POMS PRODUCTION AND OPERATIONS MANAGEMENT

PRODUCTION AND OPERATIONS MANAGEMENT

Vol. 30, No. 1, January 2021, pp. 85–102
ISSN 1059-1478 | EISSN 1937-5956 | 21 | 3001 | 0085

DOI 10.1111/poms.13258

Dynamic Assortment Planning Under Nested Logit Models

Xi Chen
Leonard N. Stern School of Business, New York University, New York, New York 10012, USA, xc13@stern.nyu.edu

Chao Shi
School of Information Management and Engineering, Shanghai University of Finance and Economics, Shanghai 200433, China, shi.chao@sufe.edu.cn

Yining Wang
Warrington College of Business, University of Florida, Gainesville, Florida 32611, USA, yining.wang@warrington.ufl.edu

Yuan Zhou
Department of Industrial and Enterprise Systems Engineering, Department of Computer Science (Affiliate), University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA, yuanz@illinois.edu

We study a stylized dynamic assortment planning problem during a selling season of finite length T. At each time period, the seller offers an arriving customer an assortment of substitutable products and the customer makes the purchase among offered products according to a discrete choice model. The goal of the seller is to maximize the expected revenue, or equivalently, to minimize the worst-case expected regret. One key challenge is that utilities of products are unknown to the seller and need to be learned. Although the dynamic assortment planning problem has received increasing attention in revenue management, most existing work is based on the multinomial logit choice models (MNL). In this paper, we study the problem of dynamic assortment planning under a more general choice model—the nested logit model, which models hierarchical choice behavior and is "the most widely used member of the GEV (generalized extreme value) family" (Train 2009). By leveraging the revenue-ordered structure of the optimal assortment within each nest, we develop a novel upper confidence bound (UCB) policy with an aggregated estimation scheme. Our policy simultaneously learns customers' choice behavior and makes dynamic decisions on assortments based on the current knowledge. It achieves the accumulated regret at the order of $\tilde{O}(\sqrt{MNT})$, where M is the number of nests and N is the number of products in each nest. We further provide a lower bound result of $\Omega(\sqrt{MT})$, which shows the near optimality of the upper bound when T is much larger than M and N. When the number of items per nest N is large, we further provide a discretization heuristic for better performance of our algorithm. Numerical results are presented to demonstrate the empirical performance of our proposed algorithms.

Key words: dynamic assortment optimization; nested logit models; regret analysis; upper confidence bound
History: Received: August 2019; Accepted: July 2020 by Dan Zhang after two rounds of revision.

1. Introduction

Assortment planning has a wide range of applications in retailing and online advertising. Given a large number of substitutable products, the assortment planning problem refers to the selection of a subset of products (a.k.a., an assortment) offered to a customer such that the expected revenue is maximized. To model customers' choice behavior when facing a set of offered products, discrete choice models, which capture demand for each product as a function of the entire assortment, have been widely used. One of the most popular discrete choice models is the *multinomial logit model (MNL)*, which naturally results from the random utility theory where a customer's preference of a product is represented by the mean utility of the product with a random factor (McFadden 1974). An important extension of the MNL is the nested logit model (Borch-Supan 1990, McFadden 1980, Williams 1977) that models a customer's choice in a hierarchical way: a customer first selects a category of products (known as a nest), and then a product within the category. When the mean utilities of the products are given, the static assortment optimization problem under MNL or nested logit models can be efficiently solved (Davis et al. 2014, Talluri and van Ryzin 2004).

In many scenarios, customers' choice behavior (e.g., mean utilities of products) is not given as *a priori* and cannot be easily estimated due to the insufficiency of historical data (e.g., fast fashion sale or online advertising). To address this challenge, dynamic assortment planning that simultaneously learns choice behavior and makes decisions about the assortment has received a lot of attention (Agrawal et al. 2017, 2019, Caro and Gallien 2007, Chen and Wang 2018, Rusmevichientong et al. 2010, Saure and Zeevi 2013, Wang et al. 2018). More specifically, in a dynamic assortment planning problem, the seller offers an assortment (or a set of assortments for different nests in a nested logit model) to each arriving customer in a finite time horizon T, observes the purchase behavior of the customer, and then updates the learned information about the underlying demand function. The goal of the seller is to maximize the cumulative expected revenue over T periods. In the literature, the *regret* is often adopted to measure the performance of a given dynamic assortment planning policy, which is defined as the gap between the expected revenue generated by the policy and the oracle expected revenue when the mean utility for each product is known as *a priori*.

In existing dynamic assortment literature, the underlying choice model is usually assumed to be an MNL model (Agrawal et al. 2017, 2019, Rusmevichientong et al. 2010, Saure and Zeevi 2013, Wang et al. 2018). (The work of (Saure and Zeevi 2013) also considered other forms of choice models, in addition to the MNL model.) In this article, we study this problem under a more general choice model—the two-level nested logit model. Indeed, the nested logit model is considered as "the most widely used member of the GEV (generalized extreme value) family" and "has been applied by many researchers in a variety of situations" (see Chapter 4 from Train (2009)). It is well known that the standard MNL suffers from the independence of irrelevant alternatives (IIA), which implies proportional substitution across alternatives (see Chapter 4 from Train (2009)). The nested logit model relaxes the IIA assumption on alternatives in different nests, and thus provides a richer set of substitution patterns. Despite the importance of the nested logit model, the dynamic assortment planning question under nested logit models remains an open problem in revenue management due to the complicated structure of nested logit models.

The main contribution of this study is to develop computationally efficient policies for addressing this problem. Assume that there are M nests and each nest has N possible products to recommend. By leveraging the revenue-ordered structure of optimal assortments and the idea of aggregate estimation of next-level utilities, we propose the first upper confidence bound (UCB)-based policy, which leads to a non-asymptotic regret bound, in which the dominating term involving T is $\tilde{O}(\sqrt{MNT})$ (see Corollary 1 for a more precise bound). Here, $\tilde{O}$ hides the logarithmic dependence on T,N, and M.

Our second contribution is to understand the information-theoretical limitation of the problem. In particular, we further provide a lower bound on the regret $\Omega(\sqrt{MT})$ (see Theorem 2). First, this lower bound shows that when the time horizon T is sufficiently large, our upper bound is within a factor of $\sqrt{N}$ of the lower bound, where N is the number products within each nest and is smaller than the total number of products. The optimal dependence on N is, however, a technically very challenging question and is beyond the scope of this study. Nevertheless, for the case of N being large, we introduce a discretization technique, which provides a useful heuristic leading to a much improved dependence on N. Through simulation studies, we found the discretization heuristic to be very effective with improved performance when there are many items per nest. Second, this lower bound also demonstrates a fundamental difference between the nested logit models and standard (plain) MNL models. According to Wang et al. (2018), the standard MNL admits a tight lower bound of $\Omega(\sqrt{T})$, independent of other problem parameters (e.g., the number of products). In contrast, for nested logit models, our lower bound shows that, in addition to $\sqrt{T}$, dependency on the number of nests M is unavoidable.

The details of the proposed policies will be presented in the main paper and here we briefly highlight the key technical points in the proposed policies:

1. Leveraging the revenue-ordered structure: For N products in each of the M nests, the total number of possible assortment combinations (i.e., the size of the action space) will be $(2^N)^M$, which is exponentially large. By leveraging the revenue-ordered structure of the optimal assortment within each nest (see Lemma 1 and Davis et al. (2014), Li et al. (2015)), the size of the action space can be effectively reduced to $O(N^M)$. However, the size of this reduced action space is still too large if one directly applies existing bandit learning algorithms that treat each assortment in the action space independently, which will incur a regret related to N^M. To address this challenge, we propose an aggregate estimation technique as follows.
2. Aggregate estimation: A key point of the paper is that estimating utility parameters for each individual product (that will incur a large regret) is unnecessary for dynamic assortment planning. Instead, we propose an aggregate estimation technique that only estimates the preference and revenue parameters on a nest

level. More specifically, in our algorithm only *level sets* of assortments within each nest are considered, which have both *unknown* aggregated revenue and utility parameters.

Another advantage of our algorithm is that it shows that estimation of exponent parameters $\{\gamma_i\}_{i=1}^M$ (see Equation (7) in the nested logit model specification) is not necessary. Instead, we directly estimate "nested-level utility" $V_i(\cdot)^{\gamma_i}$ (see Equation (7) and the discussions above Equation (7)).

3. UCB policy: We propose an upper confidence bound (UCB) algorithm using an epoch-based strategy from Agrawal et al. (2019), which leads to a worst-case expected regret of $\tilde{O}(\sqrt{MNT})$. Although the UCB has been a well-known technique for bandit problems, adopting this high-level idea to solve a problem with specific structures certainly requires technical innovations (e.g., how to build a confidence bound on a carefully designed parameter, see Lemma 4). We further note that our UCB policy generalizes the one in Agrawal et al. (2019) because in our model the "level sets" are constructed within each nest, and therefore both their revenue and utility parameters are unknown (see Equation (6) in section 2.2) This contrasts the setting in Agrawal et al. (2019) in which the revenue parameters of each single item are known.
4. Discretization technique: When N is large, we introduce a discretization technique to reduce the size of the action space to $O((1/\delta)^M)$, where δ is discretization granularity. Our policy without discretization corresponds to a special case of δ=0. We are able to show that the proposed space reduction techniques lose very little in terms of optimal expected revenue, that is, the gap of the optimal expected revenue between all the possible assortment combinations and the reduced action space is at most δ (see Lemma 8). Simulation studies confirm the effectiveness of this discretization heuristic.

To the best of our knowledge, our policies are the first policies for dynamic assortment planning under the nested logit model, which presents unique challenges compared to the standard MNL model as the nest-level revenues of assortment selections are not known and have to be estimated on the fly. It is also worthwhile noting that due to the complicated structure of the nested logit model, it is technically challenging to derive a tight lower bound on the regret in terms of N, and we suspect that the current lower bound $\Omega(\sqrt{MT})$ is not tight and misses a $\sqrt{N}$ factor (see more detailed discussions in Remark 4 in section 4). Since the main focus of the paper is to derive the first efficient policy for dynamic assortment planning under nested logit models, we leave this challenging technical problem for future works.

1.1. Related Works

Static assortment planning with known choice behavior has been an active research area since the seminal works by van Ryzin and Mahajan (1999) and Mahajan and van Ryzin (2001). When the customer makes the choice according to the MNL model, Talluri and van Ryzin (2004) and Gallego et al. (2004) proved an optimal assortment will belong to revenue-ordered assortments (a.k.a. nested-by-revenue assortments). An alternative proof is provided in Liu and van Ryzin (2008). This important structural result enables the efficient computation of static assortment planning under the MNL model, which reduces the number of candidate assortments from 2^N to N, where N is the number of products per nest. When there is a set constraint on the assortment set, an efficient polynomial-time algorithm (with running time $O(N^2)$) was proposed in Rusmevichientong et al. (2010). For nested logit models, Davis et al. (2014) proved an important structural result that the optimal assortment within each nest is revenue-ordered, which will also be used in designing our dynamic policies. Assuming that there are M nests and N products within each nest, Li and Rusmevichientong (2014) further proposed an efficient greedy algorithm to find an optimal assortment set with $O(NM \log M)$ time complexity. Kök and Xu (2011) considered the joint assortment optimization and pricing problem with a restricted number of nests. There are several recent works on static assortment planning under variants of nested logit models. For example, Gallego and Topaloglu (2014) studied the constrained nested logit model; Li et al. (2015) extended the popular two-level nested logit model to a d-level nested logit model with $d \geq 2$; Zhang et al. (2020) studied the paired combinatorial nested logit model. In addition, there are extensive research on static assortment optimization for more complex choice models, for example, a robust version of MNL (Rusmevichientong and Topaloglu 2012), the mixture of logit models (Bront et al. 2009, Méndez-Díaz et al. 2014, Rusmevichientong et al. 2014), Markov chain-based choice models (Blanchet et al. 2016, Désir et al. 2020), the generalized attraction model (Wang 2013), Mallows-based choice models (Désir et al. 2016), a multiple attempt model (Chung et al. 2019), contextual MNL (Cheung and Simchi-Levi 2017), and a general class of choice models based on a distribution over permutations (Farias et al. 2013).

Davis et al. (2014) considered the nested logit model studied in this study, with both the cases of $\gamma_i \leq 1$ and $\gamma_i > 1$. In the case of $\gamma_i \leq 1$, they established the revenue-order property within each nest, but considered an alternative linear programming type algorithm to solve for optimal assortments efficiently. In Li et al. (2015) the optimization question of d-level nested logit models is considered, and efficient fractional programming based methods are developed. Our optimization subroutine (see section 3.1) turns out to be similar as the one in Li et al. (2015) in the special case of d=2, with the difference being that upper confidence estimates of $R_i(S_i)$ and $V_i(S_i)^{\gamma_i}$ are used in our optimization, while in Li et al. (2015) the full-information parameter values were used.

Due to increasing popularity of data-driven revenue management, researchers have started to relax the assumption about fully available prior knowledge of customers' choice behavior and investigate dynamic assortment planning. Motivated by fast-fashion retailing, the work by Caro and Gallien (2007) was among the first to study the dynamic assortment planning problem, which assumes that the demand for products is independent of each other. Bertsimas and Mišić (2019) studied a two-step problem with separate demand estimation and assortment planning, where the first step estimates a generic ranking-based choice model and the second step solves a mixed-integer optimization for assortment planning. Rusmevichientong et al. (2010), Saure and Zeevi (2013), Agrawal et al. (2017, 2019), Wang et al. (2018) and Chen et al. (2018) incorporated choice models of MNL into dynamic assortment planning, formulating the problem into an online regret minimization problem. However, the extension of the plain MNL model to nested logit models is highly nontrivial and requires several technical innovations. For example, instead of estimating utility parameters for each product, we estimate nest-level aggregated quantities (see more discussions in the introduction). Furthermore, we introduce a discretization technique to alleviate the effect of having many items per nest.

There is another line of recent research on investigating the assortment planning question in which each arriving customer could have a different choice behavior. For example, Golrezaei et al. (2014) and Chen et al. (2019) assumed that each customer's choice behavior is known but that the customers' arriving sequence can be adversarially chosen, and took into account both the revenue and inventory levels. Since the arriving sequence can be arbitrary, there is no learning component in the problem and both Golrezaei et al. (2014) and Chen et al. (2019) adopted the competitive ratio as the performance evaluation metric. In addition, there are a few recent works studying joint assortment planning and pricing under MNL models (see, e.g., Besbes and Saure (2016), Miao and Chao (2018), and Wang (2012)). It would also be an interesting future work to consider dynamic joint assortment planning and pricing under nested MNL models.

1.2. Notations and Paper Organizations

Throughout the paper, we use $f(\cdot) \lesssim g(\cdot)$ to denote that $f(\cdot)=O(g(\cdot))$, or more specifically $\limsup_{T\to\infty} |f(T)|/|g(T)| < \infty$. Similarly, by $f(\cdot) \gtrsim g(\cdot)$, we denote $f(\cdot) = \Omega(g(\cdot))$. We also use $f(\cdot) \approx g(\cdot)$ for $f(\cdot) = \Theta(g(\cdot))$. In the paper, $\tilde{O}(\cdot)$ is used to hide logarithmic factors on T, N, and M. The rest of the paper is organized as follows: In section 2, we first provide the background of nested logit models and introduce an important structural result on optimal assortments (Davis et al. 2014, Li et al. 2015). In section 3, we propose our UCB policy and establish the corresponding regret bound. A lower bound on regret is provided in section 4. The numerical results are provided in section 5, followed by the conclusion in section 6.

2. Model Specifications and Assortment Space Reductions

In this section, we formally introduce the nested logit assortment choice model considered in this study. We restrict ourselves to two-level nested logit models, where items are organized as M known commodity *nests* and customers' purchasing actions are modeled by a *hierarchical* multinomial logit model (more details given in section 2.1).

2.1. The Nested Logit Model

We use $[M]=\{1,2,\cdots,M\}$ to label the M nests. For each nest $i \in [M]$, label the items in nest i by $[N_i] = \{1, 2, \cdots, N_i\}$. Each item $j \in [N_i]$ is associated with a *known* revenue parameter r_{ij} and an *unknown* mean utility parameter v_{ij}. We assume each nest has an equal number of items, that is, $N_1 = \cdots = N_M = N$. Furthermore, let $\{\gamma_i\}_{i\in[M]} \subseteq [0,1]$ be a collection of *unknown* correlation parameters for different nests. Each parameter γ_i is a measure of the degree of independence among the items in nest i: a larger value of γ_i indicates less correlation (see Chapter 4 of Train (2009)).

At each time period $t \in \{1,2,\cdots,T\}$, the retailer offers the arriving customer an assortment $S_i^{(t)} \in \mathbb{S}_i = 2^{[N]}$ for every nest $i \in [M]$, conveniently denoted as $\mathbf{S}^{(t)} = (S_1^{(t)}, \cdots, S_M^{(t)})$. The retailer then observes a nest-level purchase option $i_t \in [M] \cup \{0\}$. If $i_t \in [M]$, an item $j_t \in [N]$ is purchased within the nest i_t. Moreover, $i_t = 0$ means no purchase occurs at time t. The probabilistic model for the purchasing option (i_t, j_t) can be formulated as below:

$$\Pr[i_t = i|\mathbf{S}^{(t)}] = \frac{V_i(S_i^{(t)})^{\gamma_i}}{V_0 + \sum_{i'=1}^{M} V_{i'}(S_{i'}^{(t)})^{\gamma_{i'}}}, \text{ where } V_0 \equiv 1, V_i(S_i^{(t)}) = \sum_{j \in S_i^{(t)}} v_{ij} \text{ for } i \in [M]; \tag{1}$$

$$\Pr[j_t = j|i_t = i, \mathbf{S}^{(t)}] = \frac{v_{ij}}{\sum_{j' \in S_i^{(t)}} v_{ij'}} \text{ for } i \in [M], j \in S_i^{(t)}. \tag{2}$$

Note that when $\gamma_i = 1$ for all $i \in [M]$, the nested logit model reduces to the standard MNL model.

The retailer then collects revenue r_{i_t,j_t} provided that $i_t \neq 0$. The expected revenue $R(\mathbf{S}^{(t)})$ given the assortment combination $\mathbf{S}^{(t)}$ can then be written as

$$\begin{aligned} R(\mathbf{S}^{(t)}) &= \sum_{i=1}^{M} \Pr[i_t = i|\mathbf{S}^{(t)}] \sum_{j \in S_i^{(t)}} r_{ij} \Pr[j_t = j|i_t = i, \mathbf{S}^{(t)}] \\ &= \frac{\sum_{i=1}^{M} R_i(S_i^{(t)}) V_i(S_i^{(t)})^{\gamma_i}}{1 + \sum_{i=1}^{M} V_i(S_i^{(t)})^{\gamma_i}}; \text{ where} \\ R_i(S_i^{(t)}) &= \frac{\sum_{j \in S_i^{(t)}} r_{ij} v_{ij}}{\sum_{j \in S_i^{(t)}} v_{ij}}. \end{aligned} \tag{3}$$

The objective of the seller is to minimize *expected (accumulated) regret*, defined as follows:

$$\text{Regret}(\{\mathbf{S}^{(t)}\}_{t=1}^{T}) := \sum_{t=1}^{T} R^* - \mathbb{E}[R(\mathbf{S}^{(t)})], \text{ where } R^* = \max_{\mathbf{S} \in \mathbb{S} = \mathbb{S}_1 \times \cdots \times \mathbb{S}_M} R(\mathbf{S}). \tag{4}$$

Throughout the paper, we make the following boundedness assumptions on revenue and utility parameters:

(A1) The revenue parameters satisfy $0 \leq r_{ij} \leq 1$ for all $i \in [M]$ and $j \in [N]$.
(A2) The utility parameters satisfy $0 < v_{ij} \leq C_V$ for all $i \in [M]$ and $j \in [N]$ with some constant $C_V \geq 1$.

The first boundedness assumption on revenue parameters is standard in the literature (see e.g., Theorem 1 in Agrawal et al. (2019)). It is also worthwhile noting that assumption (A2) is weaker than the common assumption that no purchase (with $V_0 = 1$) is the most frequent outcome. Both assumptions can be regarded as without loss of generality as the parameter values could be normalized.

We remark that in the original nested-logit model assortment planning paper (Davis et al. 2014), it is allowed that $\gamma_i > 1$ and furthermore there is a no-purchase option *within each nest*. We assumed $\gamma_i \leq 1$ because it is the setting in which the full-information combinatorial optimization problem is easy to solve, which is the foundation of our theoretical regret analysis. Indeed, when γ_i exceeds one, it is proved in the work of Davis et al. (2014) that the combinatorial optimization question (when all parameters are known) is NP-hard, and only approximation algorithms can be developed.

We do not allow for a no-purchase option within each nest, moreover, for a more technical reason. In our proposed learning-while-doing algorithm, it is critical to count the number of times that each nest i is selected by customers until a no-purchase action *on the nest level* occurs. If we allow for no-purchase options within each nest, our algorithm will no longer be able to distinguish between the events of no-purchase on the nest level or within nests. This leads to biased estimates of $V_i(S_i)^{\gamma_i}$ parameters and potentially linear regret. Hence, we choose not to include no-purchase options within nests for a cleaner algorithm and analysis.

2.2. Assortment Space Reductions

For nested logit models, the complete assortment selection space (a.k.a. action space) $\mathbb{S} = \mathbb{S}_1 \times \mathbb{S}_2 \times \cdots \times \mathbb{S}_M$ is extremely large, consisting of an exponential number of candidate assortment selections (on the order of $(2^N)^M$). Existing bandit learning approaches treating each assortment set in $\mathbb{S}$ independently would easily incur a regret also exponentially large. It is, thus, mandatory to reduce the number of candidate assortment sets in $\mathbb{S}$.

Fortunately, existing results on the structure of optimal $\mathbf{S}$ show that it suffices to consider *level sets* $\mathcal{L}_i(\theta_i) := \{j \in [N] : r_{ij} \geq \theta_i\}$ for each nest i. In other words, $\mathcal{L}_i(\theta_i)$ is the set of products in nest i with revenue larger than or equal to a given threshold $\theta_i \geq 0$. Define $\mathbb{P}_i := \{\mathcal{L}_i(\theta_i) : \theta_i \geq 0\} \subseteq \mathbb{S}_i$ to be all the possible level sets of $\mathbb{S}_i$ and let

$$\mathbb{P} := \mathbb{P}_1 \times \mathbb{P}_2 \times \cdots \times \mathbb{P}_M \subseteq \mathbb{S}. \tag{5}$$

The following lemma from Davis et al. (2014) and Li et al. (2015) shows that one can restrict the assortment selections to $\mathbb{P}$ without loss of any optimality in terms of expected revenue.

LEMMA 1. (DAVIS ET AL. (2014), LI ET AL. (2015)). *There exists level set threshold parameters* $(\theta_1^*, \ldots, \theta_M^*)$ *and* $\mathbf{S}^* = (\mathcal{L}_1(\theta_1^*), \cdots, \mathcal{L}_M(\theta_M^*)) \in \mathbb{P}$ *such that the following hold:*

1. $R(\mathbf{S}^*) = \max_{\mathbf{S} \in \mathbb{S}} R(\mathbf{S}) = R^*$;
2. $\theta_i^* \geq \gamma_i R^* + (1 - \gamma_i) R_i(S_i^*)$ *for all* $i \in [M]$, *where* $S_i^* = \mathcal{L}_i(\theta_i^*)$.

The first item in Lemma 1 is an important structural result showing that the optimal assortments are "revenue-ordered" within each nest. The second item is a technical result, which will be used in the proof. Compared to the original action space $\mathbb{S}$, the reduced "level set" space $\mathbb{P}$ is much smaller, with each $\mathbb{P}_i$ consisting of N instead of 2^N candidate assortments.

With Lemma 1, an assortment combination $\mathbf{S} = (S_1, \cdots, S_M) \in \mathbb{P}$ can then be parameterized by a vector $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_M) \in ([0,1] \cup \{\infty\})^M$, such that $\mathbf{S}(\boldsymbol{\theta}) = (\mathcal{L}_1(\theta_1), \cdots, \mathcal{L}_M(\theta_M))$. Note that $\mathcal{L}_i(\infty) = \emptyset$ indicates the empty set for nest i. Denote $\mathcal{K}_i = [0,1] \cup \{\infty\}$, and for any $i \in [M]$, $\theta_i \in \mathcal{K}_i$ define

$$u_{i,\theta_i} := V_i(\mathcal{L}_i(\theta_i))^{\gamma_i} \text{ and } \phi_{i,\theta} := R_i(\mathcal{L}_i(\theta_i)), \qquad (6)$$

where $V_i(\cdot)$ and $R_i(\cdot)$ are nest-level utility parameter and expected revenue associated with the level set $\mathcal{L}_i(\theta_i)$ (see definitions of V_i and R_i in Equations (1) and (3), respectively). We note that it is fundamentally different from the standard MNL: the nest-level expected revenue $\phi_{i,\theta}$, which depends on utility parameters, is unknown and needs to be learned; while the revenue of each product in a standard MNL is known to the seller prior to the first selling period. By our assumptions (A1) and (A2), it is easy to verify that $\phi_{i,\theta_i} \in [0,1]$ and $u_{i,\theta_i} \in [0, (NC_V)^{\gamma_i}] \subseteq [0, NC_V]$ for all $i \in [M]$ and $\theta_i \in \mathcal{K}_i$. We also note that $(NC_V)^{\gamma_i} \le NC_V$, since $\gamma_i \in [0,1]$ and $C_V \ge 1$. Furthermore, because each nest consists of at most N products, the sets $\mathcal{K}_i$ can be made *finite* by considering only levels θ_i corresponding to revenue parameters of the N products.

Let $i_t \in [M] \cup \{0\}$ be the nest the customer selects at time t and r_t be the collected revenue. The expected revenue for assortment $\mathbf{S}^{(t)}$ parameterized by $\boldsymbol{\theta}^{(t)}$ can then be expressed as

$$\Pr[i_t = i | \boldsymbol{\theta}^{(t)}] = \frac{u_{i,\theta_i}}{1 + \sum_{i'=1}^{M} u_{i',\theta_{i'}}}; \mathbb{E}[r_t | i_t = i] = \phi_{i,\theta_i}; \quad r_t = 0 \ a.s. \text{ if } i_t = 0.$$

Therefore, the expected revenue for an assortment combination parameterized by $\boldsymbol{\theta}^{(t)}$ takes the following form:

$$R'(\boldsymbol{\theta}^{(t)}) := \sum_{i=1}^{M} \Pr[i_t = i | \boldsymbol{\theta}^{(t)}] \cdot \mathbb{E}[r_t | i_t = i] = \frac{\sum_{i=1}^{M} \phi_{i,\theta_i} u_{i,\theta_i}}{1 + \sum_{i=1}^{M} u_{i,\theta_i}},$$

and the regret in Equation (4) can be equivalently written as,

$$\text{Regret}(\{\boldsymbol{\theta}^{(t)}\}_{t=1}^{T}) := \mathbb{E} \sum_{t=1}^{T} R'(\boldsymbol{\theta}^*) - R'(\boldsymbol{\theta}^{(t)}) \text{ where } R'(\boldsymbol{\theta}^*) = \max_{\boldsymbol{\theta} \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M} R'(\boldsymbol{\theta}). \qquad (8)$$

Algorithm 1: The upper confidence bound (UCB) policy for dynamic assortment planning.

Input: Parameter space of $\boldsymbol{\theta}$: $\mathcal{K}_1, \cdots, \mathcal{K}_M$, upper bound U on $\{u_{i,\theta}\}$ in Equation (6).

Output: assortment sequences $\boldsymbol{\theta}^{(1)}, \cdots, \boldsymbol{\theta}^{(T)} \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M$.

Initialization: τ=1, $\{\mathcal{E}_\tau\}_{\tau=1}^{\infty} = \emptyset$, t=1; for every $i \in [M]$ and $\theta \in \mathcal{K}_i$, set $\mathcal{T}(i,\theta) = \emptyset$, $T(i,\theta)$=0, $\hat{\phi}_{i,\theta} = \bar{\phi}_{i,\theta} = 1$, $\hat{u}_{i,\theta} = \bar{u}_{i,\theta} = U$; for all $i \in [M]$ and $\theta \in \mathcal{K}_i$ corresponding to the empty assortment (i.e., $\mathcal{L}_i(\theta) = \emptyset$), set $\bar{\phi}_{i,\theta} = \phi_{i,\theta} = \bar{u}_{i,\theta} = u_{i,\theta} = 0$;
while $t \le T$ **do**
Find $\hat{\boldsymbol{\theta}}^{(\tau)} = \hat{\boldsymbol{\theta}} \leftarrow \arg\max_{\boldsymbol{\theta} \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M} \bar{R}'(\boldsymbol{\theta})$, where $\bar{R}'(\boldsymbol{\theta}) = [\sum_{i=1}^{M} \bar{\phi}_{i,\theta_i} \bar{u}_{i,\theta_i}]/[1 + \sum_{i=1}^{M} \bar{u}_{i,\theta_i}]$;
$\triangleright$ *This optimization problem can be solved in polynomial time; see section 3.1*;
repeat
  Pick $\boldsymbol{\theta}^{(t)} = \hat{\boldsymbol{\theta}}$ and observe i_t, r_t in Equation (7) and update $\mathcal{E}_\tau \leftarrow \mathcal{E}_\tau \cup \{t\}$, t=t+1;
until $i_{t-1}0$ *or* $t>T$;
for *each* $i \in [M]$ *with* $\mathcal{L}_i(\hat{\theta}_i) \neq \emptyset$ **do**
  Compute $\hat{n}_{i,\tau} = \sum_{t' \in \mathcal{E}_\tau} \mathbb{I}[i_{t'} = i]$ and $\hat{r}_{i,\tau} = \sum_{t' \in \mathcal{E}_\tau} r_{t'} \mathbb{I}[i_{t'} = i]$;
  Let $\theta = \hat{\theta}_i$ (for notational simplicity) and update:
    $\mathcal{T}(i,\theta) \leftarrow \mathcal{T}(i,\theta) \cup \{\tau\}, T(i,\theta) \leftarrow T(i,\theta) + 1$;
  Update the utility and mean revenue estimates and as well as their associated confidence bounds:
  $\hat{u}_{i,\theta} = \frac{1}{T(i,\theta)} \sum_{\tau' \in \mathcal{T}(i,\theta)} \hat{n}_{i,\tau'}$, $\hat{\phi}_{i,\theta} = \frac{\sum_{\tau' \in \mathcal{T}(i,\theta)} \hat{r}_{i,\tau'}}{\sum_{\tau' \in \mathcal{T}(i,\theta)} \hat{n}_{i,\tau'}}$;
  if $T(i,\theta) \ge 96 \ln(2MTK)$ **then**
    $\bar{u}_{i,\theta} = \min\{U, \hat{u}_{i,\theta} + \sqrt{\frac{96 \max(\hat{u}_{i,\theta}, \hat{u}_{i,\theta}^2) \ln(2MTK)}{T(i,\theta)}} + \frac{144 \ln(2MTK)}{T(i,\theta)}\}$,
    $\bar{\phi}_{i,\theta} = \min\{1, \hat{\phi}_{i,\theta} + \sqrt{\frac{\ln(2MTK)}{T(i,\theta) \hat{u}_{i,\theta}}}\}$;
  else
    $\bar{u}_{i,\theta} = U, \bar{\phi}_{i,\theta} = 1$;
  end
end
$\tau \leftarrow \tau + 1$;
end

3. UCB-Based Dynamic Assortment Planning Policies

In this section, we design dynamic planning policies under the nested logit model using an *upper-confidence-bound (UCB)* approach. The details and pseudocode of our proposed policy are given in Algorithm 1.

The high-level idea behind Algorithm 1 is as follows: for every nest i and level set $\theta \in \mathcal{K}_i$, a pair of upper confidence estimates $\bar{\phi}_{i,\theta}$ and $\bar{u}_{i,\theta}$ are constructed and maintained, estimating the nest-level revenue and utility parameters $R_i(\mathcal{L}_i(\theta)) = (\sum_{j \in \mathcal{L}_i(\theta)} r_{ij} v_{ij}) / (\sum_{j \in \mathcal{L}_i(\theta)} v_{ij})$, $V_i(\mathcal{L}_i(\theta))^{\gamma_i} = (\sum_{j \in \mathcal{L}_i(\theta)} v_{ij})^{\gamma_i}$. For every potential customer, an optimal assortment combination based on current (upper) parameter estimates $\bar{\phi}_{i,\theta}, \bar{u}_{i,\theta}$ are computed, which is then offered to the customers repetitively until a no-purchase action occurs. Afterwards, the parameter estimates $\bar{\phi}_{i,\theta}, \bar{u}_{i,\theta}$ are updated for all assortments provided in each nest, and the dynamic assortment planning procedure continues until a total of T customers are served.

We next explain a few notations used in the algorithm and then describe the details of the algorithm. The proofs of the results in this section are provided in the supplementary material.

1. $\mathcal{E}_\tau$: all iterations in epoch τ where the same assortment combination $\boldsymbol{\theta}$ is provided. Each epoch (corresponding to Steps 4-6 in Algorithm 1) terminates whenever the no-purchase action is observed. In other words, one and only one "no-purchase" action $i_t = 0$ appears at the last iteration of each epoch $\mathcal{E}_\tau$.
2. $\mathcal{T}(i, \theta)$: the indices of epochs in which $\theta \in \mathcal{K}_i$ is supplied in nest i; $T(i, \theta) = |\mathcal{T}(i, \theta)|$ denotes the cardinality of $\mathcal{T}(i, \theta)$;
3. $\hat{n}_{i,\tau}$: the number of iterations in the epoch τ (i.e., $\mathcal{E}_\tau$) in which an item in nest i is purchased;
4. $\hat{r}_{i,\tau}$: the total revenue collected for all iterations in $\mathcal{E}_\tau$ in which an item in nest i is purchased;
5. $\hat{u}_{i,\theta}, \hat{\phi}_{i,\theta}, \bar{u}_{i,\theta}, \bar{\phi}_{i,\theta}$: estimates of $u_{i,\theta}, \phi_{i,\theta}$, and their upper confidence bounds.

The epoch-based strategy (i.e., offering the same assortment until no-purchase is observed) in Algorithm 1 was first introduced by Agrawal et al. (2019) and enjoys the favorable properties stated in the next lemma.

LEMMA 2. *For each epoch $\mathcal{E}_\tau$ and nest $i \in [M]$, let $\hat{\theta}_i \in \mathcal{K}_i$ be such that assortment $\mathcal{L}_i(\hat{\theta}_i)$ is provided in nest i in epoch τ. The expectations of the number of iterations and total revenues collected in which nest i is purchased (denoted by $\hat{n}_{i,\tau}$ and $\hat{r}_{i,\tau}$, respectively, in Algorithm 1) satisfy the following regardless of the other offered assortments $\hat{\theta}_{i'}$ for $i' \neq i$ in the same epoch:*

1. $\mathbb{E}[\hat{n}_{i,\tau}] = u_{i,\hat{\theta}_i}$;
2. $\mathbb{E}[\hat{r}_{i,\tau}|\hat{n}_{i,\tau}] = \hat{n}_{i,\tau}\phi_{i,\hat{\theta}_i}$.

The above properties motivate intuitive parameter estimators $\hat{u}_{i,\theta}, \hat{\phi}_{i,\theta}$ of $u_{i,\theta}$ and $\phi_{i,\theta}$ for $\theta = \hat{\theta}_i$, which are taken to be the sample averages of $\hat{n}_{i,\tau}$ and $\hat{r}_{i,\tau}$ over all prior epochs $\mathcal{E}_\tau$ in which the assortment corresponding to level set $\mathcal{L}_i(\hat{\theta}_i)$ in nest i is offered. It is worth noting that in those epochs, the offered assortments in nests other than the i-th nest (i.e., the nests i' for $i' \neq i$) can be arbitrary since the distributions of $\hat{n}_{i,\tau}$ and $\hat{r}_{i,\tau}$ are *independent* of $\hat{\theta}_{i'}$ for $i' \neq i$. This key independence property enables us to combine purchasing information of vastly different assortment combinations (provided that $\hat{\theta}_i$ remains the same), which forms an important *aggregation strategy* that avoids exponentially large regret if assortment combinations are treated independently.

3.1. Efficient Computation of $\hat{\boldsymbol{\theta}}$

Our policy in Algorithm 1 involves a combinatorial optimization problem over all $\boldsymbol{\theta} \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M$ (see Step 3 in Algorithm 1). A brute-force algorithm that enumerates all such $\boldsymbol{\theta}$ takes $O(K^M)$ time and is computationally intractable even for small M values, where $K = \max_i |\{r_{ij} : j \in [N_i]\}| \leq N + 1$. In this section, we introduce a computationally efficient procedure to compute $\hat{\boldsymbol{\theta}}$ using a binary search technique. The idea behind our procedure is similar to the one introduced in (Rusmevichientong et al. 2010) for dynamic assortment optimization in MNL models, which can also be traced to the fractional programming work as early as (Megiddo 1978).

For any $\lambda \in [0,1]$ and $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_M) \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M$ define potential function

$$\psi_\lambda(\boldsymbol{\theta}) := \sum_{i=1}^{M} (\bar{\phi}_{i,\theta_i} - \lambda)\bar{u}_{i,\theta_i}. \tag{9}$$

Recall the definition of $\bar{R}'(\boldsymbol{\theta}) = \frac{\sum_{i=1}^{M} \bar{\phi}_{i,\theta_i}\bar{u}_{i,\theta_i}}{1+\sum_{i=1}^{M} \bar{u}_{i,\theta_i}}$ in Step 3 of Algorithm 1. The following lemma characterizes the properties of $\psi_\lambda(\boldsymbol{\theta})$ and its relationship with $\bar{R}^* = \max_{\boldsymbol{\theta} \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M} \bar{R}'(\boldsymbol{\theta})$:

LEMMA 3. *The following holds for all $\lambda \in [0,1]$:*

1. *If $\bar{R}^* \geq \lambda$, then there exists a $\boldsymbol{\theta} \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M$ such that $\psi_\lambda(\boldsymbol{\theta}) \geq \lambda$; furthermore if $\bar{R}^* > \lambda$, then the inequality is strict;*
2. *If $\bar{R}^* \leq \lambda$, then for all $\boldsymbol{\theta} \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M$, $\psi_\lambda(\boldsymbol{\theta}) \leq \lambda$; furthermore if $\bar{R}^* < \lambda$, then the inequalities are strict.*

Based on Lemma 3, an efficient optimization algorithm computing the maximizer $\hat{\boldsymbol{\theta}}^{(\tau)}$ can be designed by a *binary search* over $\lambda \in [0,1]$. In particular, for each fixed value of λ, the $\boldsymbol{\theta}^*(\lambda) = (\theta_1^*(\lambda), \cdots, \theta_M^*(\lambda)) \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M$ that maximizes $\psi_\lambda(\boldsymbol{\theta})$ can be found by setting $\theta_i^*(\lambda) \in \arg\max_{\theta_i \in \mathcal{K}_i} (\bar{\phi}_{i,\theta_i} - \lambda)\bar{u}_{i,\theta_i}$. If $\psi_\lambda(\boldsymbol{\theta}^*(\lambda)) > \lambda$, then $\bar{R}^* > \lambda$, because otherwise it violates the second property in Lemma 3. Similarly, if $\psi_\lambda(\boldsymbol{\theta}^*(\lambda)) \leq \lambda$, then $\bar{R}^* \leq \lambda$, because otherwise it violates the second part of the first property in Lemma 3 (note that since $\boldsymbol{\theta}^*(\lambda)$ is the maximizer of $\psi_\lambda(\boldsymbol{\theta})$, $\psi_\lambda(\boldsymbol{\theta}^*(\lambda)) \leq \lambda$ implies that $\psi_\lambda(\boldsymbol{\theta}) \leq \lambda$ for all $\boldsymbol{\theta}$). Thus, whether $\bar{R}^* > \lambda$ or $\bar{R}^* \leq \lambda$ can be determined by solely comparing $\psi_\lambda(\boldsymbol{\theta}^*(\lambda))$ with λ.

We remark that each evaluation of $\psi_\lambda(\boldsymbol{\theta}^*(\lambda))$ takes $O(MK)$ time, and the entire binary search procedure takes time $O(MK \log(1/\varepsilon))$ to approximate $\bar{R}^*$ up to arbitrarily small error ε. This is much faster than the brute force algorithm that takes $O(K^M)$ time.

We also remark that, similar to other bisection type algorithms, the computation procedure outlined above computes *approximate* solutions only, with $O(\log(1/\varepsilon))$ iterations required if an error level of $\varepsilon>0$ is desired. We suggest setting the accuracy level ε to $\varepsilon=1/T$, which would inflate an additional $O(1)$ term in the cumulative regret upper bound, while the running time of the binary search routine is strictly polynomial in T. When the time horizon T is unknown before hand, a doubling trick can be used to consider epochs of lengths $1,2,4,\cdots,2^\tau,\cdots$, and within epoch τ (of length 2^τ) an error level of $\epsilon_\tau=2^{-\tau}$ can be used.

3.2. Regret Analysis

Below is our main regret theorem for Algorithm 1.

THEOREM 1. *For each nest i let $\mathcal{K}_i=\{r_{ij}:j\in[N_i]\}$. The assortment sequence $\{\boldsymbol{\theta}^{(t)}\}_{t=1}^T$ produced by Algorithm 1 has the regret upper bounded as*

$$\mathrm{Regret}(\{\boldsymbol{\theta}^{(t)}\}_{t=1}^T)\lesssim\sqrt{MKT\log(MKT)} + MKU\log^2(MKT)+O(1), \tag{10}$$

where $K=\max_i|\mathcal{K}_i|$ and $U=\max_{i\in[M]}\max_{\theta\in\mathcal{K}_i}u_{i,\theta}$.

COROLLARY 1. *With $K=|\mathcal{K}_i|=N+1$ (for any $i\in[M]$) and $U\le NC_V$, the regret upper bound in Theorem 1 can be simplified to*

$$\begin{aligned}\mathrm{Regret}(\{\mathbf{S}^{(t)}\}_{t=1}^T)&\lesssim\sqrt{MNT\log(MNT)}\\ &\quad+MN^2C_V\log^2(MNT)+O(1)\\ &=\tilde{O}(\sqrt{MNT}+MN^2)\end{aligned} \tag{11}$$

We make several remarks on the regret upper bound in Corollary 1. In online and bandit learning literature, the time horizon T is usually considered to be the dominating term asymptotically. Therefore, when $T>M$ and the number of items per nest N is small as compared to T, the dominating term in Equation (11) is $\tilde{O}(\sqrt{MNT})$. This matches the lower bound result $\Omega(\sqrt{MT})$ in Theorem 2 within a factor of $\sqrt{N}$. We give further discussion on this gap of $O(\sqrt{N})$ in section 4. We will also show later in section 3.3 how to deal with a large N case by considering a "discretization" heuristic.

In the rest of the section we sketch key steps and lemmas toward the proof of Theorem 1. The detailed proofs of these lemmas are provided in the supplementary material. First, the following lemma shows that the estimates $\hat{\phi}_{i,\theta},\hat{u}_{i,\theta}$ concentrate around the true values $\phi_{i,\theta},u_{i,\theta}$.

LEMMA 4. *Suppose $T(i,\theta)\ge 96\ln(2MTK)$. With probability $1-T^{-1}$ uniformly over all $i\in[M]$, $\theta\in\mathcal{K}_i$ and $t\in[T]$*

$$|\hat{u}_{i,\theta}-u_{i,\theta}|\le\min\left\{U,3\sqrt{\frac{48\max(\hat{u}_{i,\theta},\hat{u}_{i,\theta}^2)\ln(2MTK)}{T(i,\theta)}}+\frac{144\ln(2MTK)}{T(i,\theta)}\right\}; \tag{12}$$

$$|\hat{\phi}_{i,\theta}-\phi_{i,\theta}|\le\min\left\{1,\sqrt{\frac{\ln(2MTK)}{T(i,\theta)\hat{u}_{i,\theta}}}\right\}. \tag{13}$$

In addition, if $u_{i,\theta}\ge 1$ then $\hat{u}_{i,\theta}\in[0.5u_{i,\theta},2u_{i,\theta}]$.

The following corollary is an immediate consequence of Lemma 4:

COROLLARY 2. *Suppose $T(i,\theta)\ge 96\ln(2MTK)$. With probability $1-T^{-1}$, $\bar{u}_{i,\theta}\ge u_{i,\theta}$ and $\bar{\phi}_{i,\theta}\ge\phi_{i,\theta}$ for all $i\in[M]$, $\theta\in\mathcal{K}_1\times\cdots\times\mathcal{K}_M$.*

Corollary 2 shows that (with high probability) $\bar{u}_{i,\theta}$ and $\bar{\phi}_{i,\theta}$ are valid upper bounds for $u_{i,\theta}$ and $\phi_{i,\theta}$. Our next corollary shows that $\bar{R}'$ is also an upper bound for R' at maximizers of $\bar{R}'$ and $\bar{R}$. Recall that $\bar{R}'(\boldsymbol{\theta})=[\sum_{i=1}^M\bar{\phi}_{i,\theta_i}\bar{u}_{i,\theta_i}]/[1+\sum_{i=1}^M\bar{u}_{i,\theta_i}]$ and $R'(\boldsymbol{\theta})=[\sum_{i=1}^M\phi_{i,\theta_i}u_{i,\theta_i}]/[1+\sum_{i=1}^M u_{i,\theta_i}]$.

We defer its proof to the online supplement.

COROLLARY 3. *With probability $1-T^{-1}$, $\bar{R}'(\hat{\boldsymbol{\theta}})\ge R'(\hat{\boldsymbol{\theta}})$ and $\bar{R}'(\boldsymbol{\theta}^*)\ge R'(\boldsymbol{\theta}^*)$, where $\hat{\boldsymbol{\theta}},\boldsymbol{\theta}^*\in\mathcal{K}_1\times\cdots\times\mathcal{K}_M$ are maximizers of $\bar{R}'$ and R', respectively.*

We are now ready to sketch the proof of Theorem 1. The first step is to use the classical regret decomposition for UCB-type policies ($\mathcal{A}$ denotes the success event in Corollary 3).

$$\begin{aligned}\mathrm{Regret}(\{\hat{\boldsymbol{\theta}}^{(t)}\}_{t=1}^T)&=\mathbb{E}\sum_{t=1}^T R'(\boldsymbol{\theta}^*)-R'(\boldsymbol{\theta}^{(t)})\\ &\le\mathbb{E}\left[\sum_{t=1}^T R'(\boldsymbol{\theta}^*)-R'(\boldsymbol{\theta}^{(t)})\middle|\mathcal{A}\right]\Pr[\mathcal{A}]\\ &\quad+O(T)\cdot\Pr[\mathcal{A}^c]\\ &\le O(1)+\mathbb{E}\left[\sum_{t=1}^T\bar{R}'(\boldsymbol{\theta}^*)-\bar{R}'(\boldsymbol{\theta}^{(t)})\right.\\ &\quad\left.+\bar{R}'(\boldsymbol{\theta}^{(t)})-R'(\boldsymbol{\theta}^{(t)})\middle|\mathcal{A}\right]\end{aligned} \tag{14}$$

$$\leq O(1) + \mathbb{E}\left[\sum_{t=1}^{T} \bar{R}'(\boldsymbol{\theta}^{(t)}) - R'(\boldsymbol{\theta}^{(t)})\middle|\mathcal{A}\right]. \quad (15)$$

$$= O(1) + \mathbb{E}\left[\sum_{\tau} |\mathcal{E}_\tau| \cdot (\bar{R}'(\hat{\boldsymbol{\theta}}^{(\tau)}) - R'(\hat{\boldsymbol{\theta}}^{(\tau)}))\middle|\mathcal{A}\right]. \quad (16)$$

Here, $\hat{\boldsymbol{\theta}}^{(\tau)}$ denotes any $\boldsymbol{\theta}^{(t)}$ in the τ-th epoch $\mathcal{E}_\tau$[1]. We also note that Equation (14) holds because $\Pr[\mathcal{A}^c] \leq T^{-1}$ and $\bar{R}'(\boldsymbol{\theta}^*) \geq R'(\boldsymbol{\theta}^*)$, and Equation (15) holds because $\bar{R}'(\boldsymbol{\theta}^{(t)}) \geq \bar{R}'(\boldsymbol{\theta}^*)$, since $\boldsymbol{\theta}^{(t)}$ is the maximizer of $\bar{R}'$ at time t.

It remains to upper bound the discrepancy between $\bar{R}'(\hat{\boldsymbol{\theta}}^{(\tau)})$ and $R'(\hat{\boldsymbol{\theta}}^{(\tau)})$ at every epoch τ. This is accomplished by the following "aggregation lemma," which is proved in the online supplement.

LEMMA 5. *With probability $1 - T^{-1}$, for all $t \in [T]$, $i \in [M]$ and $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_M) \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M$,*

$$\bar{R}'(\boldsymbol{\theta}) - R'(\boldsymbol{\theta}) \leq \frac{1}{1 + \sum_{i=1}^{M} u_{i,\theta_i}} \left[\sum_{i=1}^{M} \frac{\bar{u}_{i,\theta_i} - u_{i,\theta_i}}{1 + u_{i,\theta_i}} + \sum_{i=1}^{M} u_{i,\theta_i}(\bar{\phi}_{i,\theta_i} - \phi_{i,\theta_i})\right]. \quad (17)$$

REMARK 1. Comparing Lemma 5 with Lemma A.4 from (Agrawal et al. 2019), we can see that there is an additional $1/[1 + \sum_{i \in \mathcal{M}} u_{i,\theta}]$ multiplication term in the error upper bounds. Such an improvement is made possible by our more careful analysis and insights into the mathematical structures of the MNL choice model, and is important in dealing with preference parameters v_{ij} larger than one.

Note that $\mathbb{E}|\mathcal{E}_\tau| = 1 + \sum_{i=1}^{M} \mathbb{E}[\hat{n}_{i,\tau}] = 1 + \sum_{i=1}^{M} u_{i,\theta_i}$. Combining Lemma 5 with Equation (16) we obtain

$$\text{Regret}(\{\hat{\boldsymbol{\theta}}^{(t)}\}_{t=1}^{T}) \leq O(1) + \sum_{\tau} \mathbb{E}\left[\sum_{i=1}^{M} \frac{\bar{u}_{i,\hat{\theta}_i^{(\tau)}} - u_{i,\hat{\theta}_i^{(\tau)}}}{1 + u_{i,\hat{\theta}_i^{(\tau)}}} + \sum_{i=1}^{M} u_{i,\hat{\theta}_i^{(\tau)}}(\bar{\phi}_{i,\hat{\theta}_i^{(\tau)}} - \phi_{i,\hat{\theta}_i^{(\tau)}})\middle|\mathcal{A}\right]. \quad (18)$$

The following lemmas upper bound (asymptotically) the two terms in Equation (18) separately.

LEMMA 6. *Conditioned on event $\mathcal{A}$, it holds that*

$$\sum_{\tau} \sum_{i=1}^{M} \frac{\bar{u}_{i,\hat{\theta}_i^{(\tau)}} - u_{i,\hat{\theta}_i^{(\tau)}}}{1 + u_{i,\hat{\theta}_i^{(\tau)}}} \lesssim \sqrt{MKT \log(MTK)} + MKU \log^2(MTK). \quad (19)$$

LEMMA 7. *Conditioned on event $\mathcal{A}$, it holds that*

$$\sum_{\tau} \sum_{i=1}^{M} u_{i,\hat{\theta}_i^{(\tau)}}(\bar{\phi}_{i,\hat{\theta}_i^{(\tau)}} - \phi_{i,\hat{\theta}_i^{(\tau)}}) \lesssim \sqrt{MKT \log(MTK)} + MKU \log^2(MTK). \quad (20)$$

Lemmas 6 and 7 are proved in the supplementary material. Combining both lemmas and Equation (18), we complete the proof of Theorem 1.

3.3. A Discretization Heuristic

When the number of items N per nest is large, we present a useful discretization heuristic that *discretizes* the parameter sets $\mathcal{K}_i$ into small finite subsets. In other words, instead of considering level sets defined for thresholds $\theta = r_{ij}$ for all $j \in [N]$ so that $|\mathcal{K}_i| = N + 1$, we only include level sets whose thresholds are on a finite grid. Our simulation studies (see section 5) demonstrate the effectiveness of this method.

More specifically, let $\delta \in (0,1)$ be a granularity parameter to be specified by the retailer. Recall the definition of the level set $\mathcal{L}_i(\theta) = \{j \in [N] : r_{ij} \geq \theta\}$. In the discretized framework, we only consider level set threshold parameters θ that are multiples of $1/\delta$. Let $\mathbb{N}$ be the set of non-negative integers and define

$$\tilde{\mathcal{K}}_i^\delta := \{\theta : 0 \leq \theta \leq 1, \theta/\delta \in \mathbb{N}, \mathcal{L}_i(\theta)' \text{sare distinct}\} \cup \{\infty\}, \quad \text{for } i \in [M] \quad (21)$$

where each $\theta \in \tilde{\mathcal{K}}_i^\delta$ corresponds to a *unique* level set $\mathcal{L}_i(\theta)$. When there are multiple θ's leading to the same level set, we keep any one of these θ's in $\tilde{\mathcal{K}}_i^\delta$, and thus the level sets induced by $\tilde{\mathcal{K}}_i^\delta$ (i.e., $\{\mathcal{L}_i(\theta) : \theta \in \tilde{\mathcal{K}}_i^\delta\}$) are unique. Since duplicate assortment sets are removed in $\tilde{\mathcal{K}}_i^\delta$, we have $\tilde{\mathcal{K}}_i^\delta \subseteq \mathcal{K}_i$, and thus $|\tilde{\mathcal{K}}_i^\delta| \leq |\mathcal{K}_i| = K = N + 1$. Moreover, we also have $|\tilde{\mathcal{K}}_i^\delta| \leq \lfloor 1/\delta \rfloor + 2$ because level set thresholds in $\tilde{\mathcal{K}}_i^\delta$ must be an integer multiple of δ. On one hand, when δ is not too small, the size of $\tilde{\mathcal{K}}_i^\delta$ could be significantly smaller than N. On the other hand, when $\delta \to 0$, we recover the original set $\mathcal{K}_i$, which gives the full level sets. We shall, thus, define $\tilde{\mathcal{K}}_i^\delta := \mathcal{K}_i$ when δ=0.

The following discretized reduction lemma shows that by restricting ourselves to $\tilde{\mathcal{K}}_i^\delta$ instead of $\mathcal{K}_i$, the approximation error in terms of expected revenue can be upper bounded by δ, which goes to zero as we take $\delta \to 0$.

LEMMA 8. (DISCRETIZED REDUCTION LEMMA). *Fix an arbitrary $\delta \in (0,1)$. Then*

$$\max_{\boldsymbol{\theta}\in\mathcal{K}_1\times\cdots\times\mathcal{K}_M} R'(\boldsymbol{\theta}) - \max_{\boldsymbol{\theta}\in\tilde{\mathcal{K}}_1^\delta\times\cdots\times\tilde{\mathcal{K}}_M^\delta} R'(\boldsymbol{\theta}) \le \delta,$$

where $R'(\boldsymbol{\theta}) := [\sum_{i=1}^M \phi_{i,\theta_i} u_{i,\theta_i}]/[1+\sum_{i=1}^M u_{i,\theta_i}]$.

With a pre-specified δ, we run the policy in Algorithm 1 on the parameter space $\tilde{\mathcal{K}}_1^\delta\times\cdots\times\tilde{\mathcal{K}}_M^\delta$. As a result of Lemma 8, the value of δ can be thought of as a trade-off between additive bias and multiplicative terms in the final regret. With a small value of δ, there is almost no additive terms arising from Lemma 8, yet the number of items N per nest will not be reduced too much. Moreover, when δ is large the regret bound in Corollary 1 is improved as the number of items N per nest is now upper bounded by $\lfloor 1/\delta \rfloor$+2. However, a large δ value will introduce a large additive bias from Lemma 8. Hence, a balance has to be achieved for an appropriate value of δ to deliver the best performance. We further demonstrate the performance for different choices of δ in our simulation studies (see section 5).

4. A Regret Lower Bound

We establish the following lower bound on the regret of any dynamic assortment planning policy under nested logit models.

THEOREM 2. *Suppose the number of nests M is divisible by* 4 *and* $\gamma_1 = \cdots = \gamma_M = 0.5$. *Assume also that* (*A*1) *and* (*A*2) *hold. Then there exists a numerical constant* $C_0 > 0$ *such that for any dynamic assortment planning policy* π,

$$\sup_{\{r_{ij},v_{ij}\}} \sum_{t=1}^T R^* - \mathbb{E}^\pi[R(\mathbf{S}^{(t)})] \ge C_0\sqrt{MT} \text{ where } R^* = \max_{\mathbf{S}\in\mathbb{S}} R(\mathbf{S}). \quad (22)$$

REMARK 2. The condition that M is divisible by 4 is only a technical condition and does not affect the main message delivered in Theorem 2, which shows necessary dependency on M asymptotically when M is large.

REMARK 3. (Discussion on the dependency of M) Comparing Theorem 2 with the regret upper bound in Corollary 1, we notice that when T (time horizon) is large compared to M (the number of nests), both regret bounds have an $O(\sqrt{M})$ dependency on M. This suggests that our algorithm and regret analysis delivers *optimal* dependency of regret on the number of nests M in a dynamic nested assortment planning problem.

REMARK 4. (Discussion on the dependency of N) Comparing Theorem 2 with the regret upper bound in Corollary 1, we notice that there is a gap of $\sqrt{N}$ between the upper and lower bounds.

We conjecture that the *upper bound* with an additional $O(\sqrt{N})$ factor is in fact tight. Actually, because our proposed algorithm treats each "level set" assortments (within each nest) as standalone estimation units, it is intuitive to see that the regret that *our algorithm incurs* has to scale polynomially with N. We conjecture that *any* possible dynamic strategy for nested logit models has to suffer at least an $O(\sqrt{N})$ term in regret bound.

Unfortunately, due to technical difficulty of constructing lower bounds for problem instances, we are unable to extend our lower bound constructions to more than N=3 items per nest. This is because our lower bound construction (to be presented later) uses only N=3 items per nest and therefore cannot deliver a lower bound depending on N. We, thus, leave the question of proving a matching $O(\sqrt{MNT})$ lower bound as an interesting yet challenging open problem.

In the rest of this section, we provide the proof of Theorem 2, while deferring proofs of several technical lemmas to the supplementary material.

4.1. Construction of Adversarial Model Parameters

Let ε>0 be a small positive parameter depending on M and T, which will be specified later. Each nest $i \in [M]$ in our construction consists of N=3 items and is classified into two categories: "Type A" and "Type B," with parameter configurations detailed in Table 1. Note that regardless of which type of nest $i \in [M]$ is, the three items in nest i have preference parameters $(1+\epsilon)/M^2$, $(1-\epsilon)/M^2$ and $1/M^2$. Hence it is impossible to decide the type of a nest without observations of customers' purchasing actions. Given the model parameters in Table 1, it is easy to verify that for a Type A nest, the optimal assortment is {1,2}, while for a Type B nest, the optimal assortment is {1,2,3}.

The following lemma shows that any assortment S_i that does not equal {1,2} for Type A nests or {1,2,3} for Type B nests incurs an $\Omega(\varepsilon/M)$ regret. It is proved in the supplementary material.

LEMMA 9. *Let* $U\subseteq[M]$ *be the set of Type A nests, and by construction* $[M]\setminus U$ *are all Type B nests. For any* $\mathbf{S} = (S_1,\cdots,S_M) \in [N]^M$, *define* $m_U^\sharp(\mathbf{S}) := \sum_{i\in U}\mathbf{1}\{S_i \ne \{1,2\}\} + \sum_{i\notin U}\mathbf{1}\{S_i \ne \{1,2,3\}\}$. *Then there exists*

Table 1 Adversarial Construction of Two Types of Nests

	Type A Nest			Type B Nest		
	Item 1	Item 2	Item 3	Item 1	Item 2	Item 3
Revenues r_{ij}	1	0.8	ρ	1	0.8	ρ
Preferences v_{ij}	$(1+\epsilon)/M^2$	$(1-\epsilon)/M^2$	$1/M^2$	$(1-\epsilon)/M^2$	$(1+\epsilon)/M^2$	$1/M^2$

Notes: The revenue parameter ρ is set to $\rho = 9\sqrt{2}/(1+\sqrt{2}) \approx 0.694774$.

a numerical constant $C>0$ such that for all $\mathbf{S}$, $R(\mathbf{S}^*) - R(\mathbf{S}) \geq m_U^\sharp(\mathbf{S}) \cdot C\epsilon/M$, where $\mathbf{S}^* \in \arg\max_{\mathbf{S}} R(\mathbf{S})$ is the optimal assortment combination under U.

To avoid confusion, we emphasize that in our *lower bound proof* the notation U refers to a particular type of nest, instead of upper confidence bounds in algorithm descriptions and the upper bound proof.

4.2. Reduction to Average-Case Regret

For any policy π, we want to show a lower bound on the *worst-case* regret

$$\sup_{\{r_{ij},v_{ij}\}} \sum_{t=1}^{T} R^* - \mathbb{E}^\pi\Big[R(\mathbf{S}^{(t)})\Big]. \tag{23}$$

Recall that in our adversarial construction, $U\subseteq[M]$ denotes the set of all Type A nests and the remaining nests $[M]\setminus U$ are Type B. The following inequalities show a reduction to average-case regret:

$$\begin{aligned} &\sup_{\{r_{ij},v_{ij}\}} \sum_{t=1}^{T} R^* - \mathbb{E}^\pi[R(\mathbf{S}^{(t)})] \\ &\geq \sup_{U\subseteq[M]} \sum_{t=1}^{T} R^* - \mathbb{E}_U^\pi\Big[R(\mathbf{S}^{(t)})\Big] \qquad (24) \\ &\geq \frac{1}{2^M} \sum_{U\subseteq[M]} \sum_{t=1}^{T} R^* - \mathbb{E}_U^\pi\Big[R(\mathbf{S}^{(t)})\Big], \end{aligned}$$

where in $\sup_{U\subseteq[M]}$ or $\sum_{U\subseteq[M]}$ we are optimizing or summing over all 2^M subsets of $[M]=\{1,2,\cdots,M\}$. Here, we also use the $\mathbb{E}_U^\pi$ notation to emphasize that the distribution of $\{\mathbf{S}^{(t)}\}$ (and hence the expectation) depends on both the parameter setting (uniquely determined by the set of Type A nests $U\subseteq[M]$) and the policy π itself.

For any $i \in [M]$ and $S\subseteq[N]$, denote $\mathfrak{n}_{\mathfrak{S}}(\mathfrak{i}) = \sum_{\mathfrak{t}=}^{\mathfrak{T}} \mathbf{1}\{\mathfrak{S}_{\mathfrak{i}}^{(\mathfrak{t})} = \mathfrak{S}\}$ as the random variable of the number of times assortment S is offered in nest i. Let $\mathbb{E}_U^\pi[\mathfrak{n}_{\mathfrak{S}}(\mathfrak{i})]$ be the expectation of $\mathfrak{n}_{\mathfrak{S}}(\mathfrak{i})$, with expectation taken under model parameters setting U (recall that U is the set of all Type A nests) and policy π. Invoking Lemma 9 and noting that $\sum_{S\subseteq[N]} \mathbb{E}_U^\pi[\mathfrak{n}_{\mathfrak{S}}(\mathfrak{i})] = \mathfrak{T}$ for any $U\subseteq[M]$, $i \in [M]$ and policy π, the right-hand side of Equation (24) can be lower bounded by,

$$\begin{aligned} &\frac{1}{2^M} \sum_{U\subseteq[M]} \sum_{t=1}^{T} \mathbb{E}^\pi\left[m_U^\sharp(\mathbf{S}^{(t)}) \cdot \frac{C\epsilon}{M}\right] \\ &= \frac{C\epsilon}{M} \frac{1}{2^M} \sum_{U\subseteq[M]} \left[\sum_{i\in U} \sum_{S\neq\{1,2\}} \mathbb{E}_U^\pi[\mathfrak{n}_{\mathfrak{S}}(\mathfrak{i})] + \sum_{\mathfrak{i}\notin\mathfrak{U}} \sum_{\mathfrak{S}\neq\{,,\}} \mathbb{E}_{\mathfrak{U}}^\pi[\mathfrak{n}_{\mathfrak{S}}(\mathfrak{i})]\right] \\ &\geq \frac{C\epsilon}{M} \frac{1}{2^M} \sum_{U\subseteq[M]} \left[\sum_{i\in U} \sum_{S\neq\{1,2\}} \mathbb{E}_U^\pi[\mathfrak{n}_{\mathfrak{S}}(\mathfrak{i})] + \sum_{\mathfrak{i}\notin\mathfrak{U}} \mathbb{E}_{\mathfrak{U}}^\pi[\mathfrak{n}_{\{,\}}(\mathfrak{i})]\right] \\ &= \frac{C\epsilon}{M} \frac{1}{2^M} \sum_{U\subseteq[M]} \left[\sum_{i\in U} (T - \mathbb{E}_U^\pi[\mathfrak{n}_{\{,\}}(\mathfrak{i})]) + \sum_{\mathfrak{i}\notin\mathfrak{U}} \mathbb{E}_{\mathfrak{U}}^\pi[\mathfrak{n}_{\{,\}}(\mathfrak{i})]\right] \end{aligned} \tag{25}$$

$$= \frac{C\epsilon}{M} \frac{1}{2^M} \left(2^M \times \frac{MT}{2} - \sum_{U\subseteq[M]} \left[\sum_{i\in U} \mathbb{E}_U^\pi[\mathfrak{n}_{\{,\}}(\mathfrak{i})] - \sum_{\mathfrak{i}\notin\mathfrak{U}} \mathbb{E}_{\mathfrak{U}}^\pi[\mathfrak{n}_{\{,\}}(\mathfrak{i})]\right]\right) \tag{26}$$

$$= \frac{C\epsilon T}{2} - \frac{C\epsilon}{M} \frac{1}{2^M} \sum_{U\subseteq[M]} \sum_{i=1}^{M} (-1)^{\mathbf{1}\{i\in U\}} \times \mathbb{E}_U^\pi[\mathfrak{n}_{\{,\}}(\mathfrak{i})]. \tag{27}$$

Here in Equation (25) we use the fact that $\sum_{S\subseteq[N]} \mathbb{E}_u^\pi[\mathfrak{n}_{\mathfrak{S}}(\mathfrak{i})] = \mathfrak{T}$; Equation (26) holds because $\sum_{U\subseteq[M]} \sum_{i\in U} T = \sum_{U\subseteq[M]} \sum_{i\notin U} T$ by symmetry, and furthermore $\sum_{U\subseteq[M]} (\sum_{i\in U} T + \sum_{i\notin U} T) = \sum_{U\subseteq[M]} \sum_{i=1}^{M} T = 2^M \times MT$.

Next, for every $U\subseteq[M]$, define $U' = U \oplus i$ as $U' = U \cup \{i\}$ if $i \notin U$, and $U' = U\setminus\{i\}$ if $i \in U$. Clearly, there is a one-to-one correspondence between $U\subseteq[M]$ and $U\oplus i\subseteq[M]$, for every fixed $i \in [M]$. The right-hand side of Equation (27) can then be simplified as

$$\begin{aligned} &\frac{C\epsilon}{2} - \frac{C\epsilon}{M} \frac{1}{2^M} \sum_{i=1}^{M} \frac{1}{2} \left[\sum_{U\subseteq[M]} (-1)^{\mathbf{1}\{i\in U\}} \mathbb{E}_U^\pi[n\{1,2\}(i)] \right. \\ &\left. + \sum_{U\subseteq[M]} (-1)^{\mathbf{1}\{i\in U\oplus i\}} \mathbb{E}_{U\oplus i}^\pi[n\{1,2\}(i)]\right] \\ &= \frac{C\epsilon}{2} - \frac{C\epsilon}{M} \frac{1}{2^{M+1}} \sum_{i=1}^{M} \sum_{U\subseteq[M]} (-1)^{\mathbf{1}\{i\in U\}} \\ &\times \left(\mathbb{E}_U^\pi[n_{\{1,2\}}(i)] - \mathbb{E}_{U\oplus i}^\pi[n_{\{1,2\}}(i)]\right). \end{aligned} \tag{28}$$

4.3. Pinsker's Inequality

Let P^π_U, P^π_W denote the probabilistic laws under U, W and policy π. Then, for any $S \subseteq [N]$,

$$\left|\mathbb{E}^\pi_U[n_S(i)] - \mathbb{E}^\pi_W[n_S(i)]\right| \le \sum_{j=0}^{T} j \cdot |P^\pi_U[n_S(i)=j] - P^\pi_W[n_S(i)=j]| \le T \cdot \sum_{j=0}^{T} \left|P^\pi_U[n_S(i)=j] - P^\pi_W[n_S(i)=j]\right|$$
$$= T\|P^\pi_U - P^\pi_W\|_{\mathrm{TV}} \le T\sqrt{\frac{1}{2}\min\{\mathrm{KL}(P^\pi_U\|P^\pi_W), \mathrm{KL}(P^\pi_W\|P^\pi_U)\}} \tag{29}$$

$$\le T\sqrt{\frac{T}{2}\min\{\max_{\mathbf{S}}\mathrm{KL}(P_U(\cdot|\mathbf{S})\|P_W(\cdot|\mathbf{S})), \max_{\mathbf{S}}\mathrm{KL}(P_W(\cdot|\mathbf{S})\|P_U(\cdot|\mathbf{S}))\}}. \tag{30}$$

Here $\|P - Q\|_{\mathrm{TV}}$ and $\mathrm{KL}(P\|Q)$ denote the total variational distance and Kullback-Leibler divergence between two probability laws P and Q. Equation (29) is known as the *Pinsker's inequality* (see e.g., Tsybakov (2009), Csiszar and Körner (2011)). Note that in the last term P_U and P_W do not have superscript π, because conditioned on a particular assortment combination $\mathbf{S}$ the KL divergence no longer depends on π.

The following lemma shows that if U and W differ by only one nest, then the KL divergence between P_U and P_W is small *for all* $\mathbf{S} = (S_1, \cdots, S_M)$.

LEMMA 10. *Suppose* $|U \triangle W|=1$, *where* $U \triangle W = (U \setminus W) \cup (W \setminus U)$ *is the symmetric difference between subsets* $U, W \subseteq [M]$. *Then there exists a universal constant* $C' > 0$ *such that for any* $\mathbf{S} = (S_1, \cdots, S_M)$, $\min\{\mathrm{KL}(P_U(\cdot|\mathbf{S})\|P_W(\cdot|\mathbf{S}), \mathrm{KL}(P_W(\cdot|\mathbf{S})\|P_U(\cdot|\mathbf{S}))\}) \le C'\epsilon^2/M$.

Invoking Lemma 10, the right-hand side of Equation (30) can be further upper bounded by

$$T\sqrt{\frac{T}{2}\cdot\frac{C'\epsilon^2}{M}} \lesssim T\sqrt{T\epsilon^2/M}. \tag{31}$$

We are now ready to prove Theorem 2 by simplifying the Equation (28) with the help of Equations (30) and (31). For every $U \subseteq [M]$ and $i \in [M]$, by Equation (30) it holds that

$$\left|\mathbb{E}^\pi_U[n_{\{1,2\}}(i)] - \mathbb{E}^\pi_{U \oplus i}[n_{\{1,2\}}(i)]\right| \lesssim T\sqrt{T\epsilon^2/M}.$$

Subsequently, Equation (28) can be lower bounded as

$$\frac{C\epsilon}{2} - \frac{C\epsilon}{M2^{M+1}}\sum_{i=1}^{M}\sum_{U \subseteq [M]} O(T\sqrt{T\epsilon^2 M})$$
$$\ge \frac{C\epsilon}{2} - C\epsilon \times O(T\sqrt{T\epsilon^2/M}).$$

Setting $\epsilon = c_0\sqrt{M/T}$ for some sufficiently small positive constant $c_0 > 0$, the above inequality is lower bounded by $\Omega(\epsilon T) = \Omega(\sqrt{M/T})$. This completes the proof of Theorem 2.

5. Numerical Results

We present numerical studies of our proposed policies for dynamic nested assortment planning on synthetic data. The main focus of our simulation is the regret of our policies under various model parameter settings of M, N, and T, as well as the effect of the discretization granularity $\delta \in [0,1]$ on the regret.

For each nest $i \in [M]$, the revenue parameters $\{r_{ij}\}_{j=1}^N$ are independently and identically distributed from the uniform distribution on [0.2,0.8] and the preference parameters $\{v_{ij}\}_{j=1}^N$ are independently and identically distributed from the uniform distribution on $[10/N(M-1), 20/N(M-1)]$, where N is the number of items in each nest. The nest discounting parameters $\{\gamma_i\}_{i=1}^M$ are independently and identically distributed from the uniform distribution on [0.5,1].

We consider the different combinations of parameters in terms of M (the number of nests), N (the number of items per nest), T (time horizon length), and δ (the granularity parameter in the heuristic discretized policy). We note that δ=0 means that no discretization is carried out. For each (M,N) settings, we generate model parameters $\{r_{ij}, v_{ij}, \gamma_i\}_{i,j=1}^{M,N}$ as described in the previous paragraph, and then run the dynamic assortment policy for 100 independent trials. The median and maximum accumulated regret over T periods are reported.

In Table 2, we compare the accumulated regret of our proposed policies with different granularity parameters δ, under a range of different parameter settings of number of nests M, number of items per nest N, and time horizon T. We also compare the performances of our algorithms with some competitive algorithm baselines, such as the Thompson Sampling (TS) algorithm and the Explore-then-Exploit (Exp-Exp) algorithm. The TS algorithm works similar to our UCB algorithm (with δ=0), where the level-set assortments are first generated for each nest and the algorithm learns and optimizes the assortment while maintaining Beta priors for the aggregate parameters, as suggested in (Agrawal et al. 2017). In the Exp-Exp algorithm, the level-set assortments are also first generated. However, the exploration phase (learning the aggregate parameters) and the exploitation phase (exploiting the estimated optimal assortment) are separated.

In Figure 1, we further plot the accumulated and average regret of our policies for time horizon when T is large (T between 10^5 and 10^7). From both Table 2

Table 2 Median (Med) and Maximum (Max) Accumulated Regret (summation over *T* periods) for Various Algorithms and Under Various Model and Parameter Settings

	δ=0		$\delta = 10^{-3}$		$\delta = 5 \times 10^{-3}$		$\delta = 10^{-2}$		$\delta = 5 \times 10^{-2}$		TS		Exp-Exp	
(*M*,*N*)	Med	Max	Med	Max	Med	Max	Med	Max	Med	Max	Med	Max	Med	Max
T=100:														
(5,100)	5.5	6.4	5.5	6.0	3.8	**4.1**	**3.2**	4.3	5.4	8.5	6.4	10.2	6.4	18.8
(10,100)	4.8	6.2	5.4	5.5	4.7	6.5	**2.3**	**3.9**	5.8	7.0	6.7	11.7	6.6	25.3
(5,250)	10.4	14.1	9.8	12.0	5.7	6.5	**3.3**	**3.4**	7.0	8.3	6.1	12.1	6.6	22.3
(10,250)	10.8	12.0	9.7	12.3	5.5	7.4	**3.0**	**4.4**	5.1	8.7	6.8	9.2	6.1	17.4
(5,1000)	22.0	25.3	16.0	18.2	6.2	7.5	**3.2**	**5.0**	6.9	10.9	6.1	10.1	5.5	21.6
(10,1000)	21.5	24.1	15.1	17.7	5.1	6.4	**3.1**	**4.9**	6.2	9.4	6.3	9.1	6.4	24.8
T=500:														
(5,100)	**14.3**	**18.5**	18.3	22.6	26.8	30.9	31.9	35.3	33.3	34.3	32.6	42.7	25.6	88.9
(10,100)	**15.7**	23.0	16.5	**22.1**	28.4	28.9	35.4	36.5	35.0	36.5	33.0	42.9	30.7	128.0
(5,250)	14.2	17.3	**12.7**	**14.9**	16.4	18.4	29.1	36.8	32.6	34.2	30.6	43.4	26.2	105.1
(10,250)	13.8	**15.9**	**13.0**	17.4	16.6	19.6	29.2	35.0	35.8	38.6	33.2	39.5	30.5	47.4
(5,1000)	41.1	46.1	22.7	25.7	**14.1**	**17.3**	29.4	37.4	33.0	35.8	30.8	39.8	27.0	94.3
(10,1000)	39.3	44.2	21.0	27.2	**13.7**	**18.7**	28.0	37.0	35.7	41.5	32.4	39.8	29.3	49.4
$T = 10000$:														
(5,100)	491.5	505.5	**489.4**	**496.5**	494.5	500.8	503.1	511.8	513.4	525.2	579.9	672.2	538.9	904.7
(10,100)	548.4	558.0	548.6	552.9	**529.3**	**534.7**	538.2	544.3	554.3	565.2	618.0	706.7	572.4	883.1
(5,250)	534.4	543.7	529.7	543.9	523.4	536.1	519.7	**525.5**	526.1	532.2	617.7	694.3	**477.9**	970.9
(10,250)	551.0	560.5	554.5	563.3	**547.4**	555.2	548.6	**555.1**	571.6	578.4	642.1	714.4	558.6	928.6
(5,1000)	669.0	704.4	570.5	584.8	538.8	552.7	532.9	**541.3**	535.8	558.4	621.6	671.8	**489.7**	863.3
(10,1000)	703.5	738.2	613.1	633.6	555.7	566.3	**549.9**	**559.5**	567.2	578.6	646.6	697.4	560.7	911.0

Notes: The minimum regret for each case is highlighted using the bold font. TS stands for Thompson Sampling, and Exp-Exp stands for Explore-then-Exploit.

and Figure 1, one can see a clear pattern of sub-linear accumulated regret. Moreover, when N is small as compared to T, a smaller discretization granularity leads to better empirical performance; while when N is large, a larger discretization granularity is better. We can observe from both Table 2 and Figure 1 that our algorithm with the appropriate granularity level δ consistently outperforms the two baseline methods. The TS algorithm performs similarly to our UCB algorithm with δ=0. It is also possible to combine the TS algorithm with our discretization heuristic (i.e., setting δ to be a positive value), and we would expect the similar performance as its UCB counterpart with the same discretization parameter δ. For simplicity and interpretability of the figures, we omit those performance curves.

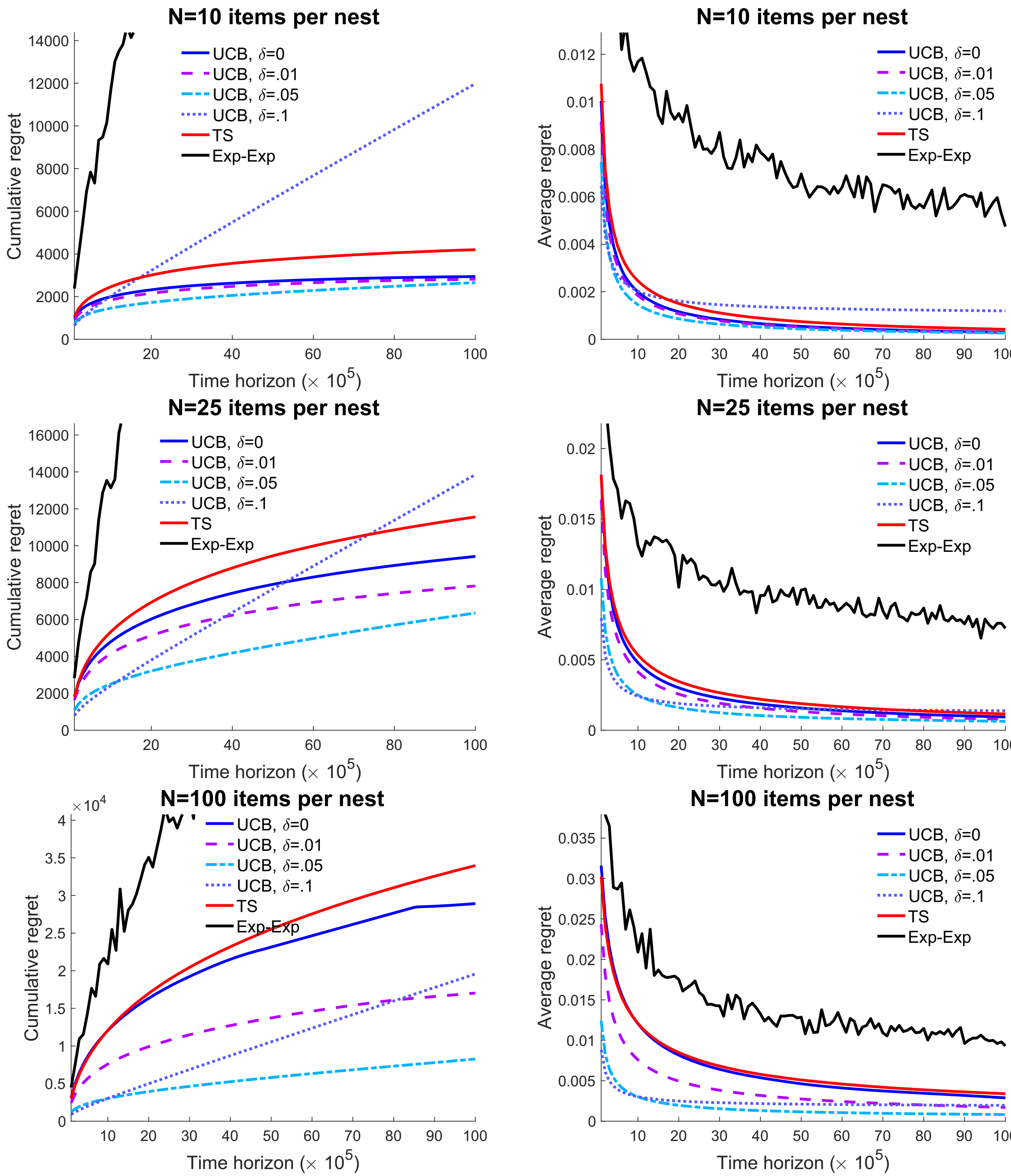
We also remark that, because of the inherent instability of the Exp-Exp algorithm (as a consequence of the fact that Exp-Exp commits to a *single* assortment for the majority of T time periods), the curve for Exp-Exp displayed in Figure 1 is much more wiggly compared to the other algorithms that are more stable with smaller variance.

One principle for choosing an appropriate value for δ is based on the time horizon T. When T is larger, the inherent bias could result in a δT cumulative regret that is linear with T, which is typically reflected by the UCB with δ=.1 curves in all settings of Figure 1. (For large enough T, a linear growth can also be observed for some of the UCB with δ=0.05 curves.) Therefore, in the long term, a smaller δ would reduce the negative impact to the cumulative regret. However, when T is smaller, a larger δ means less amount of aggregate parameters to learn, and therefore the algorithm benefits from a quick start. This is clearly reflected in the N=25 setting of Figure 1 where the UCB with δ=.1 curve enjoys the lowest regret for small T, and gradually loses its advantages as T increases. For future directions, it is a very interesting question to study how to appropriately (and maybe even dynamically) set the values of δ to derive a better theoretical regret bound.

We note that a linear growth of the regret with T would only occur when the algorithm fails to recover the optimal assortment after the discretization process. Moreover, if δ is set to be a small enough value such that the optimal assortment can be found even after the discretization process, such a linear growth of the regret will not occur. In Table 3, we report the percentage of instances in the corresponding settings of Figure 1, where the optimal assortment can be recovered after discretization. We note that a small percentage value corresponds to a linear growth curve in Figure 1.

We also remark that, when N is small, the gap between two-level set assortments in each nest is

Figure 1 **Accumulated (left) and Average (right) Regret of Our Policy and Competitive Policies with M=5 Nests, Varying the Number of Items Per Nest N and the Granularity Parameter δ. TS Stands for the Thompson Sampling Algorithm and Exp-Exp stands for the Exploration-Exploitation Algorithm, with Details in the Main Text [Color figure can be viewed at wileyonlinelibrary.com]**



potentially large and therefore the bias resulting from a large δ value could also be large. This means that when N is small, giving rise to only a few "level-set" assortments, δ cannot be set too large because otherwise the optimal level-set assortment might be missed because of the large gap between integer

Table 3 **Percentage of the Instances in the Settings of Figure 1 Where the Optimal Assortment can be Recovered After Discretization**

	δ=0	δ=.01	δ=.05	δ=.1
N = 10	100%	99%	66%	36%
N = 25	100%	99%	28%	2%
N = 100	100%	99%	1%	0%

multiples of δs. Moreover, when N is large, even if δ is big the potential bias introduced by discretization could still be smaller, because there might be a "level-set" assortment close to every levels of $i\delta$, $i \in \mathbb{N}$, at the discretization level of δ. This means that discretization at a larger value of δ is potentially more beneficial when N is large, because little additional bias is increased but the number of "level-set" assortments to be considered is significantly fewer when δ is large.

Finally, in Figure 2 we compare the cumulative regret of our proposed UCB algorithm (at different levels of discretization granularity δ) by holding M,T fixed and varying the number of products per nest (N). As we observe from Figure 2, when δ=0 (i.e., no discretization carried out), the cumulative regret of our algorithm does scale on the order of $O(\sqrt{N})$ with N, suggesting that our upper bound results in Theorem 1 and Corollary 1 are tight. Figure 2 also shows that with larger discretization granularity level δ, the regret of the proposed UCB algorithm scales more mildly with increasing number of products per nest N.

5.1. Experiments Following the Setting in Davis et al. (2014)

In this subsection, we report the simulation results on a set of different classes of the synthetic problem instances. The synthetic problem instances are generated similar as described in (Davis et al. 2014). The instance is parameterized by $\varepsilon \in (0,1)$. For each nest $i \in [M]$, we first generate the nest discounting parameter γ_i independently from the uniform distribution [0.5,1]. We then generate the first ($N-1$) items as follows. For each $j \in [N-1]$, we independently sample U_{ij} from the uniform distribution over [0,4], X_{ij} from the uniform distribution over [0.1,1], and Y_{ij} from the uniform distribution over [0.01,0.1], and set $r_{ij} = \epsilon^{U_{ij}} \cdot X_{ij}$ and $v_{ij} = \epsilon^{2-U_{ij}} \cdot Y_{ij}$. Finally, we let $r_{iN} = 0$ and $v_{iN} = \epsilon^{-1} \cdot Y_{iN}$ where Y_{iN} is also independently and uniformly sampled from [0.01,0.1]. We note that the main differences between our generating procedure and that of (Davis et al. 2014) are that the range of X_{ij} is [1,10] and the range of Y_{ij} is [0.2,1.8] in (Davis et al. 2014). While the differences for X_{ij} only affect the revenue parameters $\{r_{ij}\}$ (and therefore the revenues of all candidate assortments) up to a scaling factor, the differences for Y_{ij} are because that weight of the no-purchase option is set to 10 in (Davis et al. 2014), but normalized to 1 in our paper. Considering the typical value of the discounting parameter γ_i (which is ~ 0.75), we therefore correspondingly reduce the range of Y_{ij} by $\sim 10^{1/0.75} \approx 20$ times for the normalization purpose. We set M=5 and pick (ε,N) from $\{0.6,0.4\} \times \{25,100\}$ to generate four classes of the problem instances.

In Figure 3, we report the performance of our algorithms and the two baseline algorithms TS and Exp-Exp in these four settings. As one can observe from Figure 3, the comparison between our algorithms and the baseline algorithms are similar to the settings in Figure 1, with one difference that the TS algorithm performs better than our UCB

Figure 2 **Cumulative Regret of Our Algorithm with Varying Number of Products Per Nest (*N*), at Different levels of Discretization Granularity (δ) [Color figure can be viewed at wileyonlinelibrary.com]**

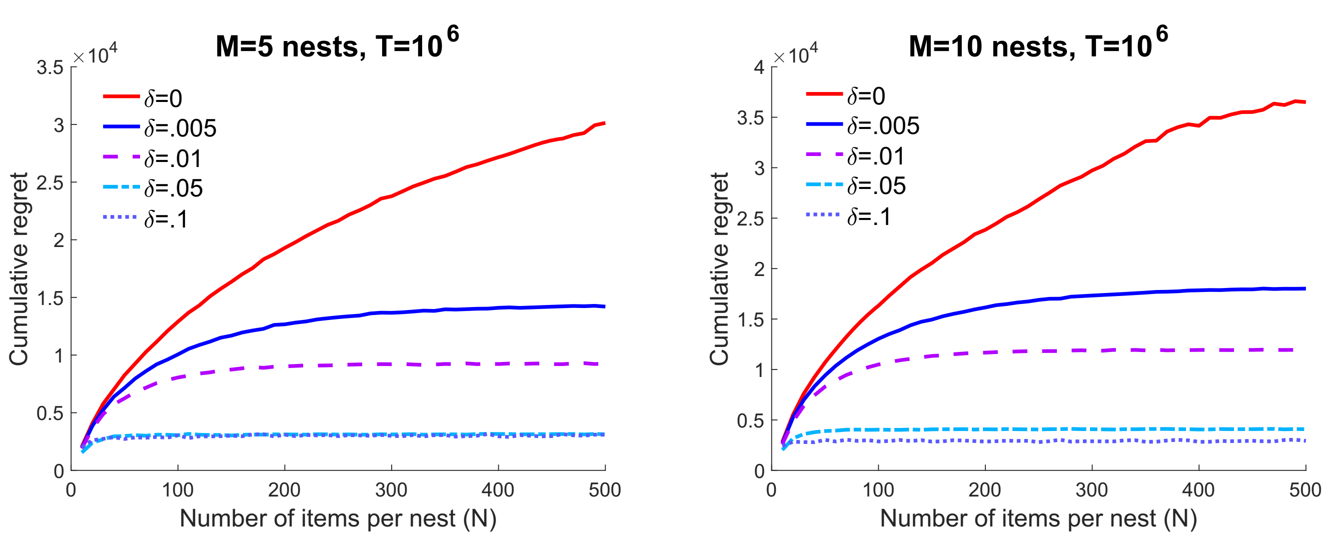
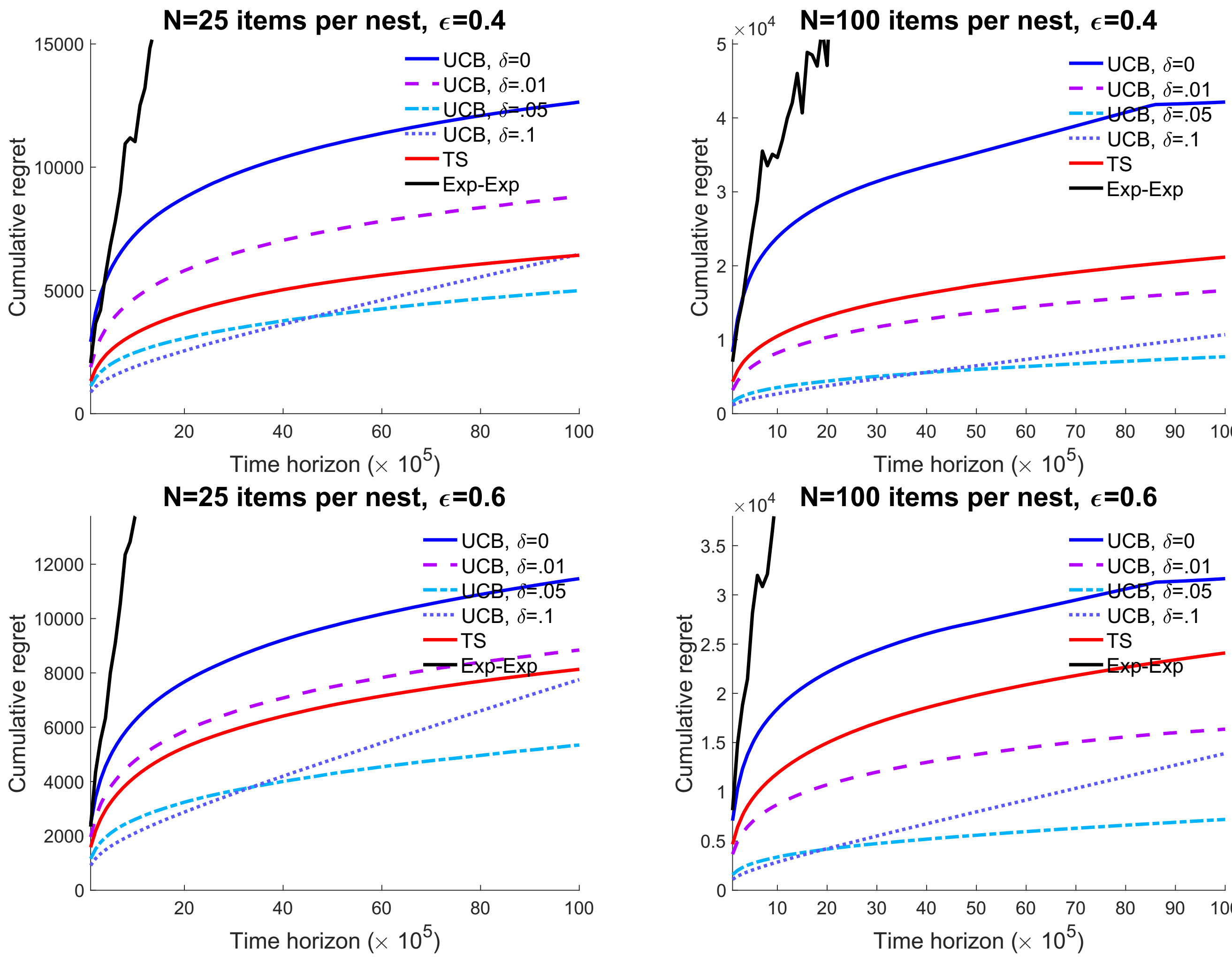


Figure 3 Cumulative Regret of Our Algorithm and Other Comparative Methods for Experiments Outlined in Section 5.1. ε is a Parameter Used in Generating Problem Instances, Which is Described in Further Details in the Main Text [Color figure can be viewed at wileyonlinelibrary.com]



algorithm with δ=0 (and sometimes our UCB algorithm with δ=0.01).

6. Conclusions

In this study, we consider the dynamic assortment planning problem under the nested logit models and we propose the UCB policy to achieve $\tilde{O}(\sqrt{MNT} + MN^2)$ accumulative regret. We also propose the discretization heuristic that shows the improved empirical performance.

There are several interesting future directions of the current work. The first technical problem is to investigate the dependency of N in the lower bound. Second, it is interesting to further extend the current two-level nested logit models to variants of nested models (e.g., constrained nested logit models (Gallego and Topaloglu 2014), d-level nested logit models (Li et al. 2015)). Third, the dynamic assortment planning is a relatively new topic in revenue management and the understanding of this problem is still limited. Therefore, most existing work (including this study) focuses on the stylized models where the assortment is the only decision variable. One potential future work is to incorporate other operational decisions and constraints, such as prices and inventory constraints.

Acknowledgment

We thank the Department Editor, the Senior Editor, and the two referees for their constructive suggestions that greatly improve the paper. Xi Chen is supported by the NSF Grant via IIS-1845444. Chao Shi thanks the National Natural Science Foundation of China (Grant 71501043). Yuan Zhou is supported by NSF CCF-2006526.

Note

[1]Recall that in Algorithm 1, $\boldsymbol{\theta}^{(t)}$ does not change within the same epoch $\mathcal{E}_\tau$. We write $\hat{\boldsymbol{\theta}}^{(\tau)}$ to highlight that $\hat{\boldsymbol{\theta}}^{(\tau)}$ is the maximizer of $\bar{R}'$ in the τ-th epoch (see Step 3 of Algorithm 1).

References

Agrawal, S., V. Avandhanula, V. Goyal, A. Zeevi. 2017. Thompson sampling for MNL-bandit. Proccedings of the Conference on Learning Theory (COLT).

Agrawal, S., V. Avadhanula, V. Goyal, A. Zeevi. 2019. MNL-bandit: A dynamic learning approach to assortment selection. *Oper. Res.* **67**(5): 1453–1485.

Bertsimas, D., V. V. Mišić. 2019Exact first-choice product line optimization. *Oper. Res.* **67**(3): 559–904.

Besbes, O., D. Saure. 2016. Product assortment and price competition under multinomial logit demand. *Prod. Oper. Manag.* **25** (1): 114–127.

Blanchet, J., G. Gallego, V. Goyal. 2016. A markov chain approximation to choice modeling. *Oper. Res.* **64**(4): 886–905.

Borch-Supan, A. 1990. On the compatibility of nested logit models with utility maximization. *J. Econom.* **43**(3): 373–388.

Bront, J. J. M., I. Méndez-Díaz, G. Vulcano. 2009. A column generation algorithm for choice-based network revenue management. *Oper. Res.* **57**(3): 769–784.

Caro, F., J. Gallien. 2007. Dynamic assortment with demand learning for seasonal consumer goods. *Management Sci.* **53**(2): 276–292.

Chen, X., Y. Wang. 2018. A note on tight lower bound for MNL-bandit assortment selection models. *Oper. Res. Lett.* **46**(5): 534–537.

Chen, X., Y. Wang, Y. Zhou. 2018. Dynamic assortment optimization with changing contextual information. *J Mach Learn Res.*

Chen, X., W. Ma, D. Simchi-Levi, L. Xin. 2019. Assortment Planning for Recommendations at Checkout Under Inventory Constraints. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_ #id=2853093 (accessed date October 20, 2020).

Cheung, W. C., D. Simchi-Levi. 2017. Thompson sampling for online personalized assortment optimization problems with multinomial logit choice models. Available at https://papers.ssrn.com>abstract_ #id=3075658 (accessed date October 20, 2020).

Chung, H., H. S. Ahn, S. Jasin. 2019. (Rescaled) multi-attempt approximation of choice model and its application to assortment optimization. *Prod. Ooper. Manag.* **28**(2): 341–353.

Csiszar, I., J. Körner. 2011. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, Cambridge.

Davis, J. M., G. Gallego, H. Topaloglu. 2014. Assortment optimization under variants of the nested logit model. *Oper. Res.* **62**(2): 250–273.

Désir, A., V. Goyal, S. Jagabathula, D. Segev. 2016. Assortment optimization under the Mallows model. *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS).*

Désir, A., V. Goyal, D. Segev, C. Ye. 2020. Capacity constrained assortment optimization under the markov chain-based choice model. *Management Sci.* **66**(2): 698–721.

Farias, V. F., S. Jagabathula, D. Shah. 2013. A nonparametric approach to modeling choice with limited data. *Management Sci.* **59**(2): 305–322.

Gallego, G., H. Topaloglu. 2014. Constrained assortment optimization for the nested logit model. *Management Sci.* **60**(10): 2583–2601.

Gallego, G., G. Iyengar, R. Phillips, A. Dubey. 2004. Managing flexible products on a network. Technical Report CORC TR-2004-01, Department of Industrial Engineering and Operations Research, Columbia University.

Golrezaei, N., H. Nazerzadeh, P. Rusmevichientong. 2014. Real-time optimization of personalized assortments. *Management Sci.* **60**(6): 1532–1551.

Kök, A. G., Y. Xu. 2011. Optimal and competitive assortments with endogenous pricing under hierarchical consumer choice models. *Management Sci.* **57**(9): 1546–1563.

Li, G., P. Rusmevichientong. 2014. A greedy algorithm for the two-level nested logit model. *Oper. Res. Lett.* **42**(5): 319–324.

Li, G., P. Rusmevichientong, H. Topaloglu. 2015. The d-level nested logit model: Assortment and price optimization problems. *Oper. Res.* **63**(2): 325–342.

Liu, Q., G. van Ryzin. 2008. On the choice-based linear programming model for network revenue management. *Manufact. Serv. Oper. Manag.* **10**(2): 288–310.

Mahajan, S., G. van Ryzin. 2001. Stocking retail assortments under dynamic consumer substitution. *Oper. Res.* **49**: 334–351.

McFadden, D. 1974. Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics* (Academic Press).

McFadden, D. 1980. Econometric models for probabilistic choice among products. *J. Bus.* **53**(3): 13–29.

Megiddo, N. 1978. Combinatorial optimization with rational objective functions. *Proceedings of the annual ACM symposium on Theory of computing (STOC).*

Méndez-Díaz, I., J. J. Miranda-Bront, G. Vulcano, P. Zabala. 2014. A branch-and-cut algorithm for the latentclass logit assortment problem. *Discrete Appl. Math.* **164**: 246–263.

Miao, S. T., X. L. Chao. 2018. Dynamic joint assortment and pricing optimization with demand learning. Technical report, University of Michigan, Ann Arbor.

Rusmevichientong, P., H. Topaloglu. 2012. Robust assortment optimization in revenue management under the multinomial logit choice model. *Oper. Res.* **60**(4): 865–882.

Rusmevichientong, P., Z. J. Shen, D. Shmoys. 2010. Dynamic assortment optimization with a multinomial logit choice model and capacity constraint. *Oper. Res.* **58**(6): 1666–1680.

Rusmevichientong, P., D. Shmoys, C. Tong, H. Topaloglu. 2014. Assortment optimization under the multinomial logit model with random choice parameters. *Prod. Oper. Manag.* **23**(11): 2023–2039.

van Ryzin, G., S. Mahajan. 1999. On the relationships between inventory costs and variety benefits in retail assortments. *Management Sci.* **45**(11): 1496–1509.

Saure, D., A. Zeevi. 2013. Optimal dynamic assortment planning with demand learning. *Manufact. Serv. Oper. Manag.* **15**(3): 387–404.

Talluri, K., G. van Ryzin. 2004. Revenue management under a general discrete choice model of consumer behavior. *Management Sci.* **50**(1): 15–33.

Train, K. 2009. *Discrete Choice Methods with Simulation*. Cambridge University Press, Cambridge, 2nd edn.

Tsybakov A. B. 2009. *Introduction to Nonparametric Estimation Springer Series in Statistics*. Springer, New York.

Wang, R. 2012. Capacitated assortment and price optimization under the multinomial logit choice model. *Oper. Res. Lett.* **40** (6): 492–497.

Wang, R. 2013. Assortment management under the generalized attraction model with a capacity constraint. *J. Rev. Pric. Manag.* **12**(3): 254–270.

Wang, Y., X. Chen, Y. Zhou. 2018. Near-optimal policies for dynamic multinomial logit assortment selection models.

Proceedings of Advances in Neural Information Processing Systems (NeurIPS).

Williams, H. C. W. L. 1977. On the formation of travel demand models and economic evaluation measures of user benefit. *Environ. Plan. A* **9**, 285–344.

Zhang, H., P. Rusmevichientong, H. Topaloglu. 2020. Assortment optimization under the paired combinatorial logit model. *Oper. Res.* **68**(3): 741–761.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Proofs of Statements