informs
https://pubsonline.informs.org/journal/opre

OPERATIONS RESEARCH

Vol. 00, No. 0, Xxxxx 0000, pp. 000–000

Submitted to *Operations Research*

ISSN 0030-364X, EISSN 1526-5463

Leveraging the Degree of Dynamic Substitution in Assortment and Inventory Planning

Jingwei Zhang
School of Data Science, The Chinese University of Hong Kong, Shenzhen (CUHK-Shenzhen), zhangjingwei@cuhk.edu.cn

Will Ma
Graduate School of Business and Data Science Institute, Columbia University, wm2428@gsb.columbia.edu

Huseyin Topaloglu
School of Operations Research and Information Engineering, Cornell Tech, topaloglu@orie.cornell.edu

Abstract. We study the joint assortment and inventory planning problem with stockout-based substitution. In this problem, we pick the number of units to stock for the products at the beginning of the selling horizon. Each arriving customer makes a choice among the set of products with remaining on-hand inventories. Our goal is to pick the stocking quantities to maximize the total expected revenue from the sales net of the stocking cost. We develop a rounding scheme that uses the solution to a fluid approximation to generate stocking quantities with performance guarantees that improve earlier results. Letting T be the number of time periods in the selling horizon and n be the number of products, when customers choose under a general choice model, we show that we can round the solution to the fluid approximation to obtain stocking quantities with an optimality gap of $O(\sqrt{nT})$, improving earlier optimality gaps by a logarithmic factor. More importantly, when customers choose under the multinomial logit model, by leveraging the degree of substitution, we show that our rounded fluid solution is within an optimality gap of $O(\log T \sqrt{T \log T})$. The optimality gap that we give under the multinomial logit model is the first one that does not depend on the number of products. Such an optimality gap has important practical implications. Earlier results cannot guarantee that the stocking quantities generated by the fluid approximation perform well when both the demand volume and the number of products are large, which is a regime becoming more relevant for online retail applications with large product variety. In contrast, we can guarantee that stocking quantities generated by our rounding scheme perform well when both the demand volume and the number of products are large.

Funding: This research was supported by [grant number, funding agency].

Key words: Assortment optimization, Multinomial logit model, Fluid approximation.

1. Introduction

One of the common challenges faced by the retailers involves deciding which products to offer and how much inventory to stock at the beginning of a selling season when the customers choose and substitute among the products that are available to them. There are multiple tradeoffs to balance in such a problem setting. Due to the substitution possibilities, when the inventory for a product runs out, its demand shifts to other products, so it is difficult to quantify the demand for each product a priori without knowing the stocking quantities. On the other hand, the stocking quantities should, in turn, depend on the demand for the products. Therefore, there is a two-way interaction between the demand for each product and stocking quantities. Moreover, even putting the inventory considerations aside, finding the right variety of products to offer to customers is a non-trivial problem. If the product variety is too low, then a large portion of the customers may not find what they are looking for and leave without a purchase. If the product variety is too high, then there may be too many low-margin products introduced into the the mix, which may end up cannibalizing the sales of high-margin products. These considerations point out that finding the right product variety and stocking quantities requires considering how the demand for each product is shaped by the inventories of all products and the substitution behavior.

Two types of substitution models have attracted attention. The first substitution model is static substitution, which is used when customers cannot observe the product availability, as in online retail. Under static substitution, customers choose within the product assortment without knowing the real-time inventory availabilities. If the customer chooses a product without on-hand inventory, then she may leave without a purchase or be served through an emergency procurement. Static substitution takes its name due to the fact that the set of products among which a customer chooses does not depend on real-time inventories. Such a substitution model simplifies the assortment and inventory planning problem significantly. Due to the goodwill cost, however, it may not be desirable to let a customer choose a product without on-hand inventory. Also, customers in brick-and-mortar retail do observe the real-time inventories on the shelf when making their choices. Thus, the second substitution model is stockout-based substitution, where customers choose among the products with on-hand inventories. Assortment and inventory planning under stockout-based substitution is more difficult because the real-time inventories influence the choice behavior.

To address the assortment and inventory planning problem under stockout-based substitution, one approach is to formulate a fluid approximation under the assumption that the choices of the customers take on their expected values and use the optimal solution to the fluid approximation to obtain stocking quantities. There are three papers of particular interest to us that use this approach. Honhon and Seshadri (2013) consider customers choosing under a general choice model. Letting T be the number of time periods in the selling horizon and n be the number of products, the authors show that implementing the optimal solution of the fluid approximation leads to an optimality gap of $O(n\sqrt{T})$. Mouchtaki et al. (2021) study the problem under the Markov chain choice model. They develop an algorithm combining a fluid approximation and a sample average approximation. Despite the optimality gap of their algorithm is $O(n+\sqrt{nT\log(nT)})$, they numerically show their algorithm benefits from the sample average approximation and outperforms a simply rounded fluid solution. Liang et al. (2022) focus on scenarios where the customers choose under the multinomial logit model. In addition, they consider a more general setting in which the initial inventory may not be zeros and inventories of some products cannot be changed. They also explore several important extensions. The authors first introduce a fluid relaxation and then develop algorithms which can efficiently solve the relaxation. They then show a rounded fluid solution can achieve an optimality gap of $O(\sqrt{nT\log(nT)})$ when the number of products is smaller than the number of time periods in the selling horizon.

Our Contributions: In this paper, we develop a new scheme for rounding the solution to the fluid approximation and give improved optimality gaps. Considering the case where the customers choose under a general choice model, we show that we can round the solution from a fluid approximation to get a solution with an optimality gap of $O(\sqrt{nT})$ comparing to the fluid approximation. This optimality gap is tight in both T and n in the sense that we give examples that attain this optimality gap. When compared with the optimality gaps established in earlier results, our optimality gap removes the rounding error $O(n)$ and also improves by a factor of $O(\sqrt{\log nT})$. We use this result as a warm-up to give much improved optimality gaps under the multinomial logit model, which is one of the most popular choice models both in practice and academic work. Under the multinomial logit model, we show that the solution from our rounding scheme has an optimality gap of $O(\log T\sqrt{T\log T})$, assuming that the preference weights of products are not infinitesimal. Non-infinitesimality corresponds to there being a sufficient "degree" of substitution between the products, and this allows for an optimality gap that is independent of the number of products n. The intuition for this result is that inventory planning should not get harder with more products as long as there is sufficient substitution between them, because overstocking of one product can compensate for understocking of another. We also consider various extensions in the Appendix to demonstrate our rounding scheme and analysis can be generalized to other choice models and more sophisticated settings.

To our knowledge, the optimality gap that we give under the multinomial logit model is the first one that is independent of the number of products. We also show the non-infinitesimality assumption to be necessary this result. Due to the fact that our optimality gap increases sublinearly in T, and also does not increase in n, our rounding scheme generates solutions with relative optimality gaps arbitrarily close to zero as both the demand volume and the number of products get large. Furthermore, we note that the optimality gap of $O(\sqrt{nT\log(nT)})$ established in earlier

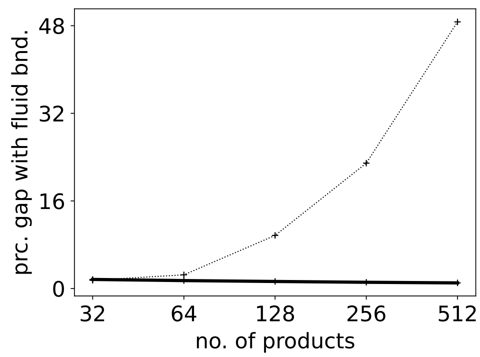

Figure 1 Optimality gaps provided by two socking quantities as a function of the number of products.

results is not an artifact of loose analysis. In Figure 1, as a function of the number of products in a problem instance, we plot the percent gap between the upper bound on the optimal total expected profit provided by the fluid approximation and the total expected profit obtained by two stocking quantities. The solid line plots the gap for the stocking quantities obtained by our rounding approach, whereas the dotted line plots the gap for the stocking quantities obtained by rounding down the fluid solution following Liang et al. (2022). Indeed, the gap for our stocking quantities remains stable, whereas the gap for the rounded-down stocking quantities increases in n. Thus, our algorithm performs better both in theory and in experiments for medium to large n.

Other Related Literature: The paper by van Ryzin and Mahajan (1999) studies joint assortment and inventory planning under the multinomial logit model with static substitution and products with identical margins, characterizing the structure of the optimal solution. Cachon et al. (2005) incorporate customer search costs. Topaloglu (2013) relaxes the assumption of identical product margins. Moving on to work under dynamic substitution, Mahajan and van Ryzin (2001) give a stochastic gradient algorithm to compute locally optimal stocking quantities. Gaur and Honhon (2006) use a locational choice model and obtain an upper bound. Hopp and Xu (2008) propose a static approximation for substitution behavior that is based on a fluid network model. Honhon et al. (2010) give a fluid approximation and solve it in $O(8^n)$ operations.

Lastly, we distinguish our work from other two streams of literature. The first stream is online assortment optimization, where the assortment of products offered to each customer can be adjusted by the retailer; see, for example, Golrezaei et al. (2014), Ma and Simchi-Levi (2020), and Rusmevichientong et al. (2020). In our problem, the set of products available for the customers is automatically implied by on-hand inventories. The second stream is assortment and inventory planning problem under stockout-based substitution with an upper bound on the total units stocked; see, for example, Goyal et al. (2016), Aouad et al. (2018), Aouad et al. (2019), and Aouad and Segev (2022). In this problem, stocking costs are ignored but there is an upper bound on the total units stocked. Exploiting the fact that the objective function does not involve a cost component, the focus is on developing multiplicative performance guarantees.

Outline: In Section 2, we formulate the problem. In Section 3, we discuss the fluid approximation that drives our stocking quantities and also our rounding scheme. In Section 4, we give our optimality gap under a general choice model. In Section 5, we give our optimality gap under the multinomial logit model. In Section 6, we illustrate our results numerically. In Section 7, we conclude.

2. Problem Formulation

We have n products indexed by $\mathcal{N} = \{1, \dots, n\}$. Selling a unit of product i yields a revenue of p_i. We procure each unit of product i at a cost of c_i. Throughout the paper, we set $\overline{p} = \max_{i \in \mathcal{N}} p_i$ and $\overline{c} = \max_{i \in \mathcal{N}} c_i$. Without loss of generality, we index the products such that $p_1 - c_1 \geq p_2 - c_2 \geq \dots \geq p_n - c_n \geq 0$. There are T time periods in the selling horizon indexed by $\mathcal{T} = \{1, \dots, T\}$. For notational brevity, we follow the convention that there is one customer arrival at each time period. At the beginning of the selling horizon, we decide how many units of each product to stock. Each customer arriving into the system makes a choice within the set of products with remaining inventories, either purchasing one unit of a product or deciding to leave without a purchase. In particular, if the set of products with remaining inventories at a certain time period is S, then the customer purchases product i with probability $\phi_i(S)$. Naturally, we have $\phi_i(S) = 0$ for all $i \notin S$. With probability $\phi_0(S) = 1 - \sum_{i \in \mathcal{N}} \phi_i(S)$, the customer leaves without a purchase. We assume that for any set $S' \subseteq S$, it holds that $\phi_i(S') \geq \phi_i(S)$ for any $i \in S'$, so that dropping products from the set of available products increases the purchase probability of all other available products. This property is called the substitutability property and it is satisfied by all choice models that are based on random utility maximization. Due to the substitutability property, the purchase probability of a product is non-decreasing over the time periods until it is sold out.

Our goal is to decide how many units of each product to stock at the beginning of the selling horizon to maximize the total expected profit. We use the vector $\boldsymbol{q} = (q_1, \dots, q_n) \in \mathbb{Z}_+^n$ to capture the inventories on-hand at the beginning of a generic time period, where q_i is the inventory on-hand for product i. If the inventories on-hand at the current time period are $\boldsymbol{q}$, then the set of available products is given by $A(\boldsymbol{q}) = \{i \in \mathcal{N} : q_i > 0\}$, in which case, the sale for product i at the current time period is a Bernoulli random variable with parameter $\phi_i(A(\boldsymbol{q}))$. Given that the inventories on-hand at the current time period are $\boldsymbol{q}$, we use the random vector $\boldsymbol{D}(\boldsymbol{q}) = (D_1(\boldsymbol{q}), \ldots, D_n(\boldsymbol{q}))$ to capture the sales for the products, where $D_i(\boldsymbol{q})$ is the sale for product i. Using $\boldsymbol{e}_i \in \mathbb{Z}_+^n$ to denote the i-th unit vector, note that we have $\boldsymbol{D}(\boldsymbol{q}) = \boldsymbol{e}_i$ with probability $\phi_i(A(\boldsymbol{q}))$. We use the random variable $X_i(\boldsymbol{q}, \tau)$ to capture the total sales for product i over τ time periods given that the inventories on-hand at the beginning of these τ time periods are $\boldsymbol{q}$. Therefore, the random variable $X_i(\boldsymbol{q}, \tau)$ is recursively defined as $X_i(\boldsymbol{q}, \tau) = D_i(\boldsymbol{q}) + X_i(\boldsymbol{q} - \boldsymbol{D}(\boldsymbol{q}), \tau - 1)$ with the boundary condition that $X_i(\boldsymbol{q}, 0) = 0$. In this case, if the stocking quantities at the beginning of the selling horizon are $\boldsymbol{q}$, then the total expected profit that we obtain is given by

$$\Pi(\boldsymbol{q}) \;=\; \sum_{i \in \mathcal{N}} p_i \, \mathbb{E}\{X_i(\boldsymbol{q}, T)\} - \sum_{i \in \mathcal{N}} c_i \, q_i. \tag{1}$$

The optimal total expected profit is $\mathsf{opt} = \max_{\boldsymbol{q} \in \mathbb{Z}_+^n} \Pi(\boldsymbol{q})$. Putting aside solving the last optimization problem, to our knowledge, computing $\Pi(\boldsymbol{q})$ at fixed stocking quantities requires solving a dynamic program with a high-dimensional state variable keeping track of the remaining inventories of each product. We use a fluid approximation to get a tractable benchmark.

3. Fluid Approximation and Rounding Procedure

Instead of comparing the total expected profit from the approximate stocking quantities that we compute with opt, we will compare it with an upper bound on opt. Consider the problem

$$V^{\mathsf{fluid}} \;=\; T \max_{S \subseteq \mathcal{N}} \left\{ \sum_{i \in \mathcal{N}} \phi_i(S)\,(p_i - c_i) \right\}. \tag{2}$$

The maximization problem above finds a subset of product to offer to maximize the expected profit extracted from the customer arriving at each time period. Using S^{fluid} to denote an optimal solution to the maximization problem above, we define the stocking quantities $\boldsymbol{q}^{\mathsf{fluid}} = (q_1^{\mathsf{fluid}}, \ldots, q_n^{\mathsf{fluid}})$ as $q_i^{\mathsf{fluid}} = T\,\phi_i(S^{\mathsf{fluid}})$ for all $i \in \mathcal{N}$. In this case, if the sale for product i at a time period were to take the deterministic and fractional value of $\phi_i(S^{\mathsf{fluid}})$, then stocking q_i^{fluid} units of product i would imply that the inventories all products are depleted simultaneously at the end of the selling horizon. Meanwhile, we would be extracting the maximum expected profit from the customer arriving at each time period. Thus, we can show that V^{fluid} is an upper bound on the optimal total expected profit. That is, we have $V^{\mathsf{fluid}} \geq \mathsf{opt}$; see Proposition 2.1 in Mouchtaki et al. (2021). This upper bound on the optimal total expected profit is computed under the assumption that the demand for each product is deterministic and fractional, so we refer to this upper bound on the optimal total expected profit as the fluid upper bound. While it is difficult to compute the optimal total expected profit efficiently, we can often compute the fluid upper bound efficiently. In particular, problem (2) is of combinatorial nature, but we can solve this problem efficiently when the choices of the customers are governed by a variety of choice models, including the multinomial logit, generalized attraction, nested logit, multi-level nested logit and Markov chain choice model; see Talluri and van Ryzin (2004), Davis et al. (2014), Gallego et al. (2015), Li et al. (2015) and Blanchet et al. (2016). For any approximate stocking quantity $\boldsymbol{q}$, the optimality gap of this approximate stocking quantity is given by $\mathsf{opt} - \Pi(\boldsymbol{q})$. Noting that $\mathsf{opt} - \Pi(\boldsymbol{q}) \leq V^{\mathsf{fluid}} - \Pi(\boldsymbol{q})$, to upper bound the optimality gap of the approximate stocking quantity, it will be enough to upper bound the $V^{\mathsf{fluid}} - \Pi(\boldsymbol{q})$.

Problem (2) not only serves as an upper bound to the original problem (1), but it also suggests $\boldsymbol{q}^{\mathsf{fluid}}$ could be a near-optimal solution. However, $\boldsymbol{q}^{\mathsf{fluid}}$ might involve fractional values and thus cannot be directly implemented. Intuitively, we want to "follow" the optimal fluid solution as close as possible in order to achieve performance close to the fluid upper bound. There are different ways to make the fluid solution feasible, for example, Liang et al. (2022) consider the rounded solution $\lfloor \boldsymbol{q}^{\mathsf{fluid}} \rfloor$ where $\lfloor \cdot \rfloor$ to denote the round down operator. Here, instead of rounding down the stocking quantities from the fluid approximation for all products, we develop a rounding scheme that can round up or down the stocking quantity for each product. Specifically, we compute the total units of fractional inventories and redistribute those inventories to products with higher margins while keeping stocking quantities for each product integral and "close" to the fluid solution. We formalize our rounding procedure below.

Rounding Procedure. The rounding budget is $\delta = \lceil \sum_{i \in S^{\mathsf{fluid}}} (q_i^{\mathsf{fluid}} - \lfloor q_i^{\mathsf{fluid}} \rfloor) \rceil$. The rounded stocking quantities $\boldsymbol{q}^{\mathsf{round}} = (q_1^{\mathsf{round}}, \ldots, q_n^{\mathsf{round}})$ are given by $q_i^{\mathsf{round}} = \lfloor q_i^{\mathsf{fluid}} \rfloor + 1$ for product i among the first δ products in the set S^{fluid} and $q_i^{\mathsf{round}} = \lfloor q_i^{\mathsf{fluid}} \rfloor$ for other products in the set S^{fluid}.

Our rounding procedure can be interpreted as follows. Every product $i \in S^{\mathsf{fluid}}$ is promised a baseline inventory of $\lfloor q_i^{\mathsf{fluid}} \rfloor$, which is the solution proposed by Liang et al. (2022). However, this results in total inventory $\sum_{i \in S^{\mathsf{fluid}}} \lfloor q_i^{\mathsf{fluid}} \rfloor$, whereas the fluid solution wanted total inventory $\sum_{i \in S^{\mathsf{fluid}}} q_i^{\mathsf{fluid}}$. We apportion this difference in inventory, captured by our δ, back among the products (taking the ceiling when defining δ). Although there are generally many ways to do this apportionment (see Cembrano et al. 2022 for a recent reference), in our setting there is a natural best apportionment: select the δ products with the highest profit margins to receive an additional unit of inventory. That is exactly our rounding procedure.

In the next lemma, we give properties of the rounded stocking quantities. Intuitively speaking, this lemma states that the rounded stocking quantities shift the stocking quantities in the solution to the fluid upper bound from products with lower margins to those with higher margins, while keeping the total stocking quantities nearly unchanged. Let S^{round} denote the set of products offered in the rounded solution, i.e., $S^{\text{round}} = \{i \in \mathcal{N} : q_i^{\text{round}} > 0\}$.

LEMMA 3.1. *The rounded stocking quantities* $\boldsymbol{q}^{\text{round}}$ *and the stocking quantities from the fluid upper bound* $\boldsymbol{q}^{\text{fluid}}$ *satisfy the inequalities*

$$0 \le \sum_{i \in S^{\text{fluid}}} (q_i^{\text{round}} - q_i^{\text{fluid}}) \le 1, \quad \sum_{i \in S^{\text{fluid}}} (p_i - c_i)\,(q_i^{\text{fluid}} - q_i^{\text{round}}) \le 0, \quad \left|S^{\text{round}}\right| \le T + 1.$$

We give the proof of the lemma in Appendix A. To see the benefit from using the stocking quantities $\boldsymbol{q}^{\text{round}}$ instead of $\lfloor \boldsymbol{q}^{\text{fluid}} \rfloor$, setting $X_0(\boldsymbol{q}, \tau) = \tau - \sum_{i \in A(\boldsymbol{q})} X_i(\boldsymbol{q}, \tau)$, we have

$$\begin{aligned}
V^{\text{fluid}} - \Pi(\boldsymbol{q}^{\text{round}}) &= \sum_{i \in S^{\text{fluid}}} (p_i - c_i)\, q_i^{\text{fluid}} - \left[\sum_{i \in S^{\text{fluid}}} p_i\, \mathbb{E}\{X_i(\boldsymbol{q}^{\text{round}}, T)\} - \sum_{i \in S^{\text{fluid}}} c_i\, q_i^{\text{round}} \right] \\
&= \sum_{i \in S^{\text{fluid}}} (p_i - c_i) \left(q_i^{\text{fluid}} - q_i^{\text{round}} \right) + \sum_{i \in S^{\text{fluid}}} p_i \left(q_i^{\text{round}} - \mathbb{E}\{X_i(\boldsymbol{q}^{\text{round}}, T)\} \right) \\
&\overset{(a)}{\le} \sum_{i \in S^{\text{fluid}}} (p_i - c_i) \left(q_i^{\text{fluid}} - q_i^{\text{round}} \right) + \overline{p} \left[\sum_{i \in S^{\text{fluid}}} q_i^{\text{round}} + \mathbb{E}\left\{X_0(\boldsymbol{q}^{\text{round}}, T)\right\} - T \right] \\
&\overset{(b)}{=} \sum_{i \in S^{\text{fluid}}} (p_i - c_i) \left(q_i^{\text{fluid}} - q_i^{\text{round}} \right) + \overline{p} \left\{ \mathbb{E}\{X_0(\boldsymbol{q}^{\text{round}}, T)\} - T\,\phi_0(S^{\text{fluid}}) \right\} + \overline{p} \sum_{i \in S^{\text{fluid}}} (q_i^{\text{round}} - q_i^{\text{fluid}}) \\
&\overset{(c)}{\le} \overline{p} \left\{ \mathbb{E}\{X_0(\boldsymbol{q}^{\text{round}}, T)\} - T\,\phi_0(S^{\text{fluid}}) \right\} + \overline{p},
\end{aligned}$$

where (a) uses the definition of $X_0(\boldsymbol{q}, T)$, (b) holds because we have $q_i^{\text{fluid}} = T\,\phi_i(S^{\text{fluid}})$ and $\sum_{i \in S^{\text{fluid}}} \phi_i(S^{\text{fluid}}) + \phi_0(S^{\text{fluid}}) = 1$ and (c) uses the first two inequalities in Lemma 3.1.

By the chain of inequalities above, the difference $\sum_{i \in S^{\text{fluid}}} (q_i^{\text{round}} - q_i^{\text{fluid}})$ is $O(1)$, so this difference contributes the term $\overline{p}$ to the optimality gap $V^{\text{fluid}} - \Pi(\boldsymbol{q}^{\text{round}})$, which is independent of the number of products. In contrast, the difference $\sum_{i \in S^{\text{fluid}}} (q_i^{\text{fluid}} - \lfloor q_i^{\text{fluid}} \rfloor)$ is $O(n)$. Thus, using the stocking quantities $\boldsymbol{q}^{\text{round}}$ instead of $\lfloor \boldsymbol{q}^{\text{fluid}} \rfloor$ eliminates the contribution of the rounding error to the optimality gap. In practice, when the number of products n is relatively large, we expect our solution lead to a better performance, e.g., as shown in Figure 1. Furthermore, by the chain of inequalities above, to bound the optimality gap $V^{\text{fluid}} - \Pi(\boldsymbol{q}^{\text{round}})$, it is enough to upper bound the difference $\mathbb{E}\{X_0(\boldsymbol{q}^{\text{round}}, T)\} - T\,\phi_0(S^{\text{fluid}})$, where $\mathbb{E}\{X_0(\boldsymbol{q}^{\text{round}}, T)\}$ is the expected number of customers without a purchase when the stocking quantities are $\boldsymbol{q}^{\text{round}}$, and $T\phi_0(S^{\text{fluid}})$ is the deterministic and fractional number of customers without a purchase under the fluid approximation.

In the next section, we consider the case where the customers choose under a general choice model and give a performance bound for the solutions from the fluid upper bound.

4. Performance Guarantee under a General Choice Model

We consider the case where the choices of the customers are governed by a general choice model and give a performance guarantee for stocking quantities from the fluid upper bound. We obtain the fluid upper bound by using stocking quantities that are depleted simultaneously for all products at the end of the selling horizon. Thus, the set of products with on-hand inventories do not change throughout the selling horizon. Motivated by this observation, we focus on a problem that is formulated under the assumption that the customers always make a choice among the products that are stocked at the beginning of the selling horizon, but if the product they choose does not have on-hand inventory anymore, then they leave without a purchase. If the stocking quantities are $\boldsymbol{q}$, then the set of stocked products is given by $A(\boldsymbol{q})$. A customer chooses product i among this set of products with probability $\phi_i(A(\boldsymbol{q}))$. We use the random vector $\boldsymbol{C}_t(\boldsymbol{q}) = (C_{1t}(\boldsymbol{q}), \ldots, C_{nt}(\boldsymbol{q}))$ to capture the choice of the customer arriving at time period t among the set of products $A(\boldsymbol{q})$. Thus, we have $C_t(\boldsymbol{q}) = \boldsymbol{e}_i$ with probability $\phi_i(A(\boldsymbol{q}))$. Under the assumption that the customers always make a choice among the products that are stocked at the beginning of the selling horizon, if we stock the quantities $\boldsymbol{q}$, then the total expected profit is

$$\Pi^{\mathsf{static}}(\boldsymbol{q}) \;=\; \sum_{i \in \mathcal{N}} p_i \, \mathbb{E}\left\{ \min\left\{ q_i, \sum_{t \in \mathcal{T}} C_{it}(\boldsymbol{q}) \right\} \right\} - \sum_{i \in \mathcal{N}} c_i \, q_i. \tag{3}$$

In the expression in (3), the set of products among which the customers choose does not change during the course of the selling horizon, so we refer (3) as the static approximation. We can argue that the purchase probability of each product at each time period in (3) is no larger than its counterpart in (1). In particular, the customers choose among the set of products stocked at the beginning of the selling horizon in (3), whereas the customers choose among the set of products with on-hand inventories in (1). By the assumption that $\phi_i(S) \leq \phi_i(S')$ for all $S' \subseteq S$ and $i \in S'$, the purchase probability of each product at each time period in (3) is no larger than its counterpart in (1). In this case, the total expected profit in (3) is no larger than the total expected profit in (1), so $\Pi^{\mathsf{static}}(\boldsymbol{q}) \leq \Pi(\boldsymbol{q})$ for all $\boldsymbol{q} \in \mathbb{Z}_+^n$. Recalling that S^{fluid} is an optimal solution to problem (2) and $q_i^{\mathsf{fluid}} = T\,\phi_i(S^{\mathsf{fluid}})$, let $\boldsymbol{q}^{\mathsf{round}}$ be the rounded solution of $\boldsymbol{q}^{\mathsf{fluid}}$. In the next theorem, we give a performance guarantee for the stocking quantities $\boldsymbol{q}^{\mathsf{round}}$.

THEOREM 4.1. *If $\overline{p}$ and $\overline{c}$ are independent of n and T, then there exists an absolute constant C_1 such that*

$$V^{\mathsf{fluid}} - \Pi(\boldsymbol{q}^{\mathsf{round}}) \le V^{\mathsf{fluid}} - \Pi^{\mathsf{static}}(\boldsymbol{q}^{\mathsf{round}}) \ \le \ C_1 \sqrt{nT}.$$

We give the proof of the theorem in Appendix B. Using the fact that $V^{\mathsf{fluid}} \ge \mathsf{opt}$ and $\Pi^{\mathsf{static}}(\boldsymbol{q}) \le \Pi(\boldsymbol{q})$ for all $\boldsymbol{q} \in \mathbb{Z}_+^n$, Theorem 4.1 implies that $\mathsf{opt} - \Pi(\boldsymbol{q}^{\mathsf{round}}) \le C_1 \sqrt{nT}$, so the optimality gap of the stocking quantities $\boldsymbol{q}^{\mathsf{round}}$ is upper bounded by $C_1 \sqrt{nT}$. We briefly outline the proof of the theorem. In the fluid upper bound, we have the total revenue $\sum_{i \in \mathcal{N}} (p_i - c_i) q_i^{\mathsf{fluid}}$. By Lemma 3.1, the fluid revenue is upper bound by $\sum_{i \in \mathcal{N}} (p_i - c_i) q_i^{\mathsf{round}}$. In the static lower bound, corresponding to the stocking quantities $\boldsymbol{q}^{\mathsf{round}}$, we incur the total procurement cost $\sum_{i \in \mathcal{N}} c_i\, q_i^{\mathsf{round}}$. So we are left to analyze the difference in the total expected profit. In the static lower bound, corresponding to the stocking quantities $\boldsymbol{q}^{\mathsf{round}}$, the total expected profit from product i is $p_i\, \mathbb{E}\{\min\{q_i^{\mathsf{round}}, \sum_{t \in \mathcal{T}} C_{it}(\boldsymbol{q}^{\mathsf{round}})\}\}$. Thus, the difference between the total expected profit from product $i \in S^{\mathsf{round}}$ is given by

$$p_i \left(q_i^{\mathsf{round}} - \mathbb{E}\left\{ \min\left\{ q_i^{\mathsf{round}}, \sum_{t \in \mathcal{T}} C_{it}(\boldsymbol{q}^{\mathsf{round}}) \right\} \right\} \right) = p_i \mathbb{E}\left\{ \left[q_i^{\mathsf{round}} - \sum_{t \in \mathcal{T}} C_{it}(\boldsymbol{q}^{\mathsf{round}}) \right]^+ \right\}.$$

Using the fact that $q_i^{\mathsf{round}} \le q_i^{\mathsf{fluid}} + 1$, we incur a total rounding error $\bar{p} \left| S^{\mathsf{round}} \right|$ and are left to analyze the difference

$$\mathbb{E}\left\{ \left[q_i^{\mathsf{fluid}} - \sum_{t \in \mathcal{T}} C_{it}(\boldsymbol{q}^{\mathsf{round}}) \right]^+ \right\}.$$

Since $S^{\mathsf{round}} \subseteq S^{\mathsf{fluid}}$, the purchase probability of each product given S^{round} at each time period is no less than its counterpart given S^{fluid}, therefore,

$$\mathbb{E}\left\{ \left[q_i^{\mathsf{fluid}} - \sum_{t \in \mathcal{T}} C_{it}(\boldsymbol{q}^{\mathsf{round}}) \right]^+ \right\} \le \mathbb{E}\left\{ \left[q_i^{\mathsf{fluid}} - \sum_{t \in \mathcal{T}} C_{it}(\boldsymbol{q}^{\mathsf{fluid}}) \right]^+ \right\}.$$

Using the Jensen inequality, we then upper bound the difference by $\sqrt{\phi_i(S^{\mathsf{fluid}})\, T}$, in which case, using this upper bound for all products with Cauchy-Schwarz inequality yields the result.

Theorem 4.1 gives an optimality gap of $O(\sqrt{nT})$. We shortly give an example where this optimality gap is tight, followed by another example where we improve this optimality gap when the products are highly substitutable. Theorem 4.1 makes use of the static upper bound, which ignores the substitution behavior based on the on-hand product inventories. Mouchtaki et al. (2021) argue that ignoring substitution behavior based on the on-hand product inventories can lead to large optimality gaps and use sample average approximation to get an optimality gap of $O(n+\sqrt{nT}\,\log(nT))$. Theorem 4.1 improves the last optimality gap by removing the linear dependence on n and also a factor of $O(\sqrt{\log(nT)})$.

Independent Demand Model:

We give an example where the optimality gap in Theorem 4.1 is tight comparing to the fluid approximation. Using $\mathbf{1}(\cdot)$ to denote the indicator function, the choice probabilities are given by $\phi_i(S) = \mathbf{1}(i \in S)\frac{1}{1+n}$. Thus, if a product is available, its choice probability is $\frac{1}{1+n}$, irrespective of what other products are available. The unit revenues and procurement cost are respectively $p_i = 2$ and $c_i = 1$ for all $i \in \mathcal{N}$. We assume that $\frac{T}{1+n}$ is an integer. The optimal solution to problem (2) is $S^{\textsf{fluid}} = \mathcal{N}$, so we have $q_i^{\textsf{fluid}} = \frac{T}{1+n}$ for all $i \in \mathcal{N}$ and $V^{\textsf{fluid}} = T\,\frac{n}{1+n}$. We compute the static approximation to the total expected profit at the stocking quantities $\boldsymbol{q}^{\textsf{round}} = \boldsymbol{q}^{\textsf{fluid}} \in \mathbb{Z}_+^n$, which is given by $\Pi^{\textsf{static}}(\boldsymbol{q}^{\textsf{round}})$. Since $q_i^{\textsf{round}} = q_i^{\textsf{fluid}} > 0$ for all $i \in \mathcal{N}$, $A(\boldsymbol{q}^{\textsf{round}}) = \mathcal{N}$, so $\phi_i(A(\boldsymbol{q}^{\textsf{round}})) = \frac{1}{1+n}$. Therefore, $C_{it}(\boldsymbol{q}^{\textsf{round}})$ is a Bernoulli random variable with parameter $\frac{1}{1+n}$, so $\sum_{t\in\mathcal{T}} C_{it}(\boldsymbol{q}^{\textsf{round}})$ is a binomial random variable with parameters $(T, \frac{1}{1+n})$. Using $\textsf{binomial}(k,p)$ to denote a binomial random variable with parameters (k,p), Lemma B.1 in Appendix B shows that $\mathbb{E}\{[k\,p - \textsf{binomial}(k,p)]^+\} \geq \frac{1}{\sqrt{2\pi}}\sqrt{k\,p\,(1-p)} - O(1)$. Thus, we obtain

$$\begin{aligned} V^{\textsf{fluid}} - \Pi^{\textsf{static}}(\boldsymbol{q}^{\textsf{round}}) &= \frac{Tn}{1+n} - n\left[2\,\mathbb{E}\left\{\min\left\{\frac{T}{1+n}, \textsf{binomial}\left(T, \frac{1}{1+n}\right)\right\}\right\} - \frac{T}{1+n}\right] \\ &= 2\frac{Tn}{1+n} - 2n\,\mathbb{E}\left\{\min\left\{\frac{T}{1+n}, \textsf{binomial}\left(T, \frac{1}{1+n}\right)\right\}\right\} \\ &= 2n\,\mathbb{E}\left\{\left[\frac{T}{1+n} - \textsf{binomial}\left(T, \frac{1}{1+n}\right)\right]^+\right\} \geq 2n\left(\frac{1}{\sqrt{2\pi}}\sqrt{T\frac{n}{(1+n)^2}} - O(1)\right) = \Omega(\sqrt{nT} - n), \end{aligned}$$

so if the number of time periods in the selling horizon is larger than the number of products, then we have $V^{\textsf{fluid}} - \Pi^{\textsf{static}}(\boldsymbol{q}^{\textsf{round}}) = \Omega(\sqrt{nT})$. Moreover, in this example, we have

$$\Pi(\boldsymbol{q}) = \Pi^{\textsf{static}}(\boldsymbol{q}) = 2\sum_{i\in\mathcal{N}} \mathbb{E}\left\{\min\left\{q_i, \textsf{binomial}\left(T, \frac{1}{1+n}\right)\right\}\right\} - \sum_{i\in\mathcal{N}} q_i,$$

which is a newsvendor problem for each product. Therefore, let F denote the cumulative probability function of the binomial distribution, the optimal solution to problem (1) is $q_i^* = F^{-1}\left(\frac{1}{2}\right) = \frac{T}{n+1}$ for all $i \in \mathcal{N}$. That is, $\boldsymbol{q}^{\textsf{round}}$ is optimal to problem (1) and thus the optimality gap of the fluid approximation, in general, cannot be improved beyond $O(\sqrt{nT})$.

Fully Substitutable Demand Model:

We give an example where we can improve the optimality gap in Theorem 4.1 under a choice model where the products are fully substitutable. This example suggests that the optimality gap in Theorem 4.1 is, in general, tight, but by focusing on choice models with a special structure, we can improve the optimality gap. The choice probabilities are given by $\phi_i(S) = \frac{n}{1+n}\,\mathbf{1}(i \in S)\,\frac{1}{|S|}$. We interpret these choice probabilities as follows. A customer is interested in making a purchase with probability $\frac{n}{1+n}$. If interested in making a purchase, then the customer chooses any of the available products with equal probability. We have $p_i = 2$ and $c_i = 1$ for all $i \in \mathcal{N}$ for the unit revenues and procurement costs. We continue assuming that $\frac{T}{1+n}$ is an integer. Considering problem (2), any non-empty subset of products is an optimal solution, so we use $S^{\mathsf{fluid}} = \{1\}$. In this case, noting that $\phi_1(\{1\}) = \frac{n}{1+n}$, we get $V^{\mathsf{fluid}} = T\frac{n}{1+n}$ and $q_1^{\mathsf{fluid}} = \frac{Tn}{1+n}$, $q_i^{\mathsf{fluid}} = 0$ for all $i = 2, \ldots, n$. Furthermore, $\sum_{t\in\mathcal{T}} C_{1t}(\boldsymbol{q}^{\mathsf{round}})$ is a binomial random variable with parameters $(T, \frac{n}{1+n})$. Lemma 1 in Gallego and Moon (1993) shows that $\mathbb{E}\{[k\,p - \mathsf{binomial}(k,p)]^+\} \le \frac{1}{2}\sqrt{k\,p\,(1-p)}$. Thus, we obtain

$$
\begin{aligned}
V^{\mathsf{fluid}} - \Pi^{\mathsf{static}}(\boldsymbol{q}^{\mathsf{round}}) &= \frac{Tn}{1+n} - \left[2\,\mathbb{E}\left\{\min\left\{\frac{Tn}{1+n}, \mathsf{binomial}\Big(T, \frac{n}{1+n}\Big)\right\}\right\} - \frac{Tn}{1+n}\right] \\
&= 2\frac{Tn}{1+n} - 2\,\mathbb{E}\left\{\min\left\{\frac{Tn}{1+n}, \mathsf{binomial}\Big(T, \frac{n}{1+n}\Big)\right\}\right\} \\
&= 2\,\mathbb{E}\left\{\left[\frac{Tn}{1+n} - \mathsf{binomial}\Big(T, \frac{n}{1+n}\Big)\right]^+\right\} \le \sqrt{T\frac{n}{(1+n)^2}} = O(\sqrt{T/n}),
\end{aligned}
$$

so the optimality gap of the stocking quantities $\boldsymbol{q}^{\mathsf{round}}$ is $O(\sqrt{T/n})$. Intuitively speaking, for n large, each customer makes a purchase with probability one, making the problem much easier.

The preceding two examples (independent and fully substitutable demand models) demonstrate the extremes of how much customers might substitute between products. The independent demand example suggests that the optimality gap compared to fluid, in general, cannot be improved beyond $O(\sqrt{nT})$, i.e. our result in this section is tight. The second example suggests, however, that with a sufficient degree of substitution, $\Theta(\sqrt{nT})$ is not the right answer and the optimality gap could potentially be improved for specific choice models. That is exactly what we do in subsequent sections.

5. Performance Guarantee under the Multinomial Logit Model

We focus on the case where the choices of the customers are governed by the multinomial logit model and show that we can obtain stocking quantities from the fluid upper bound with optimality gap independent of the number of products. By the discussion right after Theorem 4.1, the fluid upper bound assumes that the demand for product i at each time period takes the deterministic and fractional value $\phi_i(S^{\mathsf{fluid}})$, whereas the static approximation assumes that the demand for product i at each time period is a Bernoulli random variable with parameter $\phi_i(S^{\mathsf{round}})$. Under the multinomial logit model, we tighten our analysis to deal with the optimality gap more effectively. We use the specific substitution behavior under the multinomial logit model to analyze the stochastic evolution of the demand over the selling horizon.

Multinomial logit model is arguably one of the most popular choice models, both in practical applications and academic research. Under this choice model, a customer associates a preference weight of w_i with product i and a preference weight of w_0 with the no-purchase option. If the set of available products is S, then a customer chooses product i with probability $\phi_i(S) = \mathbf{1}(i \in S)\frac{w_i}{w_0+\sum_{j\in S} w_j}$. Consider computing the fluid upper bound in (2) when customers choose according to the multinomial logit model. It is known that the optimal solution to problem (2) is margin-ordered in the sense that S^{fluid} offers a certain number of products with the largest margins. Because we assume products are indexed in the margin order, the optimal solution to problem (2) is of the form $S^{\mathsf{fluid}} = \{1, 2, \dots, j\}$ for some $j \in \mathcal{N}$. Thus, we can solve problem (2) by checking the objective value provided by each solution of the form $\{1, 2, \dots, j\}$ and there are at most n solutions of this form. Once we obtain S^{fluid} efficiently in this fashion, we continue computing the stocking quantities from the fluid approximation as $q_i^{\mathsf{fluid}} = T\,\phi_i(S^{\mathsf{fluid}}) = T\,\mathbf{1}(i \in S^{\mathsf{fluid}})\,\frac{w_i}{w_0+\sum_{j\in S^{\mathsf{fluid}}} w_j}$. Applying the rounding procedure to $\boldsymbol{q}^{\mathsf{fluid}}$, we obtain integer stocking quantities $\boldsymbol{q}^{\mathsf{round}}$.

Following the discussion right after Lemma 3.1, it is sufficient to upper bound the difference $\mathbb{E}\{X_0(\boldsymbol{q}^{\mathsf{round}}, T)\} - T\,\phi_0(S^{\mathsf{fluid}})$ in order to upper bound the optimality gap of the stocking quantities $\boldsymbol{q}^{\mathsf{round}}$. We will upper bound the difference by an expression that is independent of the number of products. In this case, the optimality gap will be independent of the number of products as well.

In the next theorem, we give a performance guarantee for the stocking quantities $\boldsymbol{q}^{\mathsf{round}}$ under the multinomial logit model. Throughout the rest of the paper, we set $\underline{w} = \min_{i\in\mathcal{N}} w_i$.

THEOREM 5.1. *Considering customers choosing under the multinomial logit model, if $\overline{p}$ and $w_0/\underline{w}$ are independent of n and T, then there exist absolute constants C_2 and C_3 such that*

$$V^{\mathsf{fluid}} - \Pi(\boldsymbol{q}^{\mathsf{round}}) \;\le\; C_2 \log T\,\sqrt{T \log T} + C_3.$$

It is necessary to assume w_0/w is independent of n and T, because otherwise, we can replicate the independent demand model by having all w_i's approach 0, for which we know the optimality gap against fluid is $\Omega(\sqrt{nT})$. We give the proof of the theorem in Appendix C. Noting that $V^{\mathsf{fluid}} \ge \mathsf{opt}$, Theorem 5.1 implies that $\mathsf{opt} - \Pi(\boldsymbol{q}^{\mathsf{round}}) \le C_2 \log T \sqrt{T \log T} + C_3$, so optimality gap of the stocking quantities $\boldsymbol{q}^{\mathsf{round}}$ is $O(\log T \sqrt{T \log T})$, which is independent of the number of product. We briefly outline the proof of the theorem. Under the multinomial logit model, we bound the expected number of customers without a purchase as $\mathbb{E}\{X_0(\boldsymbol{q}, T)\} \le 1 + w_0 \max_{i \in \mathcal{N}} \left\{\frac{q_i}{w_i}\right\} \log \sum_{i \in \mathcal{N}} q_i + [T - \sum_{i \in \mathcal{N}} q_i]^+$ for any stocking quantities $\boldsymbol{q} \in \mathbb{Z}_+$. This inequality is the key driver of the proof of the theorem. If the number of customer arrivals is small in the sense that $T \le K\,|S^{\mathsf{fluid}}| \log T$ for some constant K independent of n and T, then using this inequality along with the fact that $\sum_{i \in \mathcal{N}} q_i^{\mathsf{fluid}} = T \sum_{i \in \mathcal{N}} \phi_i(S^{\mathsf{fluid}}) \le T$ in our key inequality allows us to bound $\mathbb{E}\{X_0(\boldsymbol{q}, T)\}$.

If, on the other hand, $T \ge K\,|S^{\mathsf{fluid}}| \log T$, then we proceed as follows. Using Chernoff bound, we characterize a time period $\widetilde{T}$ close to T such that no product is sold out by time period $\widetilde{T}$ with high probability. Given that none of the products is sold out by time period $\widetilde{T}$, the total number of customers without a purchase by time period $\widetilde{T}$ is a binomial random variable, which allows us to show that the conditional total expected number of customers without a purchase by time period $\widetilde{T}$ is no larger than $T\,\phi_0(S^{\mathsf{fluid}})$. Lastly, given that none of the products is sold out by time period $\widetilde{T}$, we show that the on-hand inventory of each product at time period $\widetilde{T}$ is $O(\sqrt{T})$. In this case, using our key inequality over the remaining $T - \widetilde{T}$ time periods, we show that the expected number of customers without a purchase over these time periods is $O(\log T \sqrt{T \log T})$.

6. Numerical Study

In this section, we first provide a numerical study on the symmetric multinomial logit model to illustrate the dependence on the number of products of performances of different solutions. We then conduct numerical experiments for problems under different settings. We implement all experiments on a personal computer with 2 GHz Intel i5 processor and 12 GB of RAM.

6.1. Symmetric MNL Model

We give a numerical study to test the practical performance of the solutions $\boldsymbol{q}^{\mathsf{round}}$ and $\lfloor \boldsymbol{q}^{\mathsf{fluid}} \rfloor$ when the choices of the customers are governed by the multinomial logit model. Recall that the former stocking quantities have an optimality gap of $O(\log T \sqrt{T \log T})$, whereas the latter stocking quantities have an optimality gap of $\Omega(n + \sqrt{T})$. In all of our test problems, the unit revenue and procurements costs of all products are $p_i = 2$ and $c_i = 1$ for all $i \in \mathcal{N}$. The preference weights of all products are $w_i = 1$ for all $i \in \mathcal{N}$. Finally, the preference weight of the no-purchase option is $w_0 = 1$. With these parameters, the optimal assortment is to offer all products, which allows us to investigate the relationship between the optimality gap and the number of products. In addition, although our focus here is to compare the performance of the $\boldsymbol{q}^{\mathsf{round}}$ and $\lfloor \boldsymbol{q}^{\mathsf{fluid}} \rfloor$ heuristics, for the symmetric MNL model, we can prove that it is always optimal to split inventories "evenly" across all products. This means that we can compute the optimal solution, because given any total stocking quantity Q, we know how to optimally split it, and hence we just need to do a line search over Q. We also include this in the numerical results as a benchmark, deferring the proof of this even splitting property for symmetric MNL to Appendix F.

We vary the number of products n over $\{2^k : k = 1, \ldots, 9\}$ and the number of time periods T over $\{10 \times 2^k : k = 1, \ldots, 9\}$. For each test problem, we compute the fluid upper bound V^{fluid} and stocking quantities $\lfloor \boldsymbol{q}^{\mathsf{fluid}} \rfloor$ and $\boldsymbol{q}^{\mathsf{round}}$. We estimate the total expected profit obtained by these two sets of stocking quantities using Monte Carlo simulation with 10,000 sample paths.

In Table 1, we fix the number of time periods in the selling horizon at $T = 1000$ and vary the number of products. The first row gives the number of products. The second row gives the value of the fluid upper bound. The block of next three rows focus on the stocking quantities $\boldsymbol{q}^{\mathsf{round}}$. The first row gives the total expected profit obtained by using the stocking quantities $\boldsymbol{q}^{\mathsf{round}}$, the second row gives the absolute gap between the fluid upper bound and the total expected profit from the stocking quantities $\boldsymbol{q}^{\mathsf{round}}$ and the third row gives the same gap in relative percentage terms. The block of last three rows focus on the stocking quantities $\lfloor \boldsymbol{q}^{\mathsf{fluid}} \rfloor$ and provide the same performance measures. Consistent with Theorem 5.1, the absolute gap between the fluid upper bound and the total expected profit from the stocking quantities $\boldsymbol{q}^{\mathsf{round}}$ remains stable as the number of products increases. In contrast, consistent with Theorem 4.1, the absolute gap between the fluid upper bound and the total expected profit from the stocking quantities $\lfloor \boldsymbol{q}^{\mathsf{fluid}} \rfloor$ increases as the number of products increases. For $n = 256$ or 512, the stocking quantities $\lfloor \boldsymbol{q}^{\mathsf{fluid}} \rfloor$ perform rather poorly, with optimality gaps exceeding 20%, whereas the optimality gaps of the stocking quantities $\boldsymbol{q}^{\mathsf{round}}$ is close to 1%. Other work, such as Liang et al. (2022), works with the stocking quantities $\lfloor \boldsymbol{q}^{\mathsf{fluid}} \rfloor$, but this solution is relevant when the number of time periods is large relative to number of products. When the number of products is large, the stocking quantities $\lfloor \boldsymbol{q}^{\mathsf{fluid}} \rfloor$ can be poor.

In practice, as the number of products grows, the no-purchase probability of a customer might not be close to zero. Therefore, the parameter settings for Table 1 could be unrealistic especially when n is relatively large. In Table 2, we provide numerical results where we keep the same parameters as in Table 1, except we set $w_0 = n/9$ so that the no-purchase probability is always at least 0.1. Because w_0 grows with the number of products n, the technical assumption in Theorem 5.1 is violated and

$T = 1000$

n	2	4	8	16	32	64	128	256	512
V^{fluid}	666.7	800.0	888.9	941.2	969.7	984.6	992.2	996.1	998.1
opt	650.7	782.1	870.8	924.8	955.4	972.4	981.8	987.1	990.1
$\Pi(\boldsymbol{q}^{\text{round}})$	650.3	781.3	870.1	923.0	953.6	970.3	979.4	984.7	987.8
$V^{\text{fluid}} - \Pi(\boldsymbol{q}^{\text{round}})$	16.43	18.69	18.75	18.20	16.12	14.32	12.81	11.40	10.23
$100 \times \frac{1}{V^{\text{fluid}}}(V^{\text{fluid}} - \Pi(\boldsymbol{q}^{\text{round}}))$	2.46	2.34	2.11	1.93	1.66	1.45	1.29	1.14	1.03
$\Pi(\lfloor \boldsymbol{q}^{\text{fluid}} \rfloor)$	650.3	781.6	870.1	923.6	955.1	960.0	896.0	768.0	512.0
$V^{\text{fluid}} - \Pi(\lfloor \boldsymbol{q}^{\text{fluid}} \rfloor)$	16.41	18.42	18.75	17.56	14.60	24.65	96.25	228.1	486.1
$100 \times \frac{1}{V^{\text{fluid}}}(V^{\text{fluid}} - \Pi(\lfloor \boldsymbol{q}^{\text{fluid}} \rfloor))$	2.46	2.30	2.11	1.87	1.51	2.50	9.70	22.9	48.7

Table 1 Optimality gaps for the stocking quantities $\boldsymbol{q}^{\text{round}}$ and $\lfloor \boldsymbol{q}^{\text{fluid}} \rfloor$ for varying n.

the optimality gap is no longer guaranteed to be bounded by a constant. Nonetheless, the optimality gap for the stocking quantities $\boldsymbol{q}^{\text{round}}$ grows at the rate $O(\sqrt{n})$, consistent with Theorem 4.1. Similar to Table 1, For $n = 256$ or 512, the stocking quantities $\boldsymbol{q}^{\text{round}}$ outperforms $\lfloor \boldsymbol{q}^{\text{fluid}} \rfloor$ due to the benefit of our rounding scheme. In conclusion, our solution $\boldsymbol{q}^{\text{round}}$ provides a comparable performance to $\lfloor \boldsymbol{q}^{\text{fluid}} \rfloor$ in both experiments and can perform much better when n is large.

$T = 1000$

n	2	4	8	16	32	64	128	256	512
V^{fluid}	900	900	900	900	900	900	900	900	900
opt	890.5	887.5	883.2	876.6	866.7	852.0	829.8	794.1	738.7
$\Pi(\boldsymbol{q}^{\text{round}})$	890.5	887.1	882.0	874.8	864.2	848.2	824.5	786.7	731.3
$V^{\text{fluid}} - \Pi(\boldsymbol{q}^{\text{round}})$	9.52	12.93	17.97	25.17	35.81	51.80	75.48	113.3	168.7
$100 \times \frac{1}{V^{\text{fluid}}}(V^{\text{fluid}} - \Pi(\boldsymbol{q}^{\text{round}}))$	1.06	1.44	2.00	2.80	3.98	5.76	8.39	12.59	18.74
$\Pi(\lfloor \boldsymbol{q}^{\text{fluid}} \rfloor)$	890.4	887.4	883.2	875.9	865.5	849.6	826.1	760.8	511.8
$V^{\text{fluid}} - \Pi(\lfloor \boldsymbol{q}^{\text{fluid}} \rfloor)$	9.65	12.87	16.85	24.08	34.52	50.40	73.94	139.2	388.2
$100 \times \frac{1}{V^{\text{fluid}}}(V^{\text{fluid}} - \Pi(\lfloor \boldsymbol{q}^{\text{fluid}} \rfloor))$	1.07	1.43	1.87	2.68	3.84	5.60	8.22	15.47	43.14

Table 2 Optimality gaps for the stocking quantities $\boldsymbol{q}^{\text{round}}$ and $\lfloor \boldsymbol{q}^{\text{fluid}} \rfloor$ for varying n with a fixed no-purchase probability.

In Table 3, we fix the number of products at $n = 100$ and vary the number of time periods in the selling horizon, going back to the setting of Table 1 where the no-purchase weight does not grow with n. By Theorem 5.1, the optimality gap of the stocking quantities $\boldsymbol{q}^{\text{round}}$ is $O(\log T \sqrt{T \log T})$. Accordingly, the absolute gap between the fluid upper bound and the total expected profit obtained by using the stocking quantities $\boldsymbol{q}^{\text{round}}$ increases slightly with T, but noting that V^{fluid} increase linearly in T, the relative gap between the two quantities decrease. When $T = 20, 40, 80$ or 160 so that the number of products is large relative to the number of time periods, the stocking quantities $\lfloor \boldsymbol{q}^{\text{fluid}} \rfloor$ are not practically useful as they have optimality gap exceeding 30%. When $T = 2560$ or 5120, the optimality gaps of the stocking quantities $\lfloor \boldsymbol{q}^{\text{fluid}} \rfloor$ are just over 1%, but the optimality gaps of the stocking quantities $\boldsymbol{q}^{\text{round}}$ are below 1% for these numbers of time periods in the selling horizon. Overall, the superior theoretical performance guarantee that we can give for $\boldsymbol{q}^{\text{round}}$ over $\lfloor \boldsymbol{q}^{\text{fluid}} \rfloor$ carries over to the practical performance of these stocking quantities.

$n = 100$

$T/10$	2	4	8	16	32	64	128	256	512
V^{fluid}	19.80	39.60	79.21	158.4	316.8	633.7	1267	2535	5069
opt	16.10	35.31	74.55	152.6	309.6	624.4	1255	2518	5047
$\Pi(\boldsymbol{q}^{\text{round}})$	15.69	34.70	73.61	151.6	308.5	622.7	1252	2515	5042
$V^{\text{fluid}} - \Pi(\boldsymbol{q}^{\text{round}})$	4.11	4.91	5.60	6.85	8.31	10.98	14.83	19.66	27.12
$100 \times \frac{1}{V^{\text{fluid}}}(V^{\text{fluid}} - \Pi(\boldsymbol{q}^{\text{round}}))$	20.76	12.39	7.06	4.33	2.62	1.73	1.17	0.78	0.53
$\Pi(\lfloor \boldsymbol{q}^{\text{fluid}} \rfloor)$	0	0	0	100	300	600	1200	2499	5000
$V^{\text{fluid}} - \Pi(\lfloor \boldsymbol{q}^{\text{fluid}} \rfloor)$	19.80	39.60	79.21	58.42	16.84	33.66	67.33	34.66	69.31
$100 \times \frac{1}{V^{\text{fluid}}}(V^{\text{fluid}} - \Pi(\lfloor \boldsymbol{q}^{\text{fluid}} \rfloor))$	100	100	100	36.87	5.31	5.31	5.31	1.37	1.37

Table 3 Optimality gaps for the stocking quantities $\boldsymbol{q}^{\text{round}}$ and $\lfloor \boldsymbol{q}^{\text{fluid}} \rfloor$ for varying T.

6.2. Additional Numerical Experiments

In this section, we provide more comprehensive numerical experiments. Throughout this section, we assume that the number of customers follows a Poisson distribution. Additionally, we vary the number of products in $\{20, 40, 60, 80\}$ and vary the expected number of customers in $\{100, 500, 1000\}$.

In our first experiment, we consider the MNL model and a single capacity constraint, using a setup very similar to Liang et al. (2022). The unit revenue p_i of all products is generated using a log-normal distribution with mean 0 and standard deviation 0.3. The ratio $(p_i - c_i)/c_i$ of all products are generated from a uniform distribution on ranges $[0.2, 0.6]$. Finally, the preference weight w_i of all products are uniformly distributed among $[0.1, 5.1]$ and the preference weight of the no-purchase option is set to $w_0 = 1$. The total capacity is set in the following way: we first compute the total purchase probability of the optimal assortment from problem (3), and then we set $K = \lceil 0.8 \times \text{total purchase probability} \times T \rceil$. Table 4 presents the numerical results for the capacitated problem. As the expected number of customers increases, the optimality gaps for both solutions decrease, consistent with our theoretical finding. Interestingly, the optimality gap does not appear to increase with the number of products. This can be attributed to the fact that the assortment size offered by the fluid solution remains relatively stable and small (often less than 6), even as the number of products grows larger to 80. Therefore, the rounding error is not the main contributor to the optimality gap.

Nonetheless, our experiments suggest that rounding still performs comparably well to flooring; i.e., the optimality gap of our solution $\boldsymbol{q}^{\text{round}}$ is almost as small as that of $\lfloor \boldsymbol{q}^{\text{fluid}} \rfloor$ in most cases of Table 4. The low values of $(p_i - c_i)/c_i$ from the setup of Liang et al. (2022) encourage understocking, in which case it makes sense to take the floor $\lfloor \boldsymbol{q}^{\text{fluid}} \rfloor$. We conduct a further set of experiments for higher-margin products, where we change $(p_i - c_i)/c_i$ to be drawn uniformly from [1.2,1.5]

		$n=20$	$n=40$	$n=60$	$n=80$
$T=100$	$100\times\frac{1}{V^{\text{fluid}}}(V^{\text{fluid}}-\Pi(\boldsymbol{q}^{\text{round}}))$	9.354	11.41	14.06	8.597
	$100\times\frac{1}{V^{\text{fluid}}}(V^{\text{fluid}}-\Pi(\lfloor\boldsymbol{q}^{\text{fluid}}\rfloor))$	**9.344**	**11.14**	**13.15**	**8.298**
$T=500$	$100\times\frac{1}{V^{\text{fluid}}}(V^{\text{fluid}}-\Pi(\boldsymbol{q}^{\text{round}}))$	**2.079**	**3.553**	6.030	3.638
	$100\times\frac{1}{V^{\text{fluid}}}(V^{\text{fluid}}-\Pi(\lfloor\boldsymbol{q}^{\text{fluid}}\rfloor))$	2.111	3.645	**5.856**	**3.489**
$T=1000$	$100\times\frac{1}{V^{\text{fluid}}}(V^{\text{fluid}}-\Pi(\boldsymbol{q}^{\text{round}}))$	**0.806**	1.940	**4.298**	2.567
	$100\times\frac{1}{V^{\text{fluid}}}(V^{\text{fluid}}-\Pi(\lfloor\boldsymbol{q}^{\text{fluid}}\rfloor))$	0.809	**1.907**	4.305	**2.462**

Table 4 Optimality gaps of the capacitated problem for the stocking quantities $\boldsymbol{q}^{\text{round}}$ and $\lfloor\boldsymbol{q}^{\text{fluid}}\rfloor$ for varying (T,n). The better algorithmic performance (smaller optimality gap) in each cell is **bolded**.

(instead of [0.2,0.6]). The results are shown in Table 5, and our procedure which rounds $\boldsymbol{q}^{\text{fluid}}$ while considering the ordering in the optimal assortment now performs better in more cells.

		$n=20$	$n=40$	$n=60$	$n=80$
$T=100$	$100\times\frac{1}{V^{\text{fluid}}}(V^{\text{fluid}}-\Pi(\boldsymbol{q}^{\text{round}}))$	**6.632**	6.679	**7.724**	**7.885**
	$100\times\frac{1}{V^{\text{fluid}}}(V^{\text{fluid}}-\Pi(\lfloor\boldsymbol{q}^{\text{fluid}}\rfloor))$	6.889	**6.689**	8.027	8.155
$T=500$	$100\times\frac{1}{V^{\text{fluid}}}(V^{\text{fluid}}-\Pi(\boldsymbol{q}^{\text{round}}))$	**2.853**	**2.127**	**3.345**	**3.533**
	$100\times\frac{1}{V^{\text{fluid}}}(V^{\text{fluid}}-\Pi(\lfloor\boldsymbol{q}^{\text{fluid}}\rfloor))$	2.873	2.218	3.492	3.732
$T=1000$	$100\times\frac{1}{V^{\text{fluid}}}(V^{\text{fluid}}-\Pi(\boldsymbol{q}^{\text{round}}))$	**2.029**	**1.125**	**2.436**	2.543
	$100\times\frac{1}{V^{\text{fluid}}}(V^{\text{fluid}}-\Pi(\lfloor\boldsymbol{q}^{\text{fluid}}\rfloor))$	2.034	1.159	2.471	**2.524**

Table 5 Optimality gaps of the capacitated problem for the stocking quantities $\boldsymbol{q}^{\text{round}}$ and $\lfloor\boldsymbol{q}^{\text{fluid}}\rfloor$ for varying (T,n). The better algorithmic performance (smaller optimality gap) in each cell is **bolded**.

7. Conclusions

We analyzed fluid relaxations for the assortment and inventory planning problem under stockout-based substitution. Under a general choice model, we tightened the optimality gap of a solution from the fluid approximation by a factor of $O(\sqrt{\log(nT)})$ when compared with the earlier literature. More interestingly, under the multinomial logit model, we showed that we can obtain a solution from the fluid approximation with an optimality gap of $O(\log T\sqrt{T\log T})$, which is independent of the number of products. The result is significant for both theoretical and practical reasons. In particular, our solutions perform well under large demand volume even when the number of products is large, where the solutions in earlier results perform poorly when the number of products is large compared to the demand volume. It would be interesting to obtain such tighter performance guarantees under more sophisticated settings. For example, for problems with initial inventories, whether an optimality gap independent of the number of products can be obtained remains open.

We are also curious about whether the optimal inventory division involves even splitting for "symmetric" choice models beyond MNL. In general, one can define a random-utility choice model to be symmetric if conditional on the number m of products preferred over the no-purchase option, these products in ranked order is equally likely to be any of the $n!/(n-m)!$ ordered subsets of $[n]$ of size m. Even in the special case of symmetric choice models where the consideration set has a deterministic size of $n-1$, we are unsure how to prove that even splitting is optimal, even though it appears to be intuitively true. In particular, assuming the total stocking quantity Q to be at most the number of customers T, every customer is guaranteed to make a purchase until there is a lone product among the n remaining (after which each customer will make a purchase w.p. $(n-1)/n$). Letting U denote the number of products sold before this lone product emerges, the expected number of products sold conditional on U can be expressed as

$$U+\mathbb{E}\left[\min\left\{\mathsf{binomial}\left(T-U,\frac{n-1}{n}\right),Q-U\right\}\right],$$

which is clearly increasing in U. Therefore, maximizing revenue is equivalent to maximizing the expectation of U, which we conjecture is achieved when each of the n products is initially stocked to quantity either $\lfloor Q/n\rfloor$ or $\lceil Q/n\rceil$. Some progress has been made on this problem recently by Gutiérrez and Subercaseaux (2024).

Acknowledgments

The authors thank three anonymous referees, an anonymous Associate Editor, and the Area Editor (Gustavo Vulcano) for insightful comments and suggestions that helped to improve the paper. The authors thank Omar Mouchtaki for sharing code related to NC-SAA.

References

Aouad A, Levi R, Segev D (2018) Greedy-like algorithms for dynamic assortment planning under multinomial logit preferences. *Operations Research* 66(5):1321–1345.

Aouad A, Levi R, Segev D (2019) Approximation algorithms for dynamic assortment optimization models. *Mathematics of Operations Research* 44(2):487–511.

Aouad A, Segev D (2022) The stability of mnl-based demand under dynamic customer substitution and its algorithmic implications. *Operations Research* .

Aven T (1985) Upper (lower) bounds on the mean of the maximum (minimum) of a number of random variables. *Journal of Applied Probability* 22(3):723–728.

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

Cachon G, Terwiesch C, Xu Y (2005) Retail assortment planning in the presence of consumer search. *Manufacturing & Service Operations Management* 7(4):330–346.

Cembrano J, Correa J, Verdugo V (2022) Multidimensional political apportionment. *Proceedings of the National Academy of Sciences* 119(15):e2109305119.

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

Gallego G, Moon IK (1993) The distribution-free newsboy problem - Review and extensions. *Journal of the Operational Research Society* 44(8):825–834.

Gallego G, Ratliff R, Shebalov S (2015) A general attraction model and sales-based linear programming formulation for network revenue management under customer choice. *Operations Research* 63(1):212–232.

Gaur V, Honhon D (2006) Assortment planning and inventory decisions under a locational choice model. *Management Science* 52(10):1528–1543.

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

Goyal V, Levi R, Segev D (2016) Near-optimal algorithms for the assortment planning problem under dynamic substitution and stochastic demand. *Operations Research* 64(1):219–235.

Gutiérrez F, Subercaseaux B (2024) Assortment optimization for conference goodies with indifferent attendees. *arXiv preprint arXiv:2403.03330* .

Honhon D, Gaur V, Seshadri S (2010) Assortment planning and inventory decisions under stockout-based substitution. *Operations Research* 58(5):1364–1379.

Honhon D, Seshadri S (2013) Fixed vs. random proportions demand models for the assortment planning problem under stockout-based substitution. *Manufacturing & Service Operations Management* 15(3):378–386.

Hopp WJ, Xu X (2008) A static approximation for dynamic demand substitution with applications in a competitive market. *Operations Research* 56(3):630–645.

Li G, Rusmevichientong P, Topaloglu H (2015) The d-level nested logit model: Assortment and price optimization problems. *Operations Research* 62(2):325–342.

Liang AJ, Jasin S, Uichanco J (2022) Assortment and inventory planning under dynamic substitution with MNL model: An LP approach and an asymptotically optimal policy. Technical report, University of Michigan, Ann Arbor, Michigan.

Luce RD (2012) *Individual choice behavior: A theoretical analysis* (Courier Corporation).

Ma W, Simchi-Levi D (2020) Algorithms for online matching, assortment, and pricing with tight weight-dependent competitive ratios. *Operations Research* 68(6):1787–1803.

Mahajan S, van Ryzin G (2001) Stocking retail assortments under dynamic consumer substitution. *Operations Research* 49(3):334–351.

Mouchtaki O, El Housni O, Gallego G, Goyal V, Humair S, Kim S, Sadighian A, Wu J (2021) Joint assortment and inventory planning under the markov chain choice model. Technical report, Columbia Business School, New York, NY.

Nagaev SV (1965) Some limit theorems for large deviations. *Theory of Probability & Its Applications* 10(2):214–235.

Rusmevichientong P, Sumida M, Topaloglu H (2020) Dynamic assortment optimization for reusable products with random usage durations. *Management Science* 66(7):2820–2844.

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

Topaloglu H (2013) Joint stocking and product offer decisions under the multinomial logit model. *Production and Operations Management* 22(5):1182–1199.

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

Appendix A: Results in Section 3

Proof of Lemma 3.1:

We first note that $\delta = \lceil \sum_{i\in S^{\text{fluid}}} (q_i^{\text{fluid}} - \lfloor q_i^{\text{fluid}} \rfloor) \rceil \le |S^{\text{fluid}}|$, thus,

$$\sum_{i\in S^{\text{fluid}}} (q_i^{\text{round}} - q_i^{\text{fluid}}) = \sum_{i\in S^{\text{fluid}}} (\lfloor q_i^{\text{fluid}} \rfloor - q_i^{\text{fluid}}) + \delta = \lceil \sum_{i\in S^{\text{fluid}}} (q_i^{\text{fluid}} - \lfloor q_i^{\text{fluid}} \rfloor) \rceil - \sum_{i\in S^{\text{fluid}}} (q_i^{\text{fluid}} - \lfloor q_i^{\text{fluid}} \rfloor) \in [0,1].$$

Next, let $\sigma(i)$ denote the i-th product in the set S^{fluid} following the index order, we have

$$\begin{aligned}
\sum_{i\in S^{\text{fluid}}} (p_i - c_i)(q_i^{\text{fluid}} - q_i^{\text{round}}) &= \sum_{i\in S^{\text{fluid}}} (p_i - c_i) q_i^{\text{fluid}} - \sum_{i=1}^{\delta} (p_{\sigma(i)} - c_{\sigma(i)})(\lfloor q_{\sigma(i)}^{\text{fluid}} \rfloor + 1) - \sum_{i=\delta+1}^{|S^{\text{fluid}}|} (p_{\sigma(i)} - c_{\sigma(i)}) \lfloor q_{\sigma(i)}^{\text{fluid}} \rfloor \\
&\le \sum_{i\in S^{\text{fluid}}} (p_i - c_i) q_i^{\text{fluid}} - \sum_{i=1}^{\delta} (p_{\sigma(i)} - c_{\sigma(i)}) q_{\sigma(i)}^{\text{fluid}} - (p_{\sigma(\delta)} - c_{\sigma(\delta)}) \left(\delta + \sum_{i=1}^{\delta} (\lfloor q_{\sigma(i)}^{\text{fluid}} \rfloor - q_{\sigma(i)}^{\text{fluid}}) \right) \\
&\quad - \sum_{i=\delta+1}^{|S^{\text{fluid}}|} (p_{\sigma(i)} - c_{\sigma(i)}) \lfloor q_{\sigma(i)}^{\text{fluid}} \rfloor \\
&= \sum_{i=\delta+1}^{|S^{\text{fluid}}|} (p_{\sigma(i)} - c_{\sigma(i)}) (q_{\sigma(i)}^{\text{fluid}} - \lfloor q_{\sigma(i)}^{\text{fluid}} \rfloor) - (p_{\sigma(\delta)} - c_{\sigma(\delta)}) \left(\delta - \sum_{i=1}^{\delta} (q_{\sigma(i)}^{\text{fluid}} - \lfloor q_{\sigma(i)}^{\text{fluid}} \rfloor) \right) \\
&\le (p_{\sigma(\delta)} - c_{\sigma(\delta)}) \left(\sum_{i\in S^{\text{fluid}}} (q_i^{\text{fluid}} - \lfloor q_i^{\text{fluid}} \rfloor) - \delta \right) \le 0.
\end{aligned}$$

Finally, we have

$$\left| S^{\text{round}} \right| = \left| \{ i : q_i^{\text{round}} > 0 \} \right| \le \sum_{i\in S^{\text{fluid}}} q_i^{\text{round}} \le \sum_{i\in S^{\text{fluid}}} q_i^{\text{fluid}} + 1 \le T + 1.$$

■

Appendix B: Results in Section 4

Proof of Theorem 4.1: Recall $\boldsymbol{C}_t(\boldsymbol{q}^{\text{round}})$ denotes the choice of the customer arriving at time period t among the set of products $A(\boldsymbol{q}^{\text{round}})$ and a customer chooses product i among this set of products with probability $\phi_i(S^{\text{round}})$.

We first present the formal proof for $\Pi^{\text{static}}(\boldsymbol{q}) \le \Pi(\boldsymbol{q})$ for any inventory $\boldsymbol{q}$. It holds that

$$\Pi(\boldsymbol{q}) - \Pi^{\text{static}}(\boldsymbol{q}) = \sum_{i\in\mathcal{N}} p_i \left(\mathbb{E}\left\{ \sum_{t\in\mathcal{T}} C_{it}(\boldsymbol{q}_t) \right\} - \mathbb{E}\left\{ \min\left\{ q_i, \sum_{t\in\mathcal{T}} C_{it}(\boldsymbol{q}) \right\} \right\} \right),$$

where $\boldsymbol{q}_t$ denotes the inventory level at time period t, which is a random variable. By definition, it holds that $\boldsymbol{q}_t \le \boldsymbol{q}$ for any time t, where the inequality holds component-wisely. Note that both

$C_{it}(\boldsymbol{q}_t)$ and $C_{it}(\boldsymbol{q})$ are Bernoulli random variables and $C_{it}(\boldsymbol{q}_t)$ has a larger success probability if $q_{it} > 0$, because of the substitution assumption and $A(\boldsymbol{q}_t) \subseteq A(\boldsymbol{q})$, that is, $C_{it}(\boldsymbol{q}_t) \geq C_{it}(\boldsymbol{q})$ if $q_{it} > 0$, where $\geq_{\text{st}}$ represents the first order stochastic dominance. Note that if $q_{it} = 0$, then product i is sold out, so we achieve the maximum possible revenue from product i, which must be no worse than the revenue obtained in the static approximation (i.e., $\sum_{t\in\mathcal{T}} C_{it}(\boldsymbol{q}_t) = q_i \geq \min\{q_i, \sum_{t\in\mathcal{T}} C_{it}(\boldsymbol{q})\}$). More formally, for fixed product i with $q_i > 0$, we consider the following stochastic process: in each period t, if $q_{it} > 0$, a seed ξ_t is generated uniformly over the interval $[0,1]$, and another seed ξ'_t is generated uniformly over the interval $[\phi_i(A(\boldsymbol{q})), 1]$. Let $C_{it}(\boldsymbol{q}) = \mathbf{1}(\xi_t \leq \phi_i(A(\boldsymbol{q})))$ and

$$C_{it}(\boldsymbol{q}_t) = \mathbf{1}(\xi_t \leq \phi_i(A(\boldsymbol{q}))) + \mathbf{1}(\xi_t > \phi_i(A(\boldsymbol{q}))) \cdot \mathbf{1}(\xi'_t \leq \phi_i(A(\boldsymbol{q}_t)) - \phi_t(A(\boldsymbol{q}))).$$

In this case, it holds that $\mathbb{P}(C_{it}(\boldsymbol{q}) = 1) = \phi_i(A(\boldsymbol{q}))$ and $\mathbb{P}(C_{it}(\boldsymbol{q}_t) = 1) = \phi_i(A(\boldsymbol{q}_t))$. Under this construction, we have $C_{it}(\boldsymbol{q}_t) \geq C_{it}(\boldsymbol{q})$ if $q_{it} > 0$ and thus $q_{it} \leq q_{it}^{\text{static}}$. This process is repeated until $q_{it} = 0$ at some time point t. At that point, we let $C_{it}(\boldsymbol{q}_t) = 0$ and continue to generate the seed ξ_t as described above, selling product i in the static approximation whenever inventory remains. Note that in this scenario, the total number of units of product i sold in the original problem is q_i and the maximum possible amount of product i sold in the static approximation is at most q_i. Therefore, we can conclude that under this construction, it holds that $\sum_{t\in\mathcal{T}} C_{it}(\boldsymbol{q}_t) \geq \min\{q_i, \sum_{t\in\mathcal{T}} C_{it}(\boldsymbol{q})\}$ for every sample path. Therefore, we can always couple the original problem and the static approximation in a way such that for any sample path, the number of product i purchased in the original problem is larger, which implies that $\Pi(\boldsymbol{q}) \geq \Pi^{\text{static}}(\boldsymbol{q})$.

We now provide the optimality gap by focusing on analyze the static approximation (3), i.e., the set of products among which the customers choose does not change over time. By equation (2) and (3), we have

$$\begin{aligned}
V^{\text{fluid}} - \Pi^{\text{static}}(\boldsymbol{q}^{\text{round}}) &= \sum_{i\in\mathcal{N}} (p_i - c_i) q_i^{\text{fluid}} - \sum_{i\in\mathcal{N}} p_i \mathbb{E}\left\{\min\left\{q_i^{\text{round}}, \sum_{t\in\mathcal{T}} C_{it}(\boldsymbol{q}^{\text{round}})\right\}\right\} + \sum_{i\in\mathcal{N}} c_i q_i^{\text{round}} \\
&= \sum_{i\in\mathcal{N}} (p_i - c_i)\left(q_i^{\text{fluid}} - q_i^{\text{round}}\right) + \sum_{i\in\mathcal{N}} p_i \mathbb{E}\left\{q_i^{\text{round}} - \min\left\{q_i^{\text{round}}, \sum_{t\in\mathcal{T}} C_{it}(\boldsymbol{q}^{\text{round}})\right\}\right\} \\
&\leq \sum_{i\in S^{\text{round}}} p_i \mathbb{E}\left\{q_i^{\text{round}} - \min\left\{q_i^{\text{round}}, \sum_{t\in\mathcal{T}} C_{it}(\boldsymbol{q}^{\text{round}})\right\}\right\} = \sum_{i\in S^{\text{round}}} p_i \mathbb{E}\left\{\left[q_i^{\text{round}} - \sum_{t\in\mathcal{T}} C_{it}(\boldsymbol{q}^{\text{round}})\right]^+\right\},
\end{aligned}$$

where the inequality follows from Lemma 3.1. Note that $S^{\text{round}} \subseteq S^{\text{fluid}}$, therefore, we have $C_{it}(\boldsymbol{q}^{\text{round}}) \geq_{\text{st}} C_{it}(\boldsymbol{q}^{\text{fluid}})$. It then implies that

$$V^{\text{fluid}} - \Pi^{\text{static}}(\boldsymbol{q}^{\text{round}}) \leq \sum_{i\in S^{\text{round}}} p_i \mathbb{E}\left\{\left[q_i^{\text{round}} - \sum_{t\in\mathcal{T}} C_{it}(\boldsymbol{q}^{\text{round}})\right]^+\right\}$$

$$
\begin{aligned}
&\leq \sum_{i\in S^{\mathsf{round}}} p_i \mathbb{E}\left\{\left[q_i^{\mathsf{round}} - \sum_{t\in\mathcal{T}} C_{it}(\boldsymbol{q}^{\mathsf{fluid}})\right]^+\right\} \leq \sum_{i\in S^{\mathsf{round}}} p_i \mathbb{E}\left\{\left[q_i^{\mathsf{fluid}} - \sum_{t\in\mathcal{T}} C_{it}(\boldsymbol{q}^{\mathsf{fluid}})\right]^+\right\} + \bar{p}\left|S^{\mathsf{round}}\right| \\
&\leq \sum_{i\in S^{\mathsf{round}}} p_i \mathbb{E}\left\{\left|q_i^{\mathsf{fluid}} - \sum_{t\in\mathcal{T}} C_{it}(\boldsymbol{q}^{\mathsf{fluid}})\right|\right\} + \bar{p}\left|S^{\mathsf{round}}\right| \\
&\overset{(a)}{\leq} \sum_{i\in S^{\mathsf{round}}} p_i \sqrt{\mathbb{E}\left\{\left(q_i^{\mathsf{fluid}} - \sum_{t\in\mathcal{T}} C_{it}(\boldsymbol{q}^{\mathsf{fluid}})\right)^2\right\}} + \bar{p}\left|S^{\mathsf{round}}\right| \\
&\overset{(b)}{=} \sum_{i\in S^{\mathsf{round}}} p_i \sqrt{\mathbb{V}\mathrm{ar}\left\{\sum_{t\in\mathcal{T}} C_{it}(\boldsymbol{q}^{\mathsf{fluid}})\right\}} + \bar{p}\left|S^{\mathsf{round}}\right| \\
&= \sum_{i\in S^{\mathsf{round}}} p_i \sqrt{T\phi_i(S^{\mathsf{fluid}})(1-\phi_i(S^{\mathsf{fluid}}))} + \bar{p}\left|S^{\mathsf{round}}\right| \leq \sum_{i\in S^{\mathsf{round}}} p_i \sqrt{T\phi_i(S^{\mathsf{fluid}})} + \bar{p}\left|S^{\mathsf{round}}\right| \\
&\overset{(c)}{\leq} \bar{p}\sqrt{\left|S^{\mathsf{round}}\right| T} + \bar{p}\left|S^{\mathsf{round}}\right| \overset{(d)}{\leq} 2\bar{p}\sqrt{\left|S^{\mathsf{round}}\right|(T+1)},
\end{aligned}
$$

where (a) holds by Jensen's inequality, (b) holds because the customer t will purchase product i with probability $\phi_i(S^{\mathsf{fluid}})$ when the set S^{fluid} is offered and thus $\mathbb{E}\left\{\sum_{t\in\mathcal{T}} C_{it}(\boldsymbol{q}^{\mathsf{fluid}})\right\} = \phi_i(S^{\mathsf{fluid}})T = q_i^{\mathsf{fluid}}$, (c) follows from Cauchy-Schwarz inequality and $\sum_{i\in S^{\mathsf{fluid}}} \phi_i(S^{\mathsf{fluid}}) \leq 1$ and (d) follows from Lemma 3.1 that $\left|S^{\mathsf{round}}\right| \leq T+1$. ■

LEMMA B.1. *Suppose $k\,p$ is integral, it holds that*

$$
\mathbb{E}\{[k\,p - \mathsf{binomial}(k,p)]^+\} \geq \frac{1}{\sqrt{2\pi}}\sqrt{k\,p\,(1-p)} - O(1).
$$

Proof: Let $q = 1-p$ and $(X_i)_{i=1}^k$ denote k independent and identical distributed binary random variable such that $X_i = \mathsf{Bernoulli}(p) - p$, then we have

$$
\begin{aligned}
&\mathbb{E}\left\{[k\,p - \mathsf{binomial}(k,p)]^+\right\} = \frac{1}{2}\mathbb{E}\left\{|k\,p - \mathsf{binomial}(k,p)| + k\,p - \mathsf{binomial}(k,p)\right] \\
&= \frac{1}{2}\mathbb{E}\left\{|k\,p - \mathsf{binomial}(k,p)|\right\} = \mathbb{E}\left\{[\mathsf{binomial}(k,p) - k\,p]^+\right\} = \sqrt{k\,p\,q}\,\mathbb{E}\left\{\left[\frac{\sum_{i=1}^k X_i}{\sqrt{k\,q}}\right]^+\right\} \\
&= \sqrt{k\,p\,q}\int_0^\infty \mathbb{P}\left(\frac{\sum_{i=1}^k X_i}{\sqrt{k\,p\,q}} > x\right) dx \geq \sqrt{k\,p\,q}\int_0^\infty \left(\mathbb{P}\left(\mathsf{Norm}(0,1) > x\right) - \frac{C}{(1+x^3)\sqrt{k\,p\,q}}\right) dx \\
&= \sqrt{\frac{k\,p\,q}{2\pi}} - C\int_0^\infty \frac{1}{1+x^3}\,dx,
\end{aligned}
$$

where C is a constant and the last inequality follows from the nonuniform Berry – Esseen inequality (Nagaev 1965). Note that

$$
\int_0^\infty \frac{1}{1+x^3}\,dx \leq 1 + \int_1^\infty \frac{1}{1+x^3}\,dx \leq 1 + \int_1^\infty \frac{1}{x^3}\,dx \leq \frac{3}{2},
$$

therefore,

$$\mathbb{E}\{[k\,p - \mathsf{binomial}(k,p)]^+\} \geq \sqrt{\frac{k\,p\,q}{2\pi}} - \frac{3}{2}C.$$

■

Appendix C: Results in Section 5

Lemma C.1. *Given any nonzero inventory $\boldsymbol{q} \in \mathbb{Z}_+^N$ with T customers, it holds that*

$$\mathbb{E}\{X_0(\boldsymbol{q},T)\} \leq 1 + w_0 \max_{i\in\mathcal{N}} \left\{\frac{q_i}{w_i}\right\} \log\left(\sum_{i\in\mathcal{N}} q_i\right) + \left[T - \sum_{i\in\mathcal{N}} q_i\right]^+.$$

Proof: Note that for each customer t, the no-purchase random variable follows the Bernoulli distribution with success probability $1 - \sum_{i\in\mathcal{N}} \phi_i(S_t)$ where S_t is the set of available items observed by the customer and $\phi_i(S_t)$ is the purchase probability of item i. Note that here $\boldsymbol{q}$ is a given initial inventory level. Let $\boldsymbol{q}' \in \mathbb{Z}_+^n$ denote the state variable representing the inventory level, the total expected number of no-purchase $\mathbb{E}\{X_0(\boldsymbol{q},T)\}$ can be recursively computed through:

$$V_t^0(\boldsymbol{q}') = \sum_{i\in A(\boldsymbol{q}')} \phi_i(A(\boldsymbol{q}'))V_{t+1}^0(\boldsymbol{q}' - \boldsymbol{e}_i) + \phi_0(A(\boldsymbol{q}'))\left(1 + V_{t+1}^0(\boldsymbol{q}')\right),$$

where $V_{T+1}^0(\boldsymbol{q}') = 0$ for any $\boldsymbol{q}'$, so that $\mathbb{E}\{X_0(\boldsymbol{q},T)\} = V_1^0(\boldsymbol{q})$.

Since each customer can buy at most one item, to upper bound $\mathbb{E}\{X_0(\boldsymbol{q},T)\}$, we claim it is sufficient to consider the problem where the decision maker removes one unit of product per period in order to maximize the total expected no-purchase probability $\bar{V}_1^0(\boldsymbol{q})$, i.e., $\bar{V}_1^0(\boldsymbol{q}) \geq V_1^0(\boldsymbol{q}) = \mathbb{E}\{X_0(\boldsymbol{q},T)\}$, in which $\bar{V}_1^0(\boldsymbol{q})$ can be recursively computed as:

$$\bar{V}_t^0(\boldsymbol{q}') = \phi_0(A(\boldsymbol{q}')) + \max_{i\in A(\boldsymbol{q}')} \bar{V}_{t+1}^0(\boldsymbol{q}' - \boldsymbol{e}_i),$$

where $\bar{V}_{T+1}^0(\boldsymbol{q}') = 0$ for any $\boldsymbol{q}'$. We prove this claim by induction. Note that $V_{T+1}^0(\boldsymbol{q}') = \bar{V}_{T+1}^0(\boldsymbol{q}')$ for any $\boldsymbol{q}'$. Now we assume $V_{t+1}^0(\boldsymbol{q}') \leq \bar{V}_{t+1}^0(\boldsymbol{q}')$ for any $\boldsymbol{q}'$. Then at time t and given any inventory $\boldsymbol{q}'$, we have

$$\begin{aligned}
V_t^0(\boldsymbol{q}') &= \sum_{i\in A(\boldsymbol{q}')} \phi_i(A(\boldsymbol{q}'))V_{t+1}^0(\boldsymbol{q}' - \boldsymbol{e}_i) + \phi_0(A(\boldsymbol{q}'))(1 + V_{t+1}^0(\boldsymbol{q}')) \\
&\leq \phi_0(A(\boldsymbol{q}')) + \phi_0(A(\boldsymbol{q}'))V_{t+1}^0(\boldsymbol{q}') + \left(\sum_{i\in A(\boldsymbol{q}')} \phi_i(A(\boldsymbol{q}'))\right) \max_{i\in A(\boldsymbol{q}')} V_{t+1}^0(\boldsymbol{q}' - \boldsymbol{e}_i) \\
&\overset{(a)}{\leq} \phi_0(A(\boldsymbol{q}')) + \phi_0(A(\boldsymbol{q}')) \max_{i\in A(\boldsymbol{q}')} V_{t+1}^0(\boldsymbol{q}' - \boldsymbol{e}_i) + \left(\sum_{i\in A(\boldsymbol{q}')} \phi_i(A(\boldsymbol{q}'))\right) \max_{i\in A(\boldsymbol{q}')} V_{t+1}^0(\boldsymbol{q}' - \boldsymbol{e}_i)
\end{aligned}$$

$$= \phi_0(A(\boldsymbol{q}')) + \max_{i\in A(\boldsymbol{q}')} V^0_{t+1}(\boldsymbol{q}' - \boldsymbol{e}_i)$$
$$\le \phi_0(A(\boldsymbol{q}')) + \max_{i\in A(\boldsymbol{q}')} \bar{V}^0_{t+1}(\boldsymbol{q}' - \boldsymbol{e}_i) = \bar{V}^0_t(\boldsymbol{q}'),$$

where inequality (a) holds because the no-purchase probability increases as products removed from an assortment under the MNL model. Therefore, the first claim holds.

However, analyzing $\bar{V}^0_1(\boldsymbol{q})$ is still not an easy task. In order to bound $\bar{V}^0_1(\boldsymbol{q})$, we consider an auxiliary problem:

$$\tilde{V}^0_t(\boldsymbol{q}') = \frac{w_0}{w_0 + \sum_{i:q'>0} w_i q'_i/q_i} + \max_{i:q'_i>0} \tilde{V}^0_{t+1}(\boldsymbol{q}' - \boldsymbol{e}_i), \tag{4}$$

where $\tilde{V}_{T+1}(\boldsymbol{q}') = 0$ for any $\boldsymbol{q}'$. We claim that given the initial inventory $\boldsymbol{q}$, for any time t and inventory $\boldsymbol{q}' \le \boldsymbol{q}$, $\bar{V}^0_t(\boldsymbol{q}') \le \tilde{V}^0_t(\boldsymbol{q}')$. We show the claim by induction. It is clear that the claim holds for time $T+1$. Now assume the claim holds for time $t+1$, then for time t, we have

$$\bar{V}^0_t(\boldsymbol{q}') = \frac{w_0}{w_0 + \sum_{i:q'_i>0} w_i} + \max_{i:q'_i>0} \bar{V}^0_{t+1}(\boldsymbol{q}' - \boldsymbol{e}_i) \le \frac{w_0}{w_0 + \sum_{i:q'_i>0} w_i} + \max_{i:q'_i>0} \tilde{V}^0_{t+1}(\boldsymbol{q}' - \boldsymbol{e}_i)$$
$$\le \frac{w_0}{w_0 + \sum_{i:q'_i>0} w_i q'_i/q_i} + \max_{i:q'_i>0} \tilde{V}^0_{t+1}(\boldsymbol{q}' - \boldsymbol{e}_i) = \tilde{V}^0_t(\boldsymbol{q}').$$

Therefore, the claim holds by induction, which implies that $\bar{V}^0_1(\boldsymbol{q}) \le \tilde{V}^0_1(\boldsymbol{q})$.

Note that the auxiliary problem is equivalent to the problem which splits each item with q_i inventories to q_i identical items with a preference weight w_i/q_i. Specifically, the total number of items now is $\sum_{i\in\mathcal{N}} q_i$ and each item has one unit of inventory. If an item k is split from the original product i, then its preference weight w'_k is w_i/q_i. Without loss of generality, we assume the $\sum_{i\in\mathcal{N}} q_i$ items follow the decreasing order of w'_k. Therefore, an optimal policy to the DP (4) is to select an item with largest index. This gives us

$$\tilde{V}^0_1(\boldsymbol{q}) \le \sum_{k=1}^{\sum_{i\in\mathcal{N}} q_i} \frac{w_0}{w_0 + \sum_{j=k}^{\sum_{i\in\mathcal{N}} q_i} w'_j} + \left[T - \sum_{i\in\mathcal{N}} q_i\right]^+$$
$$\le \left(\frac{w_0/\min_k\{w'_k\}}{w_0/\min_k\{w'_k\} + 1} + \cdots + \frac{w_0/\min_k\{w'_k\}}{w_0/\min_k\{w'_k\} + \sum_{i\in\mathcal{N}} q_i} \right) + \left[T - \sum_{i\in\mathcal{N}} q_i\right]^+$$
$$\le 1 + \frac{w_0}{\min_k\{w'_k\}} \left(\log\left(\frac{w_0}{\min_k\{w'_k\}} + \sum_{i\in\mathcal{N}} q_i \right) - \log\left(\frac{w_0}{\min_k\{w'_k\}} + 1 \right) \right) + \left[T - \sum_{i\in\mathcal{N}} q_i\right]^+$$
$$\le 1 + w_0 \max_{i\in\mathcal{N}} \left\{ \frac{q_i}{w_i} \right\} \log\left(\sum_{i\in\mathcal{N}} q_i \right) + \left[T - \sum_{i\in\mathcal{N}} q_i\right]^+.$$

The result then follows from $\mathbb{E}\{X_0(\boldsymbol{q},T)\} \le \tilde{V}_1^0(\boldsymbol{q})$. ■

In what follows, we let $\tilde{w} = w_0 + \sum_{i\in S^{\text{fluid}}} w_i$ and $\underline{\phi} = \min_{i\in S^{\text{fluid}}} \phi_i(S^{\text{fluid}})$. Moreover, we let $\phi_i = \phi_i(S^{\text{fluid}})$ to omit the dependence on the set S^{fluid} for purchase probability for simplicity.

Lemma C.2. *Suppose* $\tilde{T} = \left\lfloor T - \frac{1}{\underline{\phi}} - \sqrt{2\frac{1}{\underline{\phi}}T\log\left(|S^{\text{fluid}}|T\right)}\right\rfloor \ge 0$*, then it holds that*

$$\mathbb{P}\left(X_i(\boldsymbol{q}^{\text{round}},\tilde{T}) < q_i^{\text{round}}, \forall i \in S^{\text{fluid}}\right) \ge 1 - \frac{1}{T}.$$

Proof: Since $T \ge 1/\underline{\phi}$, we have $q_i^{\text{fluid}} \ge 1$ for any $i \in S^{\text{fluid}}$ and thus $A(\boldsymbol{q}^{\text{round}}) = S^{\text{fluid}}$. Note that when all products are available, the item purchased by a customer within $\tilde{T}$ periods follows a multinomial distribution $(X_0, X_1, \ldots, X_{|S^{\text{fluid}}|}) \sim (\tilde{T}, \phi_0, \ldots, \phi_{|S^{\text{fluid}}|})$. Therefore, by union bound,

$$\begin{aligned}\mathbb{P}\left(X_i(\boldsymbol{q}^{\text{round}},\tilde{T}) < q_i^{\text{round}}, \forall i \in S^{\text{fluid}}\right) &= \mathbb{P}\left(X_i < q_i^{\text{round}}, \forall i \in S^{\text{fluid}}\right) = 1 - \mathbb{P}\left(\exists i \in S^{\text{fluid}}, X_i \ge q_i^{\text{round}}\right)\\ &\ge 1 - \sum_{i\in S^{\text{fluid}}} \mathbb{P}\left(X_i \ge q_i^{\text{round}}\right).\end{aligned}$$

For any $i \in S^{\text{fluid}}$, since X_i is a binomial random variable, we have

$$\begin{aligned}\mathbb{P}\left(X_i \ge q_i^{\text{round}}\right) &\le \mathbb{P}\left(X_i \ge \phi_i T - 1\right) = \mathbb{P}\left(X_i \ge \phi_i\tilde{T} + \phi_i(T-\tilde{T}) - 1\right)\\ &= \mathbb{P}\left(X_i \ge \phi_i\tilde{T}\left(1 + \frac{T-\tilde{T}}{\tilde{T}} - \frac{1}{\phi_i\tilde{T}}\right)\right) \overset{(a)}{\le} \exp\left(-\frac{\phi_i\tilde{T}\left(\frac{T-\tilde{T}}{\tilde{T}} - \frac{1}{\phi_i\tilde{T}}\right)^2}{1 + \frac{T}{\tilde{T}} - \frac{1}{\phi_i\tilde{T}}}\right)\\ &= \exp\left(\frac{-\phi_i\left(T-\tilde{T}-\frac{1}{\phi_i}\right)^2}{\tilde{T}+T-\frac{1}{\phi_i}}\right) \le \exp\left(-\frac{\left(T-\tilde{T}-\frac{1}{\phi_i}\right)^2}{2\frac{1}{\phi_i}T}\right)\\ &\le \exp\left(-\frac{\left(T-\tilde{T}-\frac{1}{\underline{\phi}}\right)^2}{2\frac{1}{\underline{\phi}}T}\right) \le \frac{1}{|S^{\text{fluid}}|T},\end{aligned}$$

where (a) follows from Chernoff bound. ■

Lemma C.3. *Given no item is sold out till time* $\tilde{T}$*, the expected number of no purchase till time* $\tilde{T}$ *is bounded by*

$$\mathbb{E}\left\{X_0(\boldsymbol{q}^{\text{round}},\tilde{T})\middle|X_i(\boldsymbol{q}^{\text{round}},\tilde{T}) < q_i^{\text{round}}, \forall i \in S^{\text{fluid}}\right\} \le \phi_0 T.$$

Proof: Using Lemma C.2, we have

$$\mathbb{E}\left\{X_0(\boldsymbol{q}^{\text{round}},\tilde{T})\middle|X_i(\boldsymbol{q}^{\text{round}},\tilde{T}) < q_i^{\text{round}}, \forall i \in S^{\text{fluid}}\right\}$$

$$
\begin{aligned}
&= \mathbb{E}\left\{\tilde{T} - \sum_{i\in\mathcal{N}} X_i(\boldsymbol{q}^{\text{round}}, \tilde{T}) \middle| X_i(\boldsymbol{q}^{\text{round}}, \tilde{T}) < q_i^{\text{round}}, \forall i \in S^{\text{fluid}}\right\} \\
&= \mathbb{E}\left\{\tilde{T} - \sum_{i\in\mathcal{N}} X_i \middle| X_i < q_i^{\text{round}}, \forall i \in S^{\text{fluid}}\right\} = \mathbb{E}\left\{X_0 \middle| X_i < q_i^{\text{round}}, \forall i \in S^{\text{fluid}}\right\} \\
&\le \frac{\mathbb{E}\{X_0\}}{\mathbb{P}(X_i < q_i^{\text{round}}, \forall i \in S^{\text{fluid}})} \le \frac{T}{T-1}\phi_0\tilde{T} \le \phi_0 T.
\end{aligned}
$$

■

LEMMA C.4. *Suppose $\phi_0/\underline{\phi} \le \kappa$, given no item is sold out till time $\tilde{T}$, let $\tilde{\boldsymbol{q}} = \boldsymbol{q}^{\text{round}} - \boldsymbol{X}$, where $\boldsymbol{X}$ is the multinomial random variable defined in the proof of Lemma C.2, there exist constant D_1 and D_2 such that*

$$
\mathbb{E}\left\{\left[T - \tilde{T} - \sum_{i\in S^{\text{fluid}}} \tilde{q}_i\right]^+ \middle| X_i(\boldsymbol{q}^{\text{round}}, \tilde{T}) < q_i^{\text{round}}, \forall i \in S^{\text{fluid}}\right\} \le D_1\sqrt{T\log(T)} + D_2.
$$

Proof: Since $\sum_{i\in S^{\text{fluid}}} \phi_i \le 1$, we have $\underline{\phi} \le 1/|S^{\text{fluid}}|$ and thus $\phi_0 \le \kappa/|S^{\text{fluid}}|$, then

$$
\begin{aligned}
&\mathbb{E}\left\{\left[T - \tilde{T} - \sum_{i\in S^{\text{fluid}}} \tilde{q}_i\right]^+ \middle| X_i(\boldsymbol{q}^{\text{round}}, \tilde{T}) < q_i^{\text{round}}, \forall i \in S^{\text{fluid}}\right\} \\
&= \mathbb{E}\left\{\left[T - \sum_{i\in S^{\text{fluid}}} q_i^{\text{round}} - \tilde{T} + \sum_{i\in\in S^{\text{fluid}}} X_i\right]^+ \middle| X_i < q_i^{\text{round}}, \forall i \in S^{\text{fluid}}\right\} \\
&\overset{(a)}{\le} \mathbb{E}\left\{\left[T - \sum_{i\in S^{\text{fluid}}} q_i^{\text{fluid}} - \tilde{T} + \sum_{i\in S^{\text{fluid}}} X_i\right]^+ \middle| X_i < q_i^{\text{round}}, \forall i \in S^{\text{fluid}}\right\} \\
&\overset{(b)}{=} \mathbb{E}\left\{\left[\phi_0 T - \tilde{T} + \sum_{i\in S^{\text{fluid}}} X_i\right]^+ \middle| X_i < q_i^{\text{round}}, \forall i \in S^{\text{fluid}}\right\} \le \mathbb{E}\left\{\left[\phi_0 T - \tilde{T} + \sum_{i\in S^{\text{fluid}}} X_i\right]^+\right\} \\
&\le \sqrt{\mathbb{E}\left\{\left(\sum_{i\in S^{\text{fluid}}} X_i - \tilde{T} + \phi_0 T\right)^2\right\}} \overset{(c)}{=} \sqrt{\mathbb{E}\left\{(\phi_0 T - X_0)^2\right\}} \\
&\overset{(d)}{=} \sqrt{\phi_0(1-\phi_0)\tilde{T} + \left(\phi_0(T-\tilde{T})\right)^2} \\
&\le \sqrt{\phi_0^2\left(\frac{1}{\underline{\phi}} + 1 + \sqrt{2\frac{1}{\underline{\phi}} T\log(|S^{\text{fluid}}|T)}\right)^2 + \phi_0\tilde{T}} \\
&= \sqrt{\left(\frac{\phi_0}{\underline{\phi}}\right)^2 + 2\frac{\phi_0^2}{\underline{\phi}}\left(1 + T\log(|S^{\text{fluid}}|T)\right) + \phi_0^2\left(1 + 2\left(1 + \frac{1}{\underline{\phi}}\right)\sqrt{2\frac{1}{\underline{\phi}} T\log(|S^{\text{fluid}}|T)}\right) + \phi_0\tilde{T}}
\end{aligned}
$$

$$\leq \sqrt{\kappa^2 + \frac{2\kappa^2}{|S^{\text{fluid}}|}\left(1+T\log(|S^{\text{fluid}}|T)\right) + \frac{\kappa^2}{|S^{\text{fluid}}|^2} + 2\left(\sqrt{\frac{\kappa^4}{|S^{\text{fluid}}|^3}} + \sqrt{\frac{\kappa^4}{|S^{\text{fluid}}|}}\right)\sqrt{2T\log(|S^{\text{fluid}}|T)} + \frac{\kappa}{|S^{\text{fluid}}|}\tilde{T}}$$

$$\leq \sqrt{4\kappa^2 + 2\kappa^2\frac{T}{|S^{\text{fluid}}|}\log(|S^{\text{fluid}}|T) + 4\kappa^2\sqrt{2\frac{T}{|S^{\text{fluid}}|}\log(|S^{\text{fluid}}|T)} + \kappa\frac{T}{|S^{\text{fluid}}|}},$$

where (a) follows from Lemma 3.1, (b) holds because $q_i^{\text{fluid}} = \phi_i T$, (c) holds because $\boldsymbol{X}$ is a multinomial random variable with $\tilde{T}$ trials, and (d) follows from the fact that X_0 follows $\mathsf{binomial}(\tilde{T}, \phi_0)$. Since $\tilde{T} \geq 0$, we have $q_i^{\text{fluid}} \geq 1$ for any product $i \in S^{\text{fluid}}$ and thus $|S^{\text{fluid}}| \leq$T. Therefore, there exist constants D_1 and D_2 such that

$$\mathbb{E}\left\{\left[T - \tilde{T} - \sum_{i\in S^{\text{fluid}}}\tilde{q}_i\right]^+ \middle| X_i(\boldsymbol{q}^{\text{round}}, \tilde{T}) < q_i^{\text{round}}, \forall i \in S^{\text{fluid}}\right\} \leq D_1\sqrt{T\log(T)} + D_2.$$

Lemma C.5. *Suppose that $\phi_0/\underline{\phi} \leq \kappa$, given no item is sold out till time $\tilde{T}$, there exist constants D_3 and D_4 such that*

$$\phi_0\mathbb{E}\left\{\max_{i\in S^{\text{fluid}}}\left\{\frac{\tilde{q}_i}{\phi_i}\right\}\middle| X_i(\boldsymbol{q}^{\text{round}}, \tilde{T}) < q_i^{\text{round}}, \forall i \in S^{\text{fluid}}\right\} \leq D_3\sqrt{T\log(T)} + D_4.$$

Proof: By definition, we have

$$\mathbb{E}\left\{\max_{i\in S^{\text{fluid}}}\left\{\frac{\tilde{q}_i}{\phi_i}\right\}\middle| X_i(\boldsymbol{q}^{\text{round}}, \tilde{T}) < q_i^{\text{round}}, \forall i \in S^{\text{fluid}}\right\} = \mathbb{E}\left\{\max_{i\in S^{\text{fluid}}}\left\{\frac{q_i^{\text{round}} - X_i}{\phi_i}\right\}\middle| X_i < q_i^{\text{round}}, \forall i \in S^{\text{fluid}}\right\}$$

$$\leq \mathbb{E}\left\{\max_{i\in S^{\text{fluid}}}\left\{\frac{q_i^{\text{fluid}} - X_i + 1}{\phi_i}\right\}\middle| X_i < q_i^{\text{round}}, \forall i \in S^{\text{fluid}}\right\}.$$

Note that

$$\mathbb{E}\left\{\max_{i\in S^{\text{fluid}}}\left\{\frac{q_i^{\text{fluid}} - X_i}{\phi_i}\right\}\middle| X_i < q_i^{\text{round}}, \forall i \in S^{\text{fluid}}\right\}$$

$$\leq \frac{1}{\mathbb{P}\left(X_i < q_i^{\text{round}}, \forall i \in S^{\text{fluid}}\right)}\mathbb{E}\left\{\max_{i\in S^{\text{fluid}}}\left[\frac{q_i^{\text{fluid}} - X_i}{\phi_i}\right]^+\right\}$$

$$\overset{(a)}{\leq} \frac{T}{T-1}\left(\max_{i\in S^{\text{fluid}}}\left\{\mathbb{E}\left\{\left[\frac{q_i^{\text{fluid}} - X_i}{\phi_i}\right]^+\right\}\right\} + \sqrt{\sum_{i\in S^{\text{fluid}}}\mathbb{V}\text{ar}\left\{\left[\frac{q_i^{\text{fluid}} - X_i}{\phi_i}\right]^+\right\}}\right)$$

$$\leq \frac{T}{T-1}\left(\max_{i\in S^{\text{fluid}}}\left\{\sqrt{\mathbb{E}\left\{\left(\frac{q_i^{\text{fluid}} - X_i}{\phi_i}\right)^2\right\}}\right\} + \sqrt{\sum_{i\in S^{\text{fluid}}}\mathbb{E}\left\{\left(\frac{q_i^{\text{fluid}} - X_i}{\phi}\right)^2\right\}}\right),$$

where (a) follows from Theorem 2.1 in Aven (1985). Finally, we have

$$\left(\frac{\phi_0}{\phi_i}\right)^2 \mathbb{E}\left\{\left(q_i^{\mathsf{fluid}} - X_i\right)^2\right\} = \left(\frac{\phi_0}{\phi_i}\right)^2 \left(\mathbb{V}\mathrm{ar}\left\{X_i\right\} + \left(q_i^{\mathsf{fluid}} - \mathbb{E}\left\{X_i\right\}\right)^2\right)$$
$$= \frac{\phi_0^2}{\phi_i}(1-\phi_i)\tilde{T} + \phi_0^2\left(T-\tilde{T}\right)^2 \le \frac{\phi_0^2}{\phi_i}\tilde{T} + \phi_0^2\left(\frac{1}{\underline{\phi}} + 1 + \sqrt{2\frac{1}{\underline{\phi}}T\log(|S^{\mathsf{fluid}}|T)}\right)^2$$
$$\le \frac{\phi_0^2}{\phi_i}\tilde{T} + \phi_0^2\left(\frac{1}{\underline{\phi}}+1\right)^2 + 2\phi_0^2\left(\frac{1}{\underline{\phi}}+1\right)\sqrt{2\frac{1}{\underline{\phi}}T\log\left(|S^{\mathsf{fluid}}|T\right)} + 2\frac{\phi_0^2}{\underline{\phi}}T\log(|S^{\mathsf{fluid}}|T)$$
$$\le \frac{\kappa^2}{|S^{\mathsf{fluid}}|}T + 4\kappa^2 + 2\kappa^2\sqrt{2\frac{T}{|S^{\mathsf{fluid}}|}\log(|S^{\mathsf{fluid}}|T)} + 2\kappa^2\sqrt{\frac{T}{|S^{\mathsf{fluid}}|^3}\log(|S^{\mathsf{fluid}}|T)} + 2\kappa^2\frac{T}{|S^{\mathsf{fluid}}|}\log(|S^{\mathsf{fluid}}|T)$$
$$\le 4\kappa^2 + 3\kappa^2\frac{T}{|S^{\mathsf{fluid}}|}\log(|S^{\mathsf{fluid}}|T) + 6\kappa^2\sqrt{\frac{T}{|S^{\mathsf{fluid}}|}\log(|S^{\mathsf{fluid}}|T)}.$$

Therefore, we have

$$\phi_0\mathbb{E}\left\{\max_{i\in S^{\mathsf{fluid}}}\left\{\frac{\tilde{q}_i}{\phi_i}\right\}\middle| X_i(\boldsymbol{q}^{\mathsf{round}},\tilde{T}) < q_i^{\mathsf{round}}, \forall i\in S^{\mathsf{fluid}}\right\}$$
$$\le \frac{\phi_0}{\underline{\phi}} + \frac{T}{T-1}\left(1+\sqrt{|S^{\mathsf{fluid}}|}\right)\sqrt{4\kappa^2 + 3\kappa^2\frac{T}{|S^{\mathsf{fluid}}|}\log(|S^{\mathsf{fluid}}|T) + 6\kappa^2\sqrt{\frac{T}{|S^{\mathsf{fluid}}|}\log(|S^{\mathsf{fluid}}|T)}}$$
$$\le \kappa + 4\sqrt{4\kappa^2|S^{\mathsf{fluid}}| + 3\kappa^2 T\log(|S^{\mathsf{fluid}}|T) + 6\kappa^2\sqrt{T|S^{\mathsf{fluid}}|\log(|S^{\mathsf{fluid}}|T)}}.$$

Since $\tilde{T} \ge 0$, we have $q_i^{\mathsf{fluid}} \ge 1$ for any product $i \in S^{\mathsf{fluid}}$ and then $|S^{\mathsf{fluid}}| \le \sum_{i\in S^{\mathsf{fluid}}} q_i^{\mathsf{fluid}} \le T$, Therefore, there exist constants D_3 and D_4 such that

$$\phi_0\mathbb{E}\left\{\max_{i\in S^{\mathsf{fluid}}}\left\{\frac{\tilde{q}_i}{\phi_i}\right\}\middle| X_i(\boldsymbol{q}^{\mathsf{round}},\tilde{T}) < q_i^{\mathsf{round}}, \forall i\in S^{\mathsf{fluid}}\right\} \le D_3\sqrt{T\log(T)} + D_4.$$

■

Proof of Theorem 5.1:

Note that for MNL model, we have $1/\underline{\phi} = \tilde{w}/\underline{w}$ and $\phi_0/\underline{\phi} = w_0/\underline{w}$. We focus on problems with $T \ge 3$, otherwise, we have $V^{\mathsf{fluid}} - \Pi(\boldsymbol{q}^{\mathsf{round}}) \le 4\bar{p}$. We consider two possible cases:

(i) The demand is relatively small, specifically,

$$T \le \frac{1}{\underline{\phi}} + \sqrt{2\frac{1}{\underline{\phi}}T\log(|S^{\mathsf{fluid}}|T)}.$$

We claim that $T \le 8(1/\underline{\phi})\log(|S^{\mathsf{fluid}}|T)$. Suppose not, then

$$\frac{1}{\underline{\phi}} + \sqrt{2\frac{1}{\underline{\phi}}T\log(|S^{\mathsf{fluid}}|T)} < \frac{T}{8\log(|S^{\mathsf{fluid}}|T)} + \frac{T}{2} < T,$$

which leads to a contradiction and thus the claim holds. By Lemma C.1, we have

$$
\begin{aligned}
V^{\text{fluid}} - \Pi(\boldsymbol{q}^{\text{round}}) &\le \bar{p}\left(\mathbb{E}\left\{X_0(\boldsymbol{q}^{\text{round}}, T)\right\} - T\phi_0 + 1\right) \\
&\le \bar{p}\left(w_0 \max_{i\in\mathcal{N}}\left\{\frac{q_i^{\text{round}}}{w_i}\right\}\log\left(\sum_{i\in\mathcal{N}} q_i^{\text{round}}\right) + \left[T - \sum_{i\in\mathcal{N}} q_i^{\text{round}}\right]^+ - \frac{Tw_0}{\tilde{w}}\right) + 2\bar{p} \\
&\overset{(a)}{\le} \bar{p}\left(w_0 \max_{i\in\mathcal{N}}\left\{\frac{\lceil q_i^{\text{fluid}}\rceil}{w_i}\right\}\log\left(\sum_{i\in S^{\text{fluid}}} q_i^{\text{fluid}} + 1\right) + T - \sum_{i\in S^{\text{fluid}}} q_i^{\text{fluid}} - \frac{w_0}{\tilde{w}}T\right) + 2\bar{p} \\
&= \bar{p}w_0 \max_{i\in\mathcal{N}}\left\{\frac{\lceil q_i^{\text{fluid}}\rceil}{w_i}\right\}\log\left(\sum_{i\in S^{\text{fluid}}} q_i^{\text{fluid}} + 1\right) + 2\bar{p} \le \bar{p}w_0\left(\frac{T}{\tilde{w}} + \frac{1}{\underline{w}}\right)\log(T+1) + 2\bar{p} \\
&\overset{(b)}{\le} \bar{p}\kappa\left(8\log\left(|S^{\text{round}}|T\right) + 1\right)\log(1+T) + 2\bar{p} \\
&\overset{(c)}{\le} \bar{p}\kappa\left(8\log\left(T\left(T+1\right)\right) + 1\right)\log(1+T) + 2\bar{p},
\end{aligned}
$$

where (a) and (c) follows from Lemma 3.1 and (b) holds because $T \le 8(\tilde{w}/\underline{w})\log(|S^{\text{fluid}}|T)$ and if $T > \tilde{w}/\underline{w}$, then $S^{\text{round}} = S^{\text{fluid}}$.

(ii) The demand is relatively large, specifically,

$$
T > \frac{1}{\underline{\phi}} + \sqrt{2\frac{1}{\underline{\phi}}T\log(|S^{\text{fluid}}|T)}.
$$

Combining Lemma C.1, C.2, C.3, C.4 and C.5, we have

$$
\begin{aligned}
&V^{\text{fluid}} - \Pi(\boldsymbol{q}^{\text{round}}) \le \bar{p}\left(\mathbb{E}\left\{X_0(\boldsymbol{q}^{\text{round}}, T)\right\} - \frac{Tw_0}{\tilde{w}}\right) + \bar{p} \\
&\le \bar{p}\left(\mathbb{E}\left\{X_0(\boldsymbol{q}^{\text{round}}, T)\middle| X_i(\boldsymbol{q}^{\text{round}}, \tilde{T}) < q_i^{\text{round}}, \forall i \in S^{\text{fluid}}\right\}\mathbb{P}\left(X_i(\boldsymbol{q}^{\text{round}}, \tilde{T}) < q_i^{\text{round}}, \forall i \in S^{\text{fluid}}\right) - \frac{w_0}{\tilde{w}}T + 2\right) \\
&\le \bar{p}\left(\mathbb{E}\left\{X_0(\boldsymbol{q}^{\text{round}}, T)\middle| X_i(\boldsymbol{q}^{\text{round}}, \tilde{T}) < q_i^{\text{round}}, \forall i \in S^{\text{fluid}}\right\} + 2 - \frac{w_0}{\tilde{w}}T\right) \\
&= \bar{p}\left(\mathbb{E}\left\{X_0(\boldsymbol{q}^{\text{round}}, \tilde{T}) + w_0 \max_{i\in S^{\text{fluid}}}\left\{\frac{\tilde{q}_i}{w_i}\right\}\log\left(\sum_{i\in S^{\text{fluid}}}\tilde{q}_i\right)\middle| X_i(\boldsymbol{q}^{\text{round}}, \tilde{T}) < q_i^{\text{round}}, \forall i \in S^{\text{fluid}}\right\} - \frac{w_0}{\tilde{w}}T + 3\right) \\
&\quad + \bar{p}\mathbb{E}\left\{\left[T - \tilde{T} - \sum_{i\in S^{\text{fluid}}}\tilde{q}_i\right]^+\middle| X_i(\boldsymbol{q}^{\text{round}}, \tilde{T}) < q_i^{\text{round}}, \forall i \in S^{\text{fluid}}\right\} \\
&\le \bar{p}\left((D_1 + D_3)\sqrt{T\log(T)} + D_2 + D_4 + 3\right).
\end{aligned}
$$

Combining case (i) and (ii) leads to the result. ■

Appendix D: Extensions

In this section, we explore several generalizations to the MNL model, demonstrating the robustness of our solution in more realistic scenarios. We extend our analysis and results into three distinct settings: (i) Poisson demand, (ii) Alternative choice models beyond the MNL model (parsimonious GAM, Nested Logit with one nest), and (iii) Situations involving a capacity constraint.

D.1. Poisson Demand

Here we consider the problem in which the number of customers follows a Poisson distribution with mean T. In this particular setting, our approach is to solve the fluid approximation (2) where the expected demand T is utilized. Subsequently, we employ our rounding scheme to derive the corresponding solution. An essential characteristic of the Poisson distribution is that the variance of the random variable equals to its expectation. Due to this property, the fluid relaxation with expected demand continues to serve as a good approximation to the original problem, which leads to the performance guarantees similar to previous analysis in terms of order in T. We formally present the result in the following theorem. Let $\Pi^{\text{random}}(\boldsymbol{q})$ denote the total expected profit under the Poisson demand given the initial inventory $\boldsymbol{q}$.

Theorem D.1. Suppose the number of customer follows a Poisson distribution with mean T, there exists an absolute constant C_4 such that for any inventory $\boldsymbol{q}$,

$$\Pi(\boldsymbol{q}) - \Pi^{\text{random}}(\boldsymbol{q}) \le C_4\sqrt{T\log T}.$$

By standard argument (e.g., Liang et al. 2022), V^{fluid} is still an upper bound to the problem $\max_{\boldsymbol{q}\in\mathbb{Z}_+^n} \Pi^{\text{random}}(\boldsymbol{q})$. Because

$$V^{\text{fluid}} - \Pi^{\text{random}}(\boldsymbol{q}^{\text{round}}) = V^{\text{fluid}} - \Pi(\boldsymbol{q}^{\text{round}}) + \Pi(\boldsymbol{q}^{\text{round}}) - \Pi^{\text{random}}(\boldsymbol{q}^{\text{round}}),$$

Theorem D.1 implies that when the customer number follows a Poisson distribution, our solution $\boldsymbol{q}^{\text{round}}$ is still within an optimality gap of $O(\log T\sqrt{T\log T})$ under the MNL choice model.

D.2. Other Choice Models

Our analysis, initially developed for the MNL model, possesses the flexibility to be extended to other choice models that exhibit substitutability under certain assumptions. In this section, we explore the performance of our solution within the context of two distinct choice models: the parsimonious general attraction model and nested logit choice model with one nest.

The MNL choice model can be viewed as a special case of the basic attraction model (BAM) developed by Luce (2012). Gallego et al. (2015) develop a generalization of the BAM called the general attraction model (GAM). Under the parsimonious version of this choice model, a customer associates a preference weight of w_i with product i and a preference weight of w_0 with the no-purchase option. If the set of available products is S, then a customer chooses product i with probability

$$\phi_i(S) = \mathbf{1}(i \in S) \frac{w_i}{w_0 + \theta \sum_{j \in \mathcal{N}} w_j + (1-\theta) \sum_{j \in S} w_j},$$

where θ is the attraction factor. It then follows that the no-purchase probability given a set S is

$$\phi_0(S) = \frac{w_0 + \theta \sum_{i \notin S} w_i}{w_0 + \theta \sum_{i \in \mathcal{N}} w_i + (1-\theta) \sum_{i \in S} w_i}.$$

It is known that the optimal solution to problem (2) under this parsimonious model is margin-ordered.

In the next theorem, we give a performance guarantee for the stocking quantities $\boldsymbol{q}^{\text{round}}$ under the parsimonious GAM.

Theorem D.2. Considering customers choosing under the parsimonious GAM, if $\bar{p}$ and $(w_0 + \theta \sum_{i \in \mathcal{N}} w_i)/\underline{w}$ are independent of n and T, then there exist absolute constants C_5 and C_6 such that

$$V^{\text{fluid}} - \Pi(\boldsymbol{q}^{\text{round}}) \le C_5 \log T \sqrt{T \log T} + C_6.$$

The proof of the theorem is analogous to the proof of Theorem 5.1. It is necessary to assume $(w_0 + \theta \sum_{i \in \mathcal{N}} w_i)/\underline{w}$ is independent of n and T because sufficient substitutability between products is required to show results analogous to Lemma C.1, which is the key step to obtain a bound independent of n. In the extreme case where $\theta = 1$, we have $\phi_i(S) = \mathbf{1}(i \in S) \frac{w_i}{w_0 + \theta \sum_{j \in \mathcal{N}} w_j}$ and $\phi_0(S) = \frac{w_0 + \sum_{i \notin S} w_i}{w_0 + \sum_{i \in \mathcal{N} w_i}}$. Therefore, once a product becomes unavailable, customers who prefer that product will leave without substituting to other products. A special case when $\theta = 1$ is the independent demand model discussed in Section 4 by taking $w_0 = 1$ and $w_i = 1, \forall i \in \mathcal{N}$. Thus, it is impossible to achieve an optimality gap independent of n without sufficient substitutability.

The nested logit model was developed to avoid the independence of irrelevant alternative property suffered by the MNL choice model. Under this model, customers first select a nest, and then, a product within the nest. To showcase how our analysis may work for nested logit model, we consider a simple case in which there is only one nest. Under this model, given the offered set S, the probability that the customer purchases product $i \in S$ is given by

$$\phi_i(S) = \mathbf{1}(i \in S) \frac{(\sum_{i \in S} w_i)^\alpha}{w_0 + (\sum_{i \in S} w_i)^\alpha} \frac{w_i}{\sum_{i \in S} w_i},$$

where $\alpha \geq 0$ is a parameter characterizing the degree of dissimilarity of the products in the nest. The probability of no-purchase given the set S is then

$$\phi_0(S) = \frac{w_0}{w_0 + (\sum_{i\in S} w_i)^{\alpha}}.$$

In the next theorem, we give the performance guarantee for $\boldsymbol{q}^{\text{round}}$ under the special nested logit model.

THEOREM D.3. Considering customers choosing under the nested logit model with one nest, Suppose for any set S it holds that $(\sum_{i\in S} w_i)^{\alpha} \geq \alpha(\sum_{i\in S} w_i)$, and if $\bar{p}$ and $w_0/(\alpha \underline{w})$ are independent of n and T, then there exist absolute constants C_7 and C_8 such that

$$V^{\text{fluid}} - \Pi(\boldsymbol{q}^{\text{round}}) \leq C_7 \log T \sqrt{T \log T} + C_8.$$

D.3. Capacitated Problems under the Multinomial Logit Model

Due to limited storage space and warehouse capacity, it can be more realistic to consider the inventory planning problem with a capacity constraint. Specifically, we consider the following setting: the initial stocking quantities $\boldsymbol{q}$ must satisfy the constraint

$$\sum_{i\in\mathcal{N}} q_i \leq K, \tag{5}$$

where $K > 0$ denotes the total capacity. In this case, the corresponding fluid relaxation is:

$$\max_{\boldsymbol{q}\geq 0} \sum_{i\in\mathcal{N}} p_i X_i^{\mathrm{F}}(\boldsymbol{q}, T) - \sum_{i\in\mathcal{N}} c_i q_i, \text{ s.t.} \sum_{i\in\mathcal{N}} q_i \leq K, \tag{6}$$

where $X_i^{\mathrm{F}}(\boldsymbol{q}, T)$ is a deterministic value denoting the (fractional) quantities of product i sold in the fluid system given the initial inventory $\boldsymbol{q}$. Following Liang et al. (2022), the fluid relaxation (6) provides an upper bound to the original problem (1) and can be equivalently written as an LP under the MNL choice model:

$$\begin{aligned}
\max \ & \sum_{i\in\mathcal{N}} (r_i - c_i)\theta_i \\
\text{s.t. } & 0 \leq \theta_i \leq w_i \theta_0, \forall i \in \mathcal{N}, \\
& \sum_{i\in\mathcal{N}} \theta_i + \theta_0 = T, \\
& \sum_{i\in\mathcal{N}} \theta_i \leq K, \\
& \theta_0 \geq 0.
\end{aligned}$$

The decision variable $(\theta_i)_{i\in\mathcal{N}}$ can be interpreted as the quantity of product i sold under the fluid system. Note that given any θ_0, the LP is reduced to a relaxed knapsack problem. Because products are indexed by margin order, we can show there exists an optimal solution $(\theta_i^*)_{i\in\mathcal{N}\cup\{0\}}$ to the fluid LP such that there exists a product k satisfying

$$\theta_i^* \begin{cases} = w_i\theta_0^*, & \text{if } i \le k, \\ \le w_i\theta_0^*, & \text{if } i = k+1, \\ = 0, & \text{if } i > k+1. \end{cases}$$

This property implies that in the fluid system, given an optimal fluid inventory $\boldsymbol{q}^{\mathsf{fluid}} = (\theta_1^*, \ldots, \theta_N^*)$, there can be at most one stockout before time T. We restrict our attention to the case $\theta_{k+1}^* < w_{k+1}\theta_0^*$. Otherwise, the analysis is identical to the problem without the capacity constraint beacause all products offered are depleted at time T. Let $T_1 < T$ denote the first stockout time in the fluid system. Then at time T_1, product $k+1$ becomes unavailable in the fluid system. All remaining products $i \le k$ are sold out at time T. The stockout time T_1 complicates the problem and the resulting analysis. Nevertheless, we can generalize the previous approach to obtain a performance bound independent of the number of products. The next theorem formally state our results.

THEOREM D.4. Considering the retailer facing a capacity constraint (5) and customers choosing under the multinmomial logit model, if $\bar{p}$ and w_i/w_j for any $i, j \in \mathcal{N}\cup\{0\}$ are independent of n and T, then there exist feasible stocking quantities $\boldsymbol{q}^{\mathsf{fea}}$ and absolute constants C_9 and C_{10} such that

$$V^{\mathsf{fluid}} - \Pi(\boldsymbol{q}^{\mathsf{fea}}) \le C_9 \log T \sqrt{T\log T} + C_{10}.$$

We give the proof of the theorem in Appendix E. Theorem D.4 again implies that even with a capacity constraint, the optimality gap of our solution is still $O(\log T\sqrt{T\log T})$. We briefly outline the proof of the theorem. We consider three possible cases in terms of the value of the first stockout time T_1: (i) T_1 is relatively small, (ii) $T - T_1$ is relatively small and (iii) otherwise. Intuitively, for cases (i) and (ii), the problem is “close” to the one in which all products are depleted at the same time and thus can be handled analogously to the proof of Theorem 5.1. Thus, we are left to analyze case (iii). Due to the capacity constraint (5), the rounded solution $\boldsymbol{q}^{\mathsf{round}}$ to $\boldsymbol{q}^{\mathsf{fluid}}$ may not be feasible to the original problem. By Lemma 3.1, the total inventory of $\boldsymbol{q}^{\mathsf{round}}$ is at most $K+1$. Therefore, to make it implementable, we can always remove one unit inventory of the product with smallest revenue. Let $\boldsymbol{q}^{\mathsf{fea}}$ denote the resulting stocking quantities. Using Chernoff bound, we first characterize a time period $\tilde{T}$ close to T such that no product $i \le k$ is sold out by time $\tilde{T}$ with high probability. We then show that the expected number of no-purchase until time $\tilde{T}$ is upper bounded by the number of no-purchase in the fluid system. Therefore, it is sufficient to analyze the expected number of no-purchase after time $\tilde{T}$, which can be further reduced to analyze the inventory left at time $\tilde{T}$. Thus, we want to find a "good" lower bound to the number of product purchased before time $\tilde{T}$, which becomes more difficult because the stockout time of product $k+1$ is random. To overcome this challenge, we consider an auxiliary inventory $\tilde{\boldsymbol{q}}^{\mathsf{round}}$ which perturbs q_{k+1}^{round} down so that product $k+1$ is sold out before time T_1 with high probability. Let S_1 denote the set of product $i \le k+1$ and S_2 denote the set of products $i \le k$. Conditional on product $k+1$ is sold out before time T_1, the number of product i sold, $X_i(\tilde{\boldsymbol{q}}^{\mathsf{round}}, \tilde{T})$, stochastic dominates a sum of two binomial random variables $\mathsf{binomial}(T_1, \phi_i(S_1))$ and $\mathsf{binomial}(\tilde{T} - T_1, \phi_i(S_2))$. Lastly, we can bound the difference between $X_i(\boldsymbol{q}^{\mathsf{round}}, \tilde{T})$ and $X_i(\tilde{\boldsymbol{q}}^{\mathsf{round}}, \tilde{T})$ by the perturbed amount. Combining these analysis together again leads to a $O(\log T\sqrt{T\log T})$ optimality gap.

Appendix E: Results in Section D

Proof of Theorem D.1:

Let $\Pi^{\mathsf{random}}(\boldsymbol{q}, \hat{T})$ denote the total expected profit when the number of customer is realized to be $\hat{T}$, then we have

$$\begin{aligned}
\Pi(\boldsymbol{q}) - \Pi^{\mathsf{random}}(\boldsymbol{q}) &= \Pi(\boldsymbol{q}) - \mathbb{E}\left[\Pi^{\mathsf{random}}(\boldsymbol{q}, \hat{T})\right] \\
&\le \left(\Pi(\boldsymbol{q}) - \mathbb{E}\left[\Pi^{\mathsf{random}}(\boldsymbol{q}, \hat{T}) \middle| \hat{T} \le T - \delta\right]\right) \mathbb{P}\left(\hat{T} \le T - \delta\right) + \bar{p}\delta\mathbb{P}\left(\hat{T} > T - \delta\right) \\
&\le \bar{p}T\mathbb{P}\left(\hat{T} \le T - \delta\right) + \bar{p}\delta \\
&= \bar{p}\sqrt{3T\log T} + \bar{p},
\end{aligned}$$

where the last equality follows from the Poisson concentration inequality by taking $\delta = \sqrt{3T\log T}$.

■

Proof of Theorem D.2:

Suppose customers choose under the parsimonious GAM, given any inventory $\boldsymbol{q}$, we have an analogous result to Lemma C.1. Since for any set S, we have

$$\phi_0(S) = \frac{w_0 + \theta\sum_{i\notin S} w_i}{w_0 + \theta\sum_{i\notin S} w_i + \sum_{i\in S} w_i} \le \frac{w_0 + \theta\sum_{i\in\mathcal{N}} w_i}{w_0 + \theta\sum_{i\in\mathcal{N}} w_i + \sum_{i\in S} w_i}.$$

Therefore, analogous to the proof of Lemma C.1, it holds that

$$\mathbb{E}\left\{X_0(\boldsymbol{q},T)\right\} \leq 1 + \left(w_0 + \theta\sum_{i\in\mathcal{N}} w_i\right)\max_{i\in\mathcal{N}}\left\{\frac{q_i}{w_i}\right\}\log\left(\sum_{i\in\mathcal{N}} q_i\right) + \left[T - \sum_{i\in\mathcal{N}} q_i\right]^+.$$

Again, we consider the two possible cases:

(i) If

$$T \leq \frac{1}{\underline{\phi}} + \sqrt{2\frac{1}{\underline{\phi}}T\log(|S^{\text{fluid}}|T)},$$

then $T \leq 8(1/\underline{\phi})\log(|S^{\text{fluid}}|T)$. Therefore, we have

$$\begin{aligned}
V^{\text{fluid}} - \Pi(\boldsymbol{q}^{\text{round}}) &\leq \bar{p}\left(\mathbb{E}\left\{X_0(\boldsymbol{q}^{\text{round}},T)\right\} - T\phi_0 + 1\right)\\
&\leq \bar{p}\left(\left(w_0 + \theta\sum_{i\in\mathcal{N}} w_i\right)\max_{i\in\mathcal{N}}\left\{\frac{q_i^{\text{round}}}{w_i}\right\}\log\left(\sum_{i\in\mathcal{N}} q_i^{\text{round}}\right) + \left[T - \sum_{i\in\mathcal{N}} q_i^{\text{round}}\right]^+ - T\phi_0\right) + 2\bar{p}\\
&\leq \bar{p}\left(\left(w_0 + \theta\sum_{i\in\mathcal{N}} w_i\right)\max_{i\in\mathcal{N}}\left\{\frac{\lceil q_i^{\text{fluid}}\rceil}{w_i}\right\}\log\left(\sum_{i\in S^{\text{fluid}}} q_i^{\text{fluid}} + 1\right) + T - \sum_{i\in S^{\text{fluid}}} q_i^{\text{fluid}} - T\phi_0\right) + 2\bar{p}\\
&= \bar{p}\left(w_0 + \theta\sum_{i\in\mathcal{N}} w_i\right)\max_{i\in\mathcal{N}}\left\{\frac{\lceil q_i^{\text{fluid}}\rceil}{w_i}\right\}\log\left(\sum_{i\in S^{\text{fluid}}} q_i^{\text{fluid}} + 1\right) + 2\bar{p}\\
&\leq \bar{p}\left(w_0 + \theta\sum_{i\in\mathcal{N}} w_i\right)\left(\frac{T}{w_0 + \theta\sum_{i\in\mathcal{N}} w_i + (1-\theta)\sum_{i\in S^{\text{fluid}}} w_i} + \frac{1}{\underline{w}}\right)\log(T+1) + 2\bar{p}\\
&= \bar{p}\frac{w_0 + \theta\sum_{i\in\mathcal{N}} w_i}{\underline{w}}\left(T\underline{\phi} + 1\right)\log(T+1) + 2\bar{p}\\
&\leq \bar{p}\kappa\left(8\log(|S^{\text{fluid}}|T)\right)\log(T+1) + 2\bar{p}\\
&\leq \bar{p}\kappa\left(8\log\left(T\left(T+1\right)\right) + 1\right)\log\left(1+T\right) + 2\bar{p}.
\end{aligned}$$

(ii) Otherwise, since $\phi_0/\underline{\phi} = (w_0 + \theta\sum_{i\in S^{\text{fluid}}} w_i)/\underline{w} \leq (w_0 + \theta\sum_{i\in\mathcal{N}} w_i)/\underline{w} \leq \kappa$, the assumptions in Lemma C.2, C.3, C.4 and C.5 holds and thus we can combine all results.

Similar to the proof of Theorem 5.1, there exist absolute constants C_5 and C_6 such that

$$V^{\text{fluid}} - \Pi\left(\boldsymbol{q}^{\text{round}}\right) \leq C_5\log T\sqrt{T\log T} + C_6.$$

■

Proof of Theorem D.3:

Suppose customers choosing under the nested logit model with one nest, under the assumption, for any set S, we have

$$\phi_0(S) = \frac{w_0}{w_0 + (\sum_{i\in S} w_i)^\alpha} \leq \frac{w_0}{w_0 + \alpha(\sum_{i\in S} w_i)}.$$

Similarly, following the argument in the proof of Lemma C.1, it implies that

$$\mathbb{E}\left\{X_0(\boldsymbol{q},T)\right\} \le 1 + \left(\frac{w_0}{\alpha}\right) \max_{i\in\mathcal{N}} \left\{\frac{q_i}{w_i}\right\} \log\left(\sum_{i\in\mathcal{N}} q_i\right) + \left[T - \sum_{i\in\mathcal{N}} q_i\right]^+ .$$

We consider the two possible cases: (i) If

$$T \le \frac{1}{\underline{\phi}} + \sqrt{2\frac{1}{\underline{\phi}} T \log(|S^{\text{fluid}}|T)},$$

then $T \le 8(1/\underline{\phi})\log(|S^{\text{fluid}}|T)$. Similarly, we have

$$\begin{aligned}
V^{\text{fluid}} - \Pi(\boldsymbol{q}^{\text{round}}) &\le \bar{p}\left(\mathbb{E}\left\{X_0(\boldsymbol{q}^{\text{round}},T)\right\} - T\phi_0 + 1\right) \\
&\le \bar{p}\left(\left(\frac{w_0}{\alpha}\right) \max_{i\in\mathcal{N}} \left\{\frac{q_i^{\text{round}}}{w_i}\right\} \log\left(\sum_{i\in\mathcal{N}} q_i^{\text{round}}\right) + \left[T - \sum_{i\in\mathcal{N}} q_i^{\text{round}}\right]^+ - T\phi_0\right) + 2\bar{p} \\
&\le \bar{p}\left(\left(\frac{w_0}{\alpha}\right) \max_{i\in\mathcal{N}} \left\{\frac{\lceil q_i^{\text{fluid}}\rceil}{w_i}\right\} \log\left(\sum_{i\in S^{\text{fluid}}} q_i^{\text{fluid}} + 1\right) + T - \sum_{i\in S^{\text{fluid}}} q_i^{\text{fluid}} - T\phi_0\right) + 2\bar{p} \\
&= \bar{p}\left(\frac{w_0}{\alpha}\right) \max_{i\in\mathcal{N}} \left\{\frac{\lceil q_i^{\text{fluid}}\rceil}{w_i}\right\} \log\left(\sum_{i\in S^{\text{fluid}}} q_i^{\text{fluid}} + 1\right) + 2\bar{p} \\
&\le \bar{p}\left(\frac{w_0}{\alpha}\right)\left(\frac{(\sum_{i\in S^{\text{fluid}}} w_i)^\alpha}{(\sum_{i\in S^{\text{fluid}}} w_i)(w_0 + \sum_{i\in S^{\text{fluid}}} w_i)} T + \frac{1}{\underline{w}}\right) \log\left(T+1\right) + 2\bar{p} \\
&= \bar{p}\left(\frac{w_0}{\alpha\underline{w}}\right)\left(T\underline{\phi} + 1\right)\log(T+1) + 2\bar{p} \\
&\le \bar{p}\kappa\left(8\log(|S^{\text{fluid}}|T)\right)\log(T+1) + 2\bar{p} \\
&\le \bar{p}\kappa\left(8\log\left(T\left(T+1\right)\right) + 1\right)\log\left(1+T\right) + 2\bar{p}.
\end{aligned}$$

(ii) Since

$$\frac{\phi_0(S)}{\phi_i(S)} = \frac{w_0}{w_0 + (\sum_{i\in S} w_i)^\alpha} \frac{w_0 + (\sum_{i\in S} w_i)^\alpha}{(\sum_{i\in S} w_i)^\alpha} \frac{\sum_{i\in S} w_i}{w_i} = \frac{\sum_{i\in S} w_i}{(\sum_{i\in S} w_i)^\alpha} \frac{w_0}{w_i} \le \frac{w_0}{\alpha\underline{w}} \le \kappa,$$

again Lemma C.2, C.3, C.4 and C.5 hold and we can combine all results together.

To conclude, there exist absolute constants C_6 and C_7 such that

$$V^{\text{fluid}} - \Pi\left(\boldsymbol{q}^{\text{round}}\right) \le C_6 \log T \sqrt{T\log T} + C_7.$$

■

LEMMA E.1. *Under the MNL model, there exists an optimal solution $\boldsymbol{\theta}^*$ to the capacitated fluid relaxation* (6) *such that there is at most one stockout time before time T.*

Proof: Let θ_0^* be an optimal solution, then we have $\sum_{i\in\mathcal{N}}\theta_i = T-\theta_0^*\le K$. Therefore, to find an optimal solutoin (θ_i^*), it is sufficient to consider the problem

$$\begin{aligned}\max\ &\sum_{i\in\mathcal{N}}(r_i-c_i)\theta_i\\ \text{s.t. }&0\le\theta_i\le w_i\theta_0^*,\\ &\sum_{i\in\mathcal{N}}\theta_i=T-\theta_0^*.\end{aligned}$$

The optimal solution is to choose the product with higher margin as much as possible, i.e., following the index order. Thus, there exists at most one product θ_i such that the first constraint does not hold, which implies that there exists at most one stockout time before T. ■

Let $\boldsymbol{q}^{\mathsf{fluid}}=\boldsymbol{\theta}^*$ denote the corresponding fluid inventory quantities, and $\boldsymbol{q}^{\mathsf{round}}$ denote the rounded solution following our scheme. In addition, we let $\tilde{\boldsymbol{q}}^{\mathsf{round}}$ denote the solution where $\tilde{q}_i^{\mathsf{round}}=q_i^{\mathsf{round}}$ for all $i\le k$ and

$$\tilde{q}_{k+1}^{\mathsf{round}}=\max\left\{0,\left\lfloor q_{k+1}^{\mathsf{fluid}}-\sqrt{2q_{k+1}^{\mathsf{fluid}}\log((k+1)T)}\right\rfloor\right\}.$$

Note that we have $\tilde{q}_{k+1}^{\mathsf{round}}\le q_{k+1}^{\mathsf{round}}$. For simplicity, let $\Delta T=T-T_1$.

Lemma E.2. *Suppose* $\tilde{T}_1=\left\lfloor T_1-\frac{1}{\underline{\phi}}-\sqrt{2\frac{1}{\underline{\phi}}T_1\log((k+1)T_1)}\right\rfloor\ge 0$*, then it holds that*

$$\mathbb{P}\left(X_i(\boldsymbol{q}^{\mathsf{round}},\tilde{T}_1)<\lfloor\phi_i(S_1)T_1\rfloor,\forall i\le k+1\right)\ge 1-\frac{1}{T_1}.$$

Proof. Note that when all products are available, the item purchased by a customer within $\tilde{T}_1$ periods follows a multinomial distribution $(X_0,X_1,\dots,X_{k+1})\sim(\tilde{T}_1,\phi_0(S_1),\dots,\phi_{k+1}(S_1))$. Similar to the proof of Lemma C.2, we have

$$\begin{aligned}&\mathbb{P}\left(X_i(\boldsymbol{q}^{\mathsf{round}},\tilde{T}_1)<\lfloor\phi_i(S_1)T_1\rfloor,\forall i\le k+1\right)\\ &=\mathbb{P}(X_i<\lfloor\phi_i(S_1)T_1\rfloor,\forall i\le k+1)=1-\mathbb{P}\left(\exists i\le k+1,X_i\ge\lfloor\phi_i(S_1)T_1\rfloor\right)\\ &\ge 1-\sum_{i\le k+1}\mathbb{P}\left(X_i\ge\lfloor\phi_i(S_1)T_1\rfloor\right)\ge 1-\sum_{i\le k+1}\mathbb{P}\left(X_i\ge\lfloor\phi_i(S_1)T_1\rfloor\right)\\ &\ge 1-\sum_{i\le k+1}\mathbb{P}\left(X_i\ge\phi_i(S_1)T_1-1\right)\ge 1-\frac{1}{T_1}.\end{aligned}$$

■

Lemma E.3. *Suppose* $\tilde{T}=\tilde{T}_1+\left\lfloor\Delta T-\frac{1}{\underline{\phi}}-\sqrt{2\frac{1}{\underline{\phi}}\Delta T\log(k\Delta T)}\right\rfloor\ge\tilde{T}_1$*, then it holds that*

$$\mathbb{P}\left(X_i(\boldsymbol{q}^{\mathsf{round}},\tilde{T})-X_i(\boldsymbol{q}^{\mathsf{round}},\tilde{T}_1)<\lfloor\phi_i(S_2)\Delta T\rfloor,\forall i\le k\,\middle|\,X_i(\boldsymbol{q}^{\mathsf{round}},\tilde{T}_1)\le\lfloor\phi_i(S_1)T_1\rfloor,\forall i\le k+1\right)\ge 1-\frac{1}{\Delta T}.$$

Proof. Note that for any product $i \le k$, the purchase number after time $\tilde{T}_1$ is decreasing with the number of product $k+1$ left at time $\tilde{T}_1$. In addition, when the product $k+1$ is sold out before time $\tilde{T}_1$ and all products $i \le k$ are available bewteen time $\tilde{T}_1 + 1$ and $\tilde{T}$, the item purchased by a customer again follows a multinomial distribution $(X_0, X_1, \ldots, X_k) \sim (\tilde{T} - \tilde{T}_1, \phi_0(S_2), \ldots, \phi_k(S_2))$. Therefore, we have

$$\begin{aligned}
&\mathbb{P}\left(X_i(\boldsymbol{q}^{\text{round}}, \tilde{T}) - X_i(\boldsymbol{q}^{\text{round}}, \tilde{T}_1) < \lfloor \phi_i(S_2)\Delta T \rfloor, \forall i \le k \,\middle|\, X_i(\boldsymbol{q}^{\text{round}}, \tilde{T}_1) \le \lfloor \phi_i(S_1) T_1 \rfloor, \forall i \le k+1\right) \\
&\ge \mathbb{P}\left(X_i < \lfloor \phi_i(S_2)\Delta T \rfloor, \forall i \le k\right) \le 1 - \sum_{i \le k} \mathbb{P}\left(X_i \ge \lfloor \phi_i(S_2)\Delta T \rfloor\right).
\end{aligned}$$

For any $i \le k$, since X_i is a binomial random variable, we have

$$\begin{aligned}
&\mathbb{P}\left(X_i \ge \lfloor \phi_i(S_2)\Delta T \rfloor\right) \le \mathbb{P}\left(X_i \ge \phi_i(S_2)\Delta - 1\right) \\
&= \mathbb{P}\left(X_i \ge \phi_i(S_2)\Delta T\left(1 + \frac{\Delta T - \Delta\tilde{T}}{\Delta\tilde{T}} - \frac{1}{\phi_i(S_2)\Delta T}\right)\right) \le \frac{1}{k(T - T_1)},
\end{aligned}$$

where $\Delta\tilde{T} = \tilde{T} - \tilde{T}_1$. ■

Lemma E.4. *The expected number of no-purchase till time $\tilde{T}$ is bounded by*

$$\mathbb{E}\left\{X_0(\boldsymbol{q}^{\text{round}}, \tilde{T})\right\} \le \phi_0(S_1)T_1 + \phi_0(S_2)(T - T_1).$$

Moreover, there exist absolute constants D_5 and D_6 such that

$$\mathbb{E}\left\{\left[\phi_0(S_1)T_1 + \phi_0(S_2)\Delta T - X_0(\boldsymbol{q}^{\text{round}}, \tilde{T})\right]^+\right\} \le D_5\sqrt{T\log T} + D_6.$$

Proof. We consider the number of no-purchase in the two periods before $\tilde{T}_1$ and between $\tilde{T}_1 + 1$ and $\tilde{T}$. Note that

$$\begin{aligned}
&\mathbb{E}\left\{X_0(\boldsymbol{q}^{\text{round}}, \tilde{T}_1) \,\middle|\, X_i(\boldsymbol{q}^{\text{round}}, \tilde{T}_1) < q_i^{\text{round}}, \forall i \le k+1\right\} \\
&= \mathbb{E}\left\{\tilde{T}_1 - \sum_{i \le k+1} X_i(\boldsymbol{q}^{\text{round}}, \tilde{T}_1) \,\middle|\, X_i(\boldsymbol{q}^{\text{round}}, \tilde{T}_1) < q_i^{\text{round}}, \forall i \le k+1\right\} \\
&= \mathbb{E}\left\{\tilde{T}_1 - \sum_{i \le k+1} X_i \,\middle|\, X_i < q_i^{\text{round}}, \forall i \le k+1\right\} = \mathbb{E}\left\{X_0 \,\middle|\, X_i < q_i^{\text{round}}, \forall i \le k+1\right\} \\
&\le \frac{\mathbb{E}\left\{X_0\right\}}{\mathbb{P}(X_i < q_i^{\text{round}}, \forall i \le k+1)} \le \frac{T_1}{T_1 - 1}\phi_0(S_1)\tilde{T}_1 \le \phi_0(S_1)T_1,
\end{aligned}$$

where $(X_0, \ldots, X_{k+1}) \sim (\tilde{T}_1, \phi_0(S_1), \ldots, \phi_{k+1}(S_1))$. In addition,

$$\begin{aligned}
&\mathbb{E}\left\{\left[\phi_0(S_1)T_1 - X_0(\boldsymbol{q}^{\text{round}}, \tilde{T}_1)\right]^+ \middle| X_i(\boldsymbol{q}^{\text{round}}, \tilde{T}_1) < q_i^{\text{round}}, \forall i \le k+1\right\} \\
&= \mathbb{E}\left\{[\phi_0(S_1)T_1 - X_0]^+ \middle| X_i < q_i^{\text{round}}, \forall i \le k+1\right\} \le \phi_0(S_1)(T_1 - \tilde{T}_1) + \sqrt{\mathbb{V}\text{ar}[X_0]} \\
&\le \phi_0(S_1)\left(1 + \frac{1}{\underline{\phi}} + \sqrt{2\frac{1}{\underline{\phi}}T_1 \log((k+1)T)}\right) + \sqrt{\phi_0(S_1)\tilde{T}_1} \\
&\le 1 + \kappa + \sqrt{2\kappa T \log((k+1)T)} + \sqrt{T}.
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
&\mathbb{E}\left\{X_0(\boldsymbol{q}^{\text{round}}, \tilde{T}) - X_0(\boldsymbol{q}^{\text{round}}, \tilde{T}_1) \middle| X_i(\boldsymbol{q}^{\text{round}}, \tilde{T}) < q_i^{\text{round}}, \forall i \le k, X_{k+1}(\boldsymbol{q}^{\text{round}}, \tilde{T}_1) < q_{k+1}^{\text{round}}\right\} \\
&\le \frac{\Delta T}{\Delta T - 1}\phi_0(S_2)(\tilde{T} - \tilde{T}_1) \le \Delta T \phi_0(S_2).
\end{aligned}$$

Moreover,

$$\begin{aligned}
&\mathbb{E}\left\{\left[\phi_0(S_2)\Delta T - \left(X_0(\boldsymbol{q}^{\text{round}}, \tilde{T}) - X_0(\boldsymbol{q}^{\text{round}}, \tilde{T}_1)\right)\right]^+ \middle| X_i(\boldsymbol{q}^{\text{round}}, \tilde{T}) < q_i^{\text{round}}, \forall i \le k, X_{k+1}(\boldsymbol{q}^{\text{round}}, \tilde{T}_1) < q_{k+1}^{\text{round}}\right\} \\
&\le 1 + \kappa + \sqrt{2\kappa T \log(kT)} + \sqrt{T}.
\end{aligned}$$

Therefore, we can conclude that

$$\mathbb{E}\left\{X_0(\boldsymbol{q}^{\text{round}}, \tilde{T})\right\} \le \phi_0(S_1)T_1 + \phi_0(S_2)(T - T_1),$$

and

$$\mathbb{E}\left\{\left[\phi_0(S_1)T_1 + \phi_0(S_2)\Delta T - X_0(\boldsymbol{q}^{\text{round}}, \tilde{T})\right]^+\right\} \le D_5\sqrt{T \log T} + D_6.$$

■

LEMMA E.5. *Starting with initial inventory $\tilde{\boldsymbol{q}}^{\text{round}}$, it holds that*

$$\mathbb{P}\left(X_{k+1}(\tilde{\boldsymbol{q}}^{\text{round}}, T_1) = \tilde{q}_{k+1}^{\text{round}}\right) \ge 1 - \frac{1}{(k+1)T_1}.$$

Proof. Note that when the product $k+1$ is available, the probability that a customer will purchase it is at least $\phi_{k+1}(S_1)$, therefore,

$$X_{k+1}(\tilde{\boldsymbol{q}}^{\text{round}}, T_1) \ge_{\text{st}} \min\left\{\tilde{q}_{k+1}^{\text{round}}, \sum_{t=1}^{T_1} \omega_{t,k+1}\right\},$$

where $\omega_{t,k+1}$ is a binomial random variable with success probability $\phi_{k+1}(S_1)$. Thus,

$$\begin{aligned}
&\mathbb{P}\left(X_{k+1}(\tilde{\boldsymbol{q}}^{\mathsf{round}},T_1)<\tilde{q}_{k+1}^{\mathsf{round}}\right)\le\mathbb{P}\left(\min\left\{\tilde{q}_{k+1}^{\mathsf{round}},\sum_{t=1}^{T_1}\omega_{t,k+1}\right\}<\tilde{q}_{k+1}^{\mathsf{round}}\right)=\mathbb{P}\left(\sum_{t=1}^{T_1}\omega_{t,k+1}<\tilde{q}_{k+1}^{\mathsf{round}}\right)\\
&=\mathbb{P}\left(\sum_{t=1}^{T_1}\omega_{t,k+1}<\lfloor q_{k+1}^{\mathsf{round}}-\sqrt{2q_{k+1}^{\mathsf{fluid}}\log(k+1)T_1}\rfloor\right)\\
&\le\mathbb{P}\left(\sum_{t=1}^{T_1}\omega_{t,k+1}\le q_{k+1}^{\mathsf{round}}-\sqrt{2q_{k+1}^{\mathsf{fluid}}\log((k+1)T_1)}\right)\\
&\le\mathbb{P}\left(\sum_{t=1}^{T_1}\omega_{t,k+1}\le q_{k+1}^{\mathsf{fluid}}-\sqrt{2q_{k+1}^{\mathsf{fluid}}\log((k+1)T_1)}\right)\\
&\le\frac{1}{(k+1)T_1}.
\end{aligned}$$

■

Lemma E.6. *Using the adjusted rounded inventory $\tilde{\boldsymbol{q}}^{\mathsf{round}}$, there exist absolute constants D_7 and D_8 such that*

$$\mathbb{E}\left[\max_{i\le k}\left\{\frac{w_0}{w_i}\left(q_i^{\mathsf{round}}-X_i(\tilde{\boldsymbol{q}}^{\mathsf{round}},\tilde{T})\right)\right\}\right]\le D_7\sqrt{T\log T}+D_8.$$

Proof. Suppose $\tilde{T}\ge T_1$, note that given the product $k+1$ is sold out before time T_1, for any product $i\le k$, the purchase probability is at least $\phi_i(S_1)$ before time T_1 and at least $\phi_i(S_2)$ after time T_1, which implies that

$$X_i(\tilde{\boldsymbol{q}}^{\mathsf{round}},\tilde{T})\ge_{\mathrm{st}}\min\left\{q_i^{\mathsf{round}},\sum_{t=1}^{T_1}\omega_{t,i}+\sum_{t=T_1+1}^{\tilde{T}}\omega_{t,i}\right\},$$

where $\omega_{t,i}$ follows a Bernoulli distribution with success probability $\phi_i(S_1)$ before time T_1 and $\phi_i(S_2)$ after T_1. Therefore,

$$\begin{aligned}
&\mathbb{E}\left[\max_{i\le k}\left\{\frac{w_0}{w_i}\left(q_i^{\mathsf{round}}-X_i(\tilde{\boldsymbol{q}}^{\mathsf{round}},\tilde{T})\right)\right\}\middle|\tilde{q}_{k+1}^{\mathsf{round}}=X_{k+1}(\tilde{\boldsymbol{q}},T_1)\right]\\
&\le\mathbb{E}\left[\max_{i\le k}\left\{\frac{w_0}{w_i}\left(q_i^{\mathsf{round}}-\sum_{t=1}^{T_1}X_{t,i}-\sum_{t=T_1+1}^{\tilde{T}}X_{t,i}\right)^+\right\}\right]\le\mathbb{E}\left[\max_{i\le k}\left\{\frac{w_0}{w_i}\left(1+q_i^{\mathsf{fluid}}-\sum_{t=1}^{T_1}X_{t,i}-\sum_{t=T_1+1}^{\tilde{T}}X_{t,i}\right)\right\}\right]\\
&\le\mathbb{E}\left[\max_{i\le k}\left\{\frac{w_0}{w_i}\left(1+\phi_i(S_2)(T-\tilde{T})+\left(\phi_i(S_1)T_1+\phi_i(S_2)(\tilde{T}-T_1)-\sum_{t=1}^{T_1}X_{t,i}-\sum_{t=T_1+1}^{\tilde{T}}X_{t,i}\right)^+\right)\right\}\right]\\
&\le\max_{i\le k}\left\{\frac{w_0}{w_i}\left(1+\phi_i(S_2)(T-\tilde{T})\right)\right\}+\mathbb{E}\left[\max_{i\le k}\left\{\frac{w_0}{w_i}\left(\phi_i(S_1)T_1+\phi_i(S_2)(\tilde{T}-T_1)-\sum_{t=1}^{T_1}X_{t,i}-\sum_{t=T_1+1}^{\tilde{T}}X_{t,i}\right)^+\right\}\right]
\end{aligned}$$

$$
\begin{aligned}
&\leq \max_{i\leq k}\left(\frac{w_0}{w_i}\left(1+\phi_i(S_2)(T-\tilde{T})\right)\right)+\sqrt{1+k}\left(\max_{i\leq k}\sqrt{\frac{w_0}{w_i}\left(\phi_i(S_1)T_1+\phi_i(S_2)(\tilde{T}-T_1)\right)}\right)\\
&\leq \frac{w_0}{\underline{w}}+\frac{w_0}{w_0+\sum_{i\leq k}w_i}(T-\tilde{T})+\sqrt{1+k}\sqrt{\frac{w_0}{w_0+\sum_{i\leq k+1}w_i}T_1+\frac{w_0}{w_0+\sum_{i\leq k}w_i}(\tilde{T}-T_1)}\\
&\leq \kappa+\frac{\kappa}{k}(T-\tilde{T})+\sqrt{1+k}\sqrt{\frac{\kappa}{k+1}T_1+\frac{\kappa}{k}(\tilde{T}-T_1)}\\
&\leq \kappa+\frac{\kappa}{k}(T-\tilde{T})+\sqrt{2\kappa T_1+2\kappa(\tilde{T}-T_1)}\\
&\leq \kappa+\frac{\kappa}{k}\left(2+\frac{2}{\underline{\phi}}+\sqrt{\frac{2}{\underline{\phi}}T_1\log((k+1)T_1)}+\sqrt{\frac{2}{\underline{\phi}}(T-T_1)\log(k(T-T_1))}\right)+\sqrt{2\kappa T}\\
&\leq 3\kappa+2\kappa^2+2\kappa^{\frac{3}{2}}\sqrt{T\log((k+1)T)}+\sqrt{2\kappa T}.
\end{aligned}
$$

Similarly, suppose $\tilde{T}<T_1$, we can show that

$$
\begin{aligned}
&\mathbb{E}\left[\max_{i\leq k}\left\{\frac{w_0}{w_i}\left(q_i^{\text{round}}-X_i(\tilde{\boldsymbol{q}}^{\text{round}},\tilde{T})\right)\right\}\middle|\tilde{\boldsymbol{q}}_{k+1}^{\text{round}}=X_{k+1}(\tilde{\boldsymbol{q}},T_1)\right]\\
&\quad\leq \max_{i\leq k}\left\{\frac{w_0}{w_i}\left(1+\phi_i(S_1)(T_1-\tilde{T})+\phi_i(S_2)(T-T_1)\right)\right\}+\sqrt{2\kappa T}\\
&\quad\leq \max_{i\leq k}\left\{\frac{w_0}{w_i}\left(1+\phi_i(S_2)(T-\tilde{T})\right)\right\}+\sqrt{2\kappa T}\\
&\quad\leq 3\kappa+2\kappa^2+2\kappa^{\frac{3}{2}}\sqrt{T\log((k+1)T)}+\sqrt{2\kappa T}.
\end{aligned}
$$

Therefore, we can conclude that

$$
\begin{aligned}
&\mathbb{E}\left[\max_{i\leq k}\left\{\frac{w_0}{w_i}\left(q_i^{\text{round}}-X_i(\tilde{\boldsymbol{q}}^{\text{round}},\tilde{T})\right)\right\}\right]\\
&\quad\leq \mathbb{E}\left[\max_{i\leq k}\left\{\frac{w_0}{w_i}\left(q_i^{\text{round}}-X_i(\tilde{\boldsymbol{q}}^{\text{round}},\tilde{T})\right)\right\}\middle|\tilde{q}_{k+1}^{\text{round}}=X_{k+1}^{\text{round}}(\tilde{\boldsymbol{q}}^{\text{round}},T_1)\right]+T\mathbb{P}\left(\tilde{q}_{k+1}^{\text{round}}<X_{k+1}^{\text{round}}(\tilde{\boldsymbol{q}}^{\text{round}},T_1)\right)\\
&\quad\leq D_7\sqrt{T\log(T)}+D_8.
\end{aligned}
$$

■

Lemma E.7 (**Honhon and Seshadri 2013**). *For any $\varepsilon\in\mathbb{N}$, $\boldsymbol{q}\in\mathbb{Z}_+^N$ and T, it holds that for any product j and any sample path,*

$$
\sum_{i\neq j}\left(X_i(\boldsymbol{q}-\varepsilon\boldsymbol{e}_j,T)-X_i(\boldsymbol{q},T)\right)\leq\varepsilon,\quad \sum_{i\in\mathcal{N}}\left(X_i(\boldsymbol{q}-\varepsilon\boldsymbol{e}_j,T)-X_i(\boldsymbol{q},T)\right)\leq 0.
$$

Lemma E.8. *There exist absolute constants D_9 and D_{10} such that*

$$
\mathbb{E}\left[\max_{i\leq k}\left\{\frac{w_0}{w_i}\left(q_i^{\text{round}}-X_i(\boldsymbol{q}^{\text{round}},\tilde{T})\right)\right\}\right]\leq D_9\sqrt{T\log T}+D_{10}.
$$

Proof. By Lemma E.7, for any sample path,

$$X_i(\tilde{\boldsymbol{q}}^{\mathsf{round}},\tilde{T}) - X_i(\boldsymbol{q}^{\mathsf{round}},\tilde{T}) \le \sum_{i\neq k+1}\left(X_i(\tilde{\boldsymbol{q}}^{\mathsf{round}},\tilde{T}) - X_i(\boldsymbol{q}^{\mathsf{round}},\tilde{T})\right) \le q_{k+1}^{\mathsf{round}} - \tilde{q}_{k+1}^{\mathsf{round}}.$$

Therefore, we have

$$\begin{aligned}
&\mathbb{E}\left[\max_{i\le k}\left\{\frac{w_0}{w_i}\left(q_i^{\mathsf{round}} - X_i(\boldsymbol{q}^{\mathsf{round}},\tilde{T})\right)\right\}\right]\\
&\le \mathbb{E}\left[\max_{i\le k}\left\{\frac{w_0}{w_i}\left(q_i^{\mathsf{round}} - X_i(\tilde{\boldsymbol{q}}^{\mathsf{round}},\tilde{T})\right)\right\}\right] + \mathbb{E}\left\{\max_{i\le k}\left\{\frac{w_0}{w_i}\left(X_i(\tilde{\boldsymbol{q}}^{\mathsf{round}},\tilde{T}) - X_i(\boldsymbol{q}^{\mathsf{round}},\tilde{T})\right)\right\}\right\}\\
&\le D_7 + D_8\sqrt{T\log(T)} + \frac{w_0}{\underline{w}}\left(q_{k+1}^{\mathsf{round}} - \tilde{q}_{k+1}^{\mathsf{round}}\right)\\
&\le D_7 + D_8\log(T\log(T)) + \kappa\left(\lfloor q_{k+1}^{\mathsf{fluid}}\rfloor - \left\lfloor q_{k+1}^{\mathsf{fluid}} - \sqrt{2q_{k+1}^{\mathsf{fluid}}\log((k+1)T)}\right\rfloor\right)\\
&\le D_7 + D_8\sqrt{T\log(T)} + \kappa + \kappa\sqrt{2T\log((k+1)T)}\\
&\le D_9\sqrt{T\log(T)} + D_{10}.
\end{aligned}$$

■

<u>Proof of Theorem D.4</u>:

We consider the three possible cases: (i) Suppose

$$T_1 \le \frac{1}{\underline{\phi}} + \sqrt{2\frac{1}{\underline{\phi}}T_1\log((k+1)T_1)},$$

which implies that $T_1 \le 8\frac{1}{\underline{\phi}}\log((k+1)T_1)$. In this case, we consider the fluid solution $\hat{\boldsymbol{q}}^{\mathsf{fluid}} = \boldsymbol{q}^{\mathsf{fluid}} - q_{k+1}^{\mathsf{fluid}}\boldsymbol{e}_{k+1}$, which only offers the set S_2. This inventory quantity $\hat{\boldsymbol{q}}^{\mathsf{fluid}}$ will last for

$$T - T_1 + \frac{\phi_i(S_1)}{\phi_i(S_2)}T_1 = T - T_1 + \frac{w_0+\sum_{i\le k}w_i}{w_0+\sum_{i\le k+1}w_i}T_1 = T - \frac{w_{k+1}}{w_0+\sum_{i\le k+1}w_i}T_1 = T - \phi_{k+1}(S_1)T_1.$$

Let $\hat{\boldsymbol{q}}^{\mathsf{round}}$ denote the rounded solution for $\hat{\boldsymbol{q}}^{\mathsf{fluid}}$, by Theorem 5.1, there exist absolute constants C_2 and C_3 such that

$$\begin{aligned}
V^{\mathsf{fluid}} - \Pi(\hat{\boldsymbol{q}}^{\mathsf{round}}) &\le (p_{k+1} - c_{k+1})q_{k+1}^{\mathsf{fluid}} + \phi_{k+1}(S_1)T_1 + C_2\log T\sqrt{T\log T} + C_3\\
&\le (\bar{p}+1)q_{k+1}^{\mathsf{fluid}} + C_2\log T\sqrt{T\log T} + C_3\\
&\le 16(\bar{p}+1)\kappa\log(T_1) + C_2\log T\sqrt{T\log T} + C_3.
\end{aligned}$$

(ii) Suppose

$$\Delta T \le \frac{1}{\underline{\phi}} + \sqrt{2\frac{1}{\underline{\phi}}\Delta T\log(k\Delta T)},$$

which implies that $\Delta T \le 8\frac{1}{\underline{\phi}}\log(k\Delta T)$. In this case, we consider the fluid solution $\hat{\boldsymbol{q}}^{\text{fluid}}$ such that $\hat{q}_i^{\text{fluid}} = \phi_i(S_1)T$, then

$$\hat{q}_i^{\text{fluid}} - q_i^{\text{fluid}} = (\phi_i(S_1) - \phi_i(S_2))\Delta T < 0, \forall i \le k, \quad \hat{q}_{k+1}^{\text{fluid}} - q_{k+1}^{\text{fluid}} = \phi_{k+1}(S_1)\Delta T.$$

Again, let $\hat{\boldsymbol{q}}^{\text{round}}$ denote the rounded inventory. However, such inventory may violate the capacity constraint. Therefore, we first consider $\hat{\boldsymbol{q}}' = \hat{\boldsymbol{q}}^{\text{round}} - \lceil \phi_{k+1}(S_1)\Delta T \rceil \boldsymbol{e}_{k+1}$, which has at most $K+1$ total inventories because

$$\hat{q}_{k+1}^{\text{round}} = \lfloor \phi_{k+1}(S_1)T \rfloor = \lfloor \phi_{k+1}(S_1)T_1 + \phi_{k+1}(S_1)\Delta T \rfloor \ge \lfloor 1 + \phi_{k+1}(S_1)\Delta T \rfloor \ge \lceil \phi_{k+1}(S_1)\Delta T \rceil,$$

and

$$\sum_{i\in\mathcal{N}} \hat{q}_i' = \sum_{i\in\mathcal{N}} \hat{q}_i^{\text{round}} - \lceil \phi_{k+1}(S_1)\Delta T \rceil \le \sum_{i\in\mathcal{N}} \hat{q}_i^{\text{fluid}} + 1 - \lceil \phi_{k+1}(S_1)\Delta T \rceil \le \sum_{i\in\mathcal{N}} q_i^{\text{fluid}} + 1 \le K+1.$$

We then remove one unit inventory of the product with smallest revenue, and let $\boldsymbol{q}^{\text{fea}}$ denote the resulting inventory, which is always feasible. Finally, we have

$$\begin{aligned}
&V^{\text{fluid}} - \Pi(\boldsymbol{q}^{\text{fea}}) = V^{\text{fluid}} - \Pi(\hat{\boldsymbol{q}}') + \bar{p} \le \sum_{i\in\mathcal{N}} (p_i - c_i) q_i^{\text{fluid}} - \sum_{i\in\mathcal{N}} p_i \mathbb{E}\{X_i(\hat{\boldsymbol{q}}', T)\} + \sum_{i\in\mathcal{N}} c_i \hat{q}_i' + \bar{p} \\
&\le \sum_{i\in\mathcal{N}} (p_i - c_i)(q_i^{\text{fluid}} - \hat{q}_i^{\text{fluid}}) + \sum_{i\in\mathcal{N}} p_i \mathbb{E}\left\{X_i(\hat{\boldsymbol{q}}^{\text{round}}, T) - X_i(\hat{\boldsymbol{q}}', T)\right\} + \sum_{i\in\mathcal{N}} c_i (\hat{q}_i' - \hat{q}_i^{\text{round}}) + \bar{p} \\
&\quad + \sum_{i\in\mathcal{N}} (p_i - c_i)\hat{q}_i^{\text{fluid}} - \sum_{i\in\mathcal{N}} p_i \mathbb{E}\left\{X_i(\hat{\boldsymbol{q}}^{\text{round}}, T)\right\} + \sum_{i\in\mathcal{N}} c_i \hat{q}_i^{\text{round}} \\
&\le \bar{p} \sum_{i\le k} (\phi_i(S_2) - \phi_i(S_1))\,\Delta T + \bar{p}(1 + \phi_{k+1}(S_1)\Delta T) + \bar{p} \\
&\quad + \sum_{i\in\mathcal{N}} (p_i - c_i)\hat{q}_i^{\text{fluid}} - \sum_{i\in\mathcal{N}} p_i \mathbb{E}\left\{X_i(\hat{\boldsymbol{q}}^{\text{round}}, T)\right\} + \sum_{i\in\mathcal{N}} c_i \hat{q}_i^{\text{round}} \\
&= 2\bar{p}(1 + \phi_{k+1}(S_1)\Delta T) + \sum_{i\in\mathcal{N}} (p_i - c_i)\hat{q}_i^{\text{fluid}} - \sum_{i\in\mathcal{N}} p_i \mathbb{E}\left\{X_i(\hat{\boldsymbol{q}}^{\text{round}}, T)\right\} + \sum_{i\in\mathcal{N}} c_i \hat{q}_i^{\text{round}} \\
&\le 2\bar{p}(1 + 16\kappa \log(T)) + C_2 \log T \sqrt{T \log T} + C_3.
\end{aligned}$$

We decompose the performance gap into two parts:

$$\begin{aligned}
&V^{\text{fluid}} - \Pi(\boldsymbol{q}^{\text{fea}}) = V^{\text{fluid}} - \Pi(\boldsymbol{q}^{\text{round}}) + \Pi(\boldsymbol{q}^{\text{round}}) - \Pi(\boldsymbol{q}^{\text{fea}}) \\
&\le V^{\text{fluid}} - \Pi(\boldsymbol{q}^{\text{round}}) + \sum_{i\in\mathcal{N}} p_i \mathbb{E}\left\{X_i(\boldsymbol{q}^{\text{round}}, T) - X_i(\boldsymbol{q}^{\text{fea}}, T)\right\} \le V^{\text{fluid}} - \Pi(\boldsymbol{q}^{\text{round}}) + \bar{p} \\
&\le \bar{p}\left(\mathbb{E}\left\{X_0(\boldsymbol{q}^{\text{round}}, T) - T_1\phi_0(S_1) - (T - T_1)\phi_0(S_2)\right\}\right) + 2\bar{p}
\end{aligned}$$

$$\leq \bar{p}\left(\underbrace{\mathbb{E}\left\{X_0(\boldsymbol{q}^{\text{round}},\tilde{T})-T_1\phi_0(S_1)-(T-T_1)\phi_0(S_2)\right\}}_{(a)}+\underbrace{\mathbb{E}\left\{X_0(\boldsymbol{q}^{\text{round}},T)-X_0(\boldsymbol{q}^{\text{round}},\tilde{T})\right\}}_{(b)}\right)+2\bar{p},$$

where the second inequality follows from the fact that $\boldsymbol{q}^{\text{round}}$ and $\boldsymbol{q}^{\text{fea}}$ differs at most one unit of inventory.

(iii) Suppose both (i) and (ii) are not satisfied, by Lemma E.4, we have term (a) is less than zero. For term (b), we have

$$\begin{aligned}
&\mathbb{E}\left\{X_0(\boldsymbol{q}^{\text{round}},T)-X_0(\boldsymbol{q}^{\text{round}},\tilde{T})\right\}\\
&\leq \mathbb{E}\left\{\max_{i\leq k}\left\{\frac{w_0}{w_i}\left(q_i^{\text{round}}-X_i(\boldsymbol{q}^{\text{round}},\tilde{T})\right)\right\}\right\}\log T+\mathbb{E}\left\{\left[T-\tilde{T}-\sum_{i\in\mathcal{N}}\left(q_i^{\text{round}}-X_i(\boldsymbol{q}^{\text{round}},\tilde{T})\right)\right]^+\right\}\\
&\leq 1+\mathbb{E}\left\{\max_{i\leq k}\left\{\frac{w_0}{w_i}\left(q_i^{\text{round}}-X_i(\boldsymbol{q}^{\text{round}},\tilde{T})\right)\right\}\right\}\log T+\mathbb{E}\left\{\left[\phi_0(S_1)T_1+\phi_0(S_2)\Delta T-X_0(\boldsymbol{q}^{\text{round}},\tilde{T})\right]^+\right\}.
\end{aligned}$$

By Lemma E.8 and E.4, we can conclude that there exist absolute constants C_9 and C_{10} such that

$$V^{\text{fluid}}-\Pi(\boldsymbol{q}^{\text{fea}})\leq C_9\log T\sqrt{T\log T}+C_{10}.$$

■

Appendix F: Symmetric MNL Model

Here we consider the symmetric MNL model in which all products are symmetric, i.e., they share a same price p, cost c and a normalized weight $w=1$. Note that the total expected revenue by selling the products can be described using a recursion. Let $V_t(\boldsymbol{q})$ denote the expected revenue at time t with stocking quantities $\boldsymbol{q}$, then it holds that

$$V_t(\boldsymbol{q})=\sum_{i=1}^{N}\phi_i(A(\boldsymbol{q}))\left(1+V_{t+1}(\boldsymbol{q}-\boldsymbol{e}_i)\right)+\phi_0(A(\boldsymbol{q}))V_{t+1}(\boldsymbol{q}),$$

where $V_{T+1}(\boldsymbol{q})=0$ and $\phi_i(A(\boldsymbol{q}))=1/(w_0+|A(\boldsymbol{q})|)\mathbf{1}(i\in A(\boldsymbol{q}))$.

In what follows, we provide some structure properties for the value functions.

Proposition F.1. For any time t and stocking quantities $\boldsymbol{q}$, the following hold:

(a) For any product j such that $q_j>0$, $V_t(\boldsymbol{q})\leq 1+V_t(\boldsymbol{q}-\boldsymbol{e}_j)$.

(b) For any products j and k such that $q_j-q_k\geq 2$, $V_t(\boldsymbol{q}+\boldsymbol{e}_k-\boldsymbol{e}_j)\geq V_t(\boldsymbol{q})$.

Proposition F.1 (b) implies that it is optimal to split total inventories “equally” across all products. Therefore, in order to find the optimal stocking quantities for all products, it is sufficient to find the total inventory and the problem is then reduced into a single dimension optimization problem, which can be solved by binary search. However, Proposition F.1 (a), the key step to prove Proposition F.1 (b), may not hold for choice models other than MNL, thus whether “evenly split” is optimal for other symmetric models remains open. We provide the proof for Proposition F.1 below.

LEMMA F.1. *For any time t and inventory $\boldsymbol{q}$, we have*

$$V_t(\boldsymbol{q}) = V_t(\sigma(\boldsymbol{q})),$$

where $\sigma(\boldsymbol{q})$ is any permutation of inventory $\boldsymbol{q}$ over items.

Proof of Proposition F.1:

(a) We show the result by induction. For simplicity, we let $|\boldsymbol{q}| = |A(\boldsymbol{q})|$. It holds for time T because $V_T(\boldsymbol{q}) = \frac{|\boldsymbol{q}|}{w_0+|\boldsymbol{q}|} \le 1$. Now assume the statement holds for any $t+1$ and inventory $\boldsymbol{q}$, we consider two possible cases:

(i) If $q_j \ge 2$, then we have

$$\begin{aligned}
&V_t(\boldsymbol{q})\\
=&\frac{|\boldsymbol{q}|}{w_0+|\boldsymbol{q}|} + \sum_{i\in A(\boldsymbol{q})} \frac{1}{w_0+|\boldsymbol{q}|} V_{t+1}(\boldsymbol{q}-\boldsymbol{e}_i) + \frac{w_0}{w_0+|\boldsymbol{q}|} V_{t+1}(\boldsymbol{q})\\
\le&\frac{|\boldsymbol{q}|}{w_0+|\boldsymbol{q}|} + \sum_{i\in A(\boldsymbol{q})} \frac{1}{w_0+|\boldsymbol{q}|} \left(1+V_{t+2}(\boldsymbol{q}-\boldsymbol{e}_i-\boldsymbol{e}_j)\right) + \frac{w_0}{w_0+|\boldsymbol{q}|} \left(1+V_{t+2}(\boldsymbol{q}-\boldsymbol{e}_j)\right)\\
=&\frac{|\boldsymbol{q}|}{w_0+|\boldsymbol{q}|} + \frac{w_0}{w_0+|\boldsymbol{q}|} + \left(\sum_{i\in A(\boldsymbol{q})} \frac{1}{w_0+|\boldsymbol{q}|}(1+V_{t+2}(\boldsymbol{q}-\boldsymbol{e}_i-\boldsymbol{e}_j)) + \frac{w_0}{w_0+|\boldsymbol{q}|}V_{t+2}(\boldsymbol{q}-\boldsymbol{e}_j)\right)\\
=&\frac{|\boldsymbol{q}|}{w_0+|\boldsymbol{q}|} + \frac{w_0}{w_0+|\boldsymbol{q}|} + V_{t+1}(\boldsymbol{q}-\boldsymbol{e}_j)\\
\le&1+V_{t+1}(\boldsymbol{q}-\boldsymbol{e}_j).
\end{aligned}$$

(ii) If $q_j = 1$, then we have

$$\begin{aligned}
&V_t(\boldsymbol{q})\\
=&\frac{|\boldsymbol{q}|}{w_0+|\boldsymbol{q}|} + \sum_{i\in A(\boldsymbol{q})} \frac{1}{w_0+|\boldsymbol{q}|} V_{t+1}(\boldsymbol{q}-\boldsymbol{e}_i) + \frac{w_0}{w_0+|\boldsymbol{q}|} V_{t+1}(\boldsymbol{q})
\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{|\boldsymbol{q}|}{w_0+|\boldsymbol{q}|} + \sum_{i\in A(\boldsymbol{q}):i\neq j} \frac{1}{w_0+|\boldsymbol{q}|}\left(1+V_{t+2}(\boldsymbol{q}-\boldsymbol{e}_i-\boldsymbol{e}_j)\right) + \frac{1}{w_0+|\boldsymbol{q}|}V_{t+1}(\boldsymbol{q}-\boldsymbol{e}_j) \\
&\quad + \frac{w_0}{w_0+|\boldsymbol{q}|}\left(1+V_{t+2}(\boldsymbol{q}-\boldsymbol{e}_j)\right) \\
&\leq \frac{|\boldsymbol{q}|}{w_0+|\boldsymbol{q}|} + \frac{w_0+|\boldsymbol{q}|-1}{w_0+|\boldsymbol{q}|} + \frac{1}{w_0+|\boldsymbol{q}|}V_{t+1}(\boldsymbol{q}-\boldsymbol{e}_j) + \frac{w_0+|\boldsymbol{q}|-1}{w_0+|\boldsymbol{q}|}\left(V_{t+1}(\boldsymbol{q}-\boldsymbol{e}_j) - \frac{|\boldsymbol{q}|-1}{w_0+|\boldsymbol{q}|-1}\right) \\
&= \frac{|\boldsymbol{q}|}{w_0+|\boldsymbol{q}|} + \frac{w_0}{w_0+|\boldsymbol{q}|} + V_{t+1}(\boldsymbol{q}-\boldsymbol{e}_j) \\
&\leq 1 + V_{t+1}(\boldsymbol{q}-\boldsymbol{e}_j).
\end{aligned}
$$

Therefore, we can conclude the statement holds.

(b) We show the result by induction. For time T, suppose there exist j and k such that $q_j - q_k \geq 2$, we consider two possible cases:

(i) If $q_k > 0$, then

$$V_T(\boldsymbol{q}+\boldsymbol{e}_k-\boldsymbol{e}_j) = V_T(\boldsymbol{q}) = \frac{|\boldsymbol{q}|}{w_0+|\boldsymbol{q}|}.$$

(ii) If $q_k = 0$, then

$$V_T(\boldsymbol{q}+\boldsymbol{e}_k-\boldsymbol{e}_j) = \frac{1+|\boldsymbol{q}|}{w_0+1+|\boldsymbol{q}|} \geq \frac{|\boldsymbol{q}|}{w_0+|\boldsymbol{q}|} = V_T(\boldsymbol{q}).$$

Now assume the statement holds for time $t+1$, then for time t and any inventory $\boldsymbol{q}$, suppose there exist j and k such that $q_j - q_k \geq 2$, we again consider two possible cases:

(i) If $q_k > 0$, then

$$
\begin{aligned}
&V_t(\boldsymbol{q}+\boldsymbol{e}_k-\boldsymbol{e}_j) \\
&= \frac{|\boldsymbol{q}|}{w_0+|\boldsymbol{q}|} + \sum_{i\in\boldsymbol{q}} \frac{1}{w_0+|\boldsymbol{q}|}V_{t+1}(\boldsymbol{q}+\boldsymbol{e}_k-\boldsymbol{e}_j-\boldsymbol{e}_i) + \frac{w_0}{w_0+|\boldsymbol{q}|}V_{t+1}(\boldsymbol{q}+\boldsymbol{e}_k-\boldsymbol{e}_j) \\
&\geq \frac{|\boldsymbol{q}|}{w_0+|\boldsymbol{q}|} + \sum_{i\in\boldsymbol{q}:i\neq j} \frac{1}{w_0+|\boldsymbol{q}|}V_{t+1}(\boldsymbol{q}-\boldsymbol{e}_i) + \frac{1}{w_0+|\boldsymbol{q}|}V_{t+1}(\boldsymbol{q}+\boldsymbol{e}_k-2\boldsymbol{e}_j) + \frac{w_0}{w_0+|\boldsymbol{q}|}V_{t+1}(\boldsymbol{q}) \\
&\geq \frac{|\boldsymbol{q}|}{w_0+|\boldsymbol{q}|} + \sum_{i\in\boldsymbol{q}:i\neq j} \frac{1}{w_0+|\boldsymbol{q}|}V_{t+1}(\boldsymbol{q}-\boldsymbol{e}_i) + \frac{1}{w_0+|\boldsymbol{q}|}V_{t+1}(\boldsymbol{q}-\boldsymbol{e}_j) + \frac{w_0}{w_0+|\boldsymbol{q}|}V_{t+1}(\boldsymbol{q}) \\
&= V_t(\boldsymbol{q}),
\end{aligned}
$$

where the last inequality follows from the fact that if $q_j - q_k \geq 3$, then it holds by induction hypothesis, and if $q_j - q_k = 2$, then $\boldsymbol{q}+\boldsymbol{e}_k-2\boldsymbol{e}_j$ is a permutation of $\boldsymbol{q}-\boldsymbol{e}_j$, and the inequality holds from Lemma F.1.

(ii) If $q_k = 0$, then

$$
\begin{aligned}
& V_t(\boldsymbol{q}+\boldsymbol{e}_k-\boldsymbol{e}_j) \\
=& \frac{1+|\boldsymbol{q}|}{w_0+1+|\boldsymbol{q}|} + \sum_{i\in A(\boldsymbol{q})\cup\{k\}} \frac{1}{w_0+1+|\boldsymbol{q}|} V_{t+1}(\boldsymbol{q}+\boldsymbol{e}_k-\boldsymbol{e}_j-\boldsymbol{e}_i) + \frac{w_0}{w_0+1+|\boldsymbol{q}|} V_{t+1}(\boldsymbol{q}+\boldsymbol{e}_k-\boldsymbol{e}_j) \\
=& \frac{1+|\boldsymbol{q}|}{w_0+1+|\boldsymbol{q}|} + \sum_{i\in A(\boldsymbol{q})} \frac{1}{w_0+1+|\boldsymbol{q}|} V_{t+1}(\boldsymbol{q}+\boldsymbol{e}_k-\boldsymbol{e}_j-\boldsymbol{e}_i) + \frac{1}{w_0+1+|\boldsymbol{q}|} V_{t+1}(\boldsymbol{q}-\boldsymbol{e}_j) \\
& + \frac{w_0}{w_0+1+|\boldsymbol{q}|} V_{t+1}(\boldsymbol{q}+\boldsymbol{e}_k-\boldsymbol{e}_j) \\
\geq& \frac{1+|\boldsymbol{q}|}{w_0+1+|\boldsymbol{q}|} + \sum_{i\in A(\boldsymbol{q})} \frac{1}{w_0+1+|\boldsymbol{q}|} V_{t+1}(\boldsymbol{q}-\boldsymbol{e}_i) + \frac{1}{w_0+1+|\boldsymbol{q}|} V_{t+1}(\boldsymbol{q}-\boldsymbol{e}_j) + \frac{w_0}{w_0+1+|\boldsymbol{q}|} V_{t+1}(\boldsymbol{q}) \\
=& \frac{w_0+|\boldsymbol{q}|}{w_0+1+|\boldsymbol{q}|} \left(\frac{|\boldsymbol{q}|}{w_0+|\boldsymbol{q}|} + \sum_{i\in A(\boldsymbol{q})} \frac{1}{w_0+|\boldsymbol{q}|} V_{t+1}(\boldsymbol{q}-\boldsymbol{e}_i) + \frac{w_0}{w_0+|\boldsymbol{q}|} V_{t+1}(\boldsymbol{q}) \right) + \frac{1}{w_0+1+|\boldsymbol{q}|} \\
& + \frac{1}{w_0+1+|\boldsymbol{q}|} V_{t+1}(\boldsymbol{q}-\boldsymbol{e}_j) \\
=& \frac{w_0+|\boldsymbol{q}|}{w_0+1+|\boldsymbol{q}|} V_t(\boldsymbol{q}) + \frac{1}{w_0+1+|\boldsymbol{q}|} + \frac{1}{w_0+1+|\boldsymbol{q}|} V_{t+1}(\boldsymbol{q}-\boldsymbol{e}_j) \\
\geq& \frac{w_0+|\boldsymbol{q}|}{w_0+1+|\boldsymbol{q}|} V_t(\boldsymbol{q}) + \frac{1}{w_0+1+|\boldsymbol{q}|} V_t(\boldsymbol{q}) \\
=& V_t(\boldsymbol{q}),
\end{aligned}
$$

where the first inequality holds follows from the same reason as in case (i) and the last inequality follows from Proposition F.1 (a). Therefore, the statement holds by induction. ■

Appendix G: Additional Numerical Study

Here we conduct another numerical experiment to compare our solution with benchmarks in literature. We consider the Markov chain choice model, adopting the experimental setup from Mouchtaki et al. (2021). Specifically, we consider a uniform Markov chain. The arrival probabilities $(\lambda_i)_{i\in\mathcal{N}\cup\{0\}}$ are calculated as $\lambda_i = \frac{U_i}{U_0+\sum_{j\in\mathcal{N}} U_j}$ where $(U_i)_{i\in\mathcal{N}\cup\{0\}}$ are independent random variables sampled uniformly in $[0,1]$. The transition probabilities are defined as $\rho_{ij} = \frac{U_{ij}}{U_{i0}+\sum_{j\in\mathcal{N}} U_{ij}}$ where $(U_{ij})_{i,j\in\mathcal{N}\cup\{0\}}$ are independent random variables uniformly distributed in $[0,1]$. For the products, the revenue p_i is uniformly generated from $[0,1]$ and the cost is defined as $c_i = \frac{p_i}{1+M_i}$, where M_i represents the markup of product i and is sampled uniformly in $[0.5, 0.6]$. In Table 6, we report the optimality gaps for $\boldsymbol{q}^{\mathsf{round}}$, $\lfloor \boldsymbol{q}^{\mathsf{fluid}} \rfloor$ and $\boldsymbol{q}^{\mathsf{SAA}}$, where $\boldsymbol{q}^{\mathsf{SAA}}$ stands for the NC-SAA algorithm introduced in Mouchtaki et al. (2021). In the NC-SAA algorithm, we take the threshold (see details in Mouchtaki et al. 2021) to be $\{0, 0.01, \ldots, 0.1\}$ and employ $N_{\text{samp}} = 100$ samples to approximate customer choices. Table 6 reports the numerical performances for the three solutions. The NC-SAA solution performs the best when $T = 100$; however, its performance deteriorates as T increases, because $N_{\text{samp}} = 100$ samples can no longer accurately capture customer choices as T grows. Note that implementing the SAA algorithm involves solving a large linear program, which has $\Omega(N_{\text{samp}}T)$ variables and $\Omega(N_{\text{samp}}T)$ constraints. In practice, implementing the SAA algorithm with a sufficiently large N_{samp} becomes computationally challenging, especially for larger T, even though it might offer superior performance. By contrast, solving the fluid relaxation under the Markov chain choice model can be done efficiently, and both solutions $\boldsymbol{q}^{\text{round}}$ and $\lfloor \boldsymbol{q}^{\text{fluid}} \rfloor$ consistently perform well for various problem sizes as shown in Table 6.

		$n = 20$	$n = 40$	$n = 60$	$n = 80$
$T = 100$	$100 \times \frac{1}{V^{\text{fluid}}}(V^{\text{fluid}} - \Pi(\boldsymbol{q}^{\text{round}}))$	11.30	11.23	9.64	9.27
	$100 \times \frac{1}{V^{\text{fluid}}}(V^{\text{fluid}} - \Pi(\lfloor \boldsymbol{q}^{\text{fluid}} \rfloor))$	9.89	9.27	8.00	8.13
	$100 \times \frac{1}{V^{\text{fluid}}}(V^{\text{fluid}} - \Pi(\boldsymbol{q}^{\text{SAA}}))$	**9.59**	**8.89**	**7.48**	**6.69**
$T = 500$	$100 \times \frac{1}{V^{\text{fluid}}}(V^{\text{fluid}} - \Pi(\boldsymbol{q}^{\text{round}}))$	5.23	4.87	4.03	3.77
	$100 \times \frac{1}{V^{\text{fluid}}}(V^{\text{fluid}} - \Pi(\lfloor \boldsymbol{q}^{\text{fluid}} \rfloor))$	**4.61**	**3.98**	3.07	2.76
	$100 \times \frac{1}{V^{\text{fluid}}}(V^{\text{fluid}} - \Pi(\boldsymbol{q}^{\text{SAA}}))$	10.69	5.61	**2.99**	**2.74**
$T = 1000$	$100 \times \frac{1}{V^{\text{fluid}}}(V^{\text{fluid}} - \Pi(\boldsymbol{q}^{\text{round}}))$	3.47	3.25	2.65	2.53
	$100 \times \frac{1}{V^{\text{fluid}}}(V^{\text{fluid}} - \Pi(\lfloor \boldsymbol{q}^{\text{fluid}} \rfloor))$	**3.43**	**2.97**	**2.13**	**2.02**
	$100 \times \frac{1}{V^{\text{fluid}}}(V^{\text{fluid}} - \Pi(\boldsymbol{q}^{\text{SAA}}))$	15.34	8.02	3.91	2.66

Table 6 Optimality gaps for the stocking quantities $\boldsymbol{q}^{\text{round}}$, $\lfloor \boldsymbol{q}^{\text{fluid}} \rfloor$ and $\boldsymbol{q}^{\text{SAA}}$ under the Markov chain choice model for varying (T, n). The best algorithmic performance (smallest optimality gap) in each cell is **bolded**.

The markups from the setup of Mouchtaki et al. (2021) are also small and encourage understocking. We again conduct a further set of experiments where we change the markup to be sampled uniformly from [1.2,1.5] (instead of [0.5,0.6]). The results are shown in Table 7, which demonstrate that our procedure performs better than both $\lfloor \boldsymbol{q}^{\text{fluid}} \rfloor$ when margins are high and $\boldsymbol{q}^{\text{SAA}}$ when horizons are long.

		$n=20$	$n=40$	$n=60$	$n=80$
$T=100$	$100\times\frac{1}{V^{\mathsf{fluid}}}(V^{\mathsf{fluid}}-\Pi(\boldsymbol{q}^{\mathsf{round}}))$	6.176	6.251	**6.656**	6.021
	$100\times\frac{1}{V^{\mathsf{fluid}}}(V^{\mathsf{fluid}}-\Pi(\lfloor\boldsymbol{q}^{\mathsf{fluid}}\rfloor))$	6.409	6.984	9.187	7.258
	$100\times\frac{1}{V^{\mathsf{fluid}}}(V^{\mathsf{fluid}}-\Pi(\boldsymbol{q}^{\mathsf{SAA}}))$	**5.925**	**6.164**	6.806	**5.767**
$T=500$	$100\times\frac{1}{V^{\mathsf{fluid}}}(V^{\mathsf{fluid}}-\Pi(\boldsymbol{q}^{\mathsf{round}}))$	**2.727**	**2.733**	2.856	2.510
	$100\times\frac{1}{V^{\mathsf{fluid}}}(V^{\mathsf{fluid}}-\Pi(\lfloor\boldsymbol{q}^{\mathsf{fluid}}\rfloor))$	2.843	2.763	2.907	2.581
	$100\times\frac{1}{V^{\mathsf{fluid}}}(V^{\mathsf{fluid}}-\Pi(\boldsymbol{q}^{\mathsf{SAA}}))$	5.148	3.282	**2.736**	**2.412**
$T=1000$	$100\times\frac{1}{V^{\mathsf{fluid}}}(V^{\mathsf{fluid}}-\Pi(\boldsymbol{q}^{\mathsf{round}}))$	**1.914**	**1.883**	1.977	**1.733**
	$100\times\frac{1}{V^{\mathsf{fluid}}}(V^{\mathsf{fluid}}-\Pi(\lfloor\boldsymbol{q}^{\mathsf{fluid}}\rfloor))$	1.975	1.899	**1.976**	1.747
	$100\times\frac{1}{V^{\mathsf{fluid}}}(V^{\mathsf{fluid}}-\Pi(\boldsymbol{q}^{\mathsf{SAA}}))$	7.131	3.929	2.918	2.382

Table 7 Optimality gaps for the stocking quantities $\boldsymbol{q}^{\mathsf{round}}$, $\lfloor\boldsymbol{q}^{\mathsf{fluid}}\rfloor$ and $\boldsymbol{q}^{\mathsf{SAA}}$ under the Markov chain choice model for varying (T,n). The best algorithmic performance (smallest optimality gap) in each cell is **bolded**.