Product Ranking for Revenue Maximization with Multiple Purchases

Renzhe Xu[1], **Xingxuan Zhang**[1], **Bo Li**[2], **Yafeng Zhang**[3], **Xiaolong Chen**[3], **Peng Cui**[1*]
[1]Department of Computer Science and Technology, Tsinghua University, Beijing, China
[2]School of Economics and Management, Tsinghua University, Beijing, China
[3]Meituan Inc., Beijing, China
xrz199721gmail.com, xingxuanzhang@hotmail.com, libo@sem.tsinghua.edu.cn
yafengz@outlook.com, chenxiaolong13@meituan.com, cuip@tsinghua.edu.cn

Abstract

Product ranking is the core problem for revenue-maximizing online retailers. To design proper product ranking algorithms, various consumer choice models are proposed to characterize the consumers' behaviors when they are provided with a list of products. However, existing works assume that each consumer purchases at most one product or will keep viewing the product list after purchasing a product, which does not agree with the common practice in real scenarios. In this paper, we assume that each consumer can purchase multiple products at will. To model consumers' willingness to view and purchase, we set a random attention span and purchase budget, which determines the maximal amount of products that he/she views and purchases, respectively. Under this setting, we first design an optimal ranking policy when the online retailer can precisely model consumers' behaviors. Based on the policy, we further develop UCB-like algorithms with $\tilde{O}(\sqrt{T})$ regret that estimates consumers' behaviors and maximize revenue simultaneously in online settings. Experiments on both synthetic and semi-synthetic datasets prove the effectiveness of the proposed algorithms. The source code is available at https://github.com/windxrz/product-ranking.

1 Introduction

Online retailing has become increasingly popular over the last decades [17, 27, 52]. The way of product ranking is the crux for online retailers because it determines the consumers' shopping behaviors [17] and thus influences the retailers' revenue [20, 49]. For instance, the probability of consumers' purchasing from a firm or clicking an advertisement is strongly related to the display order [8, 3, 32]. Therefore, it is crucial for online retailers to design proper product ranking algorithms for revenue maximization.

The consumer choice model, when faced with a list of products, is the basis of the problem [17]. Generally speaking, existing works assume that the consumers view the list of products sequentially and select products according to a choice model [42, 41, 31, 37, 18, 25, 24, 36]. However, most of them assume that each consumer purchases at most one product and stop browsing immediately afterward, which does not agree with the common practice since a consumer may expect multiple purchases on a website or an app [33, 34, 16, 27]. Recently, several works consider a multiple-purchase setting while their targets are not revenue maximization [27, 15], or they assume that consumers always continue viewing the product list after purchasing any product, which is not practical in real scenarios [39].

*Corresponding Author

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

In this paper, we propose a more realistic consumer choice model to characterize consumer behaviors under multiple-purchase settings. Following [13, 17, 39], we assume that the consumers view the product list sequentially and purchase a product when its utility exceeds a certain threshold. Similar to [17], we assume the existence of the attention span for each consumer, which determines the maximal amount of products that the consumer is willing to view. More importantly, instead of forcing the consumer to quit viewing once she purchases one product, we allow each consumer to conduct multiple purchases at will and explore product ranking models for revenue maximization. Specifically, we constrain each consumer with a purchase budget to determine the maximal amount of purchases. Both the attention span and the purchase budget are assumed to obey the geometric distribution [13, 39].

Under this setting, we study the optimal product ranking policy to maximize online retailers' revenue. We first characterize the optimal ranking policy that achieves the maximal revenue when the online retailer knows the consumer parameters, including the purchase probabilities for each product and the distribution of the attention span and purchase budget. Afterward, we investigate the problem in online learning frameworks where consumer parameters are unknown to online retailers. In both non-contextual (*i.e.*, all consumers share the same parameters) and contextual settings (*i.e.*, consumers have personalized parameters), we provide UCB-like algorithms [1] that could fit consumers' parameters and maximize revenue simultaneously. We further theoretically prove that the algorithms achieve $\tilde{O}(\sqrt{T})$ regrets in both settings. Finally, we conduct experiments to verify the effectiveness of the algorithms.

To conclude, our contributions are highlighted as follows:

- We propose a novel consumer choice model to deal with multiple-purchase activities of consumers in online retailing scenarios.
- We characterize the optimal ranking policy and further design UCB-like algorithms with $\tilde{O}(\sqrt{T})$ regret based on them in both non-contextual and contextual online settings.
- We conduct experiments on both synthetic and semi-synthetic datasets to prove the effectiveness of our method in the multiple-purchase setting.

2 Related Works

Consumer choice model Traditional choice models adopt the multinomial logic (MNL) [12, 22, 42] models to characterize consumers' behaviors. However, these models can not describe the behaviors when the consumers are provided with a list of products. As a result, several choice models are proposed recently. In detail, Abeliuk et al. [2] add a position bias term in the traditional MNL model. Gallego et al. [29], Aouad and Segev [4] suppose that the products are framed into a set of virtual web pages. Consumers select a product, if any, from these pages following a general choice model (such as MNL). Flores et al. [28] propose the Sequential Multinomial Logit (SML), which assumed a two-stage MNL model and supposed that products are divided into two stages with prior knowledge. Liu et al. [40], Feldman and Segev [26], Gao et al. [30] further extend the setting to multiple stages. These works suppose that each stage contains multiple products while recently several works assume that a product is treated as a single stage [17, 13, 14]. These models are also closely related to the cascade model [12, 35, 37, 54, 19]. Gao et al. [31] further propose the general cascade click model. Following these works, we suppose that each stage contains only one product and the consumers view the products sequentially.

In addition, several works consider the setting where consumers may have multiple interactions with the platforms. However, most of them focus on the click model [15, 16, 33–35] or MNL-based consumer choice models [48, 53, 50]. In addition, Ferreira et al. [27] maximize the number of consumers who engage with the site. Liang et al. [39] aim to maximize the revenue in the multiple-purchase setting while they assume that consumers always continue viewing the product list after purchasing any product, which is not practical in real scenarios.

Online learning and multi-armed bandit Online retailers need to design proper algorithms to model consumers' behaviors and maximize revenue at the same time, which is closely related to the multi-armed bandit framework [47, 45]. Many works [5, 9, 11, 10, 23] have been proposed to deal with demand learning or price experimentation problems with the help of the framework in the field

of revenue management. Recently, more methods [39, 17, 13, 14, 6, 44] adopt similar techniques in the product ranking setting, which inspires the algorithm proposed in this paper.

3 Preliminaries

Let $[N] \triangleq \{1, 2, \ldots, N\}$ be the set of all products owned by an online retailer (he). The product k generates revenue r_k when a consumer purchases it. Let $r_{\max} \triangleq \max_{k\in[N]} r_k$ be the maximal revenue for all products and $\boldsymbol{r} = \{r_k\}_{k=1}^N$. We use the notation $\boldsymbol{\sigma}_t \dashv [N]$ to denote the ranking policy $\boldsymbol{\sigma}_t$ on the products for the t-th consumer (she). More precisely, $\forall k \in [N]$, $\sigma_{t,k} \in [N]$ represents the product displayed in the k-th position for her. Reversely, $\sigma_{t,k}^{-1}$ denotes the position of the k-th product in the ranking list. Furthermore, for any subset $Q \subseteq [N]$ that represents the possible positions in the ranking list, we use $\sigma_{t,Q}$ to denote the corresponding set of the products *w.r.t.* the position set Q, *i.e.*, $\sigma_{t,Q} = \{\sigma_{t,k} : k \in Q\}$. In addition, we use $\max(\cdot)$ to denote the maximal element in a set.

Consumer choice model for multiple purchases When the t-th consumer arrives, she views the list of products sequentially. The probability of product k being purchased by her is denoted as $\lambda_{t,k}$ and let $\boldsymbol{\lambda}_t = \{\lambda_{t,k}\}_{k=1}^N$. To characterize the consumers' behaviors under multiple-purchase settings, we assume that the consumer is endowed with the attention span V_t and purchase budget B_t, which determine the maximal number of the products that she will view and purchase, respectively. In reality, the consumer always has random attention span and purchase budget [17]. Therefore, following [13, 39], we assume $V_t \sim \mathrm{Geo}(1 - q_t)$ and $B_t \sim \mathrm{Geo}(1 - s_t)$ are both geometric distributions with parameters $1 - q_t$ and $1 - s_t$, respectively. As a result, $P(V_t = v) = q_t^{v-1}(1 - q_t)$ for any $v \geq 1$ and $P(B_t = b) = s_t^{b-1}(1 - s_t)$ for any $b \geq 1$. With attention span V_t and purchase budget B_t, the consumer keeps viewing the products until she views V_t or purchases B_t products.

Because of the attention span V_t, the consumer will not purchase products whose positions are greater than V_t in the ranking list. Let $Q \subseteq [V_t]$ denote the possible positions of the purchased products by the consumer. The corresponding indices of purchased products are $\sigma_{t,Q}$. Therefore, the probability that the consumer purchases the subset of products $\sigma_{t,Q}$ is given by

$$P^{\mathrm{buy}}(Q; V_t, B_t, \boldsymbol{\lambda}_t, \boldsymbol{\sigma}_t) = \begin{cases} \left(\prod_{k\in Q} \lambda_{t,\sigma_{t,k}}\right) \left(\prod_{k\in[V_t]\setminus Q} \left(1 - \lambda_{t,\sigma_{t,k}}\right)\right), & \text{if } |Q| < B_t. \\ \left(\prod_{k\in Q} \lambda_{t,\sigma_{t,k}}\right) \left(\prod_{k\in[V_t]\setminus Q, k\leq \max(Q)} \left(1 - \lambda_{t,\sigma_{t,k}}\right)\right), & \text{if } |Q| = B_t. \\ 0, & \text{if } |Q| > B_t. \end{cases} \quad (1)$$

Specifically, the probability $P^{\mathrm{buy}}(Q; V_t, B_t, \boldsymbol{\lambda}_t, \boldsymbol{\sigma}_t)$ is calculated based on the cardinality of Q. When $|Q| < B_t$, the consumer does not meet her purchase budget and stops browsing the list after viewing V_t products. When $|Q| = B_t$, the consumer stops browsing as long as she purchases the last product in the list Q. When $|Q| > B_t$, the probability becomes 0 since the number of purchases $|Q|$ exceeds her purchase budget.

Remark. The setting of the purchase budget provides a rational way to characterize the consumers' behaviors with multiple purchases and extends previous works that suppose consumers would purchase at most one product [13, 17]. Furthermore, when $s_t \to 0$, the consumers would tend to purchase at most one product and our setting degenerates to the setting proposed in [13]. Thus the previous setting can be considered a special case of our setting. In addition, Liang et al. [39] assume that consumers always continue viewing the product list after purchasing any product, which is not practical in real scenarios. By contrast, the purchase budget proposed in our paper leads to a different and more practical consumer choice model. In addition, our model introduces an extra parameter (*i.e.*, the parameter), which makes our problem more challenging.

Revenue optimization The online retailer chooses a ranking policy $\boldsymbol{\sigma}_t \dashv [N]$ on the products for the t-th consumer. For the consumer with fixed purchase probabilities for each product $\boldsymbol{\lambda}_t$, attention span V_t, and purchase budget B_t, the expected revenue could be written as

$$R(V_t, B_t; \boldsymbol{\lambda}_t, \boldsymbol{\sigma}_t) = \sum_{Q\subseteq[V_t]} P^{\mathrm{buy}}(Q; V_t, B_t, \boldsymbol{\lambda}_t, \boldsymbol{\sigma}_t) \left(\sum_{i\in Q} r_{\sigma_{t,i}}\right). \quad (2)$$

In reality, the attention span V_t and purchase budget B_t are random and can not be estimated accurately by the online retailer. As a result, the retailer calculates the expected revenue for the t-th

consumer by sampling V_t from $\text{Geo}(1-q_t)$ and sampling B_t from $\text{Geo}(1-s_t)$, *i.e.*,

$$\text{Revenue}\left(\boldsymbol{\sigma}_t; \boldsymbol{\lambda}_t, q_t, s_t\right) = \mathbb{E}_{V_t \sim \text{Geo}(1-q_t), B_t \sim \text{Geo}(1-s_t)}\left[R(V_t, B_t; \boldsymbol{\lambda}_t, \boldsymbol{\sigma}_t)\right]. \tag{3}$$

To maximize the revenue, the optimal ranking policy is given by

$$\boldsymbol{\sigma}_t^*(\boldsymbol{\lambda}_t, q_t, s_t) = \underset{\boldsymbol{\sigma}_t \dashv [N]}{\arg\max}\, \text{Revenue}\left(\boldsymbol{\sigma}_t; \boldsymbol{\lambda}_t, q_t, s_t\right). \tag{4}$$

For simplicity, we use $\boldsymbol{\sigma}_t^*$ to denote $\boldsymbol{\sigma}_t^*(\boldsymbol{\lambda}_t, q_t, s_t)$.

Basic assumption We also need the following assumption to guarantee the tractability of the proposed problem. In reality, consumers always have finite attention span, which indicates that q_t is always smaller than 1.

Assumption 3.1. There exists a known constant $\epsilon_q > 0$ such that for all $t > 0$, $q_t \leq 1 - \epsilon_q$.

4 Optimal Ranking Policy Given Consumers' Characteristics

In this section, we introduce the optimal ranking policy when the online retailer exactly knows the consumers' parameters, including the purchase probability for each product $\boldsymbol{\lambda}$, random attention span (parametrized with q), and random purchase budget (parametrized with s).

We drop the subscript t in this section for simplicity. The optimal ranking strategy stems from the following theorem.

Theorem 4.1. *Under Assumption 3.1, for any $1 \leq i < N$, the following holds for the optimal ranking strategy $\boldsymbol{\sigma}^*(\boldsymbol{\lambda}, q, s)$.*

$$\frac{\lambda_{\sigma_i^*} r_{\sigma_i^*}}{1 - q + q(1-s)\lambda_{\sigma_i^*}} \geq \frac{\lambda_{\sigma_{i+1}^*} r_{\sigma_{i+1}^*}}{1 - q + q(1-s)\lambda_{\sigma_{i+1}^*}}. \tag{5}$$

Remark. The expected revenue of a single product σ_i^* is $\lambda_{\sigma_i^*} r_{\sigma_i^*}$, which grows as the purchase probability and the revenue generated by the product increase. Yet purchasing a product reduces the purchase probability of the following products, and thus the expected revenue of other products decreases as σ_i^* increases. Furthermore, when $s \to 0$, the term in Equation (5) becomes the same as that in [13, Theorem 1] if ignoring the marketing fatigue (*i.e.*, overexposure to unwanted marketing messages) term in their result. As a result, Theorem 4.1 consistently extends the results on the single-purchase setting [13] to the multiple-purchase setting.

Liang et al. [39] assume that the online retailer will stop recommending products with a certain probability $1 - w_t$ to avoid the cost brought by marketing fatigue. As a result, the parameter w_t is a part of the ranking policy instead of consumers' characteristics and should be optimized to achieve maximal revenue. In our paper, we focus on the multiple purchases with budget setting and leave the extension to the case with the marketing fatigue as future work.

According to Theorem 4.1, the online retailer can provide the optimal ranking policy $\boldsymbol{\sigma}^*$ by sorting the products in descending order with value $(\lambda_k r_k) \,/\, (1 - q + q(1-s)\lambda_k)$ for the k-th product. The time complexity is $O(N \log N)$.

5 Online Learning of the Ranking Policy

In this section, we consider a more realistic scenario where the online retailer has no prior knowledge about consumers' characteristics. At each timestamp, a consumer arrives and the online retailer can only provide proper product ranking policies based on the feedback of previous consumers. As a result, the retailer needs to design online learning algorithms to model consumers' behaviors and maximize revenue in the meantime. We consider two settings, *i.e.*, the non-contextual setting where all consumers share the same parameters (Section 5.1) and the contextual setting where consumers have personalized behaviors (Section 5.2). Before introducing the details of the algorithms, we first formally define the notations to characterize the consumers' behaviors.

Notations for consumers' behaviors Let $\Phi_t \in \mathbb{Z}_+$ be the random variable that denotes the number of products viewed by the t-th consumer and $\boldsymbol{\Gamma}_t = \{\gamma_{t,k}\}_{k=1}^{\Phi_t}$ be the purchase activities. Specifically, $\gamma_{t,k} = 1$ if and only if the product at the k-th position (*i.e.*, $\sigma_{t,k}$) is purchased.

Let random variable $\eta_{t,k} \in \{0,1\}$ denote whether the consumer continues viewing products after she views the k-th product but does not buy it and $\boldsymbol{\eta}_t = \{\eta_{t,k}\}_{k=1}^{\Phi_t}$. Let random variable $\mu_{t,k} \in \{0,1\}$ denote whether the consumer continues viewing products if she buys the k-th product and $\boldsymbol{\mu}_t = \{\mu_{t,k}\}_{k=1}^{\Phi_t}$. So $\eta_{t,k}$ is observed only when $\gamma_{t,k} = 0$ and $\mu_{t,k}$ is observed only when $\gamma_{t,k} = 1$.

Let $\mathrm{Ber}(\cdot)$ denote the Bernoulli distribution and we have the following result on the distribution of these random variables.

Lemma 5.1. *$\gamma_{t,k}$ are i.i.d. random variables and $\gamma_{t,k} \sim \mathrm{Ber}\left(\lambda_{\sigma_{t,k}}\right)$. $\eta_{t,k}$ are i.i.d. random variables and $\eta_{t,k} \sim \mathrm{Ber}(q_t)$. $\mu_{t,k}$ are i.i.d. random variables and $\mu_{t,k} \sim \mathrm{Ber}(q_t s_t)$.*

Lemma 5.1 is the basis of the algorithms that estimate parameters $\boldsymbol{\lambda}$, q, and s from historical data. In addition, compared with s_t, it is easier to estimate $q_t s_t$ from data and we denote it as $w_t \triangleq q_t s_t$.

5.1 Non-contextual Online Setting

In this subsection, we assume that all consumers share the same parameters, including the purchase probabilities on different products and the parameters on the distribution of the attention span and purchase budget. To be specific, we assume that there exist $\boldsymbol{\lambda}, q, s$ such that $\boldsymbol{\lambda}_t = \boldsymbol{\lambda}$, $q_t = q$, and $s_t = s$.

A) Estimation of Parameters

To estimate the parameters at timestamp t, we leverage the observed behavior of consumers before timestamp t. We first calculate the following statistics that can help provide unbiased estimators of the consumer parameters.

In detail, let $C_{t,k}$ be the number of times that the product k is observed and $c_{t,k}$ be the number of times the product k is purchased, *i.e.*, $C_{t,k} \triangleq \sum_{u=1}^{t}\sum_{i=1}^{\Phi_t} \mathbb{I}\left[\sigma_{t,i} = k\right]$ and $c_{t,k} \triangleq \sum_{u=1}^{t}\sum_{i=1}^{\Phi_t} \mathbb{I}\left[\sigma_{t,i} = k\right]\mathbb{I}\left[\gamma_{t,i} = 1\right]$. In addition, Let D_t^q be the number of times that consumers view a product and do not purchase it and d_t^q be the number of times that consumers continue viewing the list after viewing a product without purchasing it, *i.e.*, $D_t^q \triangleq \sum_{u=1}^{t}\sum_{i=1}^{\Phi_t} \mathbb{I}\left[\gamma_{t,i} = 0\right]$ and $d_t^q \triangleq \sum_{u=1}^{t}\sum_{i=1}^{\Phi_t} \mathbb{I}\left[\gamma_{t,i} = 0\right]\eta_{t,i}$. Furthermore, let D_t^w be the number of times that consumers purchase a product and d_t^w be the number of times that consumers continue viewing the list after purchasing a product, *i.e.*, $D_t^w \triangleq \sum_{u=1}^{t}\sum_{i=1}^{\Phi_t} \mathbb{I}\left[\gamma_{t,i} = 1\right]$ and $d_t^w \triangleq \sum_{u=1}^{t}\sum_{i=1}^{\Phi_t} \mathbb{I}\left[\gamma_{t,i} = 1\right]\mu_{t,i}$.

With these statistics, the parameters $\boldsymbol{\lambda}$, q, and $w = qs$ are estimated as follows.

$$\hat{\lambda}_{t,k} \triangleq c_{t,k}/C_{t,k}, \quad \hat{q}_t \triangleq d_t^q/D_t^q, \quad \hat{w}_t \triangleq d_t^w/D_t^w. \tag{6}$$

According to Hoeffding's inequality, we provide the following proposition that gives the error bound of our parameter estimation approach.

Proposition 5.2. *For any $t \geq 1$, with probability at least $1 - 2(N+2)t^{-2}$, the following holds:*

$$\left|\hat{\lambda}_{t,k} - \lambda_k\right| \leq \sqrt{\frac{\log t}{C_{t,k}}}, \forall k \in [N], \quad |\hat{q}_t - q| \leq \sqrt{\frac{\log t}{D_t^q}}, \quad \textit{and } |\hat{w}_t - qs| \leq \sqrt{\frac{\log t}{D_t^w}}. \tag{7}$$

B) Algorithm

Following the classic *optimism in the face of uncertainty principle* [1], we design a UCB-like algorithm that learns consumer behaviors and maximizes revenue simultaneously. In detail, according to Proposition 5.2, we learn the optimistic estimators of the parameters given as follows

$$\tilde{\lambda}_{t,k} \triangleq \min\left\{1, \hat{\lambda}_{t,k} + \sqrt{\frac{\log t}{C_{t,k}}}\right\}, \tilde{q}_t \triangleq \min\left\{1 - \epsilon_q, \hat{q}_t + \sqrt{\frac{\log t}{D_t^q}}\right\}, \tilde{w}_t \triangleq \min\left\{\tilde{q}_t, \hat{w}_t + \sqrt{\frac{\log t}{D_t^w}}\right\}. \tag{8}$$

As shown in Algorithm 1, when the t-th consumer arrives, the online retailer displays the products based on Theorem 4.1 with $\tilde{\boldsymbol{\lambda}}_{t-1}$, $\tilde{q}_{t-1}$, and $\tilde{w}_{t-1}/\tilde{q}_{t-1}$ shown in Line 4. Afterward, the online retailer updates the statistics according to consumers' feedback as shown in Lines 5 and 6. The estimators are then updated in Line 7.

Algorithm 1: Non-contextual online policy

Input: Products revenue $\boldsymbol{r}$ and hyper-parameter ϵ_q.
Initialization: $\tilde{\lambda}_{0,k} = 1$ for $k \in [N]$, $\tilde{q}_0 = 1 - \epsilon_q$, $\tilde{w}_0 = 1 - \epsilon_q$.
for *t=1:T* **do**
  Let $\boldsymbol{\sigma}_t$ be the optimal offline policy from Theorem 4.1 with $\boldsymbol{\lambda} = \tilde{\boldsymbol{\lambda}}_{t-1}$, $q = \tilde{q}_{t-1}$, and $s = \tilde{w}_{t-1}/\tilde{q}_{t-1}$.
  Offer ranking policy $\boldsymbol{\sigma}_t$ and observe Φ_t, $\boldsymbol{\Gamma}_t$, $\boldsymbol{\eta}_t$, $\boldsymbol{\mu}_t$.
  Update statistics $C_{t,k}$, $c_{t,k}$, D_t^q, d_t^q, D_t^w, and d_t^w.
  Calculate $\tilde{\lambda}_{t,k}$, $\tilde{q}_t$, and $\tilde{w}_t$ by Equations (6) and (8).
end

C) Regret Analysis

To theoretically analyze the performance of Algorithm 1, we first introduce the regret, which measures the total difference between the maximal revenue achieved by $\boldsymbol{\sigma}^*$ and the cumulative reward by an online algorithm after T rounds.

$$\mathrm{Reg}_T \triangleq \sum_{t=1}^{T} \mathrm{Revenue}\left(\boldsymbol{\sigma}^*; \boldsymbol{\lambda}, q, s\right) - \mathrm{Revenue}\left(\boldsymbol{\sigma}_t; \boldsymbol{\lambda}, q, s\right). \tag{9}$$

Reg_T is a random variable and the randomness comes from the uncertainty in consumers' behaviors. As a result, we focus on $\mathbb{E}\left[\mathrm{Reg}_T\right]$, which is a common routine in bandit literature [46]. The performance of the algorithm is given by the following theorem.

Theorem 5.3. *Under Assumption 3.1 (with parameter ϵ_q), for any $T > 1$, the regret achieved by Algorithm 1 is bounded by*

$$\mathbb{E}\left[\mathrm{Reg}_T\right] \leq C \cdot \frac{r_{\max}}{\epsilon_q} N \sqrt{T \log T}, \tag{10}$$

where C is a absolute constant (independent of problem parameters) and $r_{\max} = \max_{k \in [N]} r_k$.

Remark. We analyze different parameters' impacts on the regret bound in Equation (10). Firstly, the regret grows at $\tilde{O}(\sqrt{T})$, which is a standard result in online learning algorithms [46]. Secondly, the regret bound is linearly related to N, which is similar to the results in cascade bandit models [54] and revenue management literature [17]. Finally, the bound depends on $r_{\max}/\epsilon_q$, which measures the maximal expected revenue the online retailer could achieve from each consumer.

5.2 Contextual Online Setting

We consider a more realistic scenario where the consumers' behaviors are various and depend on their own features. In detail, let $x_t \in \mathbb{R}^{m_x}$ be the feature of the t-th consumer with dimension m_x and $y_{t,k} \in \mathbb{R}^{m_y}$ be the joint feature of the t-th consumer and k-th product with dimension m_y. $y_{t,k}$ can be the concatenation of consumer features and product features or fusion through a transformation such as a matrix multiplication. In this paper, we adopt concatenation for simplicity. Let $\|\cdot\|$ denote the ℓ_2 norm. Following [17], we consider the classic linear bandit setting [7, 21, 38] and the ground-truth data-generating process is given by the following assumptions.

Assumption 5.1. There exists constant $\beta_\lambda \in \mathbb{R}^{m_y}$, $\beta_q \in \mathbb{R}^{m_x}$, and $\beta_s \in \mathbb{R}^{m_x}$, such that

$$\lambda_{t,k} = y_{t,k}^\top \beta_\lambda, \forall k \in [N], \quad q_t = x_t^\top \beta_q, \quad \text{and} \quad s_t = x_t^\top \beta_s. \tag{11}$$

Assumption 5.2. For any $t > 0$ and $k \in [N]$, $\|x_t\| \leq 1$ and $\|y_{t,k}\| \leq 1$. In addition, there exists a constant $U > 0$ such that $\|\beta_\lambda\|, \|\beta_q\|, \|\beta_s\| \leq U$.

Remark. These two assumptions are common in revenue maximization and linear bandits [1, 46]. Assumption 5.1 can be satisfied with kernel functions for complex data-generating processes and Assumption 5.2 can be achieved by normalization. Thus the assumptions are rational and easily attainable in real applications.

A) Estimation of Parameters

For the estimation of $\boldsymbol{\lambda}_t$ and q_t, we use the ridge regression method to estimate β_λ and β_q. Given the observation of consumer's behaviors till the t-th timestamp, the estimation $\hat{\beta}_{t,\lambda}$, $\hat{\beta}_{t,q}$ is given by

$$\hat{\beta}_{t,\lambda} = \operatorname*{argmin}_{\beta \in \mathbb{R}^{m_y}} \sum_{u=1}^{t} \sum_{k=1}^{\Phi_u} \left(y_{u,\sigma_{u,k}}^\top \beta - \gamma_{u,k} \right)^2 + \alpha_\lambda \|\beta\|_2^2 = \Sigma_{t,\lambda}^{-1} \rho_{t,\lambda}, \tag{12}$$

$$\hat{\beta}_{t,q} = \operatorname*{argmin}_{\beta \in \mathbb{R}^{m_x}} \sum_{u=1}^{t} \sum_{k=1}^{\Phi_u} \mathbb{I}[\gamma_{u,k} = 0] \left(x_u^\top \beta - \eta_{u,k} \right)^2 + \alpha_q \|\beta\|_2^2 = \Sigma_{t,q}^{-1} \rho_{t,q}. \tag{13}$$

Here $\Sigma_{t,\lambda} = \sum_{u=1}^{t} \sum_{k=1}^{\Phi_u} y_{u,\sigma_{u,k}} y_{u,\sigma_{u,k}}^\top + \alpha_\lambda \mathbf{I}_{m_y}$, $\rho_{t,\lambda} = \sum_{u=1}^{t} \sum_{k=1}^{\Phi_u} \gamma_{u,k} y_{u,\sigma_{u,k}}$, $\Sigma_{t,q} = \sum_{u=1}^{t} \sum_{k=1}^{\Phi_u} \mathbb{I}[\gamma_{u,k} = 0] x_u x_u^\top + \alpha_q \mathbf{I}_{m_x}$, and $\rho_{t,q} = \sum_{u=1}^{t} \sum_{k=1}^{\Phi_u} \mathbb{I}[\gamma_{u,k} = 0] \eta_{u,k} x_u$. $\mathbf{I}_{m_x}$ and $\mathbf{I}_{m_y}$ are the identity matrices with dimension m_x and m_y, respectively.

For the estimation of s_t, similar to Section 5.1, we estimate $w_t = q_t s_t$ instead. Note that $w_t = q_t s_t = x_t^\top \beta_q \beta_s^\top x_t = \mathrm{vec}(x_t x_t^\top)^\top \mathrm{vec}(\beta_q \beta_s^\top)$, where $\mathrm{vec}(\cdot)$ denotes the vectorization operation that maps a matrix to a vector. As a result, define $z_t \triangleq \mathrm{vec}(x_t x_t^\top)$ and $\beta_w \triangleq \mathrm{vec}(\beta_q \beta_s^\top)$ and we can get $w_t = z_t^\top \beta_w$. Similarly, we estimate β_w by conducting ridge regression as follows

$$\hat{\beta}_{t,w} = \operatorname*{argmin}_{\beta \in \mathbb{R}^{m_x^2}} \sum_{u=1}^{t} \sum_{k=1}^{\Phi_u} \mathbb{I}[\gamma_{u,k} = 1] \left(z_u^\top \beta - \mu_{u,k} \right)^2 + \alpha_w \|\beta\|_2^2 = \Sigma_{t,w}^{-1} \rho_{t,w}. \tag{14}$$

Here $\Sigma_{t,w} = \sum_{u=1}^{t} \sum_{k=1}^{\Phi_u} \mathbb{I}[\gamma_{u,k} = 1] z_u z_u^\top + \alpha_w \mathbf{I}_{m_x^2}$ and $\rho_{t,w} = \sum_{u=1}^{t} \sum_{k=1}^{\Phi_u} \mathbb{I}[\gamma_{u,k} = 1] \mu_{u,k} z_u$. $\mathbf{I}_{m_x^2}$ is the identity matrix with dimension m_x^2.

Let $\|\beta\|_A = \sqrt{\beta^\top A \beta}$ for any positive definite matrix A. Proposition 5.4 gives the estimation error bound of the coefficients.

Proposition 5.4. *Under Assumptions 5.1 and 5.2 (with parameter U), for any $t \geq 1$, with probability at least $1 - 3t^{-2}$, the following holds:*

$$\left\| \hat{\beta}_{t,\lambda} - \beta_\lambda \right\|_{\Sigma_{t,\lambda}} \leq \tau(t, m_y, \alpha_\lambda), \left\| \hat{\beta}_{t,q} - \beta_q \right\|_{\Sigma_{t,q}} \leq \tau(t, m_x, \alpha_q), \textit{and } \left\| \hat{\beta}_{t,w} - \beta_w \right\|_{\Sigma_{t,w}} \leq \tau(t, m_x^2, \alpha_w). \tag{15}$$

Here $\tau(t, m, \alpha) = 1/2\sqrt{m \log(1 + tN/(m\alpha)) + 4 \log t} + \alpha^{1/2}(U+1)^2$.

B) Algorithm

Given the estimation of β_λ, β_q and β_w we estimate $\lambda_{t,k}$, q_t and w_t as follows.

$$\begin{aligned} \tilde{\lambda}_{t,k} &= \mathrm{Proj}_{[0,1]} \left(y_{t,k}^\top \hat{\beta}_{t-1,\lambda} + \tau(t-1, m_y, \alpha_\lambda) \|y_{t,k}\|_{\Sigma_{t-1,\lambda}^{-1}} \right) \\ \tilde{q}_t &= \mathrm{Proj}_{[0,1-\epsilon_q]} \left(x_t^\top \hat{\beta}_{t-1,q} + \tau(t-1, m_x, \alpha_q) \|x_t\|_{\Sigma_{t-1,q}^{-1}} \right) \\ \tilde{w}_t &= \mathrm{Proj}_{[0,\tilde{q}_t]} \left(z_t^\top \hat{\beta}_{t-1,w} + \tau(t-1, m_x^2, \alpha_w) \|z_t\|_{\Sigma_{t-1,w}^{-1}} \right) \end{aligned} \tag{16}$$

Here $\mathrm{Proj}_{[a,b]}(\cdot)$ is a function that projects the input into the interval $[a, b]$.

As shown in Algorithm 2, when the t-th consumer arrives, the online retailer observes the consumer features x_t and $y_{t,k}$ for $k \in [N]$ and calculates the intermediate feature z_t. Afterward, he calculates $\tilde{\lambda}_{t,k}$, $\tilde{q}_t$, and $\tilde{w}_t$ according to Equation (16) in Line 5. Then the ranking policy $\boldsymbol{\sigma}_t$ is given as shown in Line 6 and the statistics are updated in Lines 7 and 8. Finally, the parameters $\hat{\beta}_{t,\lambda}$, $\hat{\beta}_{t,q}$, and $\hat{\beta}_{t,w}$ are estimated in Line 9 for the next iteration.

C) Regret Analysis

We analyze the performance of Algorithm 2 by investigating its regret bound. For a sequence of T consumers with features $\boldsymbol{X}_T = \{x_t\}_{t=1}^T$ and $\boldsymbol{Y}_T = \{\{y_{t,k}\}_{k=1}^N\}_{t=1}^T$, the regret is given by

$$\mathrm{Reg}_T(\boldsymbol{X}_T, \boldsymbol{Y}_T) = \sum_{t=1}^{T} \mathrm{Revenue}\left(\boldsymbol{\sigma}_t^*; \boldsymbol{\lambda}_t, q_t, s_t\right) - \mathrm{Revenue}\left(\boldsymbol{\sigma}_t; \boldsymbol{\lambda}_t, q_t, s_t\right). \tag{17}$$

Algorithm 2: Contextual online policy

Input: Products revenue $\boldsymbol{r}$ and hyper-parameter ϵ_q, α_λ, α_q, and α_w.
Initialization: $\hat{\beta}_{0,\lambda} = 0$, $\hat{\beta}_{0,q} = 0$, and $\hat{\beta}_{0,w} = 0$.
for *t=1:T* **do**
  Observe consumer features x_t and $y_{t,k}$ for $k \in [N]$. Let $z_t = \text{vec}(x_t x_t^\top)$.
  Calculate $\tilde{\lambda}_{t,k}$, $\tilde{q}_t$, and $\tilde{w}_t$ according to Equation (16).
  Let $\boldsymbol{\sigma}_t$ be the optimal ranking policy from Theorem 4.1 with $\boldsymbol{\lambda} = \tilde{\boldsymbol{\lambda}}_t$, $q = \tilde{q}_t$, and $s = \tilde{w}_t/\tilde{q}_t$.
  Offer ranking policy $\boldsymbol{\sigma}_t$ and observe Φ_t, $\boldsymbol{\Gamma}_t$, $\boldsymbol{\eta}_t$, $\boldsymbol{\mu}_t$.
  Calculate statistics $\Sigma_{t,\lambda}$, $\rho_{t,\lambda}$, $\Sigma_{t,q}$, $\rho_{t,q}$, $\Sigma_{t,w}$, and $\rho_{t,w}$.
  Calculate estimated parameters $\hat{\beta}_{t,\lambda}$, $\hat{\beta}_{t,q}$, and $\hat{\beta}_{t,w}$ according to Equations (12), (13) and (14).
end

Similar to the non-contextual setting, $\text{Reg}_T(\boldsymbol{X}_T, \boldsymbol{Y}_T)$ is a random variable and we focus on $\mathbb{E}[\text{Reg}_T(\boldsymbol{X}_T, \boldsymbol{Y}_T)]$. As a result, the performance of the algorithm is given by the following theorem.

Theorem 5.5. *Under Assumptions 3.1 (with parameter ϵ_q), 5.1, and 5.2 (with parameter U), for any $T > 1$, the regret achieved by Algorithm 2 is bounded by*

$$\mathbb{E}[\text{Reg}_T(\boldsymbol{X}_T, \boldsymbol{Y}_T)] \leq C \cdot \frac{r_{\max}}{\epsilon_q} \left(\chi(T, m_y, \alpha_\lambda) + \chi(T, m_x, \alpha_q) + \chi(T, m_x^2, \alpha_w) \right). \tag{18}$$

Here C is a absolute constant (independent of problem parameters), $r_{\max} = \max_{k \in [N]} r_k$, and

$$\chi(T, m, \alpha) = \left(\frac{1}{2} \sqrt{m \log \left(1 + \frac{TN}{m\alpha} \right) + 4 \log T} + \alpha^{1/2} (U+1)^2 \right) \sqrt{\frac{TN^2 m \log \left(1 + \frac{TN}{m\alpha}\right)}{\alpha \log(1 + \alpha^{-1})}}. \tag{19}$$

Remark. We analyze different parameters' impacts on the regret bound in Equation (19). Firstly, the regret grows at $\tilde{O}(\sqrt{T})$, a standard result in online learning algorithms [46]. Secondly, the regret depends on $\tilde{O}(N)$, which shares a similar result with the revenue management literature [17]. Thirdly, similar to Theorem 5.3, the bound depends linearly on $r_{\max}/\epsilon_q$ that measures the maximal expected revenue the online retailer could achieve from each consumer. Fourthly, the bound depends on the dimensions of features (*i.e.*, m_x and m_y) by $\tilde{O}(\sqrt{m_y})$ and $\tilde{O}(m_x)$ because the feature dimension for the estimation of $w_t = q_t s_t$ is m_x^2 and then the result is consistent with [17, 54]. Finally, the bound depends on $O(U^2 + U)$ since $\|\beta_w\| \leq U^2$ and $\|\beta_q\|, \|\beta_\lambda\| \leq U$, which agrees with [17].

6 Experiments

In this section, we conduct experiments on synthetic data to verify the performances of the online ranking policies. Reports and discussions of more results including the expected revenue and the ratio over the expected revenue of the optimal ranking policy are in Section B.

Baselines We adopt the following baselines. Firstly, Cao and Sun [13] considered the single-purchase setting, which is the simplified version of our multiple-purchase setting. We denote it as **Single Purchase**. Secondly, we adopt the method in [39] (denoted as **Keep Viewing**), which assumes that consumers always keep viewing the ranking list after purchasing products. Furthermore, we implement two explore-then-exploit-based algorithms (denoted as **Explore Then Exploit A** and **Explore Then Exploit B**, respectively) to verify the advantages of our method that could balance exploration and exploitation.

6.1 Synthetic Data

Data-generating processes For the non-contextual setting, we consider $N = 50$ and 300 products and $T = 100,000$ consumers. The revenue for each product r_k is uniformly distributed in $[0, 1]$ and the purchase probabilities for each product are uniformly sampled from $[0, 0.3]$. We set $q = 0.9$ and $s = 0.5$ and 0.8 to test the performance of different models on characterizing consumers' behaviors with multiple purchases.

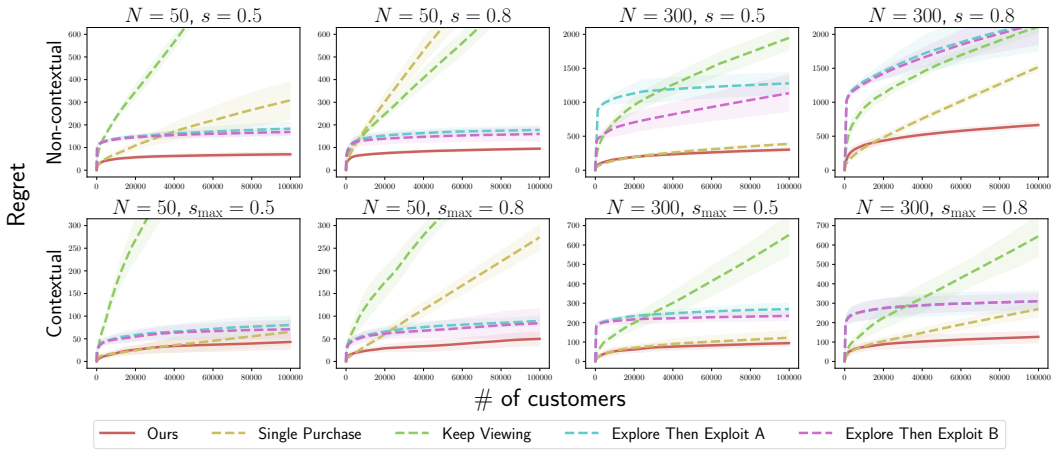

Figure 1: Experimental results on synthetic data. We plot the regret curves for different settings by varying $N = 50/300$ and $s/s_{\max} = 0.5/0.8$ in both non-contextual and contextual scenarios.

For the contextual setting, we consider $N = 50$ and 300 products and $T = 100,000$ consumers. The revenue r_k is also uniformly distributed in $[0, 1]$. Both the dimension of consumer features and product features are set to 5 (*i.e.*, $m_x = 5$). The joint feature $y_{t,k}$ (the feature of the t-th consumer and k-th product) is a concatenate of the corresponding consumer and product feature so that $m_y = 10$. The consumer features and product features are uniformly sampled from $[0.8/\sqrt{m_x}, 1/\sqrt{m_x}]$ and $[0, 1/\sqrt{m_x}]$. Afterward, the coefficients β_λ, β_q, and β_s in Equation (11) are uniformly sampled from $[0, 1]$. To ensure $\lambda_{t,k}$, q_t, s_t have similar ranges in the contextual setting, we normalize the coefficients such that the maximal elements in $\lambda_{t,k}$ and q_t (denoted as $\lambda_{\max}$ and $q_{\max}$) are 0.3 and 0.9, respectively. Similarly, by normalizing β_s, we set the maximal element in s_t (denoted as $s_{\max}$) to 0.5 and 0.8.

Results and analysis We implement the experiments for 5 distinct simulations by resampling consumers' behaviors while their basic characteristics remain unchanged. Specifically, in both the non-contextual and contextual settings, the purchase probabilities (*i.e.*, $\boldsymbol{\lambda}_t$) and parameters on the span attention and purchase budget (*i.e.*, q_t and s_t) are the same. However, the same consumer in distinct simulations may purchase different products due to randomness.

We then evaluate our method and baselines and present the regret via calculating the regret (according to Equations (9) and (17)). The results are shown in Figure 1 and our method outperforms all baselines in both settings with various parameters. On the one hand, the Single Purchase and Keep Viewing baseline considers different consumers' behavior settings and is not directly applicable here to the more realistic multiple-purchase setting. On the other hand, compared with the explore-then-exploit-based algorithms, our method integrates the advantages of traditional UCB-like algorithms, leading to a better exploration-exploitation trade-off for smaller regret. In addition, as introduced in Section B, Explore Then Exploit B incorporates learning processes during the exploration phase, leading to better performance compared with Explore Then Exploit A.

6.2 Semi-synthetic Data

We utilize the Ad Display / Click Data from Taobao[1], which contains ad display/purchase logs (26 million records) of 1,140,000 anonymized users from the website of Taobao for 8 days. Because the behavior logs only contain the purchase information of the users on a category of a brand, we view a (brand, category) pair in the dataset as a product and the average prices of the ads in this pair as the revenue for the product.

Because we can not obtain consumers' behaviors when we offer a new product ranking policy, following [14], we first estimate the parameters (*i.e.*, $\boldsymbol{\lambda}$, q, s in the non-contextual setting and β_λ, β_q, β_s in the contextual setting) with all of the data. Afterward, we use the estimated parameters to simulate consumers' behaviors when we provide them with different ranking lists of products.

[1] `https://tianchi.aliyun.com/datalab/dataSet.html?dataId=56`. License: CC BY-NC 4.0.

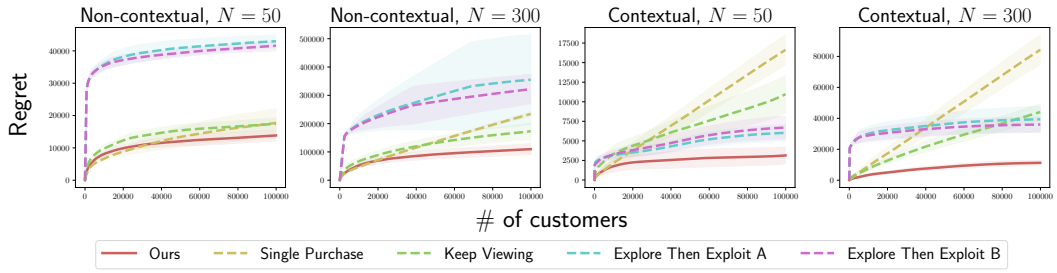

Figure 2: Experimental results on the semi-synthetic data. We plot the regret curves for different settings by varying $N = 50/300$ in both non-contextual and contextual scenarios.

Data processing Since each consumer may launch and leave the platform many times during the 8 days, we need to distinguish different launching activities of the same consumer. Specifically, we suppose that a consumer stops viewing the product list if it has been more than 10 minutes since her last interaction with the platform. Afterward, we can estimate the parameters in both non-contextual and contextual settings.

In the non-contextual setting, we calculate the statistics $C_{t,k}$, $c_{t,k}$, D_t^q, d_t^q, D_t^w, and d_t^w from the data and estimate $\boldsymbol{\lambda}$, q, s directly according to Lemma 5.1. The estimation is $q = 0.870$ and $s = 0.907$. Afterward, we choose $N = 50$ and 300 products with prices no more than 200 and purchase probabilities between $[0.05, 0.2]$. Similar to the synthetic data, we set $T = 100,000$.

In the contextual setting, we set the gender, age level, consumption grade, and shopping level as consumers' features and the number of ads, the logarithm of the average, standard deviation, maximum, and median of the ads' prices in the (brand, category) pair as products' features. With the statistics $C_{t,k}$, $c_{t,k}$, D_t^q, d_t^q, D_t^w, and d_t^w, we use the ridge regression with regularization strength 1 to estimate β_λ, β_q, and β_s. Afterward, we sample $N = 50$ and 300 products whose prices are no more than 200 and sample $T = 100,000$ consumers.

Results and analysis Similar to the synthetic experiments, we implement the experiments for 5 distinct simulations by resampling consumers' behaviors while their basic characteristics remain unchanged. The results of the regret are shown in Figure 2. As shown in the figure, our method outperforms all baselines in different scenarios. Firstly, the Single Purchase and Keep Viewing baselines are not directly applicable here since they consider different consumers' behavior settings. It is worth mentioning that the estimated s is large from the dataset. As a result, the consumers prefer to keep viewing products if they purchase any product, making the Keep Viewing baseline outperforms Single Purchase in this setting. Secondly, our method combines the benefits of existing UCB-like algorithms, resulting in a superior exploration-exploitation trade-off with less regret compared with explore-then-exploit-based methods. In addition, similar to the analysis in the synthetic experiments, Explore Then Exploit B outperforms Explore Then Exploit A in most cases.

7 Conclusion

To conclude, we propose a novel consumer choice model to deal with multiple-purchase activities of consumers in online scenarios. We characterize the optimal ranking policy and further design UCB-like algorithms with $\tilde{O}(\sqrt{T})$ regret based on them in both non-contextual and contextual online settings. We conduct extensive experiments to prove the effectiveness of our method.

Acknowledgments

Peng Cui's research was supported in part by National Key R&D Program of China (No. 2018AAA0102004, No. 2020AAA0106300), National Natural Science Foundation of China (No. U1936219, 62141607), Beijing Academy of Artificial Intelligence (BAAI). Bo Li's research was supported by the National Natural Science Foundation of China (No.72171131), the Tsinghua University Initiative Scientific Research Grant (No. 2019THZWJC11), Technology and Innovation Major Project of the Ministry of Science and Technology of China under Grants 2020AAA0108400 and 2020AAA0108403.

References

[1] Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. *Advances in neural information processing systems*, 24, 2011. 2, 5, 6, 25, 26

[2] Andrés Abeliuk, Gerardo Berbeglia, Manuel Cebrian, and Pascal Van Hentenryck. Assortment optimization under a multinomial logit model with position bias and social influence. *4OR*, 14 (1):57–75, 2016. 2

[3] Ashish Agarwal, Kartik Hosanagar, and Michael D Smith. Location, location, location: An analysis of profitability of position in online advertising markets. *Journal of marketing research*, 48(6):1057–1073, 2011. 1

[4] Ali Aouad and Danny Segev. Display optimization for vertically differentiated locations under multinomial logit preferences. *Management Science*, 67(6):3519–3550, 2021. 2

[5] Victor F Araman and René Caldentey. Dynamic pricing for nonperishable products with demand learning. *Operations research*, 57(5):1169–1188, 2009. 2

[6] Arash Asadpour, Rad Niazadeh, Amin Saberi, and Ali Shameli. Ranking an assortment of products via sequential submodular optimization. *arXiv preprint arXiv:2002.09458*, 2020. 3

[7] Peter Auer. Using confidence bounds for exploitation-exploration trade-offs. *Journal of Machine Learning Research*, 3(Nov):397–422, 2002. 6

[8] Michael R Baye, J Rupert J Gatti, Paul Kattuman, and John Morgan. Clicks, discontinuities, and firm demand online. *Journal of Economics & Management Strategy*, 18(4):935–975, 2009. 1

[9] Omar Besbes and Assaf Zeevi. Dynamic pricing without knowing the demand function: Risk bounds and near-optimal algorithms. *Operations Research*, 57(6):1407–1420, 2009. 2

[10] Omar Besbes and Assaf Zeevi. Blind network revenue management. *Operations research*, 60 (6):1537–1550, 2012. 2

[11] Josef Broder and Paat Rusmevichientong. Dynamic pricing under a general parametric choice model. *Operations Research*, 60(4):965–980, 2012. 2

[12] Juan José Miranda Bront, Isabel Méndez-Díaz, and Gustavo Vulcano. A column generation algorithm for choice-based network revenue management. *Operations research*, 57(3):769–784, 2009. 2

[13] Junyu Cao and Wei Sun. Dynamic learning of sequential choice bandit problem under marketing fatigue. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 3264–3271, 2019. 2, 3, 4, 8, 15

[14] Junyu Cao, Wei Sun, and Zuo-Jun Max Shen. Sequential choice bandits: Learning with marketing fatigue. *Available at SSRN 3355211*, 2019. 2, 3, 9, 15

[15] Junyu Cao, Wei Sun, Zuo-Jun Max Shen, and Markus Ettl. Fatigue-aware bandits for dependent click models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 3341–3348, 2020. 1, 2

[16] Olivier Chapelle and Ya Zhang. A dynamic bayesian network click model for web search ranking. In *Proceedings of the 18th international conference on World wide web*, pages 1–10, 2009. 1, 2

[17] Ningyuan Chen, Anran Li, and Shuoguang Yang. Revenue maximization and learning in products ranking. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, pages 316–317, 2021. 1, 2, 3, 6, 8, 15

[18] Yuxin Chen and Song Yao. Sequential search with refinement: Model and application with click-stream data. *Management Science*, 63(12):4345–4365, 2017. 1

[19] Wang Chi Cheung, Vincent Tan, and Zixin Zhong. A thompson sampling algorithm for cascading bandits. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 438–447. PMLR, 2019. 2

[20] Roberta Comunian, Caroline Chapain, and Nick Clifton. Location, location, location: exploring the complex relationship between creative industries and place. *Creative Industries Journal*, 3 (1):5–10, 2010. 1

[21] Varsha Dani, Thomas P Hayes, and Sham M Kakade. Stochastic linear optimization under bandit feedback. In *21st Annual Conference on Learning Theory*, pages 355–366, 2008. 6

[22] James M Davis, Guillermo Gallego, and Huseyin Topaloglu. Assortment optimization under variants of the nested logit model. *Operations Research*, 62(2):250–273, 2014. 2

[23] Arnoud V Den Boer. Dynamic pricing and learning: historical origins, current research, and new directions. *Surveys in operations research and management science*, 20(1):1–18, 2015. 2

[24] Mahsa Derakhshan, Negin Golrezaei, Vahideh Manshadi, and Vahab Mirrokni. Product ranking on online platforms. *Management Science*, 2022. 1

[25] Daria Dzyabura and John R Hauser. Recommending products when consumers learn their preference weights. *Marketing Science*, 38(3):417–441, 2019. 1

[26] Jacob Feldman and Danny Segev. Improved approximation schemes for mnl-driven sequential assortment optimization. *Available at SSRN 3440645*, 2019. 2

[27] Kris J Ferreira, Sunanda Parthasarathy, and Shreyas Sekar. Learning to rank an assortment of products. *Management Science*, 2021. 1, 2

[28] Alvaro Flores, Gerardo Berbeglia, and Pascal Van Hentenryck. Assortment optimization under the sequential multinomial logit model. *European Journal of Operational Research*, 273(3): 1052–1064, 2019. 2

[29] Guillermo Gallego, Anran Li, Van-Anh Truong, and Xinshang Wang. Approximation algorithms for product framing and pricing. *Operations Research*, 68(1):134–160, 2020. 2

[30] Pin Gao, Yuhang Ma, Ningyuan Chen, Guillermo Gallego, Anran Li, Paat Rusmevichientong, and Huseyin Topaloglu. Assortment optimization and pricing under the multinomial logit model with impatient customers: Sequential recommendation and selection. *Operations research*, 69 (5):1509–1532, 2021. 2

[31] Xiangyu Gao, Stefanus Jasin, Sajjad Najafi, and Huanan Zhang. Joint learning and optimization for multi-product pricing (and ranking) under a general cascade click model. *Management Science*, 2022. 1, 2

[32] Anindya Ghose and Sha Yang. An empirical analysis of search engine advertising: Sponsored search in electronic markets. *Management science*, 55(10):1605–1622, 2009. 1

[33] Fan Guo, Chao Liu, Anitha Kannan, Tom Minka, Michael Taylor, Yi-Min Wang, and Christos Faloutsos. Click chain model in web search. In *Proceedings of the 18th international conference on World wide web*, pages 11–20, 2009. 1, 2

[34] Fan Guo, Chao Liu, and Yi Min Wang. Efficient multiple-click models in web search. In *Proceedings of the second acm international conference on web search and data mining*, pages 124–131, 2009. 1

[35] Sumeet Katariya, Branislav Kveton, Csaba Szepesvari, and Zheng Wen. Dcm bandits: Learning to rank with multiple clicks. In *International Conference on Machine Learning*, pages 1215–1224. PMLR, 2016. 2

[36] Sergei Koulayev. Search for differentiated products: identification and estimation. *The RAND Journal of Economics*, 45(3):553–575, 2014. 1

[37] Branislav Kveton, Csaba Szepesvari, Zheng Wen, and Azin Ashkan. Cascading bandits: Learning to rank in the cascade model. In *International Conference on Machine Learning*, pages 767–776. PMLR, 2015. 1, 2

[38] Lihong Li, Wei Chu, John Langford, and Robert E Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*, pages 661–670, 2010. 6

[39] Yuan Liang, Chunlin Huang, Xiuguo Bao, and Ke Xu. Sequential dynamic event recommendation in event-based social networks: An upper confidence bound approach. *Information Sciences*, 542:1–23, 2021. 1, 2, 3, 4, 8, 15

[40] Nan Liu, Yuhang Ma, and Huseyin Topaloglu. Assortment optimization under the multinomial logit model with sequential offerings. *INFORMS Journal on Computing*, 32(3):835–853, 2020. 2

[41] Qing Liu and Neeraj Arora. Efficient choice designs for a consider-then-choose model. *Marketing Science*, 30(2):321–338, 2011. 1

[42] Siddharth Mahajan and Garrett Van Ryzin. Stocking retail assortments under dynamic consumer substitution. *Operations Research*, 49(3):334–351, 2001. 1, 2

[43] Microsoft. Neural Network Intelligence, 1 2021. URL `https://github.com/microsoft/nni`. 15

[44] Rad Niazadeh, Negin Golrezaei, Joshua R Wang, Fransisca Susan, and Ashwinkumar Badanidiyuru. Online learning via offline greedy algorithms: Applications in market design and optimization. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, pages 737–738, 2021. 3

[45] Herbert Robbins. Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58(5):527–535, 1952. 2

[46] Aleksandrs Slivkins. Introduction to multi-armed bandits. *arXiv preprint arXiv:1904.07272*, 2019. 6, 8

[47] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018. 2

[48] Theja Tulabandhula, Deeksha Sinha, and Prasoon Patidar. Multi-purchase behavior: Modeling and optimization. *arXiv preprint arXiv:2006.08055*, 2020. 2

[49] Raluca M Ursu. The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions. *Marketing Science*, 37(4):530–552, 2018. 1

[50] Mengmeng Wang, Xun Zhang, and Xiaolong Li. Multiple-purchase choice model: Estimation and optimization. *Available at SSRN 4147705*, 2022. 2

[51] Zheng Wen, Branislav Kveton, Michal Valko, and Sharan Vaswani. Online influence maximization under independent cascade model with semi-bandit feedback. *Advances in neural information processing systems*, 30, 2017. 26

[52] Renzhe Xu, Xingxuan Zhang, Peng Cui, Bo Li, Zheyan Shen, and Jiazheng Xu. Regulatory instruments for fair personalized pricing. In *Proceedings of the ACM Web Conference 2022*, pages 4–15, 2022. 1

[53] Heng Zhang, Hossein Piri, Woonghee Tim Huh, and Hongmin Li. Assortment optimization under multiple-discrete customer choices. *Available at SSRN 3988982*, 2021. 2

[54] Shi Zong, Hao Ni, Kenny Sung, Nan Rosemary Ke, Zheng Wen, and Branislav Kveton. Cascading bandits for large-scale recommendation problems. *arXiv preprint arXiv:1603.05359*, 2016. 2, 6, 8

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes] See Section A.
   (c) Did you discuss any potential negative societal impacts of your work? [Yes] See Section A.
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes] See Assumptions 3.1, 5.1, and 5.2.
   (b) Did you include complete proofs of all theoretical results? [Yes] See Section C.
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] See Section C.
   (b) Did you specify all the training details (e.g., data splits, hyper-parameters, how they were chosen)? [Yes] See Section B.
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes] See Figure 1 and Figure 2 in the main body and other figures in the appendix.
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] See Section B.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes] The dataset url is included in the paper.
   (b) Did you mention the license of the assets? [Yes] The license is included in the paper.
   (c) Did you include any new assets either in the supplemental material or as a URL? [No]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [Yes] It is a public dataset with the CC BY-NC 4.0 license.
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [Yes] The data has been anonymized.
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

A Limitations and Societal Impacts

One potential limitation is that we suppose the distribution of the attention span and purchase budget obey the geometric distribution. While the assumption is common [13, 39], we can relax it in future work, similar to [17, 14].

This paper does not have critical negative societal impacts because the purpose of the paper is to design an optimal ranking policy for online retailers. Besides the online retailers, consumers may also benefit because Equation (5) indicates that a product with a higher purchase probability may appear in a higher order in the ranking list. As a result, consumers may find the products shown in the ranking list highly relevant.

B More Experimental Results

B.1 Implementation Details

Baselines

• **Single Purchase** [13]. Cao and Sun [13] suppose that consumers always purchase at most one product. As a result, we implement them by setting $\tilde{s}_t = 0$ for all consumers. The estimations of $\tilde{\lambda}_{t,k}$ and $\tilde{q}_t$ are the same as our method.
• **Keep Viewing** [39]. Liang et al. [39] suppose that consumers always keep viewing the list if they purchase one product. To get the optimal ranking policy with the knowledge of consumers' characteristics, we can sort the products with value $(\lambda_k r_k)/(1-q-(1-q)\lambda_k)$ [39, Theorem 1]. Afterward, we use similar techniques to estimate $\tilde{\lambda}_{t,k}$ and $\tilde{q}_t$ and maximize the revenue in the online settings.
• **Explore Then Exploit A**. Following [13], we implement two explore-then-exploit-based algorithms as baselines. For the Explore Then Exploit A method, we first adopt an exploration stage to ensure that each product is displayed to consumers at least $\delta \log T$ times. (Here δ is a hyper-parameter.) To be specific, in the exploration stage, we rank the products according to their current numbers of displays. Afterward, we use the estimated parameters to provide the optimal ranking policy according to Theorem 4.1 in the main body of the paper.
• **Explore Then Exploit B**. This method is a variant of the Explore Then Exploit A method. In detail, in the exploration stage, after ranking the products according to their number of displays, we further rank the products that have been displayed to consumers at least $\delta \log T$ times according to Theorem 4.1 in the main body.

Hyper-parameters We add hyper-parameters ξ_λ, ξ_q, and ξ_w to determine the level of exploration for our method and two baselines (Single purchase and Keep viewing). The other two baselines do not include the hyper-parameter ξ because they do not have the dynamic exploration step. In detail, in the non-contextual setting, the parameters $\tilde{\lambda}_{t,k}$, $\tilde{q}_t$, and $\tilde{w}_t$ of our method are calculated as

$$\tilde{\lambda}_{t,k} \triangleq \min\left\{1, \hat{\lambda}_{t,k} + \xi_\lambda \sqrt{\frac{\log t}{C_{t,k}}}\right\}, \tilde{q}_t \triangleq \min\left\{1-\epsilon_q, \hat{q}_t + \xi_q \sqrt{\frac{\log t}{D_t^q}}\right\}, \tilde{w}_t \triangleq \min\left\{\tilde{q}_t, \hat{w}_t + \xi_w \sqrt{\frac{\log t}{D_t^w}}\right\}. \tag{20}$$

In the contextual setting, the parameters are also calculated according to the hyper-parameters ξ_λ, ξ_q, and ξ_w,

$$\begin{aligned}
\tilde{\lambda}_{t,k} &= \mathrm{Proj}_{[0,1]}\left(y_{t,k}^\top \hat{\beta}_{t-1,\lambda} + \xi_\lambda \cdot \tau(t-1, m_y, \alpha_\lambda) \left\|y_{t,k}\right\|_{\Sigma_{t-1,\lambda}^{-1}}\right), \\
\tilde{q}_t &= \mathrm{Proj}_{[0,1-\epsilon_q]}\left(x_t^\top \hat{\beta}_{t-1,q} + \xi_q \cdot \tau(t-1, m_x, \alpha_q) \left\|x_t\right\|_{\Sigma_{t-1,q}^{-1}}\right), \\
\tilde{w}_t &= \mathrm{Proj}_{[0,\tilde{q}_t]}\left(z_t^\top \hat{\beta}_{t-1,w} + \xi_w \cdot \tau(t-1, m_x^2, \alpha_w) \left\|z_t\right\|_{\Sigma_{t-1,w}^{-1}}\right).
\end{aligned} \tag{21}$$

For the synthetic dataset, we search the hyper-parameters $\xi_\lambda \in \{0.1, 0.3, 0.5\}$, $\xi_q, \xi_w \in \{0.05, 0.1, 0.2\}$, $\delta \in \{0.5, 1, 2, 5, 10\}$, and $\alpha_q = \alpha_w = \alpha_\lambda \in \{0.01, 0.1, 1\}$. For the semi-synthetic dataset, we search the hyper-parameters $\xi_\lambda \in \{0.1, 0.3, 0.5\}$, $\xi_q, \xi_w \in \{0.05, 0.1, 0.2\}$, $\delta \in \{0.5, 1, 2, 5, 10\}$, and $\alpha_q = \alpha_w = \alpha_\lambda \in \{0.1, 1, 10\}$. The grid search code is based on the NNI package [43].

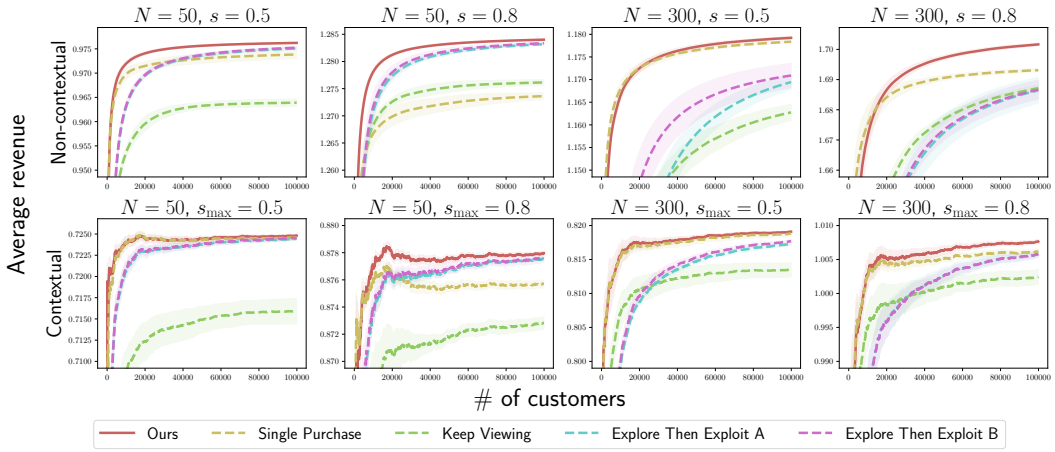

Figure 3: The average revenue metric on the synthetic data.

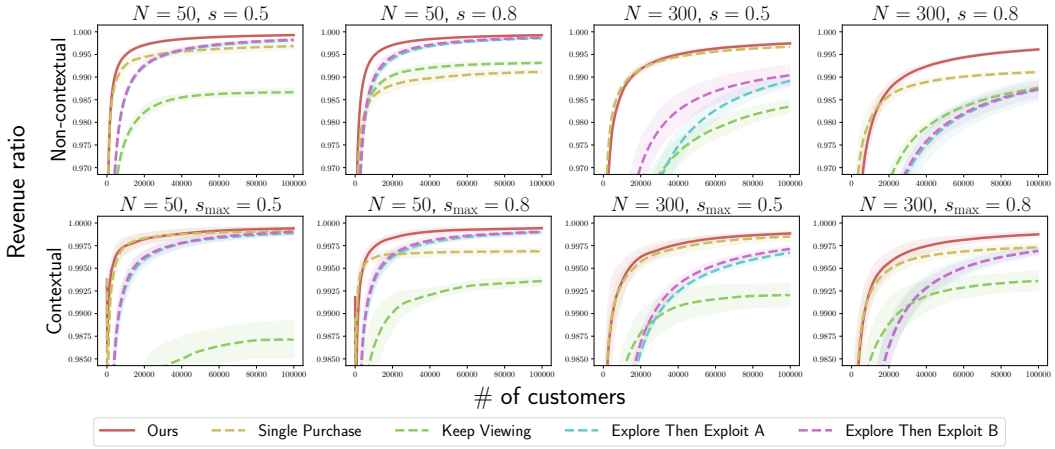

Figure 4: The revenue ratio metric on the synthetic data.

B.2 More Results on the Synthetic and Semi-synthetic Data

We further report two more metrics to verify the effectiveness of our method. In detail, Average revenue(T) measures the average revenue achieved by different algorithms till the T-th consumer. Revenue ratio(T) measures the ratio of the achieved revenue over the optimal ranking policy. Formally, the metrics are calculated as follows.

$$\begin{aligned} \text{Average revenue}(T) &= \frac{1}{T}\sum_{t=1}^{T} \text{Revenue}(\boldsymbol{\sigma}_t; \boldsymbol{\lambda}_t, q_t, s_t), \\ \text{Revenue ratio}(T) &= \frac{\sum_{t=1}^{T} \text{Revenue}(\boldsymbol{\sigma}_t; \boldsymbol{\lambda}_t, q_t, s_t)}{\sum_{t=1}^{T} \text{Revenue}(\boldsymbol{\sigma}_t^*; \boldsymbol{\lambda}_t, q_t, s_t)}. \end{aligned} \tag{22}$$

For both metrics, higher values indicate better performances. The results of these two metrics on both the synthetic and Semi-synthetic datasets are shown in Figures 3, 4, 5, and 6. These figures further demonstrate that our method outperforms all baselines in both contextual and non-contextual settings with various parameters (*i.e.*, s, $s_{\max}$, and N).

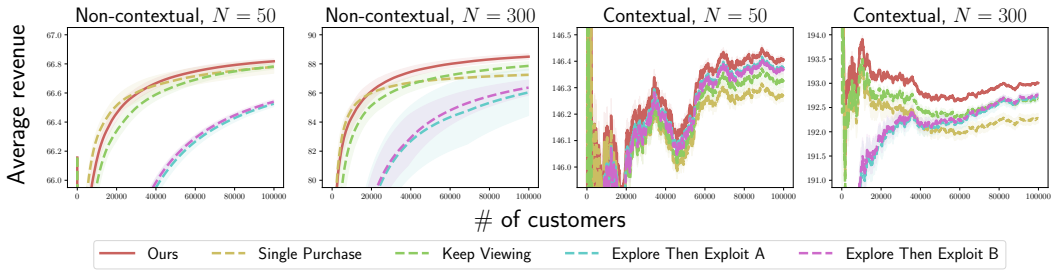

Figure 5: The average revenue metric on the semi-synthetic data.

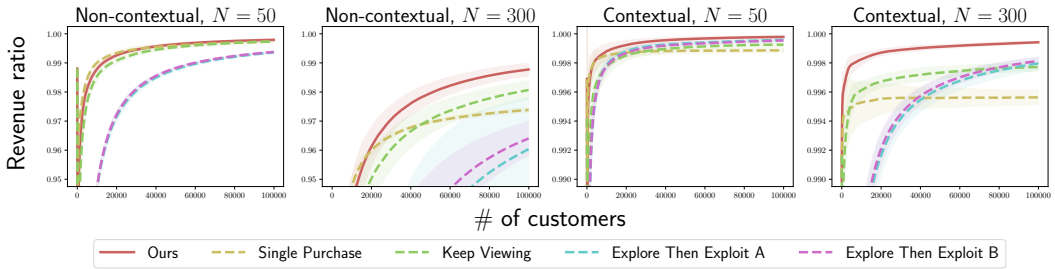

Figure 6: The revenue ratio metric on the semi-synthetic data.

C Omitted Proofs

C.1 Proof of Theorem 4.1

Proof. We drop all the subscript t for simplicity. We first write the expected revenue function.

$$
\begin{aligned}
&\text{Revenue}\left(\boldsymbol{\sigma}; \boldsymbol{\lambda}, q, s\right) \\
&= \mathbb{E}_{V\sim\text{Geo}(1-q), B\sim\text{Geo}(1-s)}\left[R(V, B; \boldsymbol{\lambda}, \boldsymbol{\sigma})\right] \\
&= \sum_{v=1}^{N}\sum_{b=1}^{N} P(V=v)P(B=b)\left(\sum_{Q\subseteq[v], |Q|\le b} P^{\text{buy}}(Q; v, b, \boldsymbol{\lambda}, \boldsymbol{\sigma})\left(\sum_{k\in Q} r_{\sigma_k}\right)\right) \\
&= \sum_{k=1}^{N} r_{\sigma_k}\left(\sum_{v=1}^{N}\sum_{b=1}^{N} P(V=v)P(B=b)\left(\sum_{Q\subseteq[N], k\in Q} P^{\text{buy}}(Q; v, b, \boldsymbol{\lambda}, \boldsymbol{\sigma})\right)\right) \\
&= \sum_{k=1}^{N} r_{\sigma_k}\left(\sum_{v=k}^{N}\sum_{b=1}^{N} P(V=v)P(B=b)\left(\sum_{Q\subseteq[N], k\in Q} P^{\text{buy}}(Q; v, b, \boldsymbol{\lambda}, \boldsymbol{\sigma})\right)\right).
\end{aligned}
\tag{23}
$$

For any $Q_k \subseteq [N]$ that satisfies $k \in Q_k$, define $Q_k^- \triangleq \{i \in Q_k : i < k\}$ and $Q_k^+ \triangleq \{i \in Q_k : i > k\}$. We further define

$$
\begin{cases}
G_k^-(Q_k^-) \triangleq \left(\prod_{i\in Q_k^-} \lambda_{\sigma_i}\right)\left(\prod_{1\le i<k, i\notin Q_k^-} (1-\lambda_{\sigma_i})\right), \\
G_k^+(Q_k^+) \triangleq \begin{cases} \left(\prod_{i\in Q_k^+} \lambda_{\sigma_i}\right)\left(\prod_{k<i<\max(Q_k^+)} (1-\lambda_{\sigma_i})\right), & \text{if } |Q_k^+| > 0, \\ 1, & \text{if } |Q_k^+| = 0. \end{cases}
\end{cases}
\tag{24}
$$

As a result,

$$
\begin{aligned}
&\sum_{v=k}^{N}\sum_{b=1}^{N} P(V=v)P(B=b)\left(\sum_{Q\subseteq[N],k\in Q} P^{\text{buy}}(Q;v,b,\boldsymbol{\lambda},\boldsymbol{\sigma})\right)\\
=&\sum_{v=k}^{N}\sum_{b=1}^{N} P(V=v)P(B=b)\left(\sum_{Q_k^-\subseteq[k-1]}\sum_{Q_k^+\subseteq[N]\setminus[k]} P^{\text{buy}}(Q;v,b,\boldsymbol{\lambda},\boldsymbol{\sigma})\right)\\
=&\lambda_{\sigma_k}\sum_{v=k}^{N}\sum_{b=1}^{N} P(V=v)P(B=b)\left(\sum_{Q_k^-\subseteq[k-1]}\sum_{Q_k^+\subseteq[N]\setminus[k]} G_k^-(Q_k^-)P_{k+1}^{\text{buy}}\left(Q_k^+;v,b-|Q_k^-|-1,\boldsymbol{\lambda},\boldsymbol{\sigma}\right)\right).
\end{aligned}
\tag{25}
$$

Here $P_{k+1}^{\text{buy}}\left(Q_k^+;v,b-|Q_k^-|-1,\boldsymbol{\lambda},\boldsymbol{\sigma}\right)$ in the second equality denotes the probability of purchasing Q_k^+ with attention span v purchase budget b after purchasing $|Q_k^-|+1$ products, *i.e.*,

$$
P_{k+1}^{\text{buy}}\left(Q_k^+;v,b-|Q_k^-|-1,\boldsymbol{\lambda},\boldsymbol{\sigma}\right)=\begin{cases}\left(\prod_{i\in Q_k^+}\lambda_{\sigma_i}\right)\left(\prod_{k<i\le v}(1-\lambda_{\sigma_i})\right), & \text{if } |Q_k^+|<b-|Q_k^-|-1.\\ \left(\prod_{i\in Q_k^+}\lambda_{\sigma_i}\right)\left(\prod_{k<i\le \max(Q_k^+)}(1-\lambda_{\sigma_i})\right), & \text{if } |Q_k^+|=b-|Q_k^-|-1.\\ 0, & \text{if } |Q_k^+|>b-|Q_k^-|-1.\end{cases}
\tag{26}
$$

As a result, when $|Q_k^-|$, v, b, $\boldsymbol{\lambda}$, and $\boldsymbol{\sigma}$ are fixed,

$$
\sum_{Q_k^+\subseteq[N]\setminus[k]} P_{k+1}^{\text{buy}}\left(Q_k^+;v,b-|Q_k^-|-1,\boldsymbol{\lambda},\boldsymbol{\sigma}\right)=1
\tag{27}
$$

Hence,

$$
\begin{aligned}
&\lambda_{\sigma_k}\sum_{v=k}^{N}\sum_{b=1}^{N} P(V=v)P(B=b)\left(\sum_{Q_k^-\subseteq[k-1]}\sum_{Q_k^+\subseteq[N]\setminus[k]} G_k^-(Q_k^-)P_{k+1}^{\text{buy}}\left(Q_k^+;v,b-|Q_k^-|-1,\boldsymbol{\lambda},\boldsymbol{\sigma}\right)\right)\\
=&\lambda_{\sigma_k}\sum_{v=k}^{N}\sum_{b=1}^{N} P(V=v)P(B=b)\left(\sum_{u=0}^{b-1}\left(\sum_{Q_k^-\subseteq[k-1],|Q_k^-|=u} G_k^-(Q_k^-)\right)\left(\sum_{Q_k^+\subseteq[N]\setminus[k]} P_{k+1}^{\text{buy}}\left(Q_k^+;v,b-u-1,\boldsymbol{\lambda},\boldsymbol{\sigma}\right)\right)\right)\\
=&\lambda_{\sigma_k}\sum_{v=k}^{N}\sum_{b=1}^{N} P(V=v)P(B=b)\left(\sum_{u=0}^{b-1}\left(\sum_{Q_k^-\subseteq[k-1],|Q_k^-|=u} G_k^-(Q_k^-)\right)\right)\\
=&\lambda_{\sigma_k}P(V\ge k)\sum_{b=1}^{N} P(B=b)\left(\sum_{u=0}^{b-1}\left(\sum_{Q_k^-\subseteq[k-1],|Q_k^-|=u} G_k^-(Q_k^-)\right)\right)\\
=&\lambda_{\sigma_k}P(V\ge k)\sum_{u=0}^{k-1} P(B\ge u+1)\left(\sum_{Q_k^-\subseteq[k-1],\left|Q_k^-\right|=u} G_k^-(Q_k^-)\right).
\end{aligned}
\tag{28}
$$

For any $k\ge 1$ and $0\le u\le k-1$, define

$$
H_{\boldsymbol{\sigma}}(k,u;\boldsymbol{\lambda})\triangleq\sum_{Q_k^-\subseteq[k-1],\left|Q_k^-\right|=u} G_k^-(Q_k^-).
\tag{29}
$$

As a result,

$$
\begin{aligned}
\text{Revenue}\left(\boldsymbol{\sigma};\boldsymbol{\lambda},q,s\right)=&\sum_{k=1}^{N} r_{\sigma_k}\lambda_{\sigma_k}P(V\ge k)\left(\sum_{u=0}^{k-1}P(B\ge u+1)H_{\boldsymbol{\sigma}}(k,u;\boldsymbol{\lambda})\right)\\
=&\sum_{k=1}^{N} r_{\sigma_k}\lambda_{\sigma_k}q^{k-1}\left(\sum_{u=0}^{k-1}s^u H_{\boldsymbol{\sigma}}(k,u;\boldsymbol{\lambda})\right)
\end{aligned}
\tag{30}
$$

In addition, it is easy to verify that $H_{\boldsymbol{\sigma}}(k,u;\boldsymbol{\lambda})$ has the following recurrence relation

$$
\begin{aligned}
&H_{\boldsymbol{\sigma}}(1,0;\boldsymbol{\lambda}) = 1.\\
&H_{\boldsymbol{\sigma}}(k,0;\boldsymbol{\lambda}) = \prod_{i=1}^{k-1}(1-\lambda_{\sigma_i}), \quad H_{\boldsymbol{\sigma}}(k,k-1;\boldsymbol{\lambda}) = \prod_{i=1}^{k-1}\lambda_{\sigma_i}, \qquad \forall k \geq 2.\\
&H_{\boldsymbol{\sigma}}(k,u;\boldsymbol{\lambda}) = \left(1-\lambda_{\sigma_{k-1}}\right) H_{\boldsymbol{\sigma}}(k-1,u;\boldsymbol{\lambda}) + \lambda_{\sigma_{k-1}} H_{\boldsymbol{\sigma}}(k-1,u-1;\boldsymbol{\lambda}), \quad \forall k \geq 2, 0 < u < k-1.
\end{aligned} \tag{31}
$$

From the definition of G_k^- and $H_{\boldsymbol{\sigma}}$, it is easy to check that the value of $H_{\boldsymbol{\sigma}}(k,u;\boldsymbol{\lambda})$ does not depend on the order of $\sigma_1, \sigma_2, \ldots, \sigma_{k-1}$. As a result, for any $1 \leq i < N$, let $\boldsymbol{\sigma}$ be a ranking policy that differs from $\boldsymbol{\sigma}^*$ only in the i-th and $(i+1)$-th position, *i.e.*,

$$
\sigma_j = \begin{cases} \sigma^*_{i+1}, & \text{if } j = i.\\ \sigma^*_i, & \text{if } j = i+1.\\ \sigma^*_j, & \text{otherwise.} \end{cases} \tag{32}
$$

Due to the optimality of $\boldsymbol{\sigma}^*$, we have

$$
\begin{aligned}
&0 \leq \text{Revenue}\left(\boldsymbol{\sigma}^*;\boldsymbol{\lambda},q,s\right) - \text{Revenue}\left(\boldsymbol{\sigma};\boldsymbol{\lambda},q,s\right)\\
&= r_{\sigma^*_i}\lambda_{\sigma^*_i} q^{i-1}\left(\sum_{u=0}^{i-1} s^u H_{\boldsymbol{\sigma}^*}(i,u;\boldsymbol{\lambda})\right) + r_{\sigma^*_{i+1}}\lambda_{\sigma^*_{i+1}} q^i \left(\sum_{u=0}^{i} s^u H_{\boldsymbol{\sigma}^*}(i+1,u;\boldsymbol{\lambda})\right)\\
&- r_{\sigma_i}\lambda_{\sigma_i} q^{i-1}\left(\sum_{u=0}^{i-1} s^u H_{\boldsymbol{\sigma}}(i,u;\boldsymbol{\lambda})\right) + r_{\sigma_{i+1}}\lambda_{\sigma_{i+1}} q^i \left(\sum_{u=0}^{i} s^u H_{\boldsymbol{\sigma}^*}(i+1,u;\boldsymbol{\lambda})\right)\\
&= r_{\sigma^*_i}\lambda_{\sigma^*_i} q^{i-1}\left(\sum_{u=0}^{i-1} s^u H_{\boldsymbol{\sigma}^*}(i,u;\boldsymbol{\lambda})\right) - r_{\sigma_i}\lambda_{\sigma_i} q^{i-1}\left(\sum_{u=0}^{i-1} s^u H_{\boldsymbol{\sigma}}(i,u;\boldsymbol{\lambda})\right)\\
&+ r_{\sigma^*_{i+1}}\lambda_{\sigma^*_{i+1}} q^i \left(\left(1-\lambda_{\sigma^*_i}\right)\sum_{u=0}^{i-1} s^u H_{\boldsymbol{\sigma}^*}(i,u;\boldsymbol{\lambda}) + \lambda_{\sigma^*_i}\sum_{u=1}^{i} s^u H_{\boldsymbol{\sigma}^*}(i,u-1;\boldsymbol{\lambda})\right)\\
&- r_{\sigma_{i+1}}\lambda_{\sigma_{i+1}} q^i \left((1-\lambda_{\sigma_i})\sum_{u=0}^{i-1} s^u H_{\boldsymbol{\sigma}}(i,u;\boldsymbol{\lambda}) + \lambda_{\sigma_i}\sum_{u=1}^{i} s^u H_{\boldsymbol{\sigma}}(i,u-1;\boldsymbol{\lambda})\right)\\
&= q^{i-1}\left(\sum_{u=0}^{i-1} s^u H_{\boldsymbol{\sigma}}(i,u;\boldsymbol{\lambda})\right)\cdot\\
&\quad \left(r_{\sigma^*_i}\lambda_{\sigma^*_i}\left(1-q+q(1-s)\lambda_{\sigma^*_{i+1}}\right) - r_{\sigma^*_{i+1}}\lambda_{\sigma^*_{i+1}}\left(1-q+q(1-s)\lambda_{\sigma^*_i}\right)\right).
\end{aligned} \tag{33}
$$

Hence,

$$
\frac{r_{\sigma^*_i}\lambda_{\sigma^*_i}}{1-q+q(1-s)\lambda_{\sigma^*_i}} \geq \frac{r_{\sigma^*_{i+1}}\lambda_{\sigma^*_{i+1}}}{1-q+q(1-s)\lambda_{\sigma^*_{i+1}}}. \tag{34}
$$

Now the claim follows. □

C.2 Proof of Lemma 5.1

Proof. The claim holds obviously for $\gamma_{t,k}$. For $\eta_{t,k}$, suppose the consumer purchased u products after viewing the k-th product, then

$$
\begin{aligned}
P\left(\eta_{t,k}=1\right) =& P\left(V_t > k, B_t > u \mid V_t \geq k, B_t > u\right)\\
=& \frac{P\left(V_t > k, B_t > u\right)}{P\left(V_t \geq k, B_t > u\right)} = \frac{P\left(V_t > k\right)P\left(B_t > u\right)}{P\left(V_t \geq k\right)P\left(B_t > u\right)} = \frac{q_t^k}{q_t^{k-1}} = q_t.
\end{aligned} \tag{35}
$$

As a result, $P\left(\eta_{t,k}=1\right)$ does not depend on the historical purchase activities and is always q_t. Hence, $\eta_{t,k}$ are *i.i.d.* Bernoulli random variables and $\eta_{t,k} \sim \text{Ber}(q_t)$.

Similarly, for $\mu_{t,k}$, suppose the consumer purchased u products after viewing the k-th product, then

$$
\begin{aligned}
P\left(\mu_{t,k}=1\right) =& P\left(V_t > k, B_t > u \mid V_t \geq k, B_t \geq u\right)\\
=& \frac{P\left(V_t > k, B_t > u\right)}{P\left(V_t \geq k, B_t \geq u\right)} = \frac{P\left(V_t > k\right)P\left(B_t > u\right)}{P\left(V_t \geq k\right)P\left(B_t \geq u\right)} = \frac{q_t^k s_t^u}{q_t^{k-1} s_t^{u-1}} = q_t s_t.
\end{aligned} \tag{36}
$$

As a result, $P(\eta_{t,k}=1)$ does not depend on the historical purchase activities. Hence, $\mu_{t,k}$ are *i.i.d.* Bernoulli random variables and $\mu_{t,k} \sim \mathrm{Ber}(q_t s_t)$. □

C.3 Proof of Proposition 5.2

Proof. According to Lemma 5.1 and Hoeffding's inequality, for any $t>0$ and $1 \le k \le N$, we have

$$\begin{aligned} P\left(\left|\hat{\lambda}_{t,k}-\lambda_k\right| \le \sqrt{\frac{\log t}{C_{t,k}}}\right) &\ge 1-2t^{-2}, \\ P\left(|\hat{q}_t - q| \le \sqrt{\frac{\log t}{D_t^q}}\right) &\ge 1-2t^{-2}, \\ P\left(|\hat{w}_t - qs| \le \sqrt{\frac{\log t}{D_t^w}}\right) &\ge 1-2t^{-2}. \end{aligned} \tag{37}$$

Now the claim follows. □

C.4 Proof of Theorem 5.3

We first define

$$O_{t,k}^{\lambda} \triangleq \mathbb{I}\left[k \le \Phi_t\right], \quad O_{t,k}^{q} \triangleq \mathbb{I}\left[k \le \Phi_t, \gamma_{t,k}=0\right], \quad O_{t,k}^{w} \triangleq \mathbb{I}\left[k \le \Phi_t, \gamma_{t,k}=1\right]. \tag{38}$$

We also need the following two propositions.

Proposition C.1. *Suppose $\tilde{\boldsymbol{\lambda}} \succeq \boldsymbol{\lambda}$, $\tilde{q} \ge q$, and $\tilde{s}\tilde{q} \ge sq$, then*

$$\mathrm{Revenue}\left(\boldsymbol{\sigma}^*; \boldsymbol{\lambda}, q, s\right) \le \mathrm{Revenue}\left(\boldsymbol{\sigma}^*; \tilde{\boldsymbol{\lambda}}, \tilde{q}, \tilde{s}\right). \tag{39}$$

Proof of Proposition C.1. For any $\boldsymbol{\sigma} \dashv [N]$, $1 \le n \le k$, $0 \le u \le k-n$, define

$$J_{\boldsymbol{\sigma}}(n,k,u;\boldsymbol{\lambda}) \triangleq \sum_{Q \subseteq [k-1]\setminus[n-1], |Q|=u} \left(\prod_{i \in Q} \lambda_{\sigma_i}\right)\left(\prod_{n \le i < k, i \notin Q} \lambda_{\sigma_i}\right). \tag{40}$$

As a result, $J_{\boldsymbol{\sigma}}(1,k,u;\boldsymbol{\lambda}) = H_{\boldsymbol{\sigma}}(k,u;\boldsymbol{\lambda})$. In addition, it is easy to check that

$$J_{\boldsymbol{\sigma}}(n,n,0;\boldsymbol{\lambda}) = 1, \quad \forall 1 \le n \le N.$$

$$J_{\boldsymbol{\sigma}}(n-1,k,u;\boldsymbol{\lambda}) = \begin{cases} (1-\lambda_{\sigma_{n-1}}) J_{\boldsymbol{\sigma}}(n,k,u;\boldsymbol{\lambda}), & \text{if } u=0. \\ \lambda_{\sigma_{n-1}} J_{\boldsymbol{\sigma}}(n,k,u-1;\boldsymbol{\lambda}), & \text{if } u=k-n+1. \\ (1-\lambda_{\sigma_{n-1}}) J_{\boldsymbol{\sigma}}(n,k,u;\boldsymbol{\lambda}) + \lambda_{\sigma_{n-1}} J_{\boldsymbol{\sigma}}(n,k,u-1;\boldsymbol{\lambda}), & \text{if } 0<u<k-n+1. \end{cases} \tag{41}$$

In addition, for any $n \ge 1$, define

$$R_{\boldsymbol{\sigma}}(n;\boldsymbol{\lambda},q,s) \triangleq \sum_{k=n}^{N} r_{\sigma_k} \lambda_{\sigma_k} q^{k-n} \left(\sum_{u=0}^{k-n} s^u J_{\boldsymbol{\sigma}}(n,k,u;\boldsymbol{\lambda})\right). \tag{42}$$

It is easy to verify that $R_{\boldsymbol{\sigma}}(1;\boldsymbol{\lambda},q,s) = \mathrm{Revenue}(\boldsymbol{\sigma};\boldsymbol{\lambda},q,s)$. Furthermore,

$$\begin{aligned} & R_{\boldsymbol{\sigma}}(n-1;\boldsymbol{\lambda},q,s) \\ =& \sum_{k=n-1}^{N} r_{\sigma_k} \lambda_{\sigma_k} q^{k-n+1} \left(\sum_{u=0}^{k-n+1} s^u J_{\boldsymbol{\sigma}}(n-1,k,u;\boldsymbol{\lambda})\right) \\ =& r_{\sigma_{n-1}} \lambda_{\sigma_{n-1}} + \sum_{k=n}^{N} r_{\sigma_k} \lambda_{\sigma_k} q^{k-n+1} \left(\sum_{u=0}^{k-n+1} s^u J_{\boldsymbol{\sigma}}(n-1,k,u;\boldsymbol{\lambda})\right) \\ =& r_{\sigma_{n-1}} \lambda_{\sigma_{n-1}} + \sum_{k=n}^{N} r_{\sigma_k} \lambda_{\sigma_k} q^{k-n+1} \left(\sum_{u=0}^{k-n} s^u (1-\lambda_{\sigma_{n-1}}) J_{\boldsymbol{\sigma}}(n,k,u;\boldsymbol{\lambda}) + \sum_{u=1}^{k-n+1} s^u \lambda_{\sigma_{n-1}} J_{\boldsymbol{\sigma}}(n,k,u-1;\boldsymbol{\lambda})\right) \\ =& r_{\sigma_{n-1}} \lambda_{\sigma_{n-1}} + q\left(1-(1-s)\lambda_{\sigma_{n-1}}\right) \left(\sum_{k=n}^{N} r_{\sigma_k} \lambda_{\sigma_k} q^{k-n} \left(\sum_{u=0}^{k-n} s^u J_{\boldsymbol{\sigma}}(n,k,u;\boldsymbol{\lambda})\right)\right) \\ =& r_{\sigma_{n-1}} \lambda_{\sigma_{n-1}} + \left(q - q(1-s)\lambda_{\sigma_{n-1}}\right) R_{\boldsymbol{\sigma}}(n;\boldsymbol{\lambda},q,s). \end{aligned} \tag{43}$$

Now we prove $R_{\boldsymbol{\sigma}^*}(n;\boldsymbol{\lambda},q,s) \le R_{\boldsymbol{\sigma}^*}(n;\tilde{\boldsymbol{\lambda}},\tilde{q},\tilde{s})$, $\forall n \in [N]$ by induction. Firstly, it is easy to check that $R_{\boldsymbol{\sigma}^*}(N;\boldsymbol{\lambda},q,s) \le R_{\boldsymbol{\sigma}^*}(N;\tilde{\boldsymbol{\lambda}},\tilde{q},\tilde{s})$. Suppose the condition holds for $R_{\boldsymbol{\sigma}^*}(n;\boldsymbol{\lambda},q,s) \le R_{\boldsymbol{\sigma}^*}(n;\tilde{\boldsymbol{\lambda}},\tilde{q},\tilde{s})$. Consider the case for $n-1$. Notice that $R_{\boldsymbol{\sigma}^*}(n-1;\boldsymbol{\lambda},q,s) \ge R_{\boldsymbol{\sigma}^*}(n;\boldsymbol{\lambda},q,s)$ (otherwise it will increase the revenue if removing the $(n-1)$-th product in the list), we can get that

$$\left(r_{\sigma_{n-1}} - q(1-s)R_{\boldsymbol{\sigma}^*}(n;\boldsymbol{\lambda},q,s)\right)\lambda_{\sigma_{n-1}} \ge (1-q)R_{\boldsymbol{\sigma}^*}(n;\boldsymbol{\lambda},q,s) \ge 0. \tag{44}$$

Hence $r_{\sigma_{n-1}} - q(1-s)R_{\boldsymbol{\sigma}^*}(n;\boldsymbol{\lambda},q,s) \ge 0$. As a result,

$$\begin{aligned}
&R_{\boldsymbol{\sigma}^*}(n-1;\boldsymbol{\lambda},q,s)\\
=&\, r_{\sigma_{n-1}}\lambda_{\sigma_{n-1}} + \left(q - q(1-s)\lambda_{\sigma_{n-1}}\right)R_{\boldsymbol{\sigma}^*}(n;\boldsymbol{\lambda},q,s)\\
=&\, qR_{\boldsymbol{\sigma}^*}(n;\boldsymbol{\lambda},q,s) + \left(r_{\sigma_{n-1}} - q(1-s)R_{\boldsymbol{\sigma}^*}(n;\boldsymbol{\lambda},q,s)\right)\lambda_{\sigma_{n-1}}\\
\le&\, qR_{\boldsymbol{\sigma}^*}(n;\boldsymbol{\lambda},q,s) + \left(r_{\sigma_{n-1}} - q(1-s)R_{\boldsymbol{\sigma}^*}(n;\boldsymbol{\lambda},q,s)\right)\tilde{\lambda}_{\sigma_{n-1}}\\
=&\, r_{\sigma_{n-1}}\tilde{\lambda}_{\sigma_{n-1}} + \left(q - q(1-s)\tilde{\lambda}_{\sigma_{n-1}}\right)R_{\boldsymbol{\sigma}^*}(n;\boldsymbol{\lambda},q,s)\\
=&\, r_{\sigma_{n-1}}\tilde{\lambda}_{\sigma_{n-1}} + \left(q\left(1-\tilde{\lambda}_{\sigma_{n-1}}\right) + qs\tilde{\lambda}_{\sigma_{n-1}}\right)R_{\boldsymbol{\sigma}^*}(n;\boldsymbol{\lambda},q,s)\\
\le&\, r_{\sigma_{n-1}}\tilde{\lambda}_{\sigma_{n-1}} + \left(\tilde{q}\left(1-\tilde{\lambda}_{\sigma_{n-1}}\right) + \tilde{q}\tilde{s}\tilde{\lambda}_{\sigma_{n-1}}\right)R_{\boldsymbol{\sigma}^*}(n;\boldsymbol{\lambda},q,s)\\
\le&\, r_{\sigma_{n-1}}\tilde{\lambda}_{\sigma_{n-1}} + \left(\tilde{q} - \tilde{q}(1-\tilde{s})\tilde{\lambda}_{\sigma_{n-1}}\right)R_{\boldsymbol{\sigma}^*}(n;\tilde{\boldsymbol{\lambda}},\tilde{q},\tilde{s}) = R_{\boldsymbol{\sigma}^*}(n-1;\tilde{\boldsymbol{\lambda}},\tilde{q},\tilde{s}).
\end{aligned} \tag{45}$$

As a result,

$$\text{Revenue}\left(\boldsymbol{\sigma}^*;\boldsymbol{\lambda},q,s\right) = R_{\boldsymbol{\sigma}^*}(1;\boldsymbol{\lambda},q,s) \le R_{\boldsymbol{\sigma}^*}(1;\tilde{\boldsymbol{\lambda}},\tilde{q},\tilde{s}) = \text{Revenue}\left(\boldsymbol{\sigma}^*;\tilde{\boldsymbol{\lambda}},\tilde{q},\tilde{s}\right). \tag{46}$$

□

Proposition C.2. *Under Assumption 3.1, suppose $\tilde{\boldsymbol{\lambda}} \succeq \boldsymbol{\lambda}$, $1-\epsilon_q \ge \tilde{q} \ge q$, and $\tilde{s}\tilde{q} \ge sq$, then*

$$\begin{aligned}
&\text{Revenue}\left(\boldsymbol{\sigma}_t;\tilde{\boldsymbol{\lambda}},\tilde{q},\tilde{s}\right) - \text{Revenue}\left(\boldsymbol{\sigma}_t;\boldsymbol{\lambda},q,s\right)\\
\le&\, \frac{r_{\max}}{\epsilon_q}\sum_{k=1}^{N}\left(|q-\tilde{q}|\,\mathbb{E}\left[O^q_{t,k}\right] + 3\left|\lambda_{\sigma_k} - \tilde{\lambda}_{\sigma_k}\right|\mathbb{E}\left[O^\lambda_{t,k}\right] + |qs - \tilde{q}\tilde{s}|\,\mathbb{E}\left[O^w_{t,k}\right]\right)
\end{aligned} \tag{47}$$

Proof of Proposition C.2. For simplicity, we use $\boldsymbol{\sigma}$ to denote $\boldsymbol{\sigma}_t$ in this proof. According to Equation (30)

$$\text{Revenue}\left(\boldsymbol{\sigma};\boldsymbol{\lambda},q,s\right) = \sum_{k=1}^{N} r_{\sigma_k}\lambda_{\sigma_k}q^{k-1}\left(\sum_{u=0}^{k-1} s^u H_{\boldsymbol{\sigma}}(k,u;\boldsymbol{\lambda})\right), \tag{48}$$

and $H_{\boldsymbol{\sigma}}(k,u;\boldsymbol{\lambda})$ is defined in Equation (29). As a result

$$\begin{aligned}
&\text{Revenue}\left(\boldsymbol{\sigma};\tilde{\boldsymbol{\lambda}},\tilde{q},\tilde{s}\right) - \text{Revenue}\left(\boldsymbol{\sigma};\boldsymbol{\lambda},q,s\right)\\
=&\sum_{k=1}^{N} r_{\sigma_k}\tilde{\lambda}_{\sigma_k}\tilde{q}^{k-1}\left(\sum_{u=0}^{k-1}\tilde{s}^u H_{\boldsymbol{\sigma}}(k,u;\tilde{\boldsymbol{\lambda}})\right) - \sum_{k=1}^{N} r_{\sigma_k}\lambda_{\sigma_k}q^{k-1}\left(\sum_{u=0}^{k-1} s^u H_{\boldsymbol{\sigma}}(k,u;\boldsymbol{\lambda})\right)\\
=&\sum_{k=1}^{N} r_{\sigma_k}\left(\tilde{\lambda}_{\sigma_k}\tilde{q}^{k-1}\left(\sum_{u=0}^{k-1}\tilde{s}^u H_{\boldsymbol{\sigma}}(k,u;\tilde{\boldsymbol{\lambda}})\right) - \lambda_{\sigma_k}q^{k-1}\left(\sum_{u=0}^{k-1} s^u H_{\boldsymbol{\sigma}}(k,u;\boldsymbol{\lambda})\right)\right).
\end{aligned} \tag{49}$$

Note that

$$\mathbb{E}\left[O^\lambda_{t,k}\right] = P(V_t \ge k)\sum_{u=0}^{k-1} P(B_t \ge u+1)H_{\boldsymbol{\sigma}}(k,u;\boldsymbol{\lambda}) = q^{k-1}\sum_{u=0}^{k-1} s^u H_{\boldsymbol{\sigma}}(k,u;\boldsymbol{\lambda}). \tag{50}$$

$$\mathbb{E}\left[O^q_{t,k}\right] = \mathbb{E}\left[O^\lambda_{t,k}\right]P\left(\gamma_{t,k}=0\right) = \left(1-\lambda_{\sigma_k}\right)q^{k-1}\sum_{u=0}^{k-1} s^u H_{\boldsymbol{\sigma}}(k,u;\boldsymbol{\lambda}). \tag{51}$$

$$\mathbb{E}\left[O^w_{t,k}\right] = \mathbb{E}\left[O^\lambda_{t,k}\right]P\left(\gamma_{t,k}=1\right) = \lambda_{\sigma_k}q^{k-1}\sum_{u=0}^{k-1} s^u H_{\boldsymbol{\sigma}}(k,u;\boldsymbol{\lambda}). \tag{52}$$

For any $1 \leq k \leq N$,

$$\begin{aligned}
&\left|\tilde{\lambda}_{\sigma_k}\tilde{q}^{k-1}\left(\sum_{u=0}^{k-1}\tilde{s}^u H_{\boldsymbol{\sigma}}(k,u;\tilde{\boldsymbol{\lambda}})\right) - \lambda_{\sigma_k} q^{k-1}\left(\sum_{u=0}^{k-1} s^u H_{\boldsymbol{\sigma}}(k,u;\boldsymbol{\lambda})\right)\right| \\
\leq &\left|\tilde{\lambda}_{\sigma_k} - \lambda_{\sigma_k}\right| \mathbb{E}\left[O_{t,k}^{\lambda}\right] + \tilde{\lambda}_{\sigma_k}\left|\tilde{q}^{k-1}\left(\sum_{u=0}^{k-1}\tilde{s}^u H_{\boldsymbol{\sigma}}(k,u;\tilde{\boldsymbol{\lambda}})\right) - q^{k-1}\left(\sum_{u=0}^{k-1} s^u H_{\boldsymbol{\sigma}}(k,u;\boldsymbol{\lambda})\right)\right|.
\end{aligned} \tag{53}$$

Here the first inequality is due to the triangle inequality of $|\cdot|$ and the second inequality is according to the fact that $\tilde{\lambda}_{\sigma_k} \leq 1$.

Now for any $1 \leq k \leq N$, define

$$A_{\boldsymbol{\sigma}}(k;q,s,\boldsymbol{\lambda}) \triangleq q^{k-1}\left(\sum_{u=0}^{k-1} s^u H_{\boldsymbol{\sigma}}(k,u;\boldsymbol{\lambda})\right). \tag{54}$$

We can get

$$\begin{aligned}
&A_{\boldsymbol{\sigma}}(k;q,s,\boldsymbol{\lambda}) \\
=&q^{k-1}\left(\sum_{u=0}^{k-1} s^u H_{\boldsymbol{\sigma}}(k,u;\boldsymbol{\lambda})\right) \\
=&q^{k-1}\left(\sum_{u=0}^{k-2} s^u \left(1-\lambda_{\sigma_{k-1}}\right) H_{\boldsymbol{\sigma}}(k-1,u;\boldsymbol{\lambda}) + \sum_{u=1}^{k-1} s^u \lambda_{\sigma_{k-1}} H_{\boldsymbol{\sigma}}(k-1,u-1;\boldsymbol{\lambda})\right) \\
=&q^{k-1}\left(\sum_{u=0}^{k-2} s^u \left(1-\lambda_{\sigma_{k-1}}\right) H_{\boldsymbol{\sigma}}(k-1,u;\boldsymbol{\lambda}) + s\sum_{u=0}^{k-2} s^u \lambda_{\sigma_{k-1}} H_{\boldsymbol{\sigma}}(k-1,u;\boldsymbol{\lambda})\right) \\
=&q\left(1-\lambda_{\sigma_{k-1}} + s\lambda_{\sigma_{k-1}}\right) A_{\boldsymbol{\sigma}}(k-1;q,s,\boldsymbol{\lambda}).
\end{aligned} \tag{55}$$

As a result, for any $1 < k \leq N$,

$$\begin{aligned}
&\left|A_{\boldsymbol{\sigma}}(k;q,s,\boldsymbol{\lambda}) - A_{\boldsymbol{\sigma}}(k;\tilde{q},\tilde{s},\tilde{\boldsymbol{\lambda}})\right| \\
=&\left|q\left(1-\lambda_{\sigma_{k-1}} + s\lambda_{\sigma_{k-1}}\right) A_{\boldsymbol{\sigma}}(k-1;q,s,\boldsymbol{\lambda}) - \tilde{q}\left(1-\tilde{\lambda}_{\sigma_{k-1}} + \tilde{s}\tilde{\lambda}_{\sigma_{k-1}}\right) A_{\boldsymbol{\sigma}}(k-1;\tilde{q},\tilde{s},\tilde{\boldsymbol{\lambda}})\right| \\
\leq&\left|q\left(1-\lambda_{\sigma_{k-1}} + s\lambda_{\sigma_{k-1}}\right) - \tilde{q}\left(1-\tilde{\lambda}_{\sigma_{k-1}} + \tilde{s}\tilde{\lambda}_{\sigma_{k-1}}\right)\right| A_{\boldsymbol{\sigma}}(k-1;q,s,\boldsymbol{\lambda}) \\
&+\tilde{q}\left(1-\tilde{\lambda}_{\sigma_{k-1}} + \tilde{s}\tilde{\lambda}_{\sigma_{k-1}}\right)\left|A_{\boldsymbol{\sigma}}(k-1;q,s,\boldsymbol{\lambda}) - A_{\boldsymbol{\sigma}}(k-1;\tilde{q},\tilde{s},\tilde{\boldsymbol{\lambda}})\right|
\end{aligned} \tag{56}$$

Here the last inequality is due the triangle inequality of $|\cdot|$. Note that

$$\begin{aligned}
&\left|q\left(1-\lambda_{\sigma_{k-1}} + s\lambda_{\sigma_{k-1}}\right) - \tilde{q}\left(1-\tilde{\lambda}_{\sigma_{k-1}} + \tilde{s}\tilde{\lambda}_{\sigma_{k-1}}\right)\right| A_{\boldsymbol{\sigma}}(k-1;q,s,\boldsymbol{\lambda}) \\
\leq&\left(\left|q\left(1-\lambda_{\sigma_{k-1}}\right) - \tilde{q}\left(1-\tilde{\lambda}_{\sigma_{k-1}}\right)\right| + \left|qs\lambda_{\sigma_{k-1}} - \tilde{q}\tilde{s}\tilde{\lambda}_{\sigma_{k-1}}\right|\right) A_{\boldsymbol{\sigma}}(k-1;q,s,\boldsymbol{\lambda}) \\
\leq&\left(|q-\tilde{q}|\left(1-\lambda_{\sigma_{k-1}}\right) + \tilde{q}\left|\lambda_{\sigma_{k-1}} - \tilde{\lambda}_{\sigma_{k-1}}\right| + |qs-\tilde{q}\tilde{s}|\,\lambda_{\sigma_{k-1}} + \tilde{q}\tilde{s}\left|\lambda_{\sigma_{k-1}} - \tilde{\lambda}_{\sigma_{k-1}}\right|\right) A_{\boldsymbol{\sigma}}(k-1;q,s,\boldsymbol{\lambda}) \\
\leq&\left(|q-\tilde{q}|\left(1-\lambda_{\sigma_{k-1}}\right) + 2\left|\lambda_{\sigma_{k-1}} - \tilde{\lambda}_{\sigma_{k-1}}\right| + |qs-\tilde{q}\tilde{s}|\,\lambda_{\sigma_{k-1}}\right) A_{\boldsymbol{\sigma}}(k-1;q,s,\boldsymbol{\lambda}) \\
=&|q-\tilde{q}|\,\mathbb{E}\left[O_{t,k-1}^{q}\right] + 2\left|\lambda_{\sigma_{k-1}} - \tilde{\lambda}_{\sigma_{k-1}}\right|\mathbb{E}\left[O_{t,k-1}^{\lambda}\right] + |qs-\tilde{q}\tilde{s}|\,\mathbb{E}\left[O_{t,k-1}^{w}\right].
\end{aligned} \tag{57}$$

Here the first two inequalities are due to the triangle inequality of $|\cdot|$ and the third inequality is due to that fact that $\tilde{q}, \tilde{s} \leq 1$. Combining Equations (56) and (57), we can get

$$\begin{aligned}
&\left|A_{\boldsymbol{\sigma}}(k; q, s, \boldsymbol{\lambda}) - A_{\boldsymbol{\sigma}}(k; \tilde{q}, \tilde{s}, \tilde{\boldsymbol{\lambda}})\right| \\
\leq &\sum_{u=1}^{k-1} \left(|q-\tilde{q}| \mathbb{E}\left[O_{t,u}^{q}\right] + 2\left|\lambda_{\sigma_u} - \tilde{\lambda}_{\sigma_u}\right| \mathbb{E}\left[O_{t,u}^{\lambda}\right] + |qs - \tilde{q}\tilde{s}| \mathbb{E}\left[O_{t,u}^{w}\right]\right) \left(\prod_{i=u+1}^{k-1} \tilde{q}\left(1 - \tilde{\lambda}_{\sigma_i} + \tilde{s}\tilde{\lambda}_{\sigma_i}\right)\right) \\
\leq &\sum_{u=1}^{k-1} \left(|q-\tilde{q}| \mathbb{E}\left[O_{t,u}^{q}\right] + 2\left|\lambda_{\sigma_u} - \tilde{\lambda}_{\sigma_u}\right| \mathbb{E}\left[O_{t,u}^{\lambda}\right] + |qs - \tilde{q}\tilde{s}| \mathbb{E}\left[O_{t,u}^{w}\right]\right) \left(\prod_{i=u+1}^{k-1} \tilde{q}\right) \\
= &\sum_{u=1}^{k-1} \left(|q-\tilde{q}| \mathbb{E}\left[O_{t,u}^{q}\right] + 2\left|\lambda_{\sigma_u} - \tilde{\lambda}_{\sigma_u}\right| \mathbb{E}\left[O_{t,u}^{\lambda}\right] + |qs - \tilde{q}\tilde{s}| \mathbb{E}\left[O_{t,u}^{w}\right]\right) \tilde{q}^{k-u-1}.
\end{aligned} \tag{58}$$

Combining the above inequality with Equation (49) and (53), we can get that

$$\begin{aligned}
&\text{Revenue}\left(\boldsymbol{\sigma}; \tilde{\boldsymbol{\lambda}}, \tilde{q}, \tilde{s}\right) - \text{Revenue}\left(\boldsymbol{\sigma}; \boldsymbol{\lambda}, q, s\right) \\
\leq &\sum_{k=1}^{N} r_{\max} \left|\tilde{\lambda}_{\sigma_k} - \lambda_{\sigma_k}\right| \mathbb{E}\left[O_{t,k}^{\lambda}\right] \\
&+ \sum_{k=2}^{N} r_{\sigma_k} \tilde{\lambda}_{\sigma_k} \left(\sum_{u=1}^{k-1} \left(|q-\tilde{q}| \mathbb{E}\left[O_{t,u}^{q}\right] + 2\left|\lambda_{\sigma_u} - \tilde{\lambda}_{\sigma_u}\right| \mathbb{E}\left[O_{t,u}^{\lambda}\right] + |qs - \tilde{q}\tilde{s}| \mathbb{E}\left[O_{t,u}^{w}\right]\right) \tilde{q}^{k-u-1}\right) \\
= &\sum_{k=1}^{N} r_{\max} \left|\tilde{\lambda}_{\sigma_k} - \lambda_{\sigma_k}\right| \mathbb{E}\left[O_{t,k}^{\lambda}\right] \\
&+ \sum_{k=1}^{N-1} \left(|q-\tilde{q}| \mathbb{E}\left[O_{t,k}^{q}\right] + 2\left|\lambda_{\sigma_k} - \tilde{\lambda}_{\sigma_k}\right| \mathbb{E}\left[O_{t,k}^{\lambda}\right] + |qs - \tilde{q}\tilde{s}| \mathbb{E}\left[O_{t,k}^{w}\right]\right) \left(\sum_{i=k+1}^{N} r_{\sigma_i} \tilde{\lambda}_{\sigma_i} \tilde{q}^{i-k-1}\right) \\
\leq &\sum_{k=1}^{N} r_{\max} \left|\tilde{\lambda}_{\sigma_k} - \lambda_{\sigma_k}\right| \mathbb{E}\left[O_{t,k}^{\lambda}\right] \\
&+ \sum_{k=1}^{N-1} r_{\max} \left(|q-\tilde{q}| \mathbb{E}\left[O_{t,k}^{q}\right] + 2\left|\lambda_{\sigma_k} - \tilde{\lambda}_{\sigma_k}\right| \mathbb{E}\left[O_{t,k}^{\lambda}\right] + |qs - \tilde{q}\tilde{s}| \mathbb{E}\left[O_{t,k}^{w}\right]\right) \left(\sum_{i=k+1}^{N} \tilde{q}^{i-k-1}\right) \\
\leq &\sum_{k=1}^{N} r_{\max} \left|\tilde{\lambda}_{\sigma_k} - \lambda_{\sigma_k}\right| \mathbb{E}\left[O_{t,k}^{\lambda}\right] \\
&+ \sum_{k=1}^{N-1} r_{\max} \left(|q-\tilde{q}| \mathbb{E}\left[O_{t,k}^{q}\right] + 2\left|\lambda_{\sigma_k} - \tilde{\lambda}_{\sigma_k}\right| \mathbb{E}\left[O_{t,k}^{\lambda}\right] + |qs - \tilde{q}\tilde{s}| \mathbb{E}\left[O_{t,k}^{w}\right]\right) \frac{1}{\epsilon_q} \\
\leq &\frac{r_{\max}}{\epsilon_q} \sum_{k=1}^{N} \left(|q-\tilde{q}| \mathbb{E}\left[O_{t,k}^{q}\right] + 3\left|\lambda_{\sigma_k} - \tilde{\lambda}_{\sigma_k}\right| \mathbb{E}\left[O_{t,k}^{\lambda}\right] + |qs - \tilde{q}\tilde{s}| \mathbb{E}\left[O_{t,k}^{w}\right]\right).
\end{aligned} \tag{59}$$

Now the claim follows. □

Proof of Theorem 5.3. For each timestamp t, define the favorable event as

$$\xi_t \triangleq \left\{\tilde{q}_t - 2\sqrt{\frac{\log t}{D_t^q}} \leq q \leq \tilde{q}_t, \tilde{w}_t - 2\sqrt{\frac{\log t}{D_t^w}} \leq qs \leq \tilde{w}_t, \tilde{\lambda}_{t,k} - 2\sqrt{\frac{\log t}{C_{t,k}}} \leq \lambda_k \leq \tilde{\lambda}_{t,k}, \forall k \in [N]\right\}. \tag{60}$$

According to Proposition 5.2, we have $P\left(\xi_t\right) \geq 1 - 2(N+2)t^{-2}$. Let ξ_t^C denote the complement event of ξ_t.

$$\begin{aligned}
\mathbb{E}[\text{Reg}_T] = &\sum_{t=1}^{T} \mathbb{E}\left[\left(\text{Revenue}\left(\boldsymbol{\sigma}^*; \boldsymbol{\lambda}, q, s\right) - \text{Revenue}\left(\boldsymbol{\sigma}_t; \boldsymbol{\lambda}, q, s\right)\right) \mathbb{I}[\xi_{t-1}]\right] \\
&+ \sum_{t=1}^{T} \mathbb{E}\left[\left(\text{Revenue}\left(\boldsymbol{\sigma}^*; \boldsymbol{\lambda}, q, s\right) - \text{Revenue}\left(\boldsymbol{\sigma}_t; \boldsymbol{\lambda}, q, s\right)\right) \mathbb{I}\left[\xi_{t-1}^C\right]\right].
\end{aligned} \tag{61}$$

On the one hand, consider the first term in the RHS of Equation (61), when the event ξ_{t-1} is satisfied, we have $q \leq \tilde{q}_{t-1}$, $z \leq \tilde{w}_{t-1}$, and $\lambda_k \leq \tilde{\lambda}_k, \forall k \in [N]$. As a result, according to Proposition C.1,

$$\text{Revenue}\left(\boldsymbol{\sigma}^*; \boldsymbol{\lambda}, q, s\right) \leq \text{Revenue}\left(\boldsymbol{\sigma}^*; \tilde{\boldsymbol{\lambda}}_{t-1}, \tilde{q}_{t-1}, \tilde{s}_{t-1}\right) \leq \text{Revenue}\left(\boldsymbol{\sigma}_t; \tilde{\boldsymbol{\lambda}}_{t-1}, \tilde{q}_{t-1}, \tilde{s}_{t-1}\right). \tag{62}$$

Hence,

$$\begin{aligned}
&\sum_{t=1}^{T} \mathbb{E}\left[\left(\text{Revenue}\left(\boldsymbol{\sigma}^*; \boldsymbol{\lambda}, q, s\right) - \text{Revenue}\left(\boldsymbol{\sigma}_t; \boldsymbol{\lambda}, q, s\right)\right) \mathbb{I}[\xi_{t-1}]\right] \\
\leq &\sum_{t=1}^{T} \mathbb{E}\left[\left(\text{Revenue}\left(\boldsymbol{\sigma}_t; \tilde{\boldsymbol{\lambda}}_{t-1}, \tilde{q}_{t-1}, \tilde{s}_{t-1}\right) - \text{Revenue}\left(\boldsymbol{\sigma}_t; \boldsymbol{\lambda}, q, s\right)\right) \mathbb{I}[\xi_{t-1}]\right] \\
\leq &\frac{r_{\max}}{\epsilon_q} \sum_{t=1}^{T} \mathbb{E}\left[\sum_{k=1}^{N} \left(|q - \tilde{q}_{t-1}| \, \mathbb{E}\left[O_{t,k}^q\right] + 3\left|\lambda_{\sigma_{t,k}} - \tilde{\lambda}_{t-1,\sigma_{t,k}}\right| \mathbb{E}\left[O_{t,k}^\lambda\right] + |qs - \tilde{q}_{t-1}\tilde{s}_{t-1}| \, \mathbb{E}\left[O_{t,k}^w\right]\right) \mathbb{I}[\xi_{t-1}]\right] \\
\leq &\frac{2r_{\max}}{\epsilon_q} \sum_{t=1}^{T} \mathbb{E}\left[\sum_{k=1}^{N} \left(\max\left\{\sqrt{\frac{\log t}{D_{t-1}^q}}, 1\right\} O_{t,k}^q + 3\max\left\{\sqrt{\frac{\log t}{C_{t-1,\sigma_{t,k}}}}, 1\right\} O_{t,k}^\lambda + \max\left\{\sqrt{\frac{\log t}{D_{t-1}^w}}, 1\right\} O_{t,k}^w\right) \mathbb{I}[\xi_{t-1}]\right] \\
\leq &\frac{2r_{\max}\sqrt{\log T}}{\epsilon_q} \sum_{t=1}^{T} \mathbb{E}\left[\sum_{k=1}^{N} \left(\max\left\{\sqrt{\frac{1}{D_{t-1}^q}}, 1\right\} O_{t,k}^q + 3\max\left\{\sqrt{\frac{1}{C_{t-1,\sigma_{t,k}}}}, 1\right\} O_{t,k}^\lambda + \max\left\{\sqrt{\frac{1}{D_{t-1}^w}}\right\} O_{t,k}^w\right)\right]
\end{aligned} \tag{63}$$

Let $\chi_t^V \triangleq \sum_{k=1}^{N} O_{t,k}^q \leq N$ and we have $D_{t,v} = D_{t-1}^V + \chi_t^V$. As a result,

$$\begin{aligned}
&\mathbb{E}\left[\sum_{t=1}^{T}\sum_{k=1}^{N} \max\left\{\sqrt{\frac{1}{D_{t-1}^q}}, 1\right\} O_{t,k}^q\right] \\
=&\mathbb{E}\left[\sum_{t=1}^{T} \max\left\{\sqrt{\frac{1}{D_{t-1}^q}}, 1\right\} \chi_t^V\right] \\
\leq&1 + N + \sum_{t=2}^{T} N\sqrt{\frac{1}{(t-1)N}} \\
\leq&1 + N + \sqrt{N}\left(1 + \int_{t=1}^{T-1} \sqrt{\frac{1}{t}}\mathrm{d}t\right) \leq C_1 N + C_2\sqrt{TN},
\end{aligned} \tag{64}$$

for some constants C_1 and C_2. Similarly, we have

$$\mathbb{E}\left[\sum_{t=1}^{T}\sum_{k=1}^{N} \max\left\{\sqrt{\frac{1}{D_{t-1}^w}}, 1\right\} O_{t,k}^q\right] \leq C_1 N + C_2\sqrt{TN}. \tag{65}$$

In addition, let $o_{t,k}^\lambda \triangleq O_{t,\sigma_{t,k}^{-1}}^\lambda$ represent whether the product k is viewed in the t-th timestamp. As a result, $C_{t,k} = C_{t-1,k} + o_{t,k}^\lambda$.

$$\begin{aligned}
&\mathbb{E}\left[\sum_{t=1}^{T}\sum_{k=1}^{N} \max\left\{\sqrt{\frac{1}{C_{t-1,\sigma_{t,k}}}}, 1\right\} O_{t,k}^\lambda\right] \\
=&\mathbb{E}\left[\sum_{k=1}^{N}\sum_{t=1}^{T} \max\left\{\sqrt{\frac{1}{C_{t-1,k}}}, 1\right\} o_{t,k}^\lambda\right] \\
\leq&\mathbb{E}\left[\sum_{k=1}^{N}\left(1 + \sum_{t=1}^{T}\sqrt{\frac{1}{t}}\right)\right] \leq C_3 N\sqrt{T},
\end{aligned} \tag{66}$$

for a constant C_3. On the other hand,

$$\begin{aligned}
&\sum_{t=1}^{T} \mathbb{E}\left[\left(\text{Revenue}\left(\boldsymbol{\sigma}^*; \boldsymbol{\lambda}, q, s\right) - \text{Revenue}\left(\boldsymbol{\sigma}_t; \boldsymbol{\lambda}, q, s\right)\right) \mathbb{I}\left[\xi_{t-1}^C\right]\right] \\
\leq&\text{Revenue}\left(\boldsymbol{\sigma}^*; \boldsymbol{\lambda}, q, s\right) \sum_{t=1}^{T} \mathbb{E}\left[\mathbb{I}\left[\xi_{t-1}^C\right]\right] \\
\leq&\frac{r_{\max}}{\epsilon_q} \sum_{t=1}^{T} \frac{2(N+2)}{t^2} \leq \frac{C_4 N r_{\max}}{\epsilon_q},
\end{aligned} \tag{67}$$

for a constant C_4. Now the claim follows after combining Equations (63), (64), (65), (66), and (67). □

C.5 Proof of Proposition 5.4

Proof. The claim follows from standard routines in linear bandit methods. We first consider the $\hat{\beta}_{t,\lambda} - \beta_\lambda$ term. We can view the tuple (t, k) as the new timestamp. Let $\epsilon_{t,k} \triangleq y_{t,\sigma_{t,k}}^\top \beta_\lambda - \gamma_{t,k}$. According to Lemma 5.1, $\mathbb{E}[\epsilon_{t,k} | y_{t,\sigma_{t,k}}, \mathcal{H}_{t,k}] = 0$ where $\mathcal{H}_{t,k}$ denotes the historical observations before the t-th consumer views the k-th product. In addition, it is bounded in $[q_t - 1, q_t]$ and hence sub-gaussian with $R = 1/2$. As a result, according to [1, Theorem 2], under Assumption 5.2 with parameter U, for any $\delta \in (0, 1)$, for any $t \geq 0$,

$$\begin{aligned} \left\| \hat{\beta}_{t,\lambda} - \beta_\lambda \right\|_{\Sigma_{t,\lambda}} &\leq \frac{1}{2} \sqrt{2 \log \frac{\det \left(\Sigma_{t,\lambda} \right)^{1/2} \det \left(\alpha_\lambda \mathbf{I}_{m_y} \right)^{-1/2}}{\delta}} + \alpha_\lambda^{1/2} U \\ &= \frac{1}{2} \sqrt{\log \frac{\det \left(\Sigma_{t,\lambda} \right)}{\delta^2 \alpha_\lambda^{m_y}}} + \alpha_\lambda^{1/2} U \end{aligned} \tag{68}$$

According to [1, Lemma 10], we have

$$\det \left(\Sigma_{t,\lambda} \right) \leq \left(\alpha_\lambda + \frac{Nt}{m_y} \right)^{m_y}. \tag{69}$$

As a result, by letting $\delta = t^{-2}$, for all $t \geq 0$, at least $1 - t^{-2}$,

$$\left\| \hat{\beta}_{t,\lambda} - \beta_\lambda \right\|_{\Sigma_{t,\lambda}} \leq \frac{1}{2} \sqrt{m_y \log \left(1 + \frac{Nt}{m_y \alpha_\lambda} \right) + 4 \log t} + \alpha_\lambda^{1/2} U \leq \tau(t, m_y, \alpha_\lambda). \tag{70}$$

Similarly, we have

$$\left\| \hat{\beta}_{t,q} - \beta_q \right\|_{\Sigma_{t,q}} \leq \frac{1}{2} \sqrt{m_x \log \left(1 + \frac{Nt}{m_x \alpha_q} \right) + 4 \log t} + \alpha_q^{1/2} U \leq \tau(t, m_x, \alpha_q). \tag{71}$$

In addition, because

$$\|x_t\| \leq 1 \Longrightarrow \|z_t\| \leq 1 \quad \text{and} \quad \|\beta_s\|, \|\beta_q\| \leq U \Longrightarrow \|\beta_w\| \leq U^2, \tag{72}$$

we have

$$\left\| \hat{\beta}_{t,w} - \beta_w \right\|_{\Sigma_{t,w}} \leq \frac{1}{2} \sqrt{m_x^2 \log \left(1 + \frac{Nt}{m_x^2 \alpha_w} \right) + 4 \log t} + \alpha_w^{1/2} U^2 \leq \tau(t, m_x^2, \alpha_w). \tag{73}$$

Now the claim follows. □

C.6 Proof of Theorem 5.5

Proof. Similar to the proof of Theorem 5.3, for each timestamp t, define the favorable event as

$$\xi_t \triangleq \left\{ \left\| \hat{\beta}_{t,\lambda} - \beta_\lambda \right\|_{\Sigma_{t,\lambda}} \leq \tau(t, m_y, \alpha_\lambda), \left\| \hat{\beta}_{t,q} - \beta_q \right\|_{\Sigma_{t,q}} \leq \tau(t, m_x, \alpha_q), \text{and } \left\| \hat{\beta}_{t,w} - \beta_w \right\|_{\Sigma_{t,w}} \leq \tau(t, m_x^2, \alpha_w) \right\}. \tag{74}$$

According to Proposition 5.4, we have $P(\xi_t) \geq 1 - 3t^{-2}$. Let ξ_t^C denote the complement event of ξ_t.

$$\begin{aligned} \mathbb{E}[\text{Reg}_T | \boldsymbol{X}_T, \boldsymbol{Y}_T] = &\sum_{t=1}^{T} \mathbb{E} \left[\left(\text{Revenue} \left(\boldsymbol{\sigma}_t^*; \boldsymbol{\lambda}_t, q_t, s_t \right) - \text{Revenue} \left(\boldsymbol{\sigma}_t; \boldsymbol{\lambda}_t, q_t, s_t \right) \right) \mathbb{I}[\xi_{t-1}] \right] \\ &+ \sum_{t=1}^{T} \mathbb{E} \left[\left(\text{Revenue} \left(\boldsymbol{\sigma}_t^*; \boldsymbol{\lambda}_t, q_t, s_t \right) - \text{Revenue} \left(\boldsymbol{\sigma}_t; \boldsymbol{\lambda}_t, q_t, s_t \right) \right) \mathbb{I} \left[\xi_{t-1}^C \right] \right]. \end{aligned} \tag{75}$$

On the one hand, consider the first term in the RHS of Equation (75), when the event ξ_t is satisfied, we have $q_t \le \tilde{q}_t$, $z_t \le \tilde{w}_t$, and $\lambda_{t,k} \le \tilde{\lambda}_{t,k}, \forall k \in [N]$. As a result, according to Proposition C.1,

$$\text{Revenue}\left(\boldsymbol{\sigma}_t^*; \boldsymbol{\lambda}_t, q_t, s_t\right) \le \text{Revenue}\left(\boldsymbol{\sigma}_t^*; \tilde{\boldsymbol{\lambda}}_t, \tilde{q}_t, \tilde{s}_t\right) \le \text{Revenue}\left(\boldsymbol{\sigma}_t; \tilde{\boldsymbol{\lambda}}_t, \tilde{q}_t, \tilde{s}_t\right). \tag{76}$$

Hence,

$$\begin{aligned}
&\sum_{t=1}^{T} \mathbb{E}\left[\left(\text{Revenue}\left(\boldsymbol{\sigma}_t^*; \boldsymbol{\lambda}_t, q_t, s_t\right) - \text{Revenue}\left(\boldsymbol{\sigma}_t; \boldsymbol{\lambda}_t, q_t, s_t\right)\right) \mathbb{I}[\xi_t]\right] \\
&\le \sum_{t=1}^{T} \mathbb{E}\left[\left(\text{Revenue}\left(\boldsymbol{\sigma}_t; \tilde{\boldsymbol{\lambda}}_t, \tilde{q}_t, \tilde{s}_t\right) - \text{Revenue}\left(\boldsymbol{\sigma}_t; \boldsymbol{\lambda}_t, q_t, s_t\right)\right) \mathbb{I}[\xi_t]\right] \\
&\le \frac{r_{\max}}{\epsilon_q} \sum_{t=1}^{T} \mathbb{E}\left[\sum_{k=1}^{N} \left(|q_t - \tilde{q}_t| \, \mathbb{E}\left[O_{t,k}^q\right] + 3\left|\lambda_{t,\sigma_{t,k}} - \tilde{\lambda}_{t,\sigma_{t,k}}\right| \mathbb{E}\left[O_{t,k}^\lambda\right] + |w_t - \tilde{w}_t| \, \mathbb{E}\left[O_{t,k}^w\right]\right) \mathbb{I}[\xi_t]\right] \\
&\le \frac{r_{\max}}{\epsilon_q} \sum_{t=1}^{T} \mathbb{E}\left[\sum_{k=1}^{N} \left(\tau(t-1, m_x, \alpha_q) \|x_t\|_{\Sigma_{t-1,q}^{-1}} O_{t,k}^q + 3\tau(t-1, m_y, \alpha_\lambda) \left\|y_{t,\sigma_{t,k}}\right\|_{\Sigma_{t-1,\lambda}^{-1}} O_{t,k}^\lambda + \tau(t-1, m_x^2, \alpha_w) \|z_t\|_{\Sigma_{t-1,w}^{-1}} O_{t,k}^w\right) \mathbb{I}[\xi_{t-1}]\right].
\end{aligned} \tag{77}$$

We consider the $\sum_{t=1}^{T} \mathbb{E}\left[\sum_{k=1}^{N} \tau(t-1, m_y, \alpha_\lambda) \left\|y_{t,\sigma_{t,k}}\right\|_{\Sigma_{t-1,\lambda}^{-1}} O_{t,k}^\lambda\right]$ term first and this part is similar to the proof of [51, Lemma 1]. Define $g_{t,k} = \left\|y_{t,\sigma_{t,k}}\right\|_{\Sigma_{t-1,\lambda}^{-1}}$ for all $t \in [T]$, $k \in [N]$ that satisfy $O_{t,k}^\lambda = 1$. Because

$$\Sigma_{t,\lambda} = \sum_{u=1}^{t} \sum_{k=1}^{\Phi_u} y_{u,\sigma_{u,k}} y_{u,\sigma_{u,k}}^\top + \alpha_\lambda \mathbf{I}_{m_y} = \Sigma_{t-1,\lambda} + \sum_{k=1}^{\Phi_t} y_{t,\sigma_{t,k}} y_{t,\sigma_{t,k}}^\top, \quad \Sigma_{0,\lambda} = \alpha_\lambda \mathbf{I}_{m_y}, \tag{78}$$

we have for all $1 \le k \le \Phi_t$,

$$\det\left(\Sigma_{t,\lambda}\right) \ge \det\left(\Sigma_{t-1,\lambda} + y_{t,\sigma_{t,k}} y_{t,\sigma_{t,k}}^\top\right) = \det\left(\Sigma_{t-1,\lambda}\right)\left(1 + g_{t,k}^2\right). \tag{79}$$

As a result

$$\left(\frac{\det\left(\Sigma_{t,\lambda}\right)}{\det\left(\Sigma_{t-1,\lambda}\right)}\right)^N \ge \left(\frac{\det\left(\Sigma_{t,\lambda}\right)}{\det\left(\Sigma_{t-1,\lambda}\right)}\right)^{\Phi_t} = \prod_{k=1}^{\Phi_t} (1 + g_{u,k}^2). \tag{80}$$

Hence,

$$\left(\det\left(\Sigma_{t,\lambda}\right)\right)^N \ge \alpha_\lambda^{N m_y} \prod_{u=1}^{t} \prod_{k=1}^{\Phi_t} (1 + g_{u,k}^2). \tag{81}$$

On the other hand, due to Assumption 5.2,

$$\begin{aligned}
\text{trace}\left(\Sigma_{t,\lambda}\right) &= \text{trace}\left(\sum_{u=1}^{t} \sum_{k=1}^{\Phi_u} y_{u,\sigma_{u,k}} y_{u,\sigma_{u,k}}^\top + \alpha_\lambda \mathbf{I}_{m_y}\right) = m_y \alpha_\lambda + \sum_{u=1}^{t} \sum_{k=1}^{\Phi_t} \|y_{u,\sigma_{u,k}}\|^2 \\
&\le m_y \alpha_\lambda + tN.
\end{aligned} \tag{82}$$

According to the [1, Lemma 10], we have $\text{trace}(\Sigma_{t,\lambda})/m_y \ge (\det(\Sigma_{t,\lambda}))^{1/m_y}$. As a result,

$$\left(\frac{m_y \alpha_\lambda + tN}{m_y}\right)^{N m_y} \ge \alpha_\lambda^{N m_y} \prod_{u=1}^{t} \prod_{k=1}^{\Phi_t} (1 + g_{u,k}^2). \tag{83}$$

By taking the logarithm, we can get that

$$\sum_{u=1}^{t} \sum_{k=1}^{\Phi_t} \log(1 + g_{u,k}^2) \le N m_y \log\left(1 + \frac{tN}{m_y \alpha_\lambda}\right). \tag{84}$$

Since

$$g_{t,k}^2 = \left\|y_{t,\sigma_{t,k}}\right\|_{\Sigma_{t-1,\lambda}^{-1}}^2 = y_{t,\sigma_{t,k}}^\top \Sigma_{t-1,\lambda}^{-1} y_{t,\sigma_{t,k}} \le y_{t,\sigma_{t,k}}^\top \Sigma_{0,\lambda}^{-1} y_{t,\sigma_{t,k}} = \alpha_\lambda^{-1} \|y_{t,\sigma_{t,k}}\|^2 \le \alpha_\lambda^{-1}, \tag{85}$$

we have $g_{t,k}^2 \leq \alpha_\lambda^{-1} \log(1+g_{t,k}^2)/\log(1+\alpha_\lambda^{-1})$. As a result,

$$\sum_{u=1}^{t}\sum_{k=1}^{\Phi_t} g_{u,k}^2 \leq \frac{1}{\alpha_\lambda \log(1+\alpha_\lambda^{-1})} \sum_{u=1}^{t}\sum_{k=1}^{\Phi_t} \log(1+g_{u,k}^2) \leq \frac{N m_y \log\left(1+\frac{tN}{m_y \alpha_\lambda}\right)}{\alpha_\lambda \log(1+\alpha_\lambda^{-1})}. \tag{86}$$

Hence, according to the Cauchy–Schwarz inequality, we have

$$\begin{aligned}\sum_{t=1}^{T}\sum_{k=1}^{N} \left\|y_{t,\sigma_{t,k}}\right\|_{\Sigma_{t-1,\lambda}^{-1}} O_{t,k}^{\lambda} &\leq \sqrt{\left(\sum_{t=1}^{T}\sum_{k=1}^{\Phi_t} g_{t,k}^2\right)\left(\sum_{t=1}^{T}\sum_{k=1}^{\Phi_t} O_{t,k}\right)} \\ &\leq \sqrt{\frac{TN^2 m_y \log\left(1+\frac{TN}{m_y\alpha_\lambda}\right)}{\alpha_\lambda \log(1+\alpha_\lambda^{-1})}}\end{aligned} \tag{87}$$

As a result, we have

$$\begin{aligned}&\sum_{t=1}^{T}\sum_{k=1}^{N} \tau(t-1, m_y, \alpha_\lambda) \left\|y_{t,k}\right\|_{\Sigma_{t-1,\lambda}^{-1}} O_{t,k}^{\lambda} \\ &\leq \tau(T-1, m_y, \alpha_\lambda) \sum_{t=1}^{T}\sum_{k=1}^{N} \left\|y_{t,k}\right\|_{\Sigma_{t-1,\lambda}^{-1}} O_{t,k}^{\lambda} \\ &\leq \tau(T-1, m_y, \alpha_\lambda) \sqrt{\frac{TN^2 m_y \log\left(1+\frac{TN}{m_y\alpha_\lambda}\right)}{\alpha_\lambda \log(1+\alpha_\lambda^{-1})}} \leq \chi(T, m_y, \alpha_\lambda).\end{aligned} \tag{88}$$

Using similar techniques, we can prove that

$$\begin{aligned}\sum_{t=1}^{T}\sum_{k=1}^{N} \tau(t-1, m_x, \alpha_q) \left\|x_t\right\|_{\Sigma_{t-1,q}^{-1}} O_{t,k}^{q} &\leq \tau(T-1, m_x, \alpha_q) \sqrt{\frac{TN^2 m_x \log\left(1+\frac{TN}{m_x\alpha_q}\right)}{\alpha_q \log(1+\alpha_q^{-1})}}, \\ &\leq \chi(T, m_x, \alpha_q) \\ \sum_{t=1}^{T}\sum_{k=1}^{N} \tau(t-1, m_x^2, \alpha_w) \left\|z_t\right\|_{\Sigma_{t-1,q}^{-1}} O_{t,k}^{w} &\leq \tau(T-1, m_x^2, \alpha_w) \sqrt{\frac{TN^2 m_x^2 \log\left(1+\frac{TN}{m_x^2\alpha_q}\right)}{\alpha_w \log(1+\alpha_w^{-1})}} \\ &\leq \chi(T, m_x^2, \alpha_w).\end{aligned} \tag{89}$$

On the other hand,

$$\begin{aligned}&\sum_{t=1}^{T} \mathbb{E}\left[\left(\text{Revenue}\left(\boldsymbol{\sigma}_t^*; \boldsymbol{\lambda}_t, q_t, s_t\right) - \text{Revenue}\left(\boldsymbol{\sigma}_t; \boldsymbol{\lambda}_t, q_t, s_t\right)\right) \mathbb{I}\left[\xi_{t-1}^C\right]\right] \\ &\leq \text{Revenue}\left(\boldsymbol{\sigma}_t^*; \boldsymbol{\lambda}_t, q_t, s_t\right) \sum_{t=1}^{T} \mathbb{E}\left[\mathbb{I}\left[\xi_{t-1}^C\right]\right] \\ &\leq \frac{r_{\max}}{\epsilon_q} \sum_{t=1}^{T} \frac{3}{t^2} \leq \frac{C_1 r_{\max}}{\epsilon_q},\end{aligned} \tag{90}$$

for a constant C_1. Now the claim follows after combining Equations (77), (88), (89), and (90). □