# A Unified Analysis for Assortment Planning with Marginal Distributions

Zeyu Sun
Institute of Operations Research and Analytics, National University of Singapore, sunzeyu@nus.edu.sg, orcid: 0000-0002-3016-6621

Selin Damla Ahipasaoglu
Mathematical Sciences and CORMSIS, University of Southampton, sda1u20@soton.ac.uk, orcid: 0000-0003-1371-315X

Xiaobo Li
Industrial Systems Engineering and Management, National University of Singapore, iselix@nus.edu.sg, orcid: 0000-0002-1909-628X

In this paper, we study assortment planning under the marginal distribution model (MDM), a semiparametric choice model that only requires information about the marginal noise in the utilities of alternatives and does not assume independence of the noise terms. It is already known in the literature that the multinomial logit (MNL) model belongs to the MDM framework. In this work, we demonstrate that some multi-purchase choice models such as the multiple-discrete-choice (MDC) model and the threshold utility model (TUM) also fall into the framework of MDM, even though MDM does not explicitly model multi-purchase behavior. For assortment problems within the MDM framework, we identify a general condition under which a strictly profit-nested assortment is optimal. While the problem is NP-hard, we show that the best strictly profit-nested assortment is a 1/2-approximate solution for all MDMs. Additionally, we present a simple example of an MDM for which the 1/2-approximate bound is tight. These results either extend or improve upon previous findings on assortment optimization under MNL, MDC, and TUM. Additionally, we present an arbitrary-close approximation algorithm for MDM, and an improved version for a class of choice models that includes MDC as a special case. Finally, we conduct experiments on real-world data and compare the predictive power of several choice models in the presence of multi-purchase behavior.

*Key words*: discrete choice models, assortment optimization, marginal distribution model, multiple discrete-continuous extreme value model, approximation ratio, approximation algorithm

## 1. Introduction

Assortment optimization is a fundamental problem in revenue management. Given a set of potential products, the firm must infer the customers' preferences and decide on a subset of products to offer to maximize its profit. Customers are expected to buy products from this subset according to their preferences. If a company only offers high-profit products, it may lose customers who are not interested in those products. On the other hand, if a company offers too many choices, customers who may have bought higher-profit products may opt for lower-profit ones instead. Therefore, to achieve high profit, the firm must carefully trade off the market share and the average profit of the products sold. Assortment optimization usually emerges as shelf-space optimization for brick-and-mortar stores, or as the search result display or product recommendation for e-commerce. With the recent growth of online retail, being able to offer customized and personalized assortments that take into account past purchase history, brand loyalty, product choice propensities (organic, premium, etc.) together with demographics is highly desired and expected to be the common industry practice in the next decade. (See Insider 2019, Bernstein et al. 2019, Sauré and Zeevi 2013, El Housni and Topaloglu 2021).

The expected profit of a subset, which is a critical step in assortment optimization, is typically calculated as the sum of the product of unit profits of each item and its choice probabilities given the subset. The choice probabilities are calculated via a discrete choice model built using the attributes mentioned above. Traditional choice models belong to the random utility model (RUM) framework, where product utilities are modeled as the addition of deterministic and stochastic components. The deterministic components represent the utility that can be explained by observable product and customer attributes, while the stochastic components capture the utility from unobserved or unaccounted attributes and perception errors. The most well-known RUM is the multinomial logit (MNL) model, where the stochastic utility components are assumed to follow i.i.d. Gumbel distributions. The MNL model is attractive because it allows for the calculation of choice probabilities in a simple, closed form. On the other hand, the substitution patterns under MNL are restrictive, resulting in limitations in modeling real-world customer choices. For instance, one well-known restriction of MNL is the Independence of Irrelevant Alternatives (IIA) property, in which the ratio of the choice probabilities of two products remains the same as long as the attributes of these two products do not change.

In this paper, we study assortment optimization under the marginal distribution model (MDM). MDM is a flexible choice model that includes the well-known multinomial logit model (MNL) as a special case. Unlike the random utility model (RUM), which requires a complete understanding of the joint distribution of noise, MDM only requires knowledge of the marginal distributions of noise. This allows for the calculation of choice probabilities using a simple bisection method for any set of marginal distributions, even if they belong to different parametric families. The marginal distributions can be chosen based on the specific data set and application.

We show that several multi-purchase choice models can be represented as special cases of MDM. These multi-purchase models are more practical for real-world applications than the classical discrete choice approach, which only allows customers to purchase one product at a time. Incorporating multiple-purchase behavior in choice modelling is relatively new, hence the related estimation, pricing, and assortment problems have not been fully examined. Our results show that some of these open questions can be addressed with the MDM framework, providing additional insights and tools into these models.

To the best of our knowledge, this is the first paper to comprehensively investigate assortment optimization under MDM. By using MDM as the choice model, we thoroughly study a broad class of assortment optimization problems through the following lens.

1. **Generality.** It is known that MDM includes MNL as a special case (Mishra et al. 2014). Are there other important special cases of MDM?

   *We show in section 3 that several choice models that capture multi-purchase behavior are also special cases of MDM. In particular, the multiple-discrete-choice (MDC) model, the deterministic variant of the well-studied Multiple Discrete Continuous Extreme Value (MDCEV) model (Kim et al. 2002, Bhat 2005), is a special case of MDM with marginal distributions belonging to the Pareto distribution family. Besides, the threshold utility model (TUM) recently proposed by Gallego and Wang (2019) is equivalent to MDM if the outside option is treated as another alternative. These results provide new insights into these multi-purchase choice models. Moreover, it allows us to study the operational and estimation problems for those choice models under the umbrella of MDM. Also, given the relation between MDC, TUM and MDM, the NP-hardness results in Zhang et al. (2021) (for MDC) and Gallego and Wang (2019) (for TUM) imply that the assortment optimization problem under MDM is NP-Hard.*

2. **Optimality.** It is well known that the product assortment optimization problem under the MNL choice model has a profit-nested structure (van Ryzin and Mahajan 1999). Is it possible to identify conditions under which the same result holds for MDM?

   *In Theorem 4, we identify a sufficient condition under which a (strictly) profit-nested assortment is optimal. This result generalizes known results for the optimality of profit-nested assortments under MNL and MDC.* Building upon this condition, we derive simple conditions that can be verified easily for some marginal distribution families.

3. **Approximation Ratio.** If we use profit-nested assortments to approximate the optimal solution of an assortment optimization problem under MDM, can we quantify how far away is the best profit-nested assortment from the optimal assortment in the worst case?

   *In Theorem 5, we show that the best (strictly) profit-nested assortment provides a 1/2-approximation for all MDMs. We further show that this bound is tight for the case in which the marginals are exponential or Pareto. These results greatly broaden the understanding of the profit-nested structure in assortment planning considering the generality of the MDM.*

4. **Approximation Algorithm.** Is it possible to obtain a more precise solution when a more accurate approximation is required?

   *In Section 4.3, inspired by Zhang et al. (2021), we present an approximation algorithm that* *provides a $1-\eta$ approximation guarantee for all marginal distribution models (MDMs) with a run time that is polynomial in $\frac{1}{\eta}$. In addition, in Section 5, we propose an improved approximation scheme under some special MDMs, such as the multiple-discrete customers choice model (MDC) model. These algorithms complement the theoretical property of profit-nested assortments for real applications.*

The rest of the paper is structured as follows: in Section 2, we provide an extensive literature review on related choice models and product assortment problems under these models; in Section 3, we discuss choice models in which customers are allowed to buy multiple products and show that some models developed to capture this behavior belong to MDM; in Section 4, we formally formulate the product assortment problem under MDM, discuss the properties of profit-nested assortments, and provide an arbitrary-close approximation algorithm for all MDMs; in Section 5, we present an improved approximation scheme under some conditions; in Section 6, we conduct numerical experiments using a real-world data set to demonstrate the predictive performance of different choice models in capturing multi-purchase choice behavior; finally conclusion is drawn in Section 7. Most proofs are presented in the appendix.

## 2. Literature Review

In this section, we provide an extensive review of relevant discrete choice models with some technical details. We then review the literature on assortment optimization under these models.

### 2.1. Random Utility Models

A discrete choice model describes how decision-makers make their choices among a set of products. Denote this set by $\mathcal{Q}$, where $\mathcal{Q}=\{1,2,\ldots,Q\}$ and the probability of choosing product $j$ by $x_{j,\mathcal{Q}}$. Customers may decide not to buy any of the products. This is modeled with the inclusion of a dummy product, called the outside market option and indexed by 0.

One common approach to modeling the utilities of different products is the random utility model (RUM), which assumes an additive noise structure. According to the RUM, the random utility for product $j$ is given by:

$$\tilde{U}_j = v_j + \tilde{\epsilon}_j, \quad \forall j \in \mathcal{Q} \cup \{0\}, \tag{1}$$

where $v_j$ represents the deterministic utility component and $\tilde{\epsilon}_j$ represents the random utility component. The multivariate random vector $\tilde{\boldsymbol{\epsilon}}$ follows a known probability distribution $\theta$. Without loss of generality, $v_0$ is usually assumed to be zero since only the difference between utilities matters. Given an offered set $\mathcal{S} \subseteq \mathcal{Q}$, the choice probability can be computed as:

$$x_{j,\mathcal{S}} = P_{\tilde{\boldsymbol{\epsilon}} \sim \theta}\left(j = \text{argmax}_{i \in \mathcal{S} \cup \{0\}} \; \tilde{U}_i\right), \quad \forall j \in \mathcal{S} \cup \{0\}.$$

It is important to note that the choice probability of a product depends on the set of products being offered for sale[1]. However, to simplify the notation, in the remainder of the paper, we will use $x_j$ directly without specifying $\mathcal{S}$ when the context makes it clear which set of products is being considered.

Different assumptions on the distribution of $\tilde{\boldsymbol{\epsilon}}$ lead to different choice models. Among them, the most popular choice model is the MNL model introduced in Luce (1959). In the MNL model, the stochastic terms, $\tilde{\epsilon}_j$'s, are assumed to follow i.i.d. Gumbel distributions with parameter $\alpha$ and the choice probability of each product can be expressed in closed form:

$$x_j = \frac{e^{\alpha v_j}}{1+\sum_{i\in\mathcal{S}} e^{\alpha v_i}}, \quad \forall j \in \mathcal{S}\cup\{0\}.$$

Thanks to the convenience of calculating the choice probabilities, the MNL model has been popular. However, the noise term distribution assumptions lead to undesirable restrictions on choice patterns, such as the well-known independence of irrelevant alternatives property. To capture more flexible choice patterns, various other choice models have been proposed. For more information on these models, we refer the reader to the textbook Train (2009).

### 2.2. Representative Agent Model and Additive Perturbed Utility Model

Another popular choice framework is the representative agent model (RAM), in which a representative agent chooses a choice probability vector on behalf of the entire population. To make her choice, the agent aims to maximize the expected utility while also having some degree of diversification in the choice vector. More precisely, given an offered set $\mathcal{S}\subseteq\mathcal{Q}$, the representative agent solves an optimization problem as follows:

$$\max\left\{\sum_{j\in\mathcal{Q}\cup\{0\}} v_j x_j - V(\boldsymbol{x}) : \sum_{j\in\mathcal{Q}\cup\{0\}} x_j = 1, x_j\geq 0, \forall j\in\mathcal{S}\cup\{0\}, x_j = 0, \forall j\notin\mathcal{S}\cup\{0\}\right\}, \tag{P1}$$

where $v_j$ and $x_j$ are the deterministic utility and choice probability corresponding to product $j$, and $V(\boldsymbol{x})$ is a convex perturbation function that punishes diversification.

It is shown in Anderson et al. (1988) that when $V(\boldsymbol{x})$ is the negative of the entropy function, i.e. $V(\boldsymbol{x}) = \eta\sum_{i=0}^{n} x_i \log x_i$, the RAM recovers the MNL choice model. Hofbauer and Sandholm (2002) further shows that for any given random utility model, there exists a $V(\boldsymbol{x})$ such that the corresponding RAM gives the same choice probability for any subset $\mathcal{S}$.

[1] Note that in RUM and most other parametric models, another way to express the absence of a product $i$ in an offered set is to let $v_i$ be negative infinity (see, e.g., Feng et al. 2018). In this paper, to make the relationship between the choice probability and the offered set clear, we define choice probability as a function of the offered set and avoid the discussion of negative infinite utility.

Fudenberg et al. (2015) introduces the additive perturbed utility (APU) model, which is a special case of RAM. Given an offered set $\mathcal{S} \subseteq \mathcal{Q}$, the APU choice probabilities correspond to the optimal solution of the following maximization problem:

$$\max\left\{\sum_{j\in\mathcal{S}\cup\{0\}}\Big(v_j x_j - c(x_j)\Big) : \sum_{j\in\mathcal{S}\cup\{0\}} x_j = 1, x_j \ge 0, \forall j \in \mathcal{S}\cup\{0\}\right\}, \tag{P2}$$

where $c(\cdot)$ is a uni-variate convex perturbation cost function. More specifically, $c(\cdot)$ is strictly convex, first-order differentiable on its domain $[0,1]$ and $\lim_{q\to 0} c'(q) = -\infty$. Fudenberg et al. (2015) shows that APU has close relations with several choice axioms, such as ordinal IIA and acyclicity. Harsanyi (1973), Machina (1985), Rosenthal (1989), Clark (1990), Mattsson and Weibull (2002) and Swait and Marley (2013) also study such perturbed functions or their variants.

Essentially, APU is a special case of RAM by letting $V(\boldsymbol{x}) = \sum_{j\in\mathcal{Q}\cup\{0\}} c(x_j)$. Note that since $V(\boldsymbol{x})$ is separable, the constraints $x_j = 0, \forall j \notin \mathcal{S}\cup\{0\}$ is irrelevant. Clearly, MNL is a special case of APU by letting $c(x) = x\log(x)$. Since $V(\boldsymbol{x})$ is separable, a natural extension of APU is to allow the perturbation function $c(x)$ to be different for different products, i.e., $V(\boldsymbol{x}) = \sum_{j\in\mathcal{Q}\cup\{0\}} c_j(x_j)$. It turns out that this extension is equivalent to the marginal distribution models proposed by Natarajan et al. (2009) from a different angle, which will be illustrated in the next subsection.

### 2.3. Marginal Distribution Models

The marginal distribution model (MDM), introduced by Natarajan et al. (2009), has gained recent attention in the literature. This model takes the marginal distributions of the noise terms as input and assumes that the distribution of the noise terms belongs to a family of distributions $\Theta$, which includes all distributions with the given marginals. Rather than assuming a particular dependency structure, such as independence, customers make choices under the extremal distribution $\theta^*$, which maximizes expected welfare among the distributions in $\Theta$. Mathematically, given an offered set $\mathcal{S} \subseteq \mathcal{Q}$, the extremal distribution $\theta^*$ solves the following optimization,

$$\theta^* = \arg\max_{\theta\in\Theta}\ E_\theta(Z(\tilde{\boldsymbol{\epsilon}})),$$

where

$$Z(\tilde{\boldsymbol{\epsilon}}) = \max\left\{\sum_{j\in\mathcal{S}\cup\{0\}}(v_j + \tilde{\epsilon}_j)x_j : \sum_{j\in\mathcal{S}\cup\{0\}} x_j = 1, x_j \in \{0,1\}, \forall j \in \mathcal{S}\cup\{0\}\right\},$$

and $\tilde{\boldsymbol{\epsilon}} = \{\tilde{\epsilon}_j, \forall j \in \mathcal{S}\cup\{0\}\}$. Let $F_j(\cdot)$ denote the $j$th marginal cumulative distribution function (c.d.f.) of the multivariate noise term. Natarajan et al. (2009) shows that the MDM choice probabilities correspond to the optimal solution to the concave maximization problem:

$$\max\left\{\sum_{j\in\mathcal{S}\cup\{0\}}\Big(v_j x_j + \int_{1-x_j}^{1} F_j^{-1}(t)dt\Big) : \sum_{j\in\mathcal{S}\cup\{0\}} x_j = 1, x_j \ge 0, \forall j \in \mathcal{S}\cup\{0\}\right\}, \tag{P3}$$

where $v_j$ is the deterministic utility corresponding to product $j$, $x_j$ is the choice probability of product $j$, and $F_j^{-1}(t)$ is the inverse of the c.d.f. of $\tilde{\epsilon}_j$.

As can be seen from (P3), MDM is a special case of RAM with a separable perturbation function. Here we show that the reverse is also true.

**Lemma 1** *For any RAM with a separable perturbation function, i.e., $V(\boldsymbol{x}) = \sum_{j\in\mathcal{Q}\cup\{0\}} c_j(x_j)$, where $c_j(\cdot)$ is strictly convex and first-order differentiable on its domain, there exist marginal distributions $F_j(\cdot), j \in \mathcal{Q}\cup\{0\}$ such that the corresponding MDM gives the same choice probabilities.*

The proof of Lemma 1 is provided in Appendix B.1. One immediate consequence of Lemma 1 is that MDM generalizes APU as a special case [2]. This generalization grants the MDM more flexibility than APU to model the stochastic choice, capture choice behavior of an agent who may be ambiguous about her true utilities to certain products, and handle incomplete choice data sets such as when data are only observed for a subset of possible menus (Fudenberg et al. 2015).

The theoretical and practical properties of MDM are studied in detail in Mishra et al. (2014). Here we note that the formulation given in Mishra et al. (2014) allows the marginal distribution $F_j(\cdot)$ be a function of $\boldsymbol{v} = \{v_0, v_1, ..., v_Q\}$ implicitly, which significantly expands the scope of the MDM. To clarify the distinction, in the following we call this model the generalized marginal distribution model (GMDM), and refer to the one without this flexibility as the marginal distribution model (MDM). To the best of our knowledge, Mishra et al. (2014) is the only paper that discusses GMDM, and it shows that GMDM includes GEV as a special case. They also show that if only the taste parameters are to be estimated, GMDM can be estimated efficiently, and the Fisher information matrix can be calculated, which helps determine the statistical significance of parameters.

In recent years, there has been further research on the properties and applications of MDM. Feng et al. (2018) shows that the MDM is always a regular choice model, which means that introducing a new product into the assortment would always decrease the choice probability of existing alternatives. Ahipaşaoğlu et al. (2019) applies the MDM to the traffic equilibrium problem. Yan et al. (2022) studies the data-driven estimation of the marginal distribution in MDM. Ruan et al. (2022) studies the problem of finding the best-fit MDM to the choice data without specifying the structure of marginal distributions. They show that checking whether a given choice dataset is consistent with MDM reduces to a linear program, while finding the best-fit MDM, in general, is difficult and can be formulated as a mixed-integer convex program.

[2] The relationship between MDM and APU has been mentioned in some papers (e.g. Ruan et al. 2022, Yan et al. 2022) without rigorous derivation. Lemma 1 provides a formal statement with proof to support this relationship.

### 2.4. Product Assortment Problems

There are many works related to assortment problems under the MNL model. The first polynomially solvable assortment problem was identified by van Ryzin and Mahajan (1999). They formulate an assortment problem by using the newsvendor model to determine the optimal order quantities and choose the MNL to capture the customers' choice behavior. They impose that the unit prices and costs are identical for all products; the demand for each product is normally distributed, and the standard deviation of the demand is a power function of the mean demand. They define assortments $\{1,2,...,m\}$ for some $m \in \mathcal{Q}$ to be *utility-nested sets*, where $v_1 \geq v_2 \geq ... \geq v_Q$, and find that there exists an optimal assortment which is a utility-nested set. Allowing products to be offered with different prices, denoted as $p_j$s below and ignoring the inventory constraints, both Gallego et al. (2004) and Liu and Van Ryzin (2008) show that there exists a *profit-nested* optimal solution, which is a subset $\{1,2,...,m\}$ for some $m \in \mathcal{Q}$, where $p_1 \geq p_2 \geq ... \geq p_Q$. Inspired by Liu and Van Ryzin (2008), Rusmevichientong et al. (2010) studies a similar problem with additional constraints. They show that Liu and Van Ryzin (2008)'s result does not hold for the constrained version of the assortment problem, nevertheless, they provide a polynomial-time algorithm to solve the constrained problem.

Beyond MNL, Davis et al. (2014) shows that the assortment problem under the nested logit (NL) choice model with dissimilarity parameter in $(0,1]$ and no no-buy options within nests is polynomially solvable and breaking any of these conditions makes the problem NP-hard. Gallego and Topaloglu (2014) studies the problem under the tractable NL model with a capacity constraint in each nest and shows that the problem is equivalent to a linear problem if the additional constraint is a cardinality constraint. Feldman and Topaloglu (2015) shows that the problem is also polynomially solvable if there is a capacity constraint on the entire assortment instead of within each nest. Zhang et al. (2020) shows that assortment problems under the paired combinatorial logit (PCL) model for both constrained and unconstrained versions are NP-hard. Cao et al. (2020) studies product assortment problems under a mixture of two choice models. They assume that a subset of customers make choices based on the MNL model and the rest of the customers follow the independent choice model, where each customer arrives with an ideal product in mind and does not purchase any product if her ideal product is unavailable. They show that the unconstrained problem is polynomially solvable and the constrained version is NP-hard.

Table 1 summarizes the results of the assortment optimization problems in the literature and in this paper. We can see that this paper fills most relevant gaps in the literature.

## 3. Relations to Multi-Purchase Choice Models

In this section, we examine the connection between the MDM and several multi-purchase choice models. Unlike discrete choice models that explicitly assume that each customer can buy at most

| source | choice model | complexity | profit-nested ratio | approx. alg. |
|---|---|---|---|---|
| Gallego et al. (2004)<br>Liu and Van Ryzin (2008) | **MNL model** | P | 1 | N.A. |
| Davis et al. (2014) | **NL model**<br>$(0,1]$ dissimilarity para.<br>no no-buy within nests | P | - | N.A. |
| | **NL model**<br>general cases | NP-hard | ? | not<br>$1-\eta$ approx. |
| Li et al. (2015) | **$d$-level NL model** | P | - | N.A. |
| Zhang et al. (2020) | **PCL model** | NP-hard | - | 0.6 approx. |
| Aouad et al. (2018) | **exponomial model** | ? | - | FPTAS |
| Feldman and Topaloglu (2017) | **markov chain model** | P | ? | N.A. |
| Berbeglia and Joret (2020) | **regular choice model** | - | 3 tight bounds,<br>$\leq 1/2$ | ? |
| Zhang et al. (2021) | **MDC model** | NP-hard | - | $1-\eta$ approx. |
| Gallego and Wang (2019)<br>Gallego and Topaloglu (2019) | **TUM-$\kappa$** | NP-hard | - | - |
| This paper | **MDM** | - | 1/2 | $1-\eta$ approx. |
| | **MEM** | - | 1/2 | $1-\eta$ approx.<br>(improved) |
| | **MDC model** | - | 1/2 | $1-\eta$ approx.<br>(improved) |
| | **TUM-$\kappa$** | - | 1/2 | $1-\eta$ approx. |

**Table 1 The computational complexity, the profit-nested heuristic approximation guarantee, and the existence of an approximation algorithm of unconstrained assortment problems under various popular choice models that are either analyzed in this paper or in previous literature. We use '-' and '?' to denote that the corresponding topic 'is not discussed in the paper but studied in others' and 'is not discussed in the paper and in the literature'. We use 'N.A.' to denote 'not applicable'. For example, if the problem is not NP-hard, then an approximation algorithm is usually unnecessary.**

one product, multi-purchase choice models allow customers to purchase multiple products or multiple units of the same products simultaneously.

There are many studies in the marketing literature that model multi-choice behavior. For example, McCardle et al. (2007) works on the bundled-products problems; Seetharaman et al. (2005) studies multi-choice behavior across various categories; Cox (1972) proposes the multivariate logit model, which has been extended by many subsequent papers (see, e.g., Russell and Petersen 2000).

There has been a growing interest in studying multi-purchase behavior within the operations management community. The bundle multivariate logit model, which can capture the complementary and substitution effects among different products, has been proposed and studied in Tulabandhula et al. (2020). Assortment problems under the multivariate MNL model has been considered in Lyu et al. (2021), in which the revenue-ordered assortment is shown not to be always optimal. Huh and Li (2022) and Zhang et al. (2021) study the pricing and assortment problem under the multiple-discrete choice model (MDC), which is a deterministic variant of the multiple discrete-continuous extreme value (MDCEV) model first introduced in Bhat (2005). Gallego and Wang (2019) recently proposes the threshold utility model (TUM), which captures the multi-purchase via the purchase quantity. Feldman et al. (2021) proposes the multi-purchase multinomial logit model, which extends the classic MNL model to a multi-purchase setting. Lin et al. (2022) studies the estimation and assortment personalization problem under the multi-choice rank list model, which is an extension of the classic rank list models (e.g., Block et al. 1959, Farias et al. 2013, van Ryzin and Vulcano 2015).

In the following discussion, we will focus on the MDC model and the TUM. Interestingly and surprisingly, these choice models are special cases of MDM, although MDM does not explicitly model multi-purchase behavior.

### 3.1. Multiple-Discrete Customer Choice Model

The MDC choice model studied in Huh and Li (2022) and Zhang et al. (2021) assumes that customers can choose any non-negative continuous quantity of products. For each product $i \in \mathcal{Q}$, the utility associated with product $i$ and the consumption quantity $y_i$ is given by

$$\nu_i(y_i) = \psi_i(y_i + \gamma_i)^{\sigma_i},$$

where $\psi_i > 0$ is the utility term which is dependent of product characteristics; $\gamma_i \geq 0$ which allows customers to have the option of not purchasing some of the products (namely, the consumption quantity of that product can be 0); and $\sigma_i \in (0,1)$ implies that all customers receive the decreasing marginal utility from the additional quantity purchased. They use option 0 to denote the outside market option and assume the purchase quantity of the outside market option is strictly positive and $\gamma_0 = 0$. Hence,

$$\nu_0(y_0) = \psi_0 y_0^{\sigma_0},$$

where $y_0 > 0$ denotes the purchase quantity associated to the outside market option; and $\sigma_0 \in (0,1)$. Given an offered set $\mathcal{S} \subseteq \mathcal{Q}$,the representative customer solves the following utility maximization problem:

$$\begin{aligned}
\max_{y_i \geq 0, \forall i \in \mathcal{S}, y_0 > 0} \quad & \psi_0 y_0^{\sigma_0} + \sum_{i \in \mathcal{S}} \psi_i (y_i + \gamma_i)^{\sigma_i} \\
\text{s.t.} \quad & y_0 + \sum_{i \in \mathcal{S}} y_i \leq \kappa,
\end{aligned}$$

where the total consumption quantity is upper-bounded by $\kappa > 0$. It turns out that this model is a special case of MDM.

**Theorem 1** *In the assortment context*[3]*, MDC belongs to MDM in which the deterministic utilities of all products are zeros and the marginals are truncated Pareto distributions. More specifically, these marginals are defined as:*

$$\begin{cases} F_i(t) = 1 + \frac{\gamma_i}{\kappa} - \frac{\left(\frac{\kappa\psi_i\sigma_i}{t}\right)^{1/(1-\sigma_i)}}{\kappa}, t \in \left[\kappa\psi_i\sigma_i(\frac{1}{\gamma_i+\kappa})^{1-\sigma_i}, \kappa\psi_i\sigma_i(\frac{1}{\gamma_i})^{1-\sigma_i}\right], \forall j \in \mathcal{Q}, \\ F_0(t) = 1 - \frac{\left(\frac{\kappa\psi_0\sigma_0}{t}\right)^{1/(1-\sigma_0)}}{\kappa}, t \in \left[\kappa\psi_0\sigma_0(\frac{1}{\kappa})^{1-\sigma_0}, \infty\right). \end{cases}$$

The proof of Theorem 1 is provided in Appendix B.2. The main idea of the proof is to treat $y_i/\kappa$ as the choice probability of product $i$ and construct the corresponding marginal c.d.f. $F_i(\cdot)$ such that $\psi_i(y_i+\gamma_i)^{\sigma_i}$ is equal to $\int_{1-y_j/\kappa}^{1} F_j^{-1}(t)dt$. Therefore, the purchasing capacity $\kappa$ appears in the marginal distribution. Note that the marginal distribution corresponding to the outside market option is the standard Pareto distribution with shape parameter $\frac{1}{1-\sigma_0}$ and scale parameter $\kappa\psi_0\sigma_0(\frac{1}{\kappa})^{1-\sigma_0}$. The marginal distribution of any other product $i$ is an upper-truncated Pareto distribution with shape parameter $\frac{1}{1-\sigma_i}$, scale parameter $\kappa\psi_i\sigma_i(\frac{1}{\gamma_i+\kappa})^{1-\sigma_i}$ and truncation point $\kappa\psi_i\sigma_i(\frac{1}{\gamma_i})^{1-\sigma_i}$. It is clear that the purchase quantity $\kappa$ does not affect the shape of the distribution.

This relationship has two important implications. First, MDM provides a natural framework to reasonably perturb or extend the existing MDC models in modeling multiple-discrete choices and diminishing return of consumption. Some other MDMs may also be capable of modeling customers' multiple choices. Second, it opens the door for new estimation methods for multi-purchase choice models. The parameter estimation of the MDC model has been studied in Bhat (2005, 2008, 2018). However, these methods are usually complicated by nature and may suffer from identification problems (Bhat 2008). On the other hand, the estimation of general MDM has been extensively explored in Mishra et al. (2014), Yan et al. (2022), and Ruan et al. (2022) in the data-driven setting. It is possible to apply these methods to estimate MDC by incorporating the marginal distribution constraints. It is even possible to develop a data-driven MDC model by allowing slight deviation from the MDC assumptions. Given the growing interest in models that allow for multiple-purchases, we believe that these are promising areas for future research.

[3] By the assortment context, we mean that the choice probabilities are defined for all different assortments. Specifically, by Theorem 1, we mean that for any instance of MDC defined on the product set $\mathcal{Q}$, we can find an instance of MDM such that for all offered sets $\mathcal{S} \subseteq \mathcal{Q}$, the choice probabilities of these two models are the same. Such a relationship perfectly fits the assortment optimization context and some other settings where the product features are fixed, but the product availability varies. However, in the pricing context where the price parameterizes the mean utilities, the MDC is not necessarily a special case of MDM. For instance, Huh and Li (2022) studies the optimal pricing problem by parameterizing $\psi$ with the product price. In this setting, MDC is not a special case of MDM because the marginal distribution changes with the price. However, suppose the mean utility is a function of the price, e.g., $v_i = a_i - b_i p_i$, but other parameters of the MDC are constants, then this pricing problem falls under the umbrella of pricing under MDM, which is recently studied in Yan et al. (2022).

Since the MDC model can be derived from the MDM, it is natural to ask whether the MDC model can also be derived from the RUM. It turns out that the answer is no. To see this, we first introduce the following known conditions for a choice model that can be represented by RUM.

**Lemma 2 (Block et al. 1959, Fiorini 2004)** *Let $x_{j,\mathcal{T}}$ denote the choice probability of alternative $j \in \mathcal{T}$ given a set of alternatives $\mathcal{T} \neq \emptyset$. A complete system of choice probabilities arises from the RUM if and only if the following inequalities hold:*

$$\sum_{\mathcal{U}:\mathcal{T}\subseteq\mathcal{U}\subseteq\mathcal{Q}^+} (-1)^{|\mathcal{U}\setminus\mathcal{T}|} x_{j,\mathcal{U}} \geq 0, \quad \forall\, \mathcal{T} \subseteq \mathcal{Q}^+, \forall\, j \in \mathcal{T},$$
$$\sum_{j\in\mathcal{T}} x_{j,\mathcal{T}} = 1, \quad \forall\, \mathcal{T} \subseteq \mathcal{Q}^+, \mathcal{T} \neq \emptyset,$$

*where $|\mathcal{U}\setminus\mathcal{T}|$ denotes the cardinality of set $\mathcal{U}\setminus\mathcal{T}$ and $\mathcal{Q}^+ = \mathcal{Q}\cup\{0\}$ is a set of all alternatives.*

The first set of inequalities is known as the Block-Marschak inequalities. We now consider a special case of the MDC model, in which $\gamma_j = 0, \forall j \in \mathcal{Q}$. We refer to this model as the *positive-MDC model* (or p-MDC for short), since all products will have strictly positive choice probabilities[4] Next, we provide an example that contradicts the aforementioned condition.

**Example 1** *In this example, we consider the MDC model with 4 products (i.e., $\mathcal{Q}^+ = \{1,2,3,4,0\}$). The MDC parameters of alternatives are given below: $\boldsymbol{\gamma} = [0,0,0,0,0]$, $\boldsymbol{\sigma} = [0.3, 0.9, 0.3, 0.9, 0.6]$, $\boldsymbol{\psi} = [1.9, 2.3, 1.6, 0.9, 0.6]$ and $\kappa = 1$. Notice that all elements in $\boldsymbol{\gamma}$ are zero, thus the model is p-MDC. We consider the case where $j = 0$ and $\mathcal{T} = \{4,0\}$, then we have*

$$x_{0,\{4,0\}} = 0.1273,\ x_{0,\{1,4,0\}} = 0.1028,\ x_{0,\{2,4,0\}} = 0.0.0126,\ x_{0,\{3,4,0\}} = 0.1088,$$
$$x_{0,\{1,2,4,0\}} = 0.0121,\ x_{0,\{1,3,4,0\}} = 0.0779,\ x_{0,\{2,3,4,0\}} = 0.0122,\ x_{0,\{1,2,3,4,0\}} = 0.0116.$$

*Then* $\sum_{\mathcal{U}:\mathcal{T}\subseteq\mathcal{U}\subseteq\mathcal{Q}^+}(-1)^{|\mathcal{U}\setminus\mathcal{T}|}(x_{j,\mathcal{U}}) = (-1)^0 \times 0.1273 + (-1)^1 \times 0.1028 + (-1)^1 \times 0.0126 + (-1)^1 \times 0.1088 + (-1)^2 \times 0.0121 + (-1)^2 \times 0.0779 + (-1)^2 \times 0.0122 + (-1)^3 \times 0.0116 = -0.0063 < 0$, *which violates Block-Marschak inequalities in Lemma 2.*

This counterexample shows that although MDC falls under the umbrella of MDM, even p-MDC does not belong to RUM. This result demonstrates that compared to the well-studied RUM framework, MDM has its unique modeling advantage.

[4] The original definition of MDC (see e.g. Zhang et al. 2021) requires $\gamma_j > 0$. However, the case where $\gamma_j = 0$ for some $j \in \mathcal{Q}$ does not conflict with any of the multi-purchase properties of the MDC model, except for the assumption that all products can have zero choice probabilities. Here we slightly generalize MDC to include this case. If this generalization is not allowed, p-MDC can still be obtained from MDC with a limiting argument, which is technically cumbersome. Later in the numerical study, we can see that this case has some salient features in estimation compared to the general MDC.

Interestingly, we find that the p-MDC model itself is equivalent to a well-studied instance of the MDM. The marginal exponential model (MEM) is an important special case of MDM in which the marginal distributions of noise terms are exponential distributions, i.e., $F_j(\epsilon) = 1 - e^{-\alpha_j \epsilon}$, $\forall j \in \mathcal{Q} \cup \{0\}$ for some positive and not necessarily identical $\alpha_j$s. The MEM is introduced in Mishra et al. (2014), establishing an interesting fact that when the scale parameters are identical, the MEM reduces to the MNL model. To see this, given an offered set $\mathcal{S} \subseteq \mathcal{Q}$, we observe that under the MEM, (P3) becomes

$$\max \left\{ \sum_{j \in \mathcal{S} \cup \{0\}} \left( v_j x_j + \frac{-x_j \log(x_j) + x_j}{\alpha_j} \right) : \sum_{j \in \mathcal{S} \cup \{0\}} x_j = 1, x_j \geq 0, \forall j \in \mathcal{S} \cup \{0\} \right\}.$$

The optimality conditions yield the following MEM choice probabilities:

$$x_j = e^{-\alpha_j \lambda + \alpha_j v_j}, \tag{2}$$

where the Lagrange multiplier $\lambda$ is the solution of the following equation:

$$\sum_{j \in \mathcal{S} \cup \{0\}} e^{-\alpha_j \lambda + \alpha_j v_j} = 1. \tag{3}$$

Given the outside market choice probability $x_0$, we can rewrite relations (2) and (3), respectively as

$$x_j(x_0) = e^{\alpha_j v_j} (x_0)^{\frac{\alpha_j}{\alpha_0}} \quad \text{and} \quad \sum_{j \in \mathcal{S}} x_j(x_0) + x_0 = 1.$$

When $\alpha_j = \alpha$ for all $j$, it is clear that

$$x_0 = \frac{1}{1 + \sum_{j \in \mathcal{S}} e^{\alpha v_j}} \quad \text{and} \quad x_j = \frac{e^{\alpha v_j}}{1 + \sum_{i \in \mathcal{S}} e^{\alpha v_i}}, \quad \forall j \in \mathcal{S},$$

which is exactly the MNL model. Mishra et al. (2014) demonstrates the superior predictive power of the MEM versus the MNL model in a case study using data from General Motors. From the modeling perspective, allowing different alternatives to have different scale parameters essentially allows them to have different variances in the utility, since the difference in the average of different exponential noise terms can be offset by shifting the mean utilities.

It turns out that MEM is equivalent to p-MDC model.

**Theorem 2** *In the assortment context*[5]*, the p-MDC model is equivalent to the MEM.*

The proof of Theorem 2 can be found in Appendix B.3. Since the exponential distribution does not belong to the Pareto family, it is somewhat surprising to see that MEM is a special case of MDC in the assortment context. Theorem 2 also implies that the MDC model includes the MNL model as a special case. This relationship was previously mentioned in Huh and Li (2022) under the pricing setup with the condition $\gamma_j \to 0$, $\sigma_j \to 0$ for all $j \in \mathcal{Q} \cup \{0\}$. In the assortment context, this condition is less stringent and only requires $\gamma_j \to 0$, $\sigma_j = \sigma$ for all $j \in \mathcal{Q} \cup \{0\}$, with $\sigma > 0$.

[5] It is not hard to find that in the pricing context, these two models are not necessarily equivalent.

### 3.2. Threshold Utility Model

Gallego and Wang (2019)[6] (see also chapter 4.12 in Gallego and Topaloglu 2019) recently develops a multi-purchase choice model, named the threshold utility model (TUM), and study the corresponding estimation, pricing and assortment optimization problems. Given an offered set $\mathcal{S} \subseteq \mathcal{Q}$, the mathematical representation of the TUM is as follows:

$$\left.\begin{array}{lll} \max\limits_{z_i \geq 0, \forall i \in \mathcal{S}\cup\{0\}} & \sum\limits_{i\in\mathcal{S}\cup\{0\}} \mathrm{E}[U_i | U_i \geq z_i]\cdot \Pr(U_i \geq z_i) \\ \text{s.t.} & \sum\limits_{i\in\mathcal{S}\cup\{0\}} \Pr(U_i \geq z_i) \leq \kappa, \end{array}\right\} \tag{P4}$$

where $\kappa \geq 1$ is a parameter that controls customers' purchase capacity, $U_i$ is the utility that a customer receives from product $i$ and $z_i$ is the utility threshold for product $i$. One may also view $\Pr(U_i \geq z_i)$ as the expected number of purchases of product $i$ made by the customer. As a result, we not only have an aggregate constraint on the expected total number of purchases, $\kappa$, but also have a purchase limit equal to 1 for each product (since the $\Pr(U_i \geq z_i) \leq 1$).

Note that in the original formulation of TUM, the outside market option is omitted. They provide two ways to incorporate the outside market option. First, given the utility corresponding to the outside market option $U_0$, they make the following transformation $U_i \leftarrow U_i - U_0$. As a result, the outside market option is chosen if the new $U_i < z_i, \forall i \in \mathcal{S}$. Second, the outside market option can be treated as an alternative. It can be checked that the choice probabilities from the first method can be reproduced from the second. Hence, we treat the outside option as an alternative in TUM to allow for maximum flexibility.

Moreover, when $\sum_{i\in\mathcal{S}\cup\{0\}} \Pr(U_i \geq z_i) < \kappa$, the model reduces to independently choosing each $z_i$ and brings limited insights. Therefore, we only consider the case when the constraint is an equality constraint, which leads to the following model

$$\left.\begin{array}{lll} \max\limits_{z_i \geq 0, \forall i \in \mathcal{S}\cup\{0\}} & \sum\limits_{i\in\mathcal{S}\cup\{0\}} \mathrm{E}[U_i | U_i \geq z_i]\cdot \Pr(U_i \geq z_i) \\ \text{s.t.} & \sum\limits_{i\in\mathcal{S}\cup\{0\}} \Pr(U_i \geq z_i) = \kappa. \end{array}\right\} \tag{P5}$$

Although (P5) is TUM with an equality constraint, for simplicity, we will refer it as TUM-$\kappa$ in the rest of the paper.

Though the derivation of the TUM appears very different from the MDM, these two models are closely related to each other. One key observation is that we can do the variable transformation by letting $x_i = \Pr(U_i \geq z_i)$, which is the choice probability of product $i$. This observation leads to the following lemma.

[6] Since this paper is a working paper (last revised on 30 Mar 2020), the models and results might be subject to change.

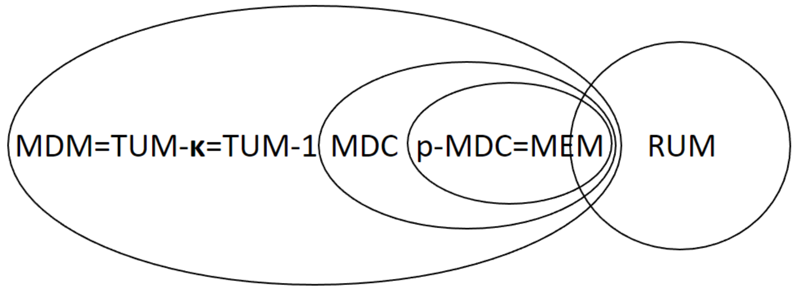


**Figure 1** **Relationship between RUM, MDM, TUM-$\kappa$ and MDC, and their special cases.**

**Lemma 3** *Suppose the c.d.f.'s of the utility functions, $U_i$s, are invertible. For all $\boldsymbol{v} \in \mathcal{R}^{Q+1}$, given assortment $S$, the optimal choice probability under the TUM-$\kappa$ is the optimal solution to the following convex optimization problem:*

$$\max\left\{\sum_{j\in\mathcal{S}\cup\{0\}}\left(v_j x_j + \int_{1-x_j}^{1} G_j^{-1}(t)dt\right) : \sum_{j\in\mathcal{S}\cup\{0\}} x_j = \kappa, 0 \le x_j \le 1, \forall j \in \mathcal{S}\cup\{0\}\right\}, \tag{P6}$$

*where $G_j^{-1}(\cdot)$ is the inverse of the c.d.f. of $U_j - v_j$*[7]*.*

The derivation of Lemma 3 is provided in Appendix B.4. It is clear that when $\kappa = 1$, formulation (P6) reduces to MDM. Therefore, MDM is equivalent to TUM-1, which is a special case of TUM-$\kappa$. It turns out the reverse is also true, meaning that TUM-$\kappa$ is also a special case of MDM.

**Theorem 3** *TUM-$\kappa$ can be derived from MDM by letting the c.d.f. of $\epsilon_j$ be*

$$F_j(x) = 1 - \frac{1}{\kappa}\left(1 - G_j\left(\frac{1}{\kappa}x\right)\right), \forall j \in \mathcal{Q}\cup\{0\}.$$

*Therefore, TUM-$\kappa$ and MDM are equivalent.*

The proof of Theorem 3 is provided in Appendix B.5. One immediate consequence of Theorem 1 and 3 is that MDC is a special case of TUM in the context of assortment optimization. This relation was not known in the literature.

The relationship between MDM, MEM, MDC and RUM, derived in Theorem 1, Theorem 2 and demonstrated in Example 1, is depicted in Figure 1. In addition, Theorem 3 shows that TUM-1 is equivalent to TUM-$\kappa$, and both are equivalent to MDM.

Above, we show that MDM can include both MDC and TUM as special cases. To the best of our knowledge, this is the first result that establishes the relations between discrete choice models

[7] Here $G_j$ appears to be dependant on $v_j$, which appears to violate the definition of MDM. However, it is not the case. In the assortment optimization context, because $G_j$ does not depend on the mean utility of other products, it is not affected by the availability of other products. Based on the discussion in Section 3.1, $\boldsymbol{v}$ is redundant and can be simply set to zero. In the context of pricing problem as is studied in Gallego and Wang (2019), one can parameterize the mean utility $v$ by the price, and let $G$ remain unchanged.

and these multi-purchase choice models. Based on this relation, the results regarding the MDM can be applied to some multi-purchase choice models as well. In the following section, we study the assortment optimization problem under MDM and obtain results that either generalize or improve the known results for these choice models.

## 4. Product Assortment under MDM

In this section, we formulate the assortment problem under the MDM. Recall that the product set is denoted by $\mathcal{Q}$, where $\mathcal{Q} = \{1, 2, \ldots, Q\}$. Recall also that we assume $v_0 = 0$ without loss of generality. To avoid potential subtle technical issues in MDM, we make the following assumption on the marginal distributions of noise terms in this section.

**Assumption 1** *For all $i \in \mathcal{Q} \cup \{0\}$, $F_i^{-1}(\cdot)$ is a continuous function. Moreover, $F_0(\cdot)$ is defined on an infinite support $(-\infty, +\infty)$ or a semi-infinite support $[\underline{\epsilon}_0, +\infty)$.*

Assumption 1 implies that for any finite $\boldsymbol{v}$, the possibility that the outside option has the largest utility is nonzero, i.e., the outside option always has a strictly positive market share[8].

In the following proposition, we show that under Assumption 1, the optimal choice probability can be computed in a semi-closed form.

**Proposition 1** *Under Assumption 1 in MDM, given any assortment $S$, the probability that consumers choose product $i \in S$ is $x_j(x_0) \triangleq 1 - F_j\left(F_0^{-1}(1 - x_0) - v_j\right)$, where $x_0 \in [0, 1]$ is the market share of outside option that satisfies:*

$$\sum_{j \in S} x_j(x_0) + x_0 = 1.$$

The proof of Proposition 1 is provided in Appendix B.6. Let $p_j > 0$ represent the unit profit of product $j$. We order the products such that $p_1 \geq p_2 \geq \cdots \geq p_Q$. Then based on Proposition 1, the product assortment problem under the MDM can be formulated as:

$$\left.\begin{array}{ll} \max\limits_{S \subseteq \mathcal{Q}, x_0 \in (0,1]} & \sum\limits_{j \in S} p_j x_j(x_0) \\ \text{s.t.} & \sum\limits_{j \in S} x_j(x_0) + x_0 = 1 \end{array}\right\}, \tag{P7}$$

[8] Note that most existing papers (e.g. Natarajan et al. 2009, Mishra et al. 2014, Yan et al. 2022, Ahipaşaoğlu et al. 2016, 2019) on MDM assume a stronger condition that $F_i(\cdot)$ is defined on an infinite support $(-\infty, +\infty)$ or a semi-finite support $[\underline{\epsilon}_0, +\infty)$ for all $i \in \mathcal{Q} \cup \{0\}$, which implies that all options have strictly positive market share. Such an assumption makes the optimization problem in (P3) slightly easier to solve since one can ignore the non-negativity constraints when deriving the Karush–Kuhn–Tucker (KKT) conditions. It is not hard to find that the c.d.f.s corresponding to the MDC model as in Theorem 1 does not satisfy that assumption, but satisfy Assumption 1.

where $x_j = 1 - F_j\big(F_0^{-1}(1-x_0) - v_j\big), \forall j \in \mathcal{Q}$. It is easy to see that $x_j(x_0)$ is a non-decreasing function.

Since Zhang et al. (2021) shows that the assortment optimization under the MDC model is NP-hard, it is clear from Theorem 1 that the problem (P7) is also NP-hard. It is worth noting that the proof of the NP-Hardness in Zhang et al. (2021) assumes that $\gamma_i = 0$ for all products, so the hardness result is essentially for p-MDC and MEM [9].

### 4.1. Conditions for Optimality of the Profit-Nested Heuristic

An assortment $\mathcal{S}$ is a (strictly) *profit-nested assortment* if there exists a price threshold $\hat{p} > 0$ such that $i \in \mathcal{S}$ if and only if $p_i \geq \hat{p}$. In other words, if product $i$ is included in the profit-nested assortment $S$, then all products $j$ with $p_j \geq p_i$ are also included. It is easy to see that the number of strictly profit-nested assortments is at most $Q$. A simple heuristic to solve the assortment optimization problem is to try all the profit-nested assortments and pick the best. This heuristic, which needs to check at most $Q$ assortments, is referred to as the *profit-nested heuristic* below. Next, we first show that the profit-nested heuristic is optimal under some conditions. Then we show in Section 4.2 that the profit-nested heuristic is guaranteed to achieve at least half of the optimal profit for all MDM.

Now we provide a sufficient condition under which the profit-nested heuristic is optimal. Recall that in the MDM, $x_j = 1 - F_j(F_0^{-1}(1-x_0) - v_j), \forall j \in \mathcal{Q}$, which is a non-decreasing function of $x_0$.

**Assumption 2** *For any $i \in \{2, ..., Q\}$, $\sum_{j\in\mathcal{Q}} x_j(x_0)\cdot(p_j - p_i)^+$ is convex in $x_0$ in the region $\{x_0 | \sum_j x_j(x_0) + x_0 \leq 1\}$, where $z^+ = \max\{z, 0\}$.*

Note that the region $\{x_0 | \sum_j x_j(x_0) + x_0 \leq 1\}$ is to make sure that we only consider relevant $x_0$. Other values of $x_0$ are never optimal and thus are irrelevant. One simple sufficient condition that implies Assumption 2 is that $x_j(x_0)$ is a convex function for all $j \in \mathcal{Q}$.

**Theorem 4** *For assortment problems under MDM, if Assumption 2 holds, then the profit-nested heuristic is optimal.*

The proof of Theorem 4 is provided in Appendix B.7, which mainly relies on convexity and duality. We essentially prove that strong duality holds for a special-structured nonlinear mixed-integer program under some conditions. We can see that the profit-nested structure comes from the dual linear integer program and once strong duality holds, the profit-nested structure naturally holds. This proof technique has the potential to apply to other assortment optimization problems and could be of independent interest on its own.

[9] This result under MEM was independently shown in an earlier version of our work with a different title, which first appeared online in Aug 2021. In that version, the relation between MEM, MDM, and MDC is not investigated.

Indeed, with this proof technique, Theorem 4 generalizes the result in Zhang et al. (2021), which shows that the profit-nested heuristic is optimal for the assortment optimization problem under the assumption that for $i = 1, 2, ..., Q, \sum_{j=1}^{i} y_j$ is increasing and convex in $y_0$. It is not hard to see that our condition is weaker than this assumption. Hence, Theorem 4 generalizes the result in Zhang et al. (2021). The main reason for this improvement is attributed to the exploitation of the Lagrangian duality and convexity, which naturally leads to Assumption 2[10].

If the marginal distributions belong to the Pareto family, we have the following simpler conditions.

**Proposition 2** *In the MDM, let all marginals of all products follow truncated Pareto distribution with scale parameter $m_j > 0$, shape parameter $\beta_j > 0$, truncation point $T_j > m_j$ and deterministic utility corresponding to the product $j$ is denoted as $v_j$, for all $j \in \mathcal{Q}$. In addition, let the marginal of the outside market option follow a standard Pareto distribution with scale parameter $m_0 > 0$, shape parameter $\beta_0 > 0$, and deterministic utility is $v_0$. Then, Assumption 2 holds if the following condition holds:*

$$\beta_j \geq \beta_0 - \frac{\min(v_j, 0)}{m_0}(\beta_0 + 1), \forall j \in \mathcal{Q}.$$

The proof of Proposition 2 is provided in Appendix B.8. This result has an immediate consequence for the MDC model.

**Corollary 1** *In the MDC model, Assumption 2 holds if $\sigma_0 \leq \sigma_j, \forall j \in \mathcal{Q}$. Hence, a strictly profit-nested assortment is optimal for the corresponding assortment problem.*

The proof of corollary 1 is provided in Appendix B.9.

Recall that the MNL model is a special case of the p-MDC model when all $\sigma$'s are the same. Thus, Corollary 1 generalizes the well-known result that the profit-nested heuristic is optimal for the MNL model. Recall that $\sigma_j, \forall j \in \mathcal{Q} \cup \{0\}$ plays an important role in reflecting the diminishing marginal return of the additional quantity of product $j$ purchased. The condition essentially requires that the utility received by the customer from choosing the outside market option increases at the lowest rate.

### 4.2. Half Approximation Guarantee of the Profit-Nested Heuristic

We have shown that the profit-nested heuristic is optimal for (P7) under some conditions. Next, we study the approximation guarantee of the profit-nested heuristic when these conditions are not satisfied.

[10] However, it is still possible to prove the same result using the proof technique in Zhang et al. (2021) with some modifications.

**Theorem 5** *For assortment problems under the MDM, the (strictly) profit-nested heuristic finds a solution which is no worse than* $1/2$ *of the optimal solution.*

The proof of Theorem 5 is provided in Appendix B.10. The idea of the proof is as follows: we first solve the linear relaxation of the original problem and observe that there is an optimal solution with a special structure (at most one set of variables that are associated with all products belonging to the same profit level is fractional, and the nonzero variables correspond to a strictly profit-nested assortment). Based on this, we construct two strictly profit-nested feasible solutions to the original problem and show that the sum of the objective functions of these solutions is greater than or equal to the optimal objective function value. Therefore, at least one of them must be greater than or equal to the half of the optimal value.

Here, we discuss the connection between our results and existing bounds in the literature. Berbeglia and Joret (2020) studies approximation guarantees provided by profit-nested assortments under regular choice models. They provide three different types of bounds, which are all tight, depending on different model parameters. These bounds in general decay fast when the number of products increases.

In a more recent paper, building upon the prophet inequalities, Gallego and Berbeglia (2021) shows that for the random utility models in which the differences of utilities of products and the outside option are independent, profit-nested assortments provide a half-approximation bound. They also show that if the condition is not satisfied, the bound can be as poor as $1/Q$ even among random utility families. Our result in Theorem 5 complements these results and provides a half approximation guarantee for a much broader class of choice models. It is worth noting that Gallego and Berbeglia (2021) studies a more general assortment personalization problem, and their bounds hold even against the clairvoyant. It is an important future research question to determine whether the half approximation holds in the personalization setting under the MDM framework as well.

Under the MDC, Zhang et al. (2021) also provides a 1/2-approximation algorithm for assortment problems. This algorithm examines all *pseudo-profit-ordered sets*, which requires $\mathcal{O}(Q^3)$ profit evaluations. Based on Theorem 1 and Theorem 5, 1/2-approximation ratio can be achieved by simply applying the profit-nested heuristic. This only requires $\mathcal{O}(Q)$ profit evaluations, which is significantly faster than the pseudo-profit-ordered heuristic.

We show via the following proposition that even for an extremely simple case of problem (P7), the approximation ratio is tight. Based on (P7), the assortment problem under the MEM can be written as:

$$\left.\begin{aligned} \max_{S\subseteq\mathcal{Q},x_0} \quad & \sum_{j\in S} p_j e^{\alpha_j v_j} (x_0)^{\frac{\alpha_j}{\alpha_0}} \\ \text{s.t.} \quad & \sum_{j\in S} e^{\alpha_j v_j} (x_0)^{\frac{\alpha_j}{\alpha_0}} + x_0 = 1 \end{aligned}\right\}. \tag{P8}$$

To show the tightness, we consider a special case of (P8) in which $\alpha_j = \alpha$ for all $j \in \mathcal{Q}$. This means that the variability of the utility for available products are identical to each other, but is different from the outside option. Let $w_j = e^{\alpha v_j}$, $c_j = w_j p_j$ for all $j \in \mathcal{Q}$, $\tau = \frac{\alpha_0}{\alpha}$ and $z = x_0^{1/\tau}$, then (P8) reduces to:

$$\begin{aligned} \max_{z\in(0,1),\boldsymbol{y}} \quad & \boldsymbol{c}^T \boldsymbol{y} z \\ \text{s.t.} \quad & \boldsymbol{w}^T \boldsymbol{y} z + z^\tau = 1, \\ & \boldsymbol{y} \in \{0,1\}^Q. \end{aligned} \tag{P9}$$

Problem (P9) has only one additional parameter than the MNL model and perhaps is the simplest case of (P8).

**Proposition 3** *Consider an instance of Problem* (P9) *with three products. Let* $0 < \epsilon < 1/2$ *and set* $[p_1, p_2, p_3] = \left[1, \epsilon(1-2\epsilon), \epsilon(1-2\epsilon)\right]$, $[w_1, w_2, w_3] = [\epsilon, 10 + \frac{2}{\epsilon}, 1-\epsilon]$, $\tau = \frac{\log(\epsilon)}{\log(1-\epsilon)}$, *the profit corresponding to the best profit-nested assortment is half of the optimal profit when* $\epsilon$ *approaches* 0.

The proof of Proposition 3 is provided in Appendix B.11. Based on Theorem 2, this 1/2-approximation ratio is also tight for the profit-nested heuristic under the MDC model.

We have identified the conditions under which the profit-nested heuristic is optimal, demonstrated that the heuristic has a $\frac{1}{2}$-approximation guarantee in general cases, and shown that this bound is tight. In the following subsection, we present an algorithm with a $1-\eta$ approximation guarantee for situations where a more accurate solution (or a solution with arbitrary accuracy) is demanded.

### 4.3. Relative Approximation Algorithm

In Zhang et al. (2021), the Lipschitz continuity property is used to devise an absolute approximation algorithm that produces an $\varepsilon$-optimal solution with a run time that is polynomial in $\frac{1}{\varepsilon}$ for the assortment problem under the MDC model, given certain technical conditions. In this subsection, we modify and extend this algorithm to be applicable for all MDMs, convert it into a relative approximation algorithm, and provide a more rigorous complexity analysis.

Recall that $x_j(x_0)$ is non-decreasing for all $j \in \mathcal{Q}$ in problem (P7). This means that problem (P7) is equivalent to the following:

$$\left.\begin{aligned} \max_{S\subseteq\mathcal{Q}, x_0\in(0,1]} \quad & \sum_{j\in S} p_j x_j(x_0) \\ \text{s.t.} \quad & \sum_{j\in S} x_j(x_0) + x_0 \le 1 \end{aligned}\right\}. \tag{P10}$$

Let $f_{kp}(\mathcal{S}, x_0)$ denote the objective function value of problem (P10), i.e., $f_{kp}(S, x_0) = \sum_{j\in\mathcal{S}} p_j x_j(x_0)$, where $x_j = 1 - F_j\left(F_0^{-1}(1-x_0) - v_j\right), \forall j \in \mathcal{S} \subseteq \mathcal{Q}$. As a choice probability, it is clear that $x_0 \in [0,1]$. Moreover, the following lemma allows us to further narrow down the search space for the optimal outside market share:

**Lemma 4** *For any assortment, the optimal outside market share lies in* $[\underline{x_0}, \overline{x_0}]$, *where*

$$\begin{cases} \underline{x_0} = 1 - F_0(\lambda_l - v_0), \ \textit{where} \displaystyle\sum_{j \in \mathcal{Q} \cup \{0\}} \big(1 - F_j(\lambda_l - v_j)\big) = 1, \text{and} \\ \overline{x_0} = \displaystyle\max_{k \in \mathcal{Q}} \{1 - F_0(\lambda_k - v_0)\}, \ \textit{where} \ \big(1 - F_k(\lambda_k - v_k)\big) + \big(1 - F_0(\lambda_k - v_0)\big) = 1, \forall k \in \mathcal{Q}. \end{cases}$$

This result follows as a corollary of Proposition 5 that is presented in Appendix B.12.

To calculate $f_{kp}(\mathcal{S}, x_0)$ for a given $\mathcal{S}$ and $x_0$, we need to evaluate the function $x_j = 1 - F_j\big(F_0^{-1}(1 - x_0) - v_j\big)$ for all $j \in \mathcal{Q}$. For the purposes of this analysis, we assume that the total computational effort to calculate $x_j(x_0)$ to the desired accuracy is upper bounded by $C$, and we express the time complexity in terms of $C$.

To develop a relative approximation algorithm, we first define the following quantity. Let

$$M = \max_{\underline{x_0} \le y_0 < x_0 \le \overline{x_0}, j \in \mathcal{Q}} \left( \frac{x_j(x_0) - x_j(y_0)}{x_j(x_0)(x_0 - y_0)} \right).$$

Inspired by Zhang et al. (2021), we present the following algorithm which uses the FPTAS for the knapsack problem (see Algorithm 3 in Appendix C.2) as a subroutine. Note that the FPTAS requires $O\Big(Q^2 \lfloor \frac{Q}{\epsilon} \rfloor + QC\Big)$ time, where $QC$ accounts for the calculation of the 'weights' and 'profits' in the corresponding knapsack problem.

**Algorithm 1** MDM relative approximation algorithm

**Input** $\eta$, $\underline{x_0}$, $\overline{x_0}$, $M$

1: $\epsilon \leftarrow \frac{\eta}{2}$
2: $\Delta \leftarrow \frac{\eta}{2M}$
3: **for** $L \leftarrow \underline{x_0}, \underline{x_0} + \Delta, \underline{x_0} + 2\Delta, ..., \underline{x_0} + \Big\lfloor \frac{\overline{x_0} - \underline{x_0}}{\Delta} \Big\rfloor \Delta$ **do**
4: Use knapsack FPTAS Algorithm 3 with parameter $\epsilon$ to approximate the inner problem of (P10) with $x_0 = L$.
5: Store the assortment and the corresponding objective function value to problem (P10).
6: **return** the assortment with the greatest objective function value

We now present our main theorem for this subsection.

**Theorem 6** *Algorithm 1 is a* $(1 - \eta)$*-approximation algorithm with a run time* $O\bigg(\Big(\Big\lfloor \frac{2M}{\eta} \Big\rfloor + 1\Big) \cdot \Big(Q^2 \Big\lfloor \frac{Q}{\eta} \Big\rfloor + QC\Big)\bigg)$ *that is polynomial in* $\frac{1}{\eta}$ *and* $M$.

The proof of Theorem 6 is provided in Appendix C.3. This theorem extends the results in Zhang et al. (2021) to all MDM cases. Although the proof appears different from Zhang et al. (2021), the key idea is the same. The difference is mainly in the definition of $M$, because we aim to provide a relative performance guarantee, while the analysis in Zhang et al. (2021) is for absolute performance guarantee.

It is worth noting that the runtime of Algorithm 1 is not necessarily polynomial in the problem parameters. In particular, the constant $M$ may depend on the problem parameters. We illustrate this point using the MDC as an example.[11] We show in Appendix A with a simple example that for MDC, $M$ could even be exponential of $\frac{1-\sigma_i}{1-\sigma_0}$. This problem could be very serious since usually $\sigma$ could be very close to 1.

## 5. Improved Approximation Algorithm

As illustrated in the previous section, the runtime of Algorithm 1 linearly depends on $M$, which may be exponential in model inputs. In this section, we develop a much faster algorithm that is polynomial in $\frac{1}{\eta}$ and $\log M$ for a general class of choice models that include MDC.

Recall our definition of $M$ in Section 4.3:

$$M = \max_{\underline{x_0} \le y_0 < x_0 \le \overline{x_0}, j \in \mathcal{Q}} \left( \frac{x_j(x_0) - x_j(y_0)}{x_j(x_0)(x_0 - y_0)} \right).$$

Our improved algorithm is based on the following assumption.

**Assumption 3** *The reciprocal of the lower bound of the optimal outside market share $\frac{1}{\underline{x_0}}$ is upper bounded by a polynomial function of $M$.*

We can verify that all MDC satisfy Assumption 3 (see the following theorem).

**Theorem 7** *The MDC model, and hence the MEM, satisfy Assumption 3.*

The proof of Theorem 7 is presented in Appendix C.4. This result shows that Assumption 3 is not restrictive, but holds for interesting models. Now we solve the assortment optimization problem (P10) under Assumption 3.

Given $x_0$, the inner problem of (P10) is a knapsack problem with non-integer parameters that can be solved approximately using the FPTAS (Algorithm 3 in Appendix C.2). The FPTAS requires $O\left(Q^2 \lfloor \frac{Q}{\epsilon} \rfloor + QC\right)$ time, where $QC$ accounts for the calculation of the 'weights' and 'profits' in the corresponding knapsack problem.

We now provide an approximation algorithm for (P10), which improves over Algorithm 1. Recall that $x_0^* \in [\underline{x_0}, \overline{x_0}]$ (Lemma 4).

[11] Due to our more rigorous analysis, the runtime of the algorithm in Zhang et al. (2021) is also not necessarily polynomial in both $\frac{1}{\eta}$ and the problem inputs unless with additional assumptions though they use a slightly different definition of $M$.

**Algorithm 2** Improved Approximation Scheme under Assumption 3

**Input** $\eta$, $\underline{x_0}$, $\overline{x_0}$, $M$

1: $\epsilon \leftarrow \frac{\eta}{2}$
2: $\psi \leftarrow \frac{\eta}{2M}$
3: **for** $x_0 \leftarrow \{\underline{x_0}(1+\psi)^l | l \in \{0, 1, ..., m\}$ `such that` $\underline{x_0}(1+\psi)^m \le \overline{x_0}$ `and` $\underline{x_0}(1+\psi)^{m+1} > \overline{x_0}$ $\}$ **do**
4: Use knapsack FPTAS with algorithm parameter $\epsilon$ to approximate the inner problem in Problem (P10)
5: Calculate and store the assortment and the corresponding objective function value of the inner problem in Problem (P10).
6: Find the assortment with the greatest objective function value, denoted as $\mathcal{S}^*$.
7: **return** $\mathcal{S}^*$

Now we present the main theorem of this section.

**Theorem 8** *Algorithm 2 computes a $(1-\eta)$-approximation solution with computational time $\mathcal{O}\left(\frac{\log(\frac{1}{\underline{x_0}})}{\log(1+\frac{\eta}{2M})}\left(Q^2\lfloor\frac{Q}{\epsilon}\rfloor + QC\right)\right)$. In addition, based on Assumption 3, it is polynomial in $1/\eta$ and $\log(M)$.*

The proof of Theorem 8 is presented in Appendix C.5 [12].

The computational complexity of Algorithm 2 is significantly improved compared to Algorithm 1 (see Appendix C.6 for a more detailed analysis). This improvement is mainly due to the use of non-even intervals in the search space of the optimal outside market share in Algorithm 2 (Step 3), which takes advantage of Assumption 3 and enhances the efficiency of the algorithm. This result provides important insights when designing similar algorithms.

## 6. Case Study with E-commerce Data

In this section, we work on an E-commerce event history data in cosmetics shops (eCommerce Events History in Cosmetics Shop, REES46 for eCommerce) to examine the predictive performance of different choice models in the presence of multi-purchase behavior.

### 6.1. Data Preparation and Summary

The data set consists of over 20 million customer transactions on products from a medium-sized cosmetics online store, covering a period of 5 months from October 2019 to February 2020. The products are organized into 12 different categories: 'appliances environment vacuum', 'stationery cartrige', 'apparel glove', 'accessories bag', 'furniture living room cabinet', 'furniture bathroom

[12] Theorem 8 for the MEM case is included in the earlier version of our paper, appeared online in August, 2021.

bath', 'appliances personal hair cutter', 'accessories cosmetic bag', 'appliances environment air conditioner', 'furniture living room chair', 'sport diving' and 'appliances personal massager'.

The purchasing process can be broken down into four stages: product browsing, adding products to the shopping cart, removing products from the shopping cart, and completing the purchase. Customers can view multiple products, and they may add and remove products multiple times before completing the purchase. In one transaction, a customer may choose not to make a purchase, buy a single product, or buy multiple products (including multiple units of the same product or multiple products from different brands). It's also possible for a customer to add, delete, and then re-add the same product during the purchasing process.

We will apply the MNL model, the MDC model, and the p-MDC model (equivalently, MEM) to this data set. Our focus will be on analyzing the behavior of customers within each category separately, rather than cross-category purchases. In addition, we will apply several MDMs that are not special cases of the MDC model: the Marginal Moment Model (MMM), MDMs with marginal Gumbel distribution (MGM), marginal logistic distribution (MLM), marginal Cauchy distribution (MCM) and marginal uniform distribution (MUM).

As an example, we will use the data set for the category 'stationery cartridge' for the month of October, and present the procedure and results. The results corresponding to the rest of the other categories are presented later. In the 'stationery cartrige' data set, there are 136 products and 3073 customers (also referred to as data points in subsequent discussion), where for each customer, we see a set of products viewed by the customer, a set of products added into the shopping cart, a set of products removed from the shopping cart and a set of products purchased eventually (notice that if no product is purchased, we assume that the dummy product 0 is purchased). Without loss of generality, we always assume that the dummy product 0 was viewed by every customer. Out of the 3073 customers in the data set, 598 made a purchase (which may have included multiple products). To capture the multiple choice behavior of these customers, we assume that each customer purchases at most $\kappa - 1$ products. In cases where the number of products purchased by a customer is less than $\kappa - 1$, we assume that they fill the remaining purchase capacity with the dummy product 0 (the outside market option). This allows us to model the multiple choice behavior of customers who purchase more than one product. We observe that only 40 customers out of the 3073 in the data set purchased more than 5 products at the same time, which represents a small portion (only 1.3%) of all the data. As a result, we decide to remove these data points and set $\kappa = 6$. We also find that the market share of the products are only 6.1%, which means that 93.9% of the market share is on the outside market. In other words, the data is very imbalanced, which is common for the real-world choice data, in particular for e-commerce data. Working with this data set may not lead to meaningful or insightful results, as a simple model that always predicts that customers will choose the outside market option would likely outperform other models. To address the issue of imbalanced data, we apply the 'upsampling' technique (see Nallamuthu 2020 for details), which involves duplicating the data points that belong to the minority (i.e., those in which a customer made a purchase). Specifically, we make 19 additional copies of each data point in the original data set in which a customer purchased at least one product. This results in a modified data set with 14435 data points in total, of which 11960 customers made a purchase and 6840 of those purchased more than 1 product. There are 800 customers who purchased more than 5 products at the same time, which is still a small portion. As before, we set $\kappa = 6$ and remove these 800 data points. As a result, the 136 products in the 'stationery cartrige' data set now capture 26.94% of the market. Additionally, we discover that there are many products with very small market shares (e.g., in the original data set, only one customer made a purchase of one unit of some products). To make the data set more manageable, we ignore these products and only consider the top 21 products that capture 23.72% of the market share. To 'ignore' these products, we simply remove them from the data set. Customers that only view the dummy product 0 are also removed. Additionally, we remove any data points with a view set containing more than 7 products, in order to keep the estimation of the MDC model tractable. The details of this decision will be discussed later.

### 6.2. Estimation on the Data

For each data point, we treat the 'view set' as the assortment of products available to the customer, and the 'purchase set' as the choice made by the customer. To estimate the model parameters, we use the 'maximum likelihood' (MLE) approach. The MLE estimation for a choice model is as follows:

$$\begin{aligned} \max_{\boldsymbol{\alpha}} \quad & \sum_{i \in I} \Big[ \sum_{j \in \mathcal{S}_i} \left[ d_{j,\mathcal{S}_i} \log(x_{j,\mathcal{S}_i}(\boldsymbol{\alpha})) \right] \Big] \\ \text{s.t.} \quad & \sum_{j \in \mathcal{S}_i} x_{j,\mathcal{S}_i}(\boldsymbol{\alpha}) = 1, \forall i \in I, \end{aligned}$$

where $\boldsymbol{\alpha}$ is the parameters involved in the corresponding choice model (e.g., linear-in-parameters of the deterministic utilities and the scale parameters for the distributions of the error terms); $I$ is the set of customer indices; $\mathcal{S}_i$ is the corresponding assortment provided in $i$th data point [13]; $d_{j,\mathcal{S}_i}$ is the realized market share (scaled by $\kappa$) of product $j$ given assortment $\mathcal{S}_i$ and $x_{j,\mathcal{S}_i}$ is the calculated choice probability of product $j$ given assortment $\mathcal{S}_i$ based on the parameters of the choice model. Recall that MNL, p-MDC and MDC are all special cases of the MDM. Note that the optimal solution of the above MLE optimization program stays the same when the constraints are '=' or '≤', because at the optimal solution the constraints are always binding.

[13] Remember that the dummy product 0 has been added into the set of products viewed by the customer, so $\mathcal{S}_i$ includes the outside market option.

The MEM (or equivalently, p-MDC model) estimation was introduced in Mishra et al. (2014) and the MNL model can be easily estimated within the MEM framework by setting all exponential parameters to 1. In fact, the MMM, the MGM, the MLM and the MCM can all be estimated using this approach with different marginal c.d.f.s. The estimation of the MDC model and the MUM are more complicated. Recall that Theorem 1 and Proposition 1 imply that in the MDC model, the choice probability of product $j, \forall j \in \mathcal{Q}$ is

$$x_j = \begin{cases} 0, & \text{if } \lambda > \kappa\psi_i\sigma_i(\frac{1}{\gamma_i})^{1-\sigma_i}, \\ 1 - F_j(\lambda), & \text{if } \lambda \in \left[\kappa\psi_i\sigma_i(\frac{1}{\gamma_i+\kappa})^{1-\sigma_i}, \kappa\psi_i\sigma_i(\frac{1}{\gamma_i})^{1-\sigma_i}\right], \\ 1, & \text{if } \lambda < \kappa\psi_i\sigma_i(\frac{1}{\gamma_i+\kappa})^{1-\sigma_i}, \end{cases}$$

and the choice probability of the outside market option is

$$x_0 = \begin{cases} 1 - F_j(\lambda), & \text{if } \lambda \geq \kappa\psi_i\sigma_i(\frac{1}{\kappa})^{1-\sigma_i}, \\ 1, & \text{if } \lambda < \kappa\psi_i\sigma_i(\frac{1}{\kappa})^{1-\sigma_i}, \end{cases}$$

where

$$\begin{cases} F_i(t) = 1 + \dfrac{\gamma_i}{\kappa} - \dfrac{(\frac{\kappa\psi_i\sigma_i}{t})^{1/(1-\sigma_i)}}{\kappa}, t \in \left[\kappa\psi_i\sigma_i(\dfrac{1}{\gamma_i+\kappa})^{1-\sigma_i}, \kappa\psi_i\sigma_i(\dfrac{1}{\gamma_i})^{1-\sigma_i}\right], \forall j \in \mathcal{Q}, \\ F_0(t) = 1 - \dfrac{\left(\frac{\kappa\psi_0\sigma_0}{t}\right)^{1/(1-\sigma_0)}}{\kappa}, t \in \left[\kappa\psi_0\sigma_0(\dfrac{1}{\kappa})^{1-\sigma_0}, \infty\right). \end{cases}$$

To solve the optimization problem, we may need to make some reformulations to eliminate the 'if statement' in the functions. First, it is worth noting that we can ignore situations where

$$\begin{aligned} &x_j = 1, \text{if } \lambda < \kappa\psi_i\sigma_i(\frac{1}{\gamma_j+\kappa})^{1-\sigma_i}, \forall j \in \mathcal{Q}, \text{ and} \\ &x_0 = 1, \text{if } \lambda < \kappa\psi_i\sigma_i(\frac{1}{\kappa})^{1-\sigma_i}. \end{aligned}$$

It is because in all of these MDMs, the constraints in the MDM-MLE estimation formulation implies that for such $\lambda$, the corresponding $x_j$ or $x_0$ can never be optimal. This is why the other models (except for the MUM) are relatively simple to estimate. However, in the MDC model, we have an additional situation, in which $x_j = 0$ if $\lambda > \kappa\psi_i\sigma_i(\frac{1}{\gamma_i})^{1-\sigma_i}$. To address this issue, we first note that in MLE approach, if the realized choice probability corresponding to a product is positive given an assortment, its calculated choice probability from the MDC model must also be positive, since otherwise the objective function is negative infinity. Second, notice that in the MLE formulation, we have the following set of constraints:

$$\sum_{j \in \mathcal{S}_i} x_{j,\mathcal{S}_i}(\boldsymbol{\alpha}) = 1, \forall i \in I.$$

| **stationery cartrige** (original) | | | |
|---|---|---|---|
| data number | product number | outside market purchases | |
| 3073 | 136 | 2475 | |
| stationery cartrige (scale-up factor = 20, $\kappa = 6$, remove small-share products) | | | |
| data number | product number | outside market share | |
| 11144 | 21 | 76.28% | |
| Model | in-sample loglklhd | out-of-sample RMSE | time/iteration number |
| the MNL model | -6241.41 | 0.10820 | 1.25s / 6 |
| the p-MDC model | -6211.47 | 0.10609 | 3.30s / 85 |
| the MDC model | -6208.69 | 0.10605 | 267.38s / 1023 |
| the MMM | -6213.90 | 0.10623 | 62.25s / 1246 |
| the MGM | -6214.38 | 0.10627 | 5.69s / 71 |
| the MLM | -6217.20 | 0.10648 | 5.84s / 64 |
| the MCM | -6217.49 | 0.10645 | 6.31s / 82 |
| the MUM | -6210.59 | 0.10624 | 60.99s / 212 |

**Table 2 Results for 'stationery cartrige' Data**

For each $i \in I$, we replace the constraint with a set of constraints as follows:

$$\sum_{j \in T} 1 - F_j(\lambda_{\mathcal{S}_i,j}) \le 1, \forall\, T \in \Theta(\mathcal{S}_i),$$

where $\Theta(\mathcal{S}_i)$ denotes the set of all possible subsets of $\mathcal{S}_i$ that contains the outside option. Clearly, $|\Theta(\mathcal{S}_i)| = 2^{|\mathcal{S}_i|} - 1$. With this set of inequalities, although $1 - F_j(\lambda_{\mathcal{S}_i,j})$ may be negative for some $j \in T \in \Theta(\mathcal{S}_i)$, the constraint $\sum_{j \in \mathcal{S}_i} x_{j,\mathcal{S}_i} \le 1, \forall i \in I$ is always obeyed. Clearly, the number of additional constraints required in the optimization problem is exponential in the number of products in the feasible assortments, which may cause the computational difficulty in the estimation procedure. This is why we removed all data points with assortments containing more than 7 products, as this helps to reduce the number of constraints and make the estimation less tractable. Our approach of estimating the MDC model is not necessarily the best, but it is easy to interpret and implement. On the other hand, alternative methods in Bhat (2008) are complicated and not immediately applicable in the assortment context. A more efficient estimation of MDC would be valuable, and therefore, this is an important future research direction. We apply the similar estimation technique for the MUM.

### 6.3. Results

The result of 'stationery cartrige' data is provided in Table 2. Initially, the outside market held a 93.9% share of the market (as shown in Section 6.1), but after using the upsampling technique, the outside market's share decreased to 76.28%. As the model becomes more complex (the other MDMs v.s., the MNL model), the in-sample log-likelihood and the out-of-sample RMSE both improve, as can be seen in Table 2. The results for other categories shown in Table 4 and Table 5

| Average Ranking | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | MNL | p-MDC | MDC | MMM | MGM | MLM | MCM | MUM |
| in-sample log-likelihood | 8 | 6.2 | 1.8 | 3.4 | 3.8 | 4.6 | 5.6 | 2.6 |
| out-of-sample RMSE | 8 | 2.8 | 3.8 | 2.4 | 4.8 | 4.8 | 4.2 | 5.2 |

**Table 3 Average Ranking of The Performance of Each Model in Both In-Sample Log-Likelihood Test and Out-of-Sample RMSE Test**

(found in Appendix D) show similar trends to the stationary cartridge data. As expected, the MNL model always performs the worst. In Table 3, we illustrate the average ranking of the performance of each model in all categories. It shows that although the MDC model aces in the in-sample log-likelihood test, it is not the best when considering the out-of-sample root-mean-square error (RMSE). This may be due to overfitting, as the dataset does not contain enough information to justify the flexibility of MDC. To ensure that the issue of imbalanced data has been addressed, we also tested a “naive model,” in which all customers only chose the outside market option. In this model, the probability of choosing the outside market was 1, while the probability for the other options was 0. The out-of-sample RMSE for this model was 0.2143, which is significantly worse than the other models. In terms of estimation time, the MNL model is the fastest to converge, followed by the other models. The MDC model and the MUM take the longest to converge and may exceed the maximum number of iterations. Furthermore, it can be seen that although not designed so, many MDMs that are not embedded by the MDC model (the MMM, the MGM, the MLM, the MCM, and the MUM) have a good performance in capturing the multi-purchase choice behaviour. Overall, the p-MDC model and the MMM strike a good balance between simplicity and the ability to model multi-purchase behavior, resulting in fast and accurate parameter estimation.

## 7. Conclusion

In this study, we examine a single-period product assortment problem in which the choice model is defined through the marginals of noise terms in utilities. We provide several theoretical results for this assortment problem under the MDC model, the MEM, and all MDMs. Our findings include sufficient conditions under which the profit-nested heuristic is optimal for these models, and the profit-nested heuristic offers a half-approximation guarantee for all MDMs. In addition, we propose an algorithm for approximating the solution to this problem for all MDMs, as well as an improved version for MDC and MEM. This framework allows for the examination of various existing models, including multi-purchase choice models, and we believe that it could be valuable to explore incorporating constraints in future research. Our technical results and proofs offer new insights into the field of discrete choice theory and assortment optimization and may inspire future theoretical developments.

## Acknowledgments

The authors gratefully acknowledge Prof. Huseyin Topaloglu for his valuable comments and suggestions on an earlier version of the paper. We also thank Prof. Karthik Natarajan for valuable discussions over the years.

## References

Ahipaşaoğlu SD, Arıkan U, Natarajan K (2016) On the flexibility of using marginal distribution choice models in traffic equilibrium. *Transportation Research Part B: Methodological* 91:130–158.

Ahipaşaoğlu SD, Arıkan U, Natarajan K (2019) Distributionally robust markovian traffic equilibrium. *Transportation Science* 53(6):1546–1562.

Anderson SP, De Palma A, Thisse JF (1988) A representative consumer theory of the logit model. *International Economic Review* 29(3):461–466.

Andonov R, Poirriez V, Rajopadhye S (2000) Unbounded knapsack problem: Dynamic programming revisited. *European Journal of Operational Research* 123(2):394–407.

Aouad A, Feldman J, Segev D (2018) The exponomial choice model: Algorithmic frameworks for assortment optimization and data-driven estimation case studies, available at SSRN 3192068.

Berbeglia G, Joret G (2020) Assortment optimisation under a general discrete choice model: A tight analysis of revenue-ordered assortments. *Algorithmica* 82(4):681–720.

Bernstein F, Modaresi S, Sauré D (2019) A dynamic clustering approach to data-driven assortment personalization. *Management Science* 65(5):2095–2115.

Bhat CR (2005) A multiple discrete–continuous extreme value model: formulation and application to discretionary time-use decisions. *Transportation Research Part B: Methodological* 39(8):679–707.

Bhat CR (2008) The multiple discrete-continuous extreme value (mdcev) model: role of utility function parameters, identification considerations, and model extensions. *Transportation Research Part B: Methodological* 42(3):274–303.

Bhat CR (2018) A new flexible multiple discrete–continuous extreme value (mdcev) choice model. *Transportation Research Part B: Methodological* 110:261–279.

Block HD, Marschak J, et al. (1959) Random orderings and stochastic theories of response. Technical report, Cowles Foundation for Research in Economics, Yale University.

Cao Y, Rusmevichientong P, Topaloglu H (2020) Revenue management under a mixture of multinomial logit and independent demand models, working paper.

Clark SA (1990) A concept of stochastic transitivity for the random utility model. *Journal of Mathematical Psychology* 34(1):95–108.

Cox DR (1972) The analysis of multivariate binary data. *Applied statistics* 113–120.

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

El Housni O, Topaloglu H (2021) Joint assortment optimization and customization under a mixture of multinomial logit models: On the value of personalized assortments. Available at SSRN 3830082.

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

Feldman J, Segev D, Topaloglu H, Wagner L, Bai Y (2021) Assortment optimization under the multi-purchase multinomial logit choice model. *Available at SSRN 3866734* .

Feldman JB, Topaloglu H (2015) Capacity constraints across nests in assortment optimization under the nested logit model. *Operations Research* 63(4):812–822.

Feldman JB, Topaloglu H (2017) Revenue management under the markov chain choice model. *Operations Research* 65(5):1322–1342.

Feng G, Li X, Wang Z (2018) On substitutability and complementarity in discrete choice models. *Operations Research Letters* 46(1):141–146.

Fiorini S (2004) A short proof of a theorem of Falmagne. *Journal of Mathematical Psychology* 48(1):80–82.

Fudenberg D, Iijima R, Strzalecki T (2015) Stochastic choice and revealed perturbed utility. *Econometrica* 83(6):2371–2409.

Gallego G, Berbeglia G (2021) Limits of personalization: Prophet inequalities for revenue-ordered assortments. *arXiv preprint arXiv:2109.14861* .

Gallego G, Iyengar G, Phillips R, Dubey A (2004) Managing flexible products on a network, available at SSRN 3567371.

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

Gallego G, Topaloglu H (2019) *Revenue management and pricing analytics*, volume 209 (Springer).

Gallego G, Wang R (2019) Threshold utility model with applications to retailing and discrete choice models, available at SSRN 3420155.

Harsanyi JC (1973) Oddness of the number of equilibrium points: a new proof. *International Journal of Game Theory* 2(1):235–250.

Hofbauer J, Sandholm WH (2002) On the global convergence of stochastic fictitious play. *Econometrica* 70(6):2265–2294.

Huh WT, Li H (2022) Optimal pricing under multiple-discrete customer choices and diminishing return of consumption. *Operations Research* 70(2):905–917.

Insider (2019) URL `https://www.insiderintelligence.com/chart/236341/Personalization-Aspects-that-Retail-Marketers-Worldwide-Currently-Use-vs-Want-Use-Q4-2019-of-respondents`.

Kim J, Allenby GM, Rossi PE (2002) Modeling consumer demand for variety. *Marketing Science* 21(3):229–250.

Lai K, Goemans M (2006) The knapsack problem and fully polynomial time approximation schemes (FPTAS). *Retrieved November* 3:2012.

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

Lin H, Li X, Wu L (2022) E-commerce assortment optimization and personalization with multiple-choice rank list model. *Available at SSRN* .

Liu Q, Van Ryzin G (2008) On the choice-based linear programming model for network revenue management. *Manufacturing & Service Operations Management* 10(2):288–310.

Luce RD (1959) *Individual choice behavior* (Wiley).

Lyu C, Jasin S, Najafi S, Zhang H (2021) Assortment optimization with multi-item basket purchase under the multivariate MNL model. *Available at SSRN 3818886* .

Machina MJ (1985) Stochastic choice functions generated from deterministic preferences over lotteries. *The Economic Journal* 95(379):575–594.

Mattsson LG, Weibull JW (2002) Probabilistic choice and procedurally bounded rationality. *Games and Economic Behavior* 41(1):61–78.

McCardle KF, Rajaram K, Tang CS (2007) Bundling retail products: Models and analysis. *European Journal of Operational Research* 177(2):1197–1217.

Mishra VK, Natarajan K, Padmanabhan D, Teo CP, Li X (2014) On theoretical and empirical aspects of marginal distribution choice models. *Management Science* 60(6):1511–1531.

Nallamuthu N (2020) Handling imbalanced data-machine learning computer vision and NLP. Data Science Blogathon.

Natarajan K, Song M, Teo CP (2009) Persistency model and its applications in choice modeling. *Management Science* 55(3):453–469.

Rosenthal RW (1989) A bounded-rationality approach to the study of noncooperative games. *International Journal of Game Theory* 18(3):273–292.

Ruan Y, Li X, Murthy K, Natarajan K (2022) The limit of the marginal distribution model in consumer choice. URL `http://dx.doi.org/10.48550/ARXIV.2208.06115`.

Rusmevichientong P, Shen ZJ, Shmoys DB (2010) Dynamic assortment optimization with a multinomial logit choice model and capacity constraint. *Operations Research* 58(6):1666–1680.

Russell GJ, Petersen A (2000) Analysis of cross category dependence in market basket selection. *Journal of Retailing* 76(3):367–392.

Sauré D, Zeevi A (2013) Optimal dynamic assortment planning with demand learning. *Manufacturing & Service Operations Management* 15(3):387–404.

Seetharaman P, Chib S, Ainslie A, Boatwright P, Chan T, Gupta S, Mehta N, Rao V, Strijnev A (2005) Models of multi-category choice behavior. *Marketing Letters* 16(3):239–254.

Swait J, Marley AA (2013) Probabilistic choice (models) as a result of balancing multiple goals. *Journal of Mathematical Psychology* 57(1-2):1–14.

Train KE (2009) *Discrete choice methods with simulation* (Cambridge university press).

Tulabandhula T, Sinha D, Patidar P (2020) Multi-purchase behavior: Modeling and optimization. *arXiv preprint arXiv:2006.08055* .

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

van Ryzin G, Vulcano G (2015) A market discovery algorithm to estimate a general class of nonparametric choice models. *Management Science* 61(2):281–300.

Yan Z, Natarajan K, Teo CP, Cheng C (2022) A representative consumer model in data-driven multiproduct pricing optimization. *Management Science* 68(8):5557–6354.

Zhang H, Piri H, Huh WT, Li H (2021) Assortment optimization under multiple-discrete customer choices. *Available at SSRN 3988982* .

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

# Appendix

## A. An Example of the Assortment problem under the MDC model

We consider an assortment problem under the MDC model with the following parameters:

$$\begin{cases} \kappa = 1, \\ \sigma_j = \sigma, \forall i \in \mathcal{Q}, \\ \psi_j = \psi, \forall i \in \mathcal{Q} \cup \{0\}, \\ \gamma_j \to 0, \forall i \in \mathcal{Q}, \\ \theta := \left(\frac{\psi\sigma}{\psi_0\sigma_0}\right)^{\frac{1}{1-\sigma}}. \end{cases}$$

Then, the choice probability of product $j$ is

$$x_j = \theta \cdot (x_0)^{\frac{1-\sigma_0}{1-\sigma}}, \forall j \in \mathcal{Q}.$$

Let $x_j'(x_0)$ be the first derivative of $x_j(x_0)$ with respect to $x_0$. Then we have

$$\begin{aligned} M &= \max_{\underline{x_0} \le y_0 < x_0 \le \overline{x_0}, j \in \mathcal{Q}} \left(\frac{x_j(x_0) - x_j(y_0)}{x_j(x_0)(x_0 - y_0)}\right) \\ &\ge \left(\frac{x_j(x_0) - x_j(y_0)}{x_j(x_0)(x_0 - y_0)}\right)\Bigg| y_0 \to \underline{x_0}, x_0 \to y_0 + \delta, \delta \to 0 \\ &= \lim_{\delta \to 0} \left(\frac{x_j(\underline{x_0} + \delta) - x_j(\underline{x_0})}{x_j(\underline{x_0} + \delta)(\delta)}\right) \\ &= \lim_{\delta \to 0} \frac{x_j'(\underline{x_0})}{x_j(\underline{x_0} + \delta)} \\ &= \frac{x_j'(\underline{x_0})}{x_j(\underline{x_0})}, \end{aligned}$$

for any $j \in \mathcal{Q}$. Hence,

$$\begin{aligned} M &\ge \frac{x_j'(\underline{x_0})}{x_j(\underline{x_0})} \\ &= \frac{\frac{1-\sigma_0}{1-\sigma}\theta \cdot (\underline{x_0})^{(\frac{1-\sigma_0}{1-\sigma} - 1)}}{\theta \cdot (\underline{x_0})^{\frac{1-\sigma_0}{1-\sigma}}} \\ &= \frac{1-\sigma_0}{1-\sigma}(\underline{x_0})^{-1}. \end{aligned}$$

From Lemma 4, we know

$$\sum_{i \in \mathcal{Q}} \left(\theta \cdot (\underline{x_0})^{\frac{1-\sigma_0}{1-\sigma}}\right) + \underline{x_0} = 1.$$

Now, we let $z$ be the solution of equation

$$\sum_{i \in \mathcal{Q}} \left(\theta \cdot (z)^{\frac{1-\sigma_0}{1-\sigma}}\right) = 1.$$

Then, clearly we have $\underline{x_0} < z$. Rearrange the terms, we have $z = \left(\frac{1}{\theta Q}\right)^{\frac{1-\sigma}{1-\sigma_0}}$. Hence, we obtain

$$\begin{aligned} M &> \frac{1-\sigma_0}{1-\sigma}(z)^{-1} \\ &= \frac{1-\sigma_0}{1-\sigma}\left(\left(\frac{1}{\theta Q}\right)^{\frac{1-\sigma}{1-\sigma_0}}\right)^{-1} \\ &= \frac{1-\sigma_0}{1-\sigma} \cdot (\theta Q)^{\frac{1-\sigma}{1-\sigma_0}} \\ &= \frac{1-\sigma_0}{1-\sigma} \cdot \left(\frac{\psi\sigma}{\psi_0\sigma_0}\right)^{\frac{1}{1-\sigma_0}} (Q)^{\frac{1-\sigma}{1-\sigma_0}}. \end{aligned}$$

It can be seen that $M$ is not polynomial in problem inputs, in particular, $\sigma$ and $\sigma_0$.

## B. Proofs

### B.1. Proof of Lemma 1.

**Proof.** Note that such $c_j'(\cdot), \forall j \in \mathcal{Q} \cup \{0\}$ are invertible. We take the following derivatives

$$\frac{d}{dx_j} \int_{1-x_j}^{1} F_j^{-1}(t)dt = F_j^{-1}(1-x_j).$$

For any valid $c_j'(\cdot)$ (the first-order derivative of a valid perturbed cost function for product $j$), we let $F_j^{-1}(y) = -c_j'(1-y)$ for $y \in [0,1]$. Hence, we have $F_j^{-1}(1-x_j) = -c_j'(x_j)$ for $x_j \in [0,1], \forall j \in \mathcal{Q} \cup \{0\}$. It is easy to see that by this construction, $F_j(\cdot), \forall j \in \mathcal{Q} \cup \{0\}$ is a valid MDM c.d.f.. Notice that if two functions with the same domain $[0,1]$ have the same derivative everywhere, then they differ by a constant. That is, we have $-c_j(x_j) + M_j = \int_{1-x_j}^{1} F_j^{-1}(t)dt$, where $M_j$ is a constant. Now, consider a separable perturbation function $V(\boldsymbol{x}) = \sum_{j \in \mathcal{Q} \cup \{0\}} c_j(x_j)$ and another function $V^*(\boldsymbol{x}) = \sum_{j \in \mathcal{Q} \cup \{0\}} c_j^*(x_j)$, where $c_j^*(\cdot) = c_j(\cdot) - M_j$ for some constants $M_j$. It is easy to see that the following problem

$$\max \left\{ \sum_{j \in \mathcal{S} \cup \{0\}} v_j x_j - V^*(\boldsymbol{x}) : \sum_{j \in \mathcal{S} \cup \{0\}} x_j = 1, x_j \geq 0, \forall j \in \mathcal{S} \cup \{0\}, x_j = 0, \forall j \notin \mathcal{S} \cup \{0\} \right\},$$

is equivalent to problem (P1), which also turns out to be equivalent to problem (P3).

■

### B.2. Proof of Theorem 1.

**Proof.** First, we show that $F_0(\cdot)$ and $F_i(\cdot), \forall i \in \mathcal{Q}$ are valid c.d.f.'s for the MDM. Notice that $F_i(t)$ is a continuous and increasing function, $F_i\left(\kappa\psi_i\sigma_i(\frac{1}{\gamma_i+\kappa})^{1-\sigma_i}\right) = 0$ and $F_i\left(\kappa\psi_i\sigma_i(\frac{1}{\gamma_i})^{1-\sigma_i}\right) = 1$, for all $i \in \mathcal{Q}$. Similarly, $F_0(t)$ is a continuous and increasing function, $F_0\left(\kappa\psi_0\sigma_0(\frac{1}{\kappa})^{1-\sigma_0}\right) = 0$ and $F_0(\infty) \to 1$. Hence, we have a valid MDM.

Next, we show that this MDM is equivalent to a MDC with corresponding parameters. To do so, let us first calculate the inverse functions of the marginal distributions, $F_i^{-1}(t) = \kappa\psi_i\sigma_i\Big(\kappa(1-t)+\gamma_i\Big)^{\sigma_i - 1}, \forall i \in \mathcal{Q} \cup \{0\}$. Then, we have

$$\begin{aligned}
\int_{1-x_i}^{1} F_i^{-1}(t)\, dt &= -\psi_i(\kappa - \kappa t + \gamma_i)^{\sigma_i}\Big|_{t=1-x_i}^{t=1} \\
&= -\psi_i\gamma_i^{\sigma_i} - \big(-\psi_i(\kappa - \kappa + \kappa x_i + \gamma_i)^{\sigma_i}\big) \\
&= -\psi_i\gamma_i^{\sigma_i} + \psi_i(\kappa x_i + \gamma_i)^{\sigma_i}.
\end{aligned}$$

Following the MDM formulation (P3), we have

$$\begin{aligned}
\max_{x_i \in [0,1], \forall i \in \mathcal{S} \cup \{0\}} \quad & \sum_{i \in \mathcal{S} \cup \{0\}} \big(\psi_i(\kappa x_i + \gamma_i)^{\sigma_i} - \psi_i\gamma_i^{\sigma_i}\big) \\
\text{s.t.} \quad & \sum_{i \in \mathcal{S} \cup \{0\}} x_i = 1.
\end{aligned}$$

Since $F_0(\cdot)$ is supported on a semi-definite interval $[\kappa\psi_0\sigma_0(\frac{1}{\kappa})^{1-\sigma_0}, \infty)$, it is easy to verify that $x_0$ must be strictly positive. What is more, notice that the second term in the objective function in each term in the summation is not a function of the decision variables, hence we can simply remove it. Moreover, since the objective function is increasing in terms of $x_j, \forall j \in \mathcal{S} \cup \{0\}$ , we have the following equivalent model,

$$\begin{aligned} \max_{x_i \in [0,1], \forall i \in \mathcal{S}, x_0 \in (0,1]} \quad & \sum_{i \in \mathcal{S} \cup \{0\}} \left(\psi_i (\kappa x_i + \gamma_i)^{\sigma_i}\right) \\ \text{s.t.} \quad & \sum_{i \in \mathcal{S} \cup \{0\}} x_i \leq 1. \end{aligned}$$

Now, we apply a *change of variables.* Let $y_j = \kappa x_j, \forall j \in \mathcal{S} \cup \{0\}$. As a result, we obtain the following equivalent model,

$$\begin{aligned} \max_{y_i \geq 0, \forall i \in \mathcal{S}, y_0 > 0} \quad & \psi_0 y_0^{\sigma_0} + \sum_{j \in \mathcal{S}} \psi_i (y_i + \gamma_i)^{\sigma_i} \\ \text{s.t.} \quad & \sum_{i \in \mathcal{S} \cup \{0\}} y_i \leq \kappa, \end{aligned}$$

which is indeed a MDC (recall that $\gamma_0 = 0$).

■

### B.3. Proof of Theorem 2.

Based on Theorem 1, Proposition 1 in Section 4 implies that the choice probability corresponding to product $j$ in the MDC model is $x_j(x_0) = \left[ \left( \frac{\psi_j \sigma_j}{\psi_0 \sigma_0} \right)^{\frac{1}{1-\sigma_j}} (\kappa)^{\frac{1-\sigma_0}{1-\sigma_j} - 1} (x_0)^{\frac{1-\sigma_0}{1-\sigma_j}} - \frac{\gamma_j}{\kappa} \right]^+$, where $[a, 0]^+ = \max\{a, 0\}$. In p-MDC, where $\gamma_j = 0, \forall j \in \mathcal{Q}$, we have $x_j(x_0) = \left( \frac{\psi_j \sigma_j}{\psi_0 \sigma_0} \right)^{\frac{1}{1-\sigma_j}} (\kappa)^{\frac{1-\sigma_0}{1-\sigma_j} - 1} (x_0)^{\frac{1-\sigma_0}{1-\sigma_j}}$. Recall from Section 3.1, under the MEM, the choice probability corresponding to product $j$ is $x_j(x_0) = e^{\alpha_j v_j} (x_0)^{\frac{\alpha_j}{\alpha_0}}$. It is easy to see that when $\alpha_j = \frac{1}{1-\sigma_j}, \forall j \in \mathcal{Q} \cup \{0\}$ and $v_j = \log(\frac{\psi_j \sigma_j}{\psi_0 \sigma_0} \cdot (\kappa)^{\sigma_j - \sigma_0}), \forall j \in \mathcal{Q}$, the choice probabilities of product $j$ from the two models are identical. One may notice that in the MEM, $\alpha_j > 0, \forall j \in \mathcal{Q} \cup \{0\}$, however, the MDC parameters $\sigma_j \in (0,1), \forall j \in \mathcal{Q} \cup \{0\}$ suggest that $\alpha_j > 1, \forall j \in \mathcal{Q} \cup \{0\}$. Note that this does not break the equivalence as for any MEM model with some $\alpha_j \in (0,1)$, we can always find an equivalent MEM in which all $\alpha_j \geq 1$ because we may always specify a value for the minimum $\alpha_j, \forall j \in \mathcal{Q} \cup \{0\}$.

Hence, for any p-MDC model, there always exists a MEM which recovers it and vice versa.

■

### B.4. Proof of Lemma 3.

**Proof.** The threshold utility model (TUM) is proposed in Gallego and Wang (2019) to model the choice behavior in retail applications. Recall the model (P5):

$$\left. \begin{aligned} \max_{z_i \geq 0, \forall i \in \mathcal{S} \cup \{0\}} \quad & \sum_{i \in \mathcal{S} \cup \{0\}} \mathrm{E}[U_i | U_i \geq z_i] \cdot \Pr(U_i \geq z_i) \\ \text{s.t.} \quad & \sum_{i \in \mathcal{S} \cup \{0\}} \Pr(U_i \geq z_i) = \kappa, \end{aligned} \right\}$$

where $\kappa$ is a predetermined bound on the expected number of products the consumer is willing to buy in a product category. Here we show that, if $\boldsymbol{U} = \boldsymbol{v} + \tilde{\boldsymbol{\epsilon}}$ and marginal distribution for $\tilde{\epsilon}_i$ is $F_i$ (assuming $F_i$ invertible), it is equivalent to formulation (P6). Start with the objective in (P5) and let $x_i = P(U_i \geq z_i) = P(\epsilon_i \geq z_i - v_i)$. Then we have

$$\sum_{i \in \mathcal{S} \cup \{0\}} E[u_i | u_i \geq z_i] P(u_i \geq z_i) = \sum_{i \in \mathcal{S} \cup \{0\}} v_i x_i + \sum_{i \in \mathcal{S} \cup \{0\}} E[\epsilon_i | \epsilon_i \geq z_i - v_i] P(\epsilon_i \geq z_i - v_i)$$

$$= \sum_{i\in\mathcal{S}\cup\{0\}} v_i x_i + \sum_{i\in\mathcal{S}\cup\{0\}} \int_{z_i-v_i}^{+\infty} \epsilon dF_i(\epsilon)$$
$$= \sum_{i\in\mathcal{S}\cup\{0\}} v_i x_i + \sum_{i\in\mathcal{S}\cup\{0\}} \int_{F_i(z_i-v_i)}^{1} F_i^{-1}(t)dt \quad (\text{letting } t = F_i(\epsilon))$$
$$= \sum_{i\in\mathcal{S}\cup\{0\}} v_i x_i + \sum_{i\in\mathcal{S}\cup\{0\}} \int_{1-x_i}^{1} F_i^{-1}(t)dt \quad (\text{because } x_i = P(\epsilon_i \geq z_i - v_i)),$$

which is the objective of (P6). The constraint in (P5) reduces to $\sum_{i=0}^{|\mathcal{S}|} x_i = \kappa$. From the definition of $x_i$, clearly we have $0 \leq x_i \leq 1$. Since the value of $z_i$ is free, we can get all possible values of $x_i \in [0,1]$. Hence, the model is equivalent to

$$\max\left\{ \sum_{j\in\mathcal{S}\cup\{0\}} \left(v_j x_j + \int_{1-x_j}^{1} F_j^{-1}(t)dt\right) : \sum_{j\in\mathcal{S}\cup\{0\}} x_j = \kappa, 0 \leq x_j \leq 1, \forall j \in \mathcal{S}\cup\{0\} \right\},$$

■

### B.5. Proof of Theorem 3.

**Proof.** Throughout the proof, we assume that the c.d.f.'s of the utility noise terms are invertible. Based on Lemma 3, when we let $\kappa = 1$, it is easy to see that TUM-1 and the MDM are equivalent.

Next, we will show that TUM-$\kappa$ can be reproduced from the MDM. In TUM-$\kappa$, let the marginal distribution for $\tilde{\epsilon}_j$ be $G_j$, where $\boldsymbol{U} = \boldsymbol{v} + \tilde{\boldsymbol{\epsilon}}$. Recall that TUM-$\kappa$ can be expressed as (Lemma 3):

$$\max_{x_j,\forall j\in\mathcal{S}\cup\{0\}} \left\{ \sum_{j\in\mathcal{S}\cup\{0\}} \left(v_j x_j + \int_{1-x_j}^{1} G_j^{-1}(t)dt\right) : \sum_{j\in\mathcal{S}\cup\{0\}} x_j = \kappa, 0 \leq x_j \leq 1, \forall j \in \mathcal{S}\cup\{0\} \right\}.$$

Let us introduce $y_j = \frac{1}{\kappa} x_j, \forall j \in \mathcal{S}\cup\{0\}$, then we obtain an equivalent model:

$$\max_{y_j,\forall j\in\mathcal{S}\cup\{0\}} \left\{ \sum_{j\in\mathcal{S}\cup\{0\}} \left(\kappa v_j y_j + \int_{1-\kappa y_j}^{1} G_j^{-1}(t)dt \cdot \mathbb{1}\{y_j \leq \frac{1}{\kappa}\}\right) : \sum_{j\in\mathcal{S}\cup\{0\}} y_j = 1, y_j \geq 0, \forall j \in \mathcal{S}\cup\{0\} \right\}, \quad \text{(P11)}$$

where

$$\mathbb{1}\left\{y_j \leq \frac{1}{\kappa}\right\} = \begin{cases} 1, \text{if } y_j \leq \frac{1}{\kappa}, \\ -\infty, \text{otherwise.} \end{cases}$$

Now, we consider a MDM where the marginal distributions, denoted as $F_j$, are defined as $F_j(x) = 1 - \frac{1}{\kappa}(1 - G_j(\frac{1}{\kappa}x))$, and the deterministic utility of alternative $j$ is $\kappa v_j$, for all $j \in \mathcal{Q}\cup\{0\}$. The corresponding model is

$$\max_{x_j,\forall j\in\mathcal{S}\cup\{0\}} \left\{ \sum_{j\in\mathcal{S}\cup\{0\}} \left(\kappa v_j x_j + \int_{1-x_j}^{1} F_j^{-1}(t)dt\right) : \sum_{j\in\mathcal{S}\cup\{0\}} x_j = 1, x_j \geq 0, \forall j \in \mathcal{S}\cup\{0\} \right\}. \quad \text{(P12)}$$

Notice that $F_j^{-1}$ exists since $G_j^{-1}$ exits. Moreover, from the following argument, we can see that $F_j$ is a valid c.d.f.. Let $X_j$ be the random variable that follows $F_j$. Then,

$$X_j = \begin{cases} Y_j \sim G_j^*, \text{ with probability } \frac{1}{\kappa}, \\ -\infty, \text{ with probability } 1 - \frac{1}{\kappa}, \end{cases}$$

where $G_j^*(x) = G_j(\frac{1}{\kappa}x)$. For $1 - F_j(x) = \frac{1}{\kappa}(1 - G_j(\frac{1}{\kappa}x))$, taking the inverse of both sides, we obtain

$$F_j^{-1}(1-x) = \kappa G_j^{-1}(1-\kappa x).$$

Notice that

$$\begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}x}\displaystyle\int_{1-\kappa x}^{1} G_j^{-1}(t)\mathrm{d}t = \kappa G_j^{-1}(1-\kappa x), \\ \dfrac{\mathrm{d}}{\mathrm{d}x}\displaystyle\int_{1-x}^{1} F_j^{-1}(t)\mathrm{d}t = F_j^{-1}(1-x). \end{cases}$$

Thus, we have $\int_{1-x}^{1} F_j^{-1}(t)\mathrm{d}t = \int_{1-\kappa x}^{1} G_j^{-1}(t)\mathrm{d}t + C_j, \forall j \in \mathcal{S}\cup\{0\}$, where $C_j$ is independent of $x$. Hence, (P12) can be written as

$$\max_{x_j,\forall j\in\mathcal{S}} \left\{ \sum_{j\in\mathcal{S}\cup\{0\}} \left(\kappa v_j x_j + \int_{1-\kappa x_j}^{1} G_j^{-1}(t)dt\right) : \sum_{j\in\mathcal{S}\cup\{0\}} x_j = 1, x_j \geq 0, \forall j \in \mathcal{S}\cup\{0\} \right\}.$$

Since $G_j(\cdot)$ is a valid c.d.f., $G_j^{-1}$ is only valid in domain $[0,1]$. Therefore, (P12) is equivalent to (P11). That is, for any given TUM-$\kappa$, we can always find an equivalent corresponding MDM.

Recall that we have shown that TUM-1 is equivalent to the MDM. Therefore, we conclude that TUM-$\kappa$, TUM-1 and MDM are equivalent given the c.d.f.'s of the utility noise terms in TUM are invertible.

■

### B.6. Proof of Proposition 1

**Proof.** Let $\lambda$ and $\mu_j$, for $j \in S\cup\{0\}$, be Lagrange multipliers. Then the optimality conditions for (P3) yield the following equations:

$$\begin{aligned} F_j^{-1}(1-x_j) &= \lambda - v_j - \mu_j, \forall j \in S\cup\{0\} \\ x_j &\geq 0, \forall j \in S\cup\{0\} \\ \mu_j &\geq 0, \forall j \in S\cup\{0\} \\ \mu_j x_j &= 0, \forall j \in S\cup\{0\} \\ \sum_{j\in S\cup\{0\}} x_j &= 1. \end{aligned}$$

Since we assume $v_0 = 0$ and $x_0 > 0$ due to Assumption 1, we must have $\mu_0 = 0$, and therefore $\lambda = F_0^{-1}(1-x_0)$.

If $\mu_j = 0$ for $j \in S$, we have $x_j(x_0) = 1 - F_j(\lambda - v_j)$. Otherwise, if $\mu_j > 0$, we have $x_j = 0$, $F_j(\lambda - v_j - \mu_j) = 1$. Then, $F_j(\lambda - v_j - \mu_j) = F_j(\lambda - v_j) = 1$ and $x_j(x_0) = 1 - F_j(\lambda - v_j - \mu_j) = 1 - F_j(\lambda - v_j)$.

Thus, $x_j(x_0) = 1 - F_j(F_0^{-1}(1-x_0) - v_j), \forall j \in S$, where $x_0$ is the market share of outside option that satisfies $\sum_{j\in S} x_j(x_0) + x_0 = 1$.

■

### B.7. Proof of Theorem 4.

**Proof.** We first introduce the following proposition and present its proof.

**Proposition 4** *For assortment problems under regular choice models, there exists an optimal solution which includes all of the most profitable products. Moreover, under strictly regular choice models, all the optimal solutions include all of the most profitable products.*

We start with the proof of the first claim by contradiction. Suppose none of the optimal solutions assort all of the most profitable products. Then we arbitrarily consider such an optimal solution, denoted as $\mathcal{S}$ and we arbitrarily denote a most profitable product, which is not included in $\mathcal{S}$, by $i$. We let $\mathcal{S}^* := \mathcal{S} \cup \{i\}$. We denote the overall profit corresponding to $\mathcal{S}$ and $\mathcal{S}^*$, by $\Pi(\mathcal{S})$ and $\Pi(\mathcal{S}^*)$. We denote the choice probabilities corresponding to products in $\mathcal{S}$ and $\mathcal{S}^*$ by $x_j$ and $x_j^*$, respectively. The outside market shares in $\mathcal{S}$ and $\mathcal{S}^*$ are denoted by $x_0$ and $x_0^*$, respectively. Let $x_j^\Delta := x_j - x_j^*, \forall j \in \mathcal{S}$ and $x_0^\Delta = x_0 - x_0^*$. The regularity condition implies that $x_j^\Delta \geq 0, \forall j \in \mathcal{S}$, $x_0^\Delta \geq 0$ and $x_i^* = \sum_{j\in\mathcal{S}} x_j^\Delta + x_0^\Delta \geq \sum_{j\in\mathcal{S}} x_j^\Delta$.

The overall profit corresponding to $\mathcal{S}^*$ is

$$\begin{aligned}
\Pi(\mathcal{S}^*) &= \sum_{j\in\mathcal{S}^*} p_j x_j^* \\
&= p_i x_i^* + \sum_{j\in\mathcal{S}} p_j x_j^* \\
&= p_i x_i^* + \sum_{j\in\mathcal{S}} p_j (x_j - x_j^\Delta) \\
&= p_i x_i^* - \sum_{j\in\mathcal{S}} p_j x_j^\Delta + \sum_{j\in\mathcal{S}} p_j x_j \\
&\geq \sum_{j\in S} p_j x_j \qquad (4)\\
&= \Pi(\mathcal{S}),
\end{aligned}$$

where the inequality (4) is due to $p_i \geq p_j, \forall j \in S$ and $x_i^* \geq \sum_{j\in\mathcal{S}} x_j^\Delta$. Hence, $\mathcal{S}^*$ is also optimal. By keeping adding the most profitable products, we would eventually obtain an assortment which includes all the most profitable products being optimal, which contradicts our assumption.

If the choice model is strictly regular (e.g. MEM), we have $x_i^* > \sum_{j\in\mathcal{S}} x_j^\Delta$, and thus $\Pi(\mathcal{S}^*) > \Pi(\mathcal{S})$ in the above argument. This contradicts with the optimality of $\mathcal{S}$ and proves that the second claim is also correct.

Now, we prove Theorem 4. The problem (P7) can be written as

$$\left.\begin{aligned}
\omega = \max_{x_0\in[0,1]} \quad \max_{\boldsymbol{y}\in\{0,1\}^Q} \quad & \sum_{j\in\mathcal{Q}} p_j x_j(x_0) y_j \\
\text{s.t.} \quad & \sum_{j\in\mathcal{Q}} x_j(x_0) y_j + x_0 \leq 1,
\end{aligned}\right\} \qquad \text{(P13)}$$

where $x_j = 1 - F_j(F_0^{-1}(1-x_0) - v_j), \forall j \in \mathcal{Q}$. Notice that the constraint holds as equality at optimality since the objective function is non-decreasing in $x_0$. Introducing the Lagrange multiplier, we have

$$\begin{aligned}
\omega &= \max_{x_0\in[0,1]} \max_{\boldsymbol{y}\in\{0,1\}^Q} \min_{\lambda\geq 0} \sum_{j\in\mathcal{Q}} p_j x_j(x_0) y_j + \lambda(1 - \sum_{j\in\mathcal{Q}} x_j(x_0) y_j - x_0) \\
&\leq \max_{x_0\in[0,1]} \min_{\lambda\geq 0} \max_{\boldsymbol{y}\in\{0,1\}^Q} \lambda - \lambda x_0 + \sum_{j\in\mathcal{Q}} x_j(x_0) y_j (p_j - \lambda) \\
&= \max_{x_0\in[0,1]} \min_{\lambda\geq 0} \lambda - \lambda x_0 + \sum_{j\in\mathcal{Q}} x_j(x_0) \cdot (p_j - \lambda)^+ \\
&:= \max_{x_0\in[0,1]} \min_{\lambda\geq 0} L(\lambda, x_0),
\end{aligned}$$

where the second inequality is due to the change from 'max min' to 'min max' and $(p_j - \lambda)^+ := \max\{p_j - \lambda, 0\}$. It is easy to see that $L(\lambda, x_0)$ is convex and piecewise linear in $\lambda$. The first order derivative of $L(\lambda, x_0)$ is given as

$$\frac{\partial}{\partial\lambda} L(\lambda, x_0) = 1 - x_0 - \sum_{j\in\mathcal{Q}} x_j(x_0) \mathbb{1}\{p_j \geq \lambda\}.$$

Let $i^*(x_0) = \min\left\{i | 1 - x_0 - \sum_{j=1}^{i} x_j(x_0) \le 0\right\}$ for reasonable values of $x_0$. Here by reasonable, we mean that $x_0$ takes the values where $i^*(x_0)$ exists. Notice that those $x_0$, with which $i^*(x_0)$ does not exist, cannot be optimal to Problem (P13). Hence we do not consider them. Then, given $x_0$, one optimal $\lambda$ can be computed as $\lambda(x_0) = p_{i^*(x_0)}$.

Let $\mathcal{Q}_k$ be strictly profit-nested, that is $\mathcal{Q}_k = \{S_1, S_2, ..., S_k\}$, where $S_k$ is a subset of all products which have the $k$th highest profit. Let $t_k$ be the outside market share corresponding to $\mathcal{Q}_k$, i.e.,

$$1 - t_k - \sum_{j \in \bigcup_{i=1}^{k} S_i} x_j(t_k) = 0.$$

Then, it is clear that for $x_0 = t_{k-1}$, $\lambda(x_0) = p_{k-1}$ is optimal and for all $x_0 \in [t_k, t_{k-1})$, $\lambda(x_0) = p_k$ is optimal. Next, we will show that when $x_0 = t_{k-1}$, $\lambda(x_0) = p_k$ is also optimal.

$$\begin{aligned}
L(p_k, t_{k-1}) &= p_k - p_k(t_{k-1}) + \sum_{j \in \mathcal{Q}} x_j(t_{k-1}) \cdot (p_j - p_k)^+ \\
&= p_k - p_k(t_{k-1}) + \sum_{j \in \bigcup_{i=1}^{k-1} S_i} x_j(t_{k-1}) \cdot (p_j - p_k) \\
&= p_k(1 - t_{k-1} - \sum_{j \in \bigcup_{i=1}^{k-1} S_i} x_j(t_{k-1})) + \sum_{j \in \bigcup_{i=1}^{k-1} S_i} x_j(t_{k-1}) p_j \\
&= p_{k-1}(1 - t_{k-1} - \sum_{j \in \bigcup_{i=1}^{k-1} S_i} x_j(t_{k-1})) + \sum_{j \in \bigcup_{i=1}^{k-1} S_i} x_j(t_{k-1}) p_j \qquad (5) \\
&= p_{k-1} - p_{k-1}(t_{k-1}) + \sum_{j \in \bigcup_{i=1}^{k-1} S_i} x_j(t_{k-1}) \cdot (p_j - p_{k-1}) \\
&= p_{k-1} - p_{k-1}(t_{k-1}) + \sum_{j \in \mathcal{Q}} x_j(t_{k-1}) \cdot (p_j - p_{k-1})^+ \\
&= L(p_{k-1}, t_{k-1}),
\end{aligned}$$

where equality (5) is because $(1 - t_{k-1} - \sum_{j \in \bigcup_{i=1}^{k-1} S_i} x_j(t_{k-1})) = 0$. Thus, $\lambda(x_0) = p_k$ is also optimal when $x_0 = t_{k-1}$. Therefore, for all $x_0 \in [t_k, t_{k-1}]$, $\lambda(x_0) = p_k$ is optimal.

From Proposition 4, we know that there exits an optimal $x_0$ which cannot exceed $t_1$. Then, we have

$$\max_{x_0 \in [0,1]} \min_{\lambda \ge 0} L(\lambda, x_0) = \max_{k=2,\dots,l} \max_{x_0 \in [t_k, t_{k-1}]} p_k - p_k x_0 + \sum_{j \in \mathcal{Q}} x_j(x_0) \cdot (p_j - p_k)^+,$$

where $l$ is the number of the profit levels. Assumption 2 implies that the objective function is convex in $x_0$, so the optimal solution of the inner maximization appears on the boundary, namely, given a $k$, the optimal $x_0$ is either $t_k$ or $t_{k-1}$. First, if $x_0 = t_k$, then

$$\begin{aligned}
\max_{x_0 \in [0,1]} \min_{\lambda \ge 0} L(\lambda, x_0) &= \max_{k=2,\dots,l} \Big(p_k - p_k t_k + \sum_{j \in \bigcup_{i=1}^{k} S_i} x_j(t_k) \cdot (p_j - p_k)\Big) \\
&= \max_{k=2,\dots,l} \Big(p_k(1 - t_k - \sum_{j \in \bigcup_{i=1}^{k} S_i} x_j(t_k)) + \sum_{j \in \bigcup_{i=1}^{k} S_i} x_j(t_k) p_j\Big) \\
&= \max_{k=2,\dots,l} \Big(\sum_{j \in \bigcup_{i=1}^{k} S_i} x_j(t_k) p_j\Big)
\end{aligned}$$

Second, if $x_0 = t_{k-1}$, then

$$\begin{aligned}\max_{x_0\in[0,1]}\min_{\lambda\geq 0} L(\lambda,x_0) &= \max_{k=2,\dots,l}\Big(p_k - p_k t_{k-1} + \sum_{j\in\bigcup_{i=1}^{k-1}S_i} x_j(t_{k-1})\cdot(p_j-p_k)\Big)\\ &= \max_{k=2,\dots,l}\Big(p_k(1-t_{k-1} - \sum_{j\in\bigcup_{i=1}^{k-1}S_i} x_j(t_{k-1})) + \sum_{j\in\bigcup_{i=1}^{k-1}S_i} x_j(t_{k-1})p_j\Big)\\ &= \max_{k=2,\dots,l}\Big(\sum_{j\in\bigcup_{i=1}^{k-1}S_i} x_j(t_{k-1})p_j\Big)\end{aligned}$$

Thus, $\max_{x_0\in[0,1]}\min_{\lambda\geq 0} L(\lambda,x_0)$ leads to a profit corresponding to a strictly profit-nested assortment. Recall that $\omega \leq \max_{x_0\in[0,1]}\min_{\lambda\geq 0} L(\lambda,x_0)$, hence, there exists a strictly profit-nested assortment that is optimal.

■

### B.8. Proof of Proposition 2.

**Proof.** Recall that showing $x_j(x_0)$ is convex for all $j\in\mathcal{Q}$ is sufficient to show that Assumption 2 holds. Recall also that in Assumption 2 we only consider reasonable $x_0$ that lies in $[0,1]$ such that $x_j(x_0)\in[0,1)$ because other values of $x_0$ are never optimal and thus are irrelevant. As a result, for each $x_j(x_0)$, there are two possibilities.

First, $x_j(x_0)=0$ for all $x_0\in[0,1]$. This trivial case may occur when $T_j$ is too small or $v_j$ is too small. Clearly, $x_j(x_0)$ is convex in this case.

Second, for any $j$ there exist a value $x_0^{\#}\in[0,1]$ such that $x_j(x_0)=0$ for all $x_0\in[0,x_0^{\#}]$ and $x_j(x_0)>0$ for all $x_0>x_0^{\#}$. We will first prove the convexity for reasonable $x_0$ that is greater than $x_0^{\#}$. Let $D_j = \frac{1}{1-\left(\frac{m_j}{T_j}\right)^{\beta_j}}$. Then, the c.d.f. of the truncated Pareto distribution with scale parameter $m_j>0$, shape parameter $\beta_j>0$, truncation point $T_j>m_j$ is

$$F_j(x) = D_j\left(1-\left(\frac{m_j}{x}\right)^{\beta_j}\right),\ \text{for } x\in[m_j,T_j],$$

where $D_j$ is clearly greater than 1. We calculate $x_j(x_0)$,

$$\begin{aligned}x_j(x_0) &= 1-F_j(F_0^{-1}(1-x_0)-v_j)\\ &= 1-D_j\left(1-\left(\frac{m_j}{\frac{m_0}{x_0^{1/\beta_0}}-v_j}\right)^{\beta_j}\right)\\ &= (1-D_j)+D_j\left(\frac{\frac{m_0}{x_0^{1/\beta_0}}-v_j}{m_j}\right)^{-\beta_j}\\ &= (1-D_j)+D_j\left(\frac{m_0}{m_j}x_0^{-1/\beta_0}-\frac{v_j}{m_j}\right)^{-\beta_j} > 0. \end{aligned} \tag{6}$$

Then, we calculate its second derivative,

$$\frac{\mathrm{d}^2}{\mathrm{d}x_0^2}x_j(x_0) = D_j\frac{\beta_j m_0\big(m_0(\beta_j-\beta_0)+(\beta_0+1)v_j x_0^{1/\beta_0}\big)\cdot\left(\frac{m_0}{m_j}x_0^{-1/\beta_0}-\frac{v_j}{m_j}\right)}{\beta_0^2 m_0^2(m_0-v_j x_0^{1/\beta_0})^2},$$

where it can be easily verified from (6) that the denominator and the second part of the numerator are both positive. Since $x_0\in[0,1]$, the proposed condition implies that $\beta_j\geq\beta_0-\frac{v_j}{m_0}(\beta_0+1)x_0^{1/\beta_0}$. Rearranging terms, we obtain

$$m_0(\beta_j-\beta_0)+(\beta_0+1)v_j x_0^{1/\beta_0}\geq 0,$$

which implies that $\frac{\mathrm{d}^2}{\mathrm{d}x_0^2}x_j(x_0)\geq 0$. It is easy to see that $x_j(x_0)$ is continuous for reasonable $x_0\in[0,1]$. What is more, since $x_j(x_0)$ is increasing and convex for reasonable $x_0$ that is greater than $x_0^{\#}$ and $x_j(x_0)=0$ for all $x_0\in[0,x_0^{\#}]$, it can be easily verified that $x_j(x_0)$ is convex for all reasonable $x_0$ that lies in $[0,1]$.

■

### B.9. Proof of Corollary 1.

**Proof.** Recall from Theorem 1 that to obtain the choice probability of the MDC model from the MDM, we use Pareto marginals and set deterministic utilities to zero. From Proposition 2, when $v_j=0,\forall j\in\mathcal{Q}$, the condition is satisfied when $\beta_0$ is smallest among $\boldsymbol{\beta}$, which is equivalent to the condition that $\sigma_0$ is the smallest among $\boldsymbol{\sigma}$ in the MDC model (this property is discussed in Zhang et al. (2021) as well).

■

### B.10. Proof of Theorem 5.

**Proof.** Recall that the assortment problems under the MDM can be formulated in the following way

$$Z_1=\left.\begin{array}{lll}\max\limits_{x_0\in(0,1]} & \max\limits_{\boldsymbol{y}\in\{0,1\}^Q} & \sum\limits_{j\in\mathcal{Q}}p_jx_j(x_0)y_j\\ & \text{s.t.} & \sum\limits_{j\in\mathcal{Q}}x_j(x_0)y_j+x_0=1,\end{array}\right\}\tag{P14}$$

where $x_j(x_0)$ calculates the choice probability of product $j$ and is non-decreasing of $x_0$.

Now, we consider a relaxation of (P14), where $y_j,\forall j\in\mathcal{Q}$ may take any value in $[0,1]$:

$$Z_2=\left.\begin{array}{lll}\max\limits_{x_0\in(0,1]} & \max\limits_{\boldsymbol{y}\in[0,1]^Q} & \sum\limits_{j\in\mathcal{Q}}p_jx_j(x_0)y_j\\ & \text{s.t.} & \sum\limits_{j\in\mathcal{Q}}x_j(x_0)y_j+x_0=1.\end{array}\right\}\tag{P15}$$

Then, we have $Z_1\leq Z_2$. Let $(x_0^*,\boldsymbol{y}^*)$ be the optimal solution to (P15), it can be easily shown that $\boldsymbol{y}^*=\{1,1,...,1,a,...,a,0,0,...,0\}$, for some $a\in[0,1]$, where all $y_i$ that receives value $a$ are corresponding to products in the same profit level (with identical unit profit). Suppose there are $l$ profit levels. W.l.o.g., we let $\boldsymbol{y}_k^*=[y_i|\text{for all } i \text{ in profit level } k]=\boldsymbol{a}=[a,...,a]$. Hence, $y_i=1$ and $y_j=0$ for all $p_i>p_k,p_j<p_k$. We can always view $\boldsymbol{y}^*$ as a combination of three parts: $\boldsymbol{y}_1^*$, $\boldsymbol{y}_2^*$, $\boldsymbol{y}_3^*$, where the first part $\boldsymbol{y}_1^*=\{y_i^*,p_i>p_k\}$, the second part $\boldsymbol{y}_2^*=\boldsymbol{y}_k^*$ and the third part $\boldsymbol{y}_3^*=\{y_j^*,p_j<p_k\}$. Notice that $k$ may be 1 or $l$, hence, it is possible that part 1 or part 3 is an empty set. We can write $Z_2$ as

$$Z_2=\sum_{j\in\mathcal{Q}}p_jx_j(x_0^*)y_j^*=\sum_{p_i>p_k}p_ix_i(x_0^*)+\sum_{j\in\text{profit level }k}p_jx_j(x_0^*)a.$$

Now, we consider two other feasible solutions to (P15): $(\overline{x_0},\overline{\boldsymbol{y}})$ and $(\underline{x_0},\underline{\boldsymbol{y}})$, where $\overline{\boldsymbol{y}}=[\boldsymbol{y}_1^*,\boldsymbol{0},\boldsymbol{y}_3^*]$, and $\underline{\boldsymbol{y}}=[\boldsymbol{y}_1^*,\boldsymbol{1},\boldsymbol{y}_3^*]$. It is easy to see that both $(\overline{x_0},\overline{\boldsymbol{y}})$ and $(\underline{x_0},\underline{\boldsymbol{y}})$ are also feasible to (P14) and both $\overline{\boldsymbol{y}}$, $\underline{\boldsymbol{y}}$ are strictly profit-nested assortments. Since $x_j(x_0)$ is non-decreasing, we have

$$0<\underline{x_0}\leq x_0^*\leq\overline{x_0}\leq 1.$$

Let $\Pi(x_0, \boldsymbol{y})$ be the objective function value of (P14) given $x_0$ and $\boldsymbol{y}$ if $(x_0, \boldsymbol{y})$ is feasible to (P14). That is, $\Pi(x_0, \boldsymbol{y})$ calculates the profit corresponding to assortment $\boldsymbol{y}$. First, we have

$$\Pi(\overline{x_0}, \overline{\boldsymbol{y}}) = \sum_{p_i > p_k} p_i x_i(\overline{x_0}) \geq \sum_{p_i > p_k} p_i x_i(x_0^*), \tag{7}$$

where the last inequality is because $x_j(x_0)$ is non-decreasing. Second,

$$\Pi(\underline{x_0}, \underline{\boldsymbol{y}}) = \sum_{p_i > p_k} p_i x_i(\underline{x_0}) + \sum_{j \in \text{profit level } k} p_j x_j(\underline{x_0}).$$

Since $x_0^* \geq \underline{x_0}$ and $x_j(x_0)$ is non-decreasing, we have

$$\begin{aligned}
\sum_{p_i > p_k} x_i(\underline{x_0}) + \underline{x_0} &\leq \sum_{p_i > p_k} x_i(x_0^*) + x_0^* \\
1 - \sum_{p_i > p_k} x_i(\underline{x_0}) - \underline{x_0} &\geq 1 - \sum_{p_i > p_k} x_i(x_0^*) - x_0^* \\
\sum_{j \in \text{profit level } k} x_j(\underline{x_0}) &\geq \sum_{j \in \text{profit level } k} x_j(x_0^*) a \\
\sum_{j \in \text{profit level } k} p_j x_j(\underline{x_0}) &\geq \sum_{j \in \text{profit level } k} p_j x_j(x_0^*) a.
\end{aligned}$$

Since $\sum_{p_i > p_k} p_i x_i(\underline{x_0}) \geq 0$, we have

$$\Pi(\underline{x_0}, \underline{\boldsymbol{y}}) \geq \sum_{j \in \text{profit level } k} p_j x_j(\underline{x_0}) \geq \sum_{j \in \text{profit level } k} p_j x_j(x_0^*) a. \tag{8}$$

Based on inequalities (7) and (8), we have

$$\Pi(\overline{x_0}, \overline{\boldsymbol{y}}) + \Pi(\underline{x_0}, \underline{\boldsymbol{y}}) \geq \sum_{p_i > p_k} p_i x_i(x_0^*) + \sum_{j \in \text{profit level } k} p_j x_j(x_0^*) a = Z_2.$$

Since both $\Pi(\overline{x_0}, \overline{\boldsymbol{y}})$ and $\Pi(\underline{x_0}, \underline{\boldsymbol{y}})$ are non-negative,

$$\max\left\{\Pi(\overline{x_0}, \overline{\boldsymbol{y}}), \Pi(\underline{x_0}, \underline{\boldsymbol{y}})\right\} \geq \frac{1}{2} Z_2.$$

Hence,

$$\max\left\{\Pi(\overline{x_0}, \overline{\boldsymbol{y}}), \Pi(\underline{x_0}, \underline{\boldsymbol{y}})\right\} \geq \frac{1}{2} Z_1.$$

Recall that $Z_1$ is the profit of the optimal solution to (P14); $\Pi(\overline{x_0}, \overline{\boldsymbol{y}})$, and $\Pi(\underline{x_0}, \underline{\boldsymbol{y}})$ are the profits corresponding to assortments $\overline{\boldsymbol{y}}$ and $\underline{\boldsymbol{y}}$, respectively, and $\overline{\boldsymbol{y}}$ and $\underline{\boldsymbol{y}}$ are both strictly profit-nested assortments. ■

### B.11. Proof of Proposition 3.

**Proof.** We use $\Pi(S)$ to denote the expected profit of the assortment $S$. We first calculate the upper bounds for expected profits of all profit-nested assortments. For assortment $\{1\}$, since $c_1 = w_1 p_1 = \epsilon$ and $z \in (0,1)$, it is easy to see that $\Pi(\{1\}) < \epsilon$. For assortment $\{1,2,3\}$, first we calculate the upper bound for the optimal $z_{123}^*$. Notice that given assortment $\{1,2,3\}$, the constraint in Problem (P9) becomes $w_1 z_{123}^* + w_2 z_{123}^* + w_3 z_{123}^* +$

$(z_{123}^*)^\tau = 1$. Since $w_1 z_{123}^*$, $w_3 z_{123}^*$ and $(z_{123}^*)^\tau$ are positive, we have $w_2 z_{123}^* < 1$. Hence, we obtain an upper bound of the optimal $z_{123}^*$: $z_{123}^* < \frac{1}{w_2} = \frac{1}{10+\frac{2}{\epsilon}}$. Therefore,

$$\begin{aligned}
\Pi(\{1,2,3\}) &= (c_1 + c_2 + c_3) z_{123}^* \\
&< \Big(\epsilon + \epsilon(1-2\epsilon)(10+\frac{2}{\epsilon}) + \epsilon(1-2\epsilon)(1-\epsilon)\Big)\Big(\frac{1}{10+\frac{2}{\epsilon}}\Big) \\
&= \epsilon(1-2\epsilon) + \Big(\epsilon + \epsilon(1-2\epsilon)(1-\epsilon)\Big)\Big(\frac{1}{10+\frac{2}{\epsilon}}\Big) \\
&< \epsilon(1-2\epsilon) + \Big(\epsilon + \epsilon(1-2\epsilon)(1-\epsilon)\Big)\Big(\frac{1}{2/\epsilon}\Big) \\
&= \epsilon(1-2\epsilon) + \frac{\epsilon^2}{2}\Big(1 + (1-2\epsilon)(1-\epsilon)\Big) \\
&< \epsilon(1-2\epsilon) + \epsilon^2 \\
&= \epsilon(1-\epsilon),
\end{aligned}$$

where in the first inequality we substitute $c_i = w_i p_i, \forall i \in \{1,2,3\}$ and $z_{123}^* < \frac{1}{10+\frac{2}{\epsilon}}$. In the last inequality, we use the fact that $1 + (1-2\epsilon)(1-\epsilon) < 2$. Similarly, for assortment $\{1,2\}$, we can also find $z_{12}^* < \frac{1}{10+\frac{2}{\epsilon}}$. Then,

$$\begin{aligned}
\Pi(\{1,2\}) &= (c_1 + c_2) z_{12}^* \\
&< (c_1 + c_2 + c_3)(\frac{1}{10+\frac{2}{\epsilon}}) \\
&< \epsilon(1-\epsilon).
\end{aligned}$$

Then, we have $\max\Big\{\Pi(\{1\}), \Pi(\{1,2\}), \Pi(\{1,2,3\})\Big\} < \epsilon$.

Next, we calculate the lower bound of the expect profit of assortment $\{1,3\}$. First, we show that $z_{13}^* = 1-\epsilon$. Notice that given assortment $\{1,3\}$, the constraint in Problem (P9) becomes $w_1 z_{13}^* + w_3 z_{13}^* + (z_{13}^*)^\tau = 1$. Substitute $z_{13}^* = 1-\epsilon$ into the equation, we obtain

$$\begin{aligned}
\text{LHS} &= \epsilon(1-\epsilon) + (1-\epsilon)(1-\epsilon) + (1-\epsilon)^\tau \\
&= (1-\epsilon) + \epsilon = 1 = \text{RHS},
\end{aligned}$$

where the second equality is due to $(1-\epsilon)^{\frac{\log(\epsilon)}{\log(1-\epsilon)}} = \epsilon$. Then, $\Pi(\{1,3\}) = (c_1 + c_3) z_{13}^* = (1-\epsilon)\Big(\epsilon + \epsilon(1-2\epsilon)(1-\epsilon)\Big)$.

Let $S^*$ denote the optimal assortment. Now, we calculate the upper bound of the ratio between the expected profits of the best profit-nested assortment and the optimal solution:

$$\begin{aligned}
\frac{\max\Big\{\Pi(\{1\}), \Pi(\{1,2\}), \Pi(\{1,2,3\})\Big\}}{\Pi(S^*)} &\leq \frac{\max\Big\{\Pi(\{1\}), \Pi(\{1,2\}), \Pi(\{1,2,3\})\Big\}}{\Pi(\{1,3\})} \\
&< \frac{\epsilon}{(1-\epsilon)\Big(\epsilon + \epsilon(1-2\epsilon)(1-\epsilon)\Big)} \\
&= \frac{1}{(1-\epsilon)\Big(1 + (1-2\epsilon)(1-\epsilon)\Big)},
\end{aligned}$$

where the first inequality is due to $\Pi(S^*) \geq \Pi(\{1,3\})$. Notice that the last function goes to 1/2 when $\epsilon$ approaches to 0. Based on Theorem 5, we also know the ratio is at least 1/2. Hence, the ratio goes to 1/2 when $\epsilon$ approaches to 0.

■

### B.12. Proposition 5 and its Proof.

**Proposition 5** *Under the MDM, the choice probabilities of alternatives in the current assortment strictly decrease if one more alternative is added.*

**Proof.** From the definition in Section 2.3, we know that in MDM,

$$x_j = 1 - F_j(\lambda^* - v_j),$$

it is clear that $x_j > 0$ unless $\lambda^* = \infty$ or $v_j = -\infty$, which is excluded from the model.

Now we consider an empty assortment, in which the outside market option captures the entire market share, which is 1. When an arbitrary product is added to this assortment, the outside market share is less than 1 since all assorted products have strictly positive choice probabilities.

Then, we consider an non-empty assortment (not an full assortment), denoted by $\mathcal{S}$. The choice probability of an arbitrary product $j$ in $\mathcal{S}$ is

$$x_{j,\mathcal{S}} = 1 - F_j(\lambda - v_j),$$

where $v_j$ is the deterministic utility of product $j$ and

$$\sum_{i \in \mathcal{S} \cup \{0\}} (1 - F_i(\lambda - v_i)) = 1.$$

Now, we consider another assortment $\mathcal{S}^+ := \mathcal{S} \cup \{k\}$, where $k$ is an arbitrary product that is not in $\mathcal{S}$. Under this assortment, the choice probability of product $j$ is

$$x_{j,\mathcal{S}^+} = 1 - F_j(\lambda^+ - v_j),$$

where $v_j$ is the deterministic utility of product $j$ and

$$\sum_{i \in \mathcal{S} \cup \{0\}} (1 - F_i(\lambda^+ - v_i)) + (1 - F_k(\lambda^+ - v_k)) = 1.$$

Since by definition, $F_i$ is a strictly increasing c.d.f., we have $\lambda < \lambda^+$ and hence, $x_{j,\mathcal{S}} > x_{j,\mathcal{S}^+}$.

■

## C. Technical Proofs for Approximation Algorithms for Assortment Problems under the MDM

### C.1. Knapsack Problem with Integer Parameters: Dynamic Programming

Suppose there are $Q$ items, where item $j$ has integer profit $r_j$ and (possibly non-integer) weight $w_j$. The total capacity of the knapsack is denoted by $C > 0$. The corresponding knapsack problem, denoted KP$(\boldsymbol{r}, \boldsymbol{w}, C)$, is formulated as

$$\begin{aligned} \max_{I_j} \quad & \sum_{j \in \mathcal{Q}} r_j I_j \\ \text{s.t.} \quad & \sum_{j \in \mathcal{Q}} w_j I_j \le C \\ & I_j \in \{0,1\}, \forall j \in \mathcal{Q}, \end{aligned}$$

where $\boldsymbol{r}=(r_1,r_2,...,r_Q)$ and $\boldsymbol{w}=(w_1,w_2,...,w_Q)$. Define $R_{\max}=\max_{j\in\mathcal{Q}} r_j$ to be the profit of the most profitable item. Thus, we know that the overall profit is upper bounded by $QR_{\max}$. Define $S_{i,r}\subseteq\{1,2,...,i\}$ to be a set whose total profit is exactly $r$ and has the least total weight among such subsets. Let $W_{i,r}$ be the total weight of $S_{i,r}$ and $W_{i,r}=\infty$ if there is no such $S_{i,r}$. Then, we can use the initial conditions:

$$W_{1,r}=\begin{cases} w_1, \text{ if } r=r_1,\\ \infty, \text{ otherwise,}\end{cases}$$

and the following recurrence to obtain a Dynamic Program (DP) for the knapsack problem:

$$W_{i+1,r}=\begin{cases} \min\{W_{i,r},W_{i,r-r_{i+1}}+w_{i+1}\}, \text{ if } r_{i+1}\le r,\\ W_{i,r}, \text{ otherwise.}\end{cases}$$

The optimal solution is a set $S_{n,r}$ with maximum $r$ such that $W_{n,r}\le C$. The total run time of the DP is $\mathcal{O}(Q^2R_{\max})$, so it is a pseudo-polynomial time algorithm. (See Andonov et al. 2000 for further details.)

### C.2. FPTAS For Knapsack Problems

Recall that in DP, we assume all items are associated with integer profits. Here, we relax this assumption by allowing non-integer unit profits. Let $r_j$ denote the profit of item $j$ and let $r_{\max}=\max_{j\in\mathcal{Q}} r_j$. The FPTAS is as follows (see Lai and Goemans 2006 for details):

**Algorithm 3** Knapsack FPTAS

**Input** $\epsilon$, $Q$, $r_{\max}$

1: $K\leftarrow\frac{\epsilon\cdot r_{\max}}{Q}$
2: For each item $j$, define $r_j'=\left\lfloor\frac{r_j}{K}\right\rfloor$.
3: Using the DP (Section C.1), solve a new knapsack problem with new unit profit $r_j'$ for each item to obtain an optimal assortment $S'$.
4: **return** $S'$

The algorithm has a run time $O(Q^2\left\lfloor\frac{Q}{\epsilon}\right\rfloor)$ and produces an $(1-\epsilon)$-approximate solution for the original Knapsack problem.

### C.3. Proof of Theorem 6.

Recall that $f_{kp}(S,x_0)=\sum_{j\in S}p_j\cdot x_j(x_0)$, where $p_j$ is the unit profit of product $j$, $S$ is an assortment and $x_0$ is the outside market share. We use Profit$(S)$ to represent the profit of assortment $S$. We denote the optimal solution by $S^*$ with the corresponding outside market share $x_0^*$. We denote the solution of Algorithm 1 by $S^A$ with the corresponding outside market share $x_0^A$. Notice that Algorithm 1 searches $x_0$ from $\underline{x_0}$ to $\overline{x_0}$. Let $x_0^\gamma$ denote a value searched by the algorithm which is no-greater than and closest to $x_0^*$ and $S^\gamma$ denote the corresponding approximate optimal assortment obtained from FPTAS for knapsack. Let $S^{\gamma*}$ be the optimal solution to the inner knapsack problem given $x_0^\gamma$. Then, there are two cases. In case 1: $x_0^\gamma\in(x_0^*-\Delta,x_0^*]$. In case 2: $x_0^\gamma=\underline{x_0}$.

Since Algorithm 1 eventually selects the one with greatest objective function value of the inner problem in Problem (P10) among searched values of $x_0$s,

$$f_{kp}(S^A, x_0^A) \geq f_{kp}(S^\gamma, x_0^\gamma). \tag{9}$$

Since $S^\gamma$ is the solution of FPTAS of knapsack for inner problem in Problem (P10) given $x_0^\gamma$,

$$f_{kp}(S^\gamma, x_0^\gamma) \geq f_{kp}(S^{\gamma *}, x_0^\gamma) \cdot (1-\epsilon). \tag{10}$$

Since $S^{\gamma *}$ is the optimal solution to the inner problem given $x_0^\gamma$,

$$f_{kp}(S^{\gamma *}, x_0^\gamma) \geq f_{kp}(S^*, x_0^\gamma). \tag{11}$$

We first discuss case 1 where $x_0^\gamma \in (x_0^* - \Delta, x_0^*]$. Notice that $f_{kp}(S, x_0)$ is a non-decreasing function of $x_0$, given $S$. Hence, we have the following inequality

$$f_{kp}(S^*, x_0^\gamma) \geq f_{kp}(S^*, x_0^* - \Delta). \tag{12}$$

Combine inequalities (9), (10), (11) and (12), we have

$$f_{kp}(S^A, x_0^A) \geq f_{kp}(S^*, x_0^* - \Delta) \cdot (1-\epsilon).$$

Notice that $\text{Profit}(S^A) \geq f_{kp}(S^A, x_0^A)$, and the equality occurs only if $\sum_{j\in S} x_j(x_0^A) = 1 - x_0^A$. Therefore, we have

$$\begin{aligned}
\text{Profit}(S^A) &\geq f_{kp}(S^A, x_0^A) \\
&\geq f_{kp}(S^*, x_0^* - \Delta) \cdot (1-\epsilon) \\
&= f_{kp}(S^*, x_0^* - \Delta) - \epsilon f_{kp}(S^*, x_0^* - \Delta) \\
&\geq f_{kp}(S^*, x_0^* - \Delta) - \epsilon \text{Profit}(S^*),
\end{aligned}$$

where the last inequality is because $\text{Profit}(S^*) = f_{kp}(S^*, x_0^*) \geq f_{kp}(S^*, x_0^* - \Delta)$. Hence,

$$\begin{aligned}
\frac{\text{Profit}(S^A)}{\text{Profit}(S^*)} &\geq \frac{f_{kp}(S^*, x_0^* - \Delta)}{\text{Profit}(S^*)} - \epsilon \\
&= \frac{f_{kp}(S^*, x_0^* - \Delta)}{f_{kp}(S^*, x_0^*)} - \epsilon \\
&= \frac{\sum_{j\in S^*} p_j x_j(x_0^* - \Delta)}{\sum_{j\in S^*} p_j x_j(x_0^*)} - \epsilon \\
&\geq \min_{j\in\mathcal{Q}} \left\{ \frac{x_j(x_0^* - \Delta)}{x_j(x_0^*)} \right\} - \epsilon.
\end{aligned}$$

Since $\forall j \in \mathcal{Q}$, we have

$$\frac{x_j(x_0^* - \Delta)}{x_j(x_0^*)} = 1 - \Big( \frac{x_j(x_0^*) - x_j(x_0^* - \Delta)}{x_j(x_0^*) \cdot \Delta} \Big) \cdot \Delta \geq 1 - M \cdot \Delta,$$

then we obtain

$$\frac{\text{Profit}(S^A)}{\text{Profit}(S^*)} \geq 1 - M \cdot \Delta - \epsilon.$$

Next, for case 2 where $x_0^\gamma = \underline{x_0}$, that is, $x_0^* < \underline{x_0} + \Delta$, following similar arguments, we obtain:

$$\frac{\text{Profit}(S^A)}{\text{Profit}(S^*)} \geq \min_{j\in\mathcal{Q}} \left\{ \frac{x_j(\underline{x_0})}{x_j(x_0^*)} \right\} - \epsilon.$$

Note that

$$\frac{x_j(x_0^*) - x_j(\underline{x_0})}{x_j(x_0^*)\cdot\Delta} \leq \frac{x_j(\underline{x_0}+\Delta) - x_j(\underline{x_0})}{x_j(\underline{x_0}+\Delta)\cdot\Delta} \leq M.$$

Hence,

$$\frac{\text{Profit}(S^A)}{\text{Profit}(S^*)} \geq \min_{j\in\mathcal{Q}} \left\{ 1 - \frac{x_j(x_0^*) - x_j(\underline{x_0})}{x_j(x_0^*)\cdot\Delta} \cdot \Delta \right\} - \epsilon \geq 1 - M\cdot\Delta - \epsilon.$$

Therefore, in both cases,

$$\frac{\text{Profit}(S^A)}{\text{Profit}(S^*)} \geq 1 - M\cdot\Delta - \epsilon.$$

Recall that in the algorithm, we let $\Delta = \frac{\eta}{2\cdot M}$ and $\epsilon = \frac{\eta}{2}$. As a result,

$$\frac{\text{Profit}(S^A)}{\text{Profit}(S^*)} \geq 1 - M\cdot\Delta - \epsilon > 1 - M\cdot\frac{\eta}{2\cdot M} - \frac{\eta}{2} = 1-\eta.$$

Notice that the FPTAS in Algorithm 1 is called at most $\left\lfloor \frac{\overline{x_0} - \underline{x_0}}{\Delta} \right\rfloor + 1 \leq \left\lfloor \frac{1}{\Delta} \right\rfloor + 1$ times. Hence, the total run time of Algorithm 1 is $O\left( \left( \left\lfloor \frac{2M}{\eta} \right\rfloor + 1 \right) \cdot \left( Q^2 \left\lfloor \frac{Q}{\eta} \right\rfloor + QC \right) \right)$, which is polynomial in both $1/\eta$ and $M$.

■

## C.4. Proof of Theorem 7.

Let $x_j'(x_0)$ be the first derivative of $x_j(x_0)$ with respect to $x_0$. Then we have

$$\begin{aligned} M &= \max_{\underline{x_0}\leq y_0 < x_0 \leq \overline{x_0}, j\in\mathcal{Q}} \left( \frac{x_j(x_0) - x_j(y_0)}{x_j(x_0)(x_0 - y_0)} \right) \\ &\geq \left( \frac{x_j(x_0) - x_j(y_0)}{x_j(x_0)(x_0 - y_0)} \right) \Bigg| y_0 \to \underline{x_0}, x_0 \to y_0 + \delta, \delta \to 0 \\ &= \lim_{\delta\to 0} \left( \frac{x_j(\underline{x_0}+\delta) - x_j(\underline{x_0})}{x_j(\underline{x_0}+\delta)(\delta)} \right) \\ &= \lim_{\delta\to 0} \frac{x_j'(\underline{x_0})}{x_j(\underline{x_0}+\delta)} \\ &= \frac{x_j'(\underline{x_0})}{x_j(\underline{x_0})}, \end{aligned}$$

for any $j \in \mathcal{Q}$.

Since MEM is a speical case of the MDC model (Theorem 2), we will only need to prove the result for the MDC model. In this case, we have $x_j(x_0) = \left(\frac{\psi_j\sigma_j}{\psi_0\sigma_0}\right)^{\frac{1}{1-\sigma_j}} (\kappa)^{\frac{1-\sigma_0}{1-\sigma_j}-1} (x_0)^{\frac{1-\sigma_0}{1-\sigma_j}}$. Note that there are the following two possible scenarios: in the first scenario, all $x_j$s remain zeros when $x_0 = \underline{x_0} = 1$. That means in all feasible assortments, no products will be chosen for sure. As a consequence, the profits of all feasible solutions are the same and equal to zero. We will not focus on this trivial case. In the second scenario, there exists at least one product, say $k$, such that $x_k(\underline{x_0}) > 0$ for some $\underline{x_0} < 1$. Then, we have

$$M \geq \frac{x_k'(\underline{x_0})}{x_k(\underline{x_0})}$$

$$
\begin{aligned}
&= \frac{\frac{1-\sigma_0}{1-\sigma_k}\left(\frac{\psi_k\sigma_k}{\psi_0\sigma_0}\right)^{\frac{1}{1-\sigma_k}}(\kappa)^{\frac{1-\sigma_0}{1-\sigma_j}-1}(\underline{x_0})^{(\frac{1-\sigma_0}{1-\sigma_k}-1)}}{\left(\frac{\psi_k\sigma_k}{\psi_0\sigma_0}\right)^{\frac{1}{1-\sigma_k}}(\kappa)^{\frac{1-\sigma_0}{1-\sigma_j}-1}(\underline{x_0})^{\frac{1-\sigma_0}{1-\sigma_k}}} \\
&= \frac{1-\sigma_0}{1-\sigma_k}(\underline{x_0})^{-1} \\
&= \frac{1-\sigma_0}{1-\sigma_{\min}}(\underline{x_0})^{-1},
\end{aligned}
$$

where $\sigma_{\min} = \min_{j\in\mathcal{Q}} \sigma_j$. Hence,

$$
\frac{1}{\underline{x_0}} \le M \cdot \frac{1-\sigma_{\min}}{1-\sigma_0}.
$$

■

### C.5. Proof of Theorem 8.

Before we present the main proof in Section C.5.1, we first provide the following analysis.

Let us say Algorithm 2 calls the FPTAS $m$ times. It is clear that $m \le \frac{\log(\frac{\overline{x_0}}{\underline{x_0}})}{\log(1+\psi)}$ and the total run time of Algorithm 2 is $\mathcal{O}\left(\frac{\log(\frac{1}{\underline{x_0}})}{\log(1+\psi)}\left(Q^2\lfloor\frac{Q}{\epsilon}\rfloor + QC\right)\right)$, since $\overline{x_0} \le 1$.

**C.5.1. Main proof.** The first part of the proof is similar to the proof of Theorem 6 in C.3. Let $f_{kp}(S,x_0) = \sum_{j\in S} p_j \cdot x_j(x_0)$, where $p_j$ is the unit profit of product $j$, $S$ is an assortment and $x_0$ is the outside market share. We use Profit$(S)$ to represent the profit of assortment $S$. We denote the optimal solution by $S^*$ with the corresponding outside market share $x_0^*$. We denote the solution of Algorithm 2 by $S^A$ with the corresponding outside market share $x_0^A$. Notice that Algorithm 2 searches $x_0$ from $\underline{x_0}$ to $\overline{x_0}$. Let $x_0^\gamma$ denote a value searched by the algorithm which is no-greater than and closest to $x_0^*$ and $S^\gamma$ denote the corresponding approximate optimal assortment obtained from FPTAS for knapsack. Let $S^{\gamma*}$ be the optimal solution to the inner knapsack problem given $x_0^\gamma$ Then, there are two cases. In case 1: $x_0^\gamma \in (x_0^* - \Delta, x_0^*]$, where $\Delta = \underline{x_0}(1+\psi)^l\psi$, where $\underline{x_0}(1+\psi)^l < x_0^* \le \underline{x_0}(1+\psi)^{l+1}$ for some $l$. In case 2: $x_0^\gamma = \underline{x_0}$.

Since Algorithm 2 eventually selects the one with greatest objective function value of the inner problem in Problem (P10) among searched values of $x_0$s,

$$
f_{kp}(S^A, x_0^A) \ge f_{kp}(S^\gamma, x_0^\gamma). \tag{13}
$$

Since $S^\gamma$ is the solution of FPTAS of knapsack for inner problem in Problem (P10) given $x_0^\gamma$,

$$
f_{kp}(S^\gamma, x_0^\gamma) \ge f_{kp}(S^{\gamma*}, x_0^\gamma)\cdot(1-\epsilon). \tag{14}
$$

Since $S^{\gamma*}$ is the optimal solution to the inner problem given $x_0^\gamma$,

$$
f_{kp}(S^{\gamma*}, x_0^\gamma) \ge f_{kp}(S^*, x_0^\gamma). \tag{15}
$$

We first discuss case 1 where $x_0^\gamma \in (x_0^* - \Delta, x_0^*]$. Notice that $f_{kp}(S,x_0)$ is a non-decreasing function of $x_0$, given $S$. Hence, we have the following inequality

$$
f_{kp}(S^*, x_0^\gamma) \ge f_{kp}(S^*, x_0^* - \Delta). \tag{16}
$$

Combine inequalities (13), (14), (15) and (16), we have

$$
f_{kp}(S^A, x_0^A) \ge f_{kp}(S^*, x_0^* - \Delta)\cdot(1-\epsilon).
$$

Notice that $\text{Profit}(S^A) \geq f_{kp}(S^A, x_0^A)$, and the equality occurs only if $\sum_{j\in S} x_j(x_0^A) = 1 - x_0^A$. Therefore, we have

$$\begin{aligned}\text{Profit}(S^A) &\geq f_{kp}(S^A, x_0^A)\\ &\geq f_{kp}(S^*, x_0^* - \Delta)\cdot(1-\epsilon)\\ &= f_{kp}(S^*, x_0^* - \Delta) - \epsilon f_{kp}(S^*, x_0^* - \Delta)\\ &\geq f_{kp}(S^*, x_0^* - \Delta) - \epsilon\text{Profit}(S^*),\end{aligned}$$

where the last inequality is because $\text{Profit}(S^*) = f_{kp}(S^*, x_0^*) \geq f_{kp}(S^*, x_0^* - \Delta)$. Hence,

$$\begin{aligned}\frac{\text{Profit}(S^A)}{\text{Profit}(S^*)} &\geq \frac{f_{kp}(S^*, x_0^* - \Delta)}{\text{Profit}(S^*)} - \epsilon\\ &= \frac{f_{kp}(S^*, x_0^* - \Delta)}{f_{kp}(S^*, x_0^*)} - \epsilon\\ &= \frac{\sum_{j\in S^*} p_j x_j(x_0^* - \Delta)}{\sum_{j\in S^*} p_j x_j(x_0^*)} - \epsilon\\ &\geq \min_{j\in\mathcal{Q}}\left\{\frac{x_j(x_0^* - \Delta)}{x_j(x_0^*)}\right\} - \epsilon.\end{aligned}$$

Since $\forall j \in \mathcal{Q}$, we have

$$\frac{x_j(x_0^* - \Delta)}{x_j(x_0^*)} = 1 - \left(\frac{x_j(x_0^*) - x_j(x_0^* - \Delta)}{x_j(x_0^*)\cdot\Delta}\right)\cdot\Delta \geq 1 - M\cdot\Delta,$$

then we obtain

$$\frac{\text{Profit}(S^A)}{\text{Profit}(S^*)} \geq 1 - M\cdot\Delta - \epsilon.$$

Next, for case 2 where $x_0^\gamma = \underline{x_0}$, that is, $x_0^* < \underline{x_0} + \Delta$, following similar arguments, we obtain:

$$\frac{\text{Profit}(S^A)}{\text{Profit}(S^*)} \geq \min_{j\in\mathcal{Q}}\left\{\frac{x_j(\underline{x_0})}{x_j(x_0^*)}\right\} - \epsilon.$$

Note that

$$\frac{x_j(x_0^*) - x_j(\underline{x_0})}{x_j(x_0^*)\cdot\Delta} \leq \frac{x_j(\underline{x_0} + \Delta) - x_j(\underline{x_0})}{x_j(\underline{x_0} + \Delta)\cdot\Delta} \leq M.$$

Hence,

$$\frac{\text{Profit}(S^A)}{\text{Profit}(S^*)} \geq \min_{j\in\mathcal{Q}}\left\{1 - \frac{x_j(x_0^*) - x_j(\underline{x_0})}{x_j(x_0^*)\cdot\Delta}\cdot\Delta\right\} - \epsilon \geq 1 - M\cdot\Delta - \epsilon.$$

Therefore, in both cases,

$$\frac{\text{Profit}(S^A)}{\text{Profit}(S^*)} \geq 1 - M\cdot\Delta - \epsilon.$$

Notice that $\Delta = \underline{x_0}(1+\psi)^l\psi$, where $\underline{x_0}(1+\psi)^l < x_0^* \leq \underline{x_0}(1+\psi)^{l+1}$. Hence, $\Delta < x_0^*\psi \leq \psi$. Since we let $\psi = \frac{\eta}{2\cdot M}$, we have $\Delta < \frac{\eta}{2\cdot M}$. In addition, we let $\epsilon = \frac{\eta}{2}$. As a result,

$$\frac{\text{Profit}(S^A)}{\text{Profit}(S^*)} \geq 1 - M\cdot\Delta - \epsilon > 1 - M\cdot\frac{\eta}{2\cdot M} - \frac{\eta}{2} = 1 - \eta,$$

and the corresponding run time of Algorithm 2 is $\mathcal{O}\left(\frac{\log(\frac{1}{\underline{x_0}})}{\log(1+\frac{\eta}{2M})}\left(Q^2\left\lfloor\frac{Q}{\epsilon}\right\rfloor + QC\right)\right)$.

■

### C.6. A Detailed Analysis of Efficiency of Algorithm 2

It should be noted that Algorithm 2 is effective for all MDMs with the current parameter selection (e.g., the selection for $\psi$ value). However, for certain MDMs, adjusting the $\psi$ value can further enhance the algorithm efficiency. Here, we will demonstrate an example of the MDC model with strictly positive choice probabilities, which is a generalization of the p-MDC model because $\gamma_i = 0, \forall i \in \mathcal{Q}$ is a sufficient but not necessary condition to achieve zero probabilities for all products. Prior to presenting the result (Lemma 7), we will first provide two lemmas that will be useful in the proof.

**Lemma 5** *Given* $0 < A^- \le A^+$, $0 < B^- \le B^+$, $\frac{A^-}{B^-} \ge \frac{A^+}{B^+}$, $M^- = \frac{B^-}{B^- - C}$, $M^+ = \frac{B^+}{B^+ - C}$, $B^- - C > 0$ *and* $B^+ - C > 0$, *then* $\frac{A^-}{B^- - C} \ge \frac{A^+}{B^+ - C}$.

**Proof of Lemma 5.** From the given condition, we have $M^- = 1 + \frac{C}{B^- - C}$ and $M^+ = 1 + \frac{C}{B^+ - C}$, hence, $M^- \ge M^+ > 0$. Since $\frac{A^-}{B^-} \ge \frac{A^+}{B^+}$, we have $\frac{A^-}{B^-} \cdot M^- \ge \frac{A^+}{B^+} \cdot M^+$. Rearrange the terms then we have $\frac{A^-}{B^- - C} \ge \frac{A^+}{B^+ - C}$. ■

**Lemma 6** *Given* $a_j > 0, b_j \ge 1, c_j >= 0, \forall j \in \mathcal{Q}$, *we have*

$$\frac{b_j a_j (x_0^*)^{b_j - 1}}{a_j (x_0^*)^{b_j - 1} - c_j} \le \frac{\underline{x_0}}{x_0^*} \cdot \frac{b_j a_j (\underline{x_0})^{b_j - 1}}{a_j (\underline{x_0})^{b_j - 1} - c_j}.$$

**Proof of Lemma 6.** Since $b_j \ge 1$, we have $a_j (\underline{x_0})^{b_j} \le a_j (x_0^*)^{b_j}$, where $x_0^* \ge \underline{x_0}$. Hence, we have

$$\frac{a_j (\underline{x_0})^{b_j} - c_j}{a_j (x_0^*)^{b_j} - c_j} \le \frac{a_j (\underline{x_0})^{b_j}}{a_j (x_0^*)^{b_j}} = \frac{(\underline{x_0})^{b_j}}{(x_0^*)^{b_j}},$$

then, multiply the same term $\left[\frac{x_0^*}{\underline{x_0}} \cdot \frac{b_j a_j (x_0^*)^{b_j - 1}}{b_j a_j (\underline{x_0})^{b_j - 1}}\right]$ on both sides of the equality:

$$\left[\frac{x_0^*}{\underline{x_0}} \cdot \frac{b_j a_j (x_0^*)^{b_j - 1}}{b_j a_j (\underline{x_0})^{b_j - 1}}\right] \cdot \frac{a_j (\underline{x_0})^{b_j} - c_j}{a_j (x_0^*)^{b_j} - c_j} \le \left[\frac{x_0^*}{\underline{x_0}} \cdot \frac{b_j a_j (x_0^*)^{b_j - 1}}{b_j a_j (\underline{x_0})^{b_j - 1}}\right] \cdot \frac{(\underline{x_0})^{b_j}}{(x_0^*)^{b_j}} = 1.$$

Rearrange the terms and then we obtain:

$$\frac{b_j a_j (x_0^*)^{b_j - 1}}{a_j (x_0^*)^{b_j - 1} - c_j} \le \frac{\underline{x_0}}{x_0^*} \cdot \frac{b_j a_j (\underline{x_0})^{b_j - 1}}{a_j (\underline{x_0})^{b_j - 1} - c_j}.$$

■

Now, we present a better selection for $\psi$ value under such a model.

**Lemma 7** *For the MDC model with strictly positive choice probabilities, Algorithm 2 with* $\psi = (1/\underline{x_0})^{\frac{1-\sigma_0}{1-\sigma_{min}}} \cdot \frac{\eta}{2M}$ *and* $\epsilon = \frac{\eta}{2}$ *provides a* $(1-\eta)$ *approximation guarantee.*

**Proof of Lemma 7.** Since choice probabilities are strictly positive, we have

$$x_j(x_0) = \left[\left(\frac{\psi_j \sigma_j}{\psi_0 \sigma_0}\right)^{\frac{1}{1-\sigma_j}} (\kappa)^{\frac{1-\sigma_0}{1-\sigma_j} - 1} (x_0)^{\frac{1-\sigma_0}{1-\sigma_j}} - \frac{\gamma_j}{\kappa}\right]^+ = \left(\frac{\psi_j \sigma_j}{\psi_0 \sigma_0}\right)^{\frac{1}{1-\sigma_j}} (\kappa)^{\frac{1-\sigma_0}{1-\sigma_j} - 1} (x_0)^{\frac{1-\sigma_0}{1-\sigma_j}} - \frac{\gamma_j}{\kappa} > 0,$$

where $[a,0]^+ = \max\{a, 0\}$. For simplicity, let us denote $x_j(x_0) = a_j (x_0)^{b_j} - c_j > 0$. That is, $\underline{x_0} > (c_j/a_j)^{1/b_j}$.

Note that we only consider the problem where $b_{\min} := \min_{j \in \mathcal{Q}} b_j < 1$ because otherwise, it is implied by Corollary 1 that the problem has a profit-nested optimal solution. Now, Recall that

$$M = \max_{j \in \mathcal{Q}} \{M_j\} := \max_{\underline{x_0} + \Delta \le x_0 \le \overline{x}_0, \Delta > 0} \left(\frac{x_j(x_0) - x_j(x_0 - \Delta)}{x_j(x_0)(\Delta)}\right) := \max_{\underline{x_0} + \Delta \le x_0 \le \overline{x}_0, \Delta > 0} h(x_0, \Delta).$$

We first calculate $M_j, \forall j \in \mathcal{Q}$. Note that $x_j(x_0)$ is increasing and can only be convex or concave. First, when it is a convex function, we compute

$$\frac{\partial}{\partial \Delta} h(x_0,\Delta) = \frac{-x_j(x_0) + x_j(x_0-\Delta) + x_j'(x_0-\Delta)\cdot\Delta}{x_j(x_0)\Delta^2} \le 0,$$

where $x_j'(\cdot)$ is the first order derivative and the inequality is because $x_j(x_0)$ is convex, hence $x_j(x_0) \ge x_j(x_0-\Delta) + x_j'(x_0-\Delta)\cdot\Delta$. Therefore, $h(x_0,\Delta)$ is maximized when $\Delta \to 0$. That is,

$$M_j = \max_{\underline{x_0} \le x_0 \le \overline{x}_0} \frac{x_j'(x_0)}{x_j(x_0)} = \max_{\underline{x_0} \le x_0 \le \overline{x}_0} \frac{b_j a_j (x_0)^{b_j-1}}{a_j(x_0)^{b_j} - c_j} = \frac{b_j a_j (\underline{x_0})^{b_j-1}}{a_j(\underline{x_0})^{b_j} - c_j},$$

where the last equality is based on Lemma 5 in which we let $A^- = b_j a_j(\underline{x_0})^{b_j-1}$, $B^- = a_j(\underline{x_0})^{b_j}$, $A^+ = b_j a_j(x_0)^{b_j-1}$, $B^+ = a_j(x_0)^{b_j}$.

When $x_j(x_0)$ is concave, it is easy to see that $h(x_0,\Delta)$ is maximized when $x_0 \to \underline{x_0}$, which also implies that $\Delta \to 0$. Hence,

$$M_j = \frac{x_j'(\underline{x_0})}{x_j(\underline{x_0})} = \frac{b_j a_j(\underline{x_0})^{b_j-1}}{a_j(\underline{x_0})^{b_j} - c_j}.$$

Therefore, in both cases, $M_j = \frac{b_j a_j(\underline{x_0})^{b_j-1}}{a_j(\underline{x_0})^{b_j} - c_j}$.

From the proof of Theorem 8, we know that $\frac{\text{Profit}(S^A)}{\text{Profit}(S^*)} \ge \min_{j\in\mathcal{Q}} \left\{ \frac{x_j(x_0^*-\Delta)}{x_j(x_0^*)} \right\} - \epsilon$ and $\Delta < x_0^*\psi$. Next, we will show that in both 'convex' and 'concave' cases, we can achieve $1-\eta$ approximation guarantee. If $x_j(x_0)$ is convex, that is $b_j \ge 1$. Then, we have

$$\frac{x_j(x_0^*-\Delta)}{x_j(x_0^*)} = 1 - \left( \frac{x_j(x_0^*) - x_j(x_0^*-\Delta)}{x_j(x_0^*)\Delta} \right)\cdot\Delta \ge 1 - \frac{\underline{\underline{x_0}}}{x_0^*}\cdot M\cdot\Delta,$$

where the last inequality is because

$$\frac{x_j(x_0^*) - x_j(x_0^*-\Delta)}{x_j(x_0^*)\Delta} \le \frac{x_j(x_0^*) - x_j(x_0^*-\Delta)}{x_j(x_0^*)\Delta}\Big|\Delta \to 0 = \frac{x_j'(x_0^*)}{x_j(x_0^*)} = \frac{b_j a_j(x_0^*)^{b_j-1}}{a_j(x_0^*)^{b_j-1} - c_j},$$

and Lemma 6 implies that

$$\frac{b_j a_j(x_0^*)^{b_j-1}}{a_j(x_0^*)^{b_j-1} - c_j} \le \frac{\underline{\underline{x_0}}}{x_0^*}\cdot M.$$

Hence,

$$\frac{\text{Profit}(S^A)}{\text{Profit}(S^*)} \ge 1 - \frac{\underline{\underline{x_0}}}{x_0^*}\cdot M\cdot\Delta - \epsilon.$$

Recall that $\psi = (\frac{1}{\underline{x_0}})^{b_{\min}} \cdot \frac{\eta}{2M} \le (\frac{1}{\underline{x_0}}) \cdot \frac{\eta}{2M}$, $\Delta < \psi x_0^*$ and $\epsilon = \frac{\eta}{2}$, then, we have

$$\frac{\text{Profit}(S^A)}{\text{Profit}(S^*)} \ge 1 - \frac{\underline{\underline{x_0}}}{x_0^*}\cdot M \cdot x_0^* \cdot (\frac{1}{\underline{x_0}})\cdot\frac{\eta}{2M} - \frac{\eta}{2} = 1-\eta.$$

When $x_j(x_0)$ is concave, that is $b_j < 1$, we have

$$\begin{aligned}
\frac{x_j(x_0^*) - x_j(x_0^*-\Delta)}{x_j(x_0^*)\Delta} &= \frac{1}{x_j(x_0^*)}\cdot(\frac{x_j(x_0^*) - x_j(x_0^*-\Delta)}{\Delta}) \\
&\le \frac{1}{x_j(x_0^*)}\cdot x_j'(\underline{x_0}) \\
&= \frac{b_j a_j(\underline{x_0})^{b_j-1}}{a_j(x_0^*)^{b_j} - c_j} \\
&= \frac{a_j(\underline{x_0})^{b_j} - c_j}{a_j(x_0^*)^{b_j} - c_j}\cdot M_j
\end{aligned}$$

$$\le \frac{a_j(\underline{x_0})^{b_j}}{a_j(x_0^*)^{b_j}} \cdot M_j$$
$$= (\frac{\underline{x_0}}{x_0^*})^{b_j} \cdot M_j$$
$$\le (\frac{\underline{x_0}}{x_0^*})^{b_j} \cdot M.$$

Recall that $\Delta < \psi x_0^*$, $\epsilon = \frac{\eta}{2}$ and

$$\psi = (\frac{1}{\underline{x_0}})^{b_{\min}} \cdot \frac{\eta}{2M}$$
$$\le (\frac{1}{\underline{x_0}})^{b_j} \cdot \frac{\eta}{2M}$$
$$= (\frac{x_0^*}{\underline{x_0}})^{b_j} \cdot (\frac{1}{x_0^*})^{b_j} \cdot \frac{\eta}{2M}$$
$$\le (\frac{x_0^*}{\underline{x_0}})^{b_j} \cdot (\frac{1}{x_0^*}) \cdot \frac{\eta}{2M},$$

Where the last inequality is due to $b_j < 1$. Hence, we have

$$\frac{\text{Profit}(S^A)}{\text{Profit}(S^*)} \ge 1 - \Big(\frac{\underline{x_0}}{x_0^*}\Big)^{b_j} \cdot M \cdot \Delta - \epsilon \ge 1 - \Big(\frac{\underline{x_0}}{x_0^*}\Big)^{b_j} \cdot M \cdot \Big((\frac{x_0^*}{\underline{x_0}})^{b_j} \cdot (\frac{1}{x_0^*}) \cdot \frac{\eta}{2M} \cdot x_0^*\Big) - \frac{\eta}{2} = 1 - \eta.$$

Therefore, in both cases, we achieve a $(1 - \eta)$ approximation guarantee. ■

To see how efficient Algorithm 2 is, we can examine a simple p-MDC case where $\overline{x_0} \to 1$, $\underline{x_0} = 0.001$, $b_{\max} = 3$, $b_{\min} = 0.7$ and $\eta = 0.1$. To achieve the $1 - \eta$ approximation guarantee, Algorithm 1 needs to conduct 39960 times knapsack approximations while Algorithm 2 only run the knapsack approximation for 2199 times, which is less than $\frac{1}{18}$ of the runtime of Algorithm 1.

## D. Additional Results for E-Commerce Data

Note that the data of the following 7 categories: 'accessories bag', 'appliances personal hair cutter', 'accessories cosmetic bag', 'appliances environment air conditioner', 'furniture living room chair', 'sport diving', and 'appliances personal massager', are not suitable for our analysis because, in these data sets, even the most popular products have extremely low market shares. For example, in the 'sport diving' category, no product was purchased during the month. Thus we did not conduct experiments on these datasets. Instead, we only work on the other four categories. Results are displayed in Tables 4 and 5.

| **apparel glove** (original) | | | |
|---|---|---|---|
| data number | product number | outside market purchases | |
| 2669 | 51 | 2012 | |
| apparel glove (scale-up factor = 20, $\kappa = 2$, remove small-share products) | | | |
| data number | product number | outside market share | |
| 13326 | 24 | 57.68% | |
| Model | in-sample loglklhd | out-of-sample RMSE | time/iteration number |
| the MNL model | -8617.50 | 0.21066 | 1.48s / 10 |
| the p-MDC model | -8265.84 | 0.19493 | 2.55s / 50 |
| the MDC model | -8132.12 | 0.19048 | 903.88s / $>4000$ |
| the MMM | -8137.05 | 0.18977 | 19.74s / 405 |
| the MGM | -8162.48 | 0.19147 | 8.98s / 178 |
| the MLM | -8155.73 | 0.19124 | 6.28s / 78 |
| the MCM | -8196.25 | 0.19141 | 6.05s / 94 |
| the MUM | -8112.25 | 0.19004 | 897.13s / $>4000$ |
| **appliances environment vacuum** (original) | | | |
| data number | product number | outside market purchases | |
| 7520 | 82 | 6863 | |
| appliances environment vacuum (scale-up factor = 20, $\kappa = 3$, remove small-share products) | | | |
| data number | product number | outside market share | |
| 15370 | 21 | 75.62% | |
| Model | in-sample loglklhd | out-of-sample RMSE | time/iteration number |
| the MNL model | -8253.02 | 0.16797 | 1.59s / 7 |
| the p-MDC model | -8113.48 | 0.16460 | 3.52s / 52 |
| the MDC model | -8084.10 | 0.16539 | 1465.45s / $>4000$ |
| the MMM | -8104.12 | 0.16457 | 13.74s / 197 |
| the MGM | -8101.64 | 0.16494 | 14.15s / 266 |
| the MLM | -8101.52 | 0.16503 | 9.14s / 148 |
| the MCM | -8109.95 | 0.16433 | 6.04s / 91 |
| the MUM | -8089.92 | 0.16538 | 235.20s / 684 |

**Table 4 Results for Data of Other Categories**

| **furniture living room cabinet** (original) | | | |
|---|---|---|---|
| data number | product number | outside market purchases | |
| 3266 | 5 | 3218 | |
| furniture living room cabinet (scale-up factor = 20, $\kappa = 3$, remove small-share products) | | | |
| data number | product number | outside market share | |
| 3380 | 4 | 90.33% | |
| Model | in-sample loglklhd | out-of-sample RMSE | time/iteration number |
| the MNL model | -919.50 | 0.14498 | 0.47s / 7 |
| the p-MDC model | -908.34 | 0.14456 | 0.48s / 59 |
| the MDC model | -906.46 | 0.14482 | 5.69s / 122 |
| the MMM | -905.55 | 0.14464 | 2.33s / 458 |
| the MGM | -906.22 | 0.14468 | 0.67s / 43 |
| the MLM | -906.54 | 0.14466 | 0.72s / 58 |
| the MCM | -906.37 | 0.14481 | 0.78s / 75 |
| the MUM | -906.91 | 0.14484 | 2.76s / 67 |
| **furniture bathroom bath** (original) | | | |
| data number | product number | outside market purchases | |
| 2145 | 48 | 1947 | |
| furniture bathroom bath (scale-up factor = 20, $\kappa = 3$, remove small-share products) | | | |
| data number | product number | outside market share | |
| 5457 | 17 | 74.10% | |
| Model | in-sample loglklhd | out-of-sample RMSE | time/iteration number |
| the MNL model | -2734.68 | 0.16611 | 0.65s / 6 |
| the p-MDC model | -2682.54 | 0.16046 | 1.07s / 72 |
| the MDC model | -2659.58 | 0.16085 | 258.78s / $>4000$ |
| the MMM | -2668.41 | 0.16087 | 4.19s / 191 |
| the MGM | -2668.38 | 0.16110 | 3.90s / 224 |
| the MLM | -2668.70 | 0.16122 | 3.86s / 269 |
| the MCM | -2673.43 | 0.16089 | 5.56s / 468 |
| the MUM | -2660.02 | 0.16143 | 25.72s / 341 |

**Table 5 Results for Data of Other Categories**