

On MIP Formulations for Logit-Based Multi-Purchase Choice Models and Applications

Taotao He, Zhongqi Wu, Yating Zhang

Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai 200030, China.

Contact: hetaotao@sjtu.edu.cn (TH), wuzhongqi@sjtu.edu.cn, ytzhang20@sjtu.edu.cn (YZ)

Abstract. We study logit-based multi-purchase choice models and develop an exact solution methodology for the resulting assortment optimization problems, which we show are NP-hard to approximate. We introduce a hypergraph representation that captures general bundle-based choice structures and subsumes several models in the literature, including the BundleMVL-K and multivariate MNL models (Tulabandhula et al. 2023, Jasin et al. 2024). Leveraging this representation, we derive mixed-integer programming (MIP) formulations by integrating polyhedral relaxations from multilinear optimization with a perspective reformulation of the logit choice model. Our approach preserves the strength of the underlying polyhedral relaxations, yielding formulations with provably tighter linear programming (LP) bounds than the prevalent Big-M approach. We further characterize structural conditions on the hypergraph under which the formulations are locally sharp, thereby generalizing existing LP characterizations for path-based models. The framework extends naturally to heterogeneous and robust settings. Computational experiments demonstrate that the proposed formulations significantly improve both solution quality and scalability.

Key words: Logit choice models; Multi-purchase behavior; Assortment optimization; Multilinear optimization; Perspective formulation

1. Introduction

Customers frequently purchase a basket of multiple items within a single transaction. This multi-purchase behavior has been consistently documented across a wide range of empirical settings (e.g., Manchanda et al. 1999, Russell and Petersen 2000). Incorporating such multi-purchase behavior into operational decisions, such as assortment planning and pricing, is crucial. By capturing cross-selling opportunities and inter-product dependencies, firms can improve demand prediction and increase total revenue (Tulabandhula et al. 2023, Jasin et al. 2024).

One prevalent approach to model multi-purchase behavior is logit-based models, referred to as Logit-MP models and formally introduced in Section 2.1 (see a comprehensive review of other models in Section 1.1). These models generalize the Multinomial Logit (MNL) model by allowing alternatives to be subsets (or bundles) of products. Notable examples include the Multivariate logit (MVMNL) (Aurier and Mejia 2014, Jasin et al. 2024), cardinality-restricted bundle logits (e.g., BundleMVL-K) (Tulabandhula et al. 2023). Theoretically, for bundles of cardinality two, these models admit a rigorous axiomatic characterization (Chambers et al. 2024) and can be derived

via spatial statistics frameworks (Russell and Petersen 2000). Empirically, these models have been evaluated and calibrated using Maximum Likelihood Estimation (MLE) (Aurier and Mejia 2014). For large-scale applications, recent works achieve estimation scalability through regularization (Vasilyev et al. 2025) or sequential choice models with latent embeddings (Ruiz et al. 2020). Consequently, these models have become a standard tool in customer behavior analysis, including cross-category effects (Song and Chintagunta 2006, Kwak et al. 2015), promotion effects (Pan et al. 2025), and price competition (Richards et al. 2018).

Motivated by their robust empirical performance, recent studies have sought to integrate these models into prescriptive decision-making processes, particularly for assortment planning (Chen et al. 2022, Tulabandhula et al. 2023, Jasin et al. 2024). For instance, Jasin et al. (2024) demonstrate numerically that incorporating multi-purchase behavior can increase a retailer’s expected revenue by up to 14%. However, to maintain computational tractability, existing research often relies on structural restrictions on the feasible bundle set, enabling the design of tailored heuristics or approximation algorithms. Specifically, Tulabandhula et al. (2023) propose a computationally efficient heuristic to identify a feasible assortment under the BundleMVL-K model with a cardinality constraint of $K = 2$. Similarly, Jasin et al. (2024) develop a fully polynomial-time approximation scheme (FPTAS) for the assortment problem under the MVMNL model, provided each bundle contains at most one product per category. These methodological limitations motivate a fundamental research question: Can we develop a rigorous global optimization methodology that accommodates arbitrary, general purchase patterns under Logit-MP models?

Addressing this question requires a computationally tractable representation of the mapping between the firm’s discrete decision space (i.e., binary assortment decisions and discrete price tiers) and the resulting multi-purchase choice probabilities. This task is inherently challenging due to the combinatorial and nonconvex nature of Logit-MP models. In the context of assortment planning, for example, the combinatorial challenge arises from modeling the set of all possible bundles considered by the customer. Simultaneously, the nonconvexity is driven by the fractional structure of the logit probabilities. These challenges are significantly compounded when allowing for general multi-purchase patterns.

We address the question by developing a mixed-integer programming (MIP) formulation approach for the Logit-MP models. Our methodology integrates convexification techniques from multilinear optimization and fractional programming through a three-stage approach. First, we build a hypergraph-based representation of the Logit-MP model, where nodes denote individual

products and hyperedges characterize the collection of bundles considered by the customer. This abstraction is sufficiently general to encompass diverse purchase patterns, including those found in MVMNL and BundleMVL-K. Second, to resolve the combinatorial dependencies within the bundle sets, we leverage results from multilinear optimization to derive valid inequalities tailored to the hypergraph structure. Last, we use a perspective transformation to linearize the fractional terms inherent in the logit probabilities. A key theoretical advantage of this transformation is that it is strength-preserving. In other words, it maps the tight polyhedral characterization achieved in the combinatorial stage directly into the fractional space without loss of relaxation quality.

We highlight the contributions of our paper as follows.

1. We provide a hypergraph-based representation of Logit-MP models, which unifies various multi-purchase patterns in literature, such as MVMNL and BundleMVL-2 in [Chen et al. \(2022\)](#), [Tulabandhula et al. \(2023\)](#), and [Jasin et al. \(2024\)](#), and enables us to leverage structural properties in the hypergraph to build strong formulations. The underlying hypergraph structures can be identified directly from transaction data via an estimation procedure, as detailed in Section 4. Furthermore, we formally characterize the computational complexity inherent in general multi-purchase patterns. Specifically, Theorem 3 shows that assortment optimization under general Logit-MP models is NP-hard to approximate. This result formally justifies the necessity of our MIP approach.

2. Our proposed MIP formulation provides an exact solution for assortment optimization under Logit-MP models with arbitrary purchase patterns, extending naturally to heterogeneous population variants (see Section 5). Beyond its flexibility, a primary theoretical contribution of our formulation is the tightness of its linear programming (LP) relaxation. Specifically, we first show that the LP relaxation of our formulation is tighter than that of the prevalent Big-M formulation obtained using the Charnes–Cooper transformation, even when both formulations use the same relaxation oracle for the combinatorial structures. We also characterize conditions on a multi-purchase hypergraph under which our LP relaxation is *locally sharp*, *i.e.*, the tightest possible linear relaxation in the space of choice probabilities. We demonstrate that our LP relaxation exactly solves the unconstrained assortment planning problem for a broad class of structures, including series–parallel graphs and specific hypergraphs with nested overlapping hyperedges. This result significantly generalizes the existing LP characterization for paths established in [Lo and Topaloglu \(2019\)](#), which relies on the sales-based linear program proposed by [Gallego et al. \(2015\)](#).

3. To solve our formulations efficiently, we leverage structural properties of the multi-purchase hypergraph and the perspective transformation to develop a cutting-plane implementation, ensuring both the tightness of LP relaxation and the computational tractability of the solution process.

We computationally demonstrate the value of our methodology across several classes of assortment optimization problems. In all tested instances, our formulations outperform both the Big-M formulations and the formulations based on the state-of-the-art conic integer optimization approach (Sen et al. 2018). In particular, when customers purchase at most two products, our formulations find a global optimal assortment for 1000 products within ten minutes. Furthermore, for more complex patterns, incorporating *running-intersection inequalities*, proposed by Del Pia and Khajavirad (2021), into our models yields substantial speedups, significantly increasing the number of instances solved within a time limit. Finally, our methodology exhibits consistent performance in heterogeneous extensions, including the mixture of Logit-MP models and their robust counterparts.

1.1. Literature Review

Multi-Purchase Choice Models. Existing studies on multi-purchase behavior modeling generally fall into three main categories. The first stream originates from the empirical need to estimate co-purchase dynamics and cross-product dependencies. Early empirical studies extended multinomial logit/probit frameworks to multiple items along with inter-product terms to capture complementarity. This led to the development of Multivariate Probit (MVP) and Multivariate Logit models (Aurier and Mejia 2014). While MVP models offer an approach to capture cross-product correlations, their practical adoption is often bottlenecked by the reliance on computationally expensive Monte Carlo Markov Chain (MCMC) methods (Manchanda et al. 1999). In contrast, due to the estimation tractability, MVMNL is widely used in empirical analyses (e.g., Kwak et al. 2015, Richards et al. 2018) and also operational decisions (e.g., Jasin et al. 2024, Pan et al. 2025). More recently, MVMNL has expanded to direct bundle-level modeling, such as the cardinality-restricted BundleMVL-K model (Tulabandhula et al. 2023). We refer to these logit-based multi-purchase formulations as Logit-MP models. Their structural flexibility and widespread adoption in empirical literature underscore the importance of developing exact MIP formulations to solve the associated large-scale optimization problems.

The second stream of literature explicitly models customers who maximize a structural utility function over a bundle of products, often subject to a budget or quantity constraint. This line of work includes the multi-discrete choice model (MDC) (Huh and Li 2022), the multiple discrete-continuous extreme value model (Bhat 2005), and threshold utility models (Gallego et al. 2019). A notable example is the Marginal Distribution Model (MDM) (Sun et al. 2020), which can identify the potential of unseen product combinations if latent marginal distributions are accurately estimated. Hence, MDM is widely used in bundle design studies (Li et al. 2022, Sun et al. 2025).

A third stream departs from bundle-level utility by focusing on ranking-based mechanisms. Under this paradigm, customers are assumed to rank individual items and select the top K products with the highest latent utilities. For instance, [Bai et al. \(2024\)](#) assume product utilities follow an MNL distribution and derive a logit-like probability form for the top K items. It is important to distinguish this from the first stream: while the resulting formula also appears logit-like, its derivation is rooted in the ordering of individual items rather than the utility assigned to a fixed bundle. Other extensions generalize this by either applying order statistics to select the top K products under general utility functions ([Abdallah et al. 2024](#)), or by assuming that customers pick the top K available items from a fixed preference list ([Lin et al. 2026](#)).

Optimization under Multi-Purchase Choice Models. A burgeoning body of literature has integrated multi-purchase choice models into operational decision-making, with a primary focus on assortment planning. However, multi-purchase dynamics introduce significant optimization challenges. Under the Logit-MP model, most studies focus on heuristic or approximation algorithms. For instance, [Jasin et al. \(2024\)](#) proposed an FPTAS under a special case of feasible bundles defined by the MVMNL model. Similarly, [Tulabandhula et al. \(2023\)](#) rely on heuristics to obtain a good assortment decision under the BundleMVL-K model. Our work contributes to this stream of literature by deriving exact MIP formulations for assortment optimization under Logit-MP models with arbitrary purchase patterns.

For the second and third streams of multi-purchase literature, the corresponding optimization problems are more difficult. Under utility maximization with budget constraints, the intractability stems from the complex bi-level structure, i.e., customers' choices and operational decisions. Under ranking-based models, the difficulty lies in formulating the relation between operational decisions and top-K rules. As a result, the literature mainly relies on approximation algorithms such as MDC ([Zhang et al. 2021](#)), MDM ([Mao and Li 2025](#)), top-K MNL utilities ([Bai et al. 2024](#)), and top-K preference lists ([Yang et al. 2025](#)). A notable exception is the MDC model, under which continuous pricing optimization admits highly tractable solutions ([Huh and Li 2022](#)).

Our methodology contributes to a growing stream of research that leverages integer programming techniques to solve optimization problems involving customer choice models. Recent work has focused on linearizing the fractional structure of logit-based choices, including convexification results for the MNL model in quick-commerce setting ([Chen et al. 2026](#)), and strong formulations for the mixture of MNL models by using second-order cones ([Sen et al. 2018](#)) and submodularity ([Atamtürk and Gómez 2020](#), [Kılınç-Karzan et al. 2025](#)). [Akçakuş and Mišić \(2025\)](#) propose

a perspective formulation for a logit-based share-of-choice product design problem, with subsequent extensions to history-dependent effects (He et al. 2025b) and resource allocation for crime prevention (He et al. 2025a). Under the random utility model, Bertsimas and Mišić (2019) propose an MIP formulation that is integral for single-segment problems, while Ma (2023) and Khalid and Sturt (2025) provide tighter inequalities for multiple segments. Finally, modeling tools for piecewise linear functions and disjunctive constraints have also been used to address pricing problems under a representative customer model (Yan et al. 2022), continuous attribute design (Huchette and Vielma 2023), and decision tree-based assortment optimization (Akchen and Mišić 2026).

1.2. Structure

The rest of the paper is organized as follows. In Section 2, we present preliminaries. Section 3 introduces our MIP formulations and establishes their theoretical properties. In Section 4, we develop a heuristic to identify sparse hypergraph structures from data. Section 5 focuses on applications to assortment planning, while Section 6 extends the framework to heterogeneous and robust settings. We conclude the paper in Section 7. All proofs are given in the ecompanion.

2. Preliminaries

2.1. Logit-Based Multi-Purchase Choice Model

We consider a set of products $V = \{1, \dots, N\}$, where N is a positive integer. In principle, a customer may purchase any bundle in the power set 2^V . However, empirical evidence suggests that customers typically evaluate only a limited subset of available options due to cognitive constraints and search costs (Hauser and Wernerfelt 1990, Roberts and Lattin 1991, Hauser 2014). In multi-purchase settings, observed baskets generally contain only a small number of products (Russell and Petersen 2000, Tulabandhula et al. 2023, Jasin et al. 2024). To capture this behavior, existing models impose structural restrictions on feasible bundles. For example, the Multivariate MNL model, proposed in Jasin et al. (2024), restricts purchases to at most one product per category, whereas the BundleMVL- K model, proposed in Tulabandhula et al. (2023), limits attention to bundles of cardinality at most K , where K is selected based on empirical fit. Motivated by these findings, we characterize customers' choices using a general *consideration set* $E \subseteq 2^V$. While we assume E contains all singleton items to ensure baseline availability, its structure remains completely flexible. Formulated in this way, E serves as a unifying framework that subsumes various existing logit-based multi-purchase models. Furthermore, we develop a data-driven heuristic to explicitly

construct E and validate its practical effectiveness using real-world transaction data (see Section 4 for details).

Given an assortment $S \subseteq V$, customers choose among the bundles in their consideration set that are feasible under S , namely $2^S \cap E$. Each bundle e is associated with a deterministic utility parameter u_e , and customer purchases follow a logit choice rule. Specifically, given an assortment S and a consideration set E , the probability of purchasing bundle e is

$$P_E(e, S) = \frac{\exp(u_e) \mathbb{1}(e \subseteq S)}{1 + \sum_{c \in 2^S \cap E} \exp(u_c)}, \quad (\text{LOGIT-MP})$$

where we normalize the utility of the no-purchase option to zero. The indicator function $\mathbb{1}(e \subseteq S)$ ensures that the purchase probability of bundle e is strictly zero unless the entire bundle is included in the offered assortment S . Model (LOGIT-MP) extends the standard MNL model to a multi-purchase setting. In particular, when $E = V$, it reduces to the classical single-item MNL model. In Section 4, we will discuss how to identify the consideration set E and estimate utilities u . We also demonstrate that (LOGIT-MP) significantly improves prediction performance compared to the MNL model via a real transaction dataset.

2.2. A Hypergraph Representation

This subsection introduces a hypergraph representation of (LOGIT-MP), which connects choice modeling to multilinear optimization, enabling us to derive strong formulations in Section 3.

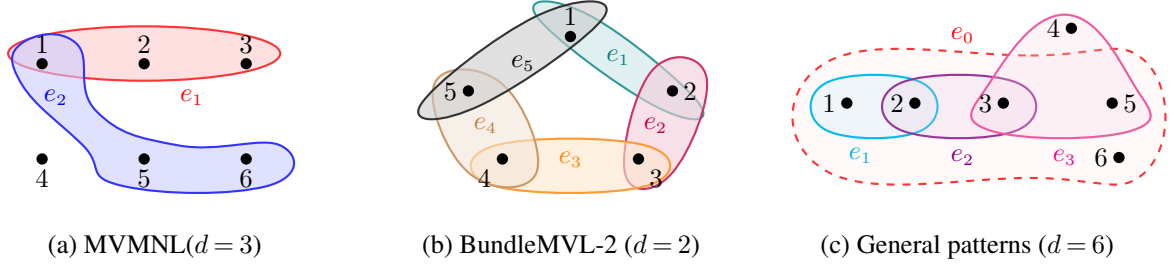
A hypergraph is a generalization of a graph in which an edge can connect a subset of nodes, and such edges are called *hyperedges*. We model multi-purchase behavior using a hypergraph $G(V, E)$, where the node set V represents the items and the hyperedge set E is exactly the bundle consideration set. The maximum bundle size, formally known as the *rank* of the hypergraph, is defined as $d := \max\{|e| : e \in E\}$. Given a predefined consideration generation rule f , e.g., a retailer offering specific suits and matching ties, we denote the corresponding structured consideration set by E_f . To quantify the concentration of customer preferences, we define a sparsity measure for the multi-purchase patterns as:

$$\theta(G) = \frac{|E \setminus V|}{|E_f \setminus V|}.$$

Next, we use hypergraphs to visualize three different bundle consideration sets. Figure 1a represents the MVMNL model (Jasin et al. 2024), where customers choose at most one product from

the k -th category for $k = 1, 2, 3$. Figure 1b illustrates the BundleMVL-2 model introduced in Tulabandhula et al. (2023), where customers buy at most two products. Figure 1c shows a general multi-purchase pattern, including disjoint, nested, and chain-structured choices. For example, choices e_1 and $\{5\}$ are disconnected. e_1, e_2 and e_3 form a chain structure included in e_0 .

Figure 1 Hypergraph representations of multipurchase behavior under different bundle consideration sets. The nodes i represent products, and colored regions represent hyperedges (customer bundles).



With a multi-purchase hypergraph G , we associate a set of choice probabilities under the logit-based multi-purchase model as follows. First, we introduce the binary variable $\mathbf{x} \in \{0, 1\}^N$ to denote an assortment, that is, $x_i = 1$ if and only if product i is available. Then, we use an indicator function $\prod_{j \in e} x_j$ to present whether a bundle $e \in E$ can be purchased under the assortment \mathbf{x} . Now, we define a *choice probability set* \mathcal{C}_G as follows.

$$\mathcal{C}_G := \left\{ (\rho, \mathbf{y}) \in \mathbb{R}^{1+|E|} \mid \begin{aligned} \rho &= \frac{1}{1 + \sum_{c \in E} v_c \prod_{i \in c} x_i}, \\ y_e &= \frac{\prod_{i \in e} x_i}{1 + \sum_{c \in E} v_c \prod_{i \in c} x_i} \quad \text{for } \mathbf{x} \in \{0, 1\}^N \end{aligned} \right\}. \quad (1)$$

where ρ is the no-purchase probability and $v_e y_e$ is the choice probability of bundle $e \in E$. We therefore refer to (ρ, \mathbf{y}) as the *choice probability variable*.

The choice probability set \mathcal{C}_G constitutes a fundamental building block for optimization models involving multi-purchase behavior. A canonical application is the assortment planning (or bundle design) problem, in which the decision-maker seeks to select a choice probability vector (ρ, \mathbf{y}) that maximizes the expected revenue $\sum_{e \in E} r_e v_e y_e$, where r_e is the revenue of selling a bundle e . This assortment optimization problem is NP-hard to approximate (see Theorem 3). Furthermore, the problem remains strongly NP-hard even when G is a complete bipartite graph, which rules out the existence of FPTAS (Chen et al. 2022). While existing methods typically rely on restrictive structural assumptions or heuristic approaches (Chen et al. 2022, Tulabandhula et al. 2023,

Jasin et al. 2024, Vasilyev et al. 2025), we propose a mixed-integer programming framework that enables global optimization over a general hypergraph structure (see Section 3). In Section 5, we demonstrate that our formulations can compute globally optimal assortments with up to 1000 items. Moreover, our formulations can be readily embedded as modular subcomponents in more complex optimization models in which the choice probability set arises as a core constraint set, as discussed in Section 6.

Before presenting our formulations, we briefly review some concepts from integer programming. Consider a set defined by linear inequalities and continuous and integer variables as follows

$$L = \{(\mathbf{x}, \mathbf{y}, \boldsymbol{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^s \times \mathbb{Z}^t \mid \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y} + \mathbf{C}\boldsymbol{\lambda} \leq \mathbf{b}\},$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{B} \in \mathbb{R}^{m \times s}$, $\mathbf{C} \in \mathbb{R}^{m \times t}$ and $\mathbf{b} \in \mathbb{R}^m$. We say that L is an MIP formulation of a set $S \subseteq \mathbb{R}^n$ if the projection of L onto the space of \mathbf{x} variables is S , that is,

$$S = \{\mathbf{x} \in \mathbb{R}^n \mid \exists \mathbf{y}, \boldsymbol{\lambda} \text{ s. t. } (\mathbf{x}, \mathbf{y}, \boldsymbol{\lambda}) \in L\}.$$

The polyhedron P obtained by dropping all integrality requirements in L is called linear programming relaxation of L . One of the factors that has a strong impact on the performance of an MIP formulation is the strength of the LP relaxation. A formulation L is said to be *locally sharp* if its LP relaxation yields the strongest possible bound, in the sense that the projection of P onto the space of \mathbf{x} variables coincides with the tightest relaxation of S , i.e. the convex hull of S , denoted as $\text{conv}(S)$ (Vielma 2015). This property is particularly valuable in our setting, as it implies that the assortment optimization problems admit an exact LP formulation.

3. MIP Formulations

In this section, we develop MIP formulations for the choice probability set \mathcal{C}_G , defined as in (1). The set \mathcal{C}_G involves two sources of nonlinearity. First, given a hypergraph representation G of multi-purchase behavior, we associate each hyperedge $e \in E$ with a monomial that indicates whether a bundle is available for purchase. Specifically, we define:

$$\mathcal{M}_G := \left\{ \mathbf{z} \in \mathbb{R}^{|E|} \mid z_e = \prod_{i \in e} x_i, \text{ for } e \in E \text{ and } \mathbf{x} \in \{0, 1\}^{|V|} \right\}. \quad (2)$$

The second source of nonlinearity is the ratio of the attraction value v_e of a bundle and that of all available bundles. To address the two challenges inherent to the logit-based multi-purchase

model, we leverage recent advances in multilinear optimization (Del Pia and Khajavirad 2021) and fractional programming (He et al. 2024), respectively.

More specifically, we propose two approaches for deriving MIP formulations. Both approaches use the same technique to linearize the monomial set \mathcal{M}_G , but differ in how the ratio terms are handled. Formulations in Section 3.1 adopt the projective transformation in He et al. (2024), whereas those in Section 3.2 are built upon the classic Charnes-Cooper transformation (Charnes and Cooper 1962) and the novel conic approach in Sen et al. (2018). We also provide a theoretical comparison of proposed approaches.

3.1. Perspective Formulations

Our approach provides flexibility in modeling the monomial set \mathcal{M}_G . We assume access to a polyhedral relaxation oracle for \mathcal{M}_G described by the linear system

$$\mathcal{R} := \{z \in [0, 1]^{|E|} \mid \exists \mathbf{w} \text{ s.t. } \mathbf{A}z + \mathbf{B}\mathbf{w} \leq \mathbf{b}\}, \quad (3)$$

where \mathbf{w} is additional auxiliary variables, \mathbf{A} and \mathbf{B} are matrices of proper dimensions and \mathbf{b} is a vector of proper dimension. Here, $[0, 1]$ is a trivial bound from the fact that \mathbf{x} is binary. The oracle \mathcal{R} is said to be *exact* if, when combined with the binary constraints $z_e \in \{0, 1\}$ for all $e \in E$, it yields an MIP formulation for \mathcal{M}_G . It is said to be *integral* if for every vertex z of \mathcal{R} , we have $z_e \in \{0, 1\}$ for all $e \in E$.

Given such a relaxation oracle, we construct the formulation

$$\mathbf{y} \in \rho \cdot \mathcal{R} \text{ and } \rho \geq 0 \quad (4a)$$

$$\rho + \sum_{e \in E} v_e y_e = 1 \quad (4b)$$

$$x_i \in \{0, 1\} \quad \text{for } i \in V \quad (4c)$$

$$x_i \geq y_i + \sum_{e \in E \mid i \in e} y_e v_e + \sum_{e \in E \mid i \notin e} v_e \max\{0, y_i + y_e - \rho\} \quad \text{for } i \in V \quad (4d)$$

$$x_i \leq y_i + \sum_{e \in E \mid i \in e} y_e v_e + \sum_{e \in E \mid i \notin e} v_e \min\{y_i, y_e\} \quad \text{for } i \in V. \quad (4e)$$

Constraint (4a) homogenizes the oracle via the nonnegative scaling variable ρ , namely

$$\rho \cdot \mathcal{R} := \{\mathbf{y} \in [0, \rho]^{|E|} \mid \exists \mathbf{w} \text{ s.t. } \mathbf{A}\mathbf{y} + \mathbf{B}\mathbf{w} \leq \mathbf{b}\rho\}.$$

Constraint (4b) enforces normalization of the purchase and non-purchase probabilities. The binary constraint (4c) corresponds to the assortment decision, that is, the i^{th} item is offered if and only

if $x_i = 1$ for $i \in V$. Finally, constraints (4d)–(4e) link the assortment decision \mathbf{x} with the choice probabilities (ρ, \mathbf{y}) and ensure consistency with our model (LOGIT-MP).

A key feature of the construction is that it preserves the polyhedral strength of the oracle \mathcal{R} , that is, improvements in the relaxation of \mathcal{M}_G translate directly into a stronger MIP formulation of the choice probability set \mathcal{C}_G . The preservation of local sharpness is important because it enables polynomial-time solvable linear programming formulations for the assortment optimization problem, as detailed in Section 5. The next result formalizes this preservation property.

THEOREM 1. *For any hypergraph G , the perspective formulation (4) is an exact (resp. locally sharp) MIP formulation for \mathcal{C}_G if the relaxation oracle \mathcal{R} is exact (resp. integral).*

Our formulation is closely related to Remark 3 of He et al. (2024), which establishes an equivalence between the convex hull of the monomial set \mathcal{M}_G and that of the choice probability set \mathcal{C}_G . Rather than pursuing a complete polyhedral description, we focus on developing MIP formulations that are compact, tight, and directly usable for solving operational problems under the logit-based multi-purchase behavior. Motivated by the perspective idea in He et al. (2024), we treat the no-purchase variable as a positive scaling variable, resulting in two systems of inequalities: the scaled polyhedral relaxation in constraint (4a) and the scaled McCormick relaxation of $x_i \times (1 + \sum_{e \in E} v_e y_e)$ in constraints (4d) and (4e).

To apply Theorem 1, we must construct a polyhedral relaxation oracle for \mathcal{M}_G under an arbitrary multi-purchase hypergraph G . This task is challenging for two reasons. First, the availability of a bundle e depends jointly on the offering decisions of all products contained in e . Second, popular products may appear in multiple bundles, creating overlapping monomial terms and non-trivial interactions across bundles. To address these challenges, we introduce three families of valid inequalities that progressively strengthen the formulation of \mathcal{M}_G . Recursive McCormick inequalities linearize individual monomial terms. Building on this foundation, odd-cycle inequalities and running-intersection inequalities further tighten the relaxation by exploiting cyclic and chain structures in the hypergraph.

3.1.1. General Structures: Recursive McCormick Inequality A standard approach for handling multilinear terms is the *recursive McCormick relaxation*. The basic idea is to reduce a multilinear expression into a sequence of bilinear terms and then linearize each bilinear term using McCormick inequalities (McCormick 1976). For an edge $e = (i, j)$, the binary vector (z_e, z_i, z_j) satisfying the indicator relation $z_e = z_i z_j$ can be exactly represented by:

$$\max\{0, z_i + z_j - 1\} \leq z_e \leq \min\{z_i, z_j\}.$$

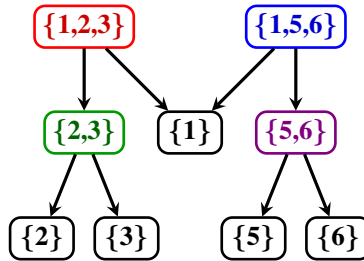
For a hyperedge e containing more than two products, we apply this idea recursively. Specifically, we split e into two disjoint subsets $p \subset e$ and $q \subset e$, and then relax the relation $z_e = z_p z_q$. This procedure is repeated until all subsets reduce to singletons. In this way, a high-order indicator can be represented through a sequence of bilinear indicators, each linearized using McCormick inequalities. The next example illustrates this using the multi-purchase graph in Figure 1a.

EXAMPLE 1. In Figure 1a, the customer considers two bundles consisting of three products, namely $\{1, 2, 3\}$ and $\{1, 5, 6\}$. The recursive decomposition can be visualized using a binary tree with two root nodes, as shown in Figure 2. The bundle set $\{1, 2, 3\}$ is first decomposed into bundles $\{1\}$ and $\{2, 3\}$, and then $\{2, 3\}$ is further split into $\{2\}$ and $\{3\}$. Similarly, the bundle $\{1, 5, 6\}$ is first decomposed into $\{1\}$ and $\{5, 6\}$, and $\{5, 6\}$ is subsequently split into $\{5\}$ and $\{6\}$. Based on the decomposition structure illustrated in Figure 2, the recursive McCormick formulation for the multi-purchase hypergraph in Figure 1a is given as follows:

$$\begin{aligned} \max\{0, z_{23} + z_1 - 1\} \leq z_{123} \leq \min\{z_{23}, z_1\} & \quad \max\{0, z_{56} + z_1 - 1\} \leq z_{156} \leq \min\{z_{56}, z_1\} \\ \max\{0, z_2 + z_3 - 1\} \leq z_{23} \leq \min\{z_2, z_3\} & \quad \max\{0, z_5 + z_6 - 1\} \leq z_{56} \leq \min\{z_5, z_6\}, \end{aligned}$$

where z_{23} and z_{56} are auxiliary variables introduced by the recursive decomposition. \square

Figure 2 Binary-tree representation of the recursive McCormick decomposition for the multi-purchase hypergraph in Figure 1a



We now formally define the procedure. For each multi-purchase hypergraph G , we can construct a binary tree, possibly with multiple roots, to decompose hyperedges E , denoted as $T(G)$. For each non-leaf node t of $T(G)$, its left (resp. right) child is denoted as t_ℓ (resp. t_r), where children t_ℓ and t_r form a partition of their parent t . Given such a binary tree $T(G)$, we use McCormick inequalities to linearize each non-leaf node, and obtain

$$\left\{ z \mid \max\{0, z_{t_\ell} + z_{t_r} - 1\} \leq z_t \leq \min\{z_{t_\ell}, z_{t_r}\} \text{ for each non-leaf node } t \in T(G) \right\}. \quad (\text{RMC})$$

Note that this formulation introduces auxiliary variables for non-leaf nodes that do not appear in the hyperedge set E . The next proposition shows that, when the relaxation oracle is defined by (RMC), this resulting perspective formulation is exact.

PROPOSITION 1. *If the oracle \mathcal{R} is given by (RMC) then the perspective formulation (4) is an MIP formulation of \mathcal{C}_G .*

For any fixed recursive decomposition, formulation (RMC) is straightforward to construct and implement. Consequently, we always include (RMC) in the oracle to ensure exactness. In practice, however, multi-purchase hypergraphs are often more complex than the examples shown in Figure 1 and may admit multiple valid decomposition sequences. The quality of the resulting relaxation can depend on the chosen sequence. In our implementation, we design a heuristic procedure to construct the binary tree and determine the recursive decomposition; see Section EC.2.1 for details.

3.1.2. Cycles: Odd-Cycle Inequality We consider a multi-purchase graph in which each customer purchases at most two products. This setting corresponds to the BundleMVL-2 model illustrated in Figure 1b. When a product appears in multiple bundles, the resulting graph may contain cycles. We leverage this cyclic structure to tighten the formulation presented in Proposition 1. To motivate our approach, we use the multi-purchase graph in Figure 1b as an illustration.

EXAMPLE 2. In Figure 1b, the edge set $C = \{e_1, e_2, e_3, e_4, e_5\}$ forms a cycle. Then, we select a subset D of the cycle that contains an odd number of edges, for example, $D = \{e_1, e_2, e_3\}$. Recall that for each edge $e = (i, j)$ we introduce a variable $z_e = z_i z_j$ to indicate whether bundle (i, j) is available, and in Section 3.1.1, these bilinear relations are linearized using the McCormick inequalities. In this example, we consider the following subset of McCormick inequalities:

$$\begin{aligned} -z_{(i,j)} &\leq -z_i - z_j + 1 \text{ and } -z_{(i,j)} \leq 0 && \text{for } (i, j) \in D \\ z_{(i,j)} &\leq z_i \text{ and } z_{(i,j)} \leq z_j && \text{for } (i, j) \in C \setminus D. \end{aligned}$$

Thus, for edges in the selected subset D , the variable z_e is underestimated, whereas for edges in $C \setminus D$, it is overestimated. Aggregating these inequalities and dividing both sides by 2 yields

$$z_{e_5} + z_{e_4} - z_{e_1} - z_{e_2} - z_{e_3} \leq z_5 - z_2 - z_3 + \frac{|D|}{2}.$$

Since $|D|$ is odd and z is binary, the inequality obtained by rounding the constant term $\frac{|D|}{2}$ down to the nearest integer preserves the validity of the inequality. \square

The inequality derived in Example 2 is referred to as an *odd-cycle inequality* in Padberg (1989). We now formalize this class of inequalities as follows. Let $C \subseteq E$ be a cycle and a subset $D \subseteq C$ that has an odd number of bundles. Let $S_{\text{in}} = \{u \in V \mid \exists e, f \in D, e \neq f, \text{ s. t. } e \cap f = u\}$ be the set of products connecting two distinct bundles in the subset D . Similarly, let $S_{\text{out}} = \{u \in V \mid \exists e, f \in C \setminus D, e \neq f, \text{ s. t. } e \cap f = u\}$ be the set of products connecting two distinct bundles in $C \setminus D$. The corresponding odd-cycle inequality is given as follows.

$$\sum_{(i,j) \in C \setminus D} z_{ij} - \sum_{(i,j) \in D} z_{ij} \leq \sum_{i \in S_{\text{out}}} z_i - \sum_{i \in S_{\text{in}}} z_i + \frac{|D| - 1}{2}. \quad (\text{ODD-CYCLE})$$

We incorporate inequality (ODD-CYCLE) to strengthen the relaxation oracle \mathcal{R} . This may provide additional computational benefits in solving assortment planning problems. More importantly, due to Theorem 1, a locally-sharp characterization of (ODD-CYCLE) is preserved under the perspective formulation (4). In particular, when the multi-purchase graph G is a series-parallel graph, inequalities (ODD-CYCLE) together with the McCormick constraints (RMC) provide an exact description of the convex hull of the monomial set \mathcal{M}_G .

A graph is series-parallel if it arises from a forest by repeatedly replacing edges by parallel edges or by edges in series. For example, the cycle in Figure 1b is a series-parallel graph. This restricted graph structure limits the combinatorial complexity of the bundle consideration set, thereby enabling a locally sharp formulation of \mathcal{C}_G .

PROPOSITION 2. *If the oracle \mathcal{R} is given by (RMC) and (ODD-CYCLE), then the perspective formulation (4) is an MIP formulation of \mathcal{C}_G for any multi-purchase hypergraph G . If G is a series-parallel graph, then (4) is also a locally sharp formulation of \mathcal{C}_G .*

The implementation of (ODD-CYCLE) is well-established in literature. Barahona and Mahjoub (1986) demonstrate that the separation problem can be solved in polynomial time. For a comprehensive review of implementation strategies and earlier developments, we refer readers to Rehfeldt et al. (2023) and the references therein. In our computational study, we propose a projection-based separation oracle in EC.2.1.2 based on methods in Barahona and Mahjoub (1986).

3.1.3. Chains: Running Intersection Inequality Here, we exploit chain-like dependencies in a multi-purchase hypergraph to derive a stronger formulation than Proposition 1. Formally, we say that a collection of hyperedges \tilde{E} has the running intersection property if there is an ordering $e_1 \cdots e_m$ of \tilde{E} such that for $2 \leq i \leq m$ there exists $j < i$ such that $e_i \cap (e_1 \cup \cdots \cup e_{i-1}) \subseteq e_j$. That

is, the intersection of each e_i with the union of the previous e_k 's is contained in one of these edges. For ease of exposition, let $N(e_i)$ be the intersection of e_i with previous edges, i.e.,

$$N(e_i) = e_i \cap (e_1 \cup \dots \cup e_{i-1}) \quad i = 2, \dots, m \quad (5)$$

and $N(e_1) = \emptyset$ for completeness. $N(e_i)$ also represents popular products appearing in multiple bundles. We use the multi-purchase graph in Figure 1c to illustrate how the running intersection property is leveraged to tighten recursive McCormick inequalities in Proposition 1.

EXAMPLE 3. For the hypergraph illustrated in Figure 1c, we use the hyperedge e_0 as a base bundle given its maximal cardinality. It has three neighbors $e_1, e_2,$ and e_3 , each intersecting with the root e_0 . We define an induced hyperedge set by the base e_0 , $\tilde{E} = \{e_k \cap e_0 \mid k = 1, 2, 3\}$. Note that \tilde{E} has a running intersection ordering, that is, $e_1 \cap e_0, e_2 \cap e_0$ and $e_3 \cap e_0$. To construct the strengthened formulation, for each $k \in \{1, 2, 3\}$, we define the relaxation of the multilinear variable z_{e_k} based on its neighborhood structure. If $N(e_k \cap e_0) \neq \emptyset$ then we relax z_{e_k} by $z_{e_k} \leq z_\nu$ for some $\nu \in N(e_k \cap e_0)$, and otherwise, we use the trivial bound $z_{e_k} \leq 1$. In our example, this yields the valid inequalities:

$$z_{e_1} \leq 1 \quad z_{e_2} \leq z_2 \quad \text{and} \quad z_{e_3} \leq z_3. \quad (6)$$

Furthermore, for node $6 \in e_0 \setminus \bigcup_{k=1}^3 e_k$, we have the valid bound

$$z_6 \leq 1. \quad (7)$$

Summing these inequalities with the lower bound $0 \leq z_{e_0}$, a naive aggregation results in the valid inequality $z_{e_1} + z_{e_2} + z_{e_3} + z_6 \leq 2 + z_{e_0} + z_2 + z_3$. Now, observe that the running intersection property in \tilde{E} allows us to show that if $z_0 = 0$ then one of the inequalities in (6) and (7) is not binding. Therefore, we can tighten the aggregated inequality by subtracting one from the right-hand side, resulting in $z_{e_1} + z_{e_2} + z_{e_3} + z_6 \leq 1 + z_{e_0} + z_2 + z_3$. \square

The inequality derived in Example 3 is referred to as a *running-intersection inequality* proposed by Del Pia and Khajavirad (2021). We now formalize this class of inequalities as follows. Let e_0 be a base bundle, and let $e_k, k \in K$, be a collection of neighboring bundles such that $\tilde{E} := \{e_0 \cap e_k \mid k \in K\}$ has the running intersection property. Consider a running intersection ordering of \tilde{E} , for $k \in K$ with $N(e \cap e_k) \neq \emptyset$, let $\mu_k \in N(e \cap e_k)$. A running intersection inequality is defined as:

$$\sum_{k \in K} z_{e_k} + \sum_{v \in e_0 \setminus \bigcup_{k \in K} e_k} z_v \leq z_{e_0} + \sum_{k \in K \mid N(e_0 \cap e_k) \neq \emptyset} z_{\mu_k} + \omega - 1, \quad (\text{RIC})$$

where $\omega = |e_0 \setminus \bigcup_{k \in K} e_k| + |\{k \in K \mid N(e_0 \cap e_k) = \emptyset\}|$, which counts the number of trivial constant bounds used in the derivation.

Including (RIC) in \mathcal{R} can tighten the polyhedral relaxation oracle for \mathcal{M}_G . Moreover, when the multi-purchase hypergraph has no β -cycles and neither kites, (RIC) together with (RMC) represents the convex hull of \mathcal{M}_G (Del Pia and Khajavirad 2021). Specifically, the β -cycle is a cycle $\mu_1, e_1, \mu_2, e_2, \dots, \mu_m, e_m, \mu_1$, where $m \geq 3$ and each node μ_k appears only in the adjacent hyperedges e_k and e_{k-1} (μ_1 is only in e_1 and e_m). The kite structure consists of three hyperedges $e_0, e_1, e_2 \in E$ such that $|e_0 \cap e_1 \cap e_2| \geq 2$, $(e_0 \cap e_1) \setminus e_2 \neq \emptyset$, and $(e_0 \cap e_2) \setminus e_1 \neq \emptyset$. If G does not have β -cycles and kite structures, the complexity of the multi-purchase hypergraph is limited, allowing a locally sharp oracle \mathcal{R} . For example, Figure 1c is kite-free and also β -acyclic, for which we have a locally sharp formulation. The corresponding result is formalized in the following proposition.

PROPOSITION 3. *If the oracle \mathcal{R} is given by (RMC) and (RIC), then the perspective formulation (4) is an MIP formulation of \mathcal{C}_G for any multi-purchase hypergraph G . If G is a kite-free β -acyclic hypergraph, then (4) is also a locally sharp formulation of \mathcal{C}_G .*

A computationally efficient, polynomial-time separation algorithm for the running intersection inequalities was originally developed by Del Pia et al. (2020) to solve a general class of mixed-integer nonlinear programs. We have integrated this separation oracle into our cutting-plane procedure to dynamically strengthen our formulations; the implementation details of this separation routine are provided in Appendix EC.2.1.3.

3.2. Alternative Formulations

In this section, we present an alternative class of formulations that do not rely on perspective reformulations. Using the Charnes–Cooper transformation (Charnes and Cooper 1962), the choice probability set \mathcal{C}_G can be equivalently expressed as:

$$\left\{ (\rho, \mathbf{y}) \mid \rho + \sum_{e \in E} v_e y_e = 1, \rho \geq 0, \mathbf{z} \in \mathcal{M}_G, y_e = \rho z_e \text{ for } e \in E \right\}.$$

This representation introduces bilinear terms of the form $y_e = \rho z_e$. Replacing \mathcal{M}_G by a relaxation oracle \mathcal{R} and linearizing the bilinear products using McCormick inequalities yields

$$\rho + \sum_{e \in E} v_e y_e = 1 \quad \rho \geq 0 \quad \mathbf{z} \in \mathcal{R} \quad \text{and } \mathbf{x} \in \{0, 1\}^{|V|} \quad (8a)$$

$$\max\{\rho_L z_e, \rho_U z_e + \rho - \rho_U\} \leq y_e \leq \min\{\rho_L z_e + \rho - \rho_L, \rho_U z_e\} \quad \text{for } e \in E \quad (8b)$$

where recall that $z_e = x_e$ for all singleton hyperedges $e \in V$. The inequalities in (8b) are obtained by using McCormick inequalities with a lower bound ρ_L and upper bound ρ_U on the no-purchase probability ρ . The strength of formulation (8) depends on these bounds. We therefore refer to (8) as the *Big-M* formulation. This formulation preserves the exactness of the oracle \mathcal{R} .

PROPOSITION 4. *If the relaxation oracle \mathcal{R} is exact, then the Big-M formulation (8) is an MIP formulation of \mathcal{C}_G .*

In contrast to the perspective formulation, the Big-M formulation does not, in general, preserve the local sharpness of the relaxation oracle. The following example illustrates this gap.

EXAMPLE 4. Consider a simple graph $G(V, E)$, where $V = \{1, 2\}$ and $E = \{(1), (2), (1, 2)\}$. Since G is a series-parallel graph, the choice availability set admits a locally sharp formulation by Proposition 2. We can obtain the convex hull of \mathcal{C}_G as follows.

$$\rho + v_1 y_1 + v_2 y_2 + v_{12} y_{12} = 1 \tag{9a}$$

$$0 \leq y_e \leq \rho \quad e \in \{(1), (2), (1, 2)\} \tag{9b}$$

$$y_1 + y_2 - \rho \leq y_{12} \leq \min\{y_1, y_2\}. \tag{9c}$$

Via the Fourier-Motzkin elimination, we project the linear relaxation of the Big-M formulation (8) to the choice probability space \mathcal{C}_G and obtain:

$$(9a), (9b)$$

$$y_1 + y_2 + \rho - 2 \leq y_{12} \leq y_i + 1 - \rho \quad i \in \{1, 2\}.$$

Suppose $v_1 = v_2 = v_{12} = 1$. The above polyhedron contains the node $(\rho, y_1, y_2, y_{12}) = (0.5, 0, 0, 0.5)$, which lies outside the convex hull of \mathcal{C}_G defined in (9). \square

Note that, relative to the perspective formulation, the Big-M formulation introduces additional variables z . An advantage of explicitly modeling z is that it enables us to use conic inequalities that further strengthen formulation (8).

REMARK 1. Following Sen et al. (2018), we augment (8) with the following second-order cone constraints:

$$\rho \left(1 + \sum_{e \in E} v_e z_e\right) \geq 1, \tag{10a}$$

$$y_e \left(1 + \sum_{e \in E} v_e z_e\right) \geq z_e^2, \quad \forall e \in E. \tag{10b}$$

The first inequality follows from the definition of the no-purchase variable ρ . The second inequality leverages the relationship $y_e \geq \rho z_e$ together with the binary identity $z_e = z_e^2$. \square

Finally, we compare the strength of the two classes of formulations. A key measure of the quality of an MIP formulation is the tightness of its natural continuous relaxation, obtained by dropping the integrality constraints. This criterion is particularly relevant in practice, since tighter relaxations typically lead to stronger bounds and faster convergence of the branch-and-bound algorithms underlying modern commercial MIP solvers. Because the two formulations are expressed in different variable spaces, we compare them by projecting their continuous relaxations onto the space of choice probability variables. Specifically, let $\mathcal{R}_{\text{pers}}$ and $\mathcal{R}_{\text{Big-M}}$ denote the projections of the natural relaxations of (4) and (8), respectively, onto the space of (ρ, \mathbf{y}) variable.

THEOREM 2. *For any relaxation oracle \mathcal{R} , $\mathcal{R}_{\text{pers}} \subseteq \mathcal{R}_{\text{Big-M}}$.*

4. Estimation

In this section, we show an empirical study of estimating (LOGIT-MP). In particular, we provide a heuristic to construct the consideration set E from data.

4.1. Data and Multi-Purchase Hypergraph Estimation

We use a daily transaction dataset from a bakery (Gimbert 2025). The data for each transaction includes a timestamp, the set of purchased products, and the total price, characterizing the daily sales of different bundles. The data could be summarized as $\{S_t, (n_{te})_{e \in H}\}_{t \in [T]}$, where S_t is the assortment on day t and n_{te} denotes the sale of bundle e on day t . H is the collection of historical bundles.

We focus on transaction data from July 2021, a month exhibiting a relatively high proportion of multi-item purchases. To mitigate potential estimation bias arising from stockouts, we restrict our attention to morning transactions. After preprocessing, the final dataset comprises 5,614 transactions. To alleviate data sparsity, we follow Akchen and Mitrofanov (2025) by retaining only the top 50% of products by sales and aggregating the remainder into the outside option. This process yields a set of 27 products across two categories. The bundle size distribution reveals that 50% of transactions involve a single product, 27.6% involve two, 15.4% involve three, and 7% involve more than three. Among the multi-purchase transactions, 51.48% consist of products from the same category, while 48.52% span both categories. This near-even split highlights the necessity of

a general multi-purchase model, i.e., the hypergraph representation, capable of capturing intra- and cross-category interactions.

Our estimation procedure for the multi-purchase hypergraph consists of three main stages: defining candidate hypergraphs, applying MLE to estimate edge weights (i.e., the bundle utilities \mathbf{u}), and selecting the best-fitting hypergraph based on predictive performance evaluated via cross-validation. Formally, we define the node set V as the collection of the 27 individual products. To construct the edge set E , we employ a heuristic governed by two parameters: the maximum bundle size d and a sparsity threshold θ . Specifically, let E_d denote the set of all possible bundles of size at most d . We then form the refined edge set $E_{d,\theta}$ by retaining all single-item options alongside the multi-item bundles whose empirical purchase frequencies rank in the top θ fraction of $E_d \setminus V$. By varying d and θ , we can generate a sequence of candidate hypergraphs $G(V, E_{d,\theta})$. This heuristic construction is motivated by the data observation that customers' purchased bundles are highly concentrated on a few specific combinations with a size of less than four. By filtering the edge set in this manner, we can limit the model complexity while preserving most information in the data.

Given a hypergraph structure $G(V, E_{d,\theta})$, the second step, estimation of utility, has been well studied in the literature. Studies often assume that the bundle utility consists of single-item and pairwise interaction effects (Russell and Petersen 2000, Aurier and Mejia 2014), i.e.,

$$u_e = \sum_{i \in e} \alpha_i + \sum_{i < j, i, j \in e} \beta_{ij} \quad (11)$$

In high-dimensional settings, regularization is used to induce sparsity (Benson et al. 2018, Vasilyev et al. 2025). Under these structure assumptions, parameter estimation via MLE is computationally tractable, leading to widespread empirical applications (e.g. Hruschka et al. 1999, Boztuğ and Hildebrandt 2008, Richards et al. 2018, etc). We follow the linear utility specification (11) and further assume that $\alpha_i = \sum_{j=1,2} \eta_j h_{ij} + \eta_r r_i$, where h_{ij} denote whether product i is in category j . η_j is the fixed effect of category j and η_r is the price elasticity.

We then employ five-fold cross-validation to evaluate the predictive performance of (LOGIT-MP) under different candidate hypergraphs. Let \mathbf{p}_t denote observed market share and $\hat{\mathbf{p}}_t$ the corresponding model predictions. Out-of-sample performance is evaluated using two standard measures of distributional fit: $\chi^2 = \sum_{t \in T} \sum_{e \in E_{d,\theta}} (p_{t,e} - \hat{p}_{t,e})^2 / p_{t,e}$ and mean squared error (MSE), $\sum_{t \in T} \sum_{e \in E_{d,\theta}} (p_{t,e} - \hat{p}_{t,e})^2$, where T is the set of days in the test dataset.

4.2. Estimation Results

Table 1 summarizes the estimation results. The first two columns correspond to the maximum bundle size and the sparsity level, respectively. When $d = 1$, (LOGIT-MP) reduces to the standard MNL choice model. Two “Improve” columns following χ^2 and MSE statistics present the relative performance gains of (LOGIT-MP) over the MNL benchmark. Overall, (LOGIT-MP) outperforms the MNL model in most specifications. Notably, when the maximum bundle size is restricted to 2, and the sparsity level is set to $\theta = 0.1$, it improves out-of-sample prediction by more than 10% in terms of the χ^2 metric. Increasing the sparsity level (i.e., considering a denser hypergraph) does not necessarily reduce prediction loss and can even deteriorate performance. This finding suggests that a sparse multi-purchase structure is sufficient to capture the main cross-item interactions in practice, thereby simplifying the optimization problem without compromising predictive accuracy.

Table 1 Out-of-sample performance

d	θ	χ^2	Improve(%)	MSE	Improve(%)
1	-	5.4898	-	0.0098	-
2	0.1	4.8560	11.54	0.0095	3.06
2	0.2	4.9600	9.65	0.0096	2.04
2	0.3	5.1468	6.25	0.0101	-3.06
3	0.1	4.8560	11.54	0.0095	3.06
3	0.2	4.9600	9.65	0.0096	2.04
3	0.3	5.1468	6.25	0.0101	-3.06

5. Assortment Optimization under the Logit-MP Model

In this section, we use the perspective formulation of the choice probability set \mathcal{C}_G , introduced in Section 3, to globally solve the assortment planning problem under the LOGIT-MP model, and investigate its computational performance.

5.1. Problem Definition and Formulation

In an assortment planning setting, a firm has access to a finite set of products, among which it needs to select a subset to offer to its customers. Customers decide either to purchase a bundle of the products or leave without purchasing according to the stochastic choice model given in (LOGIT-MP). The goal of the firm is to find a set of products to offer so as to maximize the expected revenue. Formally, consider a hypergraph $G = (V, E)$, where the node set V represents the set of potential products that can be offered, and the hyperedge set E represents the bundles that customers are willing to consider (i.e., the consideration set). The firm solves

$$\max_S \left\{ \sum_{e \in E} r_e P_E(e, S) \mid S \in \mathcal{X} \right\}, \quad (\text{AOPT})$$

where r_e denotes the revenue obtained from selling bundle e , $P_E(e, S)$ is the probability that bundle e is chosen when assortment S is offered, as defined in (LOGIT-MP), and \mathcal{X} is a set of subsets of V modeling operational constraints. When E contains only singleton bundles, problem (AOPT) reduces to the constrained assortment optimization problem under the MNL model. For unconstrained instances, the optimal solution is known to follow a revenue-ordered structure (Talluri and Van Ryzin 2004). Furthermore, Sumida et al. (2021) establish an LP formulation for cases characterized by total unimodularity. This LP approach has been generalized in Chen et al. (2026) to accommodate any constraint set provided its convex hull admits a tractable LP representation.

However, the assortment optimization problem becomes significantly more challenging when accounting for multi-purchase behavior. In this setting, the classical revenue-ordered property no longer holds. Moreover, for a general consideration set E , the problem is not only intractable but NP-hard to approximate, as established by the following complexity result.

THEOREM 3. *The assortment optimization problem (AOPT) is NP-hard to approximate within factor $O(|V|^{1-\epsilon})$ for any fixed $\epsilon \in (0, 1)$ even when this is no constraint.*

Consequently, recent literature has primarily focused on designing approximation algorithms or heuristics tailored to a specific, highly structured consideration set E . For example, under the MVMNL model, Chen et al. (2022) design a constant-approximation algorithm limited to a two-category setting, while Jasin et al. (2024) propose an FPTAS for general categories. Similarly, for the BundleMVL-2 model, Tulabandhula et al. (2023) propose a heuristic algorithm.

Next, we propose an exact MIP formulation to obtain the globally optimal solution for (AOPT) under an arbitrary consideration set. Assume that the constraint set \mathcal{X} can be described using a system of linear inequalities and binary constraints $\mathbf{x} \in \{0, 1\}^V$. By leveraging our perspective formulation (4) to model the set of all possible choice probabilities under (LOGIT-MP), we obtain the following MIP:

$$\max_{\rho, \mathbf{x}, \mathbf{y}} \left\{ \sum_{e \in E} r_e y_e v_e \mid (\rho, \mathbf{y}, \mathbf{x}) \text{ satisfies (4), } \mathbf{x} \in \mathcal{X} \right\}. \quad (\text{PERS})$$

The choice of the relaxation oracle \mathcal{R} in (4) directly affects the quality of the resulting formulation. Computationally, a stronger relaxation oracle \mathcal{R} leads to a tighter formulation and typically improves branch-and-bound performance, as shown in Section 5.2. Theoretically, when the oracle is locally sharp, the LP relaxation (PERS), obtained by dropping the binary requirement on \mathbf{x} , suffices to solve the assortment optimization problem (AOPT). As a consequence of Propositions 1, 2, and 3, we obtain the following characterization.

COROLLARY 1. *The formulation (PERS) exactly solves the assortment optimization problem (AOPT) when the relaxation oracle is given by (RMC). The LP relaxation of (PERS) solves (AOPT) when the problem is unconstrained, that is, $\mathcal{X} = \{0, 1\}^{|V|}$, the multi-purchase hypergraph G is a series-parallel (resp. kite-free β -acyclic) graph, and the relaxation oracle is derived by (RMC) and (ODD-CYCLE) (resp. (RIC)).*

REMARK 2. We relate our result to existing LP characterizations in the literature. Lo and Topaloglu (2019) investigate an assortment optimization under product synergies effects where a product’s attractiveness depends on the offering of other products. By modeling these interactions using a graph, where nodes represent products and edges denote pairwise synergy effects, they show that the problem admits an exact LP formulation when the underlying graph is a path. Our result generalizes this to a series-parallel graph, and even to a kite-free β -acyclic hypergraph. \square

5.2. Computational Performance

In this subsection, we conduct a computational study to evaluate the relative performance of three alternative formulations for (AOPT): the perspective formulation (4), the Big-M formulation (8), and the conic relaxation defined by (8) and (10). To assess the impact of different polyhedral relaxations, we test three variants of the oracle \mathcal{R} , defined as follows:

- \mathcal{R}_{RMC} : We use recursive McCormick inequalities (RMC) to construct \mathcal{R} .
- $\mathcal{R}_{\text{RMC\&ODD}}$: We add odd-cycle inequalities (ODD-CYCLE) to \mathcal{R}_{RMC} .
- $\mathcal{R}_{\text{RMC\&RIC}}$: We add running-intersection inequalities (RIC) to \mathcal{R}_{RMC} .

All numerical tests were implemented in the Julia (Bezanson et al. 2017), using modeling language JuMP (Lubin et al. 2023), on a machine with an Intel(R) Core(TM) Ultra 7 265K CPU at 3.90 GHz and 48 GB RAM. All models were solved using Gurobi 12.0.3 (Gurobi Optimization, LLC 2025).

The numerical implementation of the three proposed formulations presents two primary computational challenges. First, the perspective formulation (4) involves an exponentially large number of inequalities, specifically, $\mathcal{O}(|V|2^{|\mathbb{E} \setminus V|})$. Second, we need an efficient identification of odd-cycle or running-intersection structures in the multi-purchase hypergraph G to generate valid inequalities. To ensure computational tractability, we develop an iterative cutting-plane algorithm that avoids explicit enumeration of the full constraint set. The procedure is structured as follows:

- Solve a relaxed master formulation with a subset of constraints.
- Invoke two dedicated separation oracles to identify violated inequalities from constraints (4d) and (4e) and the relaxation oracle \mathcal{R} . These constraints can tighten the formulation.

- Once no further violations are detected, we solve the tightened formulation with binary constraints to obtain an optimal integer solution.

Comprehensive details regarding the cutting-plane implementation are provided in [EC.2](#).

We generate synthetic instances as follows. To construct the multi-purchase hypergraph, we randomly separate N products into three categories. We follow the case study observation that most customers purchase no more than three products in a single transaction, and set the maximum bundle size as $d \in \{2, 3\}$. We set the sparsity level at 0.25. Given a rank d , we generate two types of bundles: $\pi\%$ cross-category bundles and $(1 - \pi)\%$ intra-category bundles. For example, when the number of products $N = 10$, rank $d = 2$, sparsity level $\theta = 2/9$, and a consideration rule f characterized by BundleMVL-2. If $\pi = 0.5$, we have 5 cross-category bundles and 5 intra-category bundles. To assign attraction value v for each bundle, we first draw a latent value α_i from a uniform distribution $\mathcal{U}(-1.5, 0.5)$ for each $i \in V$. Then, the attraction value of a bundle $e \in E$ is given by $v_e = \exp\left(\sum_{i \in e} \alpha_i + \sum_{i < j, i, j \in e} \beta_{ij}\right)$, where $\beta_{ij} \sim \mathcal{U}(-1.25, 0.75)$. We set the range of these uniform distributions based on estimators in the case study. We normalize the no-purchase option attraction value to 1. For each product $i \in V$, the revenue is sampled independently from $r_i \sim \mathcal{U}(1, 4)$. The revenue of a bundle $e \in E$ is additive, i.e., $r_e = \sum_{i \in e} r_i$. Finally, we define the operational constraint as $\mathcal{X} = \{\mathbf{x} \in \{0, 1\}^{|V|} \mid \sum_{i \in V} x_i \leq 0.2|V|\}$.

For each parameter configuration, we generate 10 random instances following the aforementioned procedure. All experiments are conducted with a time limit of 3,600 seconds and a relative optimality tolerance of 0.05%. We evaluate the three models based on their ability to identify optimal integer solutions and the strength of their LP relaxations. Specifically, we report the following performance metrics:

- #sol: the number of instances solved to global optimality within the time limit
- Time(s): the average computation time in seconds. If a separation procedure is involved, its computational time is indicated in parentheses.
- Gap (%): the average relative optimality gap at termination, calculated as $\text{Gap} = 100\% \times \frac{R_U - R^*}{R^*}$, where R_U and R^* denote the revenue of the best upper bound and the best integer solution obtained at termination, respectively.
- RGap (%): the average root node gap, defined as $\text{RGap} = 100\% \times \frac{R_{\text{RLx}} - R^*}{R^*}$, where R_{RLx} denotes the revenue of the LP relaxation of the given formulation.
- #node: The average number of branch-and-bound nodes explored in the search tree.

Table 2 Computation results for instances with pure intra-category bundles (the maximum bundle size $d = 2$ and sparsity level $\theta = 0.25$)

Instance N	Model	Recursive McCormick					RMC & Odd Cycle				
		#sol	Time(s)	Gap(%)	RGap(%)	#node	#sol	Time(s)	Gap(%)	RGap(%)	#node
100	Big-M	10	2.9 (0.0)	0.0	12.6	9737	10	3.4 (0.1)	0.0	12.6	11705
	CONIC	10	7.7 (0.0)	0.0	10.8	5405	10	8.9 (0.9)	0.0	10.8	7841
	Pers	10	0.1 (0.1)	0.0	0.0	1	10	0.1 (0.1)	0.0	0.0	1
500	Big-M	2	1622.9 (0.0)	2.8	9.3	193394	3	1535.7 (4.0)	2.4	9.3	164121
	CONIC	0	- (0.0)	7.9	9.0	10063	0	-(3172.1)	8.5	9.9	13645
	Pers	10	10.1 (5.2)	0.0	0.0	1	10	10.0 (5.4)	0.0	0.0	1
1000	Big-M	0	-(0.0)	9.0	11.0	103476	0	-(17.7)	8.8	11.0	105271
	CONIC	0	\	\	\	\	0	\	\	\	\
	Pers	10	324.6 (60.5)	0.0	0.0	1	10	329.5 (63.1)	0.0	0.0	1

“-” in the Time(s) column indicates that the corresponding model fails to solve the MIP within 3600 seconds for all instances. “\” for all metrics indicates that the corresponding model fails to solve the linear relaxation within 3600 seconds for all instances.

Table 3 Computation results for instances with pure cross-category bundles (the maximum bundle size $d = 2$ and sparsity level $\theta = 0.25$)

Instance N	Model	Recursive McCormick					RMC & Odd Cycle				
		#sol	Time(s)	Gap(%)	RGap(%)	#node	#sol	Time(s)	Gap(%)	RGap(%)	#node
100	Big-M	10	33.1 (0.0)	0.0	11.0	34228	10	28.9 (0.1)	0.0	11.0	30762
	CONIC	10	238.5 (0.0)	0.0	10.0	50232	10	216.3 (147.4)	0.0	10.0	42329
	Pers	10	0.2 (0.2)	0.0	0.0	1	10	0.2 (0.2)	0.0	0.0	1
500	Big-M	0	-(0.0)	3.3	6.9	106343	0	-(16.8)	3.4	6.9	107076
	CONIC	0	\	\	\	\	0	\	\	\	\
	Pers	10	83.4 (22.0)	0.0	0.2	1	10	83.3 (22.4)	0.0	0.2	1
1000	Big-M	0	-(0.0)	5.1	6.6	32524	0	-(29.1)	5.1	6.6	32481
	CONIC	0	\	\	\	\	0	\	\	\	\
	Pers	10	724.1 (269.6)	0.0	0.0	1	10	725.4 (269.7)	0.0	0.0	1

“-” in the Time(s) column indicates that the corresponding model fails to solve the MIP within 3600 seconds for all instances. “\” for all metrics indicates that the corresponding model fails to solve the linear relaxation within 3600 seconds for all instances.

When the maximum bundle size is two, we construct the oracle \mathcal{R} using recursive McCormick and odd-cycle inequalities. Table 2 and 3 report computational results for instances with pure intra-category and cross-category bundles, respectively. Both tables show that the perspective formulation dominates across all instances. For small instances ($N = 100$), all three models reach optimality, but our formulation is substantially faster, requiring less than 0.2 seconds, with near-zero root gaps and a single search node. The advantage grows sharply with N : for $N = 500$ and 1000, the perspective formulation solves all instances, while the Big-M and conic formulations solve only several instances with large remaining gaps. Moreover, in all instances, the perspective

formulation can be solved to optimality at the root node. It suggests that the perspective formulation, combined with the recursive McCormick inequalities, is effective in handling the coupling between fractional and bilinear terms when the maximum bundle size is two. This observation is consistent with the limited impact of odd-cycle inequalities, which are infrequently generated and yield negligible improvements in the LP relaxation.

For a maximum bundle size of three, we represent \mathcal{R} using recursive McCormick and running-intersection inequalities. Tables 4 and 5 report computational results for instances with pure intra-category and cross-category bundles, respectively. The perspective formulation again uniformly dominates. In contrast, the Big-M formulation fails to solve any instance to optimality within the 3600-second time limit, while the conic formulation fails to solve the root node relaxation. A key additional insight from both tables is the impact of running-intersection inequalities. Incorporating these inequalities yields substantial computational gains, significantly increasing the number of instances solved within the time limit. In particular, for cross-category bundles, the perspective formulation equipped only with recursive McCormick relaxations fails to solve several instances within 3600 seconds. Augmenting the formulation with running-intersection inequalities resolves all instances and delivers speedups of up to a factor of five, even when accounting for the time required by the cutting-plane procedure.

Overall, these results demonstrate that for $d = 3$, effectively capturing higher-order interactions requires both the perspective reformulation and strong polyhedral strengthening via running-intersection inequalities.

Table 4 Computation results for instances with pure intra-category bundles (the maximum bundle size $d = 3$ and sparsity level $\theta = 0.25$)

Instance N	Model	Recursive McCormick					RMC & Running-Intersection				
		#sol	Time(s)	Gap(%)	RGap(%)	#node	#sol	Time(s)	Gap(%)	RGap(%)	#node
180	Big-M	0	- (0.0)	4.6	12.4	161636	0	- (22.4)	4.6	12.4	159971
	CONIC	0	\	\	\	\	0	\	\	\	\
	Pers	10	100.9 (5.5)	0.0	1.9	2708	10	16.9 (41.7)	0.0	0.5	78
220	Big-M	0	- (0.0)	6.3	12.6	80805	0	- (11.3)	6.3	12.6	79613
	CONIC	0	\	\	\	\	0	\	\	\	\
	Pers	10	471.4 (16.3)	0.0	1.9	5619	10	58.4 (116.9)	0.0	0.3	86
260	Big-M	0	- (0.0)	4.8	9.9	37087	0	- (14.4)	4.9	9.9	38646
	CONIC	0	\	\	\	\	0	\	\	\	\
	Pers	9	914.6 (31.3)	0.6	1.9	4933	10	308.3 (266.7)	0.0	0.4	629

“-” in the Time(s) column indicates that the corresponding model fails to solve the MIP within 3600 seconds for all instances. “\” for all metrics indicates that the corresponding model fails to solve the linear relaxation within 3600 seconds for all instances.

Table 5 Computation results for instances with pure cross-category bundles (the maximum bundle size $d = 3$ and sparsity level $\theta = 0.25$)

Instance N	Model	Recursive McCormick					RMC & Running-Intersection				
		#sol	Time(s)	Gap(%)	RGap(%)	#node	#sol	Time(s)	Gap(%)	RGap(%)	#node
100	Big-M	0	- (0.0)	3.9	12.1	181968	0	- (24.0)	4.0	12.1	182549
	CONIC	0	\	\	\	\	0	\	\	\	\
	Pers	9	348.2 (4.8)	1.1	2.7	24684	10	61.2 (67.4)	0.0	0.4	1428
120	Big-M	0	- (0.0)	4.4	10.6	87927	0	- (12.4)	4.3	10.6	89345
	CONIC	0	\	\	\	\	0	\	\	\	\
	Pers	8	671.5 (10.6)	1.3	2.5	20235	10	165.1 (197.1)	0.0	0.4	2928
140	Big-M	0	- (0.0)	5.0	10.0	45415	0	- (14.4)	5.0	10.0	45499
	CONIC	0	\	\	\	\	0	\	\	\	\
	Pers	8	1945.6 (22.1)	1.0	2.5	21319	10	89.3 (296.5)	0.0	0.5	225

“-” in the Time(s) column indicates that the corresponding model fails to solve the MIP within 3600 seconds for all instances. “\” for all metrics indicates that the corresponding model fails to solve the linear relaxation within 3600 seconds for all instances.

Table 6 examines the performance of the perspective formulation as a function of the sparsity level and the proportion of cross-category bundles (π) for instances with $N = 100$. Two variants are compared: the baseline using recursive McCormick (RMC) inequalities and the strengthened version incorporating running-intersection (RIC) inequalities.

Several patterns emerge. First, problem difficulty increases with both sparsity and the cross-category ratio. For a fixed sparsity level, increasing π consistently leads to larger root gaps and longer solution times under the RMC relaxation. Similarly, for a fixed π , denser instances (higher sparsity) are computationally more challenging. Second, adding running-intersection inequalities significantly strengthens the formulation. Across all parameter settings, the RMC&RIC variant yields substantially smaller root gaps compared to RMC alone. This improvement is particularly pronounced at higher values of π , where cross-category interactions are more prevalent. Moreover, the stronger relaxation translates into meaningful computational gains. Although the separation time increases when RIC inequalities are included, the overall solution time is typically reduced, often by a large margin. In particular, for instances with high cross-category ratios, the RMC&RIC formulation achieves notable speedups despite the additional separation overhead.

Overall, the results highlight that cross-category interactions are a primary source of difficulty, and that running-intersection inequalities are highly effective in mitigating this challenge by tightening the relaxation and improving solution times.

Table 6 Performance of perspective formulation under different sparsity levels and cross-category ratio ($N = 100$)

Parameters		Recursive McCormick			RMC & Running-Intersection		
Sparsity	Cross(π)	RGap(%)	SolTime(s)	SepTime(s)	RGap(%)	SolTime(s)	SepTime(s)
0.15	0.2	1.66	3.42	0.83	1.01	2.23	1.60
	0.4	2.38	25.24	1.52	1.70	11.18	3.51
	0.6	2.36	45.77	1.58	1.53	37.79	6.88
	0.8	2.76	112.08	2.09	1.38	50.77	16.39
0.2	0.2	3.23	126.55	1.21	2.18	47.07	5.06
	0.4	2.83	198.90	1.81	1.72	36.88	6.76
	0.6	3.00	141.11	2.71	1.59	72.59	19.26
	0.8	2.34	229.08	3.17	0.99	77.82	25.72
0.25	0.2	2.88	29.96	1.37	1.48	22.74	6.47
	0.4	2.87	85.63	2.56	1.64	43.89	11.23
	0.6	2.71	181.79	3.79	1.16	77.77	30.26
	0.8	2.55	403.55	3.60	0.76	35.67	42.75
0.3	0.2	3.07	67.85	1.81	1.88	39.38	7.09
	0.4	2.92	189.36	3.37	1.45	72.40	19.97
	0.6	3.14	315.37	3.84	1.30	100.67	35.54
	0.8	2.57	186.29	5.57	0.57	35.42	53.85

6. Extension to Heterogeneous Customers: The Mixed Logit-MP Model

Customer heterogeneity is a central consideration in choice modeling, as numerous studies document significant variation in preferences across individuals (McFadden and Train 2000). A standard approach to capturing such heterogeneity is to use a mixture of choice models. For instance, the widely used mixed MNL model assumes that the customer population consists of multiple segments, each characterized by its own set of MNL parameters. Extending this idea to the multi-purchase setting naturally yields a mixture of Logit-MP models, in which each segment follows (LOGIT-MP) with segment-specific parameters.

In this section, we consider the assortment optimization problem under the mixture of Logit-MP models with a known customer type distribution, as well as its robust counterpart, where the customer type distribution is assumed to lie within a prescribed uncertainty set. Leveraging the perspective formulations developed in Theorem 1, we derive MIP formulations for both problems and evaluate their computational performance.

6.1. Model Extensions

We assume that the market consists of K types of customers. Each type of customer is characterized by a hypergraph $G^k = (V, E^k)$ and a vector of attraction values v^k . Given an assortment

$S \subseteq V$, the type k customer selects bundle e with probability:

$$P^k(e, S) = \frac{\exp(u_e^k) \mathbb{1}(e \subseteq S)}{1 + \sum_{e \in 2^S \cap E^k} \exp(u_e^k)}.$$

For each $k \in [K]$, let λ_k denote the proportion of the type k customers. The vector $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_K)$ represents the customer type distribution and must lie in the standard simplex $\Delta^K := \{\boldsymbol{\lambda} \mid \sum_{k \in [K]} \lambda_k = 1, \lambda_k \geq 0 \text{ for } k \in [K]\}$. Under this heterogeneous model, the firm seeks to solve the following assortment optimization problem:

$$\max_S \left\{ \sum_{k \in [K]} \sum_{e \in E} r_e \lambda_k P^k(e, S) \mid S \in \mathcal{X} \right\}. \quad (\text{AOPT-MIX})$$

Incorporating customer heterogeneity introduces significant challenges for assortment optimization. First, it substantially increases computational complexity. Even in the single-purchase setting, the assortment optimization problem under a mixture of two MNL models is already NP-hard (Rusmevichientong et al. 2014). Moreover, it is not clear whether existing approximation algorithms developed for the homogeneous case (Chen et al. 2022, Jasin et al. 2024) can be extended to the heterogeneous setting.

Second, heterogeneity raises robustness concerns when the type distribution is estimated from data and subject to misspecification. In practice, segment proportions are typically inferred from limited data and may be misspecified, which can significantly affect optimal assortment decisions (Bertsimas and Mišić 2017). Therefore, it is important to develop optimization models that account for heterogeneity while remaining robust to distributional ambiguity. Following Bertsimas and Mišić (2017), we consider a polyhedral uncertainty set \mathcal{L} for the type distribution, that is,

$$\mathcal{L} := \{\boldsymbol{\lambda} \mid \mathbf{B}\boldsymbol{\lambda} \geq \mathbf{d}\},$$

where $\mathbf{B} \in \mathbb{R}^{m \times k}$ and $\mathbf{d} \in \mathbb{R}^m$. The firm therefore solves the following robust counterpart of (AOPT-MIX)

$$\max_S \min_{\boldsymbol{\lambda} \in \mathcal{L}} \left\{ \sum_{k \in [K]} \sum_{e \in E} r_e \lambda_k P^k(e, S) \mid S \in \mathcal{X} \right\}. \quad (\text{AOPT-ROBUST})$$

The perspective formulation developed in Theorem 1 enables us to globally solve both problems. For a given hypergraph G and a vector of attraction values \mathbf{v} , let $\mathcal{S}(G, \mathbf{v})$ denote the perspective formulation defined as in (4). Using this, we obtain the following MIP formulation for (AOPT-MIX):

$$\max_{\boldsymbol{\rho}, \mathbf{x}, \mathbf{y}} \left\{ \sum_{k \in [K]} \sum_{e \in E^k} r_e \lambda_k v_e^k y_e^k \mid \mathbf{x} \in \mathcal{X}, (\boldsymbol{\rho}^k, \mathbf{y}^k, \mathbf{x}) \in \mathcal{S}(G^k, \mathbf{v}^k) \text{ for } k \in [K] \right\}. \quad (\text{PERS-MIX})$$

Similarly, an MIP formulation for the robust problem (**AOPT-ROBUST**) is given by:

$$\max_{\mathbf{x}, \boldsymbol{\rho}, \mathbf{y}, \mathbf{z}} \left\{ \mathbf{d}^\top \mathbf{z} \mid \begin{array}{l} \mathbf{z}^\top B_k = \sum_{e \in E^k} r_e v_e^k y_e^k, (\boldsymbol{\rho}^k, \mathbf{y}^k, \mathbf{x}) \in \mathcal{S}(G^k, \mathbf{v}^k) \text{ for } k \in [K] \\ \mathbf{z} \in \mathbb{R}_+^m, \mathbf{x} \in \mathcal{X} \end{array} \right\}. \quad (\text{PERS-ROBUST})$$

COROLLARY 2. *Formulations (**PERS-MIX**) and (**PERS-ROBUST**) solve problems (**AOPT-MIX**) and (**AOPT-ROBUST**), respectively.*

6.2. Computational Performance

The robust formulation is the most general and computationally demanding variant considered in this paper. We therefore evaluate the computational performance of the perspective formulation for the robust mixed problem (**PERS-ROBUST**) using synthetic instances. Specifically, we consider $K \in \{2, 4\}$ customer types and follow the parameter generation procedure described in Section 5.2. When generating hyperedges, we create a hyperedge e if $0.4p_e^k + 0.6p_e^0 \leq \theta$, which ensures that similarities across different customer types arise from the products' intrinsic attributes. Here, $p_e^k \sim \mathcal{U}(0, 1)$ for $k \in \{0, 1, \dots, K\}$ and $e \in E_d$. The uncertainty set \mathcal{L} is constructed as follows:

$$\mathcal{L} = \left\{ \boldsymbol{\lambda} \mid \lambda_k \in [L_k, U_k] \text{ for } k \in [K], \sum_{k \in [K]} |\lambda_k - \hat{\lambda}_k| \leq \Gamma, \boldsymbol{\lambda} \in \Delta^K \right\}.$$

Each element of $\hat{\boldsymbol{\lambda}}$ is independently drawn from the uniform distribution $\mathcal{U}(0, 1)$, after which the vector is normalized so that it lies in the probability simplex. The bounds are defined as $L_k = 0.95 \hat{\lambda}_k$ and $U_k = 1.05 \hat{\lambda}_k$, and the budget parameter is set to $\Gamma = 0.1$. Similarly, we define the operational constraint as $\mathcal{X} = \{ \mathbf{x} \in \{0, 1\}^{|V|} \mid \sum_{i \in V} x_i \leq 0.2|V| \}$.

We vary the number of products, with $N \in \{70, 80, 90\}$ for pure intra-category bundles and $N \in \{40, 50, 60\}$ for cross-category bundles. The computational results are reported in Tables 7 and 8, respectively. As the number of customer types increases, all performance metrics deteriorate. Specifically, solution time, separation time, root gaps, and the number of branch-and-bound nodes grow substantially, reflecting the added complexity of the robust model. Nevertheless, most instances are solved to optimality within the time limit. This observation is consistent with earlier findings — running-intersection inequalities play a critical role, reducing the number of explored nodes and improving overall efficiency.

Table 7 Computation results for pure intra-category bundles with size less than three (sparsity level $\theta = 0.4$)

Instance		Recursive McCormick					RMC& Running-Intersection				
N	K	#sol	Time(s)	Gap(%)	RGap(%)	#node	#sol	Time(s)	Gap(%)	RGap(%)	#node
70	2	10	21.2 (2.1)	0.0	3.3	10401	10	17.5 (4.0)	0.0	2.2	9568
	4	10	172.0 (6.3)	0.0	4.8	12602	10	138.4 (11.9)	0.0	3.6	9667
80	2	10	22.3 (3.6)	0.0	2.9	4278	10	15.5 (6.1)	0.0	2.0	2916
	4	8	478.8 (11.6)	2.0	5.7	40781	8	370.7 (23.0)	1.6	4.0	34476
90	2	10	43.4 (6.1)	0.0	2.4	10078	10	28.8 (10.5)	0.0	1.4	4106
	4	6	845.8 (21.4)	2.6	6.4	36458	6	559.5 (39.0)	1.8	4.9	23497

Table 8 Computation results for pure cross-category bundles with size less than three (sparsity level $\theta = 0.4$)

Instance		Recursive McCormick					RMC& Running-Intersection				
N	K	#sol	Time(s)	Gap(%)	RGap(%)	#node	#sol	Time(s)	Gap(%)	RGap(%)	#node
40	2	10	8.6 (1.2)	0.0	3.3	4397	10	6.5 (2.7)	0.0	2.2	3385
	4	10	104.6 (3.9)	0.0	5.0	12023	10	61.0 (7.8)	0.0	3.6	8180
50	2	10	36.4 (4.2)	0.0	3.5	6779	10	18.6 (7.9)	0.0	2.2	3510
	4	10	423.8 (13.1)	0.0	5.9	15817	10	333.0 (26.0)	0.0	4.5	12050
60	2	10	222.0 (13.0)	0.0	3.7	16710	10	120.8 (23.9)	0.0	2.1	8818
	4	8	1903.4 (43.2)	1.0	5.8	27345	10	1812.4 (76.3)	0.0	4.2	27289

7. Conclusion

This paper develops exact MIP formulations for logit-based multi-purchase choice models, which are empirically well supported but challenging to incorporate into optimization frameworks due to product interactions and fractional choice probabilities. To address these challenges, we propose a unified hypergraph representation and derive strong formulations by integrating convexification techniques from multilinear optimization and fractional programming. Our formulations preserve the strength of the underlying polyhedral structure, yielding significantly tighter LP relaxations and substantial computational improvements over standard approaches.

Several directions for future research remain. First, the proposed exact optimization framework can be naturally extended to other complex decision problems sharing similar combinatorial features, such as optimal bundle design and joint pricing under multi-purchase behavior. Second, transitioning from optimization to empirical estimation, we currently adopt a heuristic method to construct the consideration set (i.e., the multi-purchase hypergraph). Designing scalable estimation procedures equipped with statistical guarantees is a non-trivial challenge in discrete choice studies, which needs a dedicated investigation.

References

- Abdallah T, Braverman A, Gu W (2024) Multi-purchase assortment optimization under a general random utility model. Available at SSRN 4842012 .
- Akçakuş İ, Mišić VV (2025) Exact logit-based product design. *Management Science*, Forthcoming .
- Akchen YC, Mišić VV (2026) Assortment optimization under the decision forest model. *Manufacturing & Service Operations Management*, Forthcoming .
- Akchen YC, Mitrofanov D (2025) Consider or choose? The role and power of consideration sets. *Management Sci.* ePub ahead of print, October 31, <http://dx.doi.org/10.1287/mnsc.2023.00442>.
- Atamtürk A, Gómez A (2020) Submodularity in conic quadratic mixed 0–1 optimization. *Oper. Res.* 68(2):609–630.
- Aurier P, Mejia V (2014) Multivariate logit and probit models for simultaneous purchases: Presentation, uses, appeal and limitations. *Recherche et Applications en Marketing (English Edition)* 29(2):75–94.
- Bai Y, Feldman J, Segev D, Topaloglu H, Wagner L (2024) Assortment optimization under the multi-purchase multinomial logit choice model. *Oper. Res.* 72(6):2631–2664.
- Barahona F, Mahjoub AR (1986) On the cut polytope. *Math. Programming* 36(2):157–173.
- Benson AR, Kumar R, Tomkins A (2018) A discrete choice model for subset selection. *Proceedings of the eleventh ACM international conference on web search and data mining* 37–45.
- Bertsimas D, Mišić VV (2017) Robust product line design. *Oper. Res.* 65(1):19–37.
- Bertsimas D, Mišić VV (2019) Exact first-choice product line optimization. *Oper. Res.* 67(3):651–670.
- Bezanson J, Edelman A, Karpinski S, Shah VB (2017) Julia: A fresh approach to numerical computing. *SIAM Rev.* 59(1):65–98.
- 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.
- Boztuğ Y, Hildebrandt L (2008) Modeling joint purchases with a multivariate mnl approach. *Schmalenbach Business Review* 60(4):400–422.
- Chambers CP, Masatlioglu Y, Turansick C (2024) Correlated choice. *Theoretical Economics* 19(3):1087–1117.
- Charnes A, Cooper WW (1962) Programming with linear fractional functionals. *Nav. Res. Logist. Q.* 9(3-4):181–186.
- Chen X, Li J, Li M, Zhao T, Zhou Y (2022) Assortment optimization under the multivariate mnl model. *arXiv preprint arXiv:2209.15220* .
- Chen Y, He T, Rong Y, Wang Y (2026) An integer programming approach for quick-commerce assortment planning. *Management Sci.* ePub ahead of print, February 6, <http://dx.doi.org/10.1287/mnsc.2023.02996>.
- Del Pia A, Khajavirad A (2021) The running intersection relaxation of the multilinear polytope. *Math. Oper. Res.* 46(3):1008–1037.

- Del Pia A, Khajavirad A, Sahinidis NV (2020) On the impact of running intersection inequalities for globally solving polynomial optimization problems. *Math. Programming Comput.* 12(2):165–191.
- Gallego G, Ratliff R, Shebalov S (2015) A general attraction model and sales-based linear program for network revenue management under customer choice. *Oper. Res.* 63(1):212–232.
- Gallego G, Tang Z, Wang R (2019) Threshold utility model with applications to retailing and discrete choice models. Available at SSRN 3420155 .
- Gimbert M (2025) French bakery daily sales. Accessed September 22, <https://www.kaggle.com/datasets/matthieugimbert/french-bakery-daily-sales>.
- Gurobi Optimization, LLC (2025) Gurobi Optimizer Reference Manual. <https://www.gurobi.com>.
- Hastad J (1999) Clique is hard to approximate within $n^{-1-\epsilon}$. *Acta Mathematica* 182(1):105–142.
- Hauser JR (2014) Consideration-set heuristics. *Journal of Business Research* 67(8):1688–1699.
- Hauser JR, Wernerfelt B (1990) An evaluation cost model of consideration sets. *Journal of Consumer Research* 16(4):393–408.
- He L, Li X, Zhao Y (2025a) Proactive policing: Resource allocation for crime prevention with deterrence effect. *Oper. Res.* .
- He T, Liu S, Tawarmalani M (2024) Convexification techniques for fractional programs. *Math. Programming* 1–43.
- He T, Zhang Y, Zheng H (2025b) Assortment optimization under history-dependent effects. *arXiv preprint arXiv:2408.10967* .
- Hruschka H, Lukanowicz M, Buchta C (1999) Cross-category sales promotion effects. *Journal of Retailing and Consumer Services* 6(2):99–105.
- Huchette J, Vielma JP (2023) Nonconvex piecewise linear functions: Advanced formulations and simple modeling tools. *Oper. Res.* 71(5):1835–1856.
- Huh WT, Li H (2022) Optimal pricing under multiple-discrete customer choices and diminishing return of consumption. *Oper. Res.* 70(2):905–917.
- Jasin S, Lyu C, Najafi S, Zhang H (2024) Assortment optimization with multi-item basket purchase under multivariate mnl model. *Manufacturing Service Oper. Management* 26(1):215–232.
- Khalid H, Sturt B (2025) Assortment optimization and the sample average approximation. *arXiv preprint arXiv:2510.00850* .
- Kılınç-Karzan F, Küçükyavuz S, Lee D, Shafieezadeh-Abadeh S (2025) Conic mixed-binary sets: Convex hull characterizations and applications. *Oper. Res.* 73(1):251–269.
- Kwak K, Duvvuri SD, Russell GJ (2015) An analysis of assortment choice in grocery retailing. *Journal of Retailing* 91(1):19–33.
- Li X, Sun H, Teo CP (2022) Convex optimization for bundle size pricing problem. *Management Sci.* 68(2):1095–1106.

- Lin H, Li X, Wu L (2026) Intra-category multi-choice preferences learning and assortment recommendation in e-commerce. *Production Oper. Management* 35(1):304–330.
- Lo V, Topaloglu H (2019) Assortment optimization under the multinomial logit model with product synergies. *Oper. Res. Lett.* 47(6):546–552.
- Lubin M, Dowson O, Garcia JD, Huchette J, Legat B, Vielma JP (2023) JuMP 1.0: Recent improvements to a modeling language for mathematical optimization. *Math. Programming Comput.* 15(3):581–589.
- Ma W (2023) When is assortment optimization optimal? *Management Sci.* 69(4):2088–2105.
- Manchanda P, Ansari A, Gupta S (1999) The “shopping basket”: A model for multicategory purchase incidence decisions. *Marketing science* 18(2):95–114.
- Mao R, Li X (2025) Optimizing product pool selection and pricing for size-based bundling promotions. Available at SSRN .
- McCormick GP (1976) Computability of global solutions to factorable nonconvex programs: Part i—convex underestimating problems. *Math. Programming* 10(1):147–175.
- McFadden D, Train K (2000) Mixed mnl models for discrete response. *Journal of Applied Econometrics* 15(5):447–470.
- Padberg M (1989) The boolean quadric polytope: Some characteristics, facets and relatives. *Math. Programming* 45(1):139–172.
- Pan Y, Russell G, Gruca TS, Li C (2025) Multicategory purchase behavior: Basket choice, shopping frequency, and promotional analysis. *Journal of Retailing* .
- Rehfeldt D, Koch T, Shinano Y (2023) Faster exact solution of sparse maxcut and qubo problems. *Math. Programming Comput.* 15(3):445–470.
- Richards TJ, Hamilton SF, Yonezawa K (2018) Retail market power in a shopping basket model of supermarket competition. *Journal of Retailing* 94(3):328–342.
- Roberts JH, Lattin JM (1991) Development and testing of a model of consideration set composition. *Journal of Marketing Research* 28(4):429–440.
- Ruiz FJ, Athey S, Blei DM (2020) Shopper. *The Annals of Applied Statistics* 14(1):1–27.
- Rusmevichientong P, Shmoys D, Tong C, Topaloglu H (2014) Assortment optimization under the multinomial logit model with random choice parameters. *Production Oper. Management* 23(11):2023–2039.
- Russell GJ, Petersen A (2000) Analysis of cross category dependence in market basket selection. *Journal of Retailing* 76(3):367–392.
- Sen A, Atamtürk A, Kaminsky P (2018) A conic integer optimization approach to the constrained assortment problem under the mixed multinomial logit model. *Oper. Res.* 66(4):994–1003.
- Song I, Chintagunta PK (2006) Measuring cross-category price effects with aggregate store data. *Management Sci.* 52(10):1594–1609.

- Sumida M, Gallego G, Rusmevichientong P, Topaloglu H, Davis J (2021) Revenue-utility tradeoff in assortment optimization under the multinomial logit model with totally unimodular constraints. *Management Sci.* 67(5):2845–2869.
- Sun H, Li X, Teo CP (2025) Partition and prosper: Design and pricing of single bundle. *Oper. Res.* 73(4):1983–2001.
- Sun Z, Ahipasaoglu S, Li X (2020) A unified analysis for assortment planning with marginal distributions. Available at SSRN 3638783 .
- Talluri K, Van Ryzin G (2004) Revenue management under a general discrete choice model of consumer behavior. *Management Sci.* 50(1):15–33.
- Tulabandhula T, Sinha D, Karra SR, Patidar P (2023) Multi-purchase behavior: Modeling, estimation, and optimization. *Manufacturing Service Oper. Management* 25(6):2298–2313.
- Vasilyev A, Maier S, Seifert RW (2025) Assortment optimization given basket shopping behavior using the ising model. *arXiv preprint arXiv:2502.16260* .
- Vielma JP (2015) Mixed integer linear programming formulation techniques. *SIAM Rev.* 57(1):3–57.
- Yan Z, Cheng C, Natarajan K, Teo CP (2022) A representative consumer model in data-driven multiproduct pricing optimization. *Management Sci.* 68(8):5798–5827.
- Yang L, Lyu G, Yu Y, Sun L, Teo CP (2025) Assortment optimization under multi-choice ranking preferences: Hardness and tight approximations. Available at SSRN 5942375 .
- Zhang H, Piri H, Huh WT, Li H (2021) Assortment optimization under multiple-discrete customer choices. Available at SSRN 3988982 .

This page is intentionally blank. Proper e-companion title page, with INFORMS branding and exact metadata of the main paper, will be produced by the INFORMS office when the issue is being assembled.

Additional Discussions, Implementation Details, and Missing Proofs

EC.1. Proofs

EC.1.1. Proof of Theorem 1

THEOREM 1. *For any hypergraph G , the perspective formulation (4) is an exact (resp. locally sharp) MIP formulation for \mathcal{C}_G if the relaxation oracle \mathcal{R} is exact (resp. integral).*

Proof. Let $\mathcal{F}_{\text{pers}}$ be a polyhedron defined as the linear relaxation of (4). Our goal is to formally establish the relationship between \mathcal{C}_G defined in (1) and the projection $\text{proj}_{(\rho, \mathbf{y})}(\mathcal{F}_{\text{pers}})$.

Case 1: Exactness. We first show that $\mathcal{C}_G = \text{proj}_{(\rho, \mathbf{y})}(\mathcal{F}_{\text{pers}} \cap \{\mathbf{x} : \mathbf{x} \in \{0, 1\}^{|V|}\})$.

(Part 1: $\mathcal{C}_G \subseteq \text{proj}_{(\rho, \mathbf{y})}(\mathcal{F}_{\text{pers}})$) Let $(\rho, \mathbf{y}) \in \mathcal{C}_G$. By the definition of choice probabilities, there exists a binary assortment $\mathbf{x} \in \{0, 1\}^{|V|}$ such that

$$\rho + \sum_{e \in E} v_e y_e = \frac{1}{1 + \sum_{e \in E} v_e \prod_{j \in e} x_j} + \sum_{e \in E} v_e \frac{\prod_{j \in e} x_j}{1 + \sum_{e \in E} v_e \prod_{j \in e} x_j} = 1,$$

which precisely satisfies the probability constraint (4b). Since $v_e = \exp(u_e) > 0$ for all $e \in E$, it strictly follows that $\rho \in (0, 1]$. Consequently, we have $y_e/\rho = \prod_{j \in e} x_j$ for all $e \in E$. Because the oracle \mathcal{R} is exact for \mathcal{M}_G , the vector \mathbf{z} with $z_e = y_e/\rho$ satisfies constraint (4a).

Next, we verify constraints (4d) and (4e). For any $i \in V$, let $L_i(\mathbf{y})$ and $U_i(\mathbf{y})$ denote the right-hand sides of (4d) and (4e), respectively. Noting that $y_i = \rho x_i$ and $y_e = \rho \prod_{j \in e} x_j$, we evaluate these bounds based on the binary value of x_i :

- If $x_i = 0$: Then $y_i = 0$, and for any bundle e containing i , $y_e = 0$. For any bundle e not containing i , $0 \leq y_e \leq \rho$. Thus, $\max\{0, y_i + y_e - \rho\} = 0$ and $\min\{y_i, y_e\} = 0$. Both $L_i(\mathbf{y})$ and $U_i(\mathbf{y})$ equal 0, matching $x_i = 0$.

- If $x_i = 1$: Then $y_i = \rho$. For any bundle e containing i , we have $\max\{0, y_i + y_e - \rho\} = \max\{0, y_e\} = y_e$, and $\min\{y_i, y_e\} = \min\{\rho, y_e\} = y_e$ (since $y_e \leq \rho$). In this case, both $L_i(\mathbf{y})$ and $U_i(\mathbf{y})$ reduce to $\rho + \sum_{e \in E | i \in e} v_e y_e + \sum_{e \in E | i \notin e} v_e y_e = \rho + \sum_{e \in E} v_e y_e$, which equals 1, perfectly matching $x_i = 1$.

Since $L_i(\mathbf{y}) = U_i(\mathbf{y}) = x_i$ holds for both binary cases, constraints (4d) and (4e) are satisfied. Thus, $(\rho, \mathbf{y}, \mathbf{x}) \in \mathcal{F}_{\text{pers}}$.

(Part 2: $\text{proj}_{(\rho, \mathbf{y})}(\mathcal{F}_{\text{pers}}) \subseteq \mathcal{C}_G$). Conversely, suppose $(\rho, \mathbf{y}, \mathbf{x}) \in \mathcal{F}_{\text{pers}}$ with $\mathbf{x} \in \{0, 1\}^{|V|}$. Define $\mathbf{z} \in \mathcal{R}$ and hence $y_e = \rho z_e \in [0, 1]$. Since $z_e \in [0, 1]$ and $\rho \geq 0$, we have $0 \leq y_e \leq \rho$ for all $e \in E$. Because $\mathbf{v} > \mathbf{0}$ and the maximization terms are non-negative, constraint (4d) implies:

$$x_i \geq y_i + \sum_{e \in E | i \in e} v_e y_e + \sum_{e \in E | i \notin e} v_e \max\{0, y_i + y_e - \rho\} \geq y_i. \quad (\text{EC.1})$$

Simultaneously, using the identity $\rho + \sum_{e \in E} v_e y_e = 1$, constraint (4e) implies:

$$x_i \leq y_i + \sum_{e \in E | i \in e} v_e y_e + \sum_{e \in E | i \notin e} v_e y_e = y_i + (1 - \rho). \quad (\text{EC.2})$$

Coupling $y_i \leq x_i \leq y_i + 1 - \rho$ with the binary restriction $x_i \in \{0, 1\}$, we can get $y_i = \rho x_i$ for all $i \in V$. Since \mathcal{R} is exact, $\mathbf{z} \in \mathcal{R}$ implies $z_e = \prod_{j \in e} x_j$, giving $y_e = \rho \prod_{j \in e} x_j$. Substituting this into (4b) perfectly reconstructs the definition of \mathcal{C}_G .

Case 2: Local sharpness. We now show that the $\mathcal{F}_{\text{pers}}$ projects exactly to the convex hull of \mathcal{C}_G .

When \mathbf{x} is relaxed to $x_i \in [0, 1]$, we can apply Fourier-Motzkin elimination to project out \mathbf{x} . For each x_i , the system imposes two lower bounds (0 and $L_i(\mathbf{y})$) and two upper bounds (1 and $U_i(\mathbf{y})$). The projection is valid if and only if all lower bounds are less than or equal to all upper bounds. We verify this systematically: First, $0 \leq 1$ holds trivially. Second, since $y_e \geq 0$ and $v_e > 0$, $0 \leq U_i(\mathbf{y})$ is evident. Third, because $y_i, y_e \in [0, \rho]$, we have $\max\{0, y_i + y_e - \rho\} \leq y_e$. Thus, $L_i(\mathbf{y}) \leq y_i + \sum_{e \in E} v_e y_e \leq \rho + \sum_{e \in E} v_e y_e = 1$. Finally, to see $L_i(\mathbf{y}) \leq U_i(\mathbf{y})$, we only need to verify that $\max\{0, y_i + y_e - \rho\} \leq \min\{y_i, y_e\}$. This holds naturally because $y_i, y_e \leq \rho$ implies $y_i + y_e - \rho \leq y_i$ and $y_i + y_e - \rho \leq y_e$. Therefore, projecting out \mathbf{x} imposes *no new constraints* on (ρ, \mathbf{y}) . Define $\mathbf{z} \in \mathcal{R}$ and the projection is simply:

$$\text{proj}_{(\rho, \mathbf{y})}(\mathcal{F}_{\text{pers}}) = \left\{ (\rho, \mathbf{y}) \mid z_e = y_e / \rho, \mathbf{z} \in \mathcal{R}, \rho + \sum_{e \in E} v_e y_e = 1 \right\}.$$

By Theorem 2 of [He et al. \(2024\)](#), we know

$$\text{conv}(\mathcal{C}_G) = \left\{ (\rho, \mathbf{y}) \mid (\rho, \mathbf{y}) \in \rho \cdot \text{conv}(F), \rho + \sum_{e \in E} v_e y_e = 1, \rho \geq 0 \right\},$$

where $F = \{(1, \mathbf{z}) \mid z_e = \prod_{j \in e} x_j, \mathbf{x} \in \{0, 1\}^{|V|}, e \in E\}$. Since \mathcal{R} is an integral relaxation for \mathcal{M}_G , we have $\text{conv}(F) = \{(1, \mathbf{z}) \mid \mathbf{z} \in \mathcal{R}\}$, which perfectly coincides with constraint (4a). \square

EC.1.2. Proofs of Propositions in Section 3.1

Before proving these propositions, we first introduce the *standard linear relaxation* of the choice availability set \mathcal{M}_G , defined as follows.

$$\begin{aligned} 0 \leq z_e \leq z_i & & \text{for } i \in e, e \in E \\ z_e \geq \sum_{i \in e} z_i - |e| + 1 & & \text{for } e \in E. \end{aligned}$$

The standard linear relaxation yields an exact formulation for multilinear sets. As shown in the following lemma, it is inherently implied by the recursive McCormick inequalities.

LEMMA EC.1. *The standard linear relaxation is implied by recursive McCormick inequalities.*

Proof. The upper bound $0 \leq z_e \leq z_i$ for any $i \in e$ is directly imposed by the recursive McCormick inequalities. For the lower bound, we proceed by a straightforward induction on the bundle size $|e|$. Recursively applying the McCormick lower bound inequality $z_e \geq z_j + z_{e \setminus \{j\}} - 1$ over all elements $j \in e$ systematically accumulates the terms, yielding exactly $z_e \geq \sum_{i \in e} z_i - |e| + 1$. \square

PROPOSITION 1. *If the oracle \mathcal{R} is given by (RMC) then the perspective formulation (4) is an MIP formulation of \mathcal{C}_G .*

Proof. When the decision variables are restricted to binary values ($\mathbf{x} \in \{0, 1\}^{|V|}$), the recursive McCormick inequalities exactly enforce the multilinear relationship $z_e = \prod_{i \in e} x_i$. Therefore, because the oracle \mathcal{R} perfectly captures \mathcal{M}_G at all integer points, \mathcal{R} is an exact MIP formulation of \mathcal{M}_G . The exactness of the perspective formulation then follows directly from Theorem 1. \square

PROPOSITION 2. *If the oracle \mathcal{R} is given by (RMC) and (ODD-CYCLE), then the perspective formulation (4) is an MIP formulation of \mathcal{C}_G for any multi-purchase hypergraph G . If G is a series-parallel graph, then (4) is also a locally sharp formulation of \mathcal{C}_G .*

Proof. By construction, inequalities (ODD-CYCLE) are valid for the multilinear set \mathcal{M}_G (Padberg 1989). Hence, by Theorem 1, (4) is an MIP formulation of \mathcal{C}_G .

Theorem 10 in Padberg (1989) shows that if G is a series-parallel graph, the convex hull of \mathcal{M}_G is characterized by two types of inequalities. The first class comprises the trivial inequalities (i.e., the standard linear relaxation), which are implied by (RMC) via Lemma EC.1. The second class consists of the nontrivial odd-cycle inequalities (ODD-CYCLE). The relaxation oracle by (RMC) and (ODD-CYCLE) describes the convex hull of \mathcal{M}_G , and thus is integral when G is a series-parallel graph. The local sharpness of (4) then follows immediately from Theorem 1. \square

PROPOSITION 3. *If the oracle \mathcal{R} is given by (RMC) and (RIC), then the perspective formulation (4) is an MIP formulation of \mathcal{C}_G for any multi-purchase hypergraph G . If G is a kite-free β -acyclic hypergraph, then (4) is also a locally sharp formulation of \mathcal{C}_G .*

Proof. Proposition 10 in Del Pia and Khajavirad (2021) establishes that the running-intersection inequalities (RIC) are valid for the multilinear set \mathcal{M}_G . Thus, by Theorem 1, (4) is an MIP formulation of \mathcal{C}_G .

Theorem 3 in Del Pia and Khajavirad (2021) shows that if G is a kite-free β -acyclic hypergraph, its convex hull is characterized by two types of inequalities: the trivial inequalities (implied

by **(RMC)** via Lemma **EC.1)** and the nontrivial running-intersection inequalities **(RIC)**. Consequently, **(RMC)** and **(RIC)** jointly constitute a locally sharp formulation of \mathcal{M}_G . The local sharpness of the perspective formulation follows directly from Theorem 1. \square

EC.1.3. Proof of Proposition 4

PROPOSITION 4. *If the relaxation oracle \mathcal{R} is exact, then the Big-M formulation (8) is an MIP formulation of \mathcal{C}_G .*

Proof. We first show $\text{proj}_{(\rho, \mathbf{y})}(\mathbf{8}) \subseteq (1)$. Let $(\rho, \mathbf{y}, \mathbf{x}, \mathbf{z})$ be a feasible solution in (8). The binary restriction on \mathbf{x} combined with the McCormick inequalities (8b) enforces $y_e = \rho z_e$ for each $e \in E$. Because $\mathbf{z} \in \mathcal{R}$ and \mathbf{x} is binary, the exactness of the oracle \mathcal{R} guarantees $z_e = \prod_{i \in e} x_i$. Substituting these relations into the probability constraint $\rho + \sum_{e \in E} v_e y_e = 1$ of (8), we obtain:

$$\rho + \sum_{e \in E} v_e \left(\rho \cdot \prod_{i \in e} x_i \right) = 1.$$

Since v and \mathbf{x} are non-negative, the term $1 + \sum_{e \in E} v_e \prod_{i \in e} x_i \geq 1 > 0$. This ensures the denominator is strictly positive, allowing us to safely factor out ρ and rearrange the equation as:

$$\rho = \frac{1}{1 + \sum_{e \in E} v_e \prod_{i \in e} x_i}, y_e = \frac{\prod_{i \in e} x_i}{1 + \sum_{e \in E} v_e \prod_{i \in e} x_i} \text{ for } e \in E,$$

which is exactly the definition of \mathcal{C}_G .

Conversely, we prove $\mathcal{C}_G \subseteq \text{proj}_{(\rho, \mathbf{y})}(\mathbf{8})$. Suppose (ρ, \mathbf{y}) is a feasible solution of (1). By the definition of \mathcal{C}_G , there exists a binary assortment vector $\mathbf{x} \in \{0, 1\}^{|V|}$ generating these probabilities. Consequently, we have:

$$\rho + \sum_{e \in E} v_e y_e = \frac{1}{1 + \sum_{e \in E} v_e \prod_{i \in e} x_i} + \sum_{e \in E} v_e \frac{\prod_{i \in e} x_i}{1 + \sum_{e \in E} v_e \prod_{i \in e} x_i} = 1.$$

Note that all attraction values are positive, i.e., $v_e > 0$ for $e \in E$, and hence $\rho \in (0, 1]$. Let $z_e = \prod_{i \in e} x_i$ for each $e \in E$. Because \mathbf{x} is binary and the oracle \mathcal{R} is exact, we naturally obtain $\mathbf{z} \in \mathcal{R}$, satisfying constraint (8a). By the definition of \mathbf{y} , we have $y_e = \rho z_e$ for all $e \in E$. Furthermore, \mathbf{x} explicitly enforces the integrality condition $\mathbf{z} \in \{0, 1\}^{|E|}$. Taken together, these two conditions ensure that the Big-M relaxation (8b) is satisfied. Therefore, $(\rho, \mathbf{y}, \mathbf{x}, \mathbf{z}) \in (8)$, concluding the proof. \square

EC.1.4. Proof of Theorem 2

THEOREM 2. *For any relaxation oracle \mathcal{R} , $\mathcal{R}_{\text{pers}} \subseteq \mathcal{R}_{\text{Big-M}}$.*

Proof. Suppose $(\rho, \mathbf{y}) \in \mathcal{R}_{\text{pers}}$. By definition, there exists an assortment decision \mathbf{x} such that $(\rho, \mathbf{y}, \mathbf{x})$ belongs to the linear relaxation of (4). Coupling the probability identity $\rho + \sum_{e \in E} v_e y_e = 1$ with the conditions $0 \leq y_e \leq \rho$ for all $e \in E$ and $\mathbf{v} > \mathbf{0}$, it strictly follows that $\rho \in (0, 1]$. Because ρ is strictly positive, we can well-define $z_e := y_e / \rho$ for each $e \in E \setminus V$.

Next, we demonstrate that $(\rho, \mathbf{y}, \mathbf{z}, \mathbf{x})$ satisfies all constraints of the linear relaxation of (8), thereby establishing $(\rho, \mathbf{y}) \in \mathcal{R}_{\text{Big-M}}$. The inherent condition $0 \leq y_e \leq \rho$ immediately guarantees that $z_e \in [0, 1]$. Furthermore, since $(\rho, \mathbf{y}, \mathbf{x})$ satisfies the scaled constraints of the perspective formulation $\mathcal{R}_{\text{pers}}$, unscaling these constraints by the positive scalar ρ ensures that $\mathbf{z} \in \mathcal{R}$, which naturally satisfies constraint (8a).

We now verify that $(\rho, \mathbf{y}, \mathbf{z}, \mathbf{x})$ satisfies the Big-M relaxation constraints for $y_e = \rho z_e$ by substituting the natural domain bounds $\rho_L = 0$ and $\rho_U = 1$:

- Substituting $\rho_L = 0$, the non-negativity $y_e \geq 0$ directly yields $y_e \geq \rho_L z_e$, while $y_e \leq \rho$ equivalently yields $y_e \leq \rho_L z_e + \rho - \rho_L$.
- Substituting $\rho_U = 1$, since $\rho \leq 1$ and $z_e \geq 0$, we safely obtain $y_e = \rho z_e \leq z_e = \rho_U z_e$.
- Finally, because both $\rho \leq 1$ and $z_e \leq 1$, the product $(1 - \rho)(1 - z_e)$ must be non-negative. Expanding $(1 - \rho)(1 - z_e) \geq 0$ yields $\rho z_e \geq z_e + \rho - 1$, which exactly corresponds to $y_e \geq \rho_U z_e + \rho - \rho_U$.

Since all constraints of the linear relaxation of (8) are rigorously satisfied, we conclude that $(\rho, \mathbf{y}) \in \mathcal{R}_{\text{Big-M}}$. □

EC.1.5. Proof of Theorem 3

THEOREM 3. *The assortment optimization problem (AOPT) is NP-hard to approximate within factor $O(|V|^{1-\epsilon})$ for any fixed $\epsilon \in (0, 1)$ even when there is no constraint.*

Our hardness result is established via a reduction from the Maximum Independent Set (MIS) problem. Given an undirected simple graph $G(V, E)$, the MIS problem seeks a maximum-cardinality subset of vertices such that no two nodes are adjacent. It is a well-known NP-hard problem whose approximation is difficult. Specifically, finding an approximation solution $S \subseteq V$ such that $\frac{|S^*|}{|S|} \leq O(|V|^{1-\epsilon})$ for any fixed $\epsilon \in (0, 1)$ is NP-hard unless $P = NP$, where S^* is the optimal maximum independent set. This result is formally stated in the following lemma.

LEMMA EC.2. (*Hastad 1999*) *For any $\epsilon \in (0, 1)$, it is NP-hard to approximate the Maximum Independent Set problem to within $N^{1-\epsilon}$, where N is the number of nodes.*

Proof. Given an MIS instance $G = (V, E_{MIS})$ with $|V| = N$, we construct an assortment optimization instance under the Logit-MP model as follows. We define the consideration set as $E = V \cup E_{MIS}$. For each singleton bundle $i \in V$, we set its utility $v_i = 1/N^2$ and revenue $r_i = N$. For each edge bundle $(i, j) \in E_{MIS}$, we set its utility $v_{ij} = N^3$ and revenue $r_{ij} = 0$. This construction is clearly polynomial in N .

For any offered assortment $S \subseteq V$, let $\beta(S)$ denote the number of edges induced by S , i.e.,

$$\beta(S) = |\{(i, j) \in E_{MIS} \mid i, j \in S\}|.$$

The objective function of the constructed assortment optimization problem (AOP) equals:

$$AOP(S) = \frac{N \cdot |S| \cdot (1/N^2)}{1 + |S| \cdot (1/N^2) + N^3 \cdot \beta(S)} = \frac{|S|/N}{1 + |S|/N^2 + N^3 \beta(S)},$$

where the zero-revenue edge bundles contribute exclusively to the denominator, acting as severe penalties for selecting adjacent nodes.

Next, we establish the bounds for the objective $AOP(S)$. Suppose $S \subseteq V$ is a feasible assortment. If S is an independent set, then by definition $\beta(S) = 0$, which simplifies the objective to:

$$AOP(S) = \frac{|S|/N}{1 + |S|/N^2}. \quad (\text{EC.3})$$

Since $|S| \leq N$, we have $0 \leq |S|/N^2 \leq 1/N$. This allows us to construct tight lower and upper bounds for any independent set S :

$$AOP(S) = \frac{|S|/N}{1 + |S|/N^2} \geq \frac{|S|/N}{1 + 1/N} = \frac{|S|}{N + 1}, \quad (\text{EC.4})$$

and

$$AOP(S) = \frac{|S|/N}{1 + |S|/N^2} \leq \frac{|S|}{N}. \quad (\text{EC.5})$$

Conversely, if S is *not* an independent set, then $\beta(S) \geq 1$. Consequently, we can bound the objective from above:

$$AOP(S) \leq \frac{|S|/N}{N^3 \beta(S)} \leq \frac{|S|/N}{N^3} \leq \frac{1}{N^3}, \quad (\text{EC.6})$$

where the inequalities follow from the facts that $1 + |S|/N^2 \geq 0$, $\beta(S) \geq 1$, and $|S|/N \leq 1$.

Based on these bounds, we claim that for any $N \geq 2$, the optimal assortment S_{AOP}^* must be an independent set. For the sake of contradiction, suppose S_{AOP}^* is not an independent set. By

(EC.6), we have $AOP(S_{AOP}^*) \leq 1/N^3$. However, consider a trivial assortment $\{i\}$ consisting of any single product, which is inherently an independent set of size 1. By (EC.4), its expected revenue is $AOP(\{i\}) \geq 1/(N+1)$. Since $1/(N+1) > 1/N^3$ strictly holds for all $N \geq 2$, we obtain $AOP(\{i\}) > AOP(S_{AOP}^*)$, which contradicts the optimality of S_{AOP}^* . Thus, S_{AOP}^* is guaranteed to be an independent set.

Furthermore, observe that the function $f(x) = \frac{x/N}{1+x/N^2}$ is strictly increasing for $x \in [0, N]$. Therefore, restricted to independent sets, $AOP(S)$ is strictly monotonically increasing in the cardinality $|S|$, see (EC.3). This structural property ensures that the optimal assortment exactly coincides with the maximum independent set, i.e., $S_{AOP}^* = S_{MIS}^*$.

Now, we establish the inapproximability result. We prove that any polynomial-time approximation algorithm for our assortment problem can be converted into an approximation algorithm for MIS, eventually violating Lemma EC.2. Suppose, for contradiction, there exists a polynomial-time algorithm \mathcal{A} that returns a feasible assortment \hat{S} with an approximation ratio $\gamma \geq 2N^{-(1-\epsilon)}$, such that:

$$AOP(\hat{S}) \geq \gamma \cdot AOP(S_{AOP}^*). \quad (\text{EC.7})$$

First, we must show that for sufficiently large N , \hat{S} is forced to be an independent set. Since S_{AOP}^* is optimal, $|S_{AOP}^*| \geq 1$, which implies $AOP(S_{AOP}^*) \geq 1/(N+1)$. Thus, the algorithm guarantees $AOP(\hat{S}) \geq \frac{\gamma}{N+1} = \frac{2}{N^{1-\epsilon}(N+1)}$. If \hat{S} were not an independent set, we would have $AOP(\hat{S}) \leq 1/N^3$. However, for any fixed $\epsilon > 0$ and sufficiently large N , $\frac{2}{N^{1-\epsilon}(N+1)} > \frac{1}{N^3}$, leading to a contradiction. Hence, \hat{S} must be an independent set.

Because \hat{S} is an independent set, we can safely apply the upper bound from (EC.5), yielding $|\hat{S}| \geq N \cdot AOP(\hat{S})$. We then derive:

$$|\hat{S}| \geq N \cdot AOP(\hat{S}) \geq N \cdot \gamma \cdot AOP(S_{AOP}^*) = N \cdot \gamma \cdot AOP(S_{MIS}^*).$$

Using the lower bound from (EC.4) for $AOP(S_{MIS}^*)$, we obtain:

$$|\hat{S}| \geq N \cdot \gamma \cdot \frac{|S_{MIS}^*|}{N+1} \geq \frac{\gamma}{2} \cdot |S_{MIS}^*|,$$

where the last step uses the fact that $\frac{N}{N+1} \geq \frac{1}{2}$ for all $N \geq 1$. Substituting $\gamma \geq 2N^{-(1-\epsilon)}$, we conclude that $|\hat{S}| \geq N^{-(1-\epsilon)} |S_{MIS}^*|$. This implies we have discovered a polynomial-time algorithm that approximates the MIS problem within a factor of $O(N^{1-\epsilon})$. This directly contradicts Lemma EC.2. Therefore, it is NP-hard to approximate the assortment optimization problem under the general Logit-MP model within a factor of $O(|V|^{1-\epsilon})$. \square

EC.1.6. Proofs of Corollaries 1 and 2

COROLLARY 1. *The formulation (PERS) exactly solves the assortment optimization problem (AOPT) when the relaxation oracle is given by (RMC). The LP relaxation of (PERS) solves (AOPT) when the problem is unconstrained, that is, $\mathcal{X} = \{0, 1\}^V$, the multi-purchase hypergraph G is a series-parallel (resp. kite-free β -acyclic) graph, and the relaxation oracle is derived by (RMC) and (ODD-CYCLE) (resp. (RIC)).*

Proof: For the assortment optimization problem (AOPT), define assortment decision variables $x_i = \mathbb{1}(i \in S)$, choice probability variables $y_e = P_E(e, S)$, and no-purchase probability $\rho = 1 - \sum_{e \in E} y_e$. Because $\mathbf{x} \in \{0, 1\}^{|V|}$, the bundle inclusion indicator can be exactly captured by the product $\mathbb{1}(e \subseteq S) = \prod_{i \in e} x_i$. Hence, problem (AOPT) equivalently reduces to:

$$\max_{\rho, \mathbf{x}, \mathbf{y}} \sum_{e \in E} r_e v_e y_e \quad \text{s. t. } (\rho, \mathbf{y}) \in \mathcal{C}_G. \quad (\text{EC.8})$$

By Theorem 1, (4) is an exact formulation of \mathcal{C}_G . Hence, substituting the non-linear fractional probabilities in \mathcal{C}_G with (4) yields an exact reformulation, which is exactly (PERS).

Furthermore, as established in our previous results, when the multi-purchase hypergraph G is series-parallel or kite-free β -acyclic, \mathcal{C}_G admits an exact polyhedral characterization via the linear inequalities in (RMC) and (ODD-CYCLE) (resp. (RIC)). Consequently, the convex hull of the choice probabilities is explicitly captured without requiring auxiliary integer variables, meaning the LP relaxation of (PERS) exactly solves the original problem. \square

COROLLARY 2. *Formulations (PERS-MIX) and (PERS-ROBUST) solve problems (AOPT-MIX) and (AOPT-ROBUST), respectively.*

Proof: Analogous to the single-class setting, we define the class-specific choice probability variables y_e^k and the no-purchase probability ρ^k for each customer class (or scenario) k . The proof idea remains identical: the choice probabilities for each class k must validly and independently reside within their respective probability space \mathcal{C}_{G^k} . By replacing the choice probabilities of each class with the continuous variables $(\rho^k, \mathbf{y}^k) \in \mathcal{C}_{G^k}$, and coupling these distinct probability spaces exclusively through the single, shared discrete assortment decision $\mathbf{x} \in \{0, 1\}^{|V|}$, formulations (PERS-MIX) and (PERS-ROBUST) directly emerge as exact reformulations of their original counterparts. \square

EC.2. Implementation Details

Implementing the perspective formulation (4) presents two computational challenges. First, (4) contains $\mathcal{O}(|V|2^{|\mathbb{E} \setminus V|})$ inequalities for bounding \mathbf{x} . Second, the polyhedral relaxation oracle \mathcal{R} may also include a large number of inequalities, particularly when the multi-purchase hypergraph is dense. To address this issue, we design a cutting-plane implementation. Our approach begins with a simplified base formulation containing fewer constraints. We then iteratively employ two separation oracles for the two classes of inequalities, thereby tightening the formulation. Finally, we solve the tightened formulation with binary constraints to obtain an optimal integer solution. We detail the base formulation and separation oracles for constraints (4d) and (4e), odd-cycle inequalities, and running-intersection inequalities in Section EC.2.1 and present the complete cutting-plane procedure in Section EC.2.2.

EC.2.1. Base Formulation and Separation Oracles

We construct a base formulation as follows. We first construct the base relaxation oracle \mathcal{R} by associating each bundle e with a left-deep binary tree. We build binary trees via the following Algorithm 1.

Algorithm 1 Construction of Binary Trees for Recursive McCormick Inequalities

Require: A multi-purchase hypergraph $G(V, \mathbb{E})$; an auxiliary-variable registry \mathcal{W}

Ensure: Binary trees for each hyperedge $e \in \mathbb{E}$ with $|e| \geq 2$

- 1: Initialize each hyperedge $e = \{i_1, \dots, i_k\}$ with indices sorted in ascending order $i_1 < \dots < i_k$
 - 2: Initialize an un-visited set $S = \{e : |e| \geq 2, e \in \mathbb{E}\}$ and a visited set $H \leftarrow \emptyset$
 - 3: **while** S is non-empty **do**
 - 4: Find a root node $e = \arg\min_{e \in S} |e|$ and $S \leftarrow S \setminus \{e\}, H \leftarrow H \cup \{e\}$
 - 5: Define $e' = \{i_1, i_2, \dots, i_{|e|-1}\}$ as the left child (the subtree associated with e')
 - 6: Define $j = i_{|e|}$ as the right child (the right node corresponding to x_j)
 - 7: **if** $|e'| > 1$ and $e' \notin H$ **then**
 - 8: $S \leftarrow S \cup \{e'\}$ and $\mathcal{W} \leftarrow \mathcal{W} \cup \{e'\}$
 - 9: **end if**
 - 10: **end while**
 - 11: **return** binary trees and \mathcal{W}
-

The recursive McCormick inequalities are then imposed along the edges of this tree. We refer to the relaxation oracle with these recursive inequalities as \mathcal{R}_{RMC} .

Regarding constraints (4d) and (4e), we initially incorporate only the following extreme cases:

$$x_i \geq y_i + \sum_{e \in \mathbb{E} | i \in e} y_e \quad \text{for } i \in V \quad (\text{EC.9a})$$

$$x_i \geq y_i + \sum_{e \in \mathbb{E} | i \in e} y_e + \sum_{e \in \mathbb{E} | i \notin e} v_e (y_i + y_e - \rho) \quad \text{for } i \in V \quad (\text{EC.9b})$$

$$x_i \leq y_i + \sum_{e \in \mathbb{E} | i \in e} y_e + \sum_{e \in \mathbb{E} | i \notin e} v_e y_i \quad \text{for } i \in V \quad (\text{EC.9c})$$

$$x_i \leq y_i + \sum_{e \in \mathbb{E} | i \in e} y_e + \sum_{e \in \mathbb{E} | i \notin e} v_e y_e \quad \text{for } i \in V. \quad (\text{EC.9d})$$

Combining these elements yields the following base formulation:

$$\{(\rho, \mathbf{y}, \mathbf{x}) \mid \rho + \sum_{e \in \mathbb{E}} y_e = 1, \rho \geq 0, \mathbf{y} \in \rho \cdot \mathcal{R}_{\text{RMC}}, (\text{EC.9}), \mathbf{x} \in \{0, 1\}^{|\mathbb{V}|}\}. \quad (\text{Base})$$

The next proposition establishes the validity of the base formulation.

PROPOSITION EC.1. *(Base) is an MIP formulation for (1).*

Proof. We prove this result by leveraging the established fact that (8) is an exact MIP formulation for (1) (see Proposition 4). Specifically, it suffices to show that the continuous relaxation of formulation (Base) is contained within the continuous relaxation of the Big-M formulation (8).

Let $(\rho, \mathbf{y}, \mathbf{x})$ be an arbitrary feasible point in the continuous relaxation of (Base). By the definition of the perspective scaled set $\mathbf{y} \in \rho \cdot \mathcal{R}_{\text{RMC}}$, there exists a vector $\mathbf{z} \in \mathcal{R}_{\text{RMC}}$ such that $y_e = \rho z_e$ for all $e \in \mathbb{E}$.

Because $\mathbf{z} \in \mathcal{R}_{\text{RMC}}$, the variables naturally satisfy $z_e \in [0, 1]$ for each $e \in \mathbb{E}$. Coupling this with the non-negativity constraint $\rho \geq 0$, we immediately obtain $0 \leq y_e \leq \rho$. Furthermore, since $v_e > 0$ and $y_e \geq 0$ for all $e \in \mathbb{E}$, the probability identity constraint $\rho + \sum_{e \in \mathbb{E}} v_e y_e = 1$ strictly ensures that $\rho \leq 1$.

Now, we verify that the point $(\rho, \mathbf{y}, \mathbf{x}, \mathbf{z})$ satisfies the continuous relaxation of (8). Constraint (8a) holds by construction. We verify the Big-M envelopes (8b) as follows:

- The bounds $0 \leq y_e \leq \rho$ inherently satisfy $y_e \geq \rho_L z_e$ and $y_e \leq \rho_L z_e + \rho - \rho_L$.
- Since $\rho \leq 1$ and $z_e \geq 0$, we have $y_e = \rho z_e \leq 1 \cdot z_e = z_e$, perfectly matching $y_e \leq \rho_U z_e$.
- Since both $\rho \leq 1$ and $z_e \leq 1$, their complements are non-negative, yielding $(1 - \rho)(1 - z_e) \geq 0$.

Expanding this product gives $\rho z_e \geq z_e + \rho - 1$. Substituting $y_e = \rho z_e$ directly yields $y_e \geq z_e + \rho - 1$, which aligns exactly with $y_e \geq \rho_U z_e + \rho - \rho_U$.

Since all inequalities in (8) are satisfied, the continuous relaxation of (Base) is fully contained in that of (8). Therefore, formulation (Base) is also a valid (and potentially tighter) exact MIP formulation of (1). \square

While valid, formulation (Base) can be further tightened by iteratively adding the remaining inequalities from the constraint pool of (4d) and (4e) or stronger multilinear inequalities, e.g., (ODD-CYCLE) and (RIC). We describe the separation algorithms for these constraints below.

Algorithm 2 Separation Oracle for Constraints (4d) and (4e)

Require: A point $(\hat{\rho}, \hat{\mathbf{y}}, \hat{\mathbf{x}})$; attraction value vector \mathbf{v} ; a threshold $\epsilon > 0$.

Ensure: A set of violated constraints, Cuts.

- 1: Initialize Cuts $\leftarrow \emptyset$.
 - 2: **for** $i \in V$ **do**
 - 3: Initialize $A, B, C \leftarrow \emptyset$.
 - 4: **for** each edge $e \in \{e \in E \mid i \notin e\}$ **do**
 - 5: **if** $\hat{y}_i + \hat{y}_e - \hat{\rho} > 0$ **then**
 - 6: $A \leftarrow A \cup \{e\}$
 - 7: **end if**
 - 8: **if** $\hat{y}_e < \hat{y}_i$ **then**
 - 9: $B \leftarrow B \cup \{e\}$
 - 10: **else**
 - 11: $C \leftarrow C \cup \{e\}$
 - 12: **end if**
 - 13: **end for**
 - 14: **if** $\hat{x}_i < \hat{y}_i + \sum_{e \in E \mid i \in e} \hat{y}_e v_e + \sum_{e \in A} v_e (\hat{y}_i + \hat{y}_e - \hat{\rho}) - \epsilon$ **then**
 - 15: Cuts $\leftarrow x_i \geq y_i + \sum_{e \in E \mid i \in e} y_e v_e + \sum_{e \in A} v_e (y_i + y_e - \rho)$.
 - 16: **end if**
 - 17: **if** $\hat{x}_i > \hat{y}_i + \sum_{e \in E \mid i \in e} \hat{y}_e v_e + \sum_{e \in B} v_e \hat{y}_e + \sum_{e \in C} v_e \hat{y}_i + \epsilon$ **then**
 - 18: Cuts $\leftarrow x_i \leq y_i + \sum_{e \in E \mid i \in e} y_e v_e + \sum_{e \in B} v_e y_e + \sum_{e \in C} v_e y_i$.
 - 19: **end if**
 - 20: **end for**
 - 21: **return** Cuts.
-

EC.2.1.1. Separation Oracle for Constraints (4d) and (4e) Given a point $(\hat{\rho}, \hat{\mathbf{y}}, \hat{\mathbf{x}})$, the separation oracle checks whether the lower and upper bounds for each \hat{x}_i constructed by $\hat{\rho}$ and $\hat{\mathbf{y}}$ are violated. The bounds are given by

$$x_i \geq \hat{y}_i + \sum_{e \in E | i \in e} \hat{y}_e v_e + \sum_{e \in E | i \notin e} v_e \max\{0, \hat{y}_i + \hat{y}_e - \hat{\rho}\} \quad (\text{EC.10a})$$

$$x_i \leq \hat{y}_i + \sum_{e \in E | i \in e} \hat{y}_e v_e + \sum_{e \in E | i \notin e} v_e \min\{\hat{y}_e, \hat{y}_i\}. \quad (\text{EC.10b})$$

If the current point violates the above inequalities, we add the corresponding constraint to the base model. We detail each step in Algorithm 2.

EC.2.1.2. Separation Oracle for Odd-cycle Inequalities The main difficulty in separating odd-cycle inequalities is that both the number of cycles C and the number of odd subsets $D \subseteq C$ can grow exponentially. **Barahona and Mahjoub (1986)** reformulate the odd-cycle inequality in terms of edge variables and design a polynomial-time separation oracle based on shortest-path computations. We adapt their construction to our scaled space, where inequalities are multiplied by the no-purchase variable ρ .

Given a simple graph $G(V, E)$, we introduce a virtual node o with fixed value $x_o = 1$, and connect it to every node $i \in V$. The resulting graph has nodes $V \cup \{o\}$ and edges $\{(i, o) | i \in V\} \cup \{(i, j) | (i, j) \in E\}$. For each edge, we introduce a new variable c_{ij} as follows

$$c_{io} = x_i, c_{ij} = x_i + x_j - 2z_{ij}.$$

This mapping establishes a one-to-one correspondence between (\mathbf{x}, \mathbf{z}) and \mathbf{c} . Under this transformation, the odd-cycle inequality becomes

$$\sum_{(i,j) \in D} (1 - c_{ij}) + \sum_{(i,j) \in C \setminus D} c_{ij} \geq 1, C \text{ is a cycle and } D \subseteq C, |D| \text{ is odd.}$$

The left-hand side can be interpreted as the weight of cycle C , where edges in D receive weight $1 - c_{ij}$ and the remaining edges receive weight c_{ij} . The separation problem is therefore equivalent to finding a minimum-weight cycle under the constraint that an odd number of edges are selected with weight $1 - c_{ij}$.

Following **Barahona and Mahjoub (1986)**, we solve this separation problem via an extended graph G' . For each node $i \in V \cup \{o\}$ in the original graph, G' contains two nodes i and i' . G' has four types of edges:

$$E' = \{(i, j), (i', j'), (i, j'), (i', j) | (i, j) \in E \cup \{(i, o) | i \in V\}\}.$$

Given a candidate point $(\hat{\rho}, \hat{\mathbf{y}})$, we define scaled edge lengths as follows:

$$w_{io} = w_{i'o'} = \hat{\rho} \cdot \hat{c}_{io} = \hat{\rho} \cdot \hat{x}_i = \hat{y}_i \quad \forall i \in V \quad (\text{EC.11a})$$

$$w_{ij} = w_{i'o'} = \hat{\rho} \cdot \hat{c}_{ij} = \hat{\rho} \cdot (\hat{x}_i + \hat{x}_j - 2\hat{z}_{ij}) = \hat{y}_i + \hat{y}_j - 2\hat{y}_{ij} \quad \forall (i, j) \in E \setminus V \quad (\text{EC.11b})$$

$$w_{ij'} = w_{i'j} = \hat{\rho} \cdot (1 - \hat{c}_{ij}) = \hat{\rho} \cdot (1 - \hat{x}_i - \hat{x}_j + 2\hat{z}_{ij}) = \hat{\rho} - \hat{y}_i - \hat{y}_j + 2\hat{y}_{ij} \quad \forall (i, j) \in E \setminus V. \quad (\text{EC.11c})$$

Edges connecting nodes of the same type correspond to weights $\hat{\rho}\hat{c}_{ij}$, while edges connecting primed and unprimed nodes correspond to weights $\hat{\rho}(1 - \hat{c}_{ij})$.

We then compute, for each node $i \in V$, the shortest path from i to i' in G' . By construction, any path from i to i' must traverse an odd number of ‘‘cross’’ edges (i.e., edges of type (u, v') or (u', v)). Hence, such a path corresponds exactly to selecting an odd number of edges weighted by $\hat{\rho}(1 - \hat{c}_{ij})$ in the original graph. For example, a feasible path $i \rightarrow j' \rightarrow i'$ contains one cross edge and one same-type edge, yielding a total length

$$\hat{\rho}(1 - \hat{c}_{ij}) + \hat{\rho}\hat{c}_{ij} = \hat{\rho},$$

which matches the scaled right-hand side and therefore does not violate the inequality. Consequently, if there exists a violated odd-cycle inequality (i.e., one whose left-hand side is strictly smaller than $\hat{\rho}$), then the shortest path over all pairs (i, i') will have length strictly less than $\hat{\rho}$ and corresponds to a violated cycle in the original graph.

Computing one shortest path requires $\mathcal{O}(|V|^2)$ operations. Repeating this for all $i \in V$ yields an overall separation complexity of $\mathcal{O}(|V|^3)$. Algorithm 3 summarizes this separation procedure.

When separating odd-cycle inequalities for the Big-M formulation or the Conic formulation, we require an unscaled length \mathbf{w} given as follows.

$$w_{io} = w_{i'o'} = \hat{c}_{io} = \hat{x}_i \quad \forall i \in V$$

$$w_{ij} = w_{i'o'} = \hat{c}_{ij} = \hat{x}_i + \hat{x}_j - 2\hat{z}_{ij} \quad \forall (i, j) \in E \setminus V$$

$$w_{ij'} = w_{i'j} = 1 - \hat{c}_{ij} = 1 - \hat{x}_i - \hat{x}_j + 2\hat{z}_{ij} \quad \forall (i, j) \in E \setminus V.$$

Accordingly, the length of the shortest path is compared with $1 - \epsilon$ of Step 4 in Algorithm 3.

Algorithm 3 Separation Oracle for Odd-cycle Inequality

Require: A point $(\hat{\rho}, \hat{\mathbf{y}})$; attraction value vector \mathbf{v} ; a threshold $\epsilon > 0$.

Ensure: A set of violated constraints, Cuts.

- 1: Initialize Cuts $\leftarrow \emptyset$.
- 2: Construct the extended graph G' and edge length \mathbf{w} via (EC.11).
- 3: Find shortest path $\text{path}(i)$ from i to i' over all $i \in V$.
- 4: **if** the length of path $\text{path}(i)$ is less than $\hat{\rho}(1 - \epsilon)$ **then**
- 5: $C \leftarrow (u, v)$ if (u, v) or $(u', v') \in \text{path}(i)$
- 6: $D \leftarrow (u, v)$ if (u, v') or $(u', v) \in \text{path}(i)$
- 7: Cuts \leftarrow (ODD-CYCLE) with cycle C and odd-number subset D .
- 8: **end if**
- 9: **return** Cuts.

EC.2.1.3. Separation Oracle for Running-Intersection Inequalities Finding a violated running-intersection inequality is nontrivial because a multi-purchase hypergraph can admit an exponential number of such inequalities. [Del Pia et al. \(2020\)](#) propose a separation oracle with complexity $O(|E|(d2^d|E| + 2^{d\bar{m}}\bar{m}^2d))$, where \bar{m} denotes the maximum number of neighbors of a central bundle e_0 considered by the algorithm. In many practical applications, the rank d is relatively small, i.e., $d \leq 3$ in our case study. Hence, this algorithm is polynomial in the number of candidate bundles $|E|$, making it computationally efficient for practical instances. We illustrate this algorithm for our perspective formulation as follows.

We begin by constructing candidate running-intersection structures. Given a multi-purchase hypergraph G , we identify a collection

$$S = \{(e_0, O_{e_0}, P_{e_0})\},$$

where e_0 is a central bundle. The set O_{e_0} consists of running-intersection orderings $\tilde{e}_1, \dots, \tilde{e}_k$ intersected with e_0 . A candidate running-intersection ordering $\tilde{e}_1, \dots, \tilde{e}_m$ satisfies: for any $f, g \in [m]$, $\tilde{e}_f \not\subset \tilde{e}_g$ and $\tilde{e}_f \not\supset \tilde{e}_g$. This condition is imposed to control the number of the resulting inequalities while preserving tightness. Intuitively, when $\tilde{e}_f \subset \tilde{e}_g$, the corresponding running-intersection inequality is redundant: it can be obtained by aggregating a running-intersection inequality involving only \tilde{e}_f together with a trivial constraint $z_{e_g} \leq z_{\nu_g}$, where $\nu \in N(\tilde{e}_g)$. Excluding such nested

intersections, therefore, leads to more compact inequalities without loss of strength. The mapping P_{e_0} assigns to each intersection \tilde{e}_k a set of neighbors, so that for any $e_k \in P_{e_0}(\tilde{e}_k)$, we have $e_k \cap e_0 = \tilde{e}_k$. In some cases, such as Figure 1c, $P_{e_0}(\tilde{e}_k)$ may include only one element. Algorithm 4 details the construction of all admissible running-intersection structures.

Algorithm 4 Construction of the Running Intersection Separation Structure

Require: A multi-purchase hypergraph G ; a maximum number of neighbors \bar{m} .

Ensure: Running-intersection structures.

- 1: Initialize structure $S \leftarrow \emptyset$.
 - 2: **for** $e_0 \in E$ **do**
 - 3: Initialize $O_{e_0} \leftarrow \emptyset$, $P_{e_0} \leftarrow \emptyset$, and the subgraph edge set $\tilde{E}_{e_0} \leftarrow \emptyset$.
 - 4: For each $e_k \in E \setminus \{e_0\}$, let $f \leftarrow e_0 \cap e_k$, then $\tilde{E} \leftarrow f$ and add e_k to $P_{e_0}(f)$.
 - 5: **for** each subset $H \subseteq \tilde{E}_{e_0}$ with $|H| \leq \bar{m}$ **do**
 - 6: **if** for any $\tilde{e}_f, \tilde{e}_g \in H$, $\tilde{e}_f \not\subseteq \tilde{e}_g$ and $\tilde{e}_f \not\supseteq \tilde{e}_g$. **then**
 - 7: Find a running-intersection order $(\tilde{e}_1, \dots, \tilde{e}_m)$ over H and $O_{e_0} \leftarrow (\tilde{e}_1, \dots, \tilde{e}_m)$.
 - 8: **end if**
 - 9: **end for**
 - 10: $S \leftarrow (e_0, O_{e_0}, P_{e_0})$
 - 11: **end for**
 - 12: **return** S .
-

Given running-intersection structures and a candidate solution $(\hat{\rho}, \hat{\mathbf{y}})$, the separation algorithm tries to find the most violated one. Recall the scaled running-intersection inequality is

$$\sum_{k \in [m]} y_{e_k} + \sum_{v \in e_0 \setminus \bigcup_{k \in [m]} e_k} y_v \leq z_{e_0} + \sum_{k \in [m] | N(e_0 \cap e_k) \neq \emptyset} y_{\mu_k} + \rho(\omega - 1).$$

For each central bundle e_0 and an ordering $\tilde{e}_1, \dots, \tilde{e}_m$, we can calculate ω by its definition. Hence, the separation task reduces to maximizing the left-hand side and minimizing the right-hand side variables. Specifically, for $k \in [m]$, we find the associated neighbor with the largest value

$$\bar{e}_k = \operatorname{argmax}_{c \in P_{e_0}(\tilde{e}_k)} \hat{y}_c,$$

and the smallest single element in the intersection:

$$\bar{\mu}_k = \operatorname{argmin}_{c \in N(\bar{e}_k)} \hat{y}_c.$$

If the corresponding LHS is larger than the RHS, then this running-intersection inequality should be added to the base model. We detail the separation process in Algorithm 5.

Algorithm 5 Separation Oracle for Running Intersection Inequality

Require: A point $(\hat{\rho}, \hat{\mathbf{y}})$; attraction value vector \mathbf{v} ; a threshold $\epsilon > 0$.

Ensure: A set of violated constraints, Cuts.

```

1: for  $e_0 \in \mathbf{E}$  do
2:   for  $(\tilde{e}_1, \dots, \tilde{e}_m) \in O_{e_0}$  do
3:      $\omega = |e_0 \setminus \bigcup_{k \in [m]} e_k| + |\{k \in [m] : N(\tilde{e}_k) = \emptyset\}|$ 
4:      $\bar{e}_k = \operatorname{argmax}_{c \in P_{e_0}(\tilde{e}_k)} \hat{y}_c$  for each  $k \in [m]$ 
5:      $\bar{\mu}_k = \operatorname{argmin}_{c \in N(\bar{e}_k)} \hat{y}_c$  for each  $k \in [m]$ 
6:     if  $\sum_{k \in [m]} \hat{y}_{\bar{e}_k} + \sum_{v \in e_0 \setminus \bigcup_{k \in [m]} \bar{e}_k} \hat{y}_v > \hat{y}_{e_0} + \sum_{k \in [m] | N(e_0 \cap \bar{e}_k) \neq \emptyset} \hat{y}_{\bar{\mu}_k} + \hat{\rho}(\omega - 1) + \epsilon$  then
7:       Cuts  $\leftarrow \sum_{k \in [m]} y_{\bar{e}_k} + \sum_{v \in e_0 \setminus \bigcup_{k \in [m]} \bar{e}_k} y_v \leq y_{e_0} + \sum_{k \in [m] | N(e_0 \cap \bar{e}_k) \neq \emptyset} y_{\bar{\mu}_k} + \rho(\omega - 1)$ 
8:     end if
9:   end for
10: end for
11: return Cuts.

```

EC.2.2. Cutting-Plane Implementation

We implement a cutting-plane procedure to solve optimization problems over (4), such as the assortment optimization problem with objective function $\sum_{i \in \mathbf{E}} r_e y_e$. We initialize the procedure with the exact base formulation (**Base**). The algorithm iteratively solves the continuous relaxation of the current model and augments it with valid inequalities identified by the separation oracles proposed in Section EC.2.1. We set the maximum iterations limit to $K = 500$ and the tolerance parameter to $\epsilon = 1 \times 10^{-5}$; a cut is added only if its violation exceeds this threshold. The process terminates when no new cuts are detected, the iteration limit is reached, or the cumulative separation time exceeds $T = 800$ seconds. Let Base_k denote the formulation obtained at the k^{th} iteration. We solve the final formulation with binary constraints to recover an optimal integer solution. Algorithm 6 details this procedure.

Algorithm 6 A Cutting-Plane Implementation

Require: A base formulation Base_0 (**Base**); maximum iterations K ; tolerance ϵ ; time limit T .

Ensure: An optimal integer solution.

- 1: Record $t_{\text{start}} \leftarrow \text{CurrentTime}$.
 - 2: **for** $k \leftarrow 1$ to K **do**
 - 3: $(\hat{\rho}, \hat{\mathbf{y}}, \hat{\mathbf{x}}) \leftarrow \text{SolveLP}(\text{Base}_{k-1})$.
 - 4: Cuts \leftarrow Call separation oracles Algorithm 2, Algorithm 3, and Algorithm 5 for solution $(\hat{\rho}, \hat{\mathbf{y}}, \hat{\mathbf{x}})$.
 - 5: $\text{Base}_k \leftarrow$ Add Cuts to Base_{k-1}
 - 6: **if** Cuts = \emptyset or $\text{CurrentTime} - t_{\text{start}} > T$ **then**
 - 7: **break;**
 - 8: **end if**
 - 9: **end for**
 - 10: Impose binary constraints on variables in Base_k and solve the MIP formulation.
 - 11: **return** $(\rho^*, \mathbf{y}^*, \mathbf{x}^*)$.
-