Int. J. Production Economics 265 (2023) 109010

Contents lists available at ScienceDirect

International Journal of Production Economics

journal homepage: www.elsevier.com/locate/ijpe

Multiple-purchase choice model: estimation and optimization

Mengmeng Wang [a], Xun Zhang [a,*], Xiaolong Li [b]

[a] Antai College of Economics and Management, Shanghai Jiao Tong University, 1954 Huashan Road, Shanghai, China
[b] Institute of Operations Research and Analytics, National University of Singapore, 3 Research Link, Innovation 4.0, Singapore

ARTICLE INFO

Keywords:
Multiple purchase behavior
Choice model estimation
Assortment optimization

ABSTRACT

Although multiple-purchase behavior is typical in retail practice, the choice model to portray such behavior is limited in existing research. This paper presents a new multiple-purchase (MP) choice model based on the multinomial logit (MNL) choice model, which allows customers to purchase more than one item in a single visit. We first prove that the log-likelihood function based on our MP choice model has a nice concave property such that we can efficiently estimate the parameters in the model with data. Next, we present an equivalent mixed-integer program for the multiple-purchase assortment optimization, which can be solved by state-of-the-art commercial solvers. Finally, we conduct extensive numerical experiments to evaluate the benefits from the MP choice model in both estimation and optimization problems. We first conduct a case study on a real-world dataset. The numerical results show that our MP choice model performs better in three estimation metrics and one revenue metric than the MNL choice model. Then, we demonstrate the advantage of the MP choice model on simulated data. Our model can provide significant realized revenue improvement compared with that obtained by the single-purchase MNL choice model in numerical results.

1. Introduction

Choice models are often used to model customer choice behavior in revenue management (e.g Talluri and Van Ryzin, 2004; Liu and Van Ryzin, 2008; Ma et al., 2016; Feldman and Topaloglu, 2017b). Among them, the most commonly used model in economics, marketing, and operations management is the multinomial logit (MNL) choice model (Rusmevichientong et al., 2010). However, the MNL choice model suffers from the independence of irrelevant alternatives (IIA) property, which means that the predicted market share of each product will decrease by the same relative amount if we add a new product into the offered assortment (Gallego and Topaloglu, 2014). One option to avoid IIA is to use nested multinomial logit (NMNL) model (Davis et al., 2014). In the NMNL model, we can aggregate products into nests so that the IIA holds within each nest but not across nests (Strauss et al., 2018). Another extension of the MNL model is the mixed multinomial logit (MMNL) choice model. In revenue management, the decision-maker typically seeks to find out the difference in customers' preferences and make a better decision. The MMNL choice model segments the customer population into different customer types and uses different MNL models for each type (McFadden and Train, 2000). For more about choice model, we refer readers to see Feng et al. (2022b).

However, all the above-mentioned choice models assume that customers only choose one product from the offer set, ignoring the fact that in many scenarios, customers usually choose more than one item at a time. For instance, in retailing, a considerable number of customers opt for multiple items when shopping in physical stores or online (Vend, 2019; Wang et al., 2022). In the context of advertising, customers also exhibit some choice behavior (Li et al., 2022), such as engaging with more than one ad by clicking on them, studies can gain valuable insights by taking into account customers' propensity for multiple-choice behaviors. Furthermore, a large percentage of customers may also purchase more than one item online because of the free delivery conditions (Cachon et al., 2018). For these scenarios, the MNL and other choice models based on the single-choice assumption do not perform well (Feldman et al., 2018).

This paper aims to bridge the gap between the existing research and practical purchase scenarios by relaxing the model assumption that customers only choose one item from an offer set at a single visit. Specifically, we present a multiple-purchase (MP) choice model to capture multiple-purchase behavior from a single customer visit. We show that the estimation procedure of this MP choice model is easily implemented. Then, for the corresponding assortment optimization problem, we demonstrate that the revenue-ordered property of the MNL choice model does not hold for the optimal assortment. To solve the assortment problem based on the MP choice model, we reformulate it as a mixed-integer program, which can be solved by state-of-the-art commercial solvers, such as CPLEX and GUROBI. Moreover, we conduct

* Corresponding author.
E-mail address: xunzhang@sjtu.edu.cn (X. Zhang).

https://doi.org/10.1016/j.ijpe.2023.109010
Received 3 November 2022; Received in revised form 19 July 2023; Accepted 7 August 2023
Available online 9 August 2023
0925-5273/

extensive numerical experiments to illustrate the benefits of considering customer multiple-purchase behavior and compare the results with that of the classical single-purchase based choice model (MNL) by using simulation data and a real-world dataset. To address the difficulty of observing customer's choice sequence for purchases in the real-world dataset, we propose an algorithm to determine the choice sequence of a customer during the purchase process. We show that the estimation accuracy in the multiple-purchase case is much higher than that in the single-purchase case in terms of the purchase probabilities and the expected revenue. For the corresponding assortment optimization problem, the MP choice model can even improve the realized revenue by 26% at most in the simulation case.

The remainder of this paper is organized as follows. We review the related research in Section 2. In Section 3, we formulate the choice probability model and corresponding assortment problem. Then, we show estimation procedure can be easily implemented in Section 4. In Section 5, we formulate the corresponding assortment optimization problem as a integer program. In Section 6 we present the numerical performance of the MP choice model using both the simulation data and the real world dataset. Finally, we conclude and discuss possible avenues of future research in Section 7.

Notation. Given an integer n, we define $[n] = \{1, 2, \dots, n\}$. We use $\tilde{u}$ to denote the random variable, such as ϵ with a tilde sign represents the random variable $\tilde{\epsilon}$. We use $|\cdot|$ to denote the cardinality of a finite set, such as $|\mathcal{S}|$ represents the cardinality of set $\mathcal{S}$. We use boldface lowercase letters to represent vectors, such as vector $\boldsymbol{u}$.

2. Related literature

The choice model has been widely studied in the operations management community. An expansive collection of literature discusses customer choice and the corresponding assortment optimization problems. Beside the research on solving the optimal assortment for the retailer, there is also another stream of literature studying the factors that would impact retailers' assortment decisions, such as Alibeiki et al. (2020), Flapper et al. (2010), Ren et al. (2011), Cachon and Kök (2007), Kök and Xu (2011). Since this research mainly focuses on solving the optimal assortment, and is also an extension of the multinomial logit choice model, we focus on the literature review on the logit-based choice models and the corresponding assortment optimization problems, including the MNL, MMNL and NMNL choice models.

The MNL choice model is one of the most widely used choice models and is extensively studied in the revenue management literature. McFadden et al. (1973) initially propose the conditional logit model and refer to it as the MNL choice model. Since the assortment optimization under the MNL choice model is easily tractable, substantial research has studied the estimation procedure and the assortment optimization problem. Talluri and Van Ryzin (2004) show that the corresponding assortment optimization can be solved efficiently via greedy-based methods. They find that the optimal assortment exhibits the well-known revenue-ordered property. In addition, Gallego et al. (2004) show that the optimization problem can be solved via a linear program given the choice-model-based customer demand in their setting. Rusmevichientong et al. (2010) add the capacity constraint and develop a polynomial algorithm to identify the optimal assortment. Rusmevichientong and Topaloglu (2012) study the robust formulations of assortment optimization problems under the MNL choice model. Chan et al. (2020) also investigate a robust assortment, where expected store demand and coefficient of variation are known in advance. They numerically demonstrate that their heuristic performs very well. Wang (2018) incorporate reference prices into the MNL choice model and investigate operations management problems. Katsifou et al. (2014) consider a joint product assortment, inventory and price optimization problem by combining "standard" and "special" products in the total assortment. They show that retailer can benefit from the introduction of special products if a cross-selling effect exists. Chen et al. (2021) study a multi-store location and location-dependent assortment planning problem faced by an omnichannel retailer. The numerical results show the effectiveness of their approach both in estimation and assortment optimization problem.

The MMNL choice model is developed to relax the customer homogeneity assumption of the MNL choice model by segmenting the customer population into multiple types: each customer type is modeled by a unique MNL choice model. McFadden and Train (2000) show that the MMNL model can be used to approximate any discrete choice model derived from random utility maximization as closely as required. Although MMNL is one of the most general choice model, the corresponding assortment optimization problem is not easy to solve. Rusmevichientong et al. (2014) show that the problem is NP-complete even with two customer classes and propose a polynomial-time approximation scheme. Feldman and Topaloglu (2017a) extend the result and assume a capacity constraint. They show that the problem is NP-hard and find a fully polynomial-time approximation scheme for the problem. We refer readers to Désir et al. (2014), and Jagabathula et al. (2018) for another fully polynomial-time approximation scheme (FPTAS) and the estimation problem under MMNL, respectively.

Additionally, the nested logit choice model is developed to alleviate the independence from IIA property of the MNL choice model. Under this kind of model, customers first select a nest and then choose a product within the nest (Davis et al., 2014). However, estimation under this model is difficult and requires the knowledge of key attributes and the hierarchy of consumers for proper partitioning of nests (Kök et al., 2008). Train (2009) provide an iterative maximum likelihood approach for the estimation procedure. Davis et al. (2014) find that the assortment optimization problem is NP-hard in most general cases under this choice model. They also show that it can be polynomial solvable with two conditions: (1) the nest dissimilarity parameters of the choice model are less than one; (2) customers always make a purchase within the selected nest. Wan et al. (2018) compare the nested logit model with the exogenous substitution model in estimating substitution probabilities. They numerically find that the nested logit model can obtain more accurate substitution probabilities than exogenous substitution model. Li et al. (2015) formulate a d-level nested logit model and provide an efficient algorithm that runs in $O(dn \log_n)$ time to find the optimal assortment. Gallego and Topaloglu (2014) generalize the problem to the case with cardinality and space constraints. They show that the optimal assortment under cardinality constraints can be obtained efficiently by solving a linear program. Unfortunately, the assortment optimization problem with space constraints is NP-hard. Feldman and Topaloglu (2015) develop another efficient algorithm to compute the optimal assortment. Interested readers are referred to Alptekinoğlu and Grasas (2014), Ryzin and Mahajan (1999), Anderson et al. (1992), Cachon and Kök (2007) for the nested logit model under consumer returns and other nested logit model settings.

In past decades, customers' multiple-purchase choice model has received attention in the literature. Most of them focus on the estimation process and ignore the assortment optimization problem. One line of the popularly mentioned literature on the multiple-purchase choice model is the multiple discrete-continuous (MDC) type model. Bhat (2008) first proposed the multiple discrete-continuous extreme value (MDCEV) model, which allows customers to purchase multiple products and possibly multiple units of each product. Instead of assuming that the purchased units have to be continuous, Bhat (2022) then proposed a multiple-discrete-count extreme value (MDCEV) model, which allows the purchase units to be discrete. Although MDC-type multiple purchase choice models have achieved good estimation performance, they may not be suitable for our subsequent assortment optimization problem. Firstly, the dependency between estimated parameters in the MDC-type model and the assortment prevents us from solving the subsequent assortment optimization problem. The MDC-type models implicitly assume that the assortment $\mathcal{S}$ remains unchanged over the entire dataset, indicating that each assortment has a specific estimation for the parameters. However, the subsequent assortment optimization requires that the estimation of parameters is independent of the assortment, thus avoiding the need to enumerate an exponential number of possible assortments. Secondly, in the MDC-type choice model, the budget E cannot be observed in our dataset. Thirdly, we are seeking a multiple-purchase choice model with discrete outcomes. Although the MDC-count model with discrete dependent outcomes has been proposed, the subsequent assortment problem remains challenging. Bhat (2022) shows that even if the MDC-count model has been estimated, simply applying fractions to a positive integer count would not yield integer values for the counts of individual items in the assortment. Although the authors propose a clever approach to map the fractions to discrete outcomes for individuals, incorporating such an approach into our assortment optimization problem proves to be very difficult.

Ranking-based choice models can also be used to characterize multiple-purchase behavior. Nair et al. (2018) presents a rank-ordered probit modeling approach that overcomes limitations associated with prior approaches in analyzing rank-ordered data. The model is applied to estimate the preference for adoption and usage of alternative autonomous vehicle (AV) modes and services, providing deeper insights into the differences in adoption preferences across groups. Additionally, Nair et al. (2019) demonstrates that a rank-ordered probit (ROP) model can better utilize ranking data information compared to a rank-ordered logit (ROL) model. Furthermore, Mondal and Bhat (2022) proposes a spatial rank-ordered probit (SROP) model that accommodates both spatial lag effects and spatial drift effects. The proposed model is applied in a mode choice context, and the results indicate that ignoring spatial effects substantially underestimates variable elasticity effects. However, the aforementioned ranking-based choice models assume that the ranking for the items in an assortment is known, which is not observed in our problem. Therefore, we cannot adopt a ranking-based model to solve our problem.

Apart from MDC-type models and ranking-based models, both Dubé (2004) and Harlam and Lodish (1995) propose their multiple-purchase choice model. One of the differences between these two papers is the assumption on the number of purchases. Dubé (2004) assume the number of purchases a customer must buy is generated from an exponential distribution, while the number of purchases is given in Harlam and Lodish (1995). The assumption on the number of purchases in this paper is the same as that in Harlam and Lodish (1995).

Recently, some papers pay more attention to the optimization problem. Bai et al. (2023) assumes that the customers would choose at most M products and provide two polynomial-time approximations schemes for the assortment problems. However, the complexity of estimating the proposed choice model is not discussed. Similarly, Zhang et al. (2021) investigates an assortment optimization problem based on the Multiple-Discrete-Choice model, but the estimation process is not discussed. Huh and Li (2022) also study a multiple-purchase choice assortment problem based on the MDCEV model, but the quantity of the purchased item should be continuous.

Both Luan et al. (2020) and Tulabandhula et al. (2020) study estimation and assortment optimization problems. Luan et al. (2020) assume customers first form a consideration set, select one product from the consideration set, and determine the purchased quantity of the desired product. The critical difference between this work and our paper is that they assume customers only choose one product with multiple doses, whereas we accept that customers can purchase multiple products. Tulabandhula et al. (2020) are the most closely related to our work and consider customers purchasing multiple products. They propose a bundle of multivariate logit models and develop a binary-search-based iterative strategy to solve assortment problems. In contrast, we propose an attractable MP choice model from the estimation perspective and use an integer program to solve the multiple-purchase assortment problem in this work.

3. Model formulation

In this section, we present our multiple-purchase choice model and then propose the corresponding assortment optimization problem based on our choice model. Because our model is constructed based on the MNL model, let us begin with a brief introduction to the MNL choice model and the corresponding assortment problem.

3.1. MNL model

Let $[N] = \{1, 2, \ldots, N\}$ denote a universe of substitute products, and let r_i be the revenue associated with product $i, \forall i \in [N]$. We assume that products are indexed such that $r_1 \geq r_2 \geq \cdots \geq r_N$. In the MNL choice model, the utility of product i is denoted by $\tilde{u}_i = u_i + \tilde{\epsilon}_i$, where u_i is the mean utility, and $\tilde{\epsilon}_i$ follows a Gumbel distribution. The no-purchase utility is denoted by $\tilde{u}_0^1 = u_0^1 + \tilde{\epsilon}_0^1$. Without loss of generality, we assume $u_0^1 = 0$.

Under the MNL choice model, given assortment $\mathcal{S} \subseteq [N]$, the probability that the customer purchases product i is as follows,

$$P^{MNL}(i|\mathcal{S}) = \frac{e^{u_i}}{e^{u_0^1} + \sum_{j \in \mathcal{S}} e^{u_j}}. \tag{1}$$

Then, let $v_i = e^{u_i}, i \in \mathcal{S}$, the expected revenue under the MNL choice model is denoted as

$$\pi^{MNL}(\mathcal{S}) = \sum_{i \in \mathcal{S}} r_i \left(\frac{v_i}{v_0^1 + \sum_{j \in \mathcal{S}} v_j} \right)$$

Therefore, the corresponding assortment optimization problem can be expressed as

$$[MNL] = \max_{\mathcal{S}} \sum_{i \in \mathcal{S}} r_i \left(\frac{v_i}{v_0^1 + \sum_{j \in \mathcal{S}} v_j} \right)$$

In the assortment optimization problem $[MNL]$, the objective is to determine a subset of $\mathcal{S} \subseteq [N]$ that maximizes the expected total revenue. Previous studies such as Talluri and Van Ryzin (2004) and Rusmevichientong et al. (2014) have shown that the optimal assortment follows a revenue-ordered pattern. This means that the optimal assortment includes a certain number of products with the highest revenues. Moreover, when considering the assortment optimization problem $[MNL]$ with a cardinality constraint, Gallego and Topaloglu (2014) have presented an equivalent linear program formulation. This formulation allows the problem to be solved using commercial solvers, providing an efficient approach to find the optimal assortment under such constraints. By leveraging the revenue-ordered property and utilizing the linear program formulation, decision-makers can effectively tackle the assortment optimization problem $[MNL]$ and make informed decisions on the subset of products to offer in order to maximize expected total revenue.

3.2. Multiple-purchase model

In this section, we present our multiple-purchase choice model based on the MNL model. Throughout our discussion, we will continue to use the notation $[N]$ to represent a universe of products, and r_i to denote the revenue associated with product i for all i in $[N]$.

We assume that a customer purchases at most two products in a single visit. We make this assumption for the following reasons. First, using the two-purchase choice model, we can clearly illustrate the estimation and optimization process. Second, this approach provides enough insights to demonstrate the benefit of the multiple-purchase choice model. Moreover, the proportions of customers that purchase more than two items are limited. For example, the real-world dataset we obtained from a retailer shows that only 2.21% of customers buy more than two items (see Table 5). We also assume that a customer chooses only one unit of a product. The real-world dataset obtained from a Chinese retail company shows that the proportion of customers who purchase multiple units of the same product is rare (less than 4%). Hence, the assumption that the customer can only buy one unit of each item is reasonable in our problem. However, it is worth noting that we will discuss the scenario where a customer can choose multiple units of a product later in this section.

In our setting, each customer chooses products that maximize the utility. The utility of product i is still denoted by $\tilde{u}_i = u_i + \tilde{\epsilon}_i^1$, where u_i is the mean utility and $\tilde{\epsilon}_i^1$ follows a Gumbel distribution. The no-purchase utility in the first choice is denoted by $\tilde{u}_0^1 = u_0^1 + \tilde{\epsilon}_0^1$. Suppose the customer first chooses product i, we define the no-purchase utility in the second choice as $\tilde{u}_0^2(i) = u_0^2 + u_i + \tilde{\epsilon}_i^2$, where u_0^2 is a fixed value for customers' no-purchase choice in the second purchase, and $\tilde{\epsilon}_i^2$ follows a Gumbel distribution. The definition of $u_0^2(i) = u_0^2 + u_i + \tilde{\epsilon}_0^2$ is reasonable. First, if the mean utility of product i in the first choice, denoted by u_i, is large enough, it implies that the customer is likely to be satisfied with this purchase and not buy a second product. It may naturally result from either a budget concern or the effectiveness of substitution. To capture this behavior, we define the no-purchase utility in the second choice, $u_0^2(i)$, as an affine and increasing function of u_i. Generally speaking, $u_0^2(i)$ could be any increasing function of u_i. Second, for the tractability of estimation, we preserve the concavity of the log-likelihood function by defining $u_0^2(i)$ in this way, which allows us to estimate the parameters in our multiple-purchase choice model efficiently. We will show this property in Section 4.

Next, we present our MP choice model based on the MNL choice model. As we assume a customer only chooses one unit of a product, for an assortment $S \subseteq [N]$, a product that is purchased only appears in either the first purchase or the second purchase after some other product has been chosen. Then the probability of product $i \in S$ being chosen in the first purchase is denoted as:

$$P^1(i|S) = \frac{e^{u_i}}{e^{u_0^1} + \sum_{j\in S} e^{u_j}},$$

and the probability of product $i \in S$ being chosen in the second purchase conditional on some other product $k \in S$ being purchased first is denoted as:

$$P^2(i|k,S) = \frac{e^{u_i}}{e^{u_0^2+u_k} + \sum_{j\in S\setminus\{k\}} e^{u_j}}.$$

Importantly, the formulation of the choice probability indicates that there is a sequential choice process among different products, exploring different scenarios where product i is chosen first or second. Moreover, it is worth noting that when u_0^2 is equal to 0, the order of purchase becomes irrelevant under our choice model. This observation emphasizes a key advantage of our model: its ability to be generalized to cases where the order of purchase does not matter.

Hence, the probability that product $i \in S$ is purchased can be denoted as:

$$\begin{aligned} P(i|S) &= P^1(i|S) + \sum_{k\in S\setminus\{i\}} P^1(k|S)P^2(i|k,S) \\ &= \frac{e^{u_i}}{e^{u_0^1} + \sum_{j\in S} e^{u_j}} + \sum_{k\in S\setminus\{i\}} \frac{e^{u_k}}{e^{u_0^1} + \sum_{j\in S} e^{u_j}} \cdot \frac{e^{u_i}}{e^{u_0^2+u_k} + \sum_{j\in S\setminus\{k\}} e^{u_j}}. \end{aligned}$$

Let $v_i = e^{u_i}, i \in S$, then:

$$P(i|S) = \frac{v_i}{v_0^1 + \sum_{j\in S} v_j} + \sum_{k\in S\setminus\{i\}} \frac{v_k}{v_0^1 + \sum_{j\in S} v_j} \cdot \frac{v_i}{v_0^2 v_k + \sum_{j\in S\setminus\{k\}} v_j}. \quad (2)$$

Note that our MP choice model still can be extended to accommodate the customer's purchase of two or more products of the same kind. There are two approaches taking this issue into account: First, we can keep the product purchased in the first choice staying in the offer set of the second purchase, which allows the customer to consider purchasing more than one unit of the same product. Note that our model can be extended to a multiple-time purchase model (more than two purchases), following this approach, multiple units of any offered product could be purchased with positive probability under our model; Second, we can treat a bundle of multiple units of some product as a new product in our choice model. For example, suppose a customer could purchase two units of product i, we can consider this bundle as a new product denoted by $i^\dagger$ in the model and then estimate its utility. The first approach asks for more rounds of choice which may highly increase the complexity of our decision problem later, whereas the second one suffers from the exponentially increasing number of variables with N increases.

Based on the given probability function, the expected revenue is denoted as:

$$\pi(S) = \sum_{i\in S} r_i \left(\frac{v_i}{v_0^1 + \sum_{j\in S} v_j} + \sum_{k\in S\setminus\{i\}} \frac{v_k}{v_0^1 + \sum_{j\in S} v_j} \frac{v_i}{v_0^2 v_k + \sum_{j\in S\setminus\{k\}} v_j} \right)$$

Therefore, the assortment problem can be stated as follows:

$$\begin{aligned} [MP] &= \max_S \sum_{i\in S} r_i \left(\frac{v_i}{v_0^1 + \sum_{j\in S} v_j} + \sum_{k\in S\setminus\{i\}} \frac{v_k}{v_0^1 + \sum_{j\in S} v_j} \frac{v_i}{v_0^2 v_k + \sum_{j\in S\setminus\{k\}} v_j} \right) \\ &= \max_S \sum_{i\in S} r_i \frac{v_i}{v_0^1 + \sum_{j\in S} v_j} \left(1 + \sum_{k\in S\setminus\{i\}} \frac{v_k}{v_0^2 v_k + \sum_{j\in S\setminus\{k\}} v_j} \right) \end{aligned}$$

where $S \subseteq [N]$ is the decision variable and the objective is to maximize the expected revenue.

In the assortment problem, there are two challenges: the first is to estimate the values of all the mean utilities, and the second is to find an assortment to maximize the expected revenue. In the following sections, we attempt to solve these challenges.

4. Parameter estimation

In this section, we consider the problem of how to estimate the parameters in our choice model from observed data. We start the estimation process with complete purchase information, a thorough observation of arrivals, purchase sequence, and no-purchase outcomes. We show that the maximum likelihood method can be used to estimate parameters under complete purchase information in Section 4.1. Because the purchase sequence may not be observed in reality, we present an algorithm that helps us to estimate parameters without knowing the purchase sequence in Section 4.2.

4.1. Estimation with complete purchase information

Given the complete purchase information, i.e., the observation of arrivals, purchase sequence, and no-purchase outcomes, our setting has three cases: no-purchase, single-purchase, and two-purchase scenarios. Suppose there is a total of T customers, we introduce three indicator functions to construct the likelihood function. Let

$$\mathbb{I}_t^0 = \begin{cases} 1, & \text{if customers choose nothing} \\ 0, & \text{otherwise} \end{cases}, \quad t = 1, 2, \ldots, T$$

$$\mathbb{I}_t^1 = \begin{cases} 1, & \text{if customers choose only one item} \\ 0, & \text{otherwise} \end{cases}, \quad t = 1, 2, \ldots, T$$

$$\mathbb{I}_t^2 = \begin{cases} 1, & \text{if customers choose two items} \\ 0, & \text{otherwise} \end{cases}, \quad t = 1, 2, \ldots, T$$

For the three scenarios in our setting: no-purchase, single-purchase and two-purchase cases, given set $S \subseteq [N]$, the probabilities are detailed as follows:

- The probability that a customer chooses product i and product k is: $\frac{v_i}{v_0^1 + \sum_{j\in S} v_j} \cdot \frac{v_k}{v_0^2 v_i + \sum_{j\in S\setminus\{i\}} v_j}$;
- The probability that a customer chooses only product i is: $\frac{v_i}{v_0^1 + \sum_{j\in S} v_j} \cdot \frac{v_i v_0^2}{v_0^2 v_i + \sum_{j\in S\setminus\{i\}} v_j}$;
- The probability that a customer chooses nothing (the outside option) is: $\frac{v_0^1}{v_0^1 + \sum_{j\in S} v_j}$.

Note that the purchase sequence is assumed to be known in the two-purchase cases. More specifically, we assume that the customer t first purchases item i_t^2 and then buys item k_t^2 in the two-purchase cases. Similarly, we assume that the customer t buy item i_t^1 in the single-purchase cases. Given T customers, the maximum likelihood estimates for the parameter vector $\boldsymbol{v}$ can be obtained by maximizing the log-likelihood function:

$$
\begin{aligned}
LLH(\boldsymbol{u}) &= \log \prod_t \Big(\frac{v_{i_t^2}}{v_0^1 + \sum_{j\in S_t} v_j} \cdot \frac{v_{k_t^2}}{v_0^2 v_{i_t^2} + \sum_{j\in S_t\backslash\{i_t^2\}} v_j}\Big)^{\mathbb{I}_t^2} \\
&\quad \Big(\frac{v_{i_t^1}}{v_0^1 + \sum_{j\in S_t} v_j} \cdot \frac{v_{i_t^1} v_0^2}{v_0^2 v_{i_t^1} + \sum_{j\in S_t\backslash\{i_t^1\}} v_j}\Big)^{\mathbb{I}_t^1} \Big(\frac{v_0^1}{v_0^1 + \sum_{j\in S_t} v_j}\Big)^{\mathbb{I}_t^0} \\
&= \log \prod_t \Bigg[\Big(\frac{e^{u_{i_t^2}}}{e^{u_0^1} + \sum_{j\in S_t} e^{u_j}} \cdot \frac{e^{u_{k_t^2}}}{e^{u_0^2+u_{i_t^2}} + \sum_{j\in S_t\backslash\{i_t^2\}} e^{u_j}}\Big)^{\mathbb{I}_t^2} \cdot \\
&\quad \Big(\frac{e^{u_{i_t^1}}}{e^{u_0^1} + \sum_{j\in S_t} e^{u_j}} \cdot \frac{e^{u_0^2+u_{i_t^1}}}{e^{u_0^2+u_{i_t^1}} + \sum_{j\in S_t\backslash\{i_t^1\}} e^{u_j}}\Big)^{\mathbb{I}_t^1} \Big(\frac{e^{u_0^1}}{e^{u_0^1} + \sum_{j\in S_t} e^{u_j}}\Big)^{\mathbb{I}_t^0}\Bigg]
\end{aligned}
$$

where $\boldsymbol{u} = [u_i]_{i=0}^N$. In the following Proposition 1, we prove that the log-likelihood function $LLH(\boldsymbol{u})$ is a concave function.

Proposition 1. *$LLH(\boldsymbol{u})$ is a concave function.*

Proof. To prove the concavity of the log-likelihood function, we first rewrite $LLH(\boldsymbol{u})$ as follows:

$$
\begin{aligned}
LLH(\boldsymbol{u}) = \sum_t \Bigg\{ &\mathbb{I}_t^2 \Bigg[u_{i_t^2} + u_{k_t^2} - \log(e^{u_0^1} + \sum_{j\in S_t} e^{u_j}) - \log(e^{u_0^2+u_{i_t^2}} + \sum_{j\in S_t\backslash\{i_t^2\}} e^{u_j})\Bigg] \\
&+ \mathbb{I}_t^1 \Bigg[u_{i_t^1} + (u_0^2 + u_{i_t^1}) - \log(e^{u_0^1} + \sum_{j\in S_t} e^{u_j}) - \log(e^{u_0^2+u_{i_t^1}} + \sum_{j\in S_t\backslash\{i_t^1\}} e^{u_j})\Bigg] \\
&+ \mathbb{I}_t^0 \Bigg[u_0^1 - \log(e^{u_0^1} + \sum_{j\in S_t} e^{u_j})\Bigg]\Bigg\}
\end{aligned}
$$

Because the summation of concave function is still concave, we analyze $-\log(e^{u_0^1} + \sum_{j\in S_t} e^{u_j})$ first. The Hessian matrix of $-\log(e^{u_0^1} + \sum_{j\in S_t} e^{u_j})$ is

$$
\begin{bmatrix}
-\frac{e^{u_0^1}\sum_{j\in S_t} e^{u_j}}{(e^{u_0^1}+\sum_{j\in S_t} e^{u_j})^2} & \frac{e^{u_0^1} e^{u_{j_1}}}{(e^{u_0^1}+\sum_{j\in S_t} e^{u_j})^2} & \cdots & \frac{e^{u_0^1} e^{u_{j_{|S_t|}}}}{(e^{u_0^1}+\sum_{j\in S_t} e^{u_j})^2} \\
\frac{e^{u_0^1} e^{u_{j_1}}}{(e^{u_0^1}+\sum_{j\in S_t} e^{u_j})^2} & -\frac{e^{u_{j_1}}(e^{u_0^1}\sum_{j\in S_t\backslash\{j_1\}} e^{u_j})}{(e^{u_0^1}+\sum_{j\in S_t} e^{u_j})^2} & \cdots & \frac{e^{u_0^1}+e^{u_{j_{|S_t|}}}}{(e^{u_0^1}+\sum_{j\in S_t} e^{u_j})^2} \\
\vdots & \cdots & \ddots & \vdots \\
\frac{e^{u_0^1} e^{u_{j_{|S_t|}}}}{(e^{u_0^1}+\sum_{j\in S_t} e^{u_j})^2} & \frac{e^{u_{j_1}} e^{u_{j_{|S_t|}}}}{(e^{u_0^1}+\sum_{j\in S_t} e^{u_j})^2} & \cdots & -\frac{e^{u_{j_{|S_t|}}}(e^{u_0^1}+\sum_{j\in S_t\backslash\{j_{|S_t|}\}} e^{u_j})}{(e^{u_0^1}+\sum_{j\in S_t} e^{u_j})^2}
\end{bmatrix},
$$

which is a Hermitian diagonally dominant and symmetric matrix with real non-positive diagonal entries. Hence, the Hessian matrix is negative semidefinite, which results in the concavity of $-\log(e^{u_0^1} + \sum_{j\in S_t} e^{u_j})$. By the definition of the concave function, we have, for any $0 \le \lambda \le 1$,

$$
\begin{aligned}
&-\lambda \log(e^{u_0^1} + \sum_{j\in S_t} e^{u_j}) - (1-\lambda)\log(e^{u_0^{1\prime}} + \sum_{j\in S_t} e^{u_j'}) \\
&\le -\log[e^{\lambda u_0^1 + (1-\lambda)u_0^{1\prime}} + \sum_{j\in S_t} e^{\lambda u_j + (1-\lambda)u_j'}].
\end{aligned}
$$

Obviously, the following inequality also holds:

$$
\begin{aligned}
&-\lambda \log(e^{u_0^2+u_{i_t^2}} + \sum_{j\in S_t\backslash\{i_t^2\}} e^{u_j}) - (1-\lambda)\log(e^{u_0^{2\prime}+u_{i_t^2}'} + \sum_{j\in S_t\backslash\{i_t^2\}} e^{u_j'}) \\
&\le -\log\Bigg[e^{\lambda(u_0^2+u_{i_t^2}) + (1-\lambda)(u_0^{2\prime}+u_{i_t^2}')} + \sum_{j\in S_t\backslash\{i_t^2\}} e^{\lambda u_j + (1-\lambda)u_j'}\Bigg].
\end{aligned}
$$

Hence, by the definition of concavity, the function $-\log(e^{u_0^2+u_{i_t^2}} + \sum_{j\in S_t\backslash\{i_t^2\}} e^{u_j})$ is concave, and similarly, the function $-\log(e^{u_0^2+u_{i_t^1}} + \sum_{j\in S_t\backslash\{i_t^1\}} e^{u_j})$ is also concave. Above all, $LLH(\boldsymbol{u})$ is concave. ■

Table 1
Multiple-purchase choice model estimation algorithm.

Estimation algorithm without purchase sequence
Suppose there are total n products.
Require: tolerance level $\gamma > 0$, log-likelihood function $LLH(\cdot)$
Set $m = 0$ and $\boldsymbol{u}_m = [u_i]_{i=0}^N$ where $u_i = 0, \forall i = 0,\dots,N$,
Obtain $\boldsymbol{u}_1$ by solving $LLH(\boldsymbol{u}_m)$ with a random sequence of each transaction with two items
While $\|\|\boldsymbol{u}_{m+1} - \boldsymbol{u}_m\|\|_1 > \gamma$ do
For each transaction with item (k,j)
Calculate $Pr_{\{k,j\}}, Pr_{\{j,k\}}$
Generate random number q,
If $q \le \frac{Pr_{\{k,j\}}}{Pr_{\{k,j\}}+Pr_{\{j,k\}}}$:
Set the choice sequence to be $\{k,j\}$, then update $LLH(\boldsymbol{u}_{m+1})$
else:
Set the choice sequence to be $\{j,k\}$, then update $LLH(\boldsymbol{u}_{m+1})$
Maximize $LLH(\boldsymbol{u}_{m+1})$, then obtain $\boldsymbol{u}_{m+2}$
Set $m = m+1$.

Proposition 1 indicates that we can easily obtain parameters in the MP choice model by maximizing the log-likelihood function $LLH(\boldsymbol{u})$. However, maximization of the log-likelihood function $LLH(\boldsymbol{u})$ requires the purchase sequence, which is difficult to know in reality. In the next section, we present an algorithm to estimate parameters without knowing the purchase sequence.

4.2. Estimation with incomplete purchase information

One of the problems encountered in using the real dataset to estimate customers' utilities based on our MP choice model is that we cannot observe customers' choice sequences in most cases. Ideally, when customers make multiple choices for online shopping, if we have their online click records, we can observe the add-to-cart sequence of the products during the purchase, which can be used to infer customers' choice sequences. However, such data may not be easily obtained in most cases. Therefore, we propose a practical algorithmic approach to solve the sequence problem in the multiple-purchase setting. Our approach is similar to the EM algorithm in that, we regard the purchase sequence as the latent variable in the EM algorithm. The iterative nature of this algorithm allows us to gradually reach the optimal values of customers' utilities and gracefully terminate the search for the optimal values. We define this algorithm as the estimation algorithm without purchase sequence and present the main procedure in Table 1.

Specifically, for customers who have chosen two products, we first assign a random sequence to these two products, solve the corresponding estimation log-likelihood function, then start the iteration. In each iteration, for two items $\{k,j\}$ in each transaction, we calculate $Pr_{\{k,j\}}(Pr_{\{j,k\}})$: the probability of first purchasing product $k(j)$, then product $j(k)$ based on the utility vector $\boldsymbol{u}$ obtained from the last iteration. The choice sequence of each transaction with product k and j is assumed to be $\{k,j\}$ with probability $\frac{Pr_{\{k,j\}}}{Pr_{\{k,j\}}+Pr_{\{j,k\}}}$, or $\{j,k\}$ with probability $\frac{Pr_{\{j,k\}}}{Pr_{\{k,j\}}+Pr_{\{j,k\}}}$. Based on the realized choice sequence, we implement the estimation procedure (based on Section 4), and obtain the new utility vector $\boldsymbol{u}$. Then, we calculate the gap between the estimated utility vectors obtained from two adjacent iterations. The gap is defined as the summation of differences in the absolute values of the estimated utilities of all products and the no-purchase choices. Finally, as the gap narrows, we can stop the iteration when it is below the tolerance level $\gamma = 0.001$ and obtain the final utility vector for the MP choice model.

Table 2
The expected revenue of offer set $\mathcal{S}$.

$\mathcal{S}$	$\pi^m_{\mathcal{S}}$	$\pi^s_{\mathcal{S}}$
{1}	3.950	3.950
{2}	3.636	3.636
{3}	1.950	1.950
{1,2}	4.796	3.992
{1,3}	5.721	3.933
{2,3}	4.206	3.658
{1,2,3}	5.212	3.985

Table 3
The expected revenue of offer set $\mathcal{S}$.

$\mathcal{S}$	$\pi^m_{\mathcal{S}}$	$\pi^s_{\mathcal{S}}$
{1}	3.950	3.950
{2}	3.810	3.810
{3}	2.000	2.000
{1,2}	4.454	3.995
{1,3}	5.769	3.966
{2,3}	4.135	3.818
{1,2,3}	4.730	3.996

Table 4
The expected revenue of offer set $\mathcal{S}$.

$\mathcal{S}$	$\pi^m_{\mathcal{S}}$	$\pi^s_{\mathcal{S}}$
{1}	3.950	3.950
{2}	3.333	3.333
{3}	1.333	1.333
{1,2}	5.253	3.986
{1,3}	4.449	2.975
{2,3}	3.988	3.000
{1,2,3}	5.342	3.544

5. Multiple-purchase assortment optimization problem

After obtaining the parameters in the choice model, we focus on solving the corresponding assortment problem in this section. In the assortment optimization problem based on the MNL choice model, the optimal assortment has been proven to be revenue-ordered, which means that the optimal assortment includes a certain number of products with the highest revenues (e.g Talluri and Van Ryzin, 2004; Rusmevichientong et al., 2014). Unfortunately, such an excellent property does not hold in the optimal assortment when considering multiple-purchase behavior. Intuitively, the optimal assortment under the MP choice model may be more extensive than the revenue-ordered assortment due to less cannibalization between products. However, we identify that there is no inclusion between the optimal assortment under the MP choice model and revenue-ordered assortment.

Example 1. Consider an instance of three products. The revenue of products is $[r_1, r_2, r_3] = [7.9, 4, 3.9]$. The corresponding $\boldsymbol{v}$ is $[v_1, v_2, v_3] = [1, 10, 1]$. We normalize the utility of the first no-purchase option to 0, that is, $u_0^1 = 0$, thus $v_0^1 = 1$, and then suppose that the fixed part of the utility of the second no-purchase option is arbitrarily given by $v_0^2 = 1.2$. Table 2 shows the expected revenue for each assortment $\mathcal{S}$ in the multiple-purchase case ($\pi^m_{\mathcal{S}}$) and the single-purchase case based on the MNL choice model ($\pi^s_{\mathcal{S}}$).

As shown in Table 2, the assortment $\mathcal{S}_s^* = \{1, 2\}$ maximizes the expected revenue in the MNL case, whereas in the multiple-purchase case, the optimal assortment is $\mathcal{S}_m^* = \{1, 3\}$. As this assortment skips over the second product, it is not revenue-ordered. In addition, the optimal assortment does not contain the optimal assortment obtained in the single-purchase case based on the MNL choice model, i.e., $\mathcal{S}_s^* \nsubseteq \mathcal{S}_m^*$. In the following two examples, we show that the optimal assortment based on the MP choice model can be the subset of that based on the MNL choice model, and vice versa.

Example 2. Consider an instance of three products. The revenue of the products is $[r_1, r_2, r_3] = [7.9, 4, 3.999]$, and the corresponding $[v_1, v_2, v_3]$ is $[1, 20, 1]$. The utility of the first no-purchase option is normalized to 0, that is, $v_0^1 = 1$. Then, we suppose that the fixed part of the utility of the second no-purchase option is arbitrarily given by $v_0^2 = 1.2$. Table 3 shows the revenue for each assortment $\mathcal{S}$ in the multiple-purchase case ($\pi^m_{\mathcal{S}}$) and the single-purchase case under the MNL choice model ($\pi^s_{\mathcal{S}}$).

We observe that the optimal assortment based on the MNL choice model is $\mathcal{S}_s^* = \{1, 2, 3\}$, while the optimal assortment is $\mathcal{S}_m^* = \{1, 3\}$ in the multiple-purchase case, which is a subset of the optimal assortment based on the MNL choice model, i.e., $\mathcal{S}_m^* \subseteq \mathcal{S}_s^*$. Notably, the results in Example 3 are the opposite of those obtained in Example 2.

Example 3. Consider an instance of three products. The revenue of the products is $[r_1, r_2, r_3] = [7.9, 4, 2]$, and the corresponding $[v_1, v_2, v_3]$ is $[1, 5, 2]$. We normalize the utility of the first no-purchase option to 0, that is, $v_0^1 = 1$. Then, we suppose that the fixed part of utility of the second no-purchase option is arbitrarily given by $v_0^2 = 1.2$. Table 4 shows the revenue for each assortment $\mathcal{S}$ in the multiple-purchase case ($\pi^m_{\mathcal{S}}$) and the single-purchase case under the MNL choice model ($\pi^s_{\mathcal{S}}$).

Note that, Examples 1–3 show that the optimal assortment based on MP choice model has no obvious connections with the optimal assortment obtained from the MNL choice model. The important insight here is that using MNL choice model for simplicity to capture customers' multiple choice behavior in reality may lead to sub-optimal results, especially in scenarios where customers always make multiple purchases, such as clothes or accessories. Therefore, the study of a MP choice model is necessary to capture customers' real choice behaviors.

Motivated by Examples 1–3, we focus on finding the optimal assortment when considering the multiple-purchase behavior. In the following, we reformulate the optimal assortment problem based on the MP choice model as a mixed-integer program, which can be solved by state-of-the-art solvers, such as CPLEX and GUROBI, and is also commonly used to solve assortment problems, such as Çömez-Dolgan et al. (2022), Vaagen et al. (2011).

Proposition 2. *Model $[MP]$ can be reformulated as a mixed-integer linear program $[MLP]$.*

$$[MLP] \max \sum_{i\in[N]} v_i r_i \rho_i + \sum_{k\in[N]} \sum_{i\in[N]\backslash\{k\}} v_i r_i v_k g_{ik}^k$$

$$s.t\ v_0^1 \pi + \sum_{i'\in[N]} v_{i'} \rho_{i'} = 1 \tag{3}$$

$$v_0^1(v_0^2-1)v_k h_k^k + \sum_{j'\in[N]} \Big(v_0^1 v_{j'} h_{j'}^k + (v_0^2-1) v_k v_{j'} g_{kj'}^k \Big) + \sum_{i',j'\in[N]} v_{i'} v_{j'} g_{i'j'}^k = 1 \qquad \forall k \in [N] \tag{4}$$

$$\begin{aligned}
&g_{i'j'}^k \le h_{i'}^k && \forall i', j' \in [N], k \in [N]\\
&g_{i'j'}^k \le h_{j'}^k && \forall i', j' \in [N], k \in [N]\\
&g_{i'j'}^k \ge h_{i'}^k + h_{j'}^k - w^k && \forall i', j' \in [N], k \in [N]\\
&0 \le g_{i'j'}^k \le w^k && \forall i', j' \in [N], k \in [N]\\
&x_i \in \{0, 1\} && \forall i \in [N]
\end{aligned} \tag{5}$$

$$\begin{aligned}
&\pi \ge 0\\
&w^k \ge 0 && \forall k \in [N]\\
&\rho_i \le \pi && \forall i \in [N]\\
&\rho_i \le M x_i && \forall i \in [N]\\
&\rho_i \ge \pi - M(1 - x_i) && \forall i \in [N]
\end{aligned}$$

$$
\begin{aligned}
&\max_{S} \quad \sum_{i\in S} r_i \cdot \frac{v_i}{v_0^1 + \sum_{i'\in S} v_{i'}} \cdot \Big(1 + \sum_{k\in S\setminus\{i\}} \frac{v_k}{v_0^2 v_k + \sum_{j'\in S\setminus\{k\}} v_{j'}}\Big) \\
=& \max_{x\in\{0,1\}^N} \sum_{i\in[N]} \frac{v_i r_i x_i}{v_0^1 + \sum_{i'\in[N]} v_{i'} x_{i'}} \cdot \Big(1 + \sum_{k\in[N]\setminus\{i\}} \frac{v_k x_k}{(v_0^2-1)v_k x_k + \sum_{j'\in[N]} v_{j'} x_{j'}}\Big) \\
=& \max_{x\in\{0,1\}^N} \sum_{i\in[N]} \frac{v_i r_i x_i}{v_0^1 + \sum_{i'\in[N]} v_{i'} x_{i'}} + \\
&\sum_{i\in[N]} \sum_{k\in[N]\setminus\{i\}} \frac{v_i r_i v_k x_i x_k}{v_0^1(v_0^2-1)v_k x_k + \sum_{j'\in[N]}(v_0^1 v_{j'} x_{j'} + (v_0^2-1)v_k v_{j'} x_{j'} x_k) + \sum_{i',j'\in[N]} v_{i'} v_{j'} x_{i'} x_{j'}} \\
=& \max_{x\in\{0,1\}^N, y\in\mathcal{Y}} \sum_{i\in[N]} \frac{v_i r_i x_i}{v_0^1 + \sum_{i'\in[N]} v_{i'} x_{i'}} + \\
&\sum_{i\in[N]} \sum_{k\in[N]\setminus\{i\}} \frac{v_i r_i v_k y_{ik}}{v_0^1(v_0^2-1)v_k x_k + \sum_{j'\in[N]}(v_0^1 v_{j'} x_{j'} + (v_0^2-1)v_k v_{j'} y_{kj'}) + \sum_{i',j'\in[N]} v_{i'} v_{j'} y_{i'j'}} \\
=& \max_{x\in\{0,1\}^N, y\in\mathcal{Y}} \sum_{i\in[N]} \frac{v_i r_i x_i}{v_0^1 + \sum_{i'\in[N]} v_{i'} x_{i'}} + \\
&\sum_{k\in[N]} \sum_{i\in[N]\setminus\{k\}} \frac{v_i r_i v_k y_{ik}}{v_0^1(v_0^2-1)v_k x_k + \sum_{j'\in[N]}(v_0^1 v_{j'} x_{j'} + (v_0^2-1)v_k v_{j'} y_{kj'}) + \sum_{i',j'\in[N]} v_{i'} v_{j'} y_{i'j'}}
\end{aligned}
$$

Box I.

$$
\begin{aligned}
&\rho_i \ge 0 && \forall i\in[N] \\
&h_j^k \le w^k && \forall k\in[N], j\in[N] \\
&h_j^k \le M x_j && \forall k\in[N], j\in[N] \\
&h_j^k \ge w^k - M(1-x_j) && \forall k\in[N], j\in[N] \\
&h_j^k \ge 0 && \forall k\in[N], j\in[N]
\end{aligned}
$$

Constraints (3) and (4) are derived from the transformation of the fractional objective function. Constraint (5) indicates whether product i is chosen. The remaining constraints are introduced to ensure that the program can be formulated as a mixed-integer linear program.

Proof of Proposition 2. From the original multiple-purchase assortment problem $[MP]$, we have the equation in Box I where

$$
\mathcal{Y} := \left\{ y \;\middle|\; \begin{array}{ll} y_{ij} \le x_i & \forall i,j\in[N] \\ y_{ij} \le x_j & \forall i,j\in[N] \\ y_{ij} \ge x_i + x_j - 1 & \forall i,j\in[N] \\ 0 \le y_{ij} \le 1 & \forall i,j\in[N] \end{array} \right\}, \tag{6}
$$

Note that y_{ij} is a continuous variable, but the value of y_{ij} can only be 0 or 1.

The first equality holds because we use x to define the subset of $[N]$. The second equality is the result of simple computation. When we introduce the new variables $y\in\mathcal{Y}$, the third equality holds. Obviously, the last equality holds.

Note that above program is a mixed-integer linear fractional program, where the objective function is the ratio of two linear functions and all constraints are linear. According to the results in Yue et al. (2013), we can reformulate it into an equivalent mixed-integer linear program. To this end, we introduce new decision variables,

$$
\begin{aligned}
&\pi = \frac{1}{v_0^1 + \sum_{i'\in[N]} v_{i'} x_{i'}} \\
&w^k = \frac{1}{v_0^1(v_0^2-1)v_k x_k + \sum_{j'\in[N]}(v_0^1 v_{j'} x_{j'} + (v_0^2-1)v_k v_{j'} y_{kj'}) + \sum_{i',j'\in[N]} v_{i'} v_{j'} y_{i'j'}} \quad \forall k\in[N] \\
&g_{ij}^k = y_{ij} w^k \qquad \forall i,j\in[N], k\in[N]
\end{aligned}
$$

Then, for each ratio of two linear functions in the objective function, we replace the new variables and add two sets of linear constraints, i.e., constraints (7) and (8).

$$
\begin{aligned}
\max \quad & \sum_{i\in[N]} v_i r_i \pi x_i + \sum_{k\in[N]} \sum_{i\in[N]\setminus\{k\}} v_i r_i v_k g_{ik}^k \\
\text{s.t } & v_0^1 \pi + \sum_{i'\in[N]} v_{i'} x_{i'} \pi = 1 \qquad (7) \\
& v_0^1(v_0^2-1)v_k w^k x_k \\
& + \sum_{j'\in[N]} \left(v_0^1 v_{j'} x_{j'} w^k + (v_0^2-1) v_k v_{j'} g_{kj'}^k \right) \\
& + \sum_{i',j'\in[N]} v_{i'} v_{j'} g_{i'j'}^k = 1 \qquad \forall k\in[N] \qquad (8) \\
& g_{i'j'}^k = y_{i'j'} w^k \qquad \forall i',j'\in[N], k\in[N] \qquad (9) \\
& y_{i'j'} = x_{i'} x_{j'} \qquad \forall i',j'\in[N] \qquad (10) \\
& x_i \in \{0,1\} \qquad \forall i\in[N] \\
& \pi \ge 0 \\
& w^k \ge 0 \qquad \forall k\in[N] \\
& g_{i'j'}^k \ge 0 \qquad \forall i',j'\in[N], k\in[N]
\end{aligned}
$$

Note that $\boldsymbol{y} \in \mathcal{Y}$ implies Constraint (10). Then, we can replace Constraint (9) and (10) by following constraint set

$$
\begin{aligned}
&g_{i'j'}^k \le x_{i'} w^k && \forall i',j'\in[N], k\in[N] \\
&g_{i'j'}^k \le x_{j'} w^k && \forall i',j'\in[N], k\in[N] \\
&g_{i'j'}^k \ge x_{i'} w^k + x_{j'} w^k - w^k && \forall i',j'\in[N], k\in[N] \\
&0 \le g_{i'j'}^k \le w^k && \forall i',j'\in[N], k\in[N]
\end{aligned}
$$

Therefore, the nonlinear terms in the program are $x_{i'}\pi$ and $w^k x_j$. By applying the big-M method, we obtain the $[MLP]$ model. ■

6. Numerical experiments

In this section, we perform numerical experiments to illustrate the superiority of the multiple-purchase choice model. Our experiments are conducted on both real dataset and simulated data. In Section 6.1, we run numerical experiments on the real datasets. Since the true choice

Table 5
Customer purchases statistics.

	No purchase	Single purchase	Two purchase	More than two purchase
Raw data	52.91%	36.27%	8.61%	2.21%
Final data	52.91%	36.27%	10.82%	

Table 6
Assortment statistics for each store.

Store num	Assortment num	Assortment size	
		Min	Max
1	7	2	12
2	6	2	12
3	5	2	12
4	6	2	12
5	8	3	12

model is unknown in reality, we focus on comparing the estimation performance of the MP choice model and the MNL choice model. The numerical results demonstrate that our MP choice model can achieve better estimation results. As the MDCEV model, despite its complexity when applied to assortment problems, has the capability to accommodate customers' purchases of multiple products within the assortment set. To provide a comprehensive assessment of our multiple-purchase model's performance, we also compare its estimation performance with that of the MDCEV model and present the results in Appendix.

In Section 6.2, we conduct numerical experiments on the simulated data to test the performance of the corresponding assortment problem under the MP choice model and MNL model. Under various parameter settings, we examine the performance of the solutions to the assortment problem under two choice models. The results show that assortment solutions based on the MP choice model can obtain a significant revenue improvement compared with that obtained by the solutions to the assortment problem based on the MNL model.

6.1. Performance on the real dataset

In this session, we examine the performance of the MP choice model on a real-world dataset. The dataset is provided by a retailer that sells multiple categories of products, such as snacks and groceries. We choose the snack category as the focal category to form the final dataset and split the dataset into training and test datasets. The ratio between the training and test datasets is 60 : 40. Then, we implement the estimation processes on the training dataset and compare the performance of the MP choice model and the MNL choice model on the test dataset. Specifically, we first estimate the utilities of products in the MP choice model based on the estimation algorithm proposed in Section 4.2, and initially compare the performance of the two models using the in-sample data and evaluate their fit using the AIC and BIC criteria. However, it is essential to recognize that relying solely on in-sample testing may not provide a reliable assessment of the models' true predictive ability on new data. To address this limitation and achieve a more comprehensive evaluation, we conduct out-of-sample testing. The empirical evidence derived from out-of-sample forecast performance holds higher reliability compared to evidence based on in-sample performance. Out-of-sample forecasts possess greater credibility due to their reduced sensitivity to outliers and decreased susceptibility to data mining issues. Moreover, out-of-sample forecasts provide a more accurate depiction of the information available to forecasters in real-time scenarios. Therefore, to effectively compare the performance of the MP choice model and the MNL choice model, we construct four metrics based on the out-of-sample data.

6.1.1. Data structure

The complete dataset has customer transactions from September 2017 to August 2018. We restrict our analysis to the dried fruits subcategory, which belongs to the snack category. There are 12 different products in this subcategory. In the following analysis, we estimate customers' utilities of these 12 products and compare the estimation performance of the MP choice model with the MNL choice model.

We use the following steps to formulate the final dataset. First, since we only have the information of successful transactions of customers, we cannot observe customer interactions that did not lead to sales. However, we observe customers who have purchased other products instead of dried fruits in the snack category. We posit that customers who have purchased snacks actually consist of the whole population of real target customers of the dried fruits. Therefore, it is reasonable to assume that customers who have purchased other subcategories in the snack category are the no-purchase customers in this setting.

Second, we observe the multiple-purchase behavior of customers in this dataset. Our dataset finds that 8.61% of customers purchase two products and 2.21% of customers buy more than two products. More details are shown in the "Raw data" row in Table 5. Since the customers who purchase more than two items are limited, we focus on the two-item orders in this research. For customers who have purchased more than two products in the dried fruits subcategory, we randomly select two of those products to form the two-item transactions of each customer. Then, we construct the final dataset. The statistics regarding customer purchases are shown in the "Final data" row in Table 5.

Third, five chain stores of this retailer are operating during the time range represented in our dataset, and each of them may provide different assortments of dried fruits to customers at different times. We cannot explicitly observe the assortment provided to each customer at the time of the purchase. Therefore, to derive the dynamic assortment provided to each customer, we calculate the date of the first transaction of each product in the dried fruit subcategory in each store and regard the date seven days prior to this date as the launch time of the product. Thus, we allow different assortments within or between each chain store in our dataset. As shown in Table 6, each store has 5 to 8 different numbers of assortments provided to customers. The minimum assortment size is 2 in four stores, store 5 has a minimum assortment size of 3, and the maximum assortment size is 12 in all five stores.

Finally, for each store and each day we randomly select 60% of transactions as our in-sample data. We use the in-sample data to estimate the utilities of 12 products, as well as the no-purchase utility. Then, based on the estimated utilities, we construct four metrics to compare the performance of the MP choice model and the MNL choice model on the out-of-sample data (the remaining 40% of transactions). The entire process of sample-estimation and comparison is repeated 10 times, and we use the average value of 10 sampling results as the final indicator to verify the performance of the MP choice model.

6.1.2. Comparison results

We commence by presenting the in-sample comparison results of the two models, as shown in Table 7. The AIC (BIC) gap represents the disparity between the AIC (BIC) values of the multiple-choice model and the AIC (BIC) values of the MNL choice model. Positive values in Table 7 indicate that the MNL choice model outperforms the multiple-purchase choice model in terms of the in-sample tests.

We notice that this comparison only suggests that the MNL choice model demonstrates a better fit to the in-sample data compared to the multiple-purchase choice model. However, our main concern typically

Table 7
Comparison results of two models using AIC and BIC.

Instance	1	2	3	4	5	6	7	8	9	10	Average gap
AIC Gap	7.55%	7.42%	7.48%	7.45%	7.44%	7.53%	7.44%	7.53%	7.57%	7.53%	7.49%
BIC Gap	7.55%	7.42%	7.47%	7.45%	7.44%	7.53%	7.44%	7.53%	7.57%	7.53%	7.49%

Table 8
Gap in the out-of-sample results from three measures.

Instance	WR^2			$WRMSE$			$HRMSE$			WR		
	MP	MNL	Gap	MP	MNL	Gap	MP	MNL	Gap	MP	MNL	Gap
1	0.7456	0.7105	4.95%	0.0046	0.0051	9.42%	0.3385	0.3401	0.47%	6688.52	7301.86	8.40%
2	0.7092	0.6701	5.84%	0.0047	0.0053	11.94%	0.3390	0.3406	0.48%	6792.81	7385.81	8.03%
3	0.7435	0.7069	5.17%	0.0046	0.0051	9.22%	0.3396	0.3412	0.46%	6756.02	7356.12	8.16%
4	0.7310	0.6912	5.77%	0.0046	0.0052	12.10%	0.3390	0.3407	0.47%	6742.93	7342.58	8.17%
5	0.7053	0.6754	4.43%	0.0049	0.0053	8.28%	0.3394	0.3410	0.47%	6713.46	7310.06	8.16%
6	0.7309	0.6962	4.98%	0.0047	0.0052	8.94%	0.3389	0.3405	0.47%	6713.43	7328.40	8.39%
7	0.7481	0.7104	5.30%	0.0045	0.0050	10.52%	0.3385	0.3401	0.47%	6755.06	7348.72	8.08%
8	0.7519	0.7172	4.84%	0.0044	0.0050	11.45%	0.3392	0.3409	0.49%	6664.18	7282.39	8.49%
9	0.7482	0.7129	4.95%	0.0045	0.0050	9.31%	0.3398	0.3413	0.46%	6732.31	7342.57	8.31%
10	0.7402	0.7038	5.18%	0.0045	0.0051	10.79%	0.3389	0.3405	0.48%	6682.91	7294.20	8.38%
Average	0.7354	0.6994	5.14%	0.0046	0.0051	10.20%	0.3391	0.3407	0.47%	6724.16	7329.27	8.26%

revolves around evaluating the models' performance on new data. Therefore, we employ the following four metrics in our numerical experiments to test the out-of-sample performances of these two models: (1) the aggregate level measures: weighted R square (WR^2) and (2) weighted root mean square error ($WRMSE$); (3) the customer level hard root mean square error ($HRMSE$), based on the measure developed in Berbeglia et al. (2021); (4) the revenue-related measure: weighted revenue (WR).

- WR^2: The aggregate level measures can be compared at the product level. Specifically, we construct the first measure WR^2 based on the concept R^2 in the statistics used to represent the proportion of the dependent variable that can be explained by all the variables used in a regression model. Consider a specific choice model, with θ as the observed data and ϕ as the estimation. R^2 in our scenario can be calculated as follows:

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_{j\in S}(P_\theta(j|S) - P_\phi(j|S))^2}{\sum_{j\in S}(P_\theta(j|S) - \overline{P}_\theta(j|S))^2}$$

where j represents product j, S indicates the assortment. Given assortment S, $P_\theta(j|S)$ and $P_\phi(j|S)$ are the observed and estimated choice probability of product j, respectively. We also denote $\overline{P}_\theta(j|S)$ as the sample mean of $P_\theta(j|S)$. RSS and TSS are the residual sum of squares and the total sum of squares, respectively. Since stores offer different assortments, WR^2 is the weighted average of R^2 with each assortment proportionately weighted by the number of transactions placed under this assortment. Similarly, WR^2 can represent the percentage of the variation in observed-customer choice probabilities that can be explained by the estimations of the choice model. A higher value of WR^2 infers a higher explanation power of the choice model.
The average gap of WR^2 between the MP choice model and the MNL choice model is shown in WR^2 column in Table 8. We observe that the MP choice model always outperforms the MNL choice model in all randomly chosen data samples. The average gap is 5.14%, indicating that the MP choice model usually has higher explanatory power than the MNL choice model in explaining customers' choice behavior if multiple purchases exist.
- $WRMSE$: The second measure is the weighted root mean square error ($WRMSE$). Suppose a total of M different assortments, and a total of T customers who have made purchases exists. Then, the $WRMSE$ is calculated as follows:

$$WRMSE = \sum_{m=1}^{M} \frac{T_{S_m}}{T} \cdot \sqrt{\frac{\sum_{j\in S_m}(P_\theta(j|S_m) - P_\phi(j|S_m))^2}{|S_m|}}$$

where T_{S_m} is the number of customers who have made purchase if we offer assortment S_m. Note that this measure can also be implemented on the aggregate data at the product level and allows variations in assortments provided to customers.
Table 8 shows the results of $WRMSE$ for the MP choice model and the MNL choice model, where the smaller value of $WRMSE$ represents a higher accuracy of the estimation result. Based on the results, we can conclude that MP choice model outperforms the MNL choice model in all cases, with a 10.20% average gap of $WRMSE$.
- $HRMSE$: We also construct a similar measure to that in Berbeglia et al. (2021) using customer-level data: the customer level hard root mean square error ($HRMSE$). As mentioned in Berbeglia et al. (2021), $HRMSE$ is calculated as follows:

$$HRMSE = \sqrt{\frac{\sum_{t=1}^{T}\sum_{j\in S_t}(I(j_t = j) - P_\phi(j|S_t))^2}{\sum_{t=1}^{T}|S_t|}}$$

Table 8 shows the results by using $HRMSE$. The MP choice model still has a performance advantage, with lower values of $HRMSE$ than that of the MNL choice model. A surprising result is that the average gap is only 0.47%, which seems to be smaller than we expected. However, the small gap is consistent with estimation results in the latest choice model survey (Berbeglia et al., 2021). We note that the largest gap between the best and the worst model using the hotel data in Berbeglia et al. (2021) is only 0.95%, with the smallest gap being 0.34%. Moreover, we consider a larger assortment size (12) than the assortment size (10) in Berbeglia et al. (2021). The estimation results may be worse when enlarging the offer sets (Berbeglia et al., 2021).
- WR: Note that the probability-related and revenue-related measures may not be equivalent for products with different prices. A more significant gap in $WRMSE$ may not mean a larger revenue gap. Therefore, we also construct the revenue-related measure to assess the performance of the MP choice model, as the ultimate goal of the assortment problem is maximizing revenue. Then, the measure weighted revenue (WR) is calculated as follows:

$$WR = \sum_{m=1}^{M} \frac{T_{S_m}}{T} \cdot \left| R_\theta(S_m) - T_{S_m} \sum_{j\in S_m} r_j P_\phi(j|S_m) \right|$$

Given assortment S_m, $R_\theta(S_m)$ and $\sum_{j\in S_m} r_j P_\phi(j|S_m)$ are the realized revenue and expected revenue calculated by the choice model, respectively. Since assortment S_m has been provided T_{S_m}

Table 9
Revenue gap between multiple-purchase model and MNL model using simulation data.

$\|S\|$	u_0^2					$\underline{u}$					$\overline{u}$				N		
	−2	−1	0	1	2	−5	−4	−3	−2	−1	2	3	4	5	15	25	30
2	0.17	0.07	0.12	0.05	0.03	0.13	0.09	0.07	0.03	0.04	0.03	0.02	0.07	0.11	0.07	0.11	0.06
4	0.10	0.11	0.16	0.12	0.05	0.24	0.22	0.10	0.16	0.08	0.04	0.04	0.11	0.16	0.16	0.16	0.13
6	0.05	0.11	0.14	0.13	0.08	0.26	0.23	0.20	0.12	0.12	0.04	0.06	0.16	0.12	0.13	0.16	0.17
8	0.10	0.16	0.14	0.11	0.07	0.20	0.22	0.14	0.15	0.12	0.03	0.06	0.06	0.16	0.11	0.17	0.12
10	0.04	0.05	0.08	0.12	0.07	0.12	0.15	0.09	0.12	0.09	0.02	0.05	0.06	0.12	0.04	0.09	0.11
Free	0.16	0.22	0.14	0.12	0.10	0.16	0.19	0.26	0.20	0.12	0.07	0.06	0.15	0.20	0.19	0.19	0.18

times, $T_{S_m} \sum_{j\in S_m} r_j P_\phi(j|S_m)$ is the total expected revenue under assortment S_m. This revenue-related measure represents the weighted difference between the actual revenue calculated from the observed data and the expected revenue calculated by the choice model. Table 8 shows the performances of the MP choice model and the MNL choice model, in terms of WR. We can observe that the multiple-purchase choice model still outperforms the MNL choice model, with an average gap of 8.26%.

6.2. Performance on the simulation data

In this section, we create a series of scenarios with combinations of different parameter values to test our model's performance. Using the simulation data, we compare the realized revenues of the MP choice model and the widely used MNL choice model with different parameter settings, which may correspond to different selling modes or categories. Specifically, for each scenario with one specific parameter setting, we generate 20 sample paths with 2000 customers (transactions) for each sample path. Then, for each sample path, we estimate the parameters of these two choice models and solve the corresponding assortment problems. Finally, to assess the performance of these two models, we implement the solutions to the assortment problem under MNL and MP choice model on the out-of-sample test set, which consists of 3000 new customers (transactions), and compare the average revenue of the 20 sample paths obtained from these two solutions.

6.2.1. Simulation process

We first provide a basic parameter setting, which is also termed as base scenario below. We assume that consumers purchase at most two items in a single shopping visit, which means that each transaction has two items at most. Suppose there are $N = 20$ products, and $K = 2000$ transactions for a sample path in the base scenario. We assume the mean utilities are generated from the uniform distribution on $[\underline{u}, \overline{u}]^N$ where $\underline{u} = -2$ and $\overline{u} = 5$. Similarly, we choose the product price from the uniform distribution on $[10, 30]^N$. In addition, the no-purchase utility of the first choice u_0^1 is 0, and the fixed part of the outside no-purchase utility of the second choice u_0^2 is 0.5. To generate the in-sample transaction data, we assume that each customer (transaction) first chooses the first item and then the second item according to the choice probability of the MP choice model. To observe the substitution patterns between products and obtain more accurate estimation results, we assume each customer chooses products from a randomly generated offer set S from N products. We consider 6 different offer set size, i.e., $|S| \in \{2, 4, 6, 8, 10, Free\}$. Note that $Free$ means that we do not have any restrictions on the size of the assortment.

The generated transaction data include customer ID and the first-choice and second-choice products. No purchase also be included in the transaction data. Since we have proven the log-likelihood function is concave in Section 4, we can easily use the maximum likelihood function to estimate the parameters of the MP choice model. Next, we implement the following strategy to obtain the estimation result of the MNL model. Suppose we do not know that customers make multiple purchases. Thus, we may assume that all customers choose only one product. Based on this assumption, for a transaction with two products, we treat them as two transactions with the same customer ID (Feldman et al., 2018). Note that in the transaction that includes the second purchased product, we assume the assortment for the second choice has excluded the first choice product, as we can observe customers' choice sequence of two items in the order while using simulation data, considering the choice sequence might improve the performance of the MNL choice model. We do not need to perform the transformation for customers who choose one product. As a result, for the MNL case, the number of data points (transactions) is more than 2000 but less than $2000 \times 2 = 4000$.

Given the estimation results, we solve the assortment problems based on the MP choice model and the MNL model, and obtain the optimal solutions to these two optimization problems. To assess the performance, we generate 3000 out-of-sample transactions and compute the revenue gap between these two optimal assortments, which is equal to the increase in the revenue obtained from the MP assortment solution in comparison to the revenue obtained from the MNL assortment solution.

To check the robustness of our results, we vary the values of different parameters in the data generation process, such as the total number of products $N \in \{15, 20, 25, 30\}$, the fixed part of the no-purchase utility in the second choice $u_0^2 \in \{-2, -1, 0, 1, 2\}$, the lower bound of the mean utility $\underline{u} \in \{-5, -4, -3, -2, 1\}$ and the upper bound of the mean utility $\overline{u} \in \{2, 3, 4, 5\}$.

6.2.2. Comparison results

- The impact of u_0^2: We compare the results with different values of u_0^2, while keeping other parameters the same as the base scenario. The revenue gap is presented in the column u_0^2 in Table 9. We observe the excellent performance of the MP choice model compared with the MNL choice model. Our main finding here is that with a significantly high value of u_0^2, which may correspond to the specific market where customers seldom purchase multiple items in a single shopping trip, the MP choice model continues to outperform the MNL choice model.
- The impact of u: Then, we show the impact of the mean utility u. In the base scenario, the mean utility is drawn between −2 and 5. Given the same $\overline{u}$, the small value of $\underline{u}$ may represent a low degree of similarity among products. Given the same $\underline{u}$, the small value of $\overline{u}$ may represent a high degree of similarity among products. Here, we study the performance of these two choice models when confronted with different degrees of similarity. The $\underline{u}$ and $\overline{u}$ column in Table 9 illustrates the numerical results by varying the value of $\underline{u}$, or varying the value of $\overline{u}$. The main finding is that with a relatively low degree of similarity(i.e., $\underline{u}$ is smaller or $\overline{u}$ is larger), the MP choice model performs much better than the MNL model.
- The impact of N: Next, we show the model performance under different number of products N, also keeping other parameters the same as the base scenario. The N column in Table 9 summarizes the results of the corresponding revenue gaps of these two choice models. We observe that the MP choice model also has an obvious performance advantage over the MNL choice model, especially for the scenarios with no constraints in the assortment size.

In conclusion, our numerical experiments show that considering the multiple-purchase behavior of customers is beneficial, as customers are likely to purchase multiple items in a variety of shopping scenarios.

7. Conclusion

This paper presents a MP choice model that incorporates customers' multiple-purchase behavior and studies the corresponding estimation and assortment optimization problem. We show that the optimal assortment in the multiple-purchase case is no longer revenue by order. After estimating the parameters of the multiple-choice model by using MLE, we develop an equivalent mixed-integer program for the assortment optimization problem. Moreover, we conduct extensive experiments to examine the performance of the MP choice model and compare the results with the MNL choice model. We first use a real dataset to assess the performance of the MP choice model in terms of the accuracy of estimated probabilities and revenues. Using four performance metrics, we conclude that the MP choice model has better estimation performance in terms of purchase probabilities and total revenues than the MNL choice model. The numerical results on the simulated data also show that the revenue obtained from the MP choice model is much higher than that from the single-purchase choice model.

We see several possible directions worthy of future research. One is to develop more efficient algorithms for solving the MP assortment problem in a large-scale scenario. Instead of using MNL as the based choice model, another extension is to characterize the multiple-purchase behavior based on other choice models, machine learning techniques (Peng et al., 2022) or empirical analysis (Feng et al., 2022a). Besides, it is worth noting that the quantities of products chosen by customers can also play a significant role in understanding their behavior. In this study, our focus is primarily on customers purchasing multiple different products rather than multiple units of the same product. We recommend that future studies address this interesting topic. Furthermore, future studies may incorporate customers' strategic behavior into the multiple-purchase choice model, such as customers may make choices for products sequentially, which could have an impact on their overall decision-making process (Chen et al., 2023; Feng and Wang, 2023). Finally, although we use a real dataset to assess the performance of the multiple-purchase choice model, we cannot observe the implementation of the optimal assortment obtained by the multiple-purchase assortment problem. Future research may consider conducting field experiments to compare the performances of the optimal assortments based on different choice models.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgment

We sincerely thank the editors and two anonymous reviewers for helpful and constructive comments. We also would like to thank Professor Ying Rong and Professor Huan Zheng for their insightful comments on the early version of this paper. This research has been supported by grants from the National Natural Science Foundation of China (Grant number: 72025201 and 72102194).

Appendix. Results of the MDCEV model and multiple-purchase model comparison

As mentioned in Section 2, while the MDCEV model may not specifically address the assortment optimization problem, we aim to provide a comprehensive understanding of our multiple-purchase choice model and highlight its key differences from the MDCEV model. To achieve this, we compare the estimation performance between the MDCEV model and our multiple-purchase choice model using the real dataset. This comparison allows us to assess their respective strengths and characteristics.

To better understand the limitations of the MDC-type model in addressing the assortment optimization problem, let us begin by considering the MDC-type model for an assortment or offer set $\mathcal{S}$, including K items. The corresponding consumption quantity for each item is denoted as $\boldsymbol{q}$, and the utility function is expressed as follows:

$$U(\boldsymbol{q}) = \sum_{k=1}^{K} \frac{\gamma_k}{\alpha_k} \psi_k \left\{ \left(\frac{q_k}{\gamma_k} + 1 \right)^{\alpha_k} - 1 \right\}$$

Here, γ_k, ψ_k, and $\alpha_k, \forall k = 1, \ldots, K$, need to be estimated from the data. With a given assortment or product offer set $\mathcal{S}$, the customer aims to maximize the utility function $U(\boldsymbol{q})$ while adhering to a budget constraint $\sum_{k=1}^{K} e_k q_k = E$, where e_k represents the unit cost of product k and E denotes the budget. The analyst can solve for the optimal expenditure allocations by forming the Lagrangian and applying the Kuhn–Tucker (KT) conditions. Additionally, the analyst can estimate parameters in the utility function.

The MDCEV model indicates that each assortment has a specific estimation for the parameters γ_k, ψ_k, and $\alpha_k, \forall k = 1, \ldots, K$. The dependency between parameters and assortment prevents us from optimizing the subsequent assortment problem because we cannot estimate parameters for each possible assortment, and some possible assortments are not even observed in the historical dataset. This limitation is understandable since the MDCEV model was not specifically designed for assortment problems.

Nevertheless, we are making efforts to compare estimation performance between our multiple-purchase choice model and the MDCEV model. The comparison is conducted under three assortments with the highest number of transactions in the dataset. Focusing on transactions associated with these specific assortments, we utilize 60% of the transactions for each assortment to estimate the parameters for both the MDCEV model and the multiple-purchase choice model. Subsequently, we employ the remaining 40% of transactions for each assortment to compare the performance of the two models. In our numerical experiments, we utilize the total number of items purchased by each customer as a proxy for the budget and assume that $e_k = 1, \forall k = 1, \ldots, K$. Since the total number of products purchased by customers varies, we normalize the total number to 1, and the quantity of each individual item is also scaled accordingly to fractions. The main estimation process for the MDCEV model is implemented in R language using the "Apollo" package, which provides the necessary functionalities to estimate the MDCEV model. More details about "Apollo" package can be found at http://www.apollochoicemodelling.com/index.html.

For the MDCEV model, we estimate three sets of parameters, each corresponding to one of the three assortments. for each assortment, We then calculate the purchase probabilities for products and compare the observed purchase probabilities using the transactions in the test sample corresponding to that assortment. In contrast, the multiple-purchase choice model employs a single set of parameters estimated for all three assortments, and the purchase probabilities for products within each assortment are calculated accordingly. Using the same four metrics employed in Section 6.2, we present the comparison results between the MDCEV model and the multiple-purchase choice model in Table 10. This allows us to evaluate their respective estimation performance across the different assortments.

Table 10
Comparison results of MDCEV model and multiple-purchase model.

Assortment	WR^2		$WRMSE$		$HRMSE$		WR	
	MP	MDCEV	MP	MDCEV	MP	MDCEV	MP	MDCEV
Assort1	0.2523	0.9397	0.0394	0.0112	0.1096	0.1071	1012.55	414.30
Assort2	0.9332	0.9508	0.0145	0.0124	0.0770	0.0968	1334.28	2891.95
Assort3	0.8876	0.9601	0.0184	0.0109	0.0699	0.0902	1424.89	810.28
Average value	0.7109	0.9493	0.0231	0.0117	0.0818	0.0968	1253.39	1670.50
Gap(%)	−25.12%		96.94%		−15.49%		−24.97%	

As shown in Table 10, the negative value of gaps for WR^2 and $WRMSE$ indicate that the MDCEV model outperforms the multiple-purchase choice model in terms of these two metrics. While, for $HRMSE$ and WR the multiple-purchase choice model performs better than the MDCEV model.

The relatively large gaps observed in WR^2 and $WRMSE$ can be attributed to the differences in the estimation processes between the MDCEV model and the multiple-purchase choice model. Specifically, the disparity arises from how the data samples are utilized for estimation and comparison purposes. For the MDCEV model, the estimation and comparison are performed using specific data samples corresponding to each assortment. In other words, for each assortment, we estimate a distinct set of parameters and subsequently conduct the comparisons using the test data specific to that particular assortment. In contrast, the multiple-purchase choice model adopts a pooling approach, wherein transactions from all three assortments are combined. Only one set of parameters is estimated based on this merged dataset, and the subsequent comparison is conducted using test samples specific to each assortment.

As a result, the parameters estimated using the combined dataset in the multiple-purchase choice model may exhibit poorer performance when applied to a specific assortment's test sample. Conversely, the MDCEV model achieves a better fit by estimating parameters tailored to each individual assortment. Consequently, it is expected that the MDCEV model demonstrates superior performance due to its reliance on assortment-specific estimation.

References

Alibeiki, H., Li, S., Vaidyanathan, R., 2020. Market dominance or product cost advantage: Retail power impacts on assortment decisions. Int. J. Prod. Econ. 222, 107505.

Alptekinoğlu, A., Grasas, A., 2014. When to carry eccentric products? Optimal retail assortment under consumer returns. Prod. Oper. Manage. 23 (5), 877–892.

Anderson, S.P., De Palma, A., Thisse, J.-F., 1992. Discrete Choice Theory of Product Differentiation. MIT Press.

Bai, Y., Feldman, J., Segev, D., Topaloglu, H., Wagner, L., 2023. Assortment optimization under the multi-purchase multinomial logit choice model. Oper. Res..

Berbeglia, G., Garassino, A., Vulcano, G., 2021. A comparative empirical study of discrete choice models in retail operations. Manage. Sci..

Bhat, C.R., 2008. The multiple discrete-continuous extreme value (MDCEV) model: Role of utility function parameters, identification considerations, and model extensions. Transp. Res. B 42 (3), 274–303.

Bhat, C.R., 2022. A closed-form Multiple Discrete-Count Extreme Value (MDCNTEV) model. Transp. Res. B 164, 65–86.

Cachon, G.P., Gallino, S., Xu, J., 2018. Free shipping is not free: A data-driven model to design free-shipping threshold policies. Available at SSRN 3250971.

Cachon, G.P., Kök, A.G., 2007. Category management and coordination in retail assortment planning in the presence of basket shopping consumers. Manage. Sci. 53 (6), 934–951.

Chan, R., Li, Z., Matsypura, D., 2020. Assortment optimisation problem: A distribution-free approach. Omega 95, 102083.

Chen, J., Feng, X., Kou, G., Mu, M., 2023. Multiproduct newsvendor with cross-selling and narrow-bracketing behavior using data mining methods. Transportation Research Part E: Logistics and Transportation Review 169, 102985.

Chen, J., Liang, Y., Shen, H., Shen, Z.-J.M., Xue, M., 2021. Offline-channel planning in smart omnichannel retailing. Manuf. Serv. Oper. Manag.

Çömez-Dolgan, N., Moussawi-Haidar, L., Jaber, M.Y., Cephe, E., 2022. Capacitated assortment planning of a multi-location system under transshipments. Int. J. Prod. Econ. 251, 108550.

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

Désir, A., Goyal, V., Zhang, J., 2014. Near-optimal algorithms for capacity constrained assortment optimization. Available at SSRN 2543309.

Dubé, J.-P., 2004. Multiple discreteness and product differentiation: Demand for carbonated soft drinks. Mark. Sci. 23 (1), 66–81.

Feldman, J.B., Topaloglu, H., 2015. Capacity constraints across nests in assortment optimization under the nested logit model. Oper. Res. 63 (4), 812–822.

Feldman, J., Topaloglu, H., 2017a. Capacitated assortment optimization under the multinomial logit model with nested consideration sets. Oper. Res. 66 (2), 380–391.

Feldman, J.B., Topaloglu, H., 2017b. Revenue management under the Markov chain choice model. Oper. Res. 65 (5), 1322–1342.

Feldman, J., Zhang, D., Liu, X., Zhang, N., 2018. Taking assortment optimization from theory to practice: Evidence from large field experiments on alibaba.

Feng, X., Rong, Y., Tian, X., Wang, M., Yao, Y.O., 2022a. When persuasion is too persuasive: an empirical analysis of product returns in livestream e-commerce. Available at SSRN 4079214.

Feng, Q., Shanthikumar, J.G., Xue, M., 2022b. Consumer choice models and estimation: A review and extension. Prod. Oper. Manage. 31 (2), 847–867.

Feng, X., Wang, M., 2023. Strategic driver's acceptance-or-rejection behavior and cognitive hierarchy in on-demand platforms. Transportation Research Part E: Logistics and Transportation Review 176, 103175.

Flapper, S.D.P., González-Velarde, J.L., Smith, N.R., Escobar-Saldívar, L.J., 2010. On the optimal product assortment: Comparing product and customer based strategies. Int. J. Prod. Econ. 125 (1), 167–172.

Gallego, G., Iyengar, G., Phillips, R., Dubey, A., 2004. Managing Flexible Products on a Network. Citeseer.

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

Harlam, B.A., Lodish, L.M., 1995. Modeling consumers' choices of multiple items. J. Mar. Res. 32 (4), 404–418.

Huh, W.T., Li, H., 2022. Optimal pricing under multiple-discrete customer choices and diminishing return of consumption. Oper. Res. 70 (2), 905–917.

Jagabathula, S., Subramanian, L., Venkataraman, A., 2018. A conditional gradient approach for nonparametric estimation of mixing distributions.

Katsifou, A., Seifert, R.W., Tancrez, J.-S., 2014. Joint product assortment, inventory and price optimization to attract loyal and non-loyal customers. Omega 46, 36–50.

Kök, A.G., Fisher, M.L., Vaidyanathan, R., 2008. Assortment planning: Review of literature and industry practice. In: Retail Supply Chain Management. Springer, pp. 99–153.

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

Li, X., Rong, Y., Zhang, R.P., Zheng, H., 2022. Online advertisement allocation under customer choices and algorithmic fairness. Available at SSRN 3538755.

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

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

Luan, S., Wang, R., Xu, X., Xue, W., 2020. Operations management under consumer choice models with multiple purchases. Available at SSRN 3679699.

Ma, X., Deng, T., Lan, B., 2016. Demand estimation and assortment planning in wireless communications. J. Syst. Sci. Syst. Eng. 25 (4), 398–423.

McFadden, D., Train, K., 2000. Mixed MNL models for discrete response. J. Appl. Econometrics 15 (5), 447–470.

McFadden, D., et al., 1973. Conditional Logit Analysis of Qualitative Choice Behavior. Institute of Urban and Regional Development, University of California.

Mondal, A., Bhat, C.R., 2022. A spatial rank-ordered probit model with an application to travel mode choice. Transp. Res. B 155, 374–393.

Nair, G.S., Astroza, S., Bhat, C.R., Khoeini, S., Pendyala, R.M., 2018. An application of a rank ordered probit modeling approach to understanding level of interest in autonomous vehicles. Transportation 45, 1623–1637.

Nair, G.S., Bhat, C.R., Pendyala, R.M., Loo, B.P., Lam, W.H., 2019. On the use of probit-based models for ranking data analysis. Transp. Res. Rec. 2673 (4), 229–240.

Peng, Z., Rong, Y., Zhu, T., 2022. Transformer-based choice model: a tool for assortment optimization evaluation. Available at SSRN 4298996.

Ren, C.R., Hu, Y., Hu, Y., Hausman, J., 2011. Managing product variety and collocation in a competitive environment: An empirical investigation of consumer electronics retailing. Manage. Sci. 57 (6), 1009–1024.

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

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

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

Ryzin, G.v., Mahajan, S., 1999. On the relationship between inventory costs and variety benefits in retail assortments. Manage. Sci. 45 (11), 1496–1509.

Strauss, A.K., Klein, R., Steinhardt, C., 2018. A review of choice-based revenue management: Theory and methods. European J. Oper. Res. 271 (2), 375–387.

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

Train, K.E., 2009. Discrete Choice Methods with Simulation. Cambridge University Press.

Tulabandhula, T., Sinha, D., Patidar, P., 2020. Multi-purchase behavior: Modeling and optimization. Available at SSRN 3626788.

Vaagen, H., Wallace, S.W., Kaut, M., 2011. Modelling consumer-directed substitution. Int. J. Prod. Econ. 134 (2), 388–397.

Vend, 2019. Vend's 2019 retail benchmarks report. https://www.vendhq.com/retail-benchmarks. (Last Accessed Date: 9 March 2022).

Wan, M., Huang, Y., Zhao, L., Deng, T., Fransoo, J.C., 2018. Demand estimation under multi-store multi-product substitution in high density traditional retail. European J. Oper. Res. 266 (1), 99–111.

Wang, R., 2018. When prospect theory meets consumer choice models: Assortment and pricing management with reference prices. Manuf. Serv. Oper. Manag. 20 (3), 583–600.

Wang, M., Shang, G., Rong, Y., Galbreth, M., 2022. Order basket contents and consumer returns. Available at SSRN 4283793.

Yue, D., Guillén-Gosálbez, G., You, F., 2013. Global optimization of large-scale mixed-integer linear fractional programming problems: A reformulation-linearization method and process scheduling applications. AIChE J. 59 (11), 4255–4272.

Zhang, H., Piri, H., Huh, W.T., Li, H., 2021. Assortment optimization under multiple-discrete customer choices. Available at SSRN 3988982.