us for the more complicated setting where both the bundling and unbundling strategies are available.

Mathematically, the setting fixes the strategy to be unbundling $A_t = u$ for all $t = 1, ..., T$. Prices are personalized, so the seller observes $x_t$ and sets the prices of both the focal and ancillary products, $p_{t,f}$ and $p_{t,a}$.

The regret of a certain policy/algorithm $\pi$ is

$$\text{Reg}_T^\pi(\boldsymbol{X}) := \sum_{t=1}^T r_u^*(x_t^\top \theta_f^*, x_t^\top \theta_a^*) - \mathbb{E}\left[\sum_{t=1}^T r_u(P_{t,f}, P_{t,a}; x_t^\top \theta_f^*, x_t^\top \theta_a^*)\right] \tag{1}$$

where $\boldsymbol{X} = \{x_1, ..., x_T\}$ encapsulates the covariates of all time periods, and the expectation is taken with respect to the possibly random prices $P_{t,f}, P_{t,a}$ induced by the policy $\pi$. The functions $r_u^*$ and $r_u$ are

defined in §3.2. Specifically, $r_u^*$ denotes the optimal achievable expected revenue given the knowledge of the true parameters $\theta_f^*$ and $\theta_a^*$, and thus it represents the clairvoyant optimal revenue under the unbundling strategy. The function $r_u$ denotes the expected revenue given the prices and the customer valuations.

## 4.1 Pricing Algorithm

We first give some intuition into our pricing algorithm. Recall that the purchases happen in a sequential manner. In other words, if the customer at a certain time $t$ chooses not to purchase the focal product, the seller would not know the customer's decision for the ancillary product. Technically, it means that the decision history of the focal product determines the observations/samples collection for the ancillary product. This is a key difference from single-product dynamic pricing and learning problems. The sequential structure means if the seller overprices the focal product too often, it affects the sample collection rate (equivalently, the learning rate) of the ancillary demand model. In addition, the samples used for estimating the ancillary demand model consist of only a subset of all the samples, resulting in a degeneracy in the covariance structure, leading to non-convexity when using maximum likelihood estimation (MLE).

Algorithm 1 describes our lower confidence bound (LCB)-based pricing policy for the unbundling setting, where the algorithm applies a price lower than the estimated optimal one of the focal product to encourage purchases of the focal product, and thus improves the learning rate of the ancillary demand model. At each time $t$, the algorithm estimates the parameters by regularized maximum likelihood estimation.

Let data set $\mathcal{D} = \{(p_t, x_t, d_t), t \in \mathcal{T}\}$ denote a set of available observations with price $p_t$ (for either focal, ancillary, or bundle products), customer feature vector $x_t$, utility shock distribution $F_\epsilon$ ($\epsilon = \epsilon_f, \epsilon_a$ or $\epsilon_b$ depending on the given data set), and realized demand $d_t$. Let $\mathcal{T}$ denote the index set of the observations.

Denote the data sets $\mathcal{D}_{t,f} = \{(p_{t',f}, x_{t'}, d_{t',f}), t' = 1, ..., t-1\}$ and $\mathcal{D}_{t,a} = \{(p_{t',a}, x_{t'}, d_{t',a}), t' \in \mathcal{T}_{t,a}\}$ where the time index set

$$\mathcal{T}_{t,a} \coloneqq \{t' | d_{t',f} = 1, t' = 1, ..., t-1\}$$

denotes the time periods when a focal purchase takes place.

The log-likelihood function is

$$LL(\theta; \mathcal{D}) \coloneqq \sum_{t \in \mathcal{T}, d_t = 1} \log(1 - F_\epsilon(p_t - x_t^\top \theta)) + \sum_{t \in \mathcal{T}, d_t = 0} \log(F_\epsilon(p_t - x_t^\top \theta)).$$

For time $t \geq 2$, the regularized MLEs for the focal product and for the ancillary product are given by

$$\hat{\theta}_{t,f} \coloneqq \arg\min_{\theta \in \Theta} \ -LL(\theta; \mathcal{D}_{t,f}) + \lambda \nu \|\theta\|_2^2,$$

$$\hat{\theta}_{t,a} \coloneqq \arg\min_{\theta \in \Theta} \ -LL(\theta; \mathcal{D}_{t,a}) + \lambda \nu \|\theta\|_2^2,$$

where $\lambda$ is the regularization parameter to be specified and $\nu$ is defined earlier following Assumption 3. We should note that it is possible to use different $\lambda$ values for the focal item and the ancillary item, and we only use the same $\lambda$ for both here for the sake of simplicity in notation.

Next, we construct a "data-driven" confidence set following the standard treatment in the linear bandits literature (Lattimore and Szepesvári [35]). For a regularization parameter $\lambda > 0$, define the

following design matrices for each period

$$\Sigma_{t,f} \coloneqq \sum_{t'=1}^{t} x_{t'} x_{t'}^\top + \lambda I, \quad \Sigma_{t,a} \coloneqq \sum_{t' \in \mathcal{T}_{t,a}} x_{t'} x_{t'}^\top + \lambda I,$$

where $I$ is a $d$-dimensional identity matrix.

Define the confidence sets

$$\Theta_{t,f} \coloneqq \left\{ \theta \in \Theta : \left\| \hat{\theta}_{t,f} - \theta \right\|_{\Sigma_{t-1,f}} \leq \beta(\Sigma_{t-1,f}) \right\}, \quad \Theta_{t,a} \coloneqq \left\{ \theta \in \Theta : \left\| \hat{\theta}_{t,a} - \theta \right\|_{\Sigma_{t-1,a}} \leq \beta(\Sigma_{t-1,a}) \right\},$$

where the function $\beta$ is defined by

$$\beta(\Sigma) \coloneqq 2\sqrt{\lambda}\bar{\theta} + \frac{2\mu}{\nu} \sqrt{2 \log T + \log \left( \frac{\det \Sigma}{\lambda^d} \right)}$$

with $\bar{\theta}$ as given in Assumption 2 and $\mu, \nu$ following Assumption 3. The construction of the confidence sets in our model follows the principles of self-normalized bounds [42]. Specifically, both the norm of the error term $\| \cdot \|_\Sigma$ and the magnitude of the confidence set $\beta(\Sigma)$ depend on the design matrix $\Sigma$ of the history data, which means they are normalized by the data themselves. Further discussions on the property of the estimators and justifications on the confidence set construction are referred to Online Appendix A.

We remark that the regularized MLE (with the squared Euclidean norm) has two advantages over the standard MLE: First, the Euclidean norm regularization ensures well-definedness of the optimization problem when there are not enough observations, say for the first few periods, and can further establish the error bound for the estimators (see Online Appendix A). The MLE without such regularization or with other forms of regularization, such as the $L1$ norm, cannot inherently offer this advantageous property. Second, for the parameter estimation of the ancillary product, the regularization and the construction of the confidence set based on the sample design matrix $\Sigma_{t,\cdot}$ both help to overcome the "selection bias" induced by the focal product's purchases. Specifically, we recall that the samples used for estimating the ancillary parameters are those when the focal purchase takes place. It may happen that a subset of the customers (with certain covariate structure) tend to purchase the focal product more often than others which will twist the sample space used for estimating the ancillary parameters. Again, the intuition here is aligned with the literature on linear bandits [35] where the selection bias is induced by the arm plays.

The algorithm uses the confidence sets to construct LCB and UCB for the customer valuations of the focal and ancillary products. The seller sets the prices based on the optimal pricing functions (in §3.2) as if the corresponding valuations are true. Specifically, the ancillary product's price comes directly from the estimator $\hat{\theta}_{t,a}$ and the focal product's price adopts a lower confidence principle. The rationale is that when the seller has some uncertainty about the underlying model, a price lower than the optimal price will lead to a sufficient number of focal purchases for learning the ancillary product's parameter at an appropriate rate. This strategy diverges from the UCB-style pricing algorithms in literature [52, 47]. In the following regret analysis, we demonstrate the effectiveness of this pricing technique for achieving the regret upper bounds of Algorithm 1.

## 4.2 Regret Analysis

Let

$$\eta = B + \bar{p}B'$$

**Algorithm 1** LCB Pricing for Pure Unbundling

**Input:** Regularization parameter $\lambda$.
**for** $t = 1, ..., T$ **do**
Observe $x_t$ and compute the estimators $\hat{\theta}_{t,f}$ and $\hat{\theta}_{t,a}$ and their confidence sets $\Theta_{t,f}$ and $\Theta_{t,a}$. When there are no observations available, i.e., $\mathcal{D}_{t,f} = \emptyset$ or $\mathcal{D}_{t,a} = \emptyset$, set $\hat{\theta}_{t,f/a} = 0$ and $\Theta_{t,f/a} = \Theta$.
Compute the UCB and LCB of the customer valuation:

$$\underline{v}_{t,f} = \min_{\theta \in \Theta_{t,f}} x_t^\top \theta, \quad \bar{v}_{t,a} = \max_{\theta \in \Theta_{t,a}} x_t^\top \theta.$$

Set the price by

$$p_{t,f} = p_f^* \left( \underline{v}_{t,f}, r_a^*(\bar{v}_{t,a}) \right), \tag{2}$$

$$p_{t,a} = p_a^*(x_t^\top \hat{\theta}_{t,a}), \tag{3}$$

where the optimal pricing functions $p_f^*$ and $p_a^*$ are given in §3.2.
**end for**

and

$$\bar{\beta} = 2\bar{\theta} + \frac{2\mu}{\nu} \sqrt{2 \log(T) + d \log\left(\frac{d+T}{d}\right)}.$$

**Theorem 1.** *Under Assumption 1, 2, 3 and with the regularization parameter $\lambda = 1$, the regret of the Algorithm 1 for the pure unbundling setting is bounded by*

$$2\bar{p} + 6\sqrt{2}\bar{\beta} \sqrt{dT \log\left(\frac{d+T}{d}\right)} + 2d\eta\bar{\beta}^2 \log\left(\frac{d+T}{d}\right) = O\left(d^2\sqrt{T}\log T\right).$$

Here we set $\lambda = 1$ only for the sake of simplicity in notation and can be set as other reasonable values like $\lambda = 0.1$. The above theorem shows the regret upper bound for Algorithm 1 under pure unbundling setting is $O\left(d^2\sqrt{T}\log T\right)$. The dependence on the dimension of the covariates aligns with the regret upper bound of the UCB-based algorithm in dynamic pricing with covariates [52]. However, under similar assumptions, Wang et al. [52] demonstrates that the certainty equivalent pricing policy (or greedy ILS by Keskin and Zeevi [29], Qiang and Bayati [43]) for a single item (i.e., the focal item only) can lead to a regret upper bound $O(\log^2 T)$. With this different order in $T$, it would be desirable to show a matching regret lower bound in the sequential purchase setting as it would be very informative on the tightness of the Theorem 1, but unfortunately, this looks like a very difficult problem because of the complex dependency structure in our sequential purchase problem. We leave the development of such technology as an interesting future research problem.

In the following, we will elaborate on a few key steps in deriving the regret bounds and postpone all proofs to Online Appendix B.

#### 4.2.1 Lower bound pricing

First, we define "good" events

$$\mathcal{E}_f = \left\{ \theta_f^* \in \Theta_{t,f},\ t = 1, ..., T \right\},$$

$$\mathcal{E}_a = \left\{ \theta_a^* \in \Theta_{t,a},\ t = 1, ..., T \right\},$$

under which the confidence sets cover the corresponding true coefficients throughout the entire horizon. We obtain the following probability bounds by applying a union bound to the result in Corollary 1.

**Lemma 3.** *We have*

$$\mathbb{P}(\mathcal{E}_f \cap \mathcal{E}_a) \geq 1 - 2/T.$$

Conditional on the good event $\mathcal{E}_f \cap \mathcal{E}_a$, the focal price $p_{t,f}$ used in Algorithm 1 will be a lower bound of the clairvoyant optimal price, regardless of the underlying $x_t$. Therefore, the LCB price will encourage more focal purchases (equivalently, ancillary observations) and thus achieve a faster learning rate for ancillary item's demand than the optimal prices with knowing the true parameters in a pathwise manner. In this way, the LCB price balances the exploration of the ancillary item's demand and the exploitation of the revenue from selling the focal item.

**Lemma 4** (Lower Bound Pricing)**.** *Conditional on the event* $\mathcal{E}_f \cap \mathcal{E}_a$, $\forall t = 1, ..., T$,

$$p_{t,f} \leq p_f^*(x_t^\top \theta_f^*, r_a^*(x_t^\top \theta_a^*))$$

*holds with probability* 1.

#### 4.2.2 Single-period regret bound

In this section, we present an upper bound for the single-period regret in Lemma 5. By aggregating the single-period regrets we can derive Theorem 1. Essentially, the single-period regret compares the expected revenue obtained by Algorithm 1 against the optimal revenue at time $t$,

$$\text{Reg}_t := r_u^*\left(x_t^\top \theta_f^*, x_t^\top \theta_a^*\right) - r_u\left(p_{t,f}, p_{t,a}; x_t^\top \theta_f^*, x_t^\top \theta_a^*\right).$$

Our analysis diverges from that of the single-item case [28, 52] due to the sequential purchase structure, where in our case there are uninformative covariates for the ancillary item (the main difficulty). Specifically, under the single-item setting (and for the focal item in our case), every $x_t$, coupled with its corresponding purchase/no-purchase observation, can reduce the estimation error of the unknown parameter. However, the sequential purchase structure means that if the focal item is not purchased, we cannot observe the ancillary item's purchase/no-purchase, and thus, cannot reduce the estimation error for the ancillary item, even though we still suffer regret due to the suboptimal pricing.

Lemma 5 addresses this challenge by factoring in the purchase probability of the focal product. We call it the ex-post bound since it depends on the ex-post focal item's purchase. Specifically, Lemma 5 takes into account the binary demand structure of the problem where

$$q_{t,f} = 1 - F_{\epsilon_f}\left(p_{t,f} - x_t^\top \theta_f^*\right)$$

represents the focal product's purchase probability given covariates $x_t$. Note $q_{t,f} = \mathbb{E}_{\epsilon_f}[\mathbb{1}_{\{d_{t,f}=1\}}|x_t, p_t]$, the expected single-period regret

$$\mathbb{E}[\text{Reg}_t|x_t, p_t] \leq \mathbb{E}_{\epsilon_f}[\mathbb{1}_{\{d_{t,f}=1\}}|x_t, p_t]\left(6\bar{\beta}\|x_t\|_{\Sigma_{t-1,a}^{-1}} + \eta\bar{\beta}^2\|x_t\|^2_{\Sigma_{t-1,a}^{-1}}\right)$$

is positive only when there is a focal purchase (i.e., $\mathbb{1}_{\{d_{t,f}=1\}} = 1$) by Lemma 5. Thus, there are no more uninformative covariates: the regret from the ancillary item only occurs when the focal item has been purchased. In addition, conditional on the ex-post observation of a focal purchase, the regret contains a linear term that explains the effect of ancillary estimation error on the focal's revenue, and a quadratic term that accounts for the effect of the error on the ancillary's revenue.

The proof of Theorem 1 builds on the Lemma 5. Specifically, the lemma represents the regret with an interplay between $x_t$ and $\Sigma_{t,a}$'s. Intuitively, if certain $x_t$ causes a large single-period regret, it will be an effective observation in terms of reducing $\Sigma_{t,a}$. The Elliptical Potential Lemma (Lemma 21, see its origin from Lai and Wei [34]) precisely characterizes such a trade-off.

**Lemma 5** (Ex-post Single-Period Regret Bound). *Under $\mathcal{E}_f \cap \mathcal{E}_a$, we have the following bound for the single-period regret*

$$\mathrm{Reg}_t \le q_{t,f} \cdot \left(6\bar{\beta}\|x_t\|_{\Sigma_{t-1,a}^{-1}} + \eta\bar{\beta}^2\|x_t\|^2_{\Sigma_{t-1,a}^{-1}}\right),$$

*where $\eta = B + \bar{p}B'$ and $\bar{\beta} = 2\sqrt{\lambda}\theta + \frac{2\mu}{\nu}\sqrt{2\log(T) + d\log\left(\frac{d\lambda+T}{d\lambda}\right)}$. Here the prices $p_{t,f}$ and $p_{t,a}$ are given by Algorithm 1.*

We refer to the detailed proofs for this lemma to Online Appendix B.

## 4.3 Regret Bound under I.I.D. Covariates

Theorem 1 does not make any assumptions on the generation of the covariates; in this subsection, we show that if we assume i.i.d. generated covariates (as assumed in Javanmard and Nazerzadeh [28], Xu and Wang [55]), the dependence on $T$ in the regret upper bound of Algorithm 1 can be improved from $\sqrt{T}$ (Theorem 1) to $\log T$ (Theorem 2).

**Assumption 4.** *Feature vectors $X_t$'s are generated independently from a fixed unknown distribution $F_X$ with bounded support $\mathcal{X}$ in $\mathbb{R}^d$.*

**Theorem 2.** *Under Assumption 1, 2, 3, and 4, and with the regularization parameter $\lambda = 1$, the regret of the Algorithm 1 for the pure unbundling setting is bounded by*

$$2\bar{p} + \frac{288d\eta\bar{\beta}^2}{q^*}\log\left(\frac{d+T+1}{d}\right) = O\left(\frac{d^2\log^2 T}{q^*}\right).$$

*Here $q^*$ represents the expected purchasing probability of the focal product under an optimal pricing policy that knows the true parameters, i.e.,*

$$q^* := \mathbb{E}_X\left[1 - F_{\epsilon_f}\left(p_f^*(X^\top\theta_f^*, r_a^*(X^\top\theta_a^*)) - X^\top\theta_f^*\right)\right].$$

Compared to Theorem 1, which makes no assumption on the covariates $x_t$'s, we can view Theorem 2 as a problem-dependent bound with a dependence on the underlying distribution through the parameter $q^*$. Notably, the i.i.d. assumption reduces the regret's dependence on $T$ from $\sqrt{T}$ to $\log T$. In the context of single-item pricing with i.i.d. covariates, Javanmard and Nazerzadeh [28] prove a regret lower bound of $\Omega(d\log T)$. Thus, Theorem 2 matches the order of $T$ (up to an additional logarithmic factor), but also involves an extra factor $q^*$.

The parameter $q^*$ represents the rate/probability of the focal item's purchase under the optimal pricing policy which knows the true parameters. When the true parameters are unknown, $q^*$ governs the learning rate of the ancillary product in our pricing policy. An important aspect to explore is whether the dependence on $1/q^*$ can be improved, or if it is the optimal achievable learning rate, by offering a lower regret bound. However, because of the complex nature of the sequential purchase problem, establishing such a lower regret bound remains difficult and is a challenging but valuable direction for future research.

### 4.3.1 Regret Analysis

The proof of Theorem 2 is similar to that of proving Theorem 1 except the method for bounding the single-period regret as shown in Lemma 6. We call it the ex-ante bound since it does not depend on the information of $q_{t,f}$, i.e., the focal product's purchase probability given covariates.

**Lemma 6** (Ex-ante Single-Period Regret Bound)**.** *Under $\mathcal{E}_f \cap \mathcal{E}_a$, we have the following bound for the single-period regret*

$$\mathrm{Reg}_t \le 144\eta\bar{\beta}^2 \cdot \|x_t\|^2_{\Sigma^{-1}_{t-1,a}},$$

*where $\eta = B + \bar{p}B'$ and $\bar{\beta} = 2\sqrt{\lambda}\bar{\theta} + \frac{2\mu}{\nu}\sqrt{2\log(T) + d\log\left(\frac{d\lambda+T}{d\lambda}\right)}$. The prices $p_{t,f}$ and $p_{t,a}$ are given by Algorithm 1.*

Note that the quantity $\|x_t\|_{\Sigma^{-1}_{t-1,a}}$ can be viewed as a proxy of our estimation error for both product valuations (the focal's estimation will always be more accurate than the ancillary's). Lemma 6 upper bounds the single-period regret by the quadratic of the error in valuation estimation. When $\|x_t\|_{\Sigma^{-1}_{t-1,a}} \le 1$, the linear term in Lemma 5 will be larger than the quadratic term in Lemma 6, the effect of which could be offset by the extra term $q_{t,f}$; it is difficult to say which bound is tighter.

In the proof of Theorem 2, Lemma 6 itself cannot effectively control the regret caused by the uninformative covariates for each single period, as accomplished in Lemma 5, due to the absence of $q_{t,f}$. However, under Assumption 4, the i.i.d. $X_t$'s with LCB pricing can bound the probability of encountering uninformative covariates and thus the expected regret. Specifically, Lemma 4 ensures that, with high probability, the expected purchase probability of the focal item (i.e., the informative covariates) at each time $t$ is at least $q^*$, as defined in Theorem 2. Intuitively, the expected interval between observations of ancillary items is at most $1/q^*$. Considering this, we can partition the time horizon $T$ into approximately $q^*T$ intervals, each of length $1/q^*$. At the start of each interval, if we assume we can gather a new observation for the ancillary item, set a price, and maintain this price throughout the interval, then the expected single-period regret remains the same in each interval. Thus, we can aggregate the regret occurring at the start of these intervals—treating it akin to a single-item pricing problem without uninformative covariates for the ancillary item. By multiplying this aggregate regret by the length $1/q^*$, we arrive at the regret bound articulated in Theorem 2.

While $q^*$ in Theorem 2 can be replaced by the almost sure value

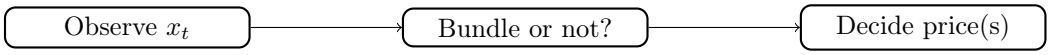
$$\underline{q}_f = \min_{x\in\mathcal{X},\, p\in[\underline{p},\bar{p}]}\Big(1 - F_{\epsilon_f}\big(p - x^\top\theta^*_f\big)\Big),$$

which allows for an analogous analysis of a certainty equivalent pricing strategy, such a substitution leads to a vacuous bound of $O\big(d^2\log^2 T/\underline{q}_f\big)$ when $\underline{q}_f \to 0$. More discussions on this issue are provided in Online Appendix F2.

# 5 Personalized Mechanism

In this section we consider the formulation where the seller has the option of choosing the selling strategy in each time period. Such a personalized bundling strategy can be effectively utilized by e-commerce platforms, which offer personalized bundle recommendations at discounted prices. In the context of physical retail stores, the modification of selling strategies can be implemented by establishing specific conditions, such as time (e.g., lunch special bundles at restaurants) and geography (e.g., exclusive bundles in certain areas).

Specifically, the selling strategy can be dependent on the covariates $x_t$. As shown in the diagram below, at each time $t$, the seller first observes the covariates $x_t$ and then decides the selling strategy and the price(s).



Under this setting, the regret of a policy/algorithm $\pi$ can be defined by

$$\mathrm{Reg}_T^{\pi}(\boldsymbol{X}) \coloneqq \sum_{t=1}^{T} \max\left\{r_u^*\left(x_t^\top \theta_f^*, x_t^\top \theta_a^*\right), r_b^*\left(x_t^\top \theta_b^*\right)\right\} - \mathbb{E}\left[\sum_{t=1}^{T} R_t^{\pi}\right]$$

where $\boldsymbol{X}$ encapsulates all the covariates as before and the expectation is taken with respect to utility shocks and the potential randomness in policy $\pi$. $R_t^{\pi}$ is the random revenue obtained by $\pi$ at time $t$. The functions $r_u^*$ and $r_b^*$ are the optimal revenue functions defined in §3.2. Since under this setting the seller can offer personalized selling strategies, we compare the online cumulative revenue against a benchmark which knows the underlying parameters and picks the more profitable one among the two selling strategies given the covariates $x_t$.

## 5.1 Algorithm

To describe the algorithm, we first introduce some notations. Recall that $A_t$ denotes the selling strategy at time $t$. Denote the sets of time periods

$$\begin{aligned}
\mathcal{T}_{t,f} &= \{t' | A_{t'} = u, t' = 1, ..., t-1\}, \\
\mathcal{T}_{t,a} &= \{t' | A_{t'} = u, d_{t',f} = 1, t' = 1, ..., t-1\}, \\
\mathcal{T}_{t,b} &= \{t' | A_{t'} = b, t' = 1, ..., t-1\}.
\end{aligned}$$

Specifically, the time periods that the strategy of bundling or unbundling is applied contribute to the corresponding parameter estimation. As before, the effective samples for the estimation of the ancillary product require a realized purchase of the focal item. Then, we can construct the same regularized MLE estimators as before by

$$\begin{aligned}
\hat{\theta}_{t,f} &\coloneqq \arg\min_{\theta \in \Theta} \; -LL(\theta; \{(p_{t',f}, x_t, d_{t',f}), t' \in \mathcal{T}_{t,f}\}) + \lambda\nu\|\theta\|_2^2, \\
\hat{\theta}_{t,a} &\coloneqq \arg\min_{\theta \in \Theta} \; -LL(\theta; \{(p_{t',a}, x_t, d_{t',a}), t' \in \mathcal{T}_{t,a}\}) + \lambda\nu\|\theta\|_2^2, \\
\hat{\theta}_{t,b} &\coloneqq \arg\min_{\theta \in \Theta} \; -LL(\theta; \{(p_{t',b}, x_t, d_{t',b}), t' \in \mathcal{T}_{t,b}\}) + \lambda\nu\|\theta\|_2^2,
\end{aligned}$$

where $\lambda$ is the regularization parameter.

The design matrices are defined by

$$\Sigma_{t,f} = \sum_{t' \in \mathcal{T}_{t,f}} x_{t'} x_{t'}^\top + \lambda I, \quad \Sigma_{t,a} = \sum_{t' \in \mathcal{T}_{t,a}} x_{t'} x_{t'}^\top + \lambda I, \quad \Sigma_{t,b} = \sum_{t' \in \mathcal{T}_{t,b}} x_{t'} x_{t'}^\top + \lambda I.$$

Then we have the following confidence sets for the estimators

$$\begin{aligned}
\Theta_{t,f} &\coloneqq \left\{\theta \in \Theta : \left\|\hat{\theta}_{t,f} - \theta\right\|_{\Sigma_{t-1,f}} \le \beta(\Sigma_{t-1,f})\right\}, \\
\Theta_{t,a} &\coloneqq \left\{\theta \in \Theta : \left\|\hat{\theta}_{t,a} - \theta\right\|_{\Sigma_{t-1,a}} \le \beta(\Sigma_{t-1,a})\right\}, \\
\Theta_{t,b} &\coloneqq \left\{\theta \in \Theta : \left\|\hat{\theta}_{t,b} - \theta\right\|_{\Sigma_{t-1,b}} \le \beta(\Sigma_{t-1,b})\right\},
\end{aligned}$$

where the function $\beta$ is defined by $\beta(\Sigma) = 2\sqrt{\lambda}\bar{\theta} + \frac{2\mu}{\nu}\sqrt{2\log T + \log\left(\frac{\det \Sigma}{\lambda^d}\right)}$ with $\bar{\theta}$, $\mu$ and $\nu$ defined in Assumption 2 and 3.

When both selling strategies are allowed, an alternative way to estimate the coefficient vector of the

ancillary product is

$$\hat{\theta}'_{t,a} \coloneqq \hat{\theta}_{t,b} - \hat{\theta}_{t,f}. \tag{4}$$

The challenge in algorithm design still comes from dealing with uninformative covariates, as discussed in §4. This issue is pertinent not only to the pricing aspect but also to the identification of the optimal selling strategy. For example, directly applying a UCB-style policy for choosing the selling strategy—selecting the option with the higher maximized optimal revenue within its confidence set—might be stuck at choosing the (potentially suboptimal) unbundling strategy. Specifically, when applying the unbundling strategy, the presence of the uninformative covariates (for the ancillary item) could lead to a near-zero reduction in the uncertainty associated with the unbundling's revenue, especially if the average purchase probability of the focal item across covariates is low. This scenario could result in the confidence set for the unbundling's revenue stopping shrinking and cause the seller to persistently opt for the unbundling strategy due to its large uncertainty bonus, and never explore the potentially more profitable bundling strategy.

To avoid the impact of uninformative covariates, one approach could be to modify the method of estimating $\theta^*_a$, perhaps by using the alternative estimator $\hat{\theta}'_{t,a} = \hat{\theta}_{t,b} - \hat{\theta}_{t,f}$, which relies exclusively on observations from the focal and bundled items. This would effectively avoid the influence of uninformative covariates by circumventing the need for $\hat{\theta}_{t,a}$. However, the challenge here is that this alternative estimator is informed by data from two different selling strategies, complicating the balance between learning $\theta^*_a$ (and thus the revenues of selling strategies) and earning by choosing the optimal selling strategy. All this to say, the issue of uninformative covariates, with the special structure of this problem, distinctly sets it apart from traditional dynamic pricing and bandit problems and requires new techniques.

To deal with this difficulty, Algorithm 2 applies the idea of maximizing "information gain" when there is uncertainty about the optimal selling strategy given $x_t$. Specifically, at each time $t$, it consists of two parts: the first part decides the selling strategy while the second part sets the price. When deciding the selling strategy, it compares the LCB revenue of each strategy with the UCB of the other. If one LCB is larger than the other UCB, it means that for this $x_t$, the seller can be quite confident about the superiority of one strategy and will adopt that strategy accordingly. Otherwise, the selling strategy will be chosen by (5). Recall that $\|x_t\|_{\Sigma^{-1}_{t-1,f}}$ and $\|x_t\|_{\Sigma^{-1}_{t-1,b}}$ represent the confidence level of the revenue of the focal and bundle product. The choice by (5) favors the strategy with larger uncertainty, and thus implements the "exploration" of the selling strategy. For the pricing part, if the seller chooses the unbundling strategy, then the problem reduces to the pure unbundling case of the previous section and the pricing policy in Algorithm 1 can be applied. Alternatively, if the seller chooses the bundling strategy, it can be viewed as a single-product dynamic pricing problem and the certainty-equivalent pricing policy can be applied.

**Remark 1.** *The choice of (5) implements the idea of information directed sampling [44]. We will see shortly that the regret induced by a wrong choice (of suboptimal selling strategy) is bounded by* $O\left(\|x_t\|_{\Sigma^{-1}_{t-1,f}} + \|x_t\|_{\Sigma^{-1}_{t-1,b}}\right)$ *for choosing either unbundling or bundling. Thus, the algorithm chooses the strategy that provides more information (for further periods) and also guarantees an information ratio [44] bounded by constant. In fact, it is easy to check that a UCB-based algorithm (for example, always choosing the selling strategy with a larger UCB of expected revenue) will not guarantee such a constant information ratio bound.*

**Remark 2.** *While Algorithm 1 directly uses the estimator* $\hat{\theta}_{t,a}$ *for the ancillary's pricing decision for (pure) unbundling pricing, Algorithm 2 computes and compares LCB/UCB of* $v_{t,a}$ *for the strategy decision. The rationale behind this difference is rooted in the different exploration–exploitation trade-offs required by each decision. When deciding whether to bundle or not, the feedback structure is analogous to*

**Algorithm 2** Confidence-based Pricing Algorithm

**Input**: Regularization parameter: $\lambda$.
**for** $t = 1, ..., T$ **do**
%%Choose the strategy:
Observe $x_t$ and compute the estimators $\hat{\theta}_{t,f}$, $\hat{\theta}_{t,a}$ and $\hat{\theta}_{t,b}$ and their confidence sets $\Theta_{t,f}$, $\Theta_{t,a}$ and $\Theta_{t,b}$. When there are no observations available, i.e., $\mathcal{D}_{t,f/a/b} = \emptyset$, set $\hat{\theta}_{t,f/a/b} = 0$ and $\Theta_{t,f/a/b} = \Theta$.
Compute UCBs and LCBs for optimal revenues under both strategies by

$$\bar{r}^*_{t,u} \coloneqq r^*_u\left(\bar{v}_{t,f}, \bar{v}'_{t,a}\right), \quad \underline{r}^*_{t,u} \coloneqq r^*_u\left(\underline{v}_{t,f}, \underline{v}'_{t,a}\right),$$

$$\bar{r}^*_{t,b} \coloneqq r^*_b\left(\bar{v}_{t,b}\right), \quad \underline{r}^*_{t,b} \coloneqq r^*_b\left(\underline{v}_{t,b}\right),$$

where

$$\bar{v}_{t,f} \coloneqq \max_{\theta\in\Theta_{t,f}} x_t^\top\theta, \quad \underline{v}_{t,f} \coloneqq \min_{\theta\in\Theta_{t,f}} x_t^\top\theta,$$

$$\bar{v}_{t,b} \coloneqq \max_{\theta\in\Theta_{t,b}} x_t^\top\theta, \quad \underline{v}_{t,b} \coloneqq \min_{\theta\in\Theta_{t,b}} x_t^\top\theta,$$

$$\bar{v}'_{t,a} \coloneqq \bar{v}_{t,b} - \underline{v}_{t,f}, \quad \underline{v}'_{t,a} \coloneqq \underline{v}_{t,b} - \bar{v}_{t,f}.$$

**if** $\underline{r}^*_{t,b} > \bar{r}^*_{t,u}$ **then**
Choose bundling, i.e., $A_t = b$.
**else if** $\underline{r}^*_{t,u} > \bar{r}^*_{t,b}$ **then**
Choose unbundling, i.e., $A_t = u$.
**else**
Choose the strategy by

$$A_t = \begin{cases} u & \text{if } \|x_t\|_{\Sigma^{-1}_{t-1,f}} \geq \|x_t\|_{\Sigma^{-1}_{t-1,b}}, \\ b & \text{if } \|x_t\|_{\Sigma^{-1}_{t-1,b}} > \|x_t\|_{\Sigma^{-1}_{t-1,f}}. \end{cases} \tag{5}$$

**end if**
%% Set the price
**if** $A_t = u$ **then**
%% Follow Algorithm 1
Compute the LCB and UCB of the customer valuation:

$$\underline{v}_{t,f} = \min_{\theta\in\Theta_{t,f}} x_t^\top\theta, \quad \bar{v}_{t,a} = \max_{\theta\in\Theta_{t,a}} x_t^\top\theta.$$

Set the prices by

$$p_{t,f} = p^*_f\left(\underline{v}_{t,f}, r^*_a(\bar{v}_{t,a})\right), \quad p_{t,a} = p^*_a\left(x_t^\top\hat{\theta}_{t,a}\right).$$

**else**
%% Follow certainty equivalent policy:
Set the price by $p_{t,b} = p^*_b\left(x_t^\top\hat{\theta}_{t,b}\right)$.
**end if**
**end for**

*a bandit problem—observations from one mechanism yield little information about the other. Therefore, Algorithm 2 uses the UCB/LCB framework (of $v_{t,a}$) to balance exploration and exploitation, similar to an arm elimination strategy (see, e.g., Chapter 6 in [35]). In contrast, for ancillary pricing under the unbundling strategy (and similarly for bundling pricing), directly using the estimator $\hat{\theta}_{t,a}$ is sufficient to achieve an $O(\log T)$ regret, as shown in Lemma 11. Intuitively, when the price coefficient is known (equal to 1 in our setting), the price decision does not affect the learning of $v_{t,a}$. This separation of pricing from parameter estimation renders the ancillary pricing problem analogous to a "full information" setting in online learning (see further discussion in Appendix A of [52]).*

## 5.2 Regret Analysis

The following theorem gives the regret bound for Algorithm 2. The regret bound consists of several parts: (i) the regret under the "bad" event when the confidence set does not cover the true parameter; (ii) the strategy regret when a suboptimal selling strategy is chosen; (iii) the pricing regret induced by

the suboptimal price when restricted to one selling strategy.

**Theorem 3.** *Under Assumption 1, 2, 3 and with the regularization parameter $\lambda = 1$, the regret of Algorithm 2 is bounded by*

$$O\left(d^2\sqrt{T}\log T\right) = \underbrace{6\bar{p}}_{\text{Regret under "bad" event}} + \underbrace{24\bar{\beta}\sqrt{dT\log\left(\frac{d+T}{d}\right)}}_{\text{Strategy regret (Lemma 9)}}$$
$$+ \underbrace{6\sqrt{2}\bar{\beta}\sqrt{dT\log\left(\frac{d+T}{d}\right)} + 2d\eta\bar{\beta}^2\log\left(\frac{d+T}{d}\right)}_{\text{Pricing regret when unbundling (Lemma 10)}} + \underbrace{2d\eta\bar{\beta}^2\log\left(\frac{d+T}{d}\right)}_{\text{Pricing regret when bundling (Lemma 11)}},$$

*where $\eta = B + \bar{p}B'$ and $\bar{\beta} = 2\bar{\theta} + \frac{2\mu}{\nu}\sqrt{2\log(T) + d\log\left(\frac{d+T}{d}\right)}$.*

In the following, we elaborate on these parts one by one. The detailed proofs can be found in Online Appendix C.

**Good event:**

First, we consider the "good" event where the confidence sets cover the true parameters. As we use the same parameter estimation routine as the previous section, the proof of the following lemma is the same as that of Lemma 3.

**Lemma 7** (Good Event)**.** *Let*

$$\mathcal{E}_f = \left\{\theta_f^* \in \Theta_{t,f} \quad \text{for } t = 1, ..., T\right\},$$
$$\mathcal{E}_a = \left\{\theta_a^* \in \Theta_{t,a} \quad \text{for } t = 1, ..., T\right\},$$
$$\mathcal{E}_b = \left\{\theta_b^* \in \Theta_{t,b} \quad \text{for } t = 1, ..., T\right\}.$$

*We have*

$$\mathbb{P}(\mathcal{E}_f \cap \mathcal{E}_a \cap \mathcal{E}_b) \geq 1 - 3/T.$$

**Strategy regret:**

Next, we consider the *strategy regret* which captures the regret induced by a suboptimal selling strategy. That is, when the bundling (resp. unbundling) strategy is more profitable at time $t$, the selling strategy $A_t$ is chosen to be unbundling (resp. bundling). Specifically, the strategy regret is defined by

$$\text{StrategyReg} := \sum_{t=1}^{T}\left|r_b^*\left(x_t^\top\theta_b^*\right) - r_u^*\left(x_t^\top\theta_f^*, x_t^\top\theta_a^*\right)\right| \cdot \mathbb{1}_{A_t \neq A_t^*}$$

where $A_t^*$ denotes the optimal selling strategy at time $t$. We note that the revenues for both strategies are taken under the optimal prices, so the definition focuses exclusively on the revenue loss induced by the suboptimal choice of the selling strategy.

**Lemma 8.** *Under the event $\mathcal{E}_f \cap \mathcal{E}_b$, the single-period strategy regret satisfies*

$$\left|r_b^*\left(x_t^\top\theta_b^*\right) - r_u^*\left(x_t^\top\theta_f^*, x_t^\top\theta_a^*\right)\right| \cdot \mathbb{1}_{A_t \neq A_t^*} \leq 6\bar{\beta}\left(\|x_t\|_{\Sigma_{t-1,f}^{-1}} + \|x_t\|_{\Sigma_{t-1,b}^{-1}}\right) \cdot \mathbb{1}_{A_t \neq A_t^*}.$$

Lemma 8 establishes an upper bound for single-period strategy regret. Intuitively, $\|x_t\|_{\Sigma_{t-1,f}^{-1}}$ and $\|x_t\|_{\Sigma_{t-1,b}^{-1}}$ represent the sizes of the confidence sets. Under the good event, $A_t \neq A_t^*$ happens only when the decision rule (5) is adopted. Consequently, the two confidence sets will overlap with each other and the revenue gap can be bounded by the summation of two confidence sets.

**Lemma 9.** *Under event $\mathcal{E}_f \cap \mathcal{E}_b$ and with any regularization parameter $\lambda \geq 1$, the total expected strategy regret satisfies*

$$\mathbb{E}_\epsilon \left[ \sum_{t=1}^{T} \left| r_b^* \left( x_t^\top \theta_b^* \right) - r_u^* \left( x_t^\top \theta_f^*, x_t^\top \theta_a^* \right) \right| \cdot \mathbb{1}_{A_t \neq A_t^*} \right] \leq 24\bar{\beta} \sqrt{dT \log\left( \frac{d\lambda + T}{d\lambda} \right)},$$

*where the expectation is taken with respect to the randomness in selling strategy $A_t$ induced by the randomness in utility shocks.*

Lemma 9 builds upon Lemma 8 and provides an upper bound for the expected strategy regret. The bound relies critically on the decision rule (5). Specifically, the decision rule aligns the right-hand-side upper bound with the choice of the selling strategy and thus makes the Elliptical Potential Lemma applicable (Lemma 21). Other than this point, the rest of the proof for Lemma 9 is similar to the previous case of pure unbundling.

**Pricing regret:**

While the previous part analyzes the regret induced by the suboptimal selling strategy, the pricing regret refers to the regret caused by the suboptimal pricing for unbundling or bundling time periods. More specifically, the expected pricing regrets are defined by:

$$\mathbb{E}_\epsilon \left[ \sum_{t \in \mathcal{T}_{T+1,f}} r_u^*(x_t^T \theta_f^*, x_t^T \theta_a^*) - \sum_{t \in \mathcal{T}_{T+1,f}} R_t^{\text{Alg2}} \right],$$

$$\mathbb{E}_\epsilon \left[ \sum_{t \in \mathcal{T}_{T+1,b}} r_b^*(x_t^T \theta_b^*) - \sum_{t \in \mathcal{T}_{T+1,b}} R_t^{\text{Alg2}} \right],$$

for the unbundling and bundling time periods respectively, where $R_t^{\text{Alg2}}$ denotes the revenue from Algorithm 2. The following two lemmas state the pricing regret upper bounds. Their analyses are similar to that of Theorem 1.

**Lemma 10.** *The expected pricing regret happened at unbundling periods (when $A_t = u$), under $\mathcal{E}_f \cap \mathcal{E}_a$ with any $\lambda \geq 1$, is upper bounded by*

$$6\sqrt{2}\bar{\beta} \sqrt{dT \log\left( \frac{d\lambda + T}{d\lambda} \right)} + 2d\eta\bar{\beta}^2 \log\left( \frac{d\lambda + T}{d\lambda} \right).$$

**Lemma 11.** *The expected pricing regret happened at bundling periods (when $A_t = b$), under $\mathcal{E}_b$ with any $\lambda \geq 1$, is upper bounded by*

$$2d\eta\bar{\beta}^2 \log\left( \frac{d\lambda + T}{d\lambda} \right).$$

# 6 One-Switch Setting

In this section, we consider the one-switch setting, where the seller is restricted to a uniform selling strategy for all customers but has the option to make a single strategic switch. In this case, the seller

must initially employ the unbundling strategy to learn the demand functions for each product (and also the bundle). This is because starting with a bundling strategy would not allow for separate estimation of the focal and ancillary item demands, and thus the revenue under an unbundling strategy.

In this sense, the seller uses the unbundling strategy in the short term to learn the market and then decides on its long-term strategy based on the learning outcome. In certain practical contexts, including customer-centric considerations like loyalty programs and brand consistency, such a one-switch policy may be preferable, compared to the policy studied in the previous section which may entail the selling strategy to change frequently and by customer. One example is Amazon Prime in the UK: it begins with the delivery prime (as the focal item) itself, which offers free and fast delivery from Amazon. But then it expands more and more other service subscriptions from other platforms, such as Deliveroo (a food delivery service) and Twitch (a streaming platform). Amazon bundles all these subscriptions into Amazon Prime, increases its price, and the customers cannot only subscribe to the delivery prime itself and have to buy the bundle.

We require the features $X_t$'s to be i.i.d. generated (Assumption 4). Such an assumption is necessary from a technical viewpoint: with only one switch of the selling strategy allowed throughout the horizon, it ensures that the seller can infer knowledge about future arrivals based on past observations.

Under this setting, the regret of a policy/algorithm $\pi$ is defined as

$$\mathrm{Reg}_T^\pi := T \cdot \max\left\{\mathbb{E}_X\left[r_u^*\left(X^\top\theta_f^*, X^\top\theta_a^*\right)\right], \mathbb{E}_X\left[r_b^*\left(X^\top\theta_b^*\right)\right]\right\} - \mathbb{E}\left[\sum_{t=1}^T R_t^\pi\right]$$

where the first two expectations are taken with respect to the covariates and the last expectation is with respect to both covariates and potential randomness in policy $\pi$. Compared to the regret definition in the last section, the benchmark sticks to one selling strategy and picks the more profitable one among these two. For the summation, $R_t^\pi$ denotes the revenue collected under the online policy $\pi$.

## 6.1 One-Switch Algorithm

Our one-switch Algorithm 3 begins by executing the unbundling strategy until it finds the bundling strategy is more profitable under certainty (when $\underline{r}_{t,b}^* \geq \bar{r}_{t,u}^*$). We name this condition the safe-switching condition since it ensures that the switch to the bundling strategy only occurs when it is "safe" enough: effectively avoiding the risk of switching to a bundling strategy that might be suboptimal and losing the chance of exploring (and switching back to) the unbundling.

The difference between Algorithm 3 and the previous personalized case is that in computing the upper and lower confidence bounds of the two strategies, we are comparing their long-term performance (since we need to stick to one strategy in the long run) rather than the advantage of one strategy in a specific time period. Specifically, the upper and lower revenue bounds for the two strategies are defined by

$$\bar{r}_{t,u}^*, \underline{r}_{t,u}^* := \frac{1}{t}\sum_{t'=1}^t r_u^*\left(X_{t'}^\top\hat{\theta}_{t+1,f}, X_{t'}^\top\hat{\theta}_{t+1,a}\right) \pm \left(4\bar{p}\sqrt{\frac{\log T}{t}} + \frac{2\bar{\beta}}{t}\sum_{t'=1}^t \|X_{t'}\|_{\Sigma_{t'-1,a}^{-1}}\right), \tag{6}$$

$$\bar{r}_{t,b}^*, \underline{r}_{t,b}^* := \frac{1}{t}\sum_{t'=1}^t r_b^*\left(X_{t'}^\top(\hat{\theta}_{t+1,f} + \hat{\theta}_{t+1,a})\right) \pm \left(4\bar{p}\sqrt{\frac{\log T}{t}} + \frac{2\bar{\beta}}{t}\sum_{t'=1}^t \|X_{t'}\|_{\Sigma_{t'-1,a}^{-1}}\right). \tag{7}$$

The first part in both definitions estimates the long-term revenue under each selling strategy, and the second part captures the size of the confidence set.

As in the case of Algorithm 1, the unbundling strategy provides observations for the learning of both the focal and ancillary products. Once it accumulates enough samples and becomes confident that the bundling strategy is better, the algorithm switches to the bundling strategy for the remaining time

periods.

**Algorithm 3** One-Switch Algorithm

**Input:** Regularization parameter $\lambda$.
Initialize $A_1 = A_2 = \cdots = A_T = u$.
**for** $t = 1, ..., T$ **do**
Observe $X_t = x_t$ and compute the estimators $\hat{\theta}_{t,f}$, $\hat{\theta}_{t,a}$ and the confidence sets $\Theta_{t,f}$, $\Theta_{t,a}$. When there are no observations available, i.e., $\mathcal{D}_{t,f/a/b} = \emptyset$, set $\hat{\theta}_{t,f/a/b} = 0$ and $\Theta_{t,f/a/b} = \Theta$.
**if** $A_t = u$ **then**
%% Follow Algorithm 1
Compute the UCB and LCB of the customer valuation (notations same as in §4):

$$\underline{v}_{t,f} = \min_{\theta \in \Theta_{t,f}} x_t^\top \theta, \quad \bar{v}_{t,a} = \max_{\theta \in \Theta_{t,a}} x_t^\top \theta,$$

Set the price by

$$p_{t,f} = p_f^* \left( \underline{v}_{t,f}, r_a^*(\bar{v}_{t,a}) \right),$$
$$p_{t,a} = p_a^* \left( x_t^\top \hat{\theta}_{t,a} \right).$$

**else if** $A_t = b$ **then**
%% Certainty-equivalent pricing
Compute $\hat{\theta}_{t,b}$ and set the price by certainty equivalent (notations same as in §5) $p_{t,b} = p_b^* \left( x_t^\top \hat{\theta}_{t,b} \right)$.
**end if**
Compute UCBs and LCBs of average revenues by (6) and (7).
**if** $\underline{r}_{t,b}^* \geq \bar{r}_{t,u}^*$ **then**
Set $A_{t+1} = \cdots = A_T = b$ (Switch the strategy to bundling).
**end if**
**end for**

## 6.2 Regret Analysis

Theorem 4 gives the regret of Algorithm 3. The first part concerns the regret caused by the revenue confidence sets in (6) and (7). The second part is the strategy regret induced by the initial unbundling exploration period; specifically, the strategy regret bounds the revenue loss when the optimal selling strategy is bundling but the algorithm starts with the unbundling strategy. This is because due to the safe-switching condition, with high probability, the algorithm will not switch to bundling if it is suboptimal (and thus will have 0 strategy regret when the unbundling is indeed optimal). The remaining two parts concern the revenue loss induced by the pricing decisions as in Theorem 3. We defer the detailed analysis to Online Appendix D.

**Theorem 4.** *Under Assumption 1, 2, 3, 4 and with the regularization parameter $\lambda = 1$, the regret of Algorithm 3 is bounded by*

$$O\left(d^2 \log T \sqrt{\frac{T}{q^*}}\right) = \underbrace{22\bar{p}}_{\text{Regret under "bad" event}} + \underbrace{16\bar{p}\sqrt{T \log T} + 8\bar{p}\sqrt{\frac{2dT}{q^*} \log\left(\frac{d+T}{d}\right)}}_{\text{Strategy regret}}$$
$$+ \underbrace{6\sqrt{2}\bar{\beta}\sqrt{dT \log\left(\frac{d+T}{d}\right)} + 2d\eta\bar{\beta}^2 \log\left(\frac{d+T}{d}\right)}_{\text{Pricing regret at unbundling (Lemma 10)}} + \underbrace{2d\eta\bar{\beta}^2 \log\left(\frac{d+T}{d}\right)}_{\text{Pricing regret at bundling (Lemma 11)}},$$

*where $\eta = B + \bar{p}B'$ and $\bar{\beta} = 2\bar{\theta} + \frac{2\mu}{\nu}\sqrt{2\log(T) + d\log\left(\frac{d+T}{d}\right)}$.*

As in Theorem 1, the regret bound involves the probability $q^*$ which governs the rate of observation samples from the ancillary product.

# 7 Numerical Experiments

In this section, we first conduct a series of illustrative simulation experiments to validate the effectiveness of our proposed algorithms. Following this, we use a real-world airline dataset to gain managerial insights to make appropriate selling decisions.

## 7.1 Simulation Experiments

In this subsection, we perform three experiments to underscore the value of key components within our algorithms. These experiments illustrate how our algorithms balance exploration and exploitation through these components under various scenarios. Specifically, we demonstrate the value of: (a) LCB pricing in Algorithm 1, (b) exploration when choosing selling strategies in Algorithm 2, and (c) safe-switching in Algorithm 3. For details about the implementation of the algorithms, please refer to Online Appendix E.

**Demand model.** In all three experiments in this subsection, we base our analysis on a logistic distribution-driven demand model. Specifically, we assume that the feature vector $x_t \in \mathbb{R}^2$, and the random utility shocks $\epsilon_f, \epsilon_a, \epsilon_b$ adhere to the logistic distribution with a price coefficient of 0.05 (which are all known to the seller). This approach aligns with the logit demand models that are commonly applied to estimate binary demands [50, 23, 6]. We also restrict the price for the focal, ancillary, and bundle to be in $[0, 50]$.

Specifically, for the unbundling case, the demand $[d_f, d_a]_u$ is defined as follows:

$$[d_f, d_a]_u = \begin{cases} [1,1], & \text{w.p. } \frac{1}{1+\exp(0.05p_f - x^\top \theta_f^*)} \cdot \frac{1}{1+\exp(0.05p_a - x^\top \theta_a^*)}, \\ [1,0], & \text{w.p. } \frac{1}{1+\exp(0.05p_f - x^\top \theta_f^*)} \cdot \frac{\exp(0.05p_a - x^\top \theta_a^*)}{1+\exp(0.05p_a - x^\top \theta_a^*)}, \\ [0,0], & \text{w.p. } \frac{\exp(0.05p_f - x^\top \theta_f^*)}{1+\exp(0.05p_f - x^\top \theta_f^*)}, \end{cases}$$

and for the bundling case:

$$d_b = \begin{cases} 1, & \text{w.p. } \frac{1}{1+\exp(0.05p_b - x^\top \theta_b^*)}, \\ 0, & \text{w.p. } \frac{\exp(0.05p_b - x^\top \theta_b^*)}{1+\exp(0.05p_b - x^\top \theta_b^*)}, \end{cases}$$

where $\theta_b^* = \theta_f^* + \theta_a^*$ and $\theta_f^*, \theta_a^* \in \mathbb{R}^2$ are parameters that will be specifically defined in each experiment.

### 7.1.1 Experiment 1: value of LCB pricing

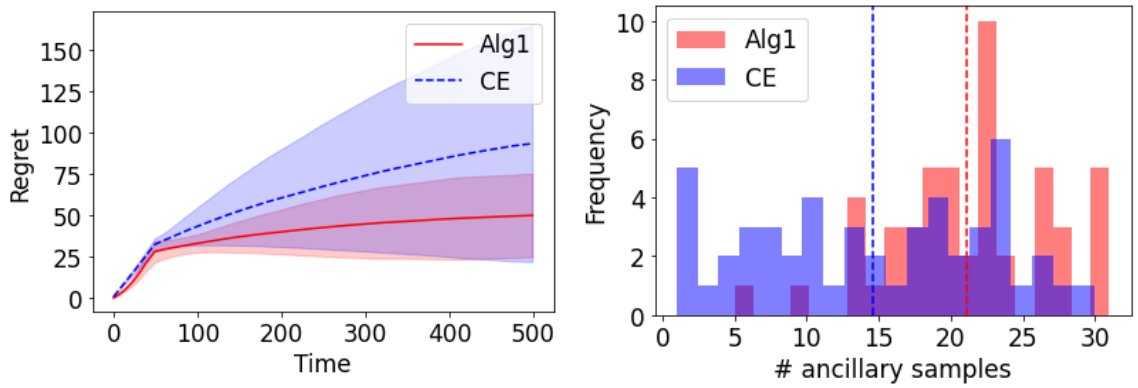
As discussed in §4, the LCB pricing for the focal product, which sets a price lower than the optimal price (with high probability), can accelerate the learning of ancillary product's demand and thus reduce the revenue loss in the future. To demonstrate the benefits of LCB pricing, we contrast Algorithm 1, which incorporates LCB pricing, with the certainty equivalent pricing policy (or greedy ILS by Keskin and Zeevi [29], Qiang and Bayati [43]) under the pure unbundling case as the following.

**Benchmark Algorithm.** The certainty equivalent (CE) policy mirrors Algorithm 1 except that the LCB pricing step (2) is replaced by CE pricing:

$$p_{t,f} = p_f^* \left( x_t^\top \hat{\theta}_{t,f}, r_a^*(x_t^\top \hat{\theta}_{t,a}) \right)$$

where the estimated optimal price is directly applied. We should note the CE pricing policy can also lead to sub-linear regret under several different sets of assumptions [43, 28, 52].

**Simulation Environment.** We consider the pure unbundling setting as in §4. We set the simulation horizon to $T = 500$, with $\theta_f^* = [0, -0.6]$ and $\theta_a^* = [-0.2, 0.4]$. The simulation environment consists of two phases based on the generation of the feature vector $x_t = [x_{t,1}, x_{t,2}]$. In the first phase ($t \leq 50$), we set




(a) (b)

Figure 1: Results of Experiment 1. (a): The cumulative regret for Algorithm 1 (Alg1) and the benchmark algorithm (CE). The results represent the mean values derived from 50 individual simulation runs with the shaded areas depicting the standard error for each data point. (b): The frequency distribution of the number of ancillary samples across the 50 simulations, with the dashed line marking the mean value.

$x_{t,1}, x_{t,2} \overset{\text{i.i.d.}}{\sim} \text{Unif}(1.5, 3.5)$ (uniform distribution within $[1.5, 3.5]$). In the second phase ($50 < t \leq 500$), we set $x_{t,1} = 0$ and $x_{t,2} \sim \text{Unif}(5, 6)$. Notably, when $t > 50$, the purchase probability of the focal product (and consequently, the learning rate for the ancillary product's demand) is extremely low ($<0.05$), even if $p_{t,f} = 0$.

**Result.** Figure 1(a) plots the cumulative regrets of both Algorithm 1 (Alg1) and the CE pricing policy (CE). The two algorithms exhibit similar performance up to $t = 50$. However, after this point, Alg1's regret is notably lower than that of CE. This difference stems from LCB pricing obtaining more observations about the ancillary product, enabling more accurate demand estimations compared to CE pricing (Figure 1(b)). Particularly in the second phase, where learning about the ancillary demand becomes more challenging, Alg1 underscores the value of additional exploration in the first phase through LCB pricing. Another interesting point is that during the first phase, LCB pricing does not sacrifice, and indeed even earns slightly more, revenue due to these explorations.

#### 7.1.2 Experiment 2: value of exploration in choosing selling strategy

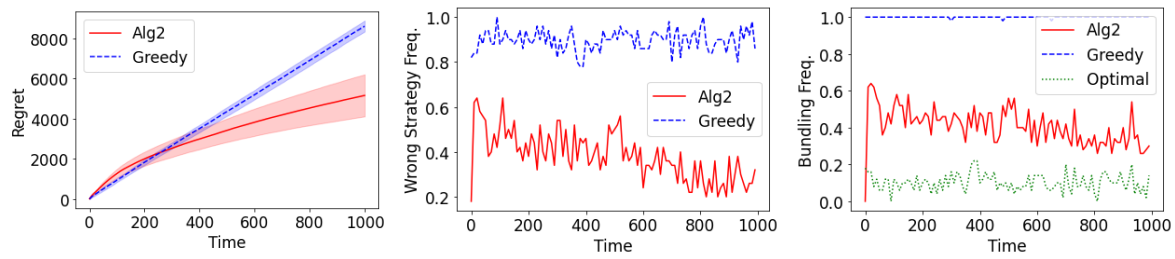
Algorithm 2 incorporates a comparison step (5) to select a selling strategy with more information gain and chooses the more profitable strategy only when the uncertainty is low. The objective of this experiment is to underscore the necessity of such exploratory tactics for achieving sub-linear regret. We achieve this by comparing Algorithm 2 with a variant that substitutes the selection step (5) with a greedy selection step as in the following benchmark algorithm.

**Benchmark Algorithm.** The greedy policy (Greedy) is similar to Algorithm 2, differing only in the strategy selection step. Here, the selling strategy $A_t$ is determined by greedily choosing the option that yields the higher estimated revenue:

$$A_t = \begin{cases} u & \text{if } \hat{r}^*_{t,b} \leq \hat{r}^*_{t,u}, \\ b & \text{if } \hat{r}^*_{t,b} > \hat{r}^*_{t,u}, \end{cases}$$

where $\hat{r}^*_{t,b} = r^*_b(x_t^\top \hat{\theta}_{t,b})$ and $\hat{r}^*_{t,u} = r^*_u(x_t^\top \hat{\theta}_{t,f}, x_t^\top \hat{\theta}_{t,a})$ represent the estimated revenues for bundling and unbundling, respectively.

**Simulation Environment.** We consider the setting where the seller can choose the personalized mechanism as in §5. For this simulation, we set the horizon to $T = 1000$. The parameters are configured




(a) (b) (c)

Figure 2: Results of Experiment 2 (a): The cumulative regret for Algorithm 2 (Alg2) and the benchmark algorithm (Greedy). (b): The frequency of selecting suboptimal selling strategy at each time step for Alg2 and Greedy. (c): The frequency of selecting bundling strategy at each time for Alg2, Greedy, and the optimal policy (which consistently selects the optimal selling strategy). All results are based on 50 independent simulation runs.

as $\theta_f^* = [-0.1, 1]$, $\theta_a^* = [0.2, -0.4]$, and $x_t = [x_{t,1}, x_{t,2}]$ with $x_{t,1}, x_{t,2} \overset{\text{i.i.d.}}{\sim} \text{Unif}(0, 5)$. In this setup, the probability of the unbundling strategy being optimal is approximately 0.9.

**Result.** Figure 2(a) displays the regret of both Algorithm 2 (Alg2) and the greedy policy (Greedy). Algorithm 2 demonstrates a sub-linear regret over time $T$. In contrast, the greedy policy exhibits a regret nearly linear in $T$ due to the modification in the strategy selection step.

Further analysis is presented in Figures 2(b) and 2(c), which plot the frequency of choosing suboptimal strategies and the bundling strategy at each timestep $t$ respectively, computed over 50 simulation runs. Figure 2(b) shows that Algorithm 2 reduces the frequency of selecting suboptimal selling strategies over time, whereas the greedy algorithm maintains a high frequency of about 0.9. This indicates that exploratory decision-making in strategy selection is crucial for learning the optimal selling approach and reduces revenue loss. As depicted in Figure 2(c), the greedy policy is almost stuck on the bundling strategy, hampering further learning about the unbundling strategy's revenue potential. This scenario is akin to the incomplete learning situation [30] in dynamic pricing for a single item with an unknown price coefficient, where pricing to greedily maximize the estimated revenue without incorporating exploratory prices can also lead to the expected regret linear in $T$.

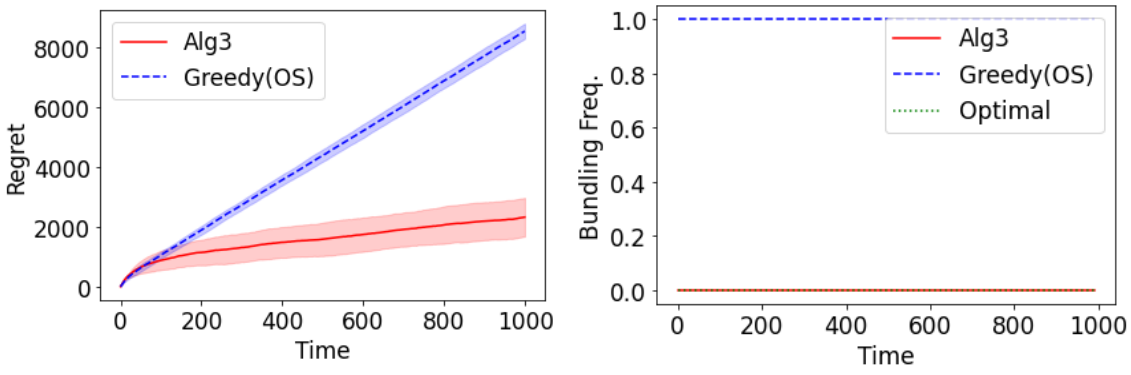
#### 7.1.3 Experiment 3: value of safe-switching

In this experiment, we show the significance of implementing safe-switching in Algorithm 3. Specifically, this involves switching (from unbundling) to the bundling strategy only when the condition $\underline{r}_{t,b}^* \geq \bar{r}_{t,u}^*$ is met, i.e., the bundling strategy is optimal with high certainty. We check the performance of Algorithm 3 and its variant where this switching condition is replaced by a greedy criterion as the following.

**Benchmark Algorithm.** The greedy policy for the one-switch setting (Greedy(OS)) is similar to Algorithm 3 except for the switching condition. Under this policy, the decision to switch from unbundling to bundling is based on a direct comparison of the estimated revenues: the switch occurs when $\sum_{t'=1}^{t} r_b^*(X_{t'}^\top(\hat{\theta}_{t+1,f} + \hat{\theta}_{t+1,a})) > \sum_{t'=1}^{t} r_u^*(X_{t'}^\top \hat{\theta}_{t+1,f}, X_{t'}^\top \hat{\theta}_{t+1,a})$, i.e., the estimated bundling's revenue is larger than the unbundling's based on the history feature vectors $\{X_{t'}\}_{t't=1,\ldots,t}$.

**Simulation Environment.** We consider the one-switch setting as in §6. The simulation environment is identical to that described in §7.1.2. In this setting, the optimal selling strategy under the one-switch condition is unbundling.

**Result.** Figure 3(a) illustrates the regret of both Algorithm 3 (Alg3) and the Greedy policy (Greedy(OS)). While Algorithm 3 exhibits sublinear regret in $T$, aligning with Theorem 4, the Greedy(OS) policy demonstrates a regret that is almost linear in $T$ caused by the modified switching condition. Moreover, Figure 3(b) indicates that the Greedy(OS) is again stuck on the bundling strategy and fails




(a) (b)

Figure 3: Results of Experiment 3, where the optimal selling strategy is unbundling. (a): The cumulative regret for Algorithm 3 (Alg3) and the benchmark algorithm (Greedy(OS)). (b): The frequency of selecting bundling strategy at each time for Alg3, Greedy(OS), and the optimal policy (which consistently selects unbundling). All results are based on 50 independent simulation runs.

to choose the optimal unbundling strategy. In the one-switch setting, once the policy switches to the suboptimal bundling strategy, it cannot revert back, thus permanently losing the opportunity to explore and earn from the unbundling strategy's revenue. This inability to revert underscores the necessity for implementing the safe-switching condition as proposed in Algorithm 3.

## 7.2 Case Study: Airline Industry

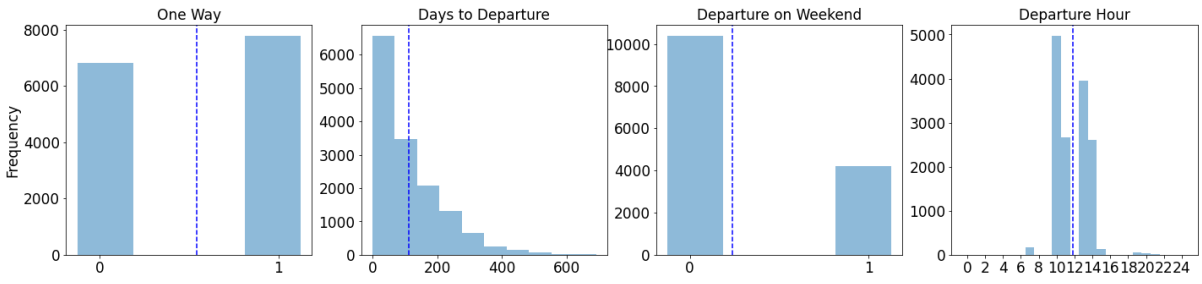
In this subsection, we validate the effectiveness of the proposed algorithms by applying them to a real-world data set from the airline industry. We first provide a detailed description of the data set and then analyze the optimal selling strategy under full information conditions (assuming known demand models) to derive management insights for decision-making. We then compare our algorithms with the benchmark algorithms, highlighting their benefits in real-world online settings.

### 7.2.1 Data description

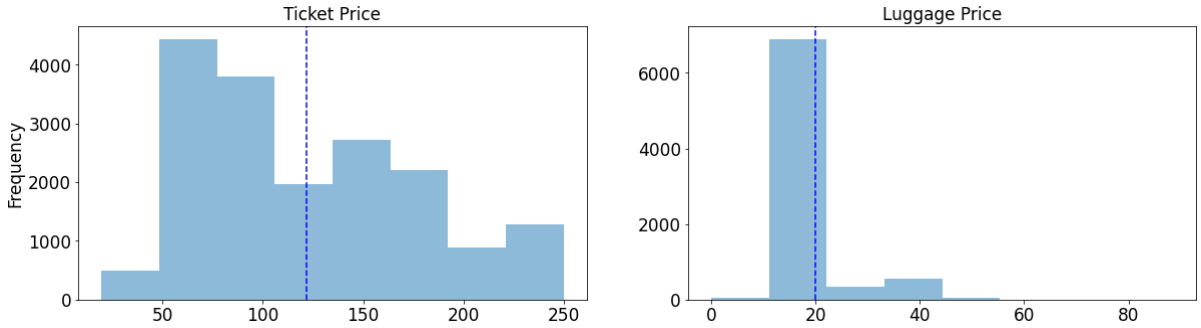
We use an airline dataset, encompassing transaction data for Economy Class customers from one airline company throughout 2015. Notably, the flight ticket, our focal item, does not include checked luggage capacity, allowing customers the option to purchase it as an ancillary product. The dataset records each transaction for purchased tickets and luggage along with their respective prices and encompasses four key features:

- *One Way* (Binary): Whether the flight is one-way (1) or round trip (0).
- *Days to Departure* (Continuous): The number of days between the ticket purchase and flight departure.
- *Departure on Weekend* (Binary): Whether the departure is on a weekend (1) or a weekday (0).
- *Departure Hour* (Continuous): The hour of the day at which the flight departs.

We also clean the dataset by removing outliers like unusually expensive bookings and erroneous entries. Histograms illustrating the distribution of features and prices are shown in Figure 4. Given that the raw dataset exclusively contains purchase records, we further generate synthetic no-purchase records for both tickets and luggage. Following these modifications, the dataset has a total of 78,135 samples. Details of these data processing steps can be found in Online Appendix E.




(a)




(b)

Figure 4: Data description for the Airline dataset. The frequency distribution of the (a) features, and (b) prices, with the dashed line marking the mean value.

**Demand Model Formulation.** To construct the demand models, we set the covariates $x \in \mathbb{R}^6$ with the following features:

$$[1, \textit{One Way}, \textit{Days to Departure}, \textit{Departure on Weekend}, \textit{Sin_ Hour}, \textit{Cos_ Hour}].$$

Here the intercept 1 is added in the first position, and to address the cyclical nature of the *Departure Hour* feature, we apply a sinusoidal transformation, resulting in the transformed features *Sine_ Hour* and *Cos_ Hour*. This transformation, achieved through the application of sine and cosine functions, effectively captures the cyclic pattern inherent in time-based data. It ensures that the model accurately represents hours, particularly those close to midnight and noon, in a continuous and coherent manner.

As in §7.1, we use logit demand models to estimate the demand functions from the samples. The purchase probabilities for the focal item (f), ancillary item (a) (conditional on purchasing the focal item), and the bundle (b) are given by

$$\frac{1}{1+\exp(c \cdot p_{f/a/b} - x^\top \theta^*_{f/a/b})}$$

respectively, where $c$ is the price coefficient to be estimated from the data.

After standardizing the features (except the binary ones) by subtracting the mean and scaling them to unit variance, we apply logistic regression to the entire dataset and get the following estimated parameters:

$$c = 0.0089,$$

$$\theta^*_f = [2.6904, 0.1237, -0.4830, -0.0444, -0.0359, -2.3884],$$

$$\theta^*_a = [0.3431, -0.6073, 0.0398, 0.4841, 0.3189, 0.0738],$$

$$\theta^*_b = [3.0335, -0.4836, -0.4432, 0.4397, 0.2830, -2.3146].$$

### 7.2.2 Optimal selling strategy and pricing under full information

We first analyze patterns among items and customers with different optimal (more profitable) selling strategies. In this context, we assume that the seller (airline company) knows the demand functions.

For each customer in our dataset, we compute the potential revenue from both unbundling and bundling strategies. This computation is to identify the strategy—bundling or unbundling—that yields higher profits for the seller for each customer based on their observed covariates.

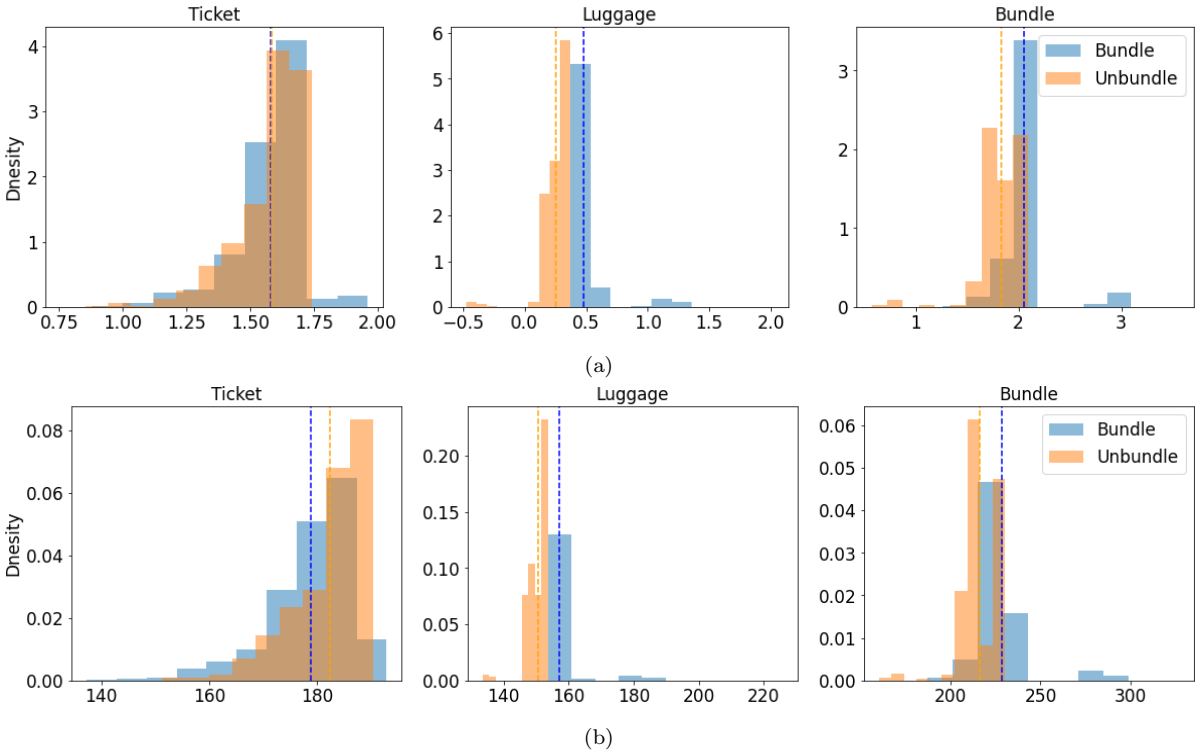



Figure 5: The frequency distribution of the (a) product values, (b) optimal prices for two distinct customer groups *Bundle* and *Unbundle*, each requiring different optimal selling strategies. The dashed line marks the mean value.

**Results.** In 78,135 customer samples, bundling is a more profitable strategy for 19,589 customers, while unbundling for 58,546 customers. We classify the former as the *Bundle* group and the latter as the *Unbundle* group. In the figures of this subsection, the *Bundle* group is indicated in blue and the *Unbundle* group in orange.

Figure 5(a) illustrates the distribution of product valuations for tickets, luggage, and bundles for two groups. It is observed that customers offered the bundling strategy show higher valuations for luggage compared to those suited for unbundling. Intuitively, when a high-value ancillary item is bundled with a focal item, it can appeal to customers who might otherwise consider the focal item alone too expensive. The perceived increased value of the bundle, due to the inclusion of the high-value ancillary item, makes the bundle more attractive. This bundling strategy aligns with the principles outlined in [2].

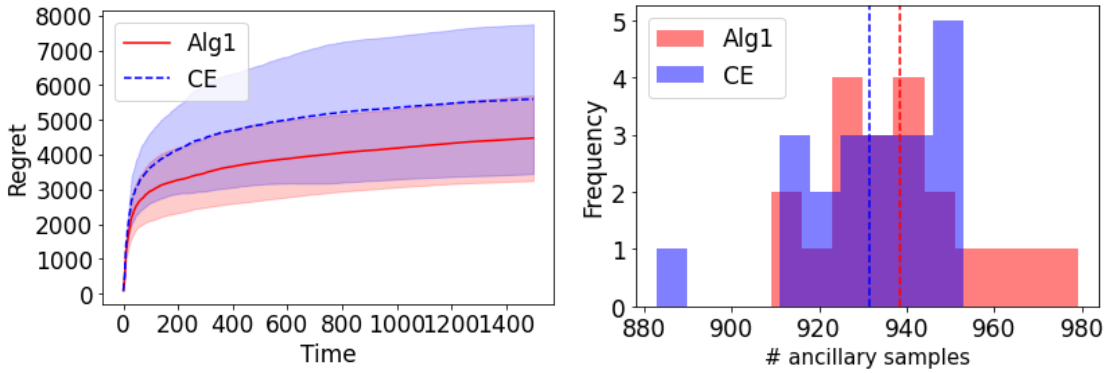
Figure 5(b) presents the distributions of the optimal prices for tickets and luggage (in cases of unbundling), as well as for bundles (in case of bundling). These price distributions are highly correlated to the product valuations, as higher valuations should have higher optimal prices to extract more customer utilities. Notably, the optimal prices for luggage under unbundling are significantly higher than the actual prices observed, whereas the ticket prices are lower as shown in Figure 4. This suggests that when employing unbundling, the seller might lower the price of the flight ticket to boost the purchases of ancillary products, a strategy in line with practices observed in real airline companies [5].

### 7.2.3 Performance under online setting

We assess the effectiveness of our algorithms in real-world online environments by examining their performance across three distinct settings: pure unbundling, personalized mechanisms, and one-switch setting, as elaborated in §4, §5 and §6, respectively. Additionally, we compare our algorithms against benchmark algorithms as described in §7.1. This comparison highlights the advantages of our methodology applied to real-world data.

**Simulation Environment.** We set the horizon $T = 1500$. To generate the covariates for sequentially arriving customers, we adopt the method used by Qiang and Bayati [43]: specifically, for each simulation run, we randomly select 1500 samples from the pool of $78,135$ and then randomly permute this subset. The demand models and their parameters adhere to those described in §7.2.1. All results are based on 20 independent simulation runs.

**Results.** The results in the airline dataset mirror the trends observed in the simulations of §7.1. In Figure 6(a), Algorithm 1 (Alg1) and the CE pricing policy (CE) both have sub-linear regrets under the pure unbundling setting, while Alg1 has more observations about the ancillary product (Figure 6(b)). Similarly, Algorithm 2 (Alg2), as shown in Figure 7(a), demonstrates sub-linear regret over time, outperforming the nearly linear regret trajectory of the greedy policy (Greedy). This is attributed to Alg2's reduced frequency in selecting suboptimal strategies and its more adaptive approach to strategy selection, unlike the greedy policy's tendency to fixate on bundling, limiting learning opportunities (Figures 7(b) and (c)). Furthermore, Algorithm 3 (Alg3), depicted in Figure 8(a), maintains sublinear regret, highlighting the importance of a safe-switching mechanism to avoid premature commitment to suboptimal strategies under the one-switch setting, as opposed to the Greedy(OS) policy's linear regret due to its incorrectly switching to bundling at the beginning. These findings emphasize the significance of exploratory decision-making and adaptability in dynamic pricing and bundling strategies in real-world online settings.

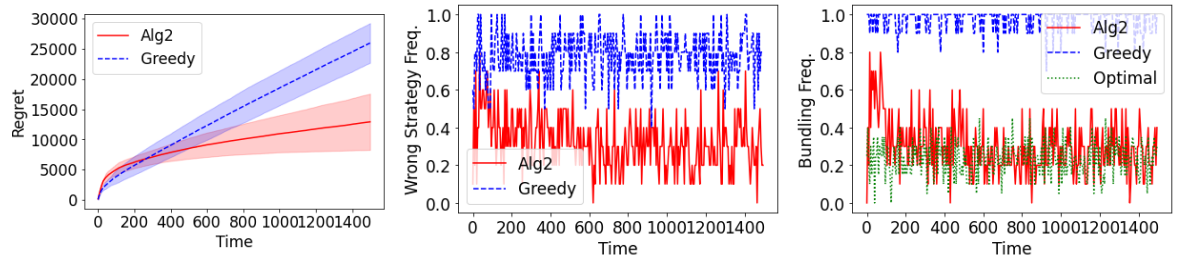



(a) (b)

Figure 6: Results for pure unbundling setting. (a): The cumulative regret for Algorithm 1 (Alg1) and the benchmark algorithm (CE). The results represent the mean values with the shaded areas depicting the standard error for each data point. (b): The frequency distribution of the number of ancillary samples, with the dashed line marking the mean value.

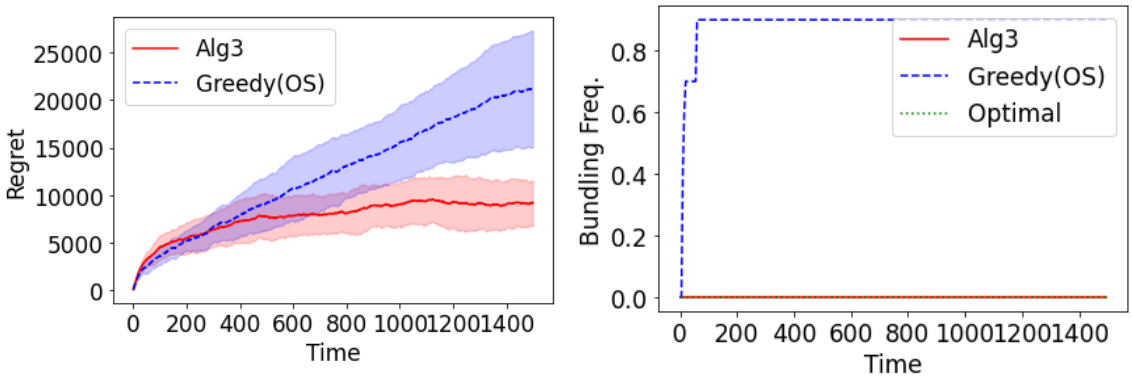
# 8 Comments and Conclusion

In this paper, we consider pricing and selling strategies for a focal-ancillary combination with unknown parameters in demand functions. Dependencies in a sequential purchase process create novel technical difficulties that we resolve. Specifically, we design algorithms for pricing and choosing a selling mechanism




(a) (b) (c)

Figure 7: Results for applying personalized mechanism. (a): The cumulative regret for Algorithm 2 (Alg2) and the benchmark algorithm (Greedy). (b): The frequency of selecting suboptimal selling strategy at each time step for Alg2 and Greedy. (c): The frequency of selecting bundling strategy at each time for Alg2, Greedy, and the optimal policy (which consistently selects the optimal selling strategy).




(a) (b)

Figure 8: Results for one-switch setting, where the optimal selling strategy is unbundling. (a): The cumulative regret for Algorithm 3 (Alg3) and the benchmark algorithm (Greedy(OS)). (b): The frequency of selecting bundling strategy at each time for Alg3, Greedy(OS), and the optimal policy (which consistently selects unbundling).

with provable bounded regrets under three settings: (a) pure unbundling, (b) personalized mechanism, and (c) unbundling and switching to bundling if necessary.

The key novelties in our analysis are the following: (a) We present two different ways to bound the single-period regret for pricing by controlling the impact from the uninformative covariates which cannot contribute to estimating the demand for the ancillary item. Such uninformative covariates are unique to the sequential purchase structure and set this pricing problem with our analysis apart from the current literature. (b) We incorporate the *Information Directed Sampling(IDS)* [44] idea to address the issue of uninformative covariates when deciding the selling strategy for exploration. We believe this is a novel approach in the area of revenue management. (c) In addition, we extend to the one-switch setting, where the seller can change the selling strategy at most once, and (d) conduct various numerical experiments to provide managerial insights and demonstrate the effectiveness of our algorithms.

We conclude the paper with discussions on future directions as follows:

- **Forward-looking model and mixed bundling strategy**: In our model, we assume that the consumer does not form rational expectations regarding the ancillary price when evaluating the focal item (as in [48]). In practice, however, customers can be *forward-looking*, simultaneously evaluating both the focal and ancillary items. In fact, this behavior is mathematically equivalent to a *mixed bundling* strategy, in which the focal item and its bundle with the ancillary are offered

together. Adapting our algorithms to accommodate forward-looking or mixed bundling behavior introduces significant challenges in learning customer choice behavior. In this setting, customers' decision processes involve a complex probabilistic model; for instance, the probability of purchasing both the focal and ancillary items can be expressed as $\mathbb{P}(v_a + \epsilon_a > p_a,\ v_a + v_f + \epsilon_a + \epsilon_f > p_f + p_a)$, which depends on the joint distribution of $(\epsilon_a, \epsilon_f + \epsilon_a)$. This complexity not only can lead to computational inefficiencies in estimation but also poses major difficulties in deriving theoretical error bounds (see the illustrative example in Online Appendix F4). To the best of our knowledge, no existing work has addressed this estimation challenge, even in an offline setting with available historical data, where there is no need to balance learning and earning as in the online setting.

- **Regret lower bound**: Establishing matching regret lower bounds for this sequential decision problem is complex due to the implicit dependencies (i.e., sale of ancillary item dependent on the purchases of the focal), and the information sharing structure (e.g., $\theta_b^*$ can be inferred from estimating $\theta_f^*$ and $\theta_a^*$ and can also be directly estimated by the observations on bundles). These unique aspects of the problem hinder the application of commonly used analytical tools, like the van Trees inequality or the Bretagnolle-Huber inequality, for regret lower bound analysis. Developing tight lower bounds for such learning problems with inherent dependencies is an intriguing area for future research and we provide more discussion in Online Appendix F3.

- **Multiple ancillary products and non-additive valuations**: It would be interesting to extend our algorithms to scenarios with multiple ancillary items, where we need to decide whether to bundle, what to bundle, and how to price them. In addition, our model, consistent with the literature on bundling [2, 38, 25, 14], assumes the valuation of a bundle is simply the sum of the valuations of the focal and the ancillary. However, this additive assumption can be restrictive in practice and exploring non-additive valuation formulations is a valuable direction for future research.

- **Unknown distributions of demand errors**: We assume the product valuations have random shocks with known distributions. While this assumption is common in the binary-demand dynamic pricing and learning literature [28, 6, 52], a few works [39, 45, 51] make progress in relaxing this assumption under different settings. Typically, these pricing algorithms involve a two-step procedure. First, they uniformly sample prices over a subset of periods to construct an unbiased estimate of the unknown parameter $\theta^*$. Then, they discretize the price space and apply bandit techniques—such as UCB-style algorithms—to learn the optimal price despite the unknown noise distribution. In our setting, however, the sequential nature of purchasing in the unbundling case means that uniformly sampled prices (and the UCB-style approach) for the focal item cannot ensure enough focal item purchases and thus influence the number of samples collected for estimating the ancillary item's valuation (where we design the LCB-pricing to guarantee that). Consequently, directly adopting these techniques would lead to poor estimation for the ancillary component, necessitating the design of a fundamentally different approach, which can be an interesting future direction.

# References


[1] Abdallah, Tarek, Arash Asadpour, Josh Reed. 2021. Large-scale bundle-size pricing: A theoretical analysis. *Operations Research* **69**(4) 1158–1185.

[2] Adams, William James, Janet L Yellen. 1976. Commodity bundling and the burden of monopoly. *The Quarterly Journal of Economics* 475–498.

[3] Allon, Gad, Achal Bassamboo, Martin Lariviere. 2011. Would the social planner let bags fly free? Https://ssrn.com/abstract=1919807.

[4] Azar, Mohammad Gheshlaghi, Alessandro Lazaric, Emma Brunskill. 2013. Sequential transfer in multi-armed bandit with finite set of models. *NIPS-Advances in Neural Information Processing Systems 25-2013*.

[5] Babić, Ružica Škurla, Maja Ozmec Ban, Jasmin Bajić. 2019. A review of recent trends in airline ancillary revenues. *EMC REVIEW-ECONOMY AND MARKET COMMUNICATION REVIEW* **17**(1).

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

[7] Bastani, Hamsa, David Simchi-Levi, Ruihao Zhu. 2022. Meta dynamic pricing: Transfer learning across experiments. *Management Science* **68**(3) 1865–1881.

[8] Bonnans, Joseph-Frédéric, Jean Charles Gilbert, Claude Lemaréchal, Claudia A Sagastizábal. 2006. *Numerical optimization: theoretical and practical aspects*. Springer Science & Business Media.

[9] Brueckner, Jan K., Darin N. Lee, Pierre M. Picard, Ethan Singer. 2015. Product unbundling in the travel industry: The economics of airline bag fees. *Journal of Economics & Management Strategy* **24**(3) 457–484.

[10] Cella, Leonardo, Alessandro Lazaric, Massimiliano Pontil. 2020. Meta-learning with stochastic linear bandits. *International Conference on Machine Learning*. PMLR, 1360–1370.

[11] Chen, Hongqiao, Ming Hu, Georgia Perakis. 2022. Distribution-free pricing. *Manufacturing & Service Operations Management* **24**(4) 1939–1958.

[12] Chen, Ningyuan, Adam N Elmachtoub, Michael Hamilton, Xiao Lei. 2020. Loot box pricing and design. *Proceedings of the 21st ACM Conference on Economics and Computation*. 291–292.

[13] Chen, Ningyuan, Anran Li, Shuoguang Yang. 2021. Revenue maximization and learning in products ranking. *Proceedings of the 22nd ACM Conference on Economics and Computation*. 316–317.

[14] Chen, Ningyuan, Xiaobo Li, Zechao Li, Chun Wang. 2022. Component pricing with a bundle size discount. *Available at SSRN 4032247* .

[15] Chen, Xi, Will Ma, David Simchi-Levi, Linwei Xin. 2023. Assortment planning for recommendations at checkout under inventory constraints. *Mathematics of Operations Research* .

[16] Chu, Chenghuan Sean, Phillip Leslie, Alan Sorensen. 2011. Bundle-size pricing as an approximation to mixed bundling. *The American Economic Review* 263–303.

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

[18] Cui, Yao, Izak Duenyas, Ozge Sahin. 2018. Unbundling of ancillary service: How does price discrimination of main service matter? *Manufacturing & Service Operations Management* **20**(3) 455–466.

[19] Ellison, Glenn. 2005. A Model of Add-On Pricing. *The Quarterly Journal of Economics* **120**(2) 585–637.

[20] Ettl, Markus, Pavithra Harsha, Anna Papush, Georgia Perakis. 2020. A data-driven approach to personalized bundle pricing and recommendation. *Manufacturing & Service Operations Management* **22**(3) 461–480.

[21] Fang, Hanming, Peter Norman. 2006. To bundle or not to bundle. *The RAND Journal of Economics* **37**(4) 946–963.

[22] Filippi, Sarah, Olivier Cappe, Aurélien Garivier, Csaba Szepesvári. 2010. Parametric bandits: The generalized linear case. *Advances in Neural Information Processing Systems* **23**.

[23] Gallego, Guillermo, Huseyin Topaloglu, et al. 2019. *Revenue management and pricing analytics*, vol. 209. Springer.

[24] Gomes, Renato, Jean Tirole. 2018. Missed sales and the pricing of ancillary goods. *The Quarterly Journal of Economics* **133**(4) 2097–2169.

[25] Hart, Sergiu, Noam Nisan. 2017. Approximate revenue maximization with multiple items. *Journal of Economic Theory* **172** 313–347.

[26] Hong, Joey, Branislav Kveton, Manzil Zaheer, Mohammad Ghavamzadeh. 2022. Hierarchical bayesian bandits. *International Conference on Artificial Intelligence and Statistics*. PMLR, 7724–7741.

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

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

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

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

[31] Kveton, Branislav, Mikhail Konobeev, Manzil Zaheer, Chih-wei Hsu, Martin Mladenov, Craig Boutilier, Csaba Szepesvari. 2021. Meta-thompson sampling. *International Conference on Machine Learning*. PMLR, 5884–5893.

[32] Kveton, Branislav, Csaba Szepesvari, Zheng Wen, Azin Ashkan. 2015. Cascading bandits: Learning to rank in the cascade model. *International Conference on Machine Learning*. PMLR, 767–776.

[33] Kveton, Branislav, Zheng Wen, Azin Ashkan, Csaba Szepesvari. 2015. Combinatorial cascading bandits. *arXiv preprint arXiv:1507.04208* .

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

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

[36] Li, Shuai, Baoxiang Wang, Shengyu Zhang, Wei Chen. 2016. Contextual combinatorial cascading bandits. *International conference on machine learning*. PMLR, 1245–1253.

[37] Li, Xiaobo, Hailong Sun, Chung Piaw Teo. 2020. Convex optimization for bundle size pricing problem. *Proceedings of the 21st ACM Conference on Economics and Computation*. 637–638.

[38] Li, Xinye, Andrew Chi-Chih Yao. 2013. On revenue maximization for selling multiple independently distributed items. *Proceedings of the National Academy of Sciences* **110**(28) 11232–11237.

[39] Luo, Yiyun, Will Wei Sun, Yufeng Liu. 2023. Distribution-free contextual dynamic pricing. *Mathematics of Operations Research* .

[40] Ma, Will, David Simchi-Levi. 2021. Reaping the benefits of bundling under high production costs. *International Conference on Artificial Intelligence and Statistics*. PMLR, 1342–1350.

[41] McAfee, R. Preston, John McMillan, Michael D. Whinston. 1989. Multiproduct monopoly, commodity bundling, and correlation of values. *The Quarterly Journal of Economics* **104**(2) 371–383.

[42] Peña, Victor H, Tze Leung Lai, Qi-Man Shao. 2009. *Self-normalized processes: Limit theory and Statistical Applications*. Springer.

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

[44] Russo, Daniel, Benjamin Van Roy. 2018. Learning to optimize via information-directed sampling. *Operations Research* **66**(1) 230–252.

[45] Shah, Virag, Jose Blanchet, Ramesh Johari. 2019. Semi-parametric dynamic contextual pricing. *arXiv preprint arXiv:1901.02045* .

[46] Shapiro, Carl. 1995. Aftermarkets and consumer welfare: Making sense of Kodak. *Antitrust Law Journal* **63**(2) 483–511.

[47] Simchi-Levi, David, Rui Sun, Huanan Zhang. 2022. Online learning and optimization for revenue management problems with add-on discounts. *Management Science* **68**(10) 7402–7421.

[48] Song, Boqian, Michael ZF Li. 2018. Dynamic pricing with service unbundling. *Production and Operations Management* **27**(7) 1334–1354.

[49] Stigler, George J. 1963. United States v. Loew's inc.:A note on block-booking. *The Supreme Court Review* **1963** 152–157.

[50] Talluri, Kalyan T, Garrett J Van Ryzin. 2006. *The theory and practice of revenue management*, vol. 68. Springer Science & Business Media.

[51] Tullii, Matilde, Solenne Gaucher, Nadav Merlis, Vianney Perchet. 2024. Improved algorithms for contextual dynamic pricing. *Advances in Neural Information Processing Systems* **37** 126088–126117.

[52] Wang, Hanzhao, Kalyan Talluri, Xiaocheng Li. 2025. On dynamic pricing with covariates. *Operations Research* .

[53] Wang, Ruxian, Maqbool Dada, Ozge Sahin. 2019. Pricing ancillary service subscriptions. *Management Science* **65**(10) 4712–4732.

[54] Wu, Shin-yi, Lorin M Hitt, Pei-yu Chen, G Anandalingam. 2008. Customized bundle pricing for information goods: A nonlinear mixed-integer programming approach. *Management Science* **54**(3) 608–622.

[55] Xu, Jianyu, Yu-xiang Wang. 2021. Logarithmic regret in feature-based dynamic pricing. *arXiv preprint arXiv:2102.10221* .