Revenue Management under a Mixture of Multinomial Logit and Independent Demand Models

Yufeng Cao[1], Paat Rusmevichientong[2], Huseyin Topaloglu[3]

[1]H. Milton Stewart School of Industrial and Systems Engineering, Georgia Tech, Atlanta, GA 30332
[2]Marshall School of Business, University of Southern California, Los Angeles, CA 90089
[3]School of Operations Research and Information Engineering, Cornell Tech, New York, NY 10044
yufeng.cao@gatech.edu, rusmevic@marshall.usc.edu, topaloglu@orie.cornell.edu

March 11, 2020

We consider assortment optimization problems when customers choose under a mixture of multinomial logit and independent demand models. In the single-shot assortment optimization problem, each product has a certain revenue associated with it. The customers choose among the products according to our mixture choice model. The goal is to find an assortment that maximizes the expected revenue from a customer. We show that we can find the optimal assortment by solving a linear program. We establish that the optimal assortment becomes larger as the relative size of the customer segment with the independent demand model increases. Moreover, we show that the Pareto-efficient assortments that maximize a weighted average of the expected revenue and the total purchase probability are nested in the sense that the Pareto-efficient assortments become larger as the weight on the total purchase probability increases. Considering the single-shot assortment optimization problem with a cardinality constraint on the offered assortment, we show that the problem is NP-hard. We give a fully polynomial-time approximation scheme. In the assortment-based network revenue management problem, we have resources with limited capacities and each product consumes a combination of resources. The goal is to find a policy for deciding which assortment of products to offer to each arriving customer to maximize the total expected revenue over a finite selling horizon. A standard linear programming approximation for this problem includes one decision variable for each subset of products. We show that this linear program can be reduced to an equivalent one of substantially smaller size. Our computational experiments indicate that using a mixture of multinomial logit and independent demand models can significantly improve our ability to capture the choice behavior of the customers. Furthermore, our linear programming formulation of smaller size can dramatically improve the computation times.
Keywords: Assortment optimization, choice model, multinomial logit, network revenue management.

1. Introduction

Over the past decade, the use of discrete choice models to capture the choice process of customers has received significant attention in the revenue management literature. By using discrete choice models, we can capture the fact that if a product is unavailable, then some customers may substitute for this product, whereas others may simply leave the system without making a purchase. A growing body of literature indicates that using choice models to capture the substitution possibilities between products can provide significant improvements in the expected revenues (Talluri and van Ryzin 2004, Vulcano et al. 2010, Dai et al. 2014). However, an inherent tension is involved in picking a choice model with which to capture the choice process of the customers. A more sophisticated choice model may capture the choice process of the customers more faithfully, whereas a simpler choice model may result in tractable optimization problems when finding the optimal assortment of products to offer or prices to charge.

We consider assortment optimization problems under a mixture of multinomial logit and independent demand models. The multinomial logit model is arguably one of the most prevalent choice models for capturing customer choice behavior. It is based on random utility maximization, so each customer associates a random utility with each product and the no-purchase option, choosing the available alternative with the largest utility. In the independent demand model, a customer arrives into the system with a particular product in mind. If this product is unavailable, then she leaves without a purchase. The independent demand model has been a reliable workhorse, because it is relatively simple to estimate and often yields tractable models for making operational decisions (van Ryzin 2005). In this paper, we mix these two very common demand models, which is, perhaps, the most natural approach to simultaneously improve the modeling flexibility of both the multinomial logit and independent demand models. Some customers make a purchase under the multinomial logit model, whereas others do so under the independent demand model. The demand emerges as a mixture of these two customer segments.

Main Contributions: We give algorithms for numerous assortment problems, characterize the structure of optimal assortments, and check the prediction effectiveness of our choice model.

Assortment Optimization. In the single-shot assortment optimization problem, we have a certain revenue for each product. Customers choose among the offered products according to our mixture choice model. The goal is to find an assortment of products that maximizes the expected revenue obtained from a customer. We show that we can solve a linear program (LP) to find the optimal assortment (Theorem 3.2). Thus, the assortment optimization problem under our mixture choice model is efficiently solvable. Assortment optimization problems under mixtures of choice models are notoriously difficult. For example, the assortment optimization problem under a mixture of just two multinomial logit models is NP-hard (Rusmevichientong et al. 2014). To our knowledge, our paper is the first to give an efficient method for assortment optimization under a mixture of choice models. Our LP has three novel components. First, it uses decision variables whose values depend on whether different pairs of products are offered. Second, its objective function is, on the surface, quite different from the objective function of the assortment problem. Third, at its extreme point solutions, the objective function gives expected revenues from different assortments.

Combinatorial Algorithm. We show that if a product, all else being equal, has a larger purchase probability in the independent demand model or a smaller preference weight in the multinomial logit model, then it becomes more attractive to offer in the optimal assortment (Theorem 3.3). Besides shedding light on the structure of the optimal assortment, this result allows us to give a combinatorial algorithm for assortment optimization. Although a combinatorial algorithm exists, our LP formulation ultimately becomes useful for network revenue management.

Comparative Statistics and Pareto-Efficient Assortments. We show that the optimal assortment becomes larger when the relative size of the customer segment with the independent demand model increases or when the revenue of each product increases by the same additive amount (Theorem 4.1). Both comparative statistics have useful implications. The customers purchasing under the independent demand model are inflexible. If the product they have in mind is unavailable, then they leave without a purchase. By the first comparative statistic, if the relative size of the inflexible customer segment increases, then it is optimal to offer a larger assortment. On the other hand, maximizing expected revenue is beneficial for the firm, whereas maximizing the total probability of purchase is beneficial for the customers, allowing a larger fraction of the customers to find a product in which they are interested. Using the second comparative statistic, we are able to argue that the Pareto-efficient assortments that maximize a weighted average of the expected revenue and the total probability of purchase are nested in the sense that the Pareto-efficient assortments become larger as the weight on the total probability of purchase increases.

Cardinality Constraints. We examine the single-shot assortment optimization problem with a cardinality constraint on the number of products that we can offer. We show that the problem is NP-hard (Theorem 5.1). It is surprising that our unconstrained assortment optimization problem is well-behaved to the extent that we can give an LP formulation for this problem, but introducing a cardinality constraint into such a well-behaved problem drastically changes its difficulty. Motivated by our complexity result, we give a fully polynomial-time approximation scheme (FPTAS) under a cardinality constraint (Theorem 5.2). Our FPTAS uses the connections of our assortment optimization problem to the cardinality-constrained knapsack problem. Other studies have used connections of various assortment optimization problems to the knapsack problem to give approximation schemes (Desir et al. 2016, Feldman and Topaloglu 2018). Our work follows a similar blueprint. We scale the preference weights of the products by the reciprocal of a fixed accuracy parameter and round them to take integer values. Thus, our main contribution for the cardinality-constrained assortment optimization problem is to establish its complexity.

Network Revenue Management. We consider assortment-based network revenue management problems, where we have resources with limited capacities and the sale of each product consumes a combination of resources. The goal is to find a policy for deciding which assortment of products to offer to each arriving customer to maximize the total expected revenue over a finite selling horizon. We consider a previously proposed LP approximation in which the decision variables are the probabilities with which we offer each subset of products to the customers. Thus, the number of decision variables increases exponentially with the number of products. We show that if the customers choose according to our mixture choice model, then we can immediately reduce the LP approximation to a compact LP whose numbers of decision variables and constraints increase only quadratically with the number of products (Theorem 6.1). We show that we can recover an optimal solution to the original LP approximation by using an optimal solution to the compact LP (Theorem 7.2). Lastly, in our computational experiments, we demonstrate that our mixture of multinomial logit and independent demand models can significantly improve our ability to predict customer purchase behavior compared to the pure multinomial logit benchmark. Also, our compact LP substantially reduces the computation times.

Our mixture choice model is a natural way to enrich the flexibility of both the multinomial logit and independent demand models, which are, perhaps, the two most prevalent demand models. Mixtures of choice models often yield computationally difficult assortment optimization problems, but under our mixture choice model, the problems remain surprisingly tractable. Our focus is mostly on assortment optimization, but our numerical work indicates that mixing the multinomial logit and independent demand models can yield better predictions of customer purchases.

Related Literature: Gallego et al. (2004) and Talluri and van Ryzin (2004) show that the optimal assortment under the multinomial logit model is revenue ordered, including a certain number of products with the largest revenues. This structure does not hold under our mixture choice model. Rusmevichientong et al. (2010), Wang (2012), and Jagabathula (2016) examine the assortment optimization problem under the multinomial logit model with various constraints on the offered assortment. Bront et al. (2009), Mendez-Diaz et al. (2014), and Rusmevichientong et al. (2014) show that the assortment optimization problem under a mixture of multinomial logit models is NP-hard even when there are only two multinomial logit models in the mixture. The authors give approximation schemes and integer programming formulations. Desir et al. (2016) show that it is NP-hard to approximate the problem within a factor of $O(1/m^{1-\epsilon})$ for any $\epsilon > 0$, where m is the number of multinomial logit models in the mixture.

Researchers have developed LP formulations for assortment optimization problems. Gallego et al. (2015) work with the generalized attraction model, whereas Feldman and Topaloglu (2017) work with the Markov chain choice model. Both papers give LP formulations for the assortment optimization problem. One can build on these LP formulations to obtain compact LP formulations for network revenue management problems. The multinomial logit model is a special case of both the generalized attraction and Markov chain choice models, but our mixture of multinomial logit and independent demand models is not a special case of these choice models. Thus, we resort to entirely different techniques to obtain the LP formulations in our paper. Topaloglu (2013) gives a compact formulation for a nonlinear program that appears when jointly making product stocking and assortment decisions under the multinomial logit model. Sumida et al. (2019) give an LP for assortment optimization under the multinomial logit model when there are constraints on the offered assortment that can be captured by a totally unimodular constraint matrix.

Motivated by online retail, in which customers examine search results page by page, Flores et al. (2019), Feldman and Segev (2019), and Liu et al. (2019) develop extensions of the multinomial logit model that allow the customers to incrementally view the products in batches. The authors give algorithms for finding the optimal sequence of product batches to offer. Wang and Sahin (2018), Feldman and Topaloglu (2018), and Aouad et al. (2019) incorporate consideration sets, where each customer focuses only on the set of products in her consideration set and chooses within the consideration set under the multinomial logit model. Aouad et al. (2018b) and Aouad and Segev (2019) focus on dynamic assortment optimization problems under the multinomial logit model, where the assortments offered to the customers are dictated by the inventory remaining on the shelf. We focus our literature review on the multinomial logit model, but assortment optimization has been studied under other choice models. For representative approaches, we refer to Farias et al. (2013), Aouad et al. (2016), Aouad et al. (2018a), and Feldman et al. (2019) for the preference-list-based choice model, Blanchet et al. (2016) for the Markov chain choice model, Davis et al. (2014), Gallego and Topaloglu (2014), Feldman and Topaloglu (2015), and Li et al. (2015) for the nested logit model, and Zhang et al. (2019) for the paired combinatorial logit model.

Incorporating customer choice into network revenue management problems is an active area of research. Gallego et al. (2004) and Liu and van Ryzin (2008) give an LP approximation for these problems. The number of decision variables in their LP approximation increases exponentially with the number of products. Under our mixture choice model, we are able to reduce the size of their LP dramatically. Other approaches to these problems are based on approximating the value functions. For such approaches, we refer to Zhang and Cooper (2005), Zhang and Adelman (2009), Kunnumkal and Topaloglu (2010), Tong and Topaloglu (2013), and Vossen and Zhang (2015).

Organization: In Section 2, we formulate our assortment optimization problem. In Section 3, we give the LP formulation for the problem. In Section 4, we give comparative statistics for the optimal assortment. In Section 5, we examine the problem with cardinality constraints. In Section 6, we give a compact LP for the network revenue management problem. In Section 7, we show that we can recover an optimal solution to the original LP approximation by using our compact LP. In Section 8, we present computational experiments. In Section 9, we give conclusions.

2. Problem Formulation

The set of products is $N=\{1,\ldots,n\}$. There are two customer segments. The customers in the first segment make a purchase according to the multinomial logit model. In the multinomial logit model, we use $v_i>0$ to denote the preference weight of product i. We normalize the preference weight of the no-purchase option to one. We let $V(S)=\sum_{i\in S} v_i$ to capture the total preference weight of the products in the subset $S\subseteq N$. In this case, if we offer the subset $S\subseteq N$ of products, then a customer in the first segment purchases product $i\in S$ with probability $v_i/(1+V(S))$. The customers in the second segment make a purchase according to the independent demand model. In the independent demand model, we use $\theta_i>0$ to denote the probability that a customer is interested in product i. In this case, if we offer the subset $S\subseteq N$ of products, then a customer in the second segment purchases product $i\in S$ with probability θ_i. The probability that an arriving customer is in the first segment is β. Thus, if we offer the subset $S\subseteq N$ of products, then a customer purchases product $i\in S$ with probability $\beta\,\frac{v_i}{1+V(S)}+(1-\beta)\,\theta_i$. For notational brevity, throughout the paper, we normalize the size of the first segment to one, in which case the size of the second segment relative to the first one is $\lambda=(1-\beta)/\beta$. Thus, if we offer the subset $S\subseteq N$ of products, then a customer purchases product $i\in S$ with the scaled probability $\frac{v_i}{1+V(S)}+\lambda\,\theta_i$. If a customer purchases product i, then we obtain a revenue of r_i. Our goal is to find a subset, or an assortment, to offer that maximizes the expected revenue from a customer, yielding the problem

$$\max_{S\subseteq N}\left\{\sum_{i\in S} r_i\left(\frac{v_i}{1+V(S)}+\lambda\,\theta_i\right)\right\}. \tag{Mixture}$$

Since we normalize the size of the first segment to one, we can have $\frac{v_i}{1+V(S)}+\lambda\,\theta_i>1$, but we can recover the purchase probabilities by scaling $\frac{v_i}{1+V(S)}+\lambda\,\theta_i$ for all $i\in N$ with β.

Working with such a mixture of the multinomial logit and independent demand models in the Mixture problem introduces nontrivial challenges. If we do not have the independent demand model in the mixture, then we can express the expected revenue under the multinomial logit model as a fraction, whose numerator and denominator are both linear functions, allowing us to use fractional programming techniques when solving the assortment optimization problem. We lose this fractional structure in the Mixture problem, but we will show that we can still solve this problem efficiently. Moreover, under only the multinomial logit model, there exists an optimal assortment that is revenue ordered, where we offer a certain number of products with the largest revenues. We lose the revenue-ordered structure of the optimal solution in the Mixture problem. In Table 1, considering a problem instance with $n=3$, $(r_1,r_2,r_3)=(50,10,5)$, $(v_1,v_2,v_3)=(0.5,5,0.01)$, $(\theta_1,\theta_2,\theta_3)=(0.05,0.25,0.7)$, and $\lambda=1$, we show the expected revenue provided by each assortment. The optimal

Assort.	Exp. Rev.	Assort.	Exp. Rev.
$\varnothing$	0	$\{1,2\}$	16.54
$\{1\}$	19.17	$\{1,3\}$	22.59
$\{2\}$	10.83	$\{2,3\}$	14.33
$\{3\}$	3.55	$\{1,2,3\}$	20.03

Table 1 Expected revenue provided by all possible assortments.

assortment is $\{1,3\}$, which does not offer the product with second largest revenue, but offers the product with the smallest revenue. In this problem instance, noting that $\theta_3 = 0.7$, the customer segment with the independent demand model is interested in product 3 with a relatively large probability, so we offer product 3 to exploit this relatively large probability. Moreover, noting that $v_2 = 5$, the customer segment with the multinomial logit model associates a relatively large preference weight with product 2, but the revenue of product 2 is much smaller than that of product 1. Thus, product 2, if offered, attracts a significant fraction of the customer segment with the multinomial logit model while providing much smaller revenue than product 1, so we do not offer product 2. In the next section, we show that, roughly speaking, an optimal solution to the Mixture problem prioritizes product i when θ_i/v_i is larger, so a larger value for θ_i and a smaller value for v_i make product i more attractive to offer, which is consistent with the observation from Table 1. Lastly, as discussed in the introduction, the multinomial logit model is a special case of the Markov chain choice model. An example in Appendix A shows that the mixture of multinomial logit and independent demand models is not a special case of the Markov chain choice model. Thus, existing results under the Markov chain choice model do not apply to our problem.

3. Assortment Optimization

In this section, we give an LP formulation for the Mixture problem and show that there exists an optimal solution that gives high priority to product i when θ_i/v_i is large. Using the decision variables x_0, $\boldsymbol{x} = \{x_i : i \in N\}$ and $\boldsymbol{y} = \{y_{ij} : i, j \in N\}$, we consider the LP

$$\max_{(x_0,\boldsymbol{x},\boldsymbol{y})\in\mathbb{R}\times\mathbb{R}_+^{n+n^2}} \left\{ \sum_{i\in N} r_i \left((v_i + \lambda\,\theta_i)\, x_i + \lambda\,\theta_i \sum_{j\in N} v_j\, y_{ij} \right) \ : \ \begin{array}{l} x_0 + \sum_{i\in N} v_i\, x_i = 1, \\ x_i \le x_0 \quad \forall\, i \in N, \\ y_{ij} \le x_i \quad \forall\, i,j\in N, \quad y_{ij} \le x_j \quad \forall\, i,j\in N \end{array} \right\}. \tag{Assortment LP}$$

Before showing that we can obtain an optimal solution to the Mixture problem by using the Assortment LP, we provide some intuition regarding the LP above. Using $\mathbf{1}(\cdot)$ to denote the indicator function, given a solution $\hat{S} \subseteq N$ to the Mixture problem, we construct a solution $(\hat{x}_0, \hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ to the

Assortment LP by setting $\hat{x}_0 = \frac{1}{1+V(\hat{S})}$, $\hat{x}_i = \mathbf{1}(i \in \hat{S})\,\hat{x}_0$, and $\hat{y}_{ij} = \mathbf{1}(i \in \hat{S},\ j \in \hat{S})\,\hat{x}_0$. Noting that $\sum_{i \in N} v_i\,\hat{x}_i = \hat{x}_0 \sum_{i \in N} v_i\,\mathbf{1}(i \in \hat{S}) = \hat{x}_0\,V(\hat{S})$, we have $\hat{x}_0 + \sum_{i \in N} v_i\,\hat{x}_i = \hat{x}_0\,(1 + V(\hat{S})) = 1$, so the solution $(\hat{x}_0, \hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ satisfies the first constraint in the Assortment LP. Moreover, since $\mathbf{1}(i \in \hat{S}) \le 1$, $\mathbf{1}(i \in \hat{S},\ j \in \hat{S}) \le \mathbf{1}(i \in \hat{S})$, and $\mathbf{1}(i \in \hat{S},\ j \in \hat{S}) \le \mathbf{1}(j \in \hat{S})$, the solution $(\hat{x}_0, \hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ satisfies the remaining constraints in the Assortment LP as well. Furthermore, for the Assortment LP, this solution provides an objective value of

$$\begin{aligned}
\sum_{i \in N} r_i \left((v_i + \lambda\,\theta_i)\,\hat{x}_i + \lambda\,\theta_i \sum_{j \in N} v_j\,\hat{y}_{ij} \right) &= \sum_{i \in N} r_i \left((v_i + \lambda\,\theta_i)\,\mathbf{1}(i \in \hat{S}) + \lambda\,\theta_i \sum_{j \in N} v_j\,\mathbf{1}(i \in \hat{S},\ j \in \hat{S}) \right) \hat{x}_0 \\
&= \sum_{i \in N} r_i \left(v_i + \lambda\,\theta_i + \lambda\,\theta_i \sum_{j \in N} v_j\,\mathbf{1}(j \in \hat{S}) \right) \mathbf{1}(i \in \hat{S})\,\hat{x}_0 \\
&= \sum_{i \in N} r_i \Big(v_i + \lambda\,\theta_i\,(1 + V(\hat{S})) \Big)\,\mathbf{1}(i \in \hat{S})\,\hat{x}_0 \;=\; \sum_{i \in N} r_i \left(\frac{v_i}{1 + V(\hat{S})} + \lambda\,\theta_i \right) \mathbf{1}(i \in \hat{S}),
\end{aligned}$$

which is the objective function of the Mixture problem evaluated at $\hat{S}$. Thus, given a solution to the Mixture problem, we can construct a feasible solution to the Assortment LP, and the objective values of the two solutions match. To show that the Assortment LP is equivalent to the Mixture problem, we need to show the converse statement as well, which is what we do next. Note how collecting the terms $\lambda\,\theta_i\,\hat{x}_i$ and $\lambda\,\theta_i \sum_{j \in N} \hat{y}_{ij}$ above surprisingly yields the purchase probability of product i in the customer segment with the independent demand model.

To establish the converse statement, we build on the next lemma, which shows an important property of the basic feasible solutions to the Assortment LP.

Lemma 3.1 (Extreme Point Solutions) *Let $(\hat{x}_0, \hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ be a basic feasible solution to the* Assortment LP*. Then, we have $\hat{x}_i \in \{0, \hat{x}_0\}$ for all $i \in N$.*

The proof of Lemma 3.1 follows by showing that if $\hat{x}_i \in (0, \hat{x}_0)$ for some $i \in N$, then we can perturb the solution $(\hat{x}_0, \hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ to obtain two feasible solutions to the Assortment LP such that $(\hat{x}_0, \hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ is a convex combination of the two feasible solutions. Note that perturbing $\hat{x}_i$ may require perturbing the other elements of the solution $(\hat{x}_0, \hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ to ensure feasibility. We give the proof in Appendix B. In the next theorem, we use the above lemma to show that we can obtain an optimal solution to the Mixture problem by using an optimal solution to the Assortment LP.

Theorem 3.2 (LP Formulation) *For a basic optimal solution $(x_0^*, \boldsymbol{x}^*, \boldsymbol{y}^*)$ to the* Assortment LP*, let $S^* = \{i \in N : x_i^* > 0\}$. Then, S^* is an optimal solution to the* Mixture *problem.*

Proof: Let $\hat{S}$ be an optimal solution to the Mixture problem providing the optimal objective value $\hat{z}$, and let z^*_{LP} be the optimal objective value of the Assortment LP. By the discussion at the beginning of this section, given the solution $\hat{S}$ to the Mixture problem, we can construct a feasible solution to the Assortment LP with the objective value of $\hat{z}$. Therefore, we have $z^*_{\mathsf{LP}} \geq \hat{z}$. On the other hand, by Lemma 3.1, we have $x^*_i = x^*_0$ for all $i \in S^*$ and $x^*_i = 0$ for all $i \in N \setminus S^*$. Since $(x^*_0, \boldsymbol{x}^*, \boldsymbol{y}^*)$ is a feasible solution to the Assortment LP, by the first constraint, we get $x^*_0 + \sum_{i \in S^*} v_i\, x^*_0 = 1$, so $x^*_0 = \frac{1}{1+V(S^*)} = x^*_i$ for all $i \in S^*$. In this case, by the last two constraints, we also get $y^*_{ij} \leq \frac{1}{1+V(S^*)}$ for all $i, j \in S^*$. If $i \notin S^*$ or $j \notin S^*$, then $x^*_i = 0$ or $x^*_j = 0$, so we have $y^*_{ij} = 0$. Let $Q^* = \{i \in S^* : r_i < 0\}$, and recall that z^*_{LP} is the optimal objective value of the Assortment LP and $\hat{z}$ is the optimal objective value of the Mixture problem. Since $x^*_i = 0$ when $i \notin S^*$ and $y^*_{ij} = 0$ when $i \notin S^*$ or $j \notin S^*$, evaluating the objective function of the Assortment LP at its optimal solution, we get

$$
\begin{aligned}
z^*_{\mathsf{LP}} &= \sum_{i \in S^*} r_i \left((v_i + \lambda\, \theta_i)\, x^*_i + \lambda\, \theta_i \sum_{j \in S^*} v_j\, y^*_{ij} \right) \\
&= \sum_{i \in S^* \setminus Q^*} r_i \left((v_i + \lambda\, \theta_i)\, x^*_i + \lambda\, \theta_i \sum_{j \in S^*} v_j\, y^*_{ij} \right) + \sum_{i \in Q^*} r_i \left((v_i + \lambda\, \theta_i)\, x^*_i + \lambda\, \theta_i \sum_{j \in S^*} v_j\, y^*_{ij} \right) \\
&\stackrel{(a)}{\leq} \sum_{i \in S^* \setminus Q^*} r_i \left(\frac{v_i + \lambda\, \theta_i}{1 + V(S^*)} + \lambda\, \theta_i \sum_{j \in S^*} \frac{v_j}{1 + V(S^*)} \right) + \sum_{i \in Q^*} r_i \left((v_i + \lambda\, \theta_i)\, x^*_i + \lambda\, \theta_i \sum_{j \in S^*} v_j\, y^*_{ij} \right) \\
&= \sum_{i \in S^* \setminus Q^*} r_i \left(\frac{v_i}{1 + V(S^*)} + \lambda\, \theta_i \right) + \sum_{i \in Q^*} r_i \left((v_i + \lambda\, \theta_i)\, x^*_i + \lambda\, \theta_i \sum_{j \in S^*} v_j\, y^*_{ij} \right) \\
&\leq \sum_{i \in S^* \setminus Q^*} r_i \left(\frac{v_i}{1 + V(S^* \setminus Q^*)} + \lambda\, \theta_i \right) + \sum_{i \in Q^*} r_i \left((v_i + \lambda\, \theta_i)\, x^*_i + \lambda\, \theta_i \sum_{j \in S^*} v_j\, y^*_{ij} \right) \\
&\stackrel{(b)}{\leq} \hat{z} + \sum_{i \in Q^*} r_i \left((v_i + \lambda\, \theta_i)\, x^*_i + \lambda\, \theta_i \sum_{j \in S^*} v_j\, y^*_{ij} \right).
\end{aligned}
$$

Here, (a) holds since $r_i \geq 0$ and $x^*_i = \frac{1}{1+V(S^*)}$ for all $i \in S^* \setminus Q^*$ and $y^*_{ij} \leq \frac{1}{1+V(S^*)}$, whereas (b) holds since $S^* \setminus Q^*$ is a feasible but not necessarily an optimal solution to the Mixture problem.

Noting that $r_i < 0$ for all $i \in Q^*$, we have $\sum_{i \in Q^*} r_i\, ((v_i + \lambda\, \theta_i)\, x^*_i + \lambda\, \theta_i \sum_{j \in S^*} v_j\, y^*_{ij}) \leq 0$. If $\sum_{i \in Q^*} r_i\, ((v_i + \lambda\, \theta_i)\, x^*_i + \lambda\, \theta_i \sum_{j \in S^*} v_j\, y^*_{ij}) < 0$, then the above chain of inequalities yields $z^*_{\mathsf{LP}} < \hat{z}$, contradicting the fact that $z^*_{\mathsf{LP}} \geq \hat{z}$, which we established at the beginning of the proof. Therefore, we have $\sum_{i \in Q^*} r_i\, ((v_i + \lambda\, \theta_i)\, x^*_i + \lambda\, \theta_i \sum_{j \in S^*} v_j\, y^*_{ij}) = 0$, but since $x^*_i > 0$, $r^*_i < 0$ and $v_i + \lambda\, \theta_i > 0$ for all $i \in Q^*$, for the last equality to hold, we must have $Q^* = \varnothing$. Since $Q^* = \varnothing$, noting that $z_{\mathsf{LP}} \geq \hat{z}$, all the inequalities in the above chain of inequalities hold as equalities. In particular, since (b) holds as an equality and $Q^* = \varnothing$, the objective value provided by the solution $S^* \setminus Q^* = S^*$ for the Mixture problem is $\hat{z}$, so S^* is an optimal solution to the Mixture problem. ∎

The proof of Theorem 3.2 also shows that the Mixture problem and the Assortment LP have the same optimal objective values. In the proof of the theorem, we do not assume that the revenues of the products are non-negative. We will build on Theorem 3.2 when examining network revenue management problems. In that setting, the revenues of the products will be adjusted by the opportunity costs of the capacities used by the products, in which case, the revenues of some of the products can be negative. Nevertheless, when we focus only on solving the Mixture problem, we can a priori drop from consideration all products with nonpositive revenues, since if we drop such products, then the expected revenue from any assortment stays at least as large.

Prioritization of Products in an Optimal Assortment:

We give a result to intuitively suggest that there exists an optimal solution to the Mixture problem that prioritizes product i when r_i is larger or when θ_i/v_i is larger. Besides providing insight into the structure of the optimal assortment, this result allows us to construct a combinatorial algorithm for solving the Mixture problem. We start by reformulating the Mixture problem. Define the decision variables $\boldsymbol{w} = \{w_i : i \in N\} \in \{0,1\}^n$, where $w_i = 1$ if and only if we offer product i. Using z^* to denote the optimal objective value of the Mixture problem, we write this problem as

$$\begin{aligned} z^* &= \max_{\boldsymbol{w}\in\{0,1\}^n} \left\{ \sum_{i\in N} r_i \left(\frac{v_i}{1+\sum_{j\in N} v_j\, w_j} + \lambda\,\theta_i \right) w_i \right\} \\ &= \max_{\boldsymbol{w}\in\{0,1\}^n} \left\{ \frac{\sum_{i\in N} r_i\,(v_i + \lambda\,\theta_i)\, w_i + \left(\sum_{i\in N} v_i\, w_i\right)\left(\sum_{i\in N} r_i\,\lambda\,\theta_i\, w_i\right)}{1+\sum_{i\in N} v_i\, w_i} \right\}, \end{aligned} \tag{1}$$

where the second equality follows by arranging the terms. As a function of $\boldsymbol{w}$, let $G(\boldsymbol{w})$ be the expression in the numerator of the fraction on the right side of (1).

By (1), we have $z^* \geq \frac{G(\boldsymbol{w})}{1+\sum_{i\in N} v_i\, w_i}$ for all $\boldsymbol{w} \in \{0,1\}^n$, and the inequality holds as equality at the optimal solution $\boldsymbol{w}^*$ to the Mixture problem. Thus, for all $\boldsymbol{w} \in \{0,1\}^n$, we have

$$\begin{aligned} z^* &\geq G(\boldsymbol{w}) - z^* \sum_{i\in N} v_i\, w_i = \sum_{i\in N} r_i\,(v_i + \lambda\,\theta_i)\, w_i + \left(\sum_{i\in N} v_i\, w_i\right)\left(\sum_{i\in N} r_i\,\lambda\,\theta_i\, w_i\right) - z^* \sum_{i\in N} v_i\, w_i \\ &= \sum_{i\in N} v_i \left(r_i + \lambda\, r_i\, \frac{\theta_i}{v_i} - z^* \right) w_i + \left(\sum_{i\in N} v_i\, w_i\right)\left(\sum_{i\in N} r_i\,\lambda\,\theta_i\, w_i\right), \end{aligned}$$

where the first equality follows by using the definition of $G(\boldsymbol{w})$. Once again, the inequality above holds for all $\boldsymbol{w} \in \{0,1\}^n$, and it holds as an equality at the optimal solution $\boldsymbol{w}^*$ to the Mixture problem. Thus, an optimal solution to the Mixture problem is a maximizer of the expression on the right side above over all $\boldsymbol{w} \in \{0,1\}^n$. In other words, letting $F(\boldsymbol{w}) = \sum_{i\in N} v_i\,(r_i + \lambda\, r_i\, \frac{\theta_i}{v_i} - z^*)\, w_i + \left(\sum_{i\in N} v_i\, w_i\right)\left(\sum_{i\in N} r_i\,\lambda\,\theta_i\, w_i\right)$ capture the expression on the right side above as a function of $\boldsymbol{w}$, $\boldsymbol{w}^*$ is an optimal solution to the problem $\max_{\boldsymbol{w}\in\{0,1\}^n} F(\boldsymbol{w})$.

In the next theorem, we use the discussion in the previous paragraph to provide insight into the structure of an optimal solution to the Mixture problem.

Theorem 3.3 (Prioritization of Products) *There exists an optimal solution S^* to the* Mixture *problem that satisfies*

$$\min_{i\in S^*}\left\{ r_i\left(1+\lambda\,\frac{\theta_i}{v_i}\,(1+V(S^*))\right)\right\} \;>\; \max_{i\in N\setminus S^*}\left\{ r_i\left(1+\lambda\,\frac{\theta_i}{v_i}\,(1+V(S^*))\right)\right\}.$$

Proof: By the discussion preceding the theorem, letting $\boldsymbol{w}^*$ be an optimal solution to problem (1), we have $\boldsymbol{w}^* = \arg\max_{\boldsymbol{w}\in\{0,1\}^n} F(\boldsymbol{w})$. If we drop all products with nonpositive revenues from an assortment, then the expected revenue from the assortment stays at least as large, so we focus on the case where $r_i > 0$ for all $i \in N$. For each $i \in N$, the only term in $F(\boldsymbol{w})$ that depends on w_i in a nonlinear fashion is $v_i\, r_i\, \lambda\, \theta_i\, w_i^2$. Thus, $F(\boldsymbol{w})$ is directionally convex. In this case, we can relax the binary constraints to get $\boldsymbol{w}^* = \arg\max_{\boldsymbol{w}\in[0,1]^n} F(\boldsymbol{w})$. Let $V^* = \sum_{i\in N} v_i\, w_i^*$ and $\Theta^* = \sum_{i\in N} r_i\, \lambda\, \theta_i\, w_i^*$ for notational brevity. Differentiating $F(\boldsymbol{w})$ by using its definition, we get

$$\begin{aligned}\frac{\partial F(\boldsymbol{w})}{\partial w_i}\bigg|_{\boldsymbol{w}=\boldsymbol{w}^*} &= v_i\left(r_i + \lambda\, r_i\,\frac{\theta_i}{v_i} - z^*\right) + v_i\left(\sum_{j\in N} r_j\,\lambda\,\theta_j\, w_j^*\right) + \left(\sum_{j\in N} v_j\, w_j^*\right) r_i\,\lambda\,\theta_i \\ &= v_i\left(r_i\left(1+\lambda\,\frac{\theta_i}{v_i}\,(1+V^*)\right) - z^* + \Theta^*\right).\end{aligned}$$

We use $f_i(\boldsymbol{w}^*)$ to denote the derivative above. To show the result by contradiction, assume that $w_k^* = 1$ and $w_\ell^* = 0$ for some $k, \ell \in N$, and $r_k(1+\lambda\,\frac{\theta_k}{v_k}\,(1+V^*)) \le r_\ell(1+\lambda\,\frac{\theta_\ell}{v_\ell}\,(1+V^*))$.

Since $\boldsymbol{w}^* = \arg\max_{\boldsymbol{w}\in[0,1]^n} F(\boldsymbol{w})$ and $w_k^* = 1$, we have $f_k(\boldsymbol{w}^*) \ge 0$. Otherwise, a small decrease in w_k strictly increases the value of $F(\boldsymbol{w}^*)$. Similarly, we have $f_\ell(\boldsymbol{w}^*) \le 0$. Thus, we obtain

$$\frac{f_\ell(\boldsymbol{w}^*)}{v_\ell} \;=\; r_\ell\left(1+\lambda\,\frac{\theta_\ell}{v_\ell}\,(1+V^*)\right) - z^* + \Theta^* \;\le\; 0 \;\le\; r_k\left(1+\lambda\,\frac{\theta_k}{v_k}\,(1+V^*)\right) - z^* + \Theta^* = \frac{f_k(\boldsymbol{w}^*)}{v_k}.$$

In this case, noting that $r_k(1+\lambda\,\frac{\theta_k}{v_k}\,(1+V^*)) \le r_\ell(1+\lambda\,\frac{\theta_\ell}{v_\ell}\,(1+V^*))$, all the inequalities above must hold as equalities. In particular, we have $r_\ell(1+\lambda\,\frac{\theta_\ell}{v_\ell}\,(1+V^*)) - z^* + \Theta^* = 0$.

Define the solution $\hat{\boldsymbol{w}} = \{\hat{w}_i : i \in N\}$ as $\hat{w}_i = w_i^*$ for all $i \in N \setminus \{\ell\}$ and $\hat{w}_\ell = 1$. Using the fact that the solutions $\hat{\boldsymbol{w}}$ and $\boldsymbol{w}^*$ differ only in the decision variable w_ℓ, we have

$$\begin{aligned}F(\hat{\boldsymbol{w}}) - F(\boldsymbol{w}^*) &= v_\ell\left(r_\ell + \lambda\, r_\ell\,\frac{\theta_\ell}{v_\ell} - z^*\right) + (V^* + v_\ell)(\Theta^* + r_\ell\,\lambda\,\theta_\ell) - V^*\,\Theta^* \\ &= v_\ell\left(r_\ell\left(1+\lambda\,\frac{\theta_\ell}{v_\ell}\,(1+V^*)\right) - z^* + \Theta^*\right) + v_\ell\, r_\ell\,\lambda\,\theta_\ell \;\overset{(a)}{=}\; v_\ell\, r_\ell\,\lambda\,\theta_\ell \;>\; 0,\end{aligned}$$

where (a) holds since $r_\ell(1+\lambda\,\frac{\theta_\ell}{v_\ell}\,(1+V^*)) - z^* + \Theta^* = 0$. Having $F(\hat{\boldsymbol{w}}) - F(\boldsymbol{w}^*) > 0$ contradicts the fact that $\boldsymbol{w}^*$ is an optimal solution to the problem $\max_{\boldsymbol{w}\in[0,1]^n} F(\boldsymbol{w})$. ■

By the theorem above, if product i has a larger value for r_i or θ_i/v_i, then the optimal assortment prioritizes this product. Besides providing insight into the structure of the optimal assortment, we can use Theorem 3.3 to construct a combinatorial algorithm for solving the Mixture problem. If we knew the value of $V(S^*)$, then letting $t = V(S^*)$, we could index the products such that $r_1(1+\lambda\frac{\theta_1}{v_1}(1+t)) \geq r_2(1+\lambda\frac{\theta_2}{v_2}(1+t)) \geq \ldots \geq r_n(1+\lambda\frac{\theta_n}{v_n}(1+t))$, in which case an optimal assortment would be of the form $\{1,\ldots,i\}$ for some $i \in N$. Thus, we would obtain an optimal assortment by checking the expected revenue from $O(n)$ candidate assortments, each of which is of the form $\{1,\ldots,i\}$ for some $i \in N$. To deal with the fact that we do not know the value of $V(S^*)$, we adopt an approach from Rusmevichientong et al. (2010). Note that $g_i(t) = r_i(1+\lambda\frac{\theta_i}{v_i}(1+t))$ is a linear function of t. The n lines $\{g_i(\cdot) : i \in N\}$ intersect at $O(n^2)$ points. Let $t^1 \leq t^2 \leq \ldots \leq t^K$ with $K = O(n^2)$ be the intersection points of the n lines $\{g_i(\cdot) : i \in N\}$. That is, for each $k = 1,\ldots,K$, we have $g_i(t^k) = g_j(t^k)$ for some $i, j \in N$. Letting $t^0 = 0$ and $t^{K+1} = \infty$ for notational uniformity, for each $k = 0,\ldots,K$, if t takes values in the interval $[t^k, t^{k+1})$, then the ordering between the values $\{g_i(t) : i \in N\}$ does not change. To capture this ordering, we let the permutation $(\sigma_1^k,\ldots,\sigma_n^k) \in N^n$ be such that $g_{\sigma_1^k}(t) \geq g_{\sigma_2^k}(t) \geq \ldots \geq g_{\sigma_n^k}(t)$ for all $t \in [t^k, t^{k+1})$. In this case, we can obtain the optimal assortment by checking the expected revenue from $O(n^3)$ candidate assortments, each of which is of the form $\{\sigma_1^k,\ldots,\sigma_i^k\}$ for some $i \in N$, $k = 0,\ldots,K$.

Thus, we can solve the Mixture problem without using an LP, but the Assortment LP becomes critical when we work with network revenue management problems.

4. Effect of Customer Segment Mix and Efficient Assortments

In this section, we show that the optimal solution to the Mixture problem becomes a larger assortment when the relative size of the customer segment with the independent demand model grows or when the revenue of each product increases by the same additive amount. The fact that the optimal solution becomes a larger assortment when the revenue of each product increases by the same additive amount has important implications when we want to find assortments that trade off expected revenue with the probability of purchase. This result also becomes useful when implementing policies in dynamic assortment optimization problems through protection levels. Throughout this section, we focus on the Mixture problem after increasing the revenue of each product by δ, which is given by

$$\max_{S \subseteq N} \left\{ \sum_{i \in S} (r_i + \delta) \left(\frac{v_i}{1+V(S)} + \lambda\,\theta_i \right) \right\}. \qquad \text{(Parametric Mixture)}$$

As a function of (λ, δ), we let $S^*(\lambda, \delta)$ be an optimal solution to the problem above. If there are multiple optima, then we choose any one that has the largest cardinality.

In the next theorem, we examine how an optimal solution to the Parametric Mixture problem changes as a function of λ and δ.

Theorem 4.1 (Sensitivity of the Optimal Assortment) *There exists an optimal solution $S^*(\lambda,\delta)$ to the* Parametric Mixture *problem that satisfies the following properties.*

(a) If $\lambda > \beta$, then $S^*(\lambda,\delta) \supseteq S^*(\beta,\delta)$.

(b) $\delta > 0$, then $S^*(\lambda,\delta) \supseteq S^*(\lambda,0)$.

Before we give a proof for the theorem, we discuss its implications. To interpret the first part of the theorem, note that a customer in the segment with the independent demand model is not willing to make substitutions. If such a customer is interested in product i and this product is unavailable, then she leaves without a purchase. In that sense, the customers in the segment with the independent demand model are inflexible. By the first part of the theorem, as the relative size of the inflexible customer segment increases, to ensure that the customers in this segment can find the product they are interested in, the optimal assortment becomes larger. To interpret the second part of the theorem, letting $\mathsf{Rev}(S) = \sum_{i\in S} r_i\big(\frac{v_i}{1+V(S)} + \lambda\,\theta_i\big)$ and $\mathsf{Dem}(S) = \sum_{i\in S}\big(\frac{v_i}{1+V(S)} + \lambda\,\theta_i\big)$, we write the objective function of the Parametric Mixture problem as $\mathsf{Rev}(S) + \delta\,\mathsf{Dem}(S)$. Observe that $\mathsf{Rev}(S)$ is the expected revenue from assortment S, whereas $\mathsf{Dem}(S)$ is the total purchase probability from assortment S. While maximizing the expected revenue is beneficial for the firm, maximizing the total purchase probability is beneficial for the customers, as a larger total purchase probability indicates that a larger fraction of the customers can find a product in which they are interested. The parameter δ characterizes the weight that we put on the total purchase probability relative to the expected revenue. Solving the problem $\max_{S\subseteq N}\big\{\mathsf{Rev}(S) + \delta\,\mathsf{Dem}(S)\big\}$ for all possible values of δ, we can construct an efficient frontier of all attainable expected revenue-total purchase probability pairs. By the second part of the theorem, the Pareto-efficient assortments that lie on the efficient frontier have a nested structure, one assortment always being included in another one. Since there can be at most n assortments that satisfy such a nested structure, there can be at most n assortments on the efficient frontier. Furthermore, the Pareto-efficient assortments get larger as the weight that we put on the total purchase probability increases. In Figure 1, we consider a problem instance with 8 products and show the expected revenue-total purchase probability pairs for the assortments on the efficient frontier. We label each assortment by the products that are included in the assortment. The assortments on the efficient frontier are nested.

Having nested Pareto-efficient assortments has other important implications. Talluri and van Ryzin (2004) investigate dynamic assortment optimization problems with a single resource, where we offer assortments of products to customers arriving over time and the sale of a product generates a revenue depending on the purchased product and consumes one unit of capacity of the resource. It turns out that if the assortments on the efficient frontier are nested, then we

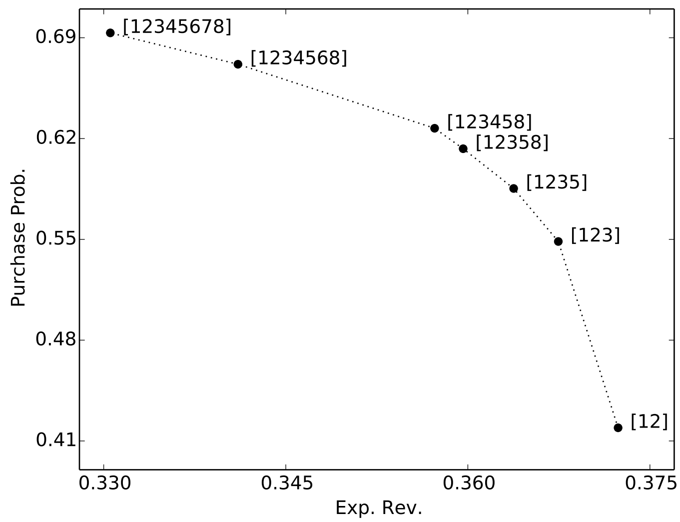

Figure 1 Efficient frontier with $n=8$, $(r_1,\ldots,r_8)=(0.96,0.81,0.34,0.29,0.19,0.09,0.04,0.03)$, $(v_1,\ldots,v_8)=(0.40,0.35,0.54,0.10,0.11,0.21,0.12,0.04)$, $(\theta_1,\ldots,\theta_8)=(0.15,0.20,0.09,0,0.16,0.15,0.07,0.18)$, $\lambda=\frac{3}{22}$.

can implement the optimal policy by associating a protection level for each product. In this case, it is optimal to offer a product only when the remaining capacity of the resource exceeds the protection level of the product. On the other hand, Ma (2019) studies assortment auctions, where each buyer submits a list of options she is willing to purchase and the seller allocates a limited amount of inventory to buyers under only probabilistic information about the preferences of the buyers. Having nested assortments on the efficient frontier plays an important role in giving a Myersonian characterization of the optimal mechanism in these auctions.

Proof of Theorem 4.1:

In the rest of this section, we give a proof of Theorem 4.1. Results similar to this theorem appear in the literature under the pure multinomial logit model, but our result, which is under a mixture of multinomial logit and independent demand models, requires techniques that, to our knowledge, have not been used. Consider two instances of the Mixture problem. In both instances, the preference weight of product i in the multinomial logit model is $v_i>0$ and a customer in the independent demand segment is interested in product i with probability $\theta_i>0$. The two instances differ in the revenues of the products and the relative size of the customer segment with the independent demand model. In the first instance, the revenue of product i is r_{i1} and the relative size of the independent demand segment is λ_1, whereas in the second instance, the revenue of product i is r_{i2} and the relative size of the independent demand segment is λ_2. For $\ell=1,2$, we use $R_\ell(S)$ to denote the expected revenue from the segment with the multinomial logit model in instance ℓ, given that we offer the assortment S. In other words, we have $R_\ell(S)=\sum_{i\in S} r_{i\ell}\, v_i/(1+V(S))$. Lastly, for

$\ell = 1,2$, we let S_ℓ^* be an optimal solution to the Mixture problem for instance ℓ. In particular, S_ℓ^* is an optimal solution to the problem

$$\max_{S\subseteq N}\left\{\frac{\sum_{i\in N} r_{i\ell}\, v_i}{1+V(S)} + \lambda_\ell \sum_{i\in N} r_{i\ell}\,\theta_i\right\} \;=\; \max_{S\subseteq N}\left\{R_\ell(S) + \lambda_\ell \sum_{i\in S} r_{i\ell}\,\theta_i\right\}.$$

If there are multiple optimal solutions, then we choose any one that has the largest cardinality. We break other ties according to a fixed lexicographical order between the assortments.

The way we will use the two instances is as follows. When showing the first part of Theorem 4.1, we will have $r_{i1} = r_i = r_{i2}$ and $\lambda_1 = \lambda > \beta = \lambda_2$, so that the two instances will differ only in the relative size of the customer segment with the independent demand model. When showing the second part of Theorem 4.1, we will have $r_{i1} = r_i$, $r_{i2} = r_i + \delta$ and $\lambda_1 = \lambda = \lambda_2$, so that two instances will differ only in the revenues of the products. In the next lemma, we give an important relationship between S_1^* and S_2^*. This result, rather than having an intuition, serves as a wrapper for a key computation that we make multiple times when giving a proof for Theorem 4.1.

Lemma 4.2 (Comparing Instances) *For $\ell = 1,2$, let S_ℓ^* be an optimal solution to the problem $\max_{S\subseteq N}\{R_\ell(S) + \lambda_\ell \sum_{i\in S} r_{i\ell}\,\theta_i\}$ and*

$$K = \{i\in N : r_{i1} \le R_1(S_1^*),\ i\notin S_1^*,\ i\in S_2^*\}.$$

Then, the assortments S_1^*, S_2^*, and K satisfy $S_1^*\cup K \supseteq S_2^*$. Furthermore, if $K\neq\varnothing$ and $r_{i2}\le R_2(S_2^*)$ for all $i\in K$, then we have

$$\frac{R_1(S_1^*) - \sum_{i\in K} r_{i1}\, v_i/V(K)}{1+V(S_1^*\cup K)} - \lambda_1\,\frac{\sum_{i\in K} r_{i1}\,\theta_i}{V(K)} \;\ge\; \frac{R_2(S_2^*) - \sum_{i\in K} r_{i2}\, v_i/V(K)}{1+V(S_2^*)} - \lambda_2\,\frac{\sum_{i\in K} r_{i2}\,\theta_i}{V(K)}. \quad (2)$$

Proof: To show the first statement, we write the objective function of the Mixture problem for instance ℓ as $R_\ell(S) + \lambda_\ell \sum_{i\in S} r_{i\ell}\,\theta_i$. As in the proof of Theorem 3.3, we consider the case $r_{i1} > 0$ for all $i\in N$. Otherwise, we can drop all products with nonpositive revenues while ensuring that the expected revenue from any assortment stays just as large. Consider $i\in S_2^*$. If $i\notin S_1^*$ and $r_{i1} > R_1(S_1^*)$, then a standard lemma for the expected revenue under the multinomial logit model, given as Lemma C.1 in Appendix C, implies that $R_1(S_1^*\cup\{i\}) \ge R_1(S_1^*)$. Moreover, since $r_{i1}\ge 0$ for all $i\in N$, we have $\sum_{j\in S_1^*\cup\{i\}} r_{j1}\,\theta_j \ge \sum_{j\in S_1^*} r_{j1}\,\theta_j$. Thus, we get $R_1(S_1^*\cup\{i\}) + \lambda_1 \sum_{j\in S_1^*\cup\{i\}} r_{j1}\,\theta_j \ge R_1(S_1^*) + \lambda_1 \sum_{j\in S_1^*} r_{j1}\,\theta_j$, which contradicts the fact that S_1^* is an optimal solution to the Mixture problem for the first instance with the largest cardinality. Therefore, if $i\in S_2^*$, then we have $i\in S_1^*$ or $r_{i1}\le R_1(S_1^*)$. Consider $i\in S_2^*$ and focus on two cases. First, if $i\in S_1^*$, then we trivially have $i\in S_1^*\cup K$. Second, if $i\notin S_1^*$, then we must have $r_{i1}\le R_1(S_1^*)$ by the earlier discussion in this paragraph, but by the definition of K, having $i\in S_2^*$, $i\notin S_1^*$ and $r_{i1}\le R_1(S_1^*)$ implies that $i\in K$, so we get $i \in S_1^* \cup K$. In both cases, having $i \in S_2^*$ implies that $i \in S_1^* \cup K$, so we have $S_1^* \cup K \supseteq S_2^*$. To show the second statement, we consider the case $K \neq \varnothing$ and $r_{i2} \leq R_2(S_2^*)$ for all $i \in K$.

Note that $S_1^* \cap K = \varnothing$ by the definition of K. Moreover, by the definition of S_1^*, we have $R_1(S_1^*) + \lambda_1 \sum_{i \in S_1^*} r_{i1}\, \theta_i \geq R_1(S_1^* \cup K) + \lambda_1 \sum_{i \in S_1^* \cup K} r_{i1}\, \theta_i$, which we write as

$$\begin{aligned}
0 \;\leq\; R_1(S_1^*) - R_1(S_1^* \cup K) - \lambda_1 \sum_{i \in K} r_{i1}\, \theta_i \;&=\; R_1(S_1^*) - \frac{\sum_{i \in S_1^* \cup K} r_{i1}\, v_i}{1 + V(S_1^* \cup K)} - \lambda_1 \sum_{i \in K} r_{i1}\, \theta_i \\
&\overset{(a)}{=}\; R_1(S_1^*) - \left(\frac{1 + V(S_1^*)}{1 + V(S_1^* \cup K)}\, R_1(S_1^*) + \frac{V(K)}{1 + V(S_1^* \cup K)}\, \frac{\sum_{i \in K} r_{i1}\, v_i}{V(K)} \right) - \lambda_1 \sum_{i \in K} r_{i1}\, \theta_i \\
&\overset{(b)}{=}\; \frac{V(K)}{1 + V(S_1^* \cup K)} \left(R_1(S_1^*) - \frac{\sum_{i \in K} r_{i1}\, v_i}{V(K)} \right) - \lambda_1 \sum_{i \in K} r_{i1}\, \theta_i,
\end{aligned}$$

where (a) uses the fact that $R_1(S_1^*) = \sum_{i \in S_1^*} r_{i1}\, v_i / (1 + V(S_1^*))$ and (b) uses the fact that $V(S_1^* \cup K) = V(S_1^*) + V(K)$ for $S_1^* \cap K = \varnothing$ and arranges the terms.

We have $K \subseteq S_2^*$ by the definition of K. Moreover, by the definition of S_2^*, we have $R_2(S_2^*) + \lambda_2 \sum_{i \in S_2^*} r_{i2}\, \theta_i \geq R_2(S_2^* \setminus K) + \lambda_2 \sum_{i \in S_2^* \setminus K} r_{i2}\, \theta_i$, which is equivalent to

$$\begin{aligned}
0 \;\geq\; R_2(S_2^* \setminus K) - R_2(S_2^*) - \lambda_2 \sum_{i \in K} r_{i2}\, \theta_i \;&=\; R_2(S_2^* \setminus K) - \frac{\sum_{i \in S_2^*} r_{i2}\, v_i}{1 + V(S_2^*)} - \lambda_2 \sum_{i \in K} r_{i2}\, \theta_i \\
&\overset{(c)}{=}\; R_2(S_2^* \setminus K) - \left(\frac{1 + V(S_2^* \setminus K)}{1 + V(S_2^*)}\, R_2(S_2^* \setminus K) + \frac{V(K)}{1 + V(S_2^*)}\, \frac{\sum_{i \in K} r_{i2}\, v_i}{V(K)} \right) - \lambda_2 \sum_{i \in K} r_{i2}\, \theta_i \\
&\overset{(d)}{=}\; \frac{V(K)}{1 + V(S_2^*)} \left(R_2(S_2^* \setminus K) - \frac{\sum_{i \in K} r_{i2}\, v_i}{V(K)} \right) - \lambda_2 \sum_{i \in K} r_{i2}\, \theta_i \\
&\overset{(e)}{\geq}\; \frac{V(K)}{1 + V(S_2^*)} \left(R_2(S_2^*) - \frac{\sum_{i \in K} r_{i2}\, v_i}{V(K)} \right) - \lambda_2 \sum_{i \in K} r_{i2}\, \theta_i \;,
\end{aligned}$$

where (c) and (d) use the same argument as do (a) and (b), whereas (e) follows by using, once again, Lemma C.1, which implies that since $r_{i2} \leq R_2(S_2^*)$ for all $i \in K$, we have $R_2(S_2^* \setminus K) \geq R_2(S_2^*)$.

Combining the two chains of inequalities displayed above and dividing both sides of the inequality by $V(K)$, the second statement in the lemma holds. ■

Here is the proof of Theorem 4.1. As in the proof of Lemma 4.2, it is sufficient to consider the case where $r_i > 0$ for all $i \in N$. To show the first part of Theorem 4.1, building on the notation introduced immediately before Lemma 4.2, we define two instances of the Mixture problem with $r_{i1} = r_i = r_{i2}$ for all $i \in N$ and $\lambda_1 = \lambda$, $\lambda_2 = \beta$ with $\lambda > \beta$. Note that since the revenues and preference weights in the two instances are the same, we have $R_1(S) = R_2(S)$ for all $S \subseteq N$. Recall that S_ℓ^*

is an optimal solution to the problem $\max_{S\subseteq N}\left\{R_\ell(S)+\lambda_\ell\sum_{i\in S} r_{i\ell}\,\theta_i\right\}$, and thus, showing that $S^*(\lambda,\delta)\supseteq S^*(\beta,\delta)$ is equivalent to showing that $S_1^*\supseteq S_2^*$. By the definition of S_1^* and S_2^*, we have

$$\begin{aligned} R_1(S_1^*)+\lambda_1\sum_{i\in S_1^*} r_{i1}\,\theta_i \;&\geq\; R_1(S_2^*)+\lambda_1\sum_{i\in S_2^*} r_{i1}\,\theta_i\\ R_2(S_2^*)+\lambda_2\sum_{i\in S_2^*} r_{i2}\,\theta_i \;&\geq\; R_2(S_1^*)+\lambda_2\sum_{i\in S_1^*} r_{i2}\,\theta_i. \end{aligned}$$

Multiplying the first and second inequalities above by λ_2 and λ_1, respectively, adding them, and noting that $r_{i1}=r_{i2}$, we have $\lambda_1\,R_2(S_2^*)-\lambda_2\,R_1(S_2^*)\geq\lambda_1\,R_2(S_1^*)-\lambda_2\,R_1(S_1^*)$.

Since $R_1(S)=R_2(S)$ for all $S\subseteq N$, the last inequality yields $(\lambda_1-\lambda_2)\,R_2(S_2^*)\geq(\lambda_1-\lambda_2)\,R_1(S_1^*)$, but noting that $\lambda_1>\lambda_2$, we get $R_2(S_2^*)\geq R_1(S_1^*)$. Define K as in Lemma 4.2, in which case we have $S_1^*\cup K\supseteq S_2^*$ by the first statement in Lemma 4.2. If $K=\varnothing$, then $S_1^*\supseteq S_2^*$, which is the desired result. To get a contradiction, assume that $K\neq\varnothing$. By the definition of K, we have $r_{i1}\leq R_1(S_1^*)$ for all $i\in K$. In this case, since $R_2(S_2^*)\geq R_1(S_1^*)$ and $r_{i1}=r_{i2}$ for all $i\in N$, we obtain $r_{i2}=r_{i1}\leq R_1(S_1^*)\leq R_2(S_2^*)$ for all $i\in K$. Observe that $\sum_{i\in K} r_{i1}\,v_i/V(K)=\sum_{i\in K} r_{i2}\,v_i/V(K)$ is a weighted average of the revenues of the products in K. Thus, since $r_{i1}=r_{i2}\leq R_1(S_1^*)\leq R_2(S_2^*)$ for all $i\in K$, we get $\sum_{i\in K} r_{i1}\,v_i/V(K)=\sum_{i\in K} r_{i2}\,v_i/V(K)\leq R_1(S_1^*)\leq R_2(S_2^*)$, which we equivalently write as $0\leq R_1(S_1^*)-\sum_{i\in K} r_{i1}\,v_i/V(K)\leq R_2(S_2^*)-\sum_{i\in K} r_{i2}\,v_i/V(K)$. Lastly, noting that $\lambda_1>\lambda_2$, $K\neq\varnothing$, and $r_{i1}=r_{i2}=r_i>0$ for all $i\in N$, we have $\lambda_1\sum_{i\in K}\theta_i\,r_{i1}>\lambda_2\sum_{i\in K}\theta_i\,r_{i2}$.

In the previous paragraph, we argue that $r_{i2}\leq R_2(S_2^*)$ for all $i\in K$, so the second statement in Lemma 4.2 holds. Thus, adding the inequality $\lambda_1\sum_{i\in K}\theta_i\,r_{i1}>\lambda_2\sum_{i\in K}\theta_i\,r_{i2}$ and (2), we get

$$\frac{R_1(S_1^*)-\sum_{i\in K} r_{i1}\,v_i/V(K)}{1+V(S_1^*\cup K)} \;>\; \frac{R_2(S_2^*)-\sum_{i\in K} r_{i2}\,v_i/V(K)}{1+V(S_2^*)}.$$

Since $0\leq R_1(S_1^*)-\sum_{i\in K} r_{i1}\,v_i/V(K)\leq R_2(S_2^*)-\sum_{i\in K} r_{i2}\,v_i/V(K)$, the inequality above holds only if $1+V(S_1^*\cup K)<1+V(S_2^*)$, which contradicts the fact that $S_1^*\cup K\supseteq S_2^*$.

To show the second part of Theorem 4.1, we define two instances of the Mixture problem with $r_{i1}=r_i$ and $r_{i2}=r_i+\delta$ for all $i\in N$ and $\lambda_1=\lambda=\lambda_2$. In this case, it is unfortunately not straightforward to show that $R_2(S_2^*)\geq R_1(S_1^*)$, and the proof of the second part becomes more complicated. We focus on two cases. First, considering the case $R_2(S_2^*)-\delta\leq R_1(S_1^*)$, we use an outline very similar to the one in the proof of the first part to show that $S_2^*\supseteq S_1^*$. Second, considering the case $R_2(S_2^*)-\delta>R_1(S_1^*)$, we show that this case cannot happen, once again, using an outline very similar to the one in the proof of the first part. The analysis of both of these cases uses Lemma 4.2. We give the detailed proof of the second part in Appendix C.

In the next section, we focus on the assortment optimization problem when there is a cardinality constraint that limits the number of offered products.

5. Assortment Optimization under a Cardinality Constraint

We show that the Mixture problem under a cardinality constraint is NP-hard and give an FPTAS. Letting C be the limit on the number of offered products, we consider the problem

$$\max_{\substack{S \subseteq N: \\ |S|=C}} \left\{ \sum_{i \in S} r_i \left(\frac{v_i}{1+V(S)} + \lambda\, \theta_i \right) \right\}. \tag{Cardinality-Mixture}$$

Note that the constraint above ensures that we offer exactly C products. If we want to offer at most k products, then we can solve the problem above with $C = 1, \ldots, k$.

To show that the Cardinality-Mixture problem is NP-hard, we use the following feasibility version of the Cardinality-Mixture problem, which we refer to as the Mixture Feasibility problem.

Mixture Feasibility: Given an instance of the Cardinality-Mixture problem and a threshold K, is there an assortment $S \subseteq N$ with $|S| = C$ that provides an expected revenue of K or more?

We use a reduction from the following Cardinality-Constrained Partition problem, which is a well-known NP-hard problem (Section A3.2, Garey and Johnson 1979).

Cardinality-Constrained Partition: Given a set of items $N = \{1, \ldots, n\}$ and their weights $\{w_i : i \in N\}$, is there a subset $S \subseteq N$ with $|S| = n/2$ such that $\sum_{i \in S} w_i = \sum_{i \notin S} w_i$?

To our knowledge, no study has built on the partition problem to establish NP-hardness for assortment optimization under a cardinality constraint, especially considering that we can solve the unconstrained assortment optimization problem in polynomial time and the problem without the cardinality constraint is well-behaved to the point that we can use an LP. In that respect, our NP-hardness result is surprising and unique. Rusmevichientong et al. (2014) use the partition problem to establish NP-hardness under a mixture of multinomial logit models without constraints. In the next theorem, we show that the Cardinality-Mixture problem is NP-hard.

Theorem 5.1 (Complexity) *The* Mixture Feasibility *problem is NP-complete.*

Proof: Given an instance of the Cardinality-Constrained Partition problem with the set of items $N = \{1, \ldots, n\}$ and weights $\{w_i : i \in N\}$, we have $\sum_{i \in S} w_i = \sum_{i \notin S} w_i$ if and only if $\sum_{i \in S} w_i = \frac{1}{2} \sum_{i \in N} w_i$. Scaling all weights by the same constant does not change the answer to the Cardinality-Constrained Partition problem, so we scale the weights so that $\sum_{i \in N} w_i = 2$. From the instance of the Cardinality-Constrained Partition problem, we define an instance of the Mixture Feasibility problem as follows. For each $i \in N$, the revenue of product i is $r_i = 1$. The preference weight of product i is $v_i = w_i$. Letting $V_{\max} = \max_{i \in N} v_i$, the purchase probability of product i in the independent demand model is $\theta_i = V_{\max} - v_i$. The relative size of the two customer segments is $\lambda = \frac{1}{4}$. We need to offer $C = \frac{n}{2}$ products. Finally, the threshold is $K = \frac{1}{4} + \frac{1}{4} C\, V_{\max}$.

We show that there exists an assortment $S \subseteq N$ with $|S| = C$ that provides an expected revenue of K or more if and only if there exists a subset of items $S \subseteq N$ with $|S| = n/2$ such that $\sum_{i\in S} w_i = \frac{1}{2}\sum_{i\in N} w_i$. Noting that $r_i = 1$ for all $i \in N$ and $\lambda = \frac{1}{4}$, the objective function of the Cardinality-Mixture problem is $\frac{V(S)}{1+V(S)} + \frac{1}{4}\sum_{i\in S}\theta_i$. Since we have $\theta_i = V_{\max} - v_i$ and we need to offer an assortment $S \subseteq N$ that satisfies $|S| = C$, this objective function is equivalent to $\frac{V(S)}{1+V(S)} + \frac{1}{4}\sum_{i\in S}(V_{\max} - v_i) = \frac{V(S)}{1+V(S)} + \frac{1}{4} C\, V_{\max} - \frac{1}{4} V(S)$.

Thus, noting that $K = \frac{1}{4} + \frac{1}{4} C\, V_{\max}$, an assortment $S \subseteq N$ with $|S| = C$ provides an expected revenue of K or more if and only if a subset of items $S \subseteq N$ with $|S| = C$ satisfies

$$\frac{V(S)}{1+V(S)} + \frac{1}{4} C\, V_{\max} - \frac{1}{4} V(S) \geq \frac{1}{4} + \frac{1}{4} C\, V_{\max}.$$

The inequality above holds if and only if $\frac{V(S)}{1+V(S)} \geq \frac{1}{4}(1 + V(S))$. Arranging the terms, the last inequality is equivalent to $(1 - V(S))^2 \leq 0$, which holds if and only if $V(S) = 1$.

By the discussion in the previous two paragraphs, there exists an assortment $S \subseteq N$ with $|S| = C$ that provides an expected revenue of K or more if and only if there exists a subset of items $S \subseteq N$ with $|S| = C$ such that $V(S) = 1$. Noting that $C = n/2$, $V(S) = \sum_{i\in S} v_i = \sum_{i\in S} w_i$ and $\sum_{i\in N} w_i = 2$, having a subset of items $S \subseteq N$ with $|S| = C$ and $V(S) = 1$ is equivalent to having a subset of items $S \subseteq N$ with $|S| = n/2$ and $\sum_{i\in S} w_i = \frac{1}{2}\sum_{i\in N} w_i$. The last statement is precisely the one that the Cardinality-Constrained Partition problem focuses on. ■

Considering the Cardinality-Mixture problem, our main contribution is to show that this problem is NP-hard, as we do in the theorem above. Once we show that the Cardinality-Mixture problem is NP-hard, it is natural to consider developing an FPTAS for this problem, and our approach to developing an FPTAS uses standard arguments that are employed when developing an FPTAS for the knapsack problem (Desir et al. 2016). In the rest of this section, we outline the three main steps of our FPTAS and defer the details to Appendix D.

In the first step, we relate the Cardinality-Mixture problem to the knapsack problem. Letting $V_{\min} = \min_{i\in N} v_i$ and $V_{\max} = \max_{i\in N} v_i$, for fixed $t \in [V_{\min}, n\, V_{\max}]$, we consider the problem

$$f(t) \;=\; \max_{S\subseteq N}\left\{ \sum_{i\in S} r_i\left(\frac{v_i}{1+t} + \lambda\,\theta_i\right) \;:\; V(S) \leq t, \quad |S| = C \right\}. \tag{3}$$

The problem above is a cardinality-constrained knapsack problem, where the set of items is N, the utility of item i is $r_i\,(\frac{v_i}{1+t} + \lambda\,\theta_i)$, the space consumption of item i is v_i, the capacity of the knapsack is t, and the number of items to carry is C. If $t < V_{\min}$, then the only potentially feasible solution to the problem above is $S = \varnothing$, whereas if $t > n\,V_{\max}$, then the first constraint above is redundant, so we consider the values of $t \in [V_{\min}, n\,V_{\max}]$. Noting that $f(t)$ is the optimal objective value of problem (3), let t^* be an optimal solution to the problem $\max_{t\in[V_{\min}, n\,V_{\max}]} f(t)$. We can show that if S^* is an optimal solution to problem (3) with $t = t^*$, then S^* is an optimal solution to the Cardinality-Mixture problem. In other words, we can recover an optimal solution to the Cardinality-Mixture problem by solving problem (3) with $t = t^*$. However, solving the problem $\max_{t\in[V_{\min}, n\,V_{\max}]} f(t)$ to compute t^* is difficult, since $f(\cdot)$ is itself hard to compute.

In the second step, we build a geometric grid over the interval $[V_{\min}, n\,V_{\max}]$. For a fixed accuracy parameter $\rho > 0$, we focus on a set of grid points that are integer powers of $1+\rho$, which are given by $\mathsf{Grid} = \{(1+\rho)^k : k = \lfloor \frac{\log V_{\min}}{\log(1+\rho)} \rfloor, \ldots, \lceil \frac{\log(n\,V_{\max})}{\log(1+\rho)} \rceil\}$. Here, $\lfloor\cdot\rfloor$ and $\lceil\cdot\rceil$ are, respectively, the round down and up functions. Let $\hat{t} \in \mathsf{Grid}$ be such that $t^* \le \hat{t} < (1+\rho)\,t^*$. We can show that if $\hat{S}$ is an optimal solution to problem (3) with $t = \hat{t}$, then the expected revenue of the assortment $\hat{S}$ deviates from the optimal expected revenue by at most a factor of $1+\rho$. Thus, there exists $\hat{t} \in \mathsf{Grid}$ such that solving problem (3) with $t = \hat{t}$ yields a $(1+\rho)$-approximate solution. However, solving problem (3) with $t = \hat{t}$ is difficult, since this problem is itself an NP-hard problem.

In the third step, we use a dynamic program to obtain an approximate solution to problem (3) with $t = \hat{t}$. Writing the constraint $V(S) = \sum_{i\in S} v_i \le t$ as $\sum_{i\in S} \frac{n}{t\rho}\, v_i \le \frac{n}{\rho}$, we consider an approximate version of problem (3) by replacing the constraint $V(S) \le t$ with $\sum_{i\in S} \lceil \frac{n}{t\rho}\, v_i \rceil \le \lceil \frac{n}{\rho} \rceil + n$. In the approximate version of problem (3), the space consumption of each item and the capacity of the knapsack are integers. Thus, we can solve the approximate version by using a dynamic program. We can show that if we let $\hat{t}$ be as in the previous paragraph and $\tilde{S}$ be an optimal solution to the approximate version of problem (3) with $t = \hat{t}$, then the expected revenue of the assortment $\tilde{S}$ deviates from the optimal expected revenue by at most a factor of $1+5\rho$.

In the next theorem, accounting for the number of operations and the errors in the second and third steps, we give an FPTAS for the Cardinality-Mixture problem.

Theorem 5.2 (FPTAS for Cardinality Constraint) *Letting z^* be the optimal objective value of the* Cardinality-Mixture *problem, there exists an algorithm where, for any $\epsilon \in (0,1)$, the algorithm runs in $O\Big(\frac{n^3}{\epsilon^2} \log\Big(\frac{n\,V_{\max}}{V_{\min}}\Big)\Big)$ operations and returns an assortment that is feasible to the* Cardinality-Mixture *problem with an expected revenue of at least $(1-\epsilon)\,z^*$.*

The proof of Theorem 5.2 is given in Appendix D.

6. Network Revenue Management

In the network revenue management setting, we have a set of resources, each with limited capacity. There is a finite number of time periods in the selling horizon. We choose the assortment of products to offer at each time period. The customer arriving at a time period chooses among the products according to a choice model. If the customer purchases a product, then we consume the capacities of a combination of resources and generate a revenue, both of which depend on the purchased product. The goal is to find a policy for choosing the assortment of products to offer at each time period to maximize the total expected revenue over the selling horizon. Gallego et al. (2004) and Liu and van Ryzin (2008) formulate an LP approximation for the network revenue management problem. In their LP approximation, we have one decision variable for each assortment of products that we can offer to the customers, capturing the frequency with which we offer each assortment. Thus, the number of decision variables increases exponentially with the number of products, and it may be computationally cumbersome to solve the LP approximation.

In this section, we show that if the customers choose according to a mixture of multinomial logit and independent demand models, then we can formulate an equivalent LP whose number of decision variables and constraints increases only quadratically with the number of products. The optimal objective values of the two formulations are equal, and we can recover an optimal solution to one formulation by using an optimal solution to the other. To pin down the network revenue management problem, let T be the number of time periods in the selling horizon. At each time period, we have one customer arrival. The set of resources is $M = \{1, \ldots, m\}$. The capacity of resource q is c_q. The set of products is $N = \{1, \ldots, n\}$. If we offer the assortment $S \subseteq N$ of products, then a customer purchases product $i \in S$ with probability $\frac{v_i}{1+V(S)} + \lambda\,\theta_i$, in which case we generate a revenue of r_i and consume a_{qi} units of the capacity of resource q. We use the decision variables $\boldsymbol{w} = \{\boldsymbol{w}(S) : S \subseteq N\}$, where $w(S)$ is the probability that we offer assortment S at a time period. The LP approximation for the network revenue management problem is

$$\max_{\boldsymbol{w} \in \mathbb{R}_+^{2^n}} \left\{ T \sum_{S \subseteq N} \sum_{i \in S} r_i \left(\frac{v_i}{1+V(S)} + \lambda\,\theta_i \right) w(S) \;\; : \right. \qquad \text{(Choice-Based LP)}$$

$$T \sum_{S \subseteq N} \sum_{i \in S} a_{qi} \left(\frac{v_i}{1+V(S)} + \lambda\,\theta_i \right) w(S) \le c_q \quad \forall\, q \in M,$$

$$\left. \sum_{S \subseteq N} w(S) = 1 \right\}.$$

Noting that $\sum_{i \in S} r_i \big(\frac{v_i}{v_0 + V(S)} + \lambda\,\theta_i \big)$ is the expected revenue at a time period at which we offer assortment S, the objective function above is the total expected revenue over the selling horizon.

The first constraint ensures that the total expected capacity consumption of a resource does not exceed its capacity. The second constraint ensures that we offer an assortment at each time period.

To give an equivalent formulation for the Choice-Based LP, using the decision variables $(x_0, \boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R} \times \mathbb{R}_+^{n+n^2}$ as in the Assortment LP, we consider the problem

$$\max_{(x_0,\boldsymbol{x},\boldsymbol{y})\in\mathbb{R}\times\mathbb{R}_+^{n+n^2}} \Bigg\{ T \sum_{i\in N} r_i \left((v_i + \lambda\,\theta_i)\,x_i + \lambda\,\theta_i \sum_{j\in N} v_j\,y_{ij} \right) \; : \qquad \text{(Compact LP)}$$
$$T \sum_{i\in N} a_{qi} \left((v_i + \lambda\,\theta_i)\,x_i + \lambda\,\theta_i \sum_{j\in N} v_j\,y_{ij} \right) \le c_q \quad \forall\, q\in M,$$
$$x_0 + \sum_{i\in N} v_i\,x_i = 1,$$
$$x_i \le x_0 \quad \forall\, i\in N,$$
$$y_{ij} \le x_i \quad \forall\, i,j\in N, \quad y_{ij} \le x_j \quad \forall\, i,j \in N \Bigg\}.$$

The objective function of the Compact LP is the total expected revenue over the selling horizon. It turns out that we can use the expression $(v_i + \lambda\,\theta_i)\,x_i + \lambda\,\theta_i \sum_{j\in N} v_j\,y_{ij}$ to capture the expected number of purchases for product i at a time period. Remarkably, we can capture the expected number of purchases for the products by associating one decision variable with each product, as well as with each pair of products, rather than by associating one decision variable with each assortment of products. The first constraint ensures that the total expected capacity consumption of a resource does not exceed its capacity. The remaining constraints ensure that the choices of the customers are governed by our mixture choice model. We have 2^n decision variables and $m+1$ constraints in the Choice-Based LP, but n^2+n+1 decision variables and $2n^2+n+m+1$ constraints in the Compact LP. While solving the Choice-Based LP almost always requires using column generation, we may directly solve the Compact LP without using column generation.

There are heuristic policies that use an optimal primal or dual solution to the Choice-Based LP to decide which assortment of products to offer at each time period. In a randomized offer policy, letting $\boldsymbol{w}^*$ be an optimal solution to the Choice-Based LP, we offer assortment S with probability $w^*(S)$, after adjusting the offered assortment to accommodate the availabilities of the resources (Jasin and Kumar 2012). On the other hand, in a bid-price policy, letting $\boldsymbol{\mu}^* = \{\mu_q^* : q \in M\}$ be the optimal values of the dual variables associated with the first constraint in the Choice-Based LP, we use μ_q^* to capture the opportunity cost of a unit of resource q. If a customer purchases product i, then the opportunity cost of the resources used by product i is $\sum_{q\in M} a_{qi}\,\mu_q^*$, so the net expected revenue from the purchase is $r_i - \sum_{q\in M} a_{qi}\,\mu_q^*$. Therefore, the expected net revenue from offering assortment S is given by $\sum_{i\in S}(\frac{v_i}{1+V(S)} + \lambda\,\theta_i)(r_i - \sum_{q\in M} a_{qi}\,\mu_q^*)$, in which case we offer an assortment that maximizes this expected net revenue, once again after adjusting the assortment to accommodate the availabilities of the resources (Zhang and Adelman 2009). The discussion in this paragraph indicates that it is important to recover both optimal primal and dual solutions to the Choice-Based LP by using the Compact LP. In the next theorem, we show that the optimal objective values of the Choice-Based LP and Compact LP are equal and we can recover an optimal dual solution to the former by using the latter. In the next section, we focus on recovering an optimal primal solution to the Choice-Based LP by using the Compact LP. It turns out that relating the primal solutions of the two formulations will require more work. Lastly, we note that the choice models that govern the choices of the customers at all time periods in the Choice-Based LP have the same parameters. If the choices of the customers at different time periods are governed by choice models with different parameters, then we need to work with the decision variables $\{w_t(S) : S \subseteq N,\ t = 1, \ldots, T\}$, where $w_t(S)$ is the probability that we offer assortment S at time period t. All of our results continue to hold in this case.

Theorem 6.1 (Equivalence of LP Formulations) *The optimal objective values of the* Choice-Based LP *and* Compact LP *are the same. Furthermore, the optimal values of the dual variables for the first constraint in the* Choice-Based LP *and* Compact LP *are the same.*

Proof: Note that Compact LP is feasible and bounded, since setting $x_0 = 1,\ x_i = 0$ for all $i \in N$ and $y_{ij} = 0$ for all $i, j \in N$ yields a feasible solution and all decision variables are bounded. The last four constraints in the Compact LP are equivalent to the four constraints in the Assortment LP. For notational brevity, we capture the polytope defined by these constraints as $\mathcal{P} = \{(x_0, \boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R} \times \mathbb{R}_+^{n+n^2} : x_0 + \sum_{i \in N} v_i\, x_i = 1,\ x_i \le x_0\ \forall\, i \in N,\ y_{ij} \le \min\{x_i, x_j\}\ \forall\, i, j \in N\}$. We construct the Lagrangian for the Compact LP by associating the dual multipliers $\boldsymbol{\mu} = \{\mu_q : q \in M\}$ with the first constraint and relaxing this constraint, so the Lagrangian is

$$\begin{aligned} L(x_0, \boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\mu}) &= \sum_{i \in N} T\, r_i \left((v_i + \lambda\, \theta_i)\, x_i + \lambda\, \theta_i \sum_{j \in N} v_j\, y_{ij} \right) \\ &\qquad + \sum_{q \in M} \mu_q \left(c_q - \sum_{i \in N} T\, a_{qi} \left((v_i + \lambda\, \theta_i)\, x_i + \lambda\, \theta_i \sum_{j \in N} v_j\, y_{ij} \right) \right) \\ &= \sum_{i \in N} T \left(r_i - \sum_{q \in M} a_{iq}\, \mu_q \right) \left((v_i + \lambda\, \theta_i)\, x_i + \lambda\, \theta_i \sum_{j \in N} v_j\, y_{ij} \right) + \sum_{q \in M} c_q\, \mu_q. \end{aligned}$$

In this case, using $D(\boldsymbol{\mu})$ to denote the dual function for the Compact LP as a function of the dual multipliers $\boldsymbol{\mu}$, we have $D(\boldsymbol{\mu}) = \max_{(x_0, \boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}} L(x_0, \boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\mu})$.

The Compact LP is feasible and bounded, so strong duality holds. Thus, we can obtain the optimal objective value of the Compact LP by solving the dual problem $\min_{\boldsymbol{\mu} \in \mathbb{R}_+^m} D(\boldsymbol{\mu})$, and an optimal solution to the last problem gives the optimal values of the dual variables for the first constraint in the Compact LP. We write the dual function as

$$\begin{aligned}
D(\boldsymbol{\mu}) &= \max_{(x_0,\boldsymbol{x},\boldsymbol{y})\in\mathcal{P}} \left\{ \sum_{i\in N} T \left(r_i - \sum_{q\in M} a_{iq}\,\mu_q \right) \left((v_i + \lambda\,\theta_i)\,x_i + \lambda\,\theta_i \sum_{j\in N} v_j\,y_{ij} \right) \right\} \; + \; \sum_{q\in M} \mu_q\,c_q \\
&\overset{(a)}{=} \max_{S\subseteq N} \left\{ \sum_{i\in S} T \left(r_i - \sum_{q\in M} a_{iq}\,\mu_q \right) \left(\frac{v_i}{1+V(S)} + \lambda\,\theta_i \right) \right\} \; + \; \sum_{q\in M} \mu_q\,c_q \\
&\overset{(b)}{=} \max_{\boldsymbol{w}\in\mathbb{R}^{2^n}_+} \left\{ \sum_{S\subseteq N} \sum_{i\in S} T \left(r_i - \sum_{q\in M} a_{iq}\,\mu_q \right) \left(\frac{v_i}{1+V(S)} + \lambda\,\theta_i \right) w(s) : \sum_{S\subseteq N} w(s) = 1 \right\} \; + \; \sum_{q\in M} \mu_q\,c_q \\
&= \max_{\boldsymbol{w}\in\mathbb{R}^{2^n}_+} \left\{ \sum_{S\subseteq N} \sum_{i\in S} T\,r_i \left(\frac{v_i}{1+V(S)} + \lambda\,\theta_i \right) w(s) \right. \\
&\qquad\qquad \left. + \sum_{q\in M} \mu_q \left(c_q - \sum_{S\subseteq N} \sum_{i\in S} T\,a_{qi} \left(\frac{v_i}{1+V(S)} + \lambda\,\theta_i \right) w(s) \right) : \sum_{S\subseteq N} w(s) = 1 \right\}.
\end{aligned}$$

In the chain of equalities above, (a) uses the fact that the LP on the left side of this equality is equivalent to the Assortment LP after replacing the revenue of product i with $r_i - \sum_{q\in M} a_{qi}\,\mu_q$. Thus, we can obtain the optimal objective value of this LP by solving the Mixture problem after replacing the revenue of product i with $r_i - \sum_{q\in M} a_{qi}\,\mu_q$. On the other hand, (b) holds, since picking one assortment that maximizes the expected revenue in the Mixture problem is equivalent to randomizing over all possible assortments. The optimal objective value of the last LP above gives the dual function for the Choice-Based LP. Therefore, the Compact LP and Choice-Based LP have the same dual functions. In this case, if we minimize the dual functions for the two LP formulations over all $\boldsymbol{\mu} \in \mathbb{R}^m_+$, then we get the same optimal objective value, so the two LP formulations have the same optimal objective value. Furthermore, the minimizers of the dual functions for the two LP formulations must be the same, which implies that the optimal values of the dual variables for the first constraint in the two LP formulations are the same. ■

By Theorem 6.1, we can solve the Compact LP to obtain the optimal dual variables of the first constraint in the Choice-Based LP, in which case we can use these dual variables to implement the bid-price policy. Considering the LP on the left side of (a) in the proof of the theorem, we do not a priori know whether product i satisfies $r_i - \sum_{q\in M} a_{qi}\,\mu_q \geq 0$. Therefore, as discussed immediately after the proof of Theorem 3.2, it is important that the Assortment LP can recover the optimal objective value of the Mixture problem even when some of the products have nonpositive revenues. To close this section, we iterate that we can have $\frac{v_i}{v_0+V(S)} + \lambda\,\theta_i > 1$, since we normalize the size of the customer segment with the multinomial logit model to one. To recover the purchase probability of product i, we need to scale $\frac{v_i}{v_0+V(S)} + \lambda\,\theta_i$ with β. Thus, the purchase probabilities in the Choice-Based LP and Compact LP have implicitly been scaled with $1/\beta$, so the resource capacities in these LP formulations have implicitly been scaled with $1/\beta$ as well.

7. Recovering a Primal Solution

We focus on recovering an optimal primal solution to the Choice-Based LP by using the Compact LP. Throughout this section, we follow the convention that if the Compact LP has multiple optimal solutions, then we pick any one that has the largest value for the decision variable x_0. It is simple to implement this convention in practice. In particular, for $\epsilon > 0$, we can add the additional term $\epsilon\, x_0$ to the objective function of the Compact LP. If ϵ is small enough, then the additional term favors an optimal solution with the largest value of x_0. If it is not clear how small ϵ should be, then another approach would be to first solve the Compact LP and obtain its optimal objective value. Letting z^*_{LP} be the optimal objective value of the Compact LP, we can then solve another LP where we maximize x_0 in the objective function, subject to the constraint that $T\sum_{i\in N} r_i\,((v_i+\lambda\,\theta_i)\,x_i+\lambda\,\theta_i\sum_{j\in N} v_j\,y_{ij}) \geq z^*_{\mathsf{LP}}$, along with all constraints in the Compact LP. In this case, we get a solution with the largest value for the decision variable x_0, providing an objective value of at least z^*_{LP}, so it must be optimal. In the next lemma, we establish a useful property of the basic optimal solutions to the Compact LP. The proof is given in Appendix E.

Lemma 7.1 (Extreme Point Optimal Solutions) *Let $(x^*_0, \boldsymbol{x}^*, \boldsymbol{y}^*)$ be a basic optimal solution to the* Compact LP. *Then, we have $y^*_{ij} = \min\{x^*_i, x^*_j\}$ for all $i,j\in N$.*

The proof, which is nontrivial, explicitly uses the fact that we pick a basic optimal solution that has the largest value for the decision variable x_0. We can generate examples to show that we may not have $y^*_{ij} = \min\{x^*_i, x^*_j\}$ for other basic optimal solutions. Also, note that the last two constraints in the Compact LP do not immediately imply that $y^*_{ij} = \min\{x^*_i, x^*_j\}$, since the first constraint in this LP may not allow setting $y^*_{ij} = \min\{x^*_i, x^*_j\}$ in a feasible solution to the Compact LP. Next, we focus on the main result of this section and give a remarkably efficient approach for recovering an optimal primal solution to the Choice-Based LP from an optimal solution to the Compact LP.

Let $(x^*_0, \boldsymbol{x}^*, \boldsymbol{y}^*)$ be a basic optimal solution to the Compact LP. We index the products so that $x^*_1 \geq x^*_2 \geq \ldots \geq x^*_n$. Defining the set $S_i = \{1,\ldots,i\}$ with $S_0 = \varnothing$, for each $i = 0,1,\ldots,n$, we set

$$\hat{w}(S_i) = (x^*_i - x^*_{i+1})\,(1+V(S_i)), \tag{Recovery}$$

where we follow the convention that $x^*_{n+1} = 0$. Noting that $x^*_0 \geq x^*_i$ for all $i\in N$ by the third constraint in the Compact LP, we have $\hat{w}(S_0) = x^*_0 - x^*_1 \geq 0$.

We define the solution $\hat{\boldsymbol{w}}$ to the Choice-Based LP as follows. For all $i = 0,1,\ldots,n$, we set $\hat{w}(S_i)$ as in the Recovery formula. For $S \notin \{S_0, S_1, \ldots, S_n\}$, we set $\hat{w}(S) = 0$.

In the next theorem, we show that the solution $\hat{\boldsymbol{w}}$ that we construct by using the Recovery formula as discussed above is an optimal solution to the Choice-Based LP.

Theorem 7.2 (Recovering an Optimal Solution) *For a basic optimal solution $(x_0^*, \boldsymbol{x}^*, \boldsymbol{y}^*)$ to the* Compact LP*, let $\hat{\boldsymbol{w}} = \{\hat{w}(S) : S \subseteq N\}$ be constructed as in the* Recovery *formula with $\hat{w}(S) = 0$ for all $S \notin \{S_0, S_1, \ldots, S_n\}$. Then, $\hat{\boldsymbol{w}}$ is an optimal solution to the* Choice-Based LP*.*

Proof: For notational brevity, let $\Lambda_i^* = T\,(v_i + \lambda\,\theta_i)\,x_i^* + T\,\lambda\,\theta_i \sum_{j \in N} v_j\, y_{ij}^*$. By Lemma 6.1, the optimal objective values of the Compact LP and Choice-Based LP are equal. Let z_{LP}^* be their common optimal objective value. Noting the objective function of the Compact LP, we have $z_{\mathsf{LP}}^* = \sum_{i \in N} r_i\, \Lambda_i^*$. Furthermore, by the first constraint in the Compact LP, we have $\sum_{i \in N} a_{qi}\, \Lambda_i^* \le c_q$. We will show that $\sum_{S \subseteq N} \hat{w}(S) = 1$ and $\sum_{S \subseteq N} \mathbf{1}(i \in S)\left(\frac{v_i}{1+V(S)} + \lambda\,\theta_i\right) \hat{w}(S) = \Lambda_i^*$. In this case, we get

$$\sum_{S \subseteq N} \sum_{i \in S} r_i \left(\frac{v_i}{v_0 + V(S)} + \lambda\,\theta_i \right) \hat{w}(S) = \sum_{i \in N} \sum_{S \subseteq N} r_i\, \mathbf{1}(i \in S) \left(\frac{v_i}{v_0 + V(S)} + \lambda\,\theta_i \right) \hat{w}(S) \; = \sum_{i \in N} r_i\, \Lambda_i^* = \; z_{\mathsf{LP}}^*.$$

Thus, the solution $\hat{\boldsymbol{w}}$ provides an objective value of z_{LP}^* for the Choice-Based LP, which is the optimal objective value of this LP. Moreover, replacing r_i with a_{qi} in the chain of equalities above and carrying out the same computation, we get $\sum_{S \subseteq N} \sum_{i \in S} a_{qi}\left(\frac{v_i}{v_0+V(S)} + \lambda\,\theta_i\right) \hat{w}(S) = \sum_{i \in N} a_{qi}\, \Lambda_i^*$. So, since $\sum_{i \in N} a_{qi}\, \Lambda_i^* \le c_q$, $\hat{\boldsymbol{w}}$ is a feasible solution to the Choice-Based LP. Noting that $\hat{\boldsymbol{w}}$ provides an objective value of z_{LP}^* for this LP, $\hat{\boldsymbol{w}}$ is an optimal solution for the Choice-Based LP, as desired.

First, we show that $\sum_{S \subseteq N} \hat{w}(S) = 1$. By the definition of S_i, we have $V(S_i) - V(S_{i-1}) = v_i$ for all $i = 1, \ldots, n$. Thus, using the Recovery formula, we get

$$\begin{aligned}
\sum_{S \subseteq N} \hat{w}(S) \; &\overset{(a)}{=} \; \sum_{i=0}^{n} \hat{w}(S_i) \; = \; \sum_{i=0}^{n} (x_i^* - x_{i+1}^*)\,(1 + V(S_i)) \; = \; \sum_{i=0}^{n} x_i^*\,(1 + V(S_i)) - \sum_{i=0}^{n} x_{i+1}^*\,(1 + V(S_i)) \\
&= \; \left(x_0^*\,(1 + V(S_0)) + \sum_{i=1}^{n} x_i^*\,(1 + V(S_i)) \right) - \left(\sum_{i=1}^{n} x_i^*\,(1 + V(S_{i-1})) + x_{n+1}^*\,(1 + V(S_n)) \right) \\
&\overset{(b)}{=} \; x_0^* + \sum_{i=1}^{n} x_i^*\,(V(S_i) - V(S_{i-1})) \; = \; x_0^* + \sum_{i=1}^{n} v_i\, x_i^* \; \overset{(c)}{=} \; 1,
\end{aligned}$$

where (a) holds since $\hat{w}(S) = 0$ for all $S \notin \{S_0, S_1, \ldots, S_n\}$, (b) holds since $x_{n+1}^* = 0$, and (c) holds since the solution $(x_0^*, \boldsymbol{x}^*, \boldsymbol{y}^*)$ satisfies the second constraint in the Compact LP.

Second, we show that $\sum_{S \subseteq N} \mathbf{1}(i \in S)\left(\frac{v_i}{1+V(S)} + \lambda\,\theta_i\right) \hat{w}(S) = \Lambda_i^*$. Noting the definition of $\hat{\boldsymbol{w}}$ in the Recovery formula, for each $k = 0, 1, \ldots, n$, we have

$$\begin{aligned}
\hat{w}(S_k) \; &= \; (x_k^* - x_{k+1}^*)\,(1 + V(S_k)) \; = \; x_k^* - x_{k+1}^* + (x_k^* - x_{k+1}^*)\,V(S_k) \\
&= \; x_k^* - x_{k+1}^* + \sum_{\ell \in N} \mathbf{1}(\ell \le k)\,(x_k^* - x_{k+1}^*)\,v_\ell,
\end{aligned}$$

where the last equality uses the fact that $S_i = \{1, \ldots, i\}$ and $V(S) = \sum_{i \in S} v_i$. By Lemma 7.1, we have $y_{ij}^* = \min\{x_i^*, x_j^*\}$ for all $i, j \in N$. Thus, since we index the products such that $x_1^* \ge x_2^* \ge \ldots \ge x_n^*$, we have $y^*_{ij} = x^*_i$ for $i \ge j$ and $y^*_{ij} = x^*_j$ for $i < j$. In other words, letting $a \vee b = \max\{a, b\}$, we have $y^*_{ij} = x^*_{i \vee j}$. Using the last chain of equalities displayed above, for each $i \in N$, we get

$$\begin{aligned}\sum_{k\in N} \mathbf{1}(k\ge i)\,\hat{w}(S_k) &= \sum_{k\in N}\mathbf{1}(k\ge i)\,(x^*_k - x^*_{k+1}) + \sum_{k\in N}\sum_{\ell\in N}\mathbf{1}(k\ge i)\,\mathbf{1}(\ell\le k)\,(x^*_k - x^*_{k+1})\,v_\ell \\ &\overset{(d)}{=} \sum_{k\in N}\mathbf{1}(k\ge i)\,(x^*_k - x^*_{k+1}) + \sum_{\ell\in N} v_\ell \sum_{k\in N}\mathbf{1}(k\ge i\vee \ell)\,(x^*_k - x^*_{k+1}) \\ &\overset{(e)}{=} x^*_i + \sum_{\ell\in N} v_\ell\, x^*_{i\vee\ell} \;=\; x^*_i + \sum_{\ell\in N} v_\ell\, y^*_{i\ell}, \end{aligned} \tag{4}$$

where (d) holds since $\mathbf{1}(k\ge i)\,\mathbf{1}(\ell\le k) = 1$ if and only if $\mathbf{1}(k \ge i\vee\ell)$ and (e) holds by canceling the telescoping terms in the first and third sums on the left side of the equality.

By the Recovery formula, we have $\sum_{k\in N}\mathbf{1}(k\ge i)\frac{1}{1+V(S_k)}\,\hat{w}(S_k) = \sum_{k\in N}\mathbf{1}(k\ge i)\,(x^*_k - x^*_{k+1}) = x^*_i$. In this case, noting that $i\in S_k$ if and only if $k\ge i$, we obtain

$$\begin{aligned}\sum_{S\subseteq N}\mathbf{1}(i\in S)\left(\frac{v_i}{1+V(S)} + \lambda\,\theta_i\right)\hat{w}(S) &= \sum_{k\in N}\mathbf{1}(i\in S_k)\left(\frac{v_i}{1+V(S_k)} + \lambda\,\theta_i\right)\hat{w}(S_k) \\ &= \sum_{k\in N}\mathbf{1}(k\ge i)\left(\frac{v_i}{1+V(S_k)} + \lambda\,\theta_i\right)\hat{w}(S_k) \\ &= v_i\sum_{k\in N}\mathbf{1}(k\ge i)\,\frac{1}{1+V(S_k)}\,\hat{w}(S_k) + \lambda\,\theta_i\sum_{k\in N}\mathbf{1}(k\ge i)\,\hat{w}(S_k) \\ &\overset{(f)}{=} v_i\,x^*_i + \lambda\,\theta_i\left(x^*_i + \sum_{\ell\in N} v_\ell\, y^*_{i\ell}\right) \overset{(g)}{=} \Lambda^*_i,\end{aligned}$$

where (f) follows from (4) and (g) holds by the definition of Λ^*_i. Thus, the two identities that we claim to hold at the beginning of the proof indeed hold. ■

By the theorem above, noting the Recovery formula, after we solve the Compact LP, recovering an optimal solution to the Choice-Based LP requires simply sorting the values of the decision variables $\{x_i : i\in N\}$. Furthermore, the Recovery formula implies that there exists an optimal solution $\boldsymbol{w}^*$ to the Choice-Based LP such that $w^*(S) = 0$ for all $S \notin \{S_0, S_1, \ldots, S_n\}$, so there exists an optimal solution to the Choice-Based LP that offers at most $n+1$ subsets. In this solution, since the sets $\{S_i : i = 0, 1, \ldots, n\}$ satisfy $S_0 \subseteq S_1 \subseteq \ldots \subseteq S_n$, if $w^*(S) > 0$ and $w^*(Q) > 0$, then we have either $S \supseteq Q$ or $Q \supseteq S$. Putting the last two observations together, there exists an optimal solution to the Choice-Based LP that offers at most $n+1$ assortments and these assortments are related to each other in the sense that one assortment is included in another one.

8. Computational Results

We provide two sets of computational experiments. In the first set, we test the ability of the mixture of multinomial logit and independent demand models to capture the choice process of the customers. In the second set, we check the computational benefits of using the Compact LP, instead of solving the Choice-Based LP directly by using column generation.

8.1 Prediction Ability of the Mixture Model

In this section, we test the benefits of using the mixture of multinomial logit and independent demand models to predict the purchase behavior of the customers.

Experimental Setup:

We generate the past purchase history of customers under the assumption that the choices of the customers are governed by a complex ground choice model that is very different from the multinomial logit model. The past purchase history includes the assortment of products offered to each customer and the product, if any, purchased within the assortment. We split the purchase history into training and testing data. We fit a mixture of multinomial logit and independent demand models to the training data and test the performance of the fitted model on the testing data. As a benchmark, we also fit a pure multinomial logit model to the same training data. In all of our test problems, we have $n = 10$ products. In the ground choice model, we have $p = 50$ customer types. Indexing the customer types by $P = \{1, \ldots, p\}$, customer type ℓ is characterized by a preference list of products $\boldsymbol{\sigma}^\ell = (\sigma^\ell(1), \sigma^\ell(2), \ldots, \sigma^\ell(k^\ell))$, with $\sigma^\ell(i) \in N$ for all $i = 1, \ldots, k^\ell$, where $\sigma^\ell(i)$ is the i-th most preferred product by a customer of type ℓ and k^ℓ is the number of products in the preference list. A customer of type ℓ arrives into the system with probability β_ℓ. An arriving customer chooses her most preferred available product in her preference list. If no product in her preference list is available, then the customer leaves without a purchase.

The parameters of the ground choice model are the collection of preference lists $\{\boldsymbol{\sigma}^\ell : \ell \in P\}$ and the arrival probabilities $\{\beta^\ell : \ell \in P\}$. To generate these parameters, we consider the case where the products have an inherent ordering $1 \succ 2 \succ \ldots \succ n$, in which product 1 has the highest quality and highest price, whereas product n has the lowest quality and lowest price. A customer of a particular type has a maximum price she can afford and minimum quality she accepts. In this case, the customer generally chooses the highest-quality product that is available within this range, but we add some noise to introduce some deviation from the inherent ordering of the products. In particular, for each customer type ℓ, we randomly choose an interval of products $[i^\ell, j^\ell]$ with $i^\ell < j^\ell$, so that a customer of type ℓ cannot afford products with price higher than that of product i^ℓ and does not accept products with quality lower than that of product j^ℓ. Considering the ordered list $(\gamma^\ell(1), \gamma^\ell(2), \ldots, \gamma^\ell(k^\ell)) = (i^\ell, i^\ell + 1, \ldots, j^\ell)$ with $k^\ell = j^\ell - i^\ell + 1$, we make 20 random swaps to change the position of a product with its successor so that the preference list only roughly follows the inherent ordering of the products. In this way, we obtain the preference list of products

$(\sigma^\ell(1), \sigma^\ell(2), \ldots, \sigma^\ell(k^\ell))$ for customers of type ℓ. Following the approach described so far in this paragraph, we generate the preference lists for $p-n=40$ customer types. The preference lists for the remaining $n=10$ customer types is a singleton, each including one of the products, so the customers of these types are unwilling to substitute between the products. To come up with the arrival probabilities, sampling ζ^ℓ from the uniform distribution over [0,1], for all $\ell = 1, \ldots, p-n$, we set $\beta^\ell = (1-\Theta)\,\zeta^\ell / \sum_{k=1}^{p-n} \zeta^k$, whereas for all $\ell = p-n+1, \ldots, p$, we set $\beta^\ell = \Theta\,\zeta^\ell / \sum_{k=p-n+1}^{p} \zeta^k$, where $\Theta \in (0,1)$ is a parameter that we vary. In this case, we have $\sum_{\ell=p-n+1}^{p} \beta^\ell = \Theta$, so Θ fraction of the customers are unwilling to substitute between the products.

Once we generate the ground choice model as in the previous paragraph, we generate the past purchase histories of customers who choose according to the ground choice model. The past purchase history consists of the pairs $\{(S_t, i_t) : t = 1, \ldots, \tau\}$, where τ is the number of customers in the history, S_t is the assortment that we offer to customer t, and i_t is the product that this customer chooses. If customer t does not purchase anything, then we have $i_t = 0$. To generate the assortment S_t, we include each product in the assortment S_t with probability $\rho \in (0,1)$, where ρ is another parameter that we vary. We sample the product i_t among the products in S_t and the no-purchase option according to the ground choice model.

We use maximum likelihood estimation to fit a mixture of multinomial logit and independent demand models to the past purchase history. To find a local maximum of the likelihood function, we use gradient search with 10 different initial points and bisection to find the best step size. All of our computational experiments are carried out in Java 1.7 on MacOS with 16 GB RAM and 2.8 GHz Intel Core i7 CPU. We compared our gradient search code with the `fmincon` routine in Matlab, and our gradient search code consistently reached local optima faster without sacrificing solution quality. Similarly, we use maximum likelihood estimation to fit a pure multinomial logit model without having an independent demand model in the mixture.

Testing Prediction Performance:

Varying $(\Theta, \rho) \in \{0.2, 0.3, 0.4, 0.5\} \times \{0.4, 0.5, 0.6\}$, we obtain 12 parameter combinations. For each parameter combination, we generate the ground choice model as described earlier in this section. Using the ground choice model, we generate the past purchase history of τ customers. We vary $\tau \in \{1250, 2500, 5000\}$ to capture three levels of data availability in the training data that we use to fit the choice models. Following the same approach to generate the training data, we generate the past purchase history for another 10,000 customers to use as the testing data. For each combination of (Θ, ρ) and τ, we replicate our results 50 times to get a better understanding of how much they change from one replication to another. We regenerate the ground choice model in each of these replications. We compare the two fitted choice models in terms of the out-of-sample

	$\tau = 1{,}250$					$\tau = 2{,}500$					$\tau = 5{,}000$				
Param. (Θ, ρ)	MIX Like.	MNL Like.	Perc. Gap	MIX Better	MNL Better	MIX Like.	MNL Like.	Perc. Gap	MIX Better	MNL Better	MIX Like.	MNL Like.	Perc. Gap	MIX Better	MNL Better
$(0.2, 0.4)$	-12,092	-12,195	0.85	46	4	-12,041	-12,167	1.05	50	0	-12,024	-12,158	1.12	50	0
$(0.2, 0.5)$	-13,753	-13,832	0.57	47	3	-13,710	-13,807	0.71	50	0	-13,690	-13,796	0.78	50	0
$(0.2, 0.6)$	-15,194	-15,250	0.37	45	5	-15,156	-15,224	0.45	49	1	-15,137	-15,214	0.51	50	0
$(0.3, 0.4)$	-12,510	-12,596	0.69	48	2	-12,461	-12,570	0.88	50	0	-12,441	-12,561	0.96	50	0
$(0.3, 0.5)$	-14,269	-14,345	0.54	49	1	-14,228	-14,321	0.66	50	0	-14,208	-14,310	0.72	50	0
$(0.3, 0.6)$	-15,841	-15,904	0.40	46	4	-15,804	-15,880	0.48	50	0	-15,784	-15,870	0.54	50	0
$(0.4, 0.4)$	-12,772	-12,867	0.74	49	1	-12,734	-12,847	0.88	50	0	-12,716	-12,838	0.96	50	0
$(0.4, 0.5)$	-14,624	-14,705	0.55	48	2	-14,588	-14,686	0.67	49	1	-14,569	-14,677	0.74	50	0
$(0.4, 0.6)$	-16,286	-16,358	0.44	48	2	-16,253	-16,341	0.54	50	0	-16,235	-16,332	0.60	50	0
$(0.5, 0.4)$	-12,881	-12,966	0.66	50	0	-12,845	-12,946	0.79	50	0	-12,826	-12,938	0.87	50	0
$(0.5, 0.5)$	-14,850	-14,932	0.55	50	0	-14,810	-14,912	0.69	50	0	-14,793	-14,903	0.74	50	0
$(0.5, 0.6)$	-16,605	-16,689	0.50	50	0	-16,567	-16,669	0.61	50	0	-16,549	-16,659	0.67	50	0
Average			0.57	48	2			0.70	49.83	0.17			0.77	50	0

Table 2 Out-of-sample log-likelihoods of the two fitted choice models.

log-likelihoods and the deviation between the choice probabilities under each fitted choice model and the exact ground choice model. Throughout this section, we use MIX to refer to the fitted mixture of multinomial logit and independent demand models, whereas MNL to refer to the fitted pure multinomial logit model.

In Table 2, we compare MIX and MNL in terms of their out-of-sample log-likelihoods. In each of the 50 replications, after generating the ground choice model, we sample training data and testing data using the ground choice model. We fit MIX and MNL to the training data, and we compute the log-likelihood of the testing data under the fitted choice models. The first column gives the parameter configuration (Θ, ρ). In the rest of the table, there are three blocks, each with five columns. The three blocks correspond to the values of $\tau \in \{1250, 2500, 5000\}$, capturing three levels of data availability. In each block, the first column gives the average out-of-sample log-likelihood of MIX, where the average is computed over 50 replications. The second column gives the average out-of-sample log-likelihood of MNL. The third column gives the average percent gap between out-of-sample log-likelihoods of MIX and MNL. The fourth column gives the number of replications out of 50 in which the out-of-sample log-likelihood of MIX is better than that of MNL, whereas the fifth column gives the number of replications in which the outcome is reversed. Our results indicate that the out-of-sample log-likelihoods of MIX are noticeably larger than those of MNL. For the smallest training data availability with $\tau = 1250$, there are a few replications in which the out-of-sample log-likelihoods of MNL are larger than those of MIX, but for $\tau = 2500$ and $\tau = 5000$, the out-of-sample log-likelihoods of MIX are quite consistently larger than those of MNL.

Note that MIX has $2n$ parameters, whereas MNL has n parameters. Thus, MIX provides more flexibility for capturing the customer choice behavior. With its larger number of parameters, however, MIX may overfit to the training data, resulting in poor out-of-sample log-likelihoods, especially when we have too little training data. Thus, it is not guaranteed that the out-of-sample

Param. (Θ, ρ)	$\tau = 1{,}250$ MIX Error	MNL Error	Perc. Gap	MIX Better	MNL Better	$\tau = 2{,}500$ MIX Error	MNL Error	Perc. Gap	MIX Better	MNL Better	$\tau = 5{,}000$ MIX Error	MNL Error	Perc. Gap	MIX Better	MNL Better
$(0.2, 0.4)$	0.022	0.027	22.43	47	3	0.019	0.026	32.96	50	0	0.018	0.025	40.09	50	0
$(0.2, 0.5)$	0.019	0.023	17.58	48	2	0.017	0.021	24.59	49	1	0.016	0.021	30.52	50	0
$(0.2, 0.6)$	0.017	0.019	12.16	47	3	0.015	0.018	17.75	48	2	0.014	0.017	22.46	50	0
$(0.3, 0.4)$	0.021	0.026	24.23	47	3	0.018	0.025	38.89	50	0	0.016	0.024	51.10	50	0
$(0.3, 0.5)$	0.018	0.022	21.06	49	1	0.016	0.021	30.00	50	0	0.014	0.020	39.76	50	0
$(0.3, 0.6)$	0.015	0.018	17.72	49	1	0.014	0.017	25.34	50	0	0.012	0.017	33.28	50	0
$(0.4, 0.4)$	0.020	0.025	30.52	50	0	0.016	0.024	47.62	50	0	0.015	0.023	60.69	50	0
$(0.4, 0.5)$	0.017	0.021	23.83	48	2	0.015	0.020	36.53	49	1	0.013	0.019	47.94	50	0
$(0.4, 0.6)$	0.014	0.018	22.38	46	4	0.012	0.017	33.75	50	0	0.011	0.016	42.85	50	0
$(0.5, 0.4)$	0.018	0.024	30.97	50	0	0.015	0.022	50.65	50	0	0.013	0.022	66.34	50	0
$(0.5, 0.5)$	0.016	0.020	26.05	49	1	0.013	0.019	44.87	50	0	0.011	0.018	58.91	50	0
$(0.5, 0.6)$	0.014	0.017	24.79	50	0	0.011	0.016	41.08	50	0	0.010	0.015	54.13	50	0
Average			22.81	48.33	1.67			35.34	49.67	0.33			45.67	50	0

Table 3 Mean absolute errors of the choice probabilities of the two fitted choice models.

log-likelihoods of MIX will be larger than those of MNL. Nevertheless, overfitting does not seem to be a concern for MIX, and the out-of-sample log-likelihoods of MIX are larger than those of MNL in an overwhelming majority of our replications. The gap between the out-of-sample log-likelihoods of MIX and MNL becomes more pronounced as the amount of training data increases, corresponding to larger values for τ. Shortly, we demonstrate that such improvements in out-of-sample log-likelihoods translate into significant improvements in expected revenues.

In Table 3, we compare MIX and MNL in terms of how closely they track the choice probabilities of the ground choice model. The layout of this table is identical to that of Table 2, with three blocks capturing three levels of data availability. In replication q, let $\phi_i^q(S)$ be the choice probability of product i within assortment S under the fitted MIX. Also, let $P_i^q(S)$ be the corresponding choice probability under the ground choice model. Letting $\{S_t : t = 1, \ldots, 10000\}$ be the assortments in the testing data, we compute the mean absolute error of the choice probabilities for MIX as $\Delta^q_{\text{MIX}} = \frac{1}{10000} \sum_{t=1}^{10000} \frac{1}{|S_t|} \sum_{i \in S_t} |\phi_i^q(S_t) - P_i^q(S_t)|$. In each block, considering the 50 replications, the first column gives the average of $\{\Delta^q_{\text{MIX}} : q = 1, \ldots, 50\}$. The second column gives the average of $\{\Delta^q_{\text{MNL}} : q = 1, \ldots, 50\}$, where Δ^q_{MNL} is the mean absolute error for MNL in replication q, computed in a fashion similar to Δ^q_{MIX}. The third column gives the percent gap between the first two columns. The fourth column gives the number of replications in which the mean absolute error of MIX is smaller than that of MNL, whereas the fifth column gives the number of replications in which the outcome is reversed. We focus on mean absolute errors rather than mean absolute percent errors, because if a product has a small purchase probability, then misestimating this purchase probability even by a small amount may increase the mean absolute percent error substantially, putting disproportionate weight on estimating small choice probabilities more accurately. Our results indicate that MIX improves the mean absolute errors in the fitted choice probabilities significantly, when compared to MNL. The improvements hold for an overwhelming majority of our

	$\tau = 1{,}250$					$\tau = 2{,}500$					$\tau = 5{,}000$				
Param. (Θ, ρ)	MIX Rev.	MNL Rev.	Perc. Gap	MIX Better	MNL Better	MIX Rev.	MNL Rev.	Perc. Gap	MIX Better	MNL Better	MIX Rev.	MNL Rev.	Perc. Gap	MIX Better	MNL Better
$(0.2, 0.4)$	5.79	5.67	2.04	48	2	5.81	5.67	2.48	50	0	5.82	5.67	2.64	50	0
$(0.2, 0.5)$	5.78	5.66	2.19	45	5	5.81	5.66	2.63	50	0	5.82	5.66	2.70	50	0
$(0.2, 0.6)$	5.78	5.66	2.05	46	4	5.81	5.66	2.47	49	1	5.82	5.66	2.67	50	0
$(0.3, 0.4)$	5.60	5.44	2.95	50	0	5.64	5.44	3.54	50	0	5.66	5.44	3.89	50	0
$(0.3, 0.5)$	5.64	5.46	3.17	50	0	5.66	5.46	3.60	50	0	5.67	5.46	3.84	50	0
$(0.3, 0.6)$	5.60	5.43	3.09	47	3	5.63	5.43	3.57	50	0	5.65	5.43	3.90	50	0
$(0.4, 0.4)$	5.51	5.25	4.62	50	0	5.53	5.26	4.98	50	0	5.54	5.26	5.19	50	0
$(0.4, 0.5)$	5.52	5.26	4.63	50	0	5.54	5.26	4.99	50	0	5.55	5.26	5.21	50	0
$(0.4, 0.6)$	5.50	5.25	4.63	50	0	5.53	5.24	5.09	50	0	5.54	5.24	5.29	50	0
$(0.5, 0.4)$	5.41	5.10	5.69	50	0	5.44	5.10	6.20	50	0	5.45	5.10	6.44	50	0
$(0.5, 0.5)$	5.41	5.10	5.75	50	0	5.43	5.10	6.22	50	0	5.45	5.09	6.48	50	0
$(0.5, 0.6)$	5.40	5.08	5.95	50	0	5.43	5.08	6.49	50	0	5.45	5.08	6.85	50	0
Average			3.90	48.83	1.17			4.35	49.92	0.08			4.59	50	0

Table 4 Expected revenues obtained by using the two fitted choice models.

replications. Furthermore, the gaps between the mean absolute errors become more pronounced as the amount of training data increases.

Testing Revenue Performance:

In Table 4, we compare MIX and MNL in terms of their revenue performance. The layout of this table is identical to that of Table 2, with three blocks capturing three levels of data availability. In replication q, let $\phi_i^q(S)$ be the choice probability of product i within assortment S under the fitted MIX. We generate 100 samples of the product revenues, sampling the revenue of each product from the uniform distribution over $[1, 10]$. In replication q, letting $(r_1^{qk}, \ldots, r_n^{qk})$ be the revenues of the products in the k-th sample, we use $\hat{S}^{qk}$ to denote the optimal assortment to offer under the assumption that the customers choose under the fitted MIX. In other words, $\hat{S}^{qk}$ is an optimal solution to the problem $\max_{S \subseteq N} \sum_{i \in S} r_i^{qk} \, \phi_i^q(S)$. The customers, however, actually choose according to the ground choice model. In replication q, letting let $P_i^q(S)$ be the choice probability of product i within assortment S under the ground choice model, we compute the actual expected revenue from assortment $\hat{S}^{qk}$ as $R^{qk}_{\mathrm{MIX}} = \sum_{i \in \hat{S}^{qk}} r_i^{qk} \, P_i^q(S)$. Averaging over all the 100 revenue samples, in replication q, we capture the expected revenue performance of the fitted MIX by $\mathrm{Rev}^q_{\mathrm{MIX}} = \frac{1}{100} \sum_{k=1}^{100} R^{qk}_{\mathrm{MIX}}$. In each block, considering the 50 replications, the first column gives the average of $\{\mathrm{Rev}^q_{\mathrm{MIX}} : q = 1, \ldots, 50\}$. The second column gives the average of $\{\mathrm{Rev}^q_{\mathrm{MNL}} : q = 1, \ldots, 50\}$, where $\mathrm{Rev}^q_{\mathrm{MNL}}$ captures the expected revenue performance of the fitted MNL in replication q, computed in a fashion similar to $\mathrm{Rev}^q_{\mathrm{MIX}}$. The third column gives the percent gap between the first two columns. The fourth column gives the number of replications in which the expected revenue performance of MIX is better than that of MNL, whereas the fifth column gives the number of replications in which the outcome is reversed.

Our results indicate that fitting MIX to the training data can provide assortments with significantly larger revenues, when compared to fitting MNL to the training data. The improvements in the expected revenue provided by MIX are consistent over an overwhelming majority of our replications and can exceed 6%. As τ gets larger and the amount of training data increases, the improvements in the expected revenue provided by MIX become more noticeable.

8.2 Computational Benefits of the Compact Formulation

In this section, we check the computational benefits of using the Compact LP in conjunction with Theorem 7.2 to get an optimal solution to the Choice-Based LP, rather than solving the Choice-Based LP directly by using column generation.

Experimental Setup:

We generate multiple instances of the network revenue management problem using the following approach. The set of products is $N = \{1, \ldots, n\}$ with $n = 100$, and the set of resources is $M = \{1, \ldots, m\}$, where m is a parameter that we vary. In the multinomial logit model, for each product i, we generate η_i from the uniform distribution over $[0,1]$ and set the preference weight of product i as $v_i = \eta_i \left(\frac{1-P_0}{P_0}\right) / \sum_{j \in N} \eta_j$, where P_0 is another parameter that we vary. In this case, if we offer all products, then the customer segment with the multinomial logit model leaves without a purchase with probability $\frac{1}{1+\sum_{i \in N} v_i} = \frac{1}{1+(1-P_0)/P_0} = P_0$. In the independent demand model, we generate γ_i from the uniform distribution over $[0,1]$ and set the probability of demand for product i as $\theta_i = \gamma_i / \sum_{j \in N} \gamma_j$. In this case, the purchase probability of product i within assortment S is $\phi_i(S) = \beta \frac{v_i}{1+\sum_{j \in S} v_j} + (1-\beta)\,\theta_i$, where β is one more parameter that we vary.

We have $T = 100$ time periods. We sample the revenue r_i of each product i from the uniform distribution over $[100, 500]$. For each product i, we randomly choose a resource q_i and set $a_{q_i,i} = 1$. For the other resources, we set $a_{qi} = 1$ with probability $1/5$ and $a_{qi} = 0$ with probability $4/5$ for all $q \in M \setminus \{q_i\}$. Thus, the expected number of resources used by a product is $1 + (m-1)/5$. To come up with the capacities of the resources, noting that $\phi_i(S)$ is the choice probability of product i within assortment S in the previous paragraph, we let S^* be an optimal solution to the problem $\max_{S \subseteq N} \sum_{i \in S} r_i\, \phi_i(S)$, which is the assortment that maximizes the expected revenue under infinite resource capacities. If we offer the assortment S^* over the entire selling horizon, then the total expected capacity consumption of resource q is $T \sum_{i \in S^*} a_{qi}\, \phi_i(S^*)$. We set the capacity of resource q as $c_q = \kappa\, T \sum_{i \in S^*} a_{qi}\, \phi_i(S^*)$, where κ is a last parameter that we vary.

Computational Results:

Varying $(m, P_0, \beta, \kappa) \in \{25, 50\} \times \{0.1, 0.2\} \times \{0.25, 0.75\} \times \{0.6, 0.8\}$, we obtain 16 parameter combinations. For each parameter combination, we generate a problem instance by using the approach in the previous two paragraphs. We obtain an optimal solution to the Choice-Based LP

Param. (m, P_0, β, κ)	COG Secs.	CLP Secs.	Secs. Ratio	1% Gp. COG Secs.	1% Gp. 1% Secs. Ratio
$(25, 0.1, 0.25, 0.6)$	62.51	4.24	14.74	22.82	5.38
$(25, 0.1, 0.25, 0.8)$	58.34	5.09	11.46	21.06	4.14
$(25, 0.1, 0.75, 0.6)$	70.50	4.79	14.72	37.98	7.93
$(25, 0.1, 0.75, 0.8)$	72.07	7.72	9.34	31.94	4.14
$(25, 0.2, 0.25, 0.6)$	55.21	4.11	13.43	23.56	5.73
$(25, 0.2, 0.25, 0.8)$	52.23	5.37	9.73	20.03	3.73
$(25, 0.2, 0.75, 0.6)$	86.15	5.02	17.16	36.20	7.21
$(25, 0.2, 0.75, 0.8)$	74.90	9.80	7.64	31.43	3.21
Average			12.28		5.18

Param. (m, P_0, β, κ)	COG Secs.	CLP Secs.	Secs. Ratio	1% Gp. COG Secs.	1% Gp. Secs. Ratio
$(50, 0.1, 0.25, 0.6)$	101.38	6.02	16.84	29.63	4.92
$(50, 0.1, 0.25, 0.8)$	93.84	4.11	22.83	27.99	6.81
$(50, 0.1, 0.75, 0.6)$	128.20	6.95	18.45	46.15	6.64
$(50, 0.1, 0.75, 0.8)$	143.04	8.78	16.29	55.12	6.28
$(50, 0.2, 0.25, 0.6)$	139.19	5.15	27.03	29.62	5.75
$(50, 0.2, 0.25, 0.8)$	149.65	5.12	29.23	46.32	9.05
$(50, 0.2, 0.75, 0.6)$	145.13	10.32	14.06	55.17	5.35
$(50, 0.2, 0.75, 0.8)$	122.47	11.07	11.06	38.22	3.45
Average			19.47		6.03

Table 5 Running times for solving the Choice-Based LP through two methods.

for each problem instance by using two methods. First, we solve the Choice-Based LP directly by using column generation. We refer to this method as COG, which stands for column generation. Second, we solve the Compact LP and build on Theorem 7.2 to use an optimal solution of this LP to recover an optimal solution of the Choice-Based LP. We refer to this method as CLP, which stands for compact LP. We show our results in Table 5. The first column gives the parameter combination. The second column gives the running time for COG to obtain an optimal solution to the Choice-Based LP through column generation. The third column gives the running time for CLP to solve the Compact LP and use an optimal solution to this LP to recover an optimal solution to the Choice-Based LP. We use Gurobi 9.0 as our LP solver. The fourth column gives the ratio of the running times in the second and third columns. Column generation may get near-optimal solutions quickly but may take a while to close the remaining portion of the optimality gap. To check for this possibility, the fifth column gives the running time for COG to solve the Choice-Based LP with a 1% optimality gap. The sixth column gives the ratio of the running times in the third and fifth columns. Our results indicate that CLP can improve the running times for COG by up to a factor of 29.23. The average improvement in the running times is a factor of 15.88. If we allow COG to terminate with a 1% optimality gap, but run CLP until it gets to the optimal solution, then CLP can still improve the running times for COG by up to a factor of 9.05. The improvements in the running times become more pronounced when m is larger, so that we have problem instances with a larger number of resources. In our test problems, most of the running time for COG is spent on solving the Compact LP. It takes less than one-tenths of a second to recover an optimal solution to the Choice-Based LP by using an optimal solution to the Compact LP through Theorem 7.2.

We also compare the performance of COG and CLP for larger test problems with $n = 500$ products and $m = 100$ resources. For such test problems, COG does not reach an optimal solution within one hour of running time. We give our results in Table 6. The first column shows the problem parameters. The interpretation of the problem parameters P_0, β, and κ is the same as the one presented earlier in this section. The second column shows the optimality gap for COG after

Param. (P_0, β, κ)	COG % Opt. Gap.	CLP Secs.
$(0.1, 0.25, 0.6)$	4.93	676.13
$(0.1, 0.25, 0.8)$	5.27	2139.15
$(0.1, 0.75, 0.6)$	14.27	909.10
$(0.1, 0.75, 0.8)$	9.51	2420.80
Average	8.49	1536.29

Param. (P_0, β, κ)	COG % Opt. Gap.	CLP Secs.
$(0.2, 0.25, 0.6)$	7.17	408.59
$(0.2, 0.25, 0.8)$	4.71	1059.42
$(0.2, 0.75, 0.6)$	14.10	950.74
$(0.2, 0.75, 0.8)$	8.53	2458.98
Average	8.63	1219.43

Table 6 Optimality gaps and running times for the two methods for solving the Choice-Based LP with $n = 500$ products and $m = 100$ resources.

one hour of running time. The third column shows the running time for CLP to get the optimal solution. Over all the test problems, the average optimality gap of the solutions obtained by COG after one hour of running time is 8.56%. There are test problems for which COG terminates with more than a 14% optimality gap. The average running time for CLP to obtain an optimal solution is about 23 minutes, the longest running time not exceeding 41 minutes.

9. Conclusions

We studied the single-shot unconstrained and cardinality-constrained assortment optimization and assortment-based network revenue management problems under a mixture of multinomial logit and independent demand models. Our mixture choice model is a natural way to simultaneously improve the flexibility of both the multinomial logit and independent demand models to capture the choice process of the customers, while ensuring that the corresponding assortment optimization problems remain tractable. There are several avenues for further research. Mixing the multinomial logit model with the independent demand model resulted in efficiently solvable assortment optimization problems, but our results closely exploited the structure of the multinomial logit model. One can mix the independent demand model with other choice models and try to tackle the corresponding assortment optimization problems. Moreover, we focused on solving assortment optimization problems, but our computational experiments indicated that our mixture choice model can improve the modeling flexibility of the pure multinomial logit model in terms of predicting the purchases of the customers. It would be useful to test the prediction ability of our mixture choice model on data generated by real-world applications. Lastly, one can work on enriching our mixture choice model by incorporating an incremental process for viewing the products in batches or by allowing customers with different consideration sets.

References

Aouad, A., V. Farias, R. Levi. 2016. Assortment optimization under consider-then-choose choice models. Tech. rep., MIT, Massachusetts, MA.

Aouad, A., V. Farias, R. Levi, D. Segev. 2018a. The approximability of assortment optimization under ranking preferences. *Operations Research* **66**(6) 1661–1669.

Aouad, A., J. Feldman, D. Segev, D. J. Zhang. 2019. Click-based MNL: Algorithmic frameworks for modeling click data in assortment optimization. Tech. rep., Washington University, St. Louis, MO.

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

Aouad, A., D. Segev. 2019. The stability of MNL-based demand under dynamic customer substitution and its algorithmic implications. Tech. rep., University of Haifa, Haifa, Israel.

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

Bront, J. J. M., I. Mendez-Diaz, G. Vulcano. 2009. A column generation algorithm for choice-based network revenue management. *Operations Research* **57**(3) 769–784.

Dai, J., W. Ding, A. J. Kleywegt, X. Wang, Y. Zhang. 2014. Choice based revenue management for parallel flights. Tech. rep., Georgia Tech, Atlanta, GA.

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

Desir, A., V. Goyal, J. Zhang. 2016. Near-optimal algorithms for capacity constrained assortment optimization. Tech. rep., Columbia University, New York, NY.

Farias, V. F., S. Jagabathula, D. Shah. 2013. A non-parametric approach to modeling choice with limited data. *Management Science* **59**(2) 305–322.

Feldman, J., A. Paul, H. Topaloglu. 2019. Assortment optimization with small consideration sets. *Operations Research* **67**(5) 1283–1299.

Feldman, J., D. Segev. 2019. Improved approximation schemes for MNL-driven sequential assortment optimization. Tech. rep., Washington University, St. Louis, MO.

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

Feldman, J., H. Topaloglu. 2018. Technical note: Capacitated assortment optimization under the multinomial logit model with nested consideration sets. *Operations Research* **66**(2) 380–391.

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

Flores, A., G. Berbeglia, P. van Hentenryck. 2019. Assortment optimization under the sequential multinomial logit model. *European Journal of Operational Research* **273**(3) 1052–1064.

Gallego, G., G. Iyengar, R. Phillips, A. Dubey. 2004. Managing flexible products on a network. CORC Technical Report TR-2004-01.

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

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

Garey, M. R., D. S. Johnson. 1979. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman and Company, New York, NY.

Jagabathula, S. 2016. Assortment optimization under general choice. Tech. rep., NYU, New York, NY.

Jasin, S., S. Kumar. 2012. A re-solving heuristic with bounded revenue loss for network revenue management with customer choice. *Mathematics of Operations Research* **37**(2) 313–345.

Kunnumkal, S., H. Topaloglu. 2010. A new dynamic programming decomposition method for the network revenue management problem with customer choice behavior. *Production and Operations Management* **19**(5) 575–590.

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

Liu, N., Y. Ma, H. Topaloglu. 2019. Assortment optimization under the multinomial logit model with sequential offerings. *INFORMS Journal on Computing* (forthcoming).

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

Ma, W. 2019. Assortment auctions: A Myersonian characterization for Markov chain based choice models. Tech. rep., Columbia University, New York, NY.

Mendez-Diaz, I., J. J. M. Bront, G. Vulcano, P. Zabala. 2014. A branch-and-cut algorithm for the latent-class logit assortment problem. *Discrete Applied Mathematics* **164**(1) 246–263.

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

Rusmevichientong, P., D. Shmoys, C. Tong, H. Topaloglu. 2014. Assortment optimization under the multinomial logit model with random choice parameters. *Production and Operations Management* **23**(11) 2023–2039.

Sumida, M., G. Gallego, P. Rusmevichientong, H. Topaloglu, J. M. Davis. 2019. Revenue-utility tradeoff in assortment optimization under the multinomial logit model with totally unimodular constraints. Tech. rep., Cornell University, Ithaca, NY.

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

Tong, C., H. Topaloglu. 2013. On the approximate linear programming approach for network revenue management problems. *INFORMS Journal on Computing* **26**(1) 121–134.

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

van Ryzin, G. J. 2005. Future of revenue management: Models of demand. *Journal of Revenue and Pricing Management* **4** 204–210.

Vossen, T. W. M., D. Zhang. 2015. Reductions of approximate linear programs for network revenue management. *Operations Research* **63**(6) 1352–1371.

Vulcano, G., G. van Ryzin, W. Chaar. 2010. Choice-based revenue management: An empirical study of estimation and optimization. *Manufacturing & Service Operations Management* **12**(3) 371–392.

Wang, R. 2012. Capacitated assortment and price optimization under the multinomial logit model. *Operations Research Letters* **40**(6) 492–497.

Wang, R., O. Sahin. 2018. The impact of consumer search cost on assortment planning and pricing. *Management Science* **64**(8) 3649–3666.

Zhang, D., D. Adelman. 2009. An approximate dynamic programming approach to network revenue management with customer choice. *Transportation Science* **43**(3) 381–394.

Zhang, D., W. L. Cooper. 2005. Revenue management for parallel flights with customer-choice behavior. *Operations Research* **53**(3) 415–431.

Zhang, H., P. Rusmevichientong, H. Topaloglu. 2019. Assortment optimization under the paired combinatorial logit model. *Operations Research* (forthcoming).

Appendix A: Comparison with the Markov Chain Choice Model

We give an example to show that our mixture of multinomial logit and independent demand models is not a special case of the Markov chain choice model. Under the Markov chain choice model, a customer arriving into the system is interested in purchasing product i with probability γ_i. If this product is available for purchase, then the customer purchases it. Otherwise, the customer transitions from product i to product j with probability ρ_{ij} and checks whether product j is available for purchase. With probability $1-\sum_{j\in N}\rho_{ij}$, the customer transitions to the no-purchase option, in which case, she leaves without a purchase. In this way, the customer transitions among the products according to a Markov chain until she visits a product that is available for purchase or she visits the no-purchase option. The parameters of the Markov chain choice model are $\{\gamma_i : i\in N\}$ and $\{\rho_{ij} : i,j\in N\}$. Given that we offer the assortment $S\subseteq N$ of products, we let $P_i(S)$ be the expected number of times that a customer visits product i during the course of her choice process. If $i\in S$, then a customer purchases product i as soon as she visits this product, so for $i\in S$, $P_i(S)$ is the purchase probability of product i when we offer the assortment S. We can compute $\{P_i(S) : i\in N\}$ by solving the system of equations

$$P_i(S)=\gamma_i+\sum_{j\notin S}\rho_{ji}\,P_j(S) \qquad \forall\, i\in N. \tag{5}$$

We can intuitively justify (5) through a balance argument (Feldman and Topaloglu 2017). On the left side, $P_i(S)$ is the expected number of times that a customer visits product i during the course of her choice process. For a customer to visit product i, she may arrive into the system with an interest to purchase product i, which happens with probability γ_i. Alternatively, she may visit some product $j\notin S$ and the expected number of visits to this product is $P_j(S)$. In this case, if she transitions from product j to product i, then the customer ends up visiting product i. The probability of transitioning from product j to product i is ρ_{ji}. If $\sum_{j\in N}\rho_{ij}<1$ for all $i\in N$, then there exists a solution to the system of equations above for any $S\subseteq N$.

We consider an instance of the mixture of multinomial logit and independent demand models with $N=\{1,2,3\}$, $(v_1,v_2,v_3)=(1,1,1)$, $(\theta_1,\theta_2,\theta_3)=(0,0,1)$ and $\beta=\frac{3}{4}$. Under this choice model, if we offer the assortment S, then a customer purchases product $i\in S$ with probability $\phi_i(S)=\beta\,\frac{v_i}{1+V(S)}+(1-\beta)\,\theta_i$. Note that we did not normalize the size of the first customer segment to one. In Table EC.1, we give the choice probabilities $\{\phi_i(S) : i\in S,\ S\subseteq N\}$ for this instance of the mixture of multinomial logit and independent demand models. We argue that there exists no Markov chain choice model such that the choice probabilities under the Markov chain choice model for all products and for all assortments match those under the mixture of multinomial

S	$\phi_1(S)$	$\phi_2(S)$	$\phi_3(S)$
$\varnothing$	0	0	0
$\{1\}$	3/8	0	0
$\{2\}$	0	3/8	0
$\{3\}$	0	0	5/8
$\{1,2\}$	1/4	1/4	0
$\{1,3\}$	1/4	0	1/2
$\{2,3\}$	0	1/4	1/2
$\{1,2,3\}$	3/16	3/16	7/16

Table EC.1 Expected revenue provided by all possible assortments.

logit and independent demand models. In other words, there exist no parameters $\{\gamma_i : i \in N\}$ and $\{\rho_{ij} : i,j \in N\}$ for the Markov chain choice model such that $P_i(S) = \phi_i(S)$ for all $S \subseteq N$, $i \in N$. To make this argument, by (5), note that $P_i(\{1,2,3\}) = \gamma_i$ for all $i \in N$. Thus, to ensure that $P_i(\{1,2,3\}) = \phi_i(\{1,2,3\})$ for all $i \in N$, we must choose $\{\gamma_i : i \in N\}$ so that $\gamma_1 = \phi_1(\{1,2,3\}) = \frac{3}{16}$, $\gamma_2 = \phi_2(\{1,2,3\}) = \frac{3}{16}$ and $\gamma_3 = \phi_3(\{1,2,3\}) = \frac{7}{16}$, fixing the values of the parameters $\{\gamma_i : i \in N\}$.

Consider an assortment of the form $N \setminus \{i\}$. Product i is the only one not in the assortment $N \setminus \{i\}$, so by (5), we get $P_k(N \setminus \{i\}) = \gamma_k + \rho_{ik}\, P_i(N \setminus \{i\})$ for all $k \in N$. Using the last equality with $k = i$, we get $(1 - \rho_{ii})\; P_i(N \setminus \{i\}) = \gamma_i$, so the equality $P_k(N \setminus \{i\}) = \gamma_k + \rho_{ik}\, P_i(N \setminus \{i\})$ is equivalent to $P_k(N \setminus \{i\}) = \gamma_k + \rho_{ik}\, \frac{\gamma_i}{1-\rho_{ii}}$, which, in turn, is equivalent to $\rho_{ik} = \frac{1-\rho_{ii}}{\gamma_i}\,(P_k(N \setminus \{i\}) - \gamma_k)$. Thus, to ensure that $P_k(N \setminus \{i\}) = \phi_k(N \setminus \{i\})$ for all $i,k \in N$, we must have

$$\rho_{ik} = \frac{1-\rho_{ii}}{\gamma_i}\,(\phi_k(N \setminus \{i\}) - \gamma_k).$$

Using the values of $\phi_k(N \setminus \{i\})$ for $i,k \in N$ in Table EC.1 and the fact that $\gamma_1 = \frac{3}{16}$, $\gamma_2 = \frac{3}{16}$ and $\gamma_3 = \frac{3}{16}$, the expression above yields $\rho_{21} = \frac{1}{3}(1-\rho_{22})$, $\rho_{31} = \frac{1}{7}(1-\rho_{33})$, $\rho_{23} = \frac{1}{3}(1-\rho_{22})$ and $\rho_{32} = \frac{1}{7}(1-\rho_{33})$. Lastly, consider the assortment $\{1\}$. By (5), we have $P_2(\{1\}) = \gamma_2 + \rho_{22}\, P_2(\{1\}) + \rho_{32}\, P_3(\{1\})$, which is equivalent to $(1-\rho_{22})\, P_2(\{1\}) = \gamma_2 + \rho_{32}\, P_3(\{1\})$. Similarly, $(1-\rho_{33})\, P_3(\{1\}) = \gamma_3 + \rho_{23}\, P_2(\{1\})$. Since $\rho_{23} = \frac{1}{3}(1-\rho_{22})$ and $\rho_{32} = \frac{1}{7}(1-\rho_{33})$, the last two equalities become

$$\begin{aligned}(1-\rho_{22})\, P_2(\{1\}) &= \gamma_2 + \tfrac{1}{7}\,(1-\rho_{33})\, P_3(\{1\})\\ (1-\rho_{33})\, P_3(\{1\}) &= \gamma_3 + \tfrac{1}{3}\,(1-\rho_{22})\, P_2(\{1\}).\end{aligned}$$

Since $\gamma_2 = \frac{3}{16}$ and $\gamma_3 = \frac{7}{16}$, solving the equalities above, we get $(1-\rho_{22})\, P_2(\{1\}) = \frac{21}{80}$ and $(1-\rho_{33})\, P_3(\{1\}) = \frac{21}{40}$. Also, by (5), we have $P_1(\{1\}) = \gamma_1 + \rho_{21}\, P_2(\{1\}) + \rho_{31}\, P_3(\{1\})$. Noting that $\gamma_1 = \frac{3}{16}$, $\rho_{21} = \frac{1}{3}(1-\rho_{22})$ and $\rho_{31} = \frac{1}{7}(1-\rho_{33})$, we get $P_1(\{1\}) = \frac{3}{16} + \frac{1}{3}(1-\rho_{22})\, P_2(\{1\}) + \frac{1}{7}(1-\rho_{33})\, P_3(\{1\})$, but since $(1-\rho_{22})\, P_2(\{1\}) = \frac{21}{80}$ and $(1-\rho_{33})\, P_3(\{1\}) = \frac{21}{40}$, plugging them in the last equality, we must have have $P_1(\{1\}) = \frac{7}{20}$, which is different from $\phi_1(\{1\}) = \frac{3}{8}$.

Thus, we cannot choose the parameters of the Markov chain choice model to make sure that its choice probabilities match those in Table EC.1. The example that we give in this section is not hard to find. Virtually for all randomly generated instances of our choice model, we cannot calibrate a Markov chain choice model to match the choice probabilities of our choice model.

Appendix B: Proof of Lemma 3.1

Let $H = \{i \in N : \hat{x}_i = \hat{x}_0\}$, $M = \{i \in N : 0 < \hat{x}_i < \hat{x}_0\}$ and $L = \{i \in N : \hat{x}_i = 0\}$. To get a contradiction, assume that $M \neq \varnothing$. We construct two distinct feasible solutions $(\tilde{x}_0, \tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})$ and $(\overline{x}_0, \overline{\boldsymbol{x}}, \overline{\boldsymbol{y}})$ to the Assortment LP such that $(\hat{x}_0, \hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) = \frac{1}{2}(\tilde{x}_0, \tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}) + \frac{1}{2}(\overline{x}_0, \overline{\boldsymbol{x}}, \overline{\boldsymbol{y}})$, contradicting the fact that $(\hat{x}_0, \hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ is a basic feasible solution. For small $\epsilon > 0$, we define the solution $(\tilde{x}_0, \tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})$ as

$$\tilde{x}_0 = \hat{x}_0 - V(M)\,\epsilon,$$
$$\tilde{x}_i = \begin{cases} \hat{x}_i - V(M)\,\epsilon & \text{if } i \in H \\ \hat{x}_i + (1 + V(H))\,\epsilon & \text{if } i \in M \\ \hat{x}_i & \text{if } i \in L, \end{cases} \qquad \tilde{y}_{ij} = \begin{cases} \min\{\tilde{x}_i, \tilde{x}_j\} & \text{if } \hat{y}_{ij} = \min\{\hat{x}_i, \hat{x}_j\} \\ \hat{y}_{ij} & \text{if } \hat{y}_{ij} < \min\{\hat{x}_i, \hat{x}_j\}. \end{cases}$$

We claim that $(\tilde{x}_0, \tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})$ is feasible to the Assortment LP. To see the claim, note that $\tilde{x}_0 + \sum_{i\in N} v_i\, \tilde{x}_i = \hat{x}_0 + \sum_{i \in N} v_i\, \hat{x}_i - V(M)\,\epsilon - \sum_{i\in H} v_i\, V(M)\,\epsilon + \sum_{i \in M} v_i\,(1 + V(H))\,\epsilon = 1$, where the last equality follows by the fact that $\hat{x}_0 + \sum_{i\in N} v_i\, \hat{x}_i = 1$, $\sum_{i \in H} v_i = V(H)$ and $\sum_{i\in M} v_i = V(M)$. Thus, $(\tilde{x}_0, \tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})$ satisfies the first constraint. Noting that $M \neq \varnothing$, we have $\hat{x}_0 > 0$. By the definitions of $\tilde{x}_i$ and $\tilde{x}_0$, for all $i \in H$, we have $\tilde{x}_i = \hat{x}_i - V(M)\,\epsilon = \hat{x}_0 - V(M)\,\epsilon = \tilde{x}_0$. For all $i \in M$, we have $\hat{x}_i < \hat{x}_0$, so for small $\epsilon > 0$, it follows that $\tilde{x}_i = \hat{x}_i + (1 + V(H))\,\epsilon < \hat{x}_0 - V(M)\,\epsilon = \tilde{x}_0$. Lastly, for all $i \in L$, noting that $\hat{x}_i = 0 < \hat{x}_0$, for small $\epsilon > 0$, we get $\tilde{x}_i = \hat{x}_i < \hat{x}_0 - V(M)\,\epsilon = \tilde{x}_0$. Thus, $(\tilde{x}_0, \tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})$ satisfies the second constraint as well. If $\hat{y}_{ij} = \min\{\hat{x}_i, \hat{x}_j\}$, then $\tilde{y}_{ij} = \min\{\tilde{x}_i, \tilde{x}_j\}$, so $\tilde{y}_{ij} \le \tilde{x}_i$ and $\tilde{y}_{ij} \le \tilde{x}_j$. If, on the other hand, $\hat{y}_{ij} < \min\{\hat{x}_i, \hat{x}_j\}$, then $\hat{y}_{ij} < \min\{\hat{x}_i, \hat{x}_j\} - V(M)\,\epsilon$ for small $\epsilon > 0$. Noting that $\tilde{x}_i \ge \hat{x}_i - V(M)\,\epsilon$ for all $i \in N$, we get $\tilde{y}_{ij} = \hat{y}_{ij} < \min\{\hat{x}_i, \hat{x}_j\} - V(M)\,\epsilon \le \min\{\tilde{x}_i, \tilde{x}_j\}$, so $\tilde{y}_{ij} \le \tilde{x}_i$ and $\tilde{y}_{ij} \le \tilde{x}_j$. Thus, $(\tilde{x}_0, \tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})$ satisfies the third and fourth constraints. Also, we have $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}) \in \mathbb{R}_+^{n+n^2}$ for small $\epsilon > 0$, establishing the claim. We define the solution $(\overline{x}_0, \overline{\boldsymbol{x}}, \overline{\boldsymbol{y}})$ as

$$\overline{x}_0 = \hat{x}_0 + V(M)\,\epsilon,$$
$$\overline{x}_i = \begin{cases} \hat{x}_i + V(M)\,\epsilon & \text{if } i \in H \\ \hat{x}_i - (1 + V(H))\,\epsilon & \text{if } i \in M \\ \hat{x}_i & \text{if } i \in L, \end{cases} \qquad \overline{y}_{ij} = \begin{cases} \min\{\overline{x}_i, \overline{x}_j\} & \text{if } \hat{y}_{ij} = \min\{\hat{x}_i, \hat{x}_j\} \\ \hat{y}_{ij} & \text{if } \hat{y}_{ij} < \min\{\hat{x}_i, \hat{x}_j\}. \end{cases}$$

Using the same argument earlier in this paragraph, we can check that $(\overline{x}_0, \overline{\boldsymbol{x}}, \overline{\boldsymbol{y}})$ is feasible to the Assortment LP. Noting that $M \neq \varnothing$, $V(M) > 0$, so $\tilde{x}_0 \neq \overline{x}_0$, which implies that $(\tilde{x}_0, \tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})$ and $(\overline{x}_0, \overline{\boldsymbol{x}}, \overline{\boldsymbol{y}})$ are distinct. By the definitions of $(\tilde{x}_0, \tilde{\boldsymbol{x}})$ and $(\overline{x}_0, \overline{\boldsymbol{x}})$, we have $(\hat{x}_0, \hat{\boldsymbol{x}}) = \frac{1}{2}(\tilde{x}_0, \tilde{\boldsymbol{x}}) + \frac{1}{2}(\overline{x}_0, \overline{\boldsymbol{x}})$, in which case, it only remains to check that $\hat{\boldsymbol{y}} = \frac{1}{2}\tilde{\boldsymbol{y}} + \frac{1}{2}\overline{\boldsymbol{y}}$.

If we have $\hat{y}_{ij} < \min\{\hat{x}_i, \hat{x}_j\}$, then $\tilde{y}_{ij} = \hat{y}_{ij} = \overline{y}_{ij}$, so $\hat{y}_{ij} = \frac{1}{2}\tilde{y}_{ij} + \frac{1}{2}\overline{y}_{ij}$, as desired. Thus, we assume that $\hat{y}_{ij} = \min\{\hat{x}_i, \hat{x}_j\}$. Note that $\tilde{y}_{ij} = \min\{\tilde{x}_i, \tilde{x}_j\}$ in this case. We consider four cases.

Case 1: Assume that $(i,j) \in H \times H$. The definition of H implies that $\hat{x}_i = \hat{x}_j = \hat{x}_0$, so $\hat{y}_{ij} = \min\{\hat{x}_i, \hat{x}_j\} = \hat{x}_0$. Furthermore, if $(i,j) \in H \times H$, then we have $\tilde{x}_i = \hat{x}_i - V(M)\,\epsilon = \hat{x}_0 - V(M)\,\epsilon$ and

$\tilde{x}_j = \hat{x}_j - V(M)\,\epsilon = \hat{x}_0 - V(M)\,\epsilon$, so $\tilde{y}_{ij} = \min\{\tilde{x}_i, \tilde{x}_j\} = \hat{x}_0 - V(M)\,\epsilon$. By the symmetric reasoning, we have $\overline{y}_{ij} = \hat{x}_0 + V(M)\,\epsilon$ as well. In this case, we get $\hat{y}_{ij} = \frac{1}{2}\,\tilde{y}_{ij} + \frac{1}{2}\,\overline{y}_{ij}$.

Case 2: Assume that $(i,j) \in (H, M)$. By the definition of H and M, $\hat{x}_i = \hat{x}_0 > \hat{x}_j$, so $\hat{y}_{ij} = \min\{\hat{x}_i, \hat{x}_j\} = \hat{x}_j$. If $(i,j) \in (H, M)$, then we have $\tilde{x}_i = \hat{x}_i - V(M)\,\epsilon$ and $\tilde{x}_j = \hat{x}_j + (1 + V(H))\,\epsilon$, but noting that $\hat{x}_i > \hat{x}_j$, we get $\tilde{x}_i > \tilde{x}_j$ for small $\epsilon > 0$, so $\tilde{y}_{ij} = \min\{\tilde{x}_i, \tilde{x}_j\} = \tilde{x}_j = \hat{x}_j + (1 + V(H))\,\epsilon$. By the symmetric reasoning, we have $\overline{y}_{ij} = \hat{x}_j - (1 + V(H))\,\epsilon$ as well. Thus, $\hat{y}_{ij} = \frac{1}{2}\,\tilde{y}_{ij} + \frac{1}{2}\,\overline{y}_{ij}$.

Case 3: Assume that $(i,j) \in (M, H)$ or $(i,j) \in (M, M)$. In this case, by using the same argument in Case 2, we can show that $\hat{y}_{ij} = \frac{1}{2}\,\tilde{y}_{ij} + \frac{1}{2}\,\overline{y}_{ij}$.

Case 4: Assume that $i \in L$ or $j \in L$. Let $\ell \in \{i,j\}$ be such that $\ell \in L$. The definition of L implies that $\hat{x}_\ell = 0$, so $\hat{y}_{ij} = \min\{\hat{x}_i, \hat{x}_j\} \le \hat{x}_\ell = 0$. Furthermore, for $\ell \in L$, we have $\tilde{x}_\ell = \hat{x}_\ell = 0$, in which case, we get $\tilde{y}_{ij} = \min\{\tilde{x}_i, \tilde{x}_j\} \le \tilde{x}_\ell = 0$. By the symmetric reasoning, we have $\overline{y}_{ij} = 0$ as well. In this case, it follows that $\hat{y}_{ij} = 0 = \frac{1}{2}\,\tilde{y}_{ij} + \frac{1}{2}\,\overline{y}_{ij}$. ■

Appendix C: Proof of Theorem 4.1

We start this section with the following lemma that is used in the proof of Lemma 4.2, which, in turn, becomes useful to give a proof for Theorem 4.1. Recall that $R(S) = \sum_{i \in S} r_i\, v_i / (1 + V(S))$.

Lemma C.1 *(a) Let the assortments $S, K \subseteq N$ be such that $K \cap S = \varnothing$ and $r_i \ge R(S)$ for all $i \in K$. Then, we have $R(S \cup K) \ge R(K)$.*

(b) Let the assortments $S, K \subseteq N$ be such that $K \subseteq S$ and $r_i \le R(S)$ for all $i \in K$. Then, we have $R(S \setminus K) \ge R(K)$.

Proof: For the first part, note that $\frac{\sum_{i \in K} v_i\, r_i}{V(K)}$ is a weighted average of the revenues of the products in K. Since $r_i \ge R(S)$ for all $i \in K$, we get $\frac{\sum_{i \in K} v_i\, r_i}{V(K)} \ge R(S)$. By the definition of $R(S)$, we have

$$R(S \cup K) = \frac{1 + V(S)}{1 + V(S \cup K)}\, R(S) + \frac{V(K)}{1 + V(S \cup K)}\, \frac{\sum_{i \in K} v_i\, r_i}{V(K)},$$

which implies that $R(S \cup K)$ is a convex combination of $R(S)$ and $\frac{\sum_{i \in K} v_i\, r_i}{V(K)}$. Thus, since $\frac{\sum_{i \in K} v_i\, r_i}{V(K)} \ge R(S)$, it must be the case that $R(S \cup K) \ge R(S)$. The proof of the second part is similar. ■

Proof of Theorem 4.1:

In the remainder of this section, we give a proof for the second part of Theorem 4.1. Let $\delta > 0$ be such that $S^*(\lambda, \alpha) = S^*(\lambda, 0)$ for all $\alpha \in [0, \delta)$ and $S^*(\lambda, \delta) \ne S^*(\lambda, 0)$. In other words, as we progressively increase the revenues of the products by larger amounts, δ is the first increment when the optimal solution to the Parametric Mixture problem changes. It is enough to show that

$S^*(\lambda,\delta) \supseteq S^*(\lambda,0)$. Once we show this result, we can set the nominal revenues of the products as $\{r_i + \delta : i \in N\}$ and progressively increase the revenues starting from these nominal values.

Having fixed δ as in the previous paragraph, letting $R(S)$ be as defined at the beginning of this section, note that there are finitely many values in the set $\{R(S) + \delta \frac{V(S)}{1+V(S)} : S \subseteq N\}$. Thus, there exists $\epsilon > 0$ such that if $R(S) + \delta \frac{V(S)}{1+V(S)} \neq R(Q) + \delta \frac{V(Q)}{1+V(Q)}$ for some $S, Q \subseteq N$, then we must have $\left|R(S) + \delta \frac{V(S)}{1+V(S)} - R(Q) - \delta \frac{V(Q)}{1+V(Q)}\right| > \epsilon$. So, if $R(S) + \delta \frac{V(S)}{1+V(S)}$ and $R(Q) + \delta \frac{V(Q)}{1+V(Q)}$ are different, then they must differ by at least $\epsilon > 0$. For $V_{\min} = \min_{i\in N} v_i$, fix $\alpha \in [0,\delta)$ as

$$0 < \left(1 + \lambda \frac{1+V(N)}{V_{\min}} \sum_{i\in N} \theta_i\right)(\delta - \alpha) \leq \epsilon. \tag{6}$$

Since $\alpha < \delta$, we have $S^*(\lambda,\alpha) = S^*(\lambda,0)$. Building on the notation introduced right before Lemma 4.2, we define two instances of the $\mathsf{Mixture}$ problem with $r_{i1} = r_i + \alpha$ and $r_{i2} = r_i + \delta$ for all $i \in N$ and $\lambda_1 = \lambda = \lambda_2$. Letting $R_\ell(S) = \sum_{i\in S} r_{i\ell}\, v_i/(1+V(S))$ and recalling that S^*_ℓ is an optimal solution to the problem $\max_{S\subseteq N}\left\{R_\ell(S) + \lambda_\ell \sum_{i\in S} r_{i\ell}\,\theta_i\right\}$, showing that $S^*(\lambda,\delta) \supseteq S^*(\lambda,0) = S^*(\lambda,\alpha)$ is equivalent to showing that $S^*_2 \supseteq S^*_1$.

We consider two cases. The first case will lead to a contradiction, so it cannot happen. In the second case, we establish the desired result. The first case is the more involved.

Case 1: Assume that $R_1(S^*_1) - \alpha < R_2(S^*_2) - \delta$. Define K as in Lemma 4.2, in which case, we have $S^*_1 \cup K \supseteq S^*_2$ by this lemma. First, we proceed under the assumption that $K \neq \varnothing$.

By the definition of K, we have $r_i + \alpha = r_{i1} \leq R_1(S^*_1)$, which we equivalently, write as $r_i \leq R_1(S^*_1) - \alpha \leq R_2(S^*_2) - \delta$, where the last inequality holds since we have $R_1(S^*_1) - \alpha < R_2(S^*_2) - \delta$ in the case we consider. Therefore, we obtain $r_{i2} = r_i + \delta \leq R_2(S^*_2)$ for all $i \in K$. In this case, the second statement in Lemma 4.2 holds. Noting that $\sum_{i\in K} r_{i1}\, v_i/V(K) = \sum_{i\in K}(r_i+\alpha)\, v_i/V(K) = \alpha + \sum_{i\in N} r_i\, v_i/V(K)$, multiplying both sides of (2) with $1 + V(S^*_1 \cup K)$, we get

$$\begin{aligned} R_1(S^*_1) - \alpha - \frac{\sum_{i\in K} r_i\, v_i}{V(K)} &\geq \frac{1+V(S^*_1\cup K)}{1+V(S^*_2)}\left(R_2(S^*_2) - \delta - \frac{\sum_{i\in K} r_i\, v_i}{V(K)}\right) - \lambda\, \frac{1+V(S^*_1\cup K)}{V(K)} \sum_{i\in K}(r_{i2} - r_{i1})\,\theta_i \\ &\geq \frac{1+V(S^*_1\cup K)}{1+V(S^*_2)}\left(R_2(S^*_2) - \delta - \frac{\sum_{i\in K} r_i\, v_i}{V(K)}\right) - \lambda\,(\delta-\alpha)\,\frac{1+V(N)}{V_{\min}} \sum_{i\in K} \theta_i, \end{aligned}$$

where the last inequality uses the fact that $r_{i2} - r_{i1} = \delta - \alpha \geq 0$, $V(N) \geq V(S^*_1 \cup K)$ and $V_{\min} \leq V(K)$. By the discussion at the beginning of this paragraph, we have $r_i + \delta \leq R_2(S^*_2)$ for all $i \in K$. Note that $\sum_{i\in K} r_i\, v_i/V(K)$ is a weighted average of the revenues $\{r_i : i \in K\}$. Since $r_i \leq R_2(S^*_2) - \delta$ for all $i \in K$, we obtain $\sum_{i\in K} r_i\, v_i/V(K) \leq R_2(S^*_2) - \delta$, which is to say that $R_2(S^*_2) - \delta - \sum_{i\in K} r_i\, v_i/V(K) \geq 0$. Also, by (6), we have $\lambda\, \frac{1+V(N)}{V_{\min}}(\delta-\alpha)\sum_{i\in K}\theta_i \leq \epsilon - (\delta - \alpha)$.

Lastly, since $S_1^* \cup K \supseteq S_2^*$, we have $1+V(S_1^* \cup K) \geq 1+V(S_2^*)$. In this case, noting the last chain of inequalities displayed above, using the fact that $R_2(S_2^*) - \delta - \sum_{i\in K} r_i\, v_i / V(K) \geq 0$, we get

$$R_1(S_1^*) - \alpha - \frac{\sum_{i\in K} r_i\, v_i}{V(K)} \geq R_2(S_2^*) - \delta - \frac{\sum_{i\in K} r_i\, v_i}{V(K)} - \epsilon + \delta - \alpha,$$

which we equivalently write as $R_1(S_1^*) \geq R_2(S_2^*) - \epsilon$. Furthermore, since $\delta \geq \alpha$, we have $r_{i2} = r_i + \delta \geq r_i + \alpha = r_{i1}$ for all $i \in N$, which implies that $R_2(S) = \frac{\sum_{i\in S} r_{i2}\, v_i}{V(S)} \geq \frac{\sum_{i\in S} r_{i1}\, v_i}{V(S)} = R_1(S)$ for all $S \subseteq N$. Thus, noting that $R_1(S_1^*) \geq R_2(S_2^*) - \epsilon$, we obtain $R_2(S_1^*) \geq R_1(S_1^*) \geq R_2(S_2^*) - \epsilon$. On the other hand, since $R_1(S_1^*) - \alpha < R_2(S_2^*) - \delta$ in the case we consider, we get $R_2(S_2^*) > R_1(S_1^*) + \delta - \alpha = \frac{\sum_{i\in S_1^*} r_{i1}\, v_i}{V(S_1^*)} + \delta - \alpha = \frac{\sum_{i\in S_1^*} (r_{i2} + \alpha - \delta)\, v_i}{V(S_1^*)} + \delta - \alpha = R_2(S_1^*) + (\delta - \alpha)\left(1 - \frac{V(S_1^*)}{V(S_1^*)}\right) = R_2(S_1^*)$. Focusing on the first and last terms in this chain of inequalities yields $R_2(S_2^*) > R_2(S_1^*)$.

By the discussion at the end of the previous paragraph, we have the two inequalities $R_2(S_1^*) \geq R_2(S_2^*) - \epsilon$ and $R_2(S_2^*) > R_2(S_1^*)$, which imply that $|R_2(S_2^*) - R_2(S_1^*)| \leq \epsilon$ and $R_2(S_2^*) \neq R_2(S_1^*)$. Letting $R(S)$ be as defined at the beginning of this section, we have $R_2(S) = \frac{\sum_{i\in S}(r_i+\delta)\, v_i}{V(S)} = R(S) + \delta \frac{V(S)}{1+V(S)}$. So, the assortments $S_1^*, S_2^* \subseteq N$ satisfy $\left| R(S_2^*) + \delta \frac{V(S_2^*)}{1+V(S_2^*)} - R(S_1^*) - \delta \frac{V(S_1^*)}{1+V(S_1^*)} \right| \leq \epsilon$, while $R(S_2^*) + \delta \frac{V(S_2^*)}{1+V(S_2^*)}$ and $R(S_1^*) + \delta \frac{V(S_1^*)}{1+V(S_1^*)}$ being distinct from each other, which contradicts the definition of ϵ at the beginning of the proof.

Second, we proceed under the assumption that $K = \varnothing$. Since $S_1^* \cup K \supseteq S_2^*$, we obtain $S_1^* \supseteq S_2^*$. To rule out the possibility that $S_1^* = S_2^*$, observe that if $S_1^* = S_2^*$, then using $\hat{S}$ to denote the common value of S_1^* and S_2^*, we obtain $R_2(\hat{S}) - R_1(\hat{S}) = \frac{\sum_{i\in \hat{S}}(\delta-\alpha) v_i}{1+V(\hat{S})} = (\delta - \alpha)\frac{V(\hat{S})}{1+V(\hat{S})} < \delta - \alpha$, which yields $R_2(S_2^*) - \delta = R_2(\hat{S}) - \delta < R_1(\hat{S}) - \alpha = R_1(S_1^*) - \alpha$, contradicting the fact that we assume $R_1(S_1^*) - \alpha < R_2(S_2^*) - \delta$ in the case we consider. Therefore, S_2^* must be a strict subset of S_1^*, which implies that $\frac{V(S_1^*)}{1+V(S_1^*)} > \frac{V(S_2^*)}{1+V(S_2^*)}$ and $\sum_{i\in S_1^*} \theta_i > \sum_{i\in S_2^*} \theta_i$.

Recall that S_ℓ^* is an optimal solution to the problem $\max_{S\subseteq N} \left\{ R_\ell(S) + \lambda_\ell \sum_{i\in N} r_{i\ell}\, \theta_i \right\}$. Noting that $r_{i1} - \alpha = r_i = r_{i2} - \delta$, we obtain the chain of inequalities

$$\begin{aligned}
R_1(S_2^*) + (\delta - \alpha)\frac{V(S_2^*)}{1+V(S_2^*)} &+ \lambda \sum_{i\in S_2^*} r_{i1}\,\theta_i + \lambda\,(\delta-\alpha)\sum_{i\in S_2^*}\theta_i \\
&= R_2(S_2^*) + \lambda \sum_{i\in S_2^*} r_{i2}\,\theta_i \;\geq\; R_2(S_1^*) + \lambda \sum_{i\in S_1^*} r_{i2}\,\theta_i \\
&= R_1(S_1^*) + (\delta-\alpha)\frac{V(S_1^*)}{1+V(S_1^*)} + \lambda \sum_{i\in S_1^*} r_{i1}\,\theta_i + \lambda\,(\delta-\alpha)\sum_{i\in S_1^*}\theta_i,
\end{aligned}$$

where the equalities follows by arranging the terms and the inequality follows from the fact that S_2^* is an optimal solution to the problem $\max_{S\subseteq N} \left\{ R_2(S) + \lambda_2 \sum_{i\in N} r_{i2}\,\theta_i \right\}$. Since $\frac{V(S_1^*)}{1+V(S_1^*)} > \frac{V(S_2^*)}{1+V(S_2^*)}$ and

$\sum_{i\in S_1^*}\theta_i > \sum_{i\in S_2^*}\theta_i$, the last chain of inequalities implies that $R_1(S_2^*)+\lambda\sum_{i\in S_2^*} r_{i1}\,\theta_i > R_1(S_1^*)+\lambda\sum_{i\in S_1^*} r_{i1}\,\theta_i$, which contradicts the definition of S_1^*.

Case 2: Assume that $R_2(S_2^*)-\delta \le R_1(S_1^*)-\alpha$. We follow an outline similar to the one in the proof of the first part of the theorem in the main text to show that $S_2^* \supseteq S_1^*$.

Observe that the choice of the first and second instances in Lemma 4.2 is arbitrary. Thus, this lemma holds when we interchange the roles of the first and second instances. Define

$$K=\{i\in N : r_{i2}\le R_2(S_2^*),\ i\notin S_2^*,\ i\in S_1^*\},$$

which is the analogue of assortment K in Lemma 4.2 after interchanging the roles of the first and second instances. In this case, by this lemma, we have $S_2^*\cup K\supseteq S_1^*$.

If $K=\varnothing$, then we get $S_2^*\supseteq S_1^*$, which is the desired result. To get a contradiction, assume that we have $K\neq\varnothing$. By the definition of K, we have $r_{i2}\le R_2(S_2^*)$ for all $i\in K$. Noting that $r_{i2}=r_i+\delta$, we obtain $r_i\le R_2(S_2^*)-\delta\le R_1(S_1^*)-\alpha$ for all $i\in K$, where the second inequality uses the fact that $R_2(S_2^*)-\delta\le R_1(S_1^*)-\alpha$ in the case we consider. Thus, we have $r_{i1}=r_i+\alpha\le R_1(S_1^*)$ for all $i\in K$. Since $r_{i1}\le R_1(S_1^*)$ for all $i\in K$, noting that $\sum_{i\in K} r_{i1}v_i/V(K)$ is a weighted average of the revenues $\{r_{i1}: i\in K\}$, we obtain $\sum_{i\in K} r_{i1}\,v_i/V(K)\le R_1(S_1^*)$.

In the previous paragraph, we show that $r_{i1}\le R_1(S_1^*)$ for all $i\in K$, so (2) in Lemma 4.2 holds after interchanging the roles of the first and second instances. Thus, by (2), we get

$$R_2(S_2^*)-\frac{\sum_{i\in K} r_{i2}\,v_i}{V(K)}\ge\frac{1+V(S_2^*\cup K)}{1+V(S_1^*)}\left(R_1(S_1^*)-\frac{\sum_{i\in K} r_{i1}\,v_i}{V(K)}\right)-\lambda\,\frac{1+V(S_2^*\cup K)}{V(K)}\sum_{i\in K}(r_{i1}-r_{i2})\,\theta_i,$$

which we obtain by multiplying both sides of (2) by $1+V(S_2^*\cup K)$, arranging the terms and noting that $\lambda_1=\lambda=\lambda_2$ in our definition of the two instances.

Since $\delta>\alpha$, we have $r_{i2}=r_i+\delta>r_i+\alpha=r_{i1}$. Thus, the last term that we subtract on the right side of the inequality above is negative, which implies that if we drop this term, then the inequality remains valid. Since $S_2^*\cup K\supseteq S_1^*$, we have $\frac{1+V(S_2^*\cup K)}{1+V(S_1^*)}\ge 1$. In this case, noting that $\frac{\sum_{i\in K} r_{i1}\,v_i}{V(K)}\le R_1(S_1^*)$ by our earlier discussion, if we drop the term $\frac{1+V(S_2^*\cup K)}{1+V(S_1^*)}$ on the right side of the inequality above, then the inequality remains valid. Thus, dropping the two terms mentioned, we get $R_2(S_2^*)-\frac{\sum_{i\in K} r_{i2}\,v_i}{V(K)}>R_1(S_1^*)-\frac{\sum_{i\in K} r_{i1}\,v_i}{V(K)}$ by the inequality above.

Note that $\sum_{i\in K} r_{i1}\,v_i=\sum_{i\in K}(r_i+\alpha)\,v_i=\sum_{i\in K} r_i\,v_i+\alpha\,V(K)$. Similarly, we have $\sum_{i\in K} r_{i2}\,v_i=\sum_{i\in K} r_i\,v_i+\delta\,V(K)$ as well. So, the inequality $R_2(S_2^*)-\frac{\sum_{i\in K} r_{i2}\,v_i}{V(K)}>R_1(S_1^*)-\frac{\sum_{i\in K} r_{i1}\,v_i}{V(K)}$ yields

$$R_2(S_2^*)-\frac{\sum_{i\in K} r_i\,v_i}{V(K)}-\delta>R_1(S_1^*)-\frac{\sum_{i\in K} r_i\,v_i}{V(K)}-\alpha,$$

which is equivalent to $R_2(S_2^*)-\delta>R_1(S_1^*)-\alpha$. Therefore, we get a contradiction to the fact that we have $R_2(S_2^*)-\delta\le R_1(S_1^*)-\alpha$ in the case we consider.

Appendix D: Proof of Theorem 5.2

To give our FPTAS, for a fixed accuracy parameter $\rho > 0$, we construct the grid points $\mathsf{Grid} = \{(1+\rho)^k : k = \lfloor \frac{\log V_{\min}}{\log(1+\rho)} \rfloor, \ldots, \lceil \frac{\log(n V_{\max})}{\log(1+\rho)} \rceil\}$. For each $t \in \mathsf{Grid}$, let $\nu_{it} = \lceil \frac{n}{t\rho} v_i \rceil$. Consider the problem

$$\max_{S \subseteq N} \left\{ \sum_{i \in S} r_i \left(\frac{v_i}{1+t} + \lambda\, \theta_i \right) \ : \ \sum_{i \in S} \nu_{it} \le \left\lceil \frac{n}{\rho} \right\rceil + n, \quad |S| = C \right\}. \tag{7}$$

Letting S^* be an optimal solution to the $\mathsf{Cardinality\text{-}Mixture}$ problem, by the construction of Grid, there exists $t \in \mathsf{Grid}$ such that $V(S^*) \le t \le (1+\rho) V(S^*)$.

In the next lemma, we show that solving problem (7) with some t such that $V(S^*) \le t \le (1+\rho)V(S^*)$ yields a $(1+5\rho)$-approximate solution to the $\mathsf{Cardinality\text{-}Mixture}$ problem.

Lemma D.1 *Using S^* to denote an optimal solution to the* $\mathsf{Cardinality\text{-}Mixture}$ *problem with the optimal objective value z^*, let $\hat t$ be such that $V(S^*) \le \hat t \le (1+\rho)V(S^*)$. Then, letting $\hat S$ be an optimal solution to problem* (7) *with $t = \hat t$, the expected revenue of the assortment $\hat S$ is at least $\frac{1}{(1+\rho)\,(1+2\rho)}\, z^*$.*

Proof: By the definition of ν_{it}, we have $\nu_{i\hat t} \le \frac{n}{\hat t \rho} v_i + 1$. Also, we have $V(S^*) \le \hat t$. In this case, we get $\sum_{i \in S^*} \nu_{i\hat t} \le \sum_{i \in S^*} (\frac{n}{\hat t \rho} v_i + 1) \le \frac{n}{\hat t\, \rho} V(S^*) + n \le \frac{n}{\rho} + n = \lceil \frac{n}{\rho} \rceil + n$. Thus, S^* is a feasible solution to problem (7) when we solve this problem with $t = \hat t$. On the other hand, we have

$$V(\hat S) = \sum_{i \in \hat S} v_i \le \frac{\hat t\, \rho}{n} \sum_{i \in \hat S} \left\lceil \frac{n}{\hat t\, \rho} v_i \right\rceil = \frac{\hat t\, \rho}{n} \sum_{i \in \hat S} \nu_{i\hat t} \overset{(a)}{\le} \left(\frac{\hat t\, \rho}{n} \right) \left(\left\lceil \frac{n}{\rho} \right\rceil + n \right) \le \left(\frac{\hat t\, \rho}{n} \right) \left(\frac{n}{\rho} + 1 + n \right) \le (1+2\rho)\, \hat t,$$

where (a) holds since $\hat S$ is an optimal solution to problem (7) when we solve this problem with $t = \hat t$, so it satisfies the first constraint in this problem.

In this case, if we evaluate the expected revenue provided by the assortment $\hat S$ and note that $V(\hat S) \le (1+2\rho)\, \hat t$ and $(1+\rho)\, V(S^*) \ge \hat t$, we obtain

$$\begin{aligned}
\sum_{i \in \hat S} r_i \left(\frac{v_i}{1 + V(\hat S)} + \lambda\, \theta_i \right) &\ge \sum_{i \in \hat S} r_i \left(\frac{v_i}{1 + (1+2\rho)\, \hat t} + \lambda\, \theta_i \right) \ge \frac{1}{1+2\rho} \sum_{i \in \hat S} r_i \left(\frac{v_i}{1+\hat t} + \lambda\, \theta_i \right) \\
&\overset{(b)}{\ge} \frac{1}{1+2\rho} \sum_{i \in S^*} r_i \left(\frac{v_i}{1+\hat t} + \lambda\, \theta_i \right) \ge \frac{1}{1+2\rho} \sum_{i \in S^*} r_i \left(\frac{v_i}{1 + (1+\rho) V(S^*)} + \lambda\, \theta_i \right) \\
&\ge \frac{1}{(1+\rho)\,(1+2\rho)} \sum_{i \in S^*} r_i \left(\frac{v_i}{1 + V(S^*)} + \lambda\, \theta_i \right) = \frac{z^*}{(1+\rho)\,(1+2\rho)},
\end{aligned}$$

where (b) uses the fact that S^* is a feasible but not necessarily an optimal solution to problem (7) when we solve this problem with $t = \hat t$. ■

For $\rho \le 1$, we have $(1+\rho)\,(1+2\rho) \le 1 + 5\rho$, showing that the assortment $\hat S$ in the lemma above is a $(1+5\rho)$-approximation when $\rho \le 1$. Note that problem (7) is a knapsack problem with a cardinality constraint, where the space consumption of each product i is ν_{it}, which is an integer. Thus, for fixed t, we can solve this problem through the dynamic program

$$J_i(p,q;t) = \max\left\{ r_i\left(\frac{v_i}{1+t} + \lambda\,\theta_i\right) + J_{i+1}(p+\nu_{it}, q+1; t)\,,\, J_{i+1}(p,q;t)\right\},$$

with the boundary condition that $J_{n+1}(p,q;t) = -\infty$ if $p > \lceil \frac{n}{\rho} \rceil + n$ or $q \neq C$. In the state variable, the first and second components, respectively, keep track of the total space consumption of the offered products and the number of offered products. The boundary condition rules out the solutions where the total capacity consumption of the offered products exceeds the capacity of the knapsack or the total number of offered products is not equal to C. Thus, if the first component of the state variable exceeds $\lceil \frac{n}{\rho} \rceil + n$, then we can immediately conclude that the value function at this state is $-\infty$. In this case, since $C \leq n$, number of states in the dynamic program above is $O(\frac{n^2}{\rho})$. There are n decision epochs, so for fixed t, we can solve the dynamic program in $O(\frac{n^3}{\rho})$ operations, giving an optimal solution to problem (7). By the definition of Grid, we have $O\Big(\frac{1}{\log(1+\rho)} \log\Big(\frac{nV_{\max}}{V_{\min}}\Big)\Big) = O\Big(\frac{1}{\rho} \log\Big(\frac{nV_{\max}}{V_{\min}}\Big)\Big)$ grid points. Thus, we can obtain an optimal solution to problem (7) for all $t \in \mathsf{Grid}$ in $O\Big(\frac{n^3}{\rho^2} \log\Big(\frac{nV_{\max}}{V_{\min}}\Big)\Big)$ operations. Here is the proof of Theorem 5.2.

Proof of Theorem 5.2:

Given $\epsilon \in (0,1)$, we choose the accuracy parameter as $\rho = \epsilon/5$. By the discussion in the previous paragraph, we can obtain an optimal solution to problem (7) for all $t \in \mathsf{Grid}$ in $O\Big(\frac{n^3}{\epsilon^2} \log\Big(\frac{nV_{\max}}{V_{\min}}\Big)\Big)$ operations. Since $\rho = \epsilon/5 < 1$, we have $(1+\rho)(1+2\rho) \leq 1+5\rho = 1+\epsilon \leq \frac{1}{1-\epsilon}$, so by Lemma D.1, the expected revenue from one of these solutions is at least $(1-\epsilon)\,z^*$. Thus, if we check the expected revenue provided by the solution to problem (7) for each $t \in \mathsf{Grid}$ and pick the best one, then the best solution provides an expected revenue of at least $(1-\epsilon)\,z^*$. The number of operations to check the expected revenue from each solution is dominated by the number of operations to get an optimal solution to problem (7) for all $t \in \mathsf{Grid}$. Thus, in $O\Big(\frac{n^3}{\epsilon^2} \log\Big(\frac{nV_{\max}}{V_{\min}}\Big)\Big)$ operations, we can obtain a solution to the Cardinality-Mixture problem with expected revenue at least $(1-\epsilon)\,z^*$.

Appendix E: Proof of Lemma 7.1

In this section, we give a proof for Lemma 7.1. By the discussion at the beginning of Section 7, recall that if the Compact LP has multiple optimal solutions, then we choose the one that has the largest value for the decision variable x_0. Furthermore, to obtain a solution that has the largest value for the decision variable for x_0, for $\epsilon > 0$, we can add the additional term $\epsilon\, x_0$ to the objective function of the Compact LP. If ϵ is small enough, then solving the Compact LP with the additional term provides an optimal solution to the original version of the Compact LP that has the largest value for the decision variable x_0. Thus, we consider the problem

$$\max_{(x_0,\boldsymbol{x},\boldsymbol{y})\in\mathbb{R}\times\mathbb{R}_+^{n+n^2}} \Bigg\{ T\sum_{i\in N} r_i \Bigg((v_i+\lambda\,\theta_i)\,x_i + \lambda\,\theta_i \sum_{j\in N} v_j\,y_{ij}\Bigg) + \epsilon\,x_0 \quad : \tag{8}$$
$$T\sum_{i\in N} a_{qi}\Bigg((v_i+\lambda\,\theta_i)\,x_i + \lambda\,\theta_i \sum_{j\in N} v_j\,y_{ij}\Bigg) \le c_q \quad \forall\, q\in M,$$
$$x_0 + \sum_{i\in N} v_i\,x_i = 1,$$
$$x_i \le x_0 \quad \forall\, i\in N,$$
$$y_{ij}\le x_i \quad \forall\, i,j\in N, \quad y_{ij}\le x_j \quad \forall\, i,j\in N \Bigg\}.$$

If ϵ is small enough, then a basic optimal solution to the problem above is also a basic optimal solution to problem the Compact LP. So, it is enough to show that if $(x_0^*,\boldsymbol{x}^*,\boldsymbol{y}^*)$ is a basic optimal solution to problem (8), then we have $y_{ij}^* = \min\{x_i^*, x_j^*\}$ for all $i,j\in N$. For notational brevity, we let $\mathcal{P} = \{(x_0,\boldsymbol{x},\boldsymbol{y})\in\mathbb{R}\times\mathbb{R}_+^{n+n^2} : x_0 + \sum_{i\in N} v_i\,x_i = 1,\ x_i\le x_0\ \forall\, i\in N,\ y_{ij}\le \min\{x_i,x_j\}\ \forall\, i,j\in N\}$, denoting the polytope captured by the last four constraints in the LP above. The proof of Lemma 7.1 uses two lemmas, which closely resemble results already established in the main text.

In the first lemma, we consider a slightly modified version the Assortment LP, where we add the additional term $\epsilon\,x_0$ to the objective function. In particular, consider the LP

$$\max_{(x_0,\boldsymbol{x},\boldsymbol{y})\in\mathbb{R}\times\mathbb{R}_+^{n+n^2}} \Bigg\{ \sum_{i\in N} r_i \Bigg((v_i+\lambda\,\theta_i)\,x_i + \lambda\,\theta_i \sum_{j\in N} v_j\,y_{ij}\Bigg) + \epsilon\,x_0 \quad : \quad (x_0,\boldsymbol{x},\boldsymbol{y})\in\mathcal{P}\Bigg\}. \tag{9}$$

In the next lemma, we relate an optimal solution to the LP above to an optimal solution of a slightly modified version of the Mixture problem.

Lemma E.1 *For a basic optimal solution $(x_0^*,\boldsymbol{x}^*,\boldsymbol{y}^*)$ to problem (9), let $S^* = \{i\in N : x_i^* > 0\}$. Then, S^* is an optimal solution to the problem*

$$\max_{S\subseteq N} \Bigg\{ \sum_{i\in S} r_i \Bigg(\frac{v_i}{1+V(S)} + \lambda\,\theta_i\Bigg) + \frac{\epsilon}{1+V(S)}\Bigg\}. \tag{10}$$

Lemma E.1 is an analogue of Theorem 3.2 and its proof follows the same line of reasoning that we used in the proof of Theorem 3.2. We skip the proof.

In the second lemma, we relate problem (8) to a slightly modified version of the Choice-Based LP. We can view this lemma as an analogue of Theorem 6.1.

Lemma E.2 *Consider the* Choice-Based LP *after adding the additional term* $\sum_{S \subseteq N} \frac{\epsilon}{1+V(S)} w(S)$ *to the objective function, which is given by*

$$\max_{\boldsymbol{w} \in \mathbb{R}_+^{2^n}} \left\{ T \sum_{S \subseteq N} \sum_{i \in S} r_i \left(\frac{v_i}{1+V(S)} + \lambda \theta_i \right) w(S) + \sum_{S \subseteq N} \frac{\epsilon}{1+V(S)} w(S) \quad : \right. \tag{11}$$
$$T \sum_{S \subseteq N} \sum_{i \in S} a_{qi} \left(\frac{v_i}{1+V(S)} + \lambda \theta_i \right) w(S) \le c_q \quad \forall q \in M,$$
$$\left. \sum_{S \subseteq N} w(S) = 1 \right\}.$$

Then, the optimal objective values of problems (8) *and* (11) *are the same. Furthermore, the optimal values of the dual variables for the first constraint in problems* (8) *and* (11) *are the same.*

The proof of the lemma above uses the same reasoning that we use in the proof of Theorem 6.1 in conjunction with Lemma E.1. We skip the proof.

Next, using the dual variables $\boldsymbol{\mu} = \{\mu_q : q \in M\}$, π, $\boldsymbol{\alpha} = \{\alpha_i : i \in N\}$, $\boldsymbol{\eta} = \{\eta_{ij} : i, j \in N\}$ and $\boldsymbol{\sigma} = \{\sigma_{ij} : i, j \in N\}$, the dual of problem (8) is given by

$$\min_{\substack{(\boldsymbol{\mu}, \pi, \boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{\sigma}) \in \\ \mathbb{R}_+^m \times \mathbb{R} \times \mathbb{R}_+^{n+2n^2}}} \left\{ \sum_{q \in M} c_q \mu_q + \pi \quad : \right. \tag{12}$$
$$\pi = \sum_{i \in N} \alpha_i + \epsilon,$$
$$v_i \pi + \alpha_i - \sum_{j \in N} \eta_{ij} - \sum_{j \in N} \sigma_{ji} \ge T (v_i + \lambda \theta_i) \left(r_i - \sum_{q \in M} a_{qi} \mu_q \right) \quad \forall i \in N,$$
$$\left. \eta_{ij} + \sigma_{ij} \ge T \lambda \theta_i v_j \left(r_i - \sum_{q \in M} a_{qi} \mu_q \right) \quad \forall i, j \in N \right\}.$$

In the next lemma, we use complementary slackness to give two useful properties that are satisfied by an optimal primal-dual solution pair for problem (8).

Lemma E.3 *Let* $(x_0^*, \boldsymbol{x}^*, \boldsymbol{y}^*)$ *and* $(\boldsymbol{\mu}^*, \pi^*, \boldsymbol{\alpha}^*, \boldsymbol{\eta}^*, \boldsymbol{\sigma}^*)$ *be a basic optimal primal-dual solution pair for problem* (8) *and* $S^* = \{i \in N : x_i^* > 0\}$*. Then, we have*

$$\pi^* = \sum_{i \in S^*} \alpha_i^* + \epsilon,$$
$$\sum_{i \in S^*} \sum_{j \in N} (\eta_{ij}^* + \sigma_{ji}^*) = \sum_{i \in S^*} \sum_{j \in S^*} (\eta_{ij}^* + \sigma_{ij}^*).$$

Proof: To see the first equality, note that $x_0^* > 0$. Otherwise, we have $x_i^* = 0$ for all $i \in N$ by the third constraint in problem (8), in which case, it is impossible to satisfy the second constraint.

Since $x_0^* > 0$ and $x_i^* = 0$ for all $i \notin S^*$, using complementary slackness on the third constraint in problem (8), we have $\alpha_i^* = 0$ for all $i \notin S^*$, in which case, by the first constraint in problem (12), we get $\pi^* = \sum_{i \in S^*} \alpha_i^* + \epsilon$. To see the second equality, if $i \in S^*$ and $j \notin S^*$, then $x_i^* > 0$ and $x_j^* = 0$, in which case, by the last two constraints in problem (8), we have $y_{ij}^* = 0$ and $y_{ji}^* = 0$. Therefore, we get $y_{ij}^* < x_i^*$ and $y_{ji}^* < x_i^*$, so using complementary slackness on the last two constraints in problem (8), we get $\eta_{ij}^* = 0$ and $\sigma_{ji}^* = 0$. Thus, if $i \in S^*$ and $j \notin S^*$, then $\eta_{ij}^* = 0$ and $\sigma_{ji}^* = 0$. In this case, the second equality in the lemma follows by noting that

$$\begin{aligned}\sum_{i \in S^*} \sum_{j \in N} (\eta_{ij}^* + \sigma_{ji}^*) &= \sum_{i \in S^*} \sum_{j \in S^*} (\eta_{ij}^* + \sigma_{ji}^*) + \sum_{i \in S^*} \sum_{j \notin S^*} (\eta_{ij}^* + \sigma_{ji}^*) \\ &= \sum_{i \in S^*} \sum_{j \in S^*} (\eta_{ij}^* + \sigma_{ji}^*) = \sum_{i \in S^*} \sum_{j \in S^*} (\eta_{ij}^* + \sigma_{ij}^*). \qquad \blacksquare\end{aligned}$$

Proof of Lemma 7.1:

Let $(x_0^*, \boldsymbol{x}^*, \boldsymbol{y}^*)$ and $(\boldsymbol{\mu}^*, \pi^*, \boldsymbol{\alpha}^*, \boldsymbol{\eta}^*, \boldsymbol{\sigma}^*)$ be a basic optimal primal-dual solution pair for problem (8). By the discussion right after problem (8), it is enough show that $y_{ij}^* = \min\{x_i^*, x_j^*\}$ for all $i, j \in N$. Let $S^* = \{i \in N : x_i^* > 0\}$. Consider $i, j \in S^*$. We have $x_i^* > 0$ and $x_j^* > 0$, so using complementary slackness on the last two constraints in problem (8), if $y_{ij}^* = 0$, then $\eta_{ij}^* = 0$ and $\sigma_{ij}^* = 0$. On the other hand, if $y_{ij}^* > 0$, then using complementary slackness on the last constraint in problem (12), we have $\eta_{ij}^* + \sigma_{ij}^* = T\,\lambda\,\theta_i\,v_j\,(r_i - \sum_{q \in M} a_{qi}\,\mu_q^*)$. Therefore, for all $i, j \in S^*$, we have $\eta_{ij}^* + \sigma_{ij}^* \le T\,\lambda\,\theta_i\,v_j\,\left(r_i - \sum_{q \in M} a_{qi}\,\mu_q^*\right)^+$, where we let $(a)^+ = \max\{a, 0\}$.

For all $i \in S^*$, $x_i^* > 0$, so using complementary slackness on the second constraint in problem (12), this constraint holds as equality for all $i \in S^*$. Adding over all $i \in S^*$ yields

$$\begin{aligned}\sum_{i \in S^*} v_i\,\pi^* + \sum_{i \in S^*} \alpha_i^* &= T \sum_{i \in S^*} (v_i + \lambda\,\theta_i) \left(r_i - \sum_{q \in M} a_{qi}\,\mu_q^* \right) + \sum_{i \in S^*} \sum_{j \in N} \eta_{ij}^* + \sum_{i \in S^*} \sum_{j \in N} \sigma_{ji}^* \\ &\overset{(a)}{=} T \sum_{i \in S^*} (v_i + \lambda\,\theta_i) \left(r_i - \sum_{q \in M} a_{qi}\,\mu_q^* \right) + \sum_{i \in S^*} \sum_{j \in S^*} (\eta_{ij}^* + \sigma_{ij}^*) \\ &\overset{(b)}{\le} T \sum_{i \in S^*} (v_i + \lambda\,\theta_i) \left(r_i - \sum_{q \in M} a_{qi}\,\mu_q^* \right) + T \sum_{i \in S^*} \sum_{j \in S^*} \lambda\,\theta_i\,v_j \left(r_i - \sum_{q \in M} a_{qi}\,\mu_q^* \right)^+ \\ &\overset{(c)}{\le} T \sum_{i \in S^*} v_i \left(r_i - \sum_{q \in M} a_{qi}\,\mu_q^* \right)^+ + T \sum_{i \in S^*} \lambda\,\theta_i\,(1 + V(S^*)) \left(r_i - \sum_{q \in M} a_{qi}\,\mu_q^* \right)^+,\end{aligned}$$

where (a) follows from Lemma E.3, (b) holds since $\eta_{ij}^* + \sigma_{ij}^* \le T\,\lambda\,\theta_i\,v_j\,\left(r_i - \sum_{q \in M} a_{qi}\,\mu_q^*\right)^+$ as in the previous paragraph and (c) holds by arranging the terms and noting that $\sum_{j \in S^*} v_j = V(S^*)$.

The expression on the left side of the chain of inequalities above is given by $\sum_{i \in S^*} v_i\,\pi + \sum_{i \in S^*} \alpha_i^* = V(S^*)\,\pi^* + \sum_{i \in S^*} \alpha_i^* = (1 + V(S^*))\,\pi^* - \epsilon$, where the last equality follows from

Lemma E.3. In this case, replacing the left side of the chain of inequalities above by $(1+V(S^*))\,\pi^* - \epsilon$ and dividing both sides of the inequality by $1+V(S^*)$, we get

$$\pi^* - \frac{\epsilon}{1+V(S^*)} \leq\; T \sum_{i\in S^*} \Big(r_i - \sum_{q\in M} a_{qi}\,\mu_q^* \Big)^+ \Big(\frac{v_i}{1+V(S^*)} + \lambda\,\theta_i \Big). \tag{13}$$

To show the result by contradiction, assume that there exists $i,j\in N$ such that $y_{ij}^* < \min\{x_i^*, x_j^*\}$. Since $y_{ij}^* \geq 0$, it must be the case that $x_i^* > 0$ and $x_j^* > 0$, so we get $i,j\in S^*$.

Letting $i,j\in S^*$ such that $y_{ij}^* < \min\{x_i^*, x_j^*\}$, using complementary slackness on the last two constraints in problem (8), it follows that $\eta_{ij}^* = 0$ and $\sigma_{ij}^* = 0$, in which case, by the last constraint in problem (12), we have $0 \geq r_i - \sum_{q\in M} a_{qi}\,\mu_q^*$. Thus, there exists $i\in S^*$ such that $r_i - \sum_{q\in M} a_{qi}\,\mu_q^* \leq 0$. Let $N^* = \{i\in S^* : r_i - \sum_{q\in M} a_{qi}\,\mu_q^* \leq 0\}$, so N^* is non-empty.

By Lemma E.2, problems (8) and (11) have the same optimal objective values. Letting z_{LP}^* be their common optimal objective value, since problem (12) is the dual of (8), we get

$$\begin{aligned}
z_{\mathsf{LP}}^* \;&=\; \sum_{q\in M} c_q\,\mu_q^* + \pi^* \;\overset{(c)}{\leq}\; \sum_{q\in M} c_q\,\mu_q^* + T\sum_{i\in S^*} \Big(r_i - \sum_{q\in M} a_{qi}\,\mu_q^* \Big)^+ \Big(\frac{v_i}{1+V(S^*)} + \lambda\,\theta_i \Big) + \frac{\epsilon}{1+V(S^*)} \\
&\overset{(d)}{=}\; \sum_{q\in M} c_q\,\mu_q^* + T \sum_{i\in S^*\setminus N^*} \Big(r_i - \sum_{q\in M} a_{qi}\,\mu_q^* \Big) \Big(\frac{v_i}{1+V(S^*)} + \lambda\,\theta_i \Big) + \frac{\epsilon}{1+V(S^*)} \\
&\overset{(e)}{<}\; \sum_{q\in M} c_q\,\mu_q^* + T \sum_{i\in S^*\setminus N^*} \Big(r_i - \sum_{q\in M} a_{qi}\,\mu_q^* \Big) \Big(\frac{v_i}{1+V(S^*\setminus N^*)} + \lambda\,\theta_i \Big) + \frac{\epsilon}{1+V(S^*\setminus N^*)} \\
&\leq\; \sum_{q\in M} c_q\,\mu_q^* + \max_{S\subseteq N} \Big\{ T \sum_{i\in S} \Big(r_i - \sum_{q\in M} a_{qi}\,\mu_q^* \Big) \Big(\frac{v_i}{1+V(S)} + \lambda\,\theta_i \Big) + \frac{\epsilon}{1+V(S)} \Big\} \\
&\overset{(f)}{=}\; \sum_{q\in M} c_q\,\mu_q^* + \max_{\boldsymbol{w}\in\mathbb{R}_+^{2^n}} \Big\{ T \sum_{S\subseteq N}\sum_{i\in S} \Big(r_i - \sum_{q\in M} a_{qi}\,\mu_q^* \Big) \Big(\frac{v_i}{1+V(S)} + \lambda\,\theta_i \Big) w(s) \\
&\qquad\qquad + \sum_{S\subseteq N} \frac{\epsilon}{1+V(S)}\, w(s) \;:\; \sum_{S\subseteq N} w(S) = 1 \Big\},
\end{aligned}$$

where (c) uses (13), (d) holds since $r_i - \sum_{q\in M} a_{qi}\,\mu_q^* \leq 0$ for all $i\in N^*$, (e) holds since $N^* \neq \varnothing$, so $V(S^*\setminus N^*) < V(S^*)$ and (f) holds by randomizing instead of picking one assortment.

Consider computing the dual function for problem (11) by associating the dual multipliers $\boldsymbol{\mu}^* = \{\mu_q^* : q\in M\}$ with the first constraint in this problem. In this case, the value of the dual function precisely corresponds to the expression on the right side of the chain of inequalities above. Furthermore, by Lemma E.2, $\boldsymbol{\mu}^*$, which gives the optimal values of the dual variables associated with the first constraint in problem (8), also gives the optimal values of the dual variables associated with the first constraint in problem (11). Thus, the expression on the right side of the chain of inequalities above is the optimal objective value of problem (11), which is also z_{LP}^*. In this case, noting the strict inequality in (e), we get a contradiction.