

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

Yufeng Cao¹, Paat Rusmevichientong², Huseyin Topaloglu³

¹Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai 200030, China

²Marshall School of Business, University of Southern California, Los Angeles, CA 90089, USA

³School of Operations Research and Information Engineering, Cornell Tech, New York, NY 10044, USA
yufeng.cao@gatech.edu, rusmevic@marshall.usc.edu, topaloglu@orie.cornell.edu

June 2, 2022

We consider assortment optimization problems when customers choose under a mixture of independent demand and multinomial logit models. In the assortment optimization setting, each product has a fixed 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 assortment optimization problem with a capacity constraint on the offered assortment, we show that the problem is NP-hard even when each product consumes unit capacity so that we have a constraint on the number of offered products. 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. We give an expectation-maximization algorithm to estimate the parameters of our mixture model. Our computational experiments indicate that our mixture model can provide improvements in predicting customer purchases and identifying profitable assortments.

Keywords: Assortment optimization, choice model, multinomial logit.

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 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 independent demand and multinomial logit 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 independent demand and multinomial logit models. Some customers make a purchase under the independent demand model, whereas others do so under the multinomial logit model. The demand emerges as a mixture of the choices of the customers in these two segments.

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

Assortment Optimization. In the assortment optimization problem, we have a fixed 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, the objective function of the LP at its extreme points gives expected revenues of different assortments.

Combinatorial Algorithm. We show that if it is optimal to offer a given product, then it is optimal to offer all other products with smaller preference weights in the multinomial logit model or

larger purchase probabilities in the independent demand model, all else being equal (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. Both the combinatorial algorithm and the LP formulation become useful throughout the paper.

Comparative Statistics and Pareto-Efficiency. We show that the optimal assortment becomes larger when the relative size of the independent demand segment increases or when the revenue of each product increases by the same additive amount (Theorem 4.1). To see the implication of the first comparative statistic, the customers in the independent demand segment are inflexible. If the product they have in mind is unavailable, then they leave without a purchase. So, if the relative size of the inflexible customer segment increases, then it is optimal to offer a larger assortment. To see the implication of the second comparative statistic, the expected revenue is a firm-centric objective. For the two choice models that we mix, maximizing the total probability of purchase is equivalent to maximizing the expected utility of the customer (Sumida et al. 2019). Thus, the total probability of purchase is a customer-centric objective. Using the second comparative statistic, we show that the Pareto-efficient assortments that maximize a weighted sum of the expected revenue and the total probability of purchase are nested in the sense that the Pareto-efficient assortments get larger as the weight on the total probability of purchase increases.

Capacity Constraints. We study the assortment optimization problem with a capacity constraint, where each product occupies a fixed amount of capacity and there is a constraint on the total capacity consumption of the offered products. We show that the problem is NP-hard even when each product occupies unit capacity so that we have a constraint on the number of offered products (Theorem 5.1). Motivated by our complexity result, we give a fully polynomial-time approximation scheme (FPTAS) under a capacity constraint (Theorem 5.2). Our FPTAS uses the connection of our assortment optimization problem to a variant of the knapsack problem. We also build on the approach that we use to develop our FPTAS to give a heuristic for the assortment optimization problem under a capacity constraint. Thus, our FPTAS guides the design of a heuristic as well. Lastly, we give an efficiently computable upper bound on the optimal expected revenue under a capacity constraint. Our heuristic performs remarkably well. In a numerical study, we compare the performance of our heuristic with the upper bound on the optimal expected revenue. The average optimality gap of the heuristic comes out to be 0.15%.

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 an equivalent compact LP whose numbers of decision variables and constraints increase only quadratically with the number of products (Theorem 6.1). We can recover an optimal solution to one LP formulation by using an optimal solution to the other.

Fitting the Choice Model. We give an expectation-maximization algorithm to estimate the parameters of our mixture model and show that we can parameterize our mixture model as a function of product features, such as price, weight, brand and color. The number of product features is often smaller than the number of products in consideration, so parameterizations as a function of product features yield a more parsimonious representation. We fit our choice model to a real-world dataset to predict the choices of diners among sushi varieties. Our mixture model provides significant improvements over using the independent demand or multinomial logit model alone. Thus, there is practical benefit in mixing the two choice models. We also compare our mixture model with the exponential and Markov chain choice models. Our mixture model consistently performs better than the exponential model. The number of parameters in the Markov chain choice model is significantly more than that in our mixture model, so the Markov chain choice model can be prone to overfitting when we have limited data. Under limited data, our mixture model indeed provides improvements over the Markov chain choice model in predicting customer purchases.

Discussion of the Mixture Model: Our mixture model does not significantly complicate the parameter estimation and assortment optimization problems under the independent demand or multinomial logit model, while increasing the flexibility of these two choice models to capture the customer choice behavior. As we also observe in our numerical experiments, our mixture model can provide improvements over more sophisticated choice models, such as the Markov chain choice model, especially when we have limited data, so that more sophisticated choice models with larger numbers of parameters end up being prone to overfitting. We can efficiently estimate the parameters of our mixture model by iteratively maximizing concave likelihood functions in an expectation-maximization algorithm, we can parameterize our mixture model as a function of product features, and we can give efficient algorithms to solve the corresponding assortment optimization problems. Due to its parsimonious nature, our mixture model can avoid overfitting. All these characteristics enhance the practical viability of our mixture model.

In our expectation-maximization algorithm, we estimate the parameters of our mixture model by iteratively maximizing concave likelihood functions that one would maximize when estimating

the parameters of the independent demand and multinomial logit models. Thus, estimating the parameters of our mixture model brings minimal burden. We can parameterize our mixture model as a function of product features. Such a parameterization reduces the number of parameters from being proportional to the number of products in consideration to being proportional to the number of product features. If a new product with no past purchase history is introduced, then using product features allows us to estimate the choice probability of the new product, as long as we know the features of the new product. While the independent demand and multinomial logit models enjoy similar parameterizations, to our knowledge, there is no work on parameterizing the Markov chain choice model as a function of product features. Also, the assortment optimization problem under our mixture model is efficiently solvable. A natural approach to enhance the prediction ability of the multinomial logit model is to mix multinomial logit models, but if we mix even just two multinomial logit models, then the corresponding assortment problem is NP-hard. Thus, our mixture model uniquely possesses the practical benefits from a variety of choice models.

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 independent demand and multinomial logit 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.

Kunnumkal and Martinez-de-Albeniz (2019) study assortment optimization problems under the multinomial logit model when there is a cost for offering a product. The objective functions in their problem and our problem can both be expressed as the sum of a linear function and a fraction with linear numerator and denominator. The linear function takes negative values in their problem, capturing the cost of offered products, whereas the linear function takes positive values in ours, capturing the expected revenue from the independent demand segment. Interestingly, due to this difference, their problem is NP-hard, whereas ours is solvable in polynomial time.

Motivated by online retail where 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. 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 offered assortments are dictated by the inventory remaining on the shelf. We focus our literature review on the multinomial logit model. For work under other choice models, 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 capacity constraints. In Section 6, we give a compact LP for the network revenue management problem. In Section 7, we present computational experiments to test the prediction ability of our mixture model, as well as an expectation-maximization algorithm for parameter estimation. In Section 8, we give conclusions.

2. Problem Formulation

The set of products is $N = \{1, \dots, n\}$. There are two customer segments. The customers in the first segment make their purchases 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 . If we offer the subset S of products, then a customer in the first segment purchases product $i \in S$ with probability θ_i . It is possible to have $\sum_{i \in N} \theta_i < 1$, in which case, with probability $1 - \sum_{i \in N} \theta_i$, a customer in the first segment is not interested in any of the products, so we interpret this customer simply as a browser. The customers in the second segment make their purchases 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 . In this case, if we offer the subset S of products, then a customer in the second segment purchases product $i \in S$ with probability $v_i / (1 + V(S))$. The probability that an arriving customer is in the first segment is $\lambda \in (0, 1)$. With the remaining probability $1 - \lambda$, an arriving customer is in the second segment. Therefore, if we offer the subset S of products, then a customer purchases product $i \in S$ with probability $\lambda \theta_i + (1 - \lambda) \frac{v_i}{1 + V(S)}$. If a customer purchases product i , then we obtain a revenue of r_i . Our goal is to find a subset, or an assortment, of products to offer that maximizes the expected revenue from a customer, yielding the assortment optimization problem

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

Thus, the choice probabilities in the problem above are driven by a mixture of the independent demand and multinomial logit models, each with weights λ and $1 - \lambda$, respectively.

Working with such a mixture of independent demand and multinomial logit models 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. Also, 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. To demonstrate that revenue-ordered assortments are not necessarily optimal for the Mixture problem, consider a problem instance with $n = 3$, $\lambda = 1/2$, $(r_1, r_2, r_3) = (50, 10, 5)$, $(\theta_1, \theta_2, \theta_3) = (0.05, 0.25, 0.7)$ and $(v_1, v_2, v_3) = (0.5, 5, 0.01)$. In Table 1, we

Assort.	Exp. Rev.	Assort.	Exp. Rev.
\emptyset	0	$\{1, 2\}$	8.27
$\{1\}$	9.58	$\{1, 3\}$	11.29
$\{2\}$	5.42	$\{2, 3\}$	7.16
$\{3\}$	1.77	$\{1, 2, 3\}$	10.01

Table 1 Expected revenue provided by all possible assortments.

show the expected revenue provided by each assortment. The optimal 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 $\theta_3 = 0.7$, a customer in the independent demand segment is interested in product 3 with a relatively large probability, so we offer product 3 to exploit this relatively large probability. Also, noting that $v_2 = 5$, a customer in the multinomial logit segment 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 customers in the multinomial logit segment 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. Next, we discuss the relationship of our mixture model to random utility maximization and other choice models.

Connections to Random Utility Maximization and Other Choice Models. Under random utility maximization, a customer associates random utilities with the products and the no-purchase option, choosing the available alternative with the largest utility. Both the multinomial logit and independent demand models are compatible with random utility maximization. In the multinomial logit model, the utility of product i has the Gumbel distribution with location-scale parameters $(\mu_i, 1)$, whereas the utility of the no-purchase option has the Gumbel distribution with location-scale parameters $(0, 1)$. All utilities are independent. In this case, if we offer the subset S of products, then a customer purchases product i with probability $e^{\mu_i}/(1 + \sum_{j \in S} e^{\mu_j})$. In the independent demand model, let X be a random variable with support N , taking value i with probability θ_i . Using $\mathbf{1}(\cdot)$ to denote the indicator function, the utility of product i is $\mathbf{1}(X = i)$. The utility of the no-purchase option is $1/2$. Thus, with probability θ_i , a customer associates a utility of 1 with product i , a utility of $1/2$ with the no-purchase option and a utility of zero with all other products, in which case, she purchases product i if it is offered, but leaves without a purchase if not.

By the discussion above, our choice model is a mixture of choice models that are compatible with random utility maximization, so it is compatible with random utility maximization as well. Furthermore, we can express our choice model as a mixture of multinomial logit models. Consider

a mixture of multinomial logit models with $n + 1$ segments indexed by $N \cup \{0\}$. For each $i \in N$, a customer in segment i associates a preference weight of w_i with product i and a preference weight of zero with all other products. A customer in segment 0 associates the preference weights $\{v_j : j \in N\}$ with the products. An arriving customer is in segment i with probability $\lambda \alpha_i$, where we have $\sum_{i \in N} \alpha_i = 1$. An arriving customer is in segment 0 with probability $1 - \lambda$. In this case, if we offer the assortment S of products, then a customer chooses product i with probability $\lambda \alpha_i \frac{w_i}{1+w_i} + (1 - \lambda) \frac{v_i}{1+\sum_{j \in S} v_j}$. If we choose $\alpha_i = \frac{\theta_i}{\sum_{j \in N} \theta_j}$ and $w_i = \frac{1}{1-\sum_{j \in N} \theta_j} - 1$ for all $i \in N$, then the last choice probability becomes $\lambda \theta_i + (1 - \lambda) \frac{v_i}{1+\sum_{j \in S} v_j}$, which is the choice probability of product i under our choice model. Being able to express our choice model as a mixture of multinomial logit models is not surprising, as any choice model that is compatible with random utility maximization can be expressed as a mixture of multinomial logit models (McFadden and Train 2000). Here, we use $n + 1$ segments in the mixture of multinomial logit models and the assortment problem under a mixture of multinomial logit models is particularly difficult to approximate when there are too many segments (Desir et al. 2016). Thus, expressing our choice model as a mixture of multinomial logit models does not yield efficient algorithms. In Appendix A, we show that it is impossible to express our choice model as a mixture of multinomial logit models with two segments.

The multinomial logit model displays the independence of irrelevant alternatives property, which refers to the fact that if we add some product k to an assortment, then the purchase probabilities of products i and j already in the assortment decrease by the same relative amount. In other words, using $\phi_i^{\text{MNL}}(S)$ to denote the choice probability of product i out of assortment S under the multinomial logit model, we have $\frac{\phi_i^{\text{MNL}}(S \cup \{k\})}{\phi_i^{\text{MNL}}(S)} = \frac{\phi_j^{\text{MNL}}(S \cup \{k\})}{\phi_j^{\text{MNL}}(S)}$ for $k \notin S$ and $i, j \in S$. Thus, adding product k into an assortment cannibalizes on the demands of products i and j by the same relative amount, which should not hold when, for example, product k is similar to product i but not similar to product j . Our choice model does not display this property. For $\epsilon \in (0, 1)$, consider our choice model with $n = 3$, $\lambda = 1/2$, $(\theta_1, \theta_2, \theta_3) = (1 - \epsilon, \epsilon^2, \epsilon - \epsilon^2)$ and $(v_1, v_2, v_3) = (\epsilon, \epsilon, 1/\epsilon)$. Using $\phi_i^{\text{Mix}}(S)$ to denote the choice probability of product i out of assortment S under our choice model, we have $\frac{\phi_1^{\text{Mix}}(\{1,2,3\})}{\phi_1^{\text{Mix}}(\{1,2\})} = \frac{1-\epsilon+[\epsilon/(1+2\epsilon+\epsilon^{-1})]}{1-\epsilon+[\epsilon/(1+2\epsilon)]}$, which approaches one as $\epsilon \rightarrow 0$, whereas we have $\frac{\phi_2^{\text{Mix}}(\{1,2,3\})}{\phi_2^{\text{Mix}}(\{1,2\})} = \frac{\epsilon^2+[\epsilon/(1+2\epsilon+\epsilon^{-1})]}{\epsilon^2+[\epsilon/(1+2\epsilon)]}$, which approaches zero as $\epsilon \rightarrow 0$. Therefore, $\frac{\phi_1^{\text{Mix}}(\{1,2,3\})}{\phi_1^{\text{Mix}}(\{1,2\})}$ and $\frac{\phi_2^{\text{Mix}}(\{1,2,3\})}{\phi_2^{\text{Mix}}(\{1,2\})}$ may be dramatically different, indicating that adding product 3 to the assortment $\{1, 2\}$ may cannibalize on the demands of products 1 and 2 to drastically different extents. One manifestation of the independence of irrelevant alternatives property is the red bus-blue bus paradox (McFadden 1980). In Appendix B, we discuss this paradox within the context of our choice model.

As discussed in the introduction, the multinomial logit model is a special case of the Markov chain choice model. An example in Appendix C shows that our mixture of independent demand and multinomial logit models is not a special case of the Markov chain choice model.

3. Assortment Optimization

We show that we can get an optimal solution to the Mixture problem by solving an LP. Using the decision variables x_0 , $\mathbf{x} = \{x_i : i \in N\}$ and $\mathbf{y} = \{y_{ij} : i, j \in N\}$, we consider the LP

$$\begin{aligned} \max_{(x_0, \mathbf{x}, \mathbf{y}) \in \mathbb{R} \times \mathbb{R}_+^{n+n^2}} & \left\{ \sum_{i \in N} r_i \left(\left[\lambda \theta_i + (1 - \lambda) v_i \right] x_i + \lambda \theta_i \sum_{j \in N} v_j y_{ij} \right) \right\} : & \text{(Assortment LP)} \\ & x_0 + \sum_{i \in N} v_i x_i = 1 \\ & x_i \leq x_0 \quad \forall i \in N, \\ & \left. \begin{aligned} y_{ij} \leq x_i \quad \forall i, j \in N, \quad y_{ij} \leq x_j \quad \forall i, j \in N \end{aligned} \right\}. \end{aligned}$$

Before showing that we can obtain an optimal solution to the Mixture problem by using the Assortment LP, we provide some intuition regarding this LP.

Given a solution $\hat{S} \subseteq N$ to the Mixture problem, we construct a solution $(\hat{x}_0, \hat{\mathbf{x}}, \hat{\mathbf{y}})$ 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 $\hat{y}_{ij} = \mathbf{1}(i \in \hat{S}, j \in \hat{S}) \hat{x}_0$. Since $\sum_{i \in N} v_i \hat{x}_i = \hat{x}_0 \sum_{i \in N} v_i \mathbf{1}(i \in \hat{S}) = \hat{x}_0 V(\hat{S})$, 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 $(\hat{x}_0, \hat{\mathbf{x}}, \hat{\mathbf{y}})$ satisfies the first constraint in the Assortment LP. Noting that $\mathbf{1}(i \in \hat{S}) \leq 1$, $\mathbf{1}(i \in \hat{S}, j \in \hat{S}) \leq \mathbf{1}(i \in \hat{S})$ and $\mathbf{1}(i \in \hat{S}, j \in \hat{S}) \leq \mathbf{1}(j \in \hat{S})$, the solution $(\hat{x}_0, \hat{\mathbf{x}}, \hat{\mathbf{y}})$ also satisfies the remaining constraints in the Assortment LP. Since $\hat{x}_i = \mathbf{1}(i \in \hat{S}) \hat{x}_0$ and $\hat{y}_{ij} = \mathbf{1}(i \in \hat{S}, j \in \hat{S}) \hat{x}_0$, for the Assortment LP, the solution $(\hat{x}_0, \hat{\mathbf{x}}, \hat{\mathbf{y}})$ provides an objective value of

$$\begin{aligned} & \sum_{i \in N} r_i \left(\left[\lambda \theta_i + (1 - \lambda) v_i \right] \mathbf{1}(i \in \hat{S}) + \lambda \theta_i \sum_{j \in N} v_j \mathbf{1}(i \in \hat{S}, j \in \hat{S}) \right) \hat{x}_0 \\ &= \sum_{i \in N} r_i \left(\left[\lambda \theta_i + (1 - \lambda) v_i \right] + \lambda \theta_i \sum_{j \in N} v_j \mathbf{1}(j \in \hat{S}) \right) \mathbf{1}(i \in \hat{S}) \hat{x}_0 \\ &= \sum_{i \in N} r_i \left(\lambda \theta_i (1 + V(\hat{S})) + (1 - \lambda) v_i \right) \mathbf{1}(i \in \hat{S}) \hat{x}_0 = \sum_{i \in N} r_i \left(\lambda \theta_i + (1 - \lambda) \frac{v_i}{1 + V(\hat{S})} \right) \mathbf{1}(i \in \hat{S}), \quad (1) \end{aligned}$$

which is the objective function of the Mixture problem evaluated at \hat{S} . Thus, given a solution \hat{S} to the Mixture problem, we can construct a feasible solution $(\hat{x}_0, \hat{\mathbf{x}}, \hat{\mathbf{y}})$ to the Assortment LP and the objective values provided by the two solutions for their respective problems 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. In the chain of equalities above, observe that collecting the terms $\lambda \theta_i$ and $\lambda \theta_i \sum_{j \in N} v_j \mathbf{1}(j \in \hat{S})$ as $\lambda \theta_i (1 + V(\hat{S}))$ and noting that $\hat{x}_0 = \frac{1}{1+V(\hat{S})}$, we get the purchase probability of product i in the independent demand segment.

To establish the converse statement, we use the next lemma, which shows an important property of the basic feasible solutions to the Assortment LP. The proof is in Appendix D.

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

In the next theorem, we use Lemma 3.1 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 $(x_0^*, \mathbf{x}^*, \mathbf{y}^*)$ 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 z_{LP}^* be the optimal objective value of the Assortment LP. In (1), given the solution \hat{S} to the Mixture problem, we construct a feasible solution to the Assortment LP with the objective value of \hat{z} , so $z_{\text{LP}}^* \geq \hat{z}$. 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 $(x_0^*, \mathbf{x}^*, \mathbf{y}^*)$ 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 $i, j \in S^*$. Lastly, 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, whereas \hat{z} is the optimal objective value of the Mixture problem. Noting that $x_i^* = 0$ 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 $(x_0^*, \mathbf{x}^*, \mathbf{y}^*)$, we get

$$\begin{aligned}
z_{\text{LP}}^* &= \sum_{i \in S^*} r_i \left(\left[\lambda \theta_i + (1 - \lambda) v_i \right] x_i^* + \lambda \theta_i \sum_{j \in S^*} v_j y_{ij}^* \right) \\
&= \sum_{i \in S^* \setminus Q^*} r_i \left(\left[\lambda \theta_i + (1 - \lambda) v_i \right] x_i^* + \lambda \theta_i \sum_{j \in S^*} v_j y_{ij}^* \right) + \sum_{i \in Q^*} r_i \left(\left[\lambda \theta_i + (1 - \lambda) v_i \right] x_i^* + \lambda \theta_i \sum_{j \in S^*} v_j y_{ij}^* \right) \\
&\stackrel{(a)}{\leq} \sum_{i \in S^* \setminus Q^*} r_i \left(\frac{\lambda \theta_i + (1 - \lambda) v_i}{1 + V(S^*)} + \lambda \theta_i \sum_{j \in S^*} \frac{v_j}{1 + V(S^*)} \right) + \sum_{i \in Q^*} r_i \left(\left[\lambda \theta_i + (1 - \lambda) v_i \right] x_i^* + \lambda \theta_i \sum_{j \in S^*} v_j y_{ij}^* \right) \\
&= \sum_{i \in S^* \setminus Q^*} r_i \left(\lambda \theta_i + (1 - \lambda) \frac{v_i}{1 + V(S^*)} \right) + \sum_{i \in Q^*} r_i \left(\left[\lambda \theta_i + (1 - \lambda) v_i \right] x_i^* + \lambda \theta_i \sum_{j \in S^*} v_j y_{ij}^* \right) \\
&\leq \sum_{i \in S^* \setminus Q^*} r_i \left(\lambda \theta_i + (1 - \lambda) \frac{v_i}{1 + V(S^* \setminus Q^*)} \right) + \sum_{i \in Q^*} r_i \left(\left[\lambda \theta_i + (1 - \lambda) v_i \right] x_i^* + \lambda \theta_i \sum_{j \in S^*} v_j y_{ij}^* \right) \\
&\stackrel{(b)}{\leq} \hat{z} + \sum_{i \in Q^*} r_i \left(\left[\lambda \theta_i + (1 - \lambda) v_i \right] x_i^* + \lambda \theta_i \sum_{j \in S^*} v_j y_{ij}^* \right)
\end{aligned}$$

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.

For the moment, assume that $Q^* \neq \emptyset$. Using the definition of S^* and Q^* , we have $x_i^* > 0$ and $r_i^* < 0$ for all $i \in Q^*$. Thus, using the fact that $\lambda \theta_i + (1 - \lambda) v_i > 0$, along with $Q^* \neq \emptyset$, we obtain

$\sum_{i \in Q^*} r_i ((\lambda \theta_i + (1 - \lambda) v_i) x_i^* + \lambda \theta_i \sum_{j \in S^*} v_j y_{ij}^*) < 0$, in which case, the chain of inequalities above yields $z_{\text{LP}}^* < \hat{z}$, contradicting the fact that $z_{\text{LP}}^* \geq \hat{z}$, which we established at the beginning of the proof. Therefore, we must have $Q^* = \emptyset$. Since $Q^* = \emptyset$, all of the sums over the index set Q^* in the chain of inequalities are zero, so noting that $z_{\text{LP}} \geq \hat{z}$, all of the inequalities in the chain of inequalities above must hold as equalities. In particular, since (b) holds as an equality and $Q^* = \emptyset$, the objective value provided by the solution $S^* \setminus Q^* = S^*$ for the Mixture problem is \hat{z} , which implies that S^* is an optimal solution to the Mixture problem. ■

The proof above also shows that the Mixture problem and the Assortment LP have the same optimal objective values. In Theorem 3.2, the revenues of the products can be negative. We will use this theorem when studying network revenue management problems, where the revenues of the products will be adjusted by the opportunity costs of the capacities used by the products, so some products may have negative revenues. When we focus only on solving the Mixture problem, we can a priori drop from consideration all products with non-positive revenues, since the expected revenue from any assortment does not degrade when we drop such products.

Characterization of an Optimal Assortment:

We give a characterization of an optimal solution to the Mixture problem to intuitively suggest that there exists an optimal solution that prioritizes product i when θ_i/v_i is larger. This result also allows us to give a combinatorial algorithm for the Mixture problem. In the rest of this section, if there are multiple optimal solutions to the Mixture problem, then we choose any one that has the largest cardinality. Furthermore, for notational brevity, we let $R(S) = \frac{\sum_{i \in S} r_i v_i}{1 + V(S)}$, which is the expected revenue from the multinomial logit segment when we offer the assortment S . In the next theorem, we give a characterization of an optimal solution.

Theorem 3.3 (Characterization of an Optimal Assortment) *Letting S^* be an optimal solution to the Mixture problem with the largest cardinality, we have*

$$S^* = \left\{ i \in N : \lambda r_i \theta_i \geq (1 - \lambda) v_i \frac{R(S^*) - r_i}{1 + V(S^*)} \right\}.$$

Before we give a proof for the theorem above, we consider its implications. Rearranging the terms, we have $\lambda r_i \theta_i \geq (1 - \lambda) v_i \frac{R(S^*) - r_i}{1 + V(S^*)}$ if and only if $r_i \left[1 + \frac{\lambda \theta_i}{(1 - \lambda) v_i} (1 + V(S^*)) \right] \geq R(S^*)$, in which case, by the theorem above, we have $S^* = \{ i \in N : r_i \left[1 + \frac{\lambda \theta_i}{(1 - \lambda) v_i} (1 + V(S^*)) \right] \geq R(S^*) \}$. So, for two products i and j such that $r_i = r_j$ but $\theta_i/v_i \geq \theta_j/v_j$, if we have $j \in S^*$, then $i \in S^*$ as well. Thus, among products having the same revenue, an optimal solution prioritizes the product that has the larger independent demand probability to preference weight ratio. Similarly, for two products i and j such that $\theta_i/v_i = \theta_j/v_j$ but $r_i \geq r_j$, if we have $j \in S^*$, then we have $i \in S^*$ as well.

In other words, among products having the same independent demand probability to preference weight ratio, an optimal solution prioritizes the product that has the larger revenue.

We can also use Theorem 3.3 to construct a combinatorial algorithm for the Mixture problem. By the discussion in the previous paragraph, there exists an optimal solution S^* to the Mixture problem that satisfies $S^* = \{i \in N : r_i [1 + \frac{\lambda \theta_i}{(1-\lambda) v_i} (1 + V(S^*))] \geq R(S^*)\}$. If we knew the value of $V(S^*)$, then letting $t = V(S^*)$, we could index the products such that $r_1(1 + \frac{\lambda \theta_1}{(1-\lambda) v_1} (1 + t)) \geq r_2(1 + \frac{\lambda \theta_2}{(1-\lambda) v_2} (1 + t)) \geq \dots \geq r_n(1 + \frac{\lambda \theta_n}{(1-\lambda) v_n} (1 + t))$, in which case, an optimal assortment would be of the form $\{1, \dots, i\}$ 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 $\{1, \dots, i\}$ 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 $g_i(t) = r_i(1 + \frac{\lambda \theta_i}{(1-\lambda) v_i} (1 + t))$ is a linear function of t . The n lines $\{g_i(\cdot) : i \in N\}$ intersect at $O(n^2)$ points. Let $t^1 \leq t^2 \leq \dots \leq t^K$ with $K = O(n^2)$ be the intersection points of the n lines $\{g_i(\cdot) : i \in N\}$. That is, for each $k = 1, \dots, K$, 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 $k = 0, \dots, K$, 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, let the permutation $(\sigma_1^k, \dots, \sigma_n^k) \in N^n$ be such that $g_{\sigma_1^k}(t) \geq g_{\sigma_2^k}(t) \geq \dots \geq g_{\sigma_n^k}(t)$ for all $t \in [t^k, t^{k+1})$. In this case, if we know that $V(S^*) \in [t^k, t^{k+1})$, then an optimal assortment is of the form $\{\sigma_1^k, \dots, \sigma_i^k\}$ for some $i \in N$. Thus, we can obtain the optimal assortment by checking the expected revenue from $O(n^3)$ candidate assortments, each of which is of the form $\{\sigma_1^k, \dots, \sigma_i^k\}$ for some $i \in N, k = 0, \dots, K$. Below is the proof of Theorem 3.3. The auxiliary lemma that we give for the proof also becomes useful later in the paper.

Proof of Theorem 3.3. In the next lemma and throughout the rest of the paper, we let $A(S) = \frac{\sum_{i \in S} r_i v_i}{V(S)}$, which we view as a weighted average of the revenues of the products in S .

Lemma 3.4 (Comparison with Independent Demand Revenue) *Letting S^* be an optimal solution to the Mixture problem, for any $K, L \subseteq N$ with $K \cap S^* = \emptyset$ and $L \subseteq S^*$, we have*

$$\lambda \sum_{i \in K} \theta_i r_i \leq (1 - \lambda) V(K) \frac{R(S^*) - A(K)}{1 + V(S^* \cup K)} \quad \text{and} \quad \lambda \sum_{i \in L} \theta_i r_i \geq (1 - \lambda) V(L) \frac{R(S^* \setminus L) - A(L)}{1 + V(S^*)}.$$

Proof. For any $S, K \subseteq N$ with $K \cap S = \emptyset$, we can express $R(S \cup K)$ as a convex combination of $R(S)$ and $A(K)$. In particular, noting the definitions of $R(S)$ and $A(S)$, we have

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

Since $K \cap S = \emptyset$, we have $\frac{1 + V(S)}{1 + V(S \cup K)} + \frac{V(K)}{1 + V(S \cup K)} = 1$, in which case, the chain of equalities above implies that $R(S \cup K)$ is a convex combination of $R(S)$ and $A(K)$. We write the objective function

of the Mixture problem as $\lambda \sum_{i \in S} r_i \theta_i + (1 - \lambda) R(S)$. Since S^* is an optimal, but $S^* \cup K$ is only a feasible, solution to the Mixture problem, we obtain

$$\begin{aligned} 0 &\geq \left\{ \lambda \sum_{i \in S^* \cup K} r_i \theta_i + (1 - \lambda) R(S^* \cup K) \right\} - \left\{ \lambda \sum_{i \in S^*} r_i \theta_i + (1 - \lambda) R(S^*) \right\} \\ &\stackrel{(a)}{=} \lambda \sum_{i \in K} r_i \theta_i + (1 - \lambda) \left\{ \frac{1 + V(S^*)}{1 + V(S^* \cup K)} R(S^*) + \frac{V(K)}{1 + V(S^* \cup K)} A(K) - R(S^*) \right\} \\ &\stackrel{(b)}{=} \lambda \sum_{i \in K} r_i \theta_i - (1 - \lambda) V(K) \frac{R(S^*) - A(K)}{1 + V(S^* \cup K)}, \end{aligned}$$

where (a) follows by noting that $\sum_{i \in S^* \cup K} r_i \theta_i - \sum_{i \in S^*} r_i \theta_i = \sum_{i \in K} r_i \theta_i$ and expressing $R(S^* \cup K)$ as a convex combination of $R(S^*)$ and $A(K)$, whereas (b) follows by rearranging the terms.

The chain of inequalities above shows that the first inequality in the lemma holds. The second inequality follows similarly by expressing $R(S^*)$ as a convex combination of $R(S^* \setminus L)$ and $A(L)$. ■

Here is the proof of Theorem 3.3. Let $H = \{i \in N : \lambda r_i \theta_i \geq (1 - \lambda) v_i \frac{R(S^*) - r_i}{1 + V(S^*)}\}$. First, we show that if $i \in S^*$, then $i \in H$. Consider some $i \in S^*$. Since $r_i \theta_i \geq 0$, by the definition of H , if $r_i \geq R(S^*)$, then we have $i \in H$, as desired. Thus, we focus on the case with $r_i \leq R(S^*)$. By the argument at the beginning of the proof of Lemma 3.4, we can express $R(S^*)$ as a convex combination of r_i and $R(S^* \setminus \{i\})$, in which case, since $r_i \leq R(S^*)$, we must have $r_i \leq R(S^*) \leq R(S^* \setminus \{i\})$. Also, since $i \in S^*$, we have $\{i\} \subseteq S^*$. Thus, using the second inequality in Lemma 3.4 with $L = \{i\}$, we get $\lambda \theta_i r_i \geq (1 - \lambda) v_i \frac{R(S^* \setminus \{i\}) - r_i}{1 + V(S^*)} \geq (1 - \lambda) v_i \frac{R(S^*) - r_i}{1 + V(S^*)}$, where the last equality holds since we have $R(S^*) \leq R(S^* \setminus \{i\})$. The last chain of inequalities implies that $i \in H$.

Second, we show that if $i \notin S^*$, then $i \notin H$. Consider some $i \notin S^*$. Since S^* is an optimal solution to the Mixture problem with the largest cardinality and $i \notin S^*$, we have $\lambda \sum_{j \in S^*} r_j \theta_j + (1 - \lambda) R(S^*) > \lambda \sum_{j \in S^* \cup \{i\}} r_j \theta_j + (1 - \lambda) R(S^* \cup \{i\})$, which is equivalent to $R(S^*) > \frac{\lambda}{1 - \lambda} r_i \theta_i + R(S^* \cup \{i\})$. Thus, we have $R(S^*) > R(S^* \cup \{i\})$. By the argument at the beginning of the proof of Lemma 3.4, we can express $R(S^* \cup \{i\})$ as a convex combination of $R(S^*)$ and r_i , so noting that $R(S^* \cup \{i\}) < R(S^*)$, we must have $r_i \leq R(S^* \cup \{i\}) < R(S^*)$. Also, since $i \notin S^*$, using the first inequality in Lemma 3.4 with $K = \{i\}$, we get $\lambda \theta_i r_i \leq (1 - \lambda) v_i \frac{R(S^*) - r_i}{1 + V(S^* \cup \{i\})} < (1 - \lambda) v_i \frac{R(S^*) - r_i}{1 + V(S^*)}$, where the last inequality holds since $R(S^*) > r_i$. The last chain of inequalities implies that $i \notin H$. ■

Our combinatorial algorithm allows us to solve the Mixture problem without using an LP, but the Assortment LP is useful when we work with network revenue management problems.

4. Effect of Customer Segment Mix and Efficient Assortments

In this section, we give sensitivity results for the optimal solution for the Mixture problem. First, we show that the optimal solution to the Mixture problem becomes a larger assortment in a nested

fashion when the the likelihood of observing a customer in the independent demand segment grows. In particular, for $\alpha > \beta$, if we solve the Mixture problem with $\lambda = \alpha$ and $\lambda = \beta$, then the optimal solution with $\lambda = \alpha$ includes all the products in the optimal solution with $\lambda = \beta$. Second, we show that the optimal solution to the Mixture problem becomes a larger assortment in a nested fashion when the revenue of each product increases by the same additive amount. The latter result has important implications when we want to find assortments that trade off expected revenue with the probability of purchase, as well as when we implement 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

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

As a function of (λ, δ) , 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 $\alpha > \beta$, then $S^*(\alpha, 0) \supseteq S^*(\beta, 0)$.*
- (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 the customers in the independent demand segment are not willing to make substitutions. In particular, 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 independent demand segment 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 $\text{Rev}(S) = \sum_{i \in S} r_i (\lambda \theta_i + (1 - \lambda) \frac{v_i}{1 + V(S)})$ and $\text{Pur}(S) = \sum_{i \in S} (\lambda \theta_i + (1 - \lambda) \frac{v_i}{1 + V(S)})$, the objective function of the Parametric Mixture problem is $\text{Rev}(S) + \delta \text{Pur}(S)$. Thus, we maximize a linear combination of $\text{Rev}(S)$ and $\text{Pur}(S)$ in the Parametric Mixture problem. Observe that $\text{Rev}(S)$ is the expected revenue from assortment S , whereas $\text{Pur}(S)$ is the total purchase probability from assortment S . Maximizing $\text{Rev}(S)$ ensures that the expected revenue that the firm obtains is largest, whereas maximizing $\text{Pur}(S)$ ensures that the total probability that a customer purchases a product is largest. The parameter δ characterizes

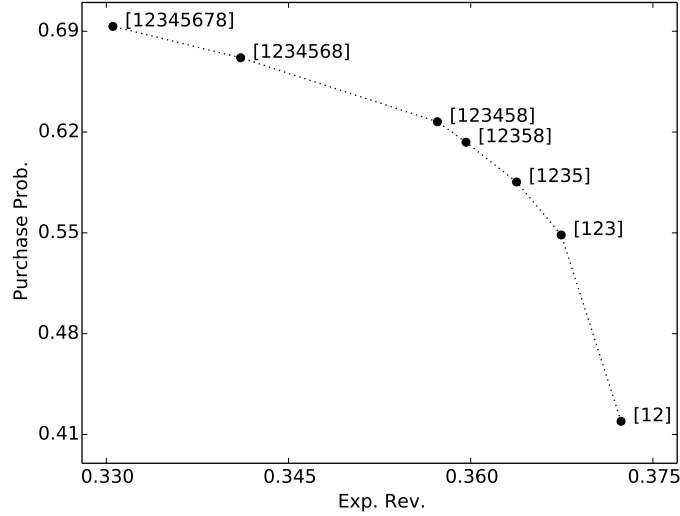


Figure 1 Efficient frontier with $n = 8$, $(r_1, \dots, r_8) = (0.96, 0.81, 0.34, 0.29, 0.19, 0.09, 0.04, 0.03)$, $(\theta_1, \dots, \theta_8) = (0.15, 0.20, 0.09, 0, 0.16, 0.15, 0.07, 0.18)$, $(v_1, \dots, v_8) = (0.40, 0.35, 0.54, 0.10, 0.11, 0.21, 0.12, 0.04)$, $\lambda = \frac{3}{25}$.

the weight that we put on the total purchase probability relative to the expected revenue. Solving the problem $\max_{S \subseteq N} \{\text{Rev}(S) + \delta \text{Pur}(S)\}$ 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 on the efficient frontier are nested, one assortment always being included in another one. Since there can be at most n such nested assortments, there can be at most n assortments on the efficient frontier. Also, 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 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. If the assortments on the efficient frontier are nested, then we 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.

Independent demand segment corresponds to the inflexible customers, so a possible conjecture is that the optimal expected revenue in the Parametric Mixture problem decreases as the relative size of the independent demand segment, measured by λ , increases. We can come up with counterexamples to demonstrate that the optimal expected revenue is neither increasing nor decreasing in λ . In particular, consider a problem instance with $n = 3$, $(r_1, r_2, r_3) = (15, 10, 6)$, $(\theta_1, \theta_2, \theta_3) = (0.3, 0.3, 0.4)$ and $(v_1, v_2, v_3) = (1, 2, 10)$. Letting z_λ^* be the optimal objective value of the Parametric Mixture problem as a function of λ , we have $z_{0.1}^* = 8.63$, $z_{0.4}^* = 8.25$ and $z_{0.7}^* = 8.97$. As λ increases from 0.1 to 0.4, the optimal expected revenue decreases, but as λ increases from 0.4 to 0.7, the optimal expected revenue increases. In the rest of this section, we give a proof for Theorem 4.1. Lemma 3.4 used in the proof of Theorem 3.3 also plays an important role in the proof of Theorem 4.1.

Proof of Theorem 4.1. To show the first part, let $S_\alpha = S^*(\alpha, 0)$ and $S_\beta = S^*(\beta, 0)$. We need to show that $S_\alpha \supseteq S_\beta$. Letting $\Theta(S) = \sum_{i \in S} r_i \theta_i$ for notational brevity, the objective function of the Parametric Mixture problem with $\delta = 0$ is $\lambda \Theta(S) + (1 - \lambda) R(S)$. We claim that $R(S_\beta) \geq R(S_\alpha)$. Since S_α is an optimal solution for the Parametric Mixture problem with $\lambda = \alpha$ and $\delta = 0$, we have $\alpha \Theta(S_\alpha) + (1 - \alpha) R(S_\alpha) \geq \alpha \Theta(S_\beta) + (1 - \alpha) R(S_\beta)$. By the same argument, we have $\beta \Theta(S_\beta) + (1 - \beta) R(S_\beta) \geq \beta \Theta(S_\alpha) + (1 - \beta) R(S_\alpha)$. We multiply the first inequality with β and the second one with α , so adding them yields $[\alpha(1 - \beta) - \beta(1 - \alpha)] R(S_\beta) \geq [\alpha(1 - \beta) - \beta(1 - \alpha)] R(S_\alpha)$. Since $\alpha > \beta$, we have $\alpha(1 - \beta) > \beta(1 - \alpha)$, so the last inequality yields $R(S_\beta) \geq R(S_\alpha)$ and the claim holds. Let $K = S_\beta \setminus S_\alpha$. If $K = \emptyset$, then $S_\alpha \supseteq S_\beta$, which is the desired result. To get a contradiction, assume that $K \neq \emptyset$. Next, we claim that $R(S_\beta \setminus K) - A(K) \geq R(S_\alpha) - A(K) > 0$.

Since S_α is an optimal solution with the largest cardinality for the Parametric Mixture problem with $\lambda = \alpha$ and $\delta = 0$, we have $\alpha \Theta(S_\alpha) + (1 - \alpha) R(S_\alpha) > \alpha \Theta(S_\alpha \cup K) + (1 - \alpha) R(S_\alpha \cup K)$, which yields $(1 - \alpha)(R(S_\alpha) - R(S_\alpha \cup K)) > \alpha \sum_{i \in K} r_i \theta_i > 0$, so $R(S_\alpha) > R(S_\alpha \cup K)$. By the argument at the beginning of the proof of Lemma 3.4, $R(S_\alpha \cup K)$ is a convex combination of $R(S_\alpha)$ and $A(K)$, so noting that $R(S_\alpha \cup K) < R(S_\alpha)$, we must have $A(K) \leq R(S_\alpha \cup K) < R(S_\alpha)$. In this case, since $R(S_\beta) \geq R(S_\alpha)$ by the claim in the previous paragraph, the last chain of inequalities also yields $A(K) \leq R(S_\beta)$. Similarly, $R(S_\beta)$ is a convex combination of $R(S_\beta \setminus K)$ and $A(K)$, so noting the fact that $A(K) \leq R(S_\beta)$, we must have $A(K) \leq R(S_\beta) \leq R(S_\beta \setminus K)$. Collecting the discussion so far in this paragraph together, we have $A(K) \leq R(S_\alpha \cup K) < R(S_\alpha) \leq R(S_\beta) \leq R(S_\beta \setminus K)$, which implies that $R(S_\beta \setminus K) - A(K) \geq R(S_\alpha) - A(K) > 0$ and the claim holds.

Since $K = S_\beta \setminus S_\alpha$, we $S_\alpha \cup K \supseteq S_\beta$, which implies that $1 + V(S_\alpha \cup K) \geq 1 + V(S_\beta) > 0$. Thus, noting that $R(S_\beta \setminus K) - A(K) \geq R(S_\alpha) - A(K) > 0$, we get $\frac{R(S_\alpha) - A(K)}{1 + V(S_\alpha \cup K)} \leq \frac{R(S_\beta \setminus K) - A(K)}{1 + V(S_\beta)}$. On the

other hand, since $K = S_\beta \setminus S_\alpha$, we get $K \cap S_\alpha = \emptyset$ and $K \subseteq S_\beta$, so the first inequality in Lemma 3.4 with $\lambda = \alpha$ and $S^* = S_\alpha$ and the second inequality in Lemma 3.4 with $\lambda = \beta$ and $S^* = S_\beta$ yield

$$\frac{\alpha \Theta(K)}{(1-\alpha)V(K)} \leq \frac{R(S_\alpha) - A(K)}{1 + V(S_\alpha \cup K)} \quad \text{and} \quad \frac{\beta \Theta(K)}{(1-\beta)V(K)} \geq \frac{R(S_\beta \setminus K) - A(K)}{1 + V(S_\beta)}.$$

Since $\alpha > \beta$, we have $\frac{\alpha}{1-\alpha} > \frac{\beta}{1-\beta}$, so by the two inequalities above, we get $\frac{R(S_\alpha) - A(K)}{1 + V(S_\alpha \cup K)} > \frac{R(S_\beta \setminus K) - A(K)}{1 + V(S_\beta)}$, contradicting the inequality $\frac{R(S_\alpha) - A(K)}{1 + V(S_\alpha \cup K)} \leq \frac{R(S_\beta \setminus K) - A(K)}{1 + V(S_\beta)}$ at the beginning of this paragraph.

To show the second part, we follow a similar outline that also builds on Lemma 3.4. We defer the details of the proof of the second part to Appendix E. ■

5. Assortment Optimization under a Capacity Constraint

We consider a capacitated version of the Mixture problem, where each product consumes a fixed amount of capacity and there is a limit on the total capacity consumption of the offered products. We show that the capacitated version of the Mixture problem is NP-hard even when the capacity consumption of each product is one so that we limit the number of products in the offered assortment. Following this complexity result, we give an FPTAS for the capacitated version of the Mixture problem. We use c_i to denote the capacity consumption of product i . If we offer the assortment S , then the total capacity consumption of the offered products is $\sum_{i \in S} c_i$. Letting C be the limit on the total capacity consumption of offered products, we consider the problem

$$\max_{\substack{S \subseteq N: \\ \sum_{i \in S} c_i \leq C}} \left\{ \sum_{i \in S} r_i \left(\lambda \theta_i + (1-\lambda) \frac{v_i}{1 + V(S)} \right) \right\}. \quad (\text{Capacitated Mixture})$$

If we have $c_i = 1$ for all $i \in N$ and C is an integer, then the Capacitated Mixture problem finds an assortment that maximizes the expected revenue while offering at most C products.

To study the complexity of the problem above, we use the following feasibility version of the Capacitated Mixture problem, which we refer to as the MIXTURE FEASIBILITY problem.

Mixture Feasibility: Given an instance of the Capacitated Mixture problem and a threshold K , is there an assortment $S \subseteq N$ with $\sum_{i \in S} c_i \leq C$ that provides an expected revenue of K or more?

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

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

In the next theorem, we show that the Capacitated Mixture problem is NP-hard even when $c_i = 1$ for all $i \in N$ so that we limit the number products in the offered assortment.

Theorem 5.1 (Complexity under a Cardinality Constraint) *The MIXTURE FEASIBILITY problem is NP-complete even when we have $c_i = 1$ for all $i \in N$.*

Proof. Given an instance of PARTITION with the set of items $N = \{1, \dots, n\}$ and weights $\{w_i : i \in N\}$, we define an instance MIXTURE FEASIBILITY as follows. The set of products is N . The revenue and preference weight of product i are $r_i = 1$ and $v_i = w_i$. Letting $V_{\max} = \max_{i \in N} v_i$, the purchase probability of product i is $\theta_i = V_{\max} - v_i$. The relative size of the independent demand segment is $\lambda = \frac{1}{5}$. The capacity consumption of product i is $c_i = 1$. The limit on the capacity consumption is $C = \frac{n}{2}$, so we offer at most $n/2$ products. Lastly, the threshold is $K = \frac{1}{5} + \frac{1}{5} C V_{\max}$. We show that there exists an assortment $S \subseteq N$ with $|S| \leq C$ 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 $\sum_{i \in S} w_i = 1$.

Noting that $r_i = 1$ for all $i \in N$ and $\lambda = \frac{1}{5}$, the objective function of the Capacitated Mixture problem is given by $\frac{1}{5} \sum_{i \in S} \theta_i + \frac{4}{5} \frac{V(S)}{1+V(S)}$. Therefore, since $\frac{x}{1+x}$ is increasing in x , there exists an assortment $S \subseteq N$ with $|S| \leq C$ that provides an expected revenue of K or more if and only if there exists an assortment $S \subseteq N$ with $|S| = C$ that provides an expected revenue of K or more. Since $\theta_i = V_{\max} - v_i$ and we need to offer an assortment $S \subseteq N$ with $|S| = C$, the objective function of the Capacitated Mixture problem becomes $\frac{1}{5} \sum_{i \in S} (V_{\max} - v_i) + \frac{4}{5} \frac{V(S)}{1+V(S)} = \frac{1}{5} C V_{\max} - \frac{1}{5} V(S) + \frac{4}{5} \frac{V(S)}{1+V(S)}$. Thus, noting that $K = \frac{1}{5} + \frac{1}{5} C V_{\max}$, 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

$$\frac{1}{5} C V_{\max} - \frac{1}{5} V(S) + \frac{4}{5} \frac{V(S)}{1+V(S)} \geq \frac{1}{5} + \frac{1}{5} C V_{\max}.$$

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 paragraph, there exists an assortment $S \subseteq N$ with $|S| \leq C$ 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 $C = n/2$ and $V(S) = \sum_{i \in S} v_i = \sum_{i \in S} w_i$, it follows that there exists an assortment $S \subseteq N$ with $|S| \leq C$ 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 $\sum_{i \in S} w_i = 1$ ■

Motivated by this complexity result, we give an FPTAS for the Capacitated Mixture problem. Our FPTAS continues to work when we do not necessarily have $c_i = 1$ for all $i \in N$.

Fully Polynomial-Time Approximation Scheme:

We borrow from ideas for developing an FPTAS for the knapsack problem (Desir et al. 2016). In the rest of this section, we outline the steps of our FPTAS and defer the details to Appendix F. For

fixed $(p, q, s) \in \mathbb{R}_+^3$, we consider finding an assortment S with the smallest capacity consumption that satisfies $\sum_{i \in S} r_i \theta_i \geq p$, $\sum_{i \in S} r_i v_i \geq q$ and $\sum_{i \in S} v_i \leq s$, yielding the problem

$$G(p, q, s) = \min_{S \subseteq N} \left\{ \sum_{i \in S} c_i : \sum_{i \in S} r_i \theta_i \geq p, \sum_{i \in S} r_i v_i \geq q, \sum_{i \in S} v_i \leq s \right\}. \quad (2)$$

If the assortment S is feasible to the problem above, then we have $\sum_{i \in S} r_i \theta_i \geq p$ and $\frac{\sum_{i \in S} r_i v_i}{1 + \sum_{i \in S} v_i} \geq \frac{q}{1+s}$, so the assortment S yields an expected revenue of at least p from the independent demand segment and an expected revenue of at least $\frac{q}{1+s}$ from the multinomial logit one. Letting $\Theta_{\min} = \min_{i \in N} \theta_i$, $R_{\min} = \min_{i \in N} r_i$, $R_{\max} = \max_{i \in N} r_i$, $V_{\min} = \min_{i \in N} v_i$ and $V_{\max} = \max_{i \in N} v_i$, for any non-empty assortment S , we have $\sum_{i \in S} r_i \theta_i \in [R_{\min} \Theta_{\min}, R_{\max}]$, $\sum_{i \in S} r_i v_i \in [R_{\min} V_{\min}, n R_{\max} V_{\max}]$ and $\sum_{i \in S} v_i \in [V_{\min}, n V_{\max}]$. In this case, using $a \wedge b = \min\{a, b\}$ and $a \vee b = \max\{a, b\}$, letting $\underline{p} = (R_{\min} \wedge 1)(V_{\min} \wedge \Theta_{\min})$ and $\bar{p} = (R_{\max} \vee 1)(n V_{\max} \vee 1)$ for notational brevity, we have $\sum_{i \in S} r_i \theta_i \in [\underline{p}, \bar{p}]$, $\sum_{i \in S} r_i v_i \in [\underline{p}, \bar{p}]$ and $\sum_{i \in S} v_i \in [\underline{p}, \bar{p}]$. Therefore, it is enough to consider the values of $(p, q, s) \in [\underline{p}, \bar{p}]^3$ in problem (2). We follow three steps.

In the first step, we relate problem (2) to the **Capacitated Mixture** problem. In particular, noting that $G(p, q, s)$ is the optimal objective value of problem (2), if $G(p, q, s) \leq C$, then there exists an assortment S that satisfies $\sum_{i \in S} c_i \leq C$ and this assortment provides expected revenues of at least p and $\frac{q}{1+s}$ from the two segments, yielding an objective value of at least $\lambda p + (1 - \lambda) \frac{q}{1+s}$ for the **Capacitated Mixture** problem. In this case, letting (p^*, q^*, s^*) be an optimal solution to the problem $\max_{(p, q, s) \in [\underline{p}, \bar{p}]^3} \left\{ \lambda p + (1 - \lambda) \frac{q}{1+s} : G(p, q, s) \leq C \right\}$, we can show that $\lambda p^* + (1 - \lambda) \frac{q^*}{1+s^*}$ is the optimal objective value of the **Capacitated Mixture** problem. Also, if S^* is an optimal solution to problem (2) with $(p, q, s) = (p^*, q^*, s^*)$, then S^* is an optimal solution to the **Capacitated Mixture** problem. However, computing (p^*, q^*, s^*) is difficult since $G(p, q, s)$ is not convex in (p, q, s) .

In the second step, we build a geometric grid over the interval $[\underline{p}, \bar{p}]$ and consider the values of (p, q, s) in the geometric grid. Using $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ to, respectively, denote the round down and up functions, for a fixed accuracy parameter $\alpha > 0$, we consider grid points that are integer powers of $1 + \alpha$, which are given by $\text{Grid} = \left\{ (1 + \alpha)^k : k = \lfloor \frac{\log \underline{p}}{\log(1 + \alpha)} \rfloor, \dots, \lceil \frac{\log \bar{p}}{\log(1 + \alpha)} \rceil \right\}$. Letting $(\hat{p}, \hat{q}, \hat{s}) \in \text{Grid}^3$ be such that $\hat{p} \leq p^* \leq (1 + \alpha) \hat{p}$, $\hat{q} \leq q^* \leq (1 + \alpha) \hat{q}$ and $\frac{1}{1 + \alpha} \hat{s} \leq s^* \leq \hat{s}$, we can show that if \hat{S} is an optimal solution to problem (2) with $(p, q, s) = (\hat{p}, \hat{q}, \hat{s})$, then the expected revenue from the assortment \hat{S} deviates from the optimal objective value of the **Capacitated Mixture** problem by at most a factor of $1 - 3\alpha$ and we have $\sum_{i \in \hat{S}} c_i \leq C$. However, solving problem (2) even at fixed (p, q, s) is difficult since this problem is at least as difficult as the knapsack problem, which is NP-hard.

In the third step, we use a dynamic program to solve an approximate version of problem (2) with $(p, q, s) = (\hat{p}, \hat{q}, \hat{s})$. Expressing the first constraint in problem (2) as $\sum_{i \in S} \frac{n}{\alpha p} r_i \theta_i \geq \frac{n}{\alpha}$, in the

approximate version of problem (2), we replace the first constraint with $\sum_{i \in S} \lceil \frac{n}{\alpha p} r_i \theta_i \rceil \geq \lfloor \frac{n}{\alpha} \rfloor$. Similarly, we replace the second constraint with $\sum_{i \in S} \lceil \frac{n}{\alpha q} r_i v_i \rceil \geq \lfloor \frac{n}{\alpha} \rfloor$. Lastly, expressing the third constraint in problem (2) as $\sum_{i \in S} \frac{n}{\alpha s} v_i \leq \frac{n}{\alpha}$, in the approximate version of problem (2), we replace the third constraint with $\sum_{i \in S} \lfloor \frac{n}{\alpha s} v_i \rfloor \leq \lceil \frac{n}{\alpha} \rceil$. In this case, the constraint coefficients and right side values of all constraints in the approximate version of problem (2) are integers, which allows us to solve the approximate version by using a dynamic program. Letting $(\hat{p}, \hat{q}, \hat{s}) \in \text{Grid}^3$ be as in the previous paragraph satisfying $\hat{p} \leq p^* \leq (1 + \alpha)\hat{p}$, $\hat{q} \leq q^* \leq (1 + \alpha)\hat{q}$ and $\frac{1}{1+\alpha}\hat{s} \leq s^* \leq \hat{s}$, we can show that if \hat{S} is an optimal solution to the approximate version of problem (2) with $(p, q, s) = (\hat{p}, \hat{q}, \hat{s})$, then the expected revenue from the assortment \hat{S} deviates from the optimal objective value of the Capacitated Mixture problem by at most a factor of $1 - 6\alpha$.

By the definitions of $\underline{\nu}$ and $\bar{\nu}$, there are $O\left(\frac{1}{\log(1+\alpha)} \log\left(\frac{(R_{\max} \vee 1)(nV_{\max} \vee 1)}{(R_{\min} \wedge 1)(V_{\min} \wedge \Theta_{\min})}\right)\right)$ points in Grid. In the next theorem, counting the number of operations in the steps above, we give our FPTAS.

Theorem 5.2 (FPTAS under a Capacity Constraint) *Letting z^* be the optimal objective value of the Capacitated Mixture problem, there exists an algorithm, where for any $\epsilon \in (0, 1)$, the algorithm runs in $O\left(\frac{n^4}{\epsilon^6} \log^3\left(\frac{(R_{\max} \vee 1)(nV_{\max} \vee 1)}{(R_{\min} \wedge 1)(V_{\min} \wedge \Theta_{\min})}\right)\right)$ operations and returns an assortment that is feasible to the Capacitated Mixture problem with an expected revenue of at least $(1 - \epsilon)z^*$.*

The proof of Theorem 5.2 is in Appendix F. Another useful feature of our FPTAS is that we can build on our FPTAS to develop effective heuristics and bounds, as discussed next.

Practical Heuristics and Bounds on the Optimal Expected Revenue:

Using the decision variables $\mathbf{x} = \{x_i : i \in N\} \in \{0, 1\}^n$, where $x_i = 1$ if and only if we offer product i , we consider a relaxation of problem (2) given by the LP

$$\tilde{G}(p, q, s) = \min_{\mathbf{x} \in [0, 1]^n} \left\{ \sum_{i \in N} c_i x_i : \sum_{i \in N} r_i \theta_i x_i \geq p, \quad \sum_{i \in N} r_i v_i x_i \geq q, \quad \sum_{i \in N} v_i x_i \leq s \right\}. \quad (3)$$

If we impose the constraint $\mathbf{x} \in \{0, 1\}^n$ on the decision variables, then the problem above is equivalent to problem (2). Since we impose the constraint $\mathbf{x} \in [0, 1]^n$, we view the problem above as an LP relaxation of (2). Other than the upper bound of one on the decision variables, there are three constraints in problem (3). Thus, it is a simple LP exercise to show that there exist at most three fractional decision variables in a basic optimal solution to problem (3). Thus, after solving the LP in (3), we can construct eight possible binary solutions by rounding the fractional decision variables. In particular, we choose a subset of the three fractional decision variables, round them down to zero and round the rest up to one. Naturally, if there are fewer than three fractional decision variables, then we can generate fewer than eight binary solutions. Motivated by this observation, we use the following approach to construct a practical heuristic for the Capacitated

Mixture problem. For each $(p, q, s) \in \text{Grid}^3$, we solve the LP in (3). Letting $\mathbf{x}^*(p, q, s)$ be a basic optimal solution corresponding to the grid point (p, q, s) , using this solution, we construct at most eight binary solutions as described earlier in this paragraph, which we denote by $\widehat{\mathbf{x}}^k(p, q, s)$ for $k = 1, \dots, 8$. In this case, considering all points in the geometric grid, the heuristic comes up with the binary solutions $\{\mathbf{x}^k(p, q, s) : (p, q, s) \in \text{Grid}^3, k = 1, \dots, 8\}$. Among these solutions, the heuristic returns the one that provides the largest expected revenue while satisfying the constraint that the total capacity consumption of the offered products does not exceed C .

To check the optimality gap of the heuristic in the previous paragraph, we also use problem (2) to construct an upper bound on the optimal objective value of the **Capacitated Mixture** problem. We consider any $K + 1$ points $\underline{p} = \tilde{p}^1 < \dots < \tilde{p}^{K+1} = \bar{p}$, any $L + 1$ points $\underline{q} = \tilde{q}^1 < \dots < \tilde{q}^{L+1} = \bar{q}$ and any $M + 1$ points $\underline{s} = \tilde{s}^1 < \dots < \tilde{s}^{M+1} = \bar{s}$. Noting that $\tilde{G}(p, q, s)$ is the optimal objective value of the LP in (3), we compute $\tilde{G}(\tilde{p}^k, \tilde{q}^\ell, \tilde{s}^{m+1})$ for all $k = 1, \dots, K, \ell = 1, \dots, L, m = 1, \dots, M$. In this case, we can show that $\max_{k=1, \dots, K, \ell=1, \dots, L, m=1, \dots, M} \{\lambda \tilde{p}^{k+1} + (1 - \lambda) \frac{\tilde{q}^{\ell+1}}{1 + \tilde{s}^m} : \tilde{G}(\tilde{p}^k, \tilde{q}^\ell, \tilde{s}^{m+1}) \leq C\}$ provides an upper bound on the optimal objective value of the **Capacitated Mixture** problem.

In Appendix G, we show that the approach described above indeed provides an upper bound on the optimal expected revenue in the **Capacitated Mixture** problem. This upper bound holds for any choice of the points $\{\tilde{p}^1, \dots, \tilde{p}^{K+1}\}$, $\{\tilde{q}^1, \dots, \tilde{q}^{L+1}\}$ and $\{\tilde{s}^1, \dots, \tilde{s}^{M+1}\}$, but it gets tighter as the separation between the points gets smaller. In the same appendix, we give a numerical study to test the performance of the heuristic. The heuristic performs remarkably well. Comparing the expected revenue of the solutions from the heuristic with the upper bound on the optimal expected revenue, the average optimality gap of the heuristic comes out to be fraction of a percent.

6. Network Revenue Management

In the network revenue management setting, we have a set of resources indexed by $M = \{1, \dots, m\}$ and a set of products indexed by $N = \{1, \dots, n\}$. The capacity of resource q is c_q . There are T time periods in the selling horizon. At each time period in the selling horizon, a customer arrives into the system and we offer an assortment of products. If we offer the assortment S of products, then a customer purchases product $i \in S$ with probability $\lambda \theta_i + (1 - \lambda) \frac{v_i}{1 + V(S)}$, in which case, we generate a revenue of r_i and consume a_{qi} units of the capacity of resource q . The goal is to find a policy to decide which assortment of products to offer to each customer so that we maximize the total expected revenue over the selling horizon. We can formulate a dynamic program to find the optimal policy but the state variable in this dynamic program has to keep track of the remaining capacity of each resource, resulting in a high-dimensional state variable when the number of resources is large. We focus on an LP approximation instead (Gallego et al. 2004, Liu and van Ryzin 2008).

We use the decision variables $\mathbf{w} = \{w(S) : S \subseteq N\}$, where $w(S)$ is the probability that we offer assortment S of products at a time period. Consider the LP

$$\begin{aligned} \max_{\mathbf{w} \in \mathbb{R}_+^{2^n}} \left\{ T \sum_{S \subseteq N} \sum_{i \in S} r_i \left(\lambda \theta_i + (1 - \lambda) \frac{v_i}{1 + V(S)} \right) w(S) \quad : \quad \right. & \text{(Choice-Based LP)} \\ T \sum_{S \subseteq N} \sum_{i \in S} a_{qi} \left(\lambda \theta_i + (1 - \lambda) \frac{v_i}{1 + V(S)} \right) w(S) \leq c_q \quad \forall q \in M, & \\ \left. \sum_{S \subseteq N} w(S) = 1 \right\}. & \end{aligned}$$

Noting that $\sum_{i \in S} r_i (\lambda \theta_i + (1 - \lambda) \frac{v_i}{v_0 + V(S)})$ is the expected revenue at a time period during which we offer the assortment S , the objective function is the total expected revenue over the selling horizon. Similarly, $\sum_{i \in S} a_{qi} (\lambda \theta_i + (1 - \lambda) \frac{v_i}{v_0 + V(S)})$ is the expected capacity consumption of resource q at a time period during which we offer the assortment S , so 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. The number of decision variables above increases exponentially with the number of products, so solving the Choice-Based LP almost always requires using column generation. The column generation subproblem has the same structure as the Mixture problem. As we discuss below, there are heuristics that use an optimal primal or dual solution to the Choice-Based LP to decide which assortment to offer at each time period.

In a randomized offer policy, letting \mathbf{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). 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 revenue from the purchase is $r_i - \sum_{q \in M} a_{qi} \mu_q^*$. Thus, the expected net revenue from offering assortment S is $\sum_{i \in S} (\lambda \theta_i + (1 - \lambda) \frac{v_i}{1 + V(S)}) (r_i - \sum_{q \in M} a_{qi} \mu_q^*)$, 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).

Next, we give an equivalent formulation of the Choice-Based LP, where the numbers of decision variables and constraints increase polynomially with the number of products and resources.

Equivalent Formulation:

Noting that we have n products and m resources, our equivalent formulation will have $n^2 + n + 1$ decision variables and $2n^2 + n + m + 1$ constraints. Thus, while solving the Choice-Based LP almost

always requires column generation, we can directly solve our equivalent formulation. Using the decision variables $(x_0, \mathbf{x}, \mathbf{y}) \in \mathbb{R} \times \mathbb{R}_+^{n+n^2}$ as in the Assortment LP, we consider the LP

$$\begin{aligned} \max_{(x_0, \mathbf{x}, \mathbf{y}) \in \mathbb{R} \times \mathbb{R}_+^{n+n^2}} & \left\{ T \sum_{i \in N} r_i \left(\left[\lambda \theta_i + (1 - \lambda) v_i \right] x_i + \lambda \theta_i \sum_{j \in N} v_j y_{ij} \right) : \right. & \text{(Compact LP)} \\ & T \sum_{i \in N} a_{qi} \left(\left[\lambda \theta_i + (1 - \lambda) v_i \right] x_i + \lambda \theta_i \sum_{j \in N} v_j y_{ij} \right) \leq c_q \quad \forall q \in M, \\ & x_0 + \sum_{i \in N} v_i x_i = 1, \\ & x_i \leq x_0 \quad \forall i \in N, \\ & \left. y_{ij} \leq x_i \quad \forall i, j \in N, \quad y_{ij} \leq x_j \quad \forall i, j \in N \right\}. \end{aligned}$$

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 $[\lambda \theta_i + (1 - \lambda) v_i] x_i + \lambda \theta_i \sum_{j \in N} v_j y_{ij}$ to capture the expected number of purchases for product i at a time period, even though we allow offering different assortments of products in the Choice-Based LP. 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. In the rest of this section, we show that the Compact LP is equivalent to the Choice-Based LP in the sense that we can use an optimal primal or dual solution to the Compact LP to obtain an optimal primal or dual solution to the Choice-Based LP. In this case, we can solve the Compact LP to implement the randomized offer, bid-price or any other policy that uses an optimal primal or dual solution to the Choice-Based LP. Our computational results indicate that using Compact LP instead of Choice-Based LP results in substantial improvements in running times. To give our equivalence result, we start by considering the question of obtaining an optimal primal solution to the Choice-Based LP by using an optimal primal solution to the Compact LP.

Let $(x_0^*, \mathbf{x}^*, \mathbf{y}^*)$ be a basic optimal solution to the Compact LP. Indexing the products so that $x_1^* \geq x_2^* \geq \dots \geq x_n^*$ and defining the set $S_i = \{1, \dots, i\}$ with $S_0 = \emptyset$, let

$$\begin{aligned} \widehat{w}(S_i) &= (x_i^* - x_{i+1}^*) (1 + V(S_i)) \quad \forall i = 0, 1, \dots, n & \text{(Recovery)} \\ \widehat{w}(S) &= 0 \quad \forall S \notin \{S_0, S_1, \dots, S_n\}, \end{aligned}$$

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 $\widehat{w}(S_0) = x_0^* - x_1^* \geq 0$.

The Recovery formula gives a closed-form expression to construct a solution $\widehat{\mathbf{w}} = \{\widehat{w}(S) : S \subseteq N\}$ to the Choice-Based LP by using a basic optimal solution $(x_0^*, \mathbf{x}^*, \mathbf{y}^*)$ to the Compact LP. Other

than being non-negative, it is not clear that the solution $\hat{\mathbf{w}}$ is feasible to the Choice-Based LP. In the next theorem, we show that this solution is not only feasible but also optimal. Also, we show that we can construct an optimal dual solution to the Choice-Based LP by using an optimal dual solution to the Compact LP. Thus, we can use an optimal primal or dual solution to the Compact LP to construct the same to the Choice-Based LP. In this result, 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 . To implement this convention in practice, for a small value of $\epsilon > 0$, we can add the additional term ϵx_0 to the objective function of the Compact LP, in which case, the additional term favors an optimal solution with the largest value of x_0 . In addition to following this convention, we associate 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 $\boldsymbol{\sigma} = \{\sigma_{ij} : i, j \in N\}$ with the constraints in the Compact LP.

Theorem 6.1 (Equivalent Formulation) *Letting $(x_0^*, \mathbf{x}^*, \mathbf{y}^*)$ and $(\boldsymbol{\mu}^*, \pi^*, \boldsymbol{\alpha}^*, \boldsymbol{\eta}^*, \boldsymbol{\sigma}^*)$ be a basic optimal primal and dual solution pair to the Compact LP, we have the following results.*

- (a) *The solution $\hat{\mathbf{w}}$ provided by the Recovery formula is optimal to the Choice-Based LP.*
- (b) *The solution $(\boldsymbol{\mu}^*, \pi^*)$ is optimal to the dual of the Choice-Based LP.*

The theorem above shows that we can efficiently recover an optimal primal or dual solution to the Choice-Based LP by using the same type of solution from the Compact LP. We give a proof for the theorem in Appendix H. The proof explicitly uses the fact that we focus on basic optimal solutions to the Compact LP that have the largest value for the decision variable x_0 . In Appendix I, we provide computational experiments to check the improvements in running times obtained by using the Compact LP in conjunction with Theorem 6.1 to get an optimal solution to the Choice-Based LP, rather than solving the Choice-Based LP directly by using column generation. Our results indicate that we can improve running times by up to a factor of 29.

7. Computational Experiments

We give computational experiments to test the ability of our mixture model to predict the choice process of the customers and to identify profitable assortments. We start by comparing our mixture model with the standard multinomial logit and independent demand models. In this way, we quantify the benefits obtained by mixing the independent demand and multinomial logit models, rather than using each of these choice models by itself. In addition, we compare our mixture model with the exponential and Markov chain choice models. Next, we outline an expectation-maximization algorithm to estimate the parameters of our mixture choice model. We describe our experimental setup. Finally, we give our computational results.

7.1 Expectation-Maximization Algorithm

The parameters of our mixture choice model are $(\lambda, \boldsymbol{\theta}, \mathbf{v})$, where λ is the relative size of the two segments, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$ is the vector of demand probabilities in the independent demand segment and $\mathbf{v} = (v_1, \dots, v_n)$ is the vector of preference weights in the multinomial logit segment. We use $\mathcal{H} = \{(S_t, i_t) : t = 1, \dots, \tau\}$ to capture the purchase history that we use to estimate the parameters of our mixture model, where τ is the number of customers in the purchase history, S_t is the assortment of products offered to customer t and i_t is the product purchased by customer t . If customer t did not purchase anything, then $i_t = 0$. Our expectation-maximization algorithm uses the following observation. In addition to the purchase history $\mathcal{H} = \{(S_t, i_t) : t = 1, \dots, \tau\}$, if we knew the segment of each customer, then we could partition the customers in the two segments and separately fit an independent demand and multinomial logit models to the purchases from the two partitions.

To pursue this observation, for the moment, assume that we have access to the segment of each customer in the purchase history. In particular, we use $\mathcal{C} = \{(z_t, S_t, i_t) : t = 1, \dots, \tau\}$ to capture the so-called complete purchase history, where we have $z_t = 1$ if customer t was in the independent demand segment, whereas we have $z_t = 0$ if customer t was in the multinomial logit segment. If $z_t = 1$, so that customer t is in the independent demand segment, then she chooses product $i \in S_t$ with probability θ_i , whereas she leaves without a purchase with probability $1 - \sum_{i \in S_t} \theta_i$. If $z_t = 0$, so that customer t is in the multinomial logit segment, then she chooses product $i \in S_t$ with probability $\frac{v_i}{1 + V(S_t)}$, whereas she leaves without a purchase with probability $\frac{1}{1 + V(S_t)}$. Thus, if we had access to the complete purchase history $\mathcal{C} = \{(z_t, S_t, i_t) : t = 1, \dots, \tau\}$, then we could estimate the parameters of our mixture model by maximizing the likelihood function given by

$$L(\lambda, \boldsymbol{\theta}, \mathbf{v}; \mathcal{C}) = \prod_{t=1}^{\tau} \left\{ \left(\lambda \prod_{i \in S_t} \theta_i^{\mathbf{1}(i_t=i)} \left(1 - \sum_{i \in S_t} \theta_i\right)^{\mathbf{1}(i_t=0)} \right)^{z_t} \left((1 - \lambda) \frac{\prod_{i \in S_t} v_i^{\mathbf{1}(i_t=i)} \mathbf{1}^{\mathbf{1}(i_t=0)}}{1 + V(S_t)} \right)^{1-z_t} \right\}. \quad (4)$$

In our expectation-maximization algorithm, we use the purchase history $\{(S_t, i_t) : t = 1, \dots, \tau\}$, rather than the complete purchase history $\mathcal{C} = \{(z_t, S_t, i_t) : t = 1, \dots, \tau\}$, but we iteratively estimate the portion $\{z_t : t = 1, \dots, \tau\}$ of the complete purchase history. We need two random variables in the estimation procedure. We use the random variable Z with support $\{0, 1\}$ to capture the segment of a generic customer, where $Z = 1$ if and only if the customer is in the independent demand segment. For any $S \subseteq N$, we use the random variable $P(S)$ with support $S \cup \{0\}$ to capture the choice of a generic customer within the assortment S , where $P(S) = i$ if and only if the customer purchases product i within the assortment S . If the choices of the customers are governed by our mixture model with parameters $(\lambda, \boldsymbol{\theta}, \mathbf{v})$, then Z is Bernoulli with parameter λ and $P(S)$ takes value $i \in S$ with probability $\lambda \theta_i + (1 - \lambda) \frac{v_i}{1 + V(S)}$. At any iteration ℓ of our expectation-maximization algorithm,

we have the current parameter estimates $(\lambda^\ell, \boldsymbol{\theta}^\ell, \mathbf{v}^\ell)$. Letting $\{\bar{z}_t^\ell : t = 1, \dots, \tau\}$ be the estimates of $\{z_t : t = 1, \dots, \tau\}$ at iteration ℓ , we compute \bar{z}_t^ℓ as the expectation of Z conditional on the fact that customer t chooses according to our mixture model with parameters $(\lambda^\ell, \boldsymbol{\theta}^\ell, \mathbf{v}^\ell)$ and her purchase decision within the assortment S_t is i_t . In this way, we estimate the complete purchase history $\mathcal{C}^\ell = \{(\bar{z}_t^\ell, S_t, i_t) : t = 1, \dots, \tau\}$ at iteration ℓ . For large $B \in \mathbb{R}_+$ such that $v_i \leq B$ for all $i \in N$, the parameters of our model are in the set $\mathcal{P} = \{(\lambda, \boldsymbol{\theta}, \mathbf{v}) \in [0, 1]^{1+n} \times [0, B]^n : \sum_{i \in N} \theta_i \leq 1\}$. Using the estimated complete purchase history, we maximize the likelihood $L(\lambda, \boldsymbol{\theta}, \mathbf{v}; \mathcal{C}^\ell)$ in (4) subject to the constraint that $(\lambda, \boldsymbol{\theta}, \mathbf{v}) \in \mathcal{P}$ to get the parameter estimates $(\lambda^{\ell+1}, \boldsymbol{\theta}^{\ell+1}, \mathbf{v}^{\ell+1})$ at the next iteration.

Below is an overview of our expectation-maximization algorithm.

Step 1. The purchase history to estimate the parameters is $\{(S_t, i_t) : t = 1, \dots, \tau\}$. Choose the initial parameter estimates $(\lambda^1, \boldsymbol{\theta}^1, \mathbf{v}^1) \in \mathcal{P}$. Initialize the iteration counter by setting $\ell = 1$.

Step 2. (Expectation) Given that the customers choose according to the mixture model with parameters $(\lambda^\ell, \boldsymbol{\theta}^\ell, \mathbf{v}^\ell)$, for all $t = 1, \dots, \tau$, set $\bar{z}_t^\ell = \mathbb{E}\{Z | P(S_t) = i_t\}$.

Step 3. (Maximization) Using the complete purchase history $\mathcal{C}^\ell = \{(\bar{z}_t^\ell, S_t, i_t) : t = 1, \dots, \tau\}$, set $(\lambda^{\ell+1}, \boldsymbol{\theta}^{\ell+1}, \mathbf{v}^{\ell+1}) = \arg \max_{(\lambda, \boldsymbol{\theta}, \mathbf{v}) \in \mathcal{P}} L(\lambda, \boldsymbol{\theta}, \mathbf{v}; \mathcal{C}^\ell)$. Increase ℓ by one and go to Step 2.

In the expectation step, we compute a conditional expectation. In the maximization step, we solve an optimization problem. We can perform both of these tasks efficiently.

Computing the Conditional Expectation. In the expectation step, given that the customers choose according to our mixture model with some parameters $(\lambda, \boldsymbol{\theta}, \mathbf{v})$, we compute an expectation of the form $\mathbb{E}\{Z | P(S) = i\}$. We have a closed-form expression for this expectation. The support of the random variable Z is $\{0, 1\}$, so $\mathbb{E}\{Z | P(S) = i\} = \mathbb{P}\{Z = 1 | P(S) = i\} = \frac{\mathbb{P}\{Z=1, P(S)=i\}}{\mathbb{P}\{P(S)=i\}}$. The expression in the numerator is the probability that a customer is in the independent demand segment and she purchases product i , so this probability is $\lambda \theta_i$. The expression in the denominator is the probability that a customer purchases product i , so this probability is $\lambda \theta_i + (1 - \lambda) \frac{v_i}{1 + V(S)}$. We can use a similar argument to compute the expectation $\mathbb{E}\{Z | P(S) = 0\}$.

Solving the Optimization Problem. In the maximization step, given some complete purchase history $\mathcal{C} = \{(z_t, S_t, i_t) : t = 1, \dots, \tau\}$, we solve a problem of the form $\max_{(\lambda, \boldsymbol{\theta}, \mathbf{v}) \in \mathcal{P}} L(\lambda, \boldsymbol{\theta}, \mathbf{v}; \mathcal{C})$. We can formulate this problem as a convex program with linear constraints. Furthermore, the objective function of this convex program is separable by the decision variables $(\lambda, \boldsymbol{\theta}, \mathbf{v})$. Therefore, we can solve the optimization problem in the maximization step efficiently. This result intuitively builds on our earlier observation that if we have access to the complete purchase history, then we can estimate the parameters of the independent demand and multinomial logit models separately by independently focusing on the customers in the two segments. To make this discussion concrete,

rather than maximizing the likelihood function, we maximize its logarithm. Taking the logarithm in (4), we can express $\log L(\lambda, \boldsymbol{\theta}, \mathbf{v}; \mathcal{C})$ as $L_1(\lambda; \mathcal{C}) + L_2(\boldsymbol{\theta}; \mathcal{C}) + L_3(\mathbf{v}; \mathcal{C})$, where we have

$$\begin{aligned} L_1(\lambda; \mathcal{C}) &= \sum_{t=1}^{\tau} \left\{ z_t \log \lambda + (1 - z_t) \log(1 - \lambda) \right\} \\ L_2(\boldsymbol{\theta}; \mathcal{C}) &= \sum_{t=1}^{\tau} z_t \left\{ \sum_{i \in S_t} \mathbf{1}(i = i_t) \log \theta_i + \mathbf{1}(i_t = 0) \log \left(1 - \sum_{i \in S_t} \theta_i \right) \right\} \\ L_3(\mathbf{v}; \mathcal{C}) &= \sum_{t=1}^{\tau} (1 - z_t) \left\{ \sum_{i \in S_t} \mathbf{1}(i = i_t) \log v_i - \log(1 + V(S_t)) \right\}. \end{aligned}$$

Thus, we can equivalently solve the problems $\max_{\lambda \in [0,1]} L_1(\lambda; \mathcal{C})$, $\max_{\boldsymbol{\theta} \in [0,1]^n: \sum_{i \in N} \theta_i \leq 1} L_2(\boldsymbol{\theta}; \mathcal{C})$ and $\max_{\mathbf{v} \in [0,B]^n} L_3(\mathbf{v}; \mathcal{C})$ in the maximization step. Since $\log x$ is concave in x , $L_1(\lambda; \mathcal{C})$ and $L_2(\boldsymbol{\theta}; \mathcal{C})$ are concave in, respectively, λ and $\boldsymbol{\theta}$. The log-sum-exp function $\log(1 + \sum_{i=1}^n e^{x_i})$ is convex in (x_1, \dots, x_n) (Example 3.1.5, Boyd and Vandenberghe 2004). In this case, if we make the change of variables $v_i = e^{\mu_i}$ in $L_3(\mathbf{v}; \mathcal{C})$, then $L_3(\boldsymbol{\mu}; \mathcal{C})$ is concave in $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)$ as well. Therefore, by the discussion in the last two paragraphs, we can compute the conditional expectations in the expectation step and solve the optimization problems in the maximization step in our expectation-maximization algorithm efficiently. In Appendix J, we combine the discussion in the last two paragraphs with the outline of our expectation-maximization algorithm to give a step-by-step specification of the algorithm. In the same appendix, we also argue that the sequence of parameter estimates $\{(\lambda^\ell, \boldsymbol{\theta}^\ell, \mathbf{v}^\ell) : \ell = 1, 2, \dots\}$ generated by the expectation-maximization algorithm monotonically increases the likelihood function built by using the purchase history $\{(S_t, i_t) : t = 1, \dots, \tau\}$. Thus, we can stop the algorithm when the increase in the value of this likelihood function in the successive iterations falls below a threshold.

We can parameterize $(\lambda, \boldsymbol{\theta}, \mathbf{v})$ in our mixture model as a function of the features of the products, which becomes useful when one desires a parsimonious model or a new product with no purchase data is introduced. The features of product i are given by $\mathbf{f}_i = (f_{i1}, \dots, f_{im})$, where f_{ik} is the value of feature k for product i . We postulate the following Poisson arrival process. Customers in the multinomial logit segment arrive with rate Λ . These customers are willing to consider any product, so their arrival rate is independent of product features, but they can leave without a purchase. If we offer the assortment S , then each chooses product i with probability $\frac{e^{\boldsymbol{\beta}^\top \mathbf{f}_i}}{1 + \sum_{j \in S} e^{\boldsymbol{\beta}^\top \mathbf{f}_j}}$. For $\Theta : \mathbb{R} \rightarrow \mathbb{R}_+$, customers in the independent demand segment for product i arrive with rate $\Theta(\boldsymbol{\alpha}^\top \mathbf{f}_i)$. These customers only consider product i . Thus, over an infinitesimal period δ , as a function of the offered assortment S , the purchase probability of product $i \in S$ is $\delta \Theta(\boldsymbol{\alpha}^\top \mathbf{f}_i) + \delta \Lambda \frac{e^{\boldsymbol{\beta}^\top \mathbf{f}_i}}{1 + \sum_{j \in S} e^{\boldsymbol{\beta}^\top \mathbf{f}_j}}$. The parameters $(\Lambda, \boldsymbol{\alpha}, \boldsymbol{\beta}) \in \mathbb{R}_+ \times \mathbb{R}^{2m}$ are to be estimated. Representing the parameters $(\lambda, \boldsymbol{\theta}, \mathbf{v})$ in our mixture model as $\lambda = 1 - \delta \Lambda$, $\theta_i = \frac{\delta}{1 - \delta \Lambda} \Theta(\boldsymbol{\alpha}^\top \mathbf{f}_i)$ and $v_i = e^{\boldsymbol{\beta}^\top \mathbf{f}_i}$, the last choice probability is $\lambda \theta_i + (1 - \lambda) \frac{v_i}{1 + \sum_{j \in S} v_j}$. Possible choices for Θ , for example, are $\Theta(x) = e^x$ or $\Theta(x) = K e^x / (1 + e^x)$ with $K > 0$, both choices are simply motivated by the fact that Θ should take values in positive reals.

7.2 Experimental Setup

In our computational experiments, we consider the purchase histories of customers making purchases according to a complex ground choice model that does not comply with the independent demand or multinomial logit models. We fit our mixture model, as well as the multinomial logit, independent demand, exponential and Markov choice models, to the purchase histories and compare the prediction ability of these fitted choice models. The ground choice model is the nonparametric choice model, where we populate the preference lists through a dataset from Kamishima (2018), including the rankings of sushi varieties by diners. In the nonparametric choice model, we have p customer types. Indexing the customer types by $P = \{1, \dots, p\}$ and the products by $N = \{1, \dots, n\}$, customers of type ℓ are characterized by a preference list of products $\sigma^\ell = (\sigma^\ell(1), \sigma^\ell(2), \dots, \sigma^\ell(k^\ell))$ with $\sigma^\ell(i) \in N$ for all $i = 1, \dots, k^\ell$, 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 with probability η^ℓ . An arriving customer chooses the most preferred product in her preference list that is also available in the offered assortment. If no product in her preference list is available, then she leaves without a purchase. Thus, the parameters of the nonparametric choice model are the preference lists $\{\sigma^\ell : \ell \in P\}$ and the arrival probabilities (η^1, \dots, η^p) .

In the dataset, we have the rankings of 10 sushi varieties declared by 5000 diners. We use $\rho^\ell = (\rho^\ell(1), \dots, \rho^\ell(10))$ to capture the ranking declared by diner ℓ , where $\rho^\ell(i)$ is the i th most preferred sushi variety for diner ℓ . We use this dataset to populate the preference lists in our ground choice model as follows. We associate one customer type with each diner and one product with each sushi variety, so the set of customer types is $P = \{1, \dots, 5000\}$ and the set of products is $N = \{1, \dots, 10\}$. To come up with the preference list $\sigma^\ell = (\sigma^\ell(1), \sigma^\ell(2), \dots, \sigma^\ell(k^\ell))$ for customer type ℓ , we sample k^ℓ from the geometric distribution with parameter $e^{-\psi} \in (0, 1)$ truncated over the interval $\{1, \dots, 10\}$ and set $(\sigma^\ell(1), \dots, \sigma^\ell(k^\ell)) = (\rho^\ell(1), \dots, \rho^\ell(k^\ell))$. In other words, we sample the length of the preference list k^ℓ for each customer type ℓ from the probability mass function $f(k) = \frac{e^{-\psi k}}{\sum_{q=1}^{10} e^{-\psi q}}$, in which case, the preference list of customer type ℓ is the ranking declared by diner ℓ cut off after the first k^ℓ sushi varieties. We vary ψ in our computational experiments. The expected length of the preference lists increases as the value of ψ decreases. Customers of each type arrive into the system with equal probability, so we have $\eta^\ell = 1/5000$ for all $\ell \in P$. The discussion so far describes our approach for using the dataset to populate the preference lists $\{\sigma^\ell : \ell \in P\}$ and the arrival probabilities (η^1, \dots, η^p) in the ground choice model.

Once we come up with the ground choice model by using the dataset as described in the previous paragraph, we generate the purchase histories of customers making purchases according to the ground choice model. The purchase history consists of the pairs $\{(S_t, i_t) : t = 1, \dots, \tau\}$, as discussed

in Section 7.1. To generate the assortment S_t , we include each product in the assortment S_t with probability 0.5. We sample the product i_t among the products in S_t and the no-purchase option according to the ground choice model. In particular, noting that customers of each type arrive into the system with equal probability, we sample the type of customer t from the uniform distribution over $\{1, \dots, 5000\}$, in which case, if customer t is of type ℓ , then this customer chooses the most preferred product in the preference list $\sigma^\ell = (\sigma^\ell(1), \sigma^\ell(2), \dots, \sigma^\ell(k^\ell))$ that is also available in the assortment S_t . If none of the products in the preference list $\sigma^\ell = (\sigma^\ell(1), \sigma^\ell(2), \dots, \sigma^\ell(k^\ell))$ is available in the assortment S_t , then customer t leaves without a purchase.

We use maximum likelihood to fit our mixture model, as well as the multinomial logit, independent demand, exponential and Markov chain choice models, to the purchase histories that we generate. We have an expectation-maximization algorithm to fit our mixture model. The log-likelihood function under the multinomial logit and independent demand models are concave in the parameters, so we use steepest ascent. To estimate the parameters of the exponential and Markov chain choice models, we use the Python code provided by Berbeglia et al. (2021). For each product i , we estimate θ_i and v_i separately without using product features. Thus, the parameters (Λ, α, β) at the end of Section 7.1 do not play a role in our estimation procedure.

7.3 Computational Results

The parameter ψ controls the length of the preference lists in our ground choice model. We vary ψ over $\{0.5, 0.6, \dots, 1.0\}$ in our computational experiments. The number of customers in the purchase history is τ . We vary τ over $\{1250, 2500, 5000\}$ to capture three levels of data availability in the training data that we use to fit the choice models. Following the same approach that we use to generate the training data, we also generate the purchase history for another 10000 customers to use as the testing data. For each combination of ψ and τ , we replicated our computational results 50 times to get a better understanding of how much they change from one replication to another. We regenerate the ground choice model, training data and testing data in each of these replications. We compare the fitted choice models in terms of the out-of-sample log-likelihoods and the expected revenues from the assortments obtained by using the fitted choice models to predict the customer purchases. Throughout the discussion of our computational experiments, we use MIX to refer to our fitted mixture choice model, MNL to refer to the fitted pure multinomial logit model, IDM to refer to the fitted pure independent demand model, EXP to refer to the fitted exponential model and MCC to refer to the fitted Markov chain choice model.

Comparing Out-of-Sample Log-Likelihoods. In Table 2, we compare MIX, MNL, IDM, EXP and MCC in terms of their out-of-sample log-likelihoods. In each of the 50 replications, after

$\tau = 1250$																	
ψ	Out-of-Sample Log-Likelihood					MIX vs. MNL			MIX vs. IDM			MIX vs. EXP			MIX vs. MCC		
						Perc.	MIX	MNL	Perc.	MIX	IDM	Perc.	MIX	EXP	Perc.	MIX	MCC
	MIX	MNL	IDM	EXP	MCC	Gap	Btr.	Btr.	Gap	Btr.	Btr.	Gap	Btr.	Btr.	Gap	Btr.	Btr.
0.5	-3,790	-3,800	-4,060	-3,829	-3,843	0.27	45	5	7.12	50	0	1.03	50	0	1.39	48	2
0.6	-3,743	-3,755	-3,941	-3,792	-3,796	0.32	46	4	5.30	50	0	1.31	50	0	1.43	49	1
0.7	-3,694	-3,710	-3,841	-3,752	-3,748	0.43	48	2	3.98	50	0	1.57	50	0	1.47	48	2
0.8	-3,655	-3,673	-3,764	-3,721	-3,706	0.48	46	4	2.97	50	0	1.79	50	0	1.40	48	2
0.9	-3,614	-3,634	-3,695	-3,687	-3,663	0.57	49	1	2.25	50	0	2.03	50	0	1.36	48	2
1.0	-3,579	-3,601	-3,639	-3,657	-3,635	0.61	49	1	1.66	49	1	2.18	50	0	1.54	47	3
Avg.	-3,679	-3,696	-3,823	-3,740	-3,732	0.45	47.2	2.8	3.88	49.8	0.2	1.65	50.0	0.0	1.43	48.0	2.0

$\tau = 2500$																	
ψ	Out-of-Sample Log-Likelihood					MIX vs. MNL			MIX vs. IDM			MIX vs. EXP			MIX vs. MCC		
						Perc.	MIX	MNL	Perc.	MIX	IDM	Perc.	MIX	EXP	Perc.	MIX	MCC
	MIX	MNL	IDM	EXP	MCC	Gap	Btr.	Btr.	Gap	Btr.	Btr.	Gap	Btr.	Btr.	Gap	Btr.	Btr.
0.5	-3,781	-3,796	-4,056	-3,824	-3,793	0.38	50	0	7.26	50	0	1.12	50	0	0.30	37	13
0.6	-3,734	-3,750	-3,937	-3,787	-3,746	0.43	49	1	5.43	50	0	1.41	50	0	0.32	33	17
0.7	-3,685	-3,705	-3,837	-3,747	-3,698	0.55	50	0	4.12	50	0	1.67	50	0	0.34	36	14
0.8	-3,643	-3,668	-3,759	-3,716	-3,657	0.69	50	0	3.19	50	0	1.99	50	0	0.40	41	9
0.9	-3,602	-3,629	-3,691	-3,682	-3,620	0.77	50	0	2.47	50	0	2.24	50	0	0.52	41	9
1.0	-3,567	-3,596	-3,634	-3,652	-3,583	0.80	50	0	1.88	50	0	2.37	50	0	0.43	41	9
Avg.	-3,669	-3,691	-3,819	-3,735	-3,683	0.60	49.8	0.2	4.06	50.0	0.0	1.80	50.0	0.0	0.39	38.2	11.8

$\tau = 5000$																	
ψ	Out-of-Sample Log-Likelihood					MIX vs. MNL			MIX vs. IDM			MIX vs. EXP			MIX vs. MCC		
						Perc.	MIX	MNL	Perc.	MIX	IDM	Perc.	MIX	EXP	Perc.	MIX	MCC
	MIX	MNL	IDM	EXP	MCC	Gap	Btr.	Btr.	Gap	Btr.	Btr.	Gap	Btr.	Btr.	Gap	Btr.	Btr.
0.5	-3,776	-3,793	-4,053	-3,821	-3,773	0.44	50	0	7.33	50	0	1.18	50	0	-0.08	15	35
0.6	-3,730	-3,748	-3,935	-3,784	-3,727	0.50	50	0	5.52	50	0	1.47	50	0	-0.08	14	36
0.7	-3,680	-3,703	-3,835	-3,745	-3,680	0.63	50	0	4.21	50	0	1.75	50	0	0.00	24	26
0.8	-3,638	-3,666	-3,757	-3,713	-3,638	0.77	50	0	3.28	50	0	2.06	50	0	-0.01	24	26
0.9	-3,597	-3,627	-3,689	-3,680	-3,598	0.83	50	0	2.54	50	0	2.30	50	0	0.02	24	26
1.0	-3,562	-3,594	-3,633	-3,649	-3,564	0.89	50	0	1.98	50	0	2.45	50	0	0.07	24	26
Avg.	-3,664	-3,689	-3,817	-3,732	-3,663	0.68	50.0	0.0	4.14	50.0	0.0	1.87	50.0	0.0	-0.01	20.8	29.2

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

generating the ground choice model, we sample training data and testing data using the ground choice model. We fit MIX, MNL, IDM, EXP and MCC to the training data and compute the log-likelihood of the testing data under the fitted choice models. There are three blocks in the table. The top, middle and bottom blocks correspond to the values of $\tau \in \{1250, 2500, 5000\}$. In each block, the first column gives the value of the parameter ψ , controlling the length of the preference lists. The second column gives the average out-of-sample log-likelihood of MIX, where the average is computed over 50 replications. Similarly, the third, fourth, fifth and sixth columns give the average log-likelihoods of the remaining four fitted choice models. The next three columns compare the performance of MIX with that of MNL. In particular, the seventh column gives the average percent gap between out-of-sample log-likelihoods of MIX and MNL. The eighth 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 ninth column gives the number of replications in which the outcome is reversed. In the remaining portion of the table, we use the same approach to compare the out-of-sample log-likelihood performance of MIX with that of IDM, EXP and MCC.

The out-of-sample log-likelihoods of MIX improve those of MNL and IDM consistently, indicating that we obtain significant benefits from mixing the independent demand and multinomial logit

models, as done in our mixture model, rather than using either of the two choice models by itself. We shortly demonstrate that these improvements in the out-of-sample log-likelihoods translate into more profitable assortments. The out-of-sample log-likelihoods of MIX are significantly better than those of EXP. The performance of MIX is competitive to that of MCC. For the training data with $\tau = 1250$ and $\tau = 2500$, corresponding to small to moderate levels of data availability, MIX has an edge over MCC, whereas for the training data with $\tau = 5000$, corresponding to large levels of data availability, MCC has an edge over MIX.

The fact that MCC lags behind MIX when we have small to moderate amounts of data is somewhat expected, as the number of parameters for MCC is significantly larger than that for MIX. In particular, the parameters for MIX are λ , $(\theta_1, \dots, \theta_n)$ and (v_1, \dots, v_n) . In contrast, MCC has one parameter δ_{ij} for each pair of products i and j , characterizing the probability that a customer considers purchasing product j when she currently considers product i and this product is not available, as well as one more parameter ζ_i for each product i , characterizing the probability that an arriving customer considers purchasing product i . Thus, MIX has $2n + 1$ parameters, whereas MCC has $n^2 + n$ parameters. With its larger number of parameters, MCC may potentially provide more flexibility for capturing the choice behavior of the customers, but its large number of parameters may also cause MCC to overfit to the training data, resulting in poor out-of-sample log-likelihoods. The concern for MCC to overfit to the training data is especially significant when we have little training data (Section 1.1, Bishop 2006). The results in Table 2 are consistent with the discussion in this paragraph. When $\tau = 1250$ and $\tau = 2500$, corresponding to smaller and moderate levels of data availability, the out-of-sample log-likelihoods of MIX are indeed larger than those of MCC. Only when $\tau = 5000$, corresponding to larger levels of data availability, MCC catches up with MIX. Thus, MIX is an appealing alternative to MCC, especially when we have small to moderate amounts of training data to estimate the parameters of the choice model that we are fitting. In addition, as discussed in Section 7.1, we can parameterize the independent demand probability θ_i and preference weight v_i of each product i in MIX as functions of the features of the product, but as far as we are aware, there is no work on parameterizing MCC.

Comparing Expected Revenues. In Table 3, we compare MIX, MNL, IDM, EXP and MCC in terms of their expected revenue performance. Recall that we carry out 50 replications for each combination of (ψ, τ) , where we regenerate the ground choice model, as well as the training and testing data, in each replication. In replication q , let $\phi_{i,\text{MIX}}^q(S)$ be the choice probability of product i within the 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 $(r_1^{qk}, \dots, r_n^{qk})$ be the product revenues in the k th sample, we use $\widehat{S}_{\text{MIX}}^{qk}$ to

$\tau = 1250$																	
ψ	Expected Revenue					MIX vs. MNL			MIX vs. IDM			MIX vs. EXP			MIX vs. MCC		
	MIX	MNL	IDM	EXP	MCC	Perc. Gap	MIX Btr.	MNL Btr.	Perc. Gap	MIX Btr.	IDM Btr.	Perc. Gap	MIX Btr.	EXP Btr.	Perc. Gap	MIX Btr.	MCC Btr.
0.5	5.760	5.689	5.504	5.499	5.785	1.23	46	4	4.44	50	0	4.53	50	0	-0.43	17	33
0.6	5.666	5.542	5.504	5.274	5.655	2.19	50	0	2.86	50	0	6.92	50	0	0.19	31	19
0.7	5.598	5.418	5.504	5.077	5.575	3.22	50	0	1.68	50	0	9.31	50	0	0.41	34	16
0.8	5.550	5.319	5.504	4.916	5.502	4.16	50	0	0.83	46	4	11.42	50	0	0.86	42	8
0.9	5.512	5.239	5.504	4.781	5.435	4.95	50	0	0.15	32	18	13.26	50	0	1.40	45	5
1.0	5.490	5.168	5.504	4.664	5.403	5.87	50	0	-0.26	19	31	15.05	50	0	1.58	50	0
Avg.	5.596	5.396	5.504	5.035	5.559	3.60	49.3	0.7	1.62	41.2	8.8	10.08	50.0	0.0	0.67	36.5	13.5

$\tau = 2500$																	
ψ	Expected Revenue					MIX vs. MNL			MIX vs. IDM			MIX vs. EXP			MIX vs. MCC		
	MIX	MNL	IDM	EXP	MCC	Perc. Gap	MIX Btr.	MNL Btr.	Perc. Gap	MIX Btr.	IDM Btr.	Perc. Gap	MIX Btr.	EXP Btr.	Perc. Gap	MIX Btr.	MCC Btr.
0.5	5.784	5.688	5.504	5.497	5.839	1.66	50	0	4.84	50	0	4.96	50	0	-0.95	4	46
0.6	5.696	5.541	5.504	5.271	5.717	2.72	50	0	3.37	50	0	7.46	50	0	-0.37	14	36
0.7	5.625	5.416	5.504	5.076	5.628	3.72	50	0	2.15	50	0	9.76	50	0	-0.05	22	28
0.8	5.584	5.317	5.504	4.918	5.569	4.78	50	0	1.43	50	0	11.93	50	0	0.27	39	11
0.9	5.548	5.237	5.504	4.781	5.514	5.61	50	0	0.79	46	4	13.82	50	0	0.61	45	5
1.0	5.518	5.168	5.504	4.665	5.476	6.34	50	0	0.25	37	13	15.46	50	0	0.76	46	4
Avg.	5.626	5.395	5.504	5.035	5.624	4.14	50.0	0.0	2.14	47.2	2.8	10.57	50.0	0.0	0.04	28.3	21.7

$\tau = 5000$																	
ψ	Expected Revenue					MIX vs. MNL			MIX vs. IDM			MIX vs. EXP			MIX vs. MCC		
	MIX	MNL	IDM	EXP	MCC	Perc. Gap	MIX Btr.	MNL Btr.	Perc. Gap	MIX Btr.	IDM Btr.	Perc. Gap	MIX Btr.	EXP Btr.	Perc. Gap	MIX Btr.	MCC Btr.
0.5	5.800	5.689	5.504	5.497	5.868	1.91	50	0	5.10	50	0	5.22	50	0	-1.17	0	50
0.6	5.713	5.541	5.504	5.268	5.748	3.01	50	0	3.66	50	0	7.79	50	0	-0.61	6	44
0.7	5.640	5.420	5.504	5.076	5.664	3.90	50	0	2.41	50	0	10.00	50	0	-0.43	9	41
0.8	5.597	5.318	5.504	4.913	5.606	4.98	50	0	1.66	50	0	12.22	50	0	-0.16	19	31
0.9	5.561	5.240	5.504	4.777	5.553	5.77	50	0	1.02	49	1	14.10	50	0	0.14	29	21
1.0	5.532	5.169	5.504	4.660	5.520	6.56	50	0	0.51	44	6	15.76	50	0	0.22	34	16
Avg.	5.641	5.396	5.504	5.032	5.660	4.36	50.0	0.0	2.39	48.8	1.2	10.85	50.0	0.0	-0.34	16.2	33.8

Table 3 Expected revenues obtained by the fitted choice models.

denote the optimal assortment to offer under the assumption that the customers choose according to the fitted MIX, so we have $\hat{S}_{\text{MIX}}^{qk} = \arg \max_{S \subseteq N} \sum_{i \in S} r_i^{qk} \phi_{i, \text{MIX}}^q(S)$. Note that customers actually choose according to the ground choice model. In replication q , letting $\phi_{i, \text{GRN}}^q(S)$ be the choice probability of product i within the assortment S under the ground choice model, we compute the actual expected revenue from the assortment $\hat{S}_{\text{MIX}}^{qk}$ as $R_{\text{MIX}}^{qk} = \sum_{i \in \hat{S}_{\text{MIX}}^{qk}} r_i^{qk} \phi_{i, \text{GRN}}^q(\hat{S}_{\text{MIX}}^{qk})$. Averaging over the 100 revenue samples, we capture the expected revenue performance of MIX in replication q by $\text{Rev}_{\text{MIX}}^q = \frac{1}{100} \sum_{k=1}^{100} R_{\text{MIX}}^{qk}$. We capture the expected revenue performance of MNL, IDM, EXP and MCC in replication q by computing $\text{Rev}_{\text{MNL}}^q$, $\text{Rev}_{\text{IDM}}^q$, $\text{Rev}_{\text{EXP}}^q$ and $\text{Rev}_{\text{MCC}}^q$ similarly.

The layout of Table 3 is identical to that of Table 2, except that this table focuses on the expected revenues rather than log-likelihoods. In this table, the first column gives the value of ψ . The second column gives average expected revenue obtained by MIX, where the average is computed over the 50 replications and 100 product revenue samples, so the second column gives the average of $\{\text{Rev}_{\text{MIX}}^q : q = 1, \dots, 50\}$. Similarly, the third, fourth, fifth and sixth columns give the average expected revenues of the remaining four fitted choice models. The next three columns compare the expected revenue performance of MIX with that of MNL. In particular, the seventh column shows the average percent gap between the expected revenues of MIX and MNL. The

eighth column gives the number of replications out of 50 in which $\{\text{Rev}_{\text{MIX}}^q : q = 1, \dots, 50\}$ is larger than $\{\text{Rev}_{\text{MNL}}^q : q = 1, \dots, 50\}$, whereas the ninth column gives the number of replications in which the outcome is reversed. That is, the eighth and ninth column give $\sum_{q=1}^{50} \mathbf{1}(\text{Rev}_{\text{MIX}}^q > \text{Rev}_{\text{MNL}}^q)$ and $\sum_{q=1}^{50} \mathbf{1}(\text{Rev}_{\text{MIX}}^q < \text{Rev}_{\text{MNL}}^q)$. In the remaining portion of the table, we use the same approach to compare the expected revenue performance of MIX with that of IDM, EXP and MCC.

Our results indicate that MIX can noticeably improve the expected revenues of MNL and IDM, the average gaps reaching 4.03% and 2.05%, respectively. The performance of IDM is competitive to that of MIX only for larger values of ψ . As the value of ψ increases, the expected length of the preference lists decreases, so intuitively speaking, the customers become less likely to substitute and the ground choice model gets closer to the independent demand model. Still, over all 900 replications, in 823 of them, the expected revenue performance of MIX is better than that of IDM. Also, we can always maximize the expected revenue under IDM by offering all products, so the expected revenue performance of IDM does not change from one replication to another. Note that EXP consistently lags behind MIX. The performance of MIX is competitive to that of MCC. In 484 out of 900 replications, the expected revenue performance of MIX is better than that of MCC, whereas the outcome is reversed in 416 replications. Similar to out-of-sample log-likelihoods, the expected revenues of MIX tend to be larger than those of MCC especially when $\tau = 1250$, corresponding to smaller levels of data availability. Thus, MIX can be an appealing alternative to MCC when we have a small amount of training data. In Appendix L, we extend our expectation-maximization algorithm to censored demands and give computational experiments. Under censored demands, if a sale does not happen for a certain duration of time, then we do not know whether a customer did not arrive or the arriving customers did not make a purchase.

8. Conclusions

Our mixture choice model is a natural way to improve the flexibility of both the multinomial logit and independent demand models in capturing the choice process of the customers, while ensuring that the corresponding assortment optimization problems remain tractable. Our numerical results indicate that the mixture choice model can be an appealing alternative to the exponential and Markov chain choice models as well. Our work opens several research paths.

Pricing Problems. One can study pricing problems under a mixture of independent demand and multinomial logit models. In this paper, our focus is on assortment optimization and a detailed treatment of the pricing problem is beyond our scope. However, to illustrate research challenges, we discuss one possible approach to incorporate pricing decisions into our mixture model. We index the set of products by $N = \{1, \dots, n\}$. We use the vector $\mathbf{p} = (p_1, \dots, p_n) \in \mathbb{R}_+^n$ to denote the prices

charged for the products. For the independent demand segment, a customer in the independent demand segment is interested in purchasing product i with probability α_i with $\sum_{i \in N} \alpha_i \leq 1$. For $\vartheta_i : \mathbb{R}_+ \rightarrow [0, 1]$, if we charge the price p_i for product i , then a customer in the independent demand segment interested in purchasing product i purchases this product with probability $\vartheta_i(p_i)$. With probability $1 - \vartheta_i(p_i)$, the customer leaves without a purchase. For the multinomial logit segment, if we charge the prices \mathbf{p} for the products, then a customer in the multinomial logit segment purchases product i with probability $\frac{e^{\gamma_i - \beta_i p_i}}{1 + \sum_{j \in N} e^{\gamma_j - \beta_j p_j}}$, where $\{(\gamma_i, \beta_i) : i \in N\}$ are fixed parameters. We can view $\{\beta_i : i \in N\}$ as price sensitivity parameters. An arriving customer is in the independent demand segment with probability λ . Using c_i to denote the marginal cost of product i , to find the prices that maximize the expected profit from a customer, we can solve the problem

$$\max_{\mathbf{p} \in \mathbb{R}_+^N} \left\{ \sum_{i \in N} \left(\lambda \alpha_i \vartheta_i(p_i) + (1 - \lambda) \frac{e^{\gamma_i - \eta_i p_i}}{1 + \sum_{j \in N} e^{\gamma_j - \eta_j p_j}} \right) (p_i - c_i) \right\}. \quad (5)$$

In the objective function above, the expected profit from product i in the independent demand segment is $\alpha_i \vartheta_i(p_i) (p_i - c_i)$, depending only on the price for product i , which is a natural requisite for the independent demand for product i when we make pricing decisions. The expected profit from product i in the multinomial logit segment is $\frac{e^{\mu_i - \beta_i p_i}}{1 + \sum_{j \in N} e^{\mu_j - \beta_j p_j}} (p_i - c_i)$, depending on the prices of all products. We expect $\vartheta_i : \mathbb{R}_+ \rightarrow [0, 1]$ to be decreasing. Possible choices for ϑ_i , for example, are $\vartheta_i(p_i) = e^{-\eta_i p_i}$ or $\vartheta_i(p_i) = \frac{e^{\mu_i - \eta_i p_i}}{1 + e^{\mu_i - \eta_i p_i}}$ with $\eta_i > 0$, both choices are simply motivated by the fact that ϑ_i should be decreasing and take values in $[0, 1]$. Here, we can view $\{\eta_i : i \in N\}$ as price sensitivity parameters. A detailed treatment of the pricing problem is beyond our scope, but in Appendix K, for the second choice of ϑ_i with equal price sensitivity parameters for the products, we give a dynamic program to obtain a solution to problem (5) with a pre-specified optimality loss.

Extra Flexibility. Customers in our independent demand segment do not make a purchase when their most preferred product is not available, but one can introduce extra flexibility into the choice process. Consider a choice model where the customer can substitute to her second most favorite product, when her most favorite product is not available. Similar to the standard multinomial logit model, the utility of product i has the Gumbel distribution with location-scale parameters $(\mu_i, 1)$, so the preference weight of product i is $v_i = e^{\mu_i}$. The utility of the no-purchase option has the Gumbel distribution with location-scale parameters $(0, 1)$. A customer associates utilities with the products and the no-purchase option. The available alternatives are the products in the offered assortment and the no-purchase option. If the alternative with the largest utility is available, then the customer chooses this alternative. Otherwise, the customer substitutes to the alternative with the second largest utility with probability β . With probability $1 - \beta$, the customer does not substitute and leaves. If the customer substitutes to the alternative with the second largest

utility and this alternative is not available either, then she leaves. In Bai et al. (2021), letting $V(S) = \sum_{i \in S} v_i$, we show that the choice probability of product i within the assortment S is

$$\phi_i(S) = \frac{v_i}{1 + V(N)} \left(1 + \beta \sum_{j \in N \setminus S} \frac{v_j}{1 + V(N \setminus \{j\})} \right).$$

Furthermore, we show that the assortment optimization problem is NP-hard under this choice model. Thus, one additional substitution opportunity substantially complicates the problem. In the same paper, for $K > 1$, we consider K substitution opportunities as opposed to only one. Naturally, the assortment optimization problem remains NP-hard in this case. We give a polynomial time approximation scheme (PTAS), where for any $\epsilon \in (0, 1)$, the PTAS obtains a $(1 - \epsilon)$ -approximate solution in running time that is polynomial in n , but exponential in $1/\epsilon$ and K . Unlike our mixture model in this paper, the assortment optimization problem under the choice model described in this paragraph does not admit a polynomial-time solution method and its structure is substantially different from that of the problem that we consider in this paper. Clearly, there may be other approaches for introducing extra substitution opportunities into the choice process and studying the corresponding assortment optimization problems is an interesting area of research.

Other Mixtures. Our results for exactly solving the unconstrained assortment optimization problem under our mixture model closely exploit the structure of the independent demand and multinomial logit models in the mixture. Exact solutions under mixtures of other choice models will likely require different approaches. Nevertheless, we built on the FPTAS for the knapsack problem to develop an FPTAS for the capacitated assortment optimization problem under our choice model. One may use a similar approach to give an FPTAS under other mixtures when the objective function involves sums of fractions, which is the case, for example, for the nested logit model. The key is to mix choice models to improve their ability to capture the choice behavior, while keeping the resulting optimization problems tractable.

Fitting Choice Models. There are challenges in fitting choice models in practice. First, demand may be censored, which refers to the fact that if we did not have sales over a period of time, then we may not know whether no customer arrivals occurred or the arriving customers did not purchase anything. In Appendix L, we extend our expectation-maximization algorithm to censored demands and test its performance. Second, the offered assortments change from customer to customer in our computational experiments, but the product prices may change too. In this case, we may parameterize the preference weights and independent demand purchase probabilities as a function of the prices as discussed at the end of Section 7.1. Alternatively, for each product i , we may associate a different preference weight v_i and purchase probability θ_i for different price levels with

the constraint that the preference weight and purchase probability for larger prices are smaller. These issues occur when working with not only our mixture model but all choice models.

Acknowledgements: We thank the area editor, associate editor and two anonymous referees whose comments helped us significantly strengthen both the technical and computational results.

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.
- Bai, Y., J. Feldman, H. Topaloglu, L. Wagner. 2021. Assortment optimization under the multinomial logit model in the presence of endogenous consideration sets. Tech. rep., Cornell University, New York, NY. Available at <http://people.orie.cornell.edu/~huseyin/publications/publications.html>.
- Berbeglia, G., A. Garassino, G. Vulcano. 2021. A comparative empirical study of discrete choice models in retail operations. *Management Science* (forthcoming).
- Bishop, C. M. 2006. *Pattern Recognition and Machine Learning*. Springer, New York, NY.
- Blanchet, J., G. Gallego, V. Goyal. 2016. A Markov chain approximation to choice modeling. *Operations Research* **64**(4) 886–905.
- Boyd, S., L. Vandenberghe. 2004. *Convex optimization*. Cambridge University Press, Cambridge, UK.
- 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.

- Kamishima, T. 2018. Sushi preference datasets. URL <https://www.kamishima.net/sushi/>. Last access date: January 8, 2021.
- Kunnunkal, S., V. Martinez-de-Albeniz. 2019. Tractable approximations for assortment planning with product costs. *Operations Research* **67**(2) 436–452.
- Kunnunkal, 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.
- McFadden, D. 1980. Econometric models for probabilistic choice among products. *The Journal of Business* **53**(3) S13–S29.
- McFadden, D., K. Train. 2000. Mixed MNL models for discrete response. *Journal of Applied Economics* **15** 447–470.
- 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.
- Wu, C. F. J. 1983. On the convergence properties of the EM algorithm. *The Annals of Statistics* **11**(1) 95–103.
- 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 a Mixture of Multinomial Logit Models

We argue that it is impossible to express our choice model as a mixture of multinomial logit models with two segments. In particular, we consider a mixture of multinomial logit models with two segments. The set of products is $N = \{1, \dots, n\}$. Customers in the first segment associate the preference v_i with product i , whereas customers in the second segment associate the preference weight w_i with product i . An arriving customer is in the first segment with probability λ . As a function of the preference weights $\mathbf{v} = (v_1, \dots, v_n)$ and $\mathbf{w} = (w_1, \dots, w_n)$ and the relative size λ of the two segments, we use $\psi_i(S; \lambda, \mathbf{v}, \mathbf{w})$ to denote the choice probability of product $i \in S$ within the assortment S . Thus, letting $V(S) = \sum_{i \in S} v_i$ and $W(S) = \sum_{i \in S} w_i$ for notational brevity, we have $\psi_i(S; \lambda, \mathbf{v}, \mathbf{w}) = \lambda \frac{v_i}{1+V(S)} + (1-\lambda) \frac{w_i}{1+W(S)}$ for all $i \in S$. We consider the independent demand model, where product i , if offered, is purchased with probability $1/(1+n)$. This independent demand model is an instance of our mixture model. Consider calibrating the parameters of the mixture of multinomial logit models so that the choice probability of each product i out of the full assortment N matches its corresponding choice probability under the independent demand model. We show that the choice probability of some product i out of the singleton assortment $\{i\}$ must deviate from its corresponding choice probability under the independent demand model by a fixed constant. In particular, we show that if we calibrate the parameters of the mixture of multinomial logit models so that $\psi_i(N; \lambda, \mathbf{v}, \mathbf{w}) = \frac{1}{1+n}$ for all $i \in N$, then we must have $\max_{i \in N} |\psi_i(\{i\}; \lambda, \mathbf{v}, \mathbf{w}) - \frac{1}{1+n}| \geq \frac{1}{8}$ as long as $n \geq 20$. Thus, we cannot calibrate the two choice models to match the choice probabilities of all products out of all assortments. Consider the problem

$$Z_n^* = \min_{(\lambda, \mathbf{v}, \mathbf{w}) \in [0,1] \times \mathbb{R}_+^{2n}} \left\{ \sum_{i \in N} \left| \psi_i(\{i\}; \lambda, \mathbf{v}, \mathbf{w}) - \frac{1}{n+1} \right| : \psi_i(N; \lambda, \mathbf{v}, \mathbf{w}) = \frac{1}{n+1} \quad \forall i \in N \right\}. \quad (6)$$

Noting that Z_n^* is the optimal objective value of the problem above, in the next theorem, we show that Z_n^* increases linearly with the number of products.

Theorem A.1 *For all $n \geq 1$, $Z_n^* \geq \frac{n}{4} - \frac{5}{2}$.*

By the theorem above, for any $(\lambda, \mathbf{v}, \mathbf{w}) \in [0, 1] \times \mathbb{R}_+^{2n}$ that satisfies $\psi_i(N; \lambda, \mathbf{v}, \mathbf{w}) = \frac{1}{n+1}$ for all $i \in N$, we have $\frac{1}{n} \sum_{i \in N} \left| \psi_i(\{i\}; \lambda, \mathbf{v}, \mathbf{w}) - \frac{1}{n+1} \right| \geq \frac{1}{4} - \frac{5}{2n}$. Since $\max_{i \in N} a_i \geq \frac{1}{n} \sum_{i \in N} a_i$, we have

$$\max_{i \in N} \left| \psi_i(\{i\}; \lambda, \mathbf{v}, \mathbf{w}) - \frac{1}{n+1} \right| \geq \frac{1}{n} \sum_{i \in N} \left| \psi_i(\{i\}; \lambda, \mathbf{v}, \mathbf{w}) - \frac{1}{n+1} \right|,$$

in which case, for any $(\lambda, \mathbf{v}, \mathbf{w}) \in [0, 1] \times \mathbb{R}_+^{2n}$ that satisfies $\psi_i(N; \lambda, \mathbf{v}, \mathbf{w}) = \frac{1}{n+1}$ for all $i \in N$, we have $\max_{i \in N} |\psi_i(\{i\}; \lambda, \mathbf{v}, \mathbf{w}) - \frac{1}{1+n}| \geq \frac{1}{4} - \frac{5}{2n}$. We have $\frac{1}{4} - \frac{5}{2n} \geq \frac{1}{8}$ for $n \geq 20$, so the desired result

holds. In the rest of this section, we use a sequence of lemmas to show Theorem A.1. We use a change of variables to give an alternative formulation for problem (6). Consider the problem

$$Y_n^* = \min_{(\lambda, x_0, \mathbf{x}) \in [0, 1] \times \mathbb{R}_+^{n+1}} \left\{ \sum_{i \in N} \left(\frac{\lambda x_i}{x_0 + x_i} + \frac{(1 - \lambda) \left(\frac{1}{n+1} - \lambda x_i \right)}{\left(\frac{1}{n+1} - \lambda x_0 \right) + \left(\frac{1}{n+1} - \lambda x_i \right)} \right) : \right. \\ \left. x_0 + \sum_{i \in N} x_i = 1, \quad x_i \leq \frac{1}{\lambda(n+1)} \quad \forall i \in N \cup \{0\} \right\}, \quad (7)$$

where we use the vector of decision variables $\mathbf{x} = (x_1, \dots, x_n)$. In the next lemma, we show that the optimal objective values of problems (6) and (7) differ by $n/(n+1)$.

Lemma A.2 *Noting that the optimal objective values of problems (6) and (7) are, respectively, Z_n^* and Y_n^* , for all $n \geq 1$, we have $Z_n^* = Y_n^* - \frac{n}{n+1}$.*

Proof. In any feasible solution to problem (6), we have $\psi_i(N; \lambda, \mathbf{v}, \mathbf{w}) = \frac{1}{n+1}$ for all $i \in N$. By the definition of $\psi_i(S; \lambda, \mathbf{v}, \mathbf{w})$, we have $\psi_i(\{i\}; \lambda, \mathbf{v}, \mathbf{w}) \geq \psi_i(N; \lambda, \mathbf{v}, \mathbf{w})$, so we get $\psi_i(\{i\}; \lambda, \mathbf{v}, \mathbf{w}) \geq \frac{1}{n+1}$ in any feasible solution to problem (6). Thus, we can drop the absolute value in the objective function, so the objective function of problem (6) is $\sum_{i \in N} \left| \psi_i(\{i\}; \lambda, \mathbf{v}, \mathbf{w}) - \frac{1}{n+1} \right| = -\frac{n}{n+1} + \sum_{i \in N} \psi_i(\{i\}; \lambda, \mathbf{v}, \mathbf{w}) = -\frac{n}{n+1} + \sum_{i \in N} \left(\lambda \frac{v_i}{1+v_i} + (1-\lambda) \frac{w_i}{1+w_i} \right)$. Also, the constraints in problem (6) are given by $\lambda \frac{v_i}{1+V(N)} + (1-\lambda) \frac{w_i}{1+W(N)} = \frac{1}{n+1}$ for all $i \in N$. Thus, problem (6) is equivalent to

$$Z_n^* = -\frac{n}{n+1} + \min_{(\lambda, \mathbf{v}, \mathbf{w}) \in [0, 1] \times \mathbb{R}_+^{2n}} \left\{ \sum_{i \in N} \left(\lambda \frac{v_i}{1+v_i} + (1-\lambda) \frac{w_i}{1+w_i} \right) : \right. \\ \left. \lambda \frac{v_i}{1+V(N)} + (1-\lambda) \frac{w_i}{1+W(N)} = \frac{1}{n+1} \quad \forall i \in N \right\} \\ \stackrel{(a)}{=} -\frac{n}{n+1} + \min_{(\lambda, x_0, \mathbf{x}, y_0, \mathbf{y}) \in [0, 1]^{2n+3}} \left\{ \sum_{i \in N} \left(\lambda \frac{x_i}{x_0 + x_i} + (1-\lambda) \frac{y_i}{y_0 + y_i} \right) : \right. \\ \left. \lambda x_i + (1-\lambda) y_i = \frac{1}{n+1} \quad \forall i \in N, \quad x_0 + \sum_{i \in N} x_i = 1, \quad y_0 + \sum_{i \in N} y_i = 1 \right\}, \quad (8)$$

where (a) follows by making the change of variables $x_i = v_i/(1+V(N))$ and $y_i = w_i/(1+W(N))$ for all $i \in N$, as well as $x_0 = 1/(1+V(N))$ and $y_0 = 1/(1+W(N))$.

It is enough to show that the optimal objective value of the problem on the right side of (8) is equal to Y_n^* . If we have $\lambda x_i + (1-\lambda) y_i = \frac{1}{n+1}$ for all $i \in N \cup \{0\}$ and $x_0 + \sum_{i \in N} x_i = 1$, then adding the first $n+1$ equalities over all $i \in N \cup \{0\}$, we obtain $\lambda(x_0 + \sum_{i \in N} x_i) + (1-\lambda)(y_0 + \sum_{i \in N} y_i) = 1$, in which case, in view of the equality $x_0 + \sum_{i \in N} x_i = 1$, we get $y_0 + \sum_{i \in N} y_i = 1$. In other words, having $\lambda x_i + (1-\lambda) y_i = \frac{1}{n+1}$ for all $i \in N \cup \{0\}$ and $x_0 + \sum_{i \in N} x_i = 1$ implies that $y_0 + \sum_{i \in N} y_i = 1$. In this case, for the problem on the right side of (8), if we replace the constraint $\lambda x_i + (1-\lambda) y_i = \frac{1}{n+1}$

for all $i \in N$ with $\lambda x_i + (1 - \lambda) y_i = \frac{1}{n+1}$ for all $i \in N \cup \{0\}$, then the constraint $y_0 + \sum_{i \in N} y_i = 1$ becomes redundant. Therefore, the problem on the right side of (8) is equivalent to

$$\begin{aligned} & \min_{(\lambda, x_0, \mathbf{x}, y_0, \mathbf{y}) \in [0, 1]^{2n+3}} \left\{ \sum_{i \in N} \left(\lambda \frac{x_i}{x_0 + x_i} + (1 - \lambda) \frac{y_i}{y_0 + y_i} \right) : \right. \\ & \qquad \left. \lambda x_i + (1 - \lambda) y_i = \frac{1}{n+1} \quad \forall i \in N \cup \{0\}, \quad x_0 + \sum_{i \in N} x_i = 1 \right\} \\ & \stackrel{(b)}{=} \min_{(\lambda, x_0, \mathbf{x}) \in [0, 1]^{n+2}} \left\{ \sum_{i \in N} \left(\frac{\lambda x_i}{x_0 + x_i} + \frac{(1 - \lambda) \left(\frac{1}{n+1} - \lambda x_i \right)}{\left(\frac{1}{n+1} - \lambda x_0 \right) + \left(\frac{1}{n+1} - \lambda x_i \right)} \right) : \right. \\ & \qquad \left. x_i \leq \frac{1}{\lambda(n+1)} \quad \forall i \in N \cup \{0\}, \quad x_0 + \sum_{i \in N} x_i = 1 \right\}, \end{aligned}$$

where (b) holds because the first constraint in the problem on the left side above implies that $y_i = \frac{1}{1-\lambda} \left(\frac{1}{n+1} - \lambda x_i \right)$. By (7), the optimal objective value of the last problem above is Y_n^* . ■

For fixed $(\lambda, x_0) \in [0, 1] \times \mathbb{R}_+$, we consider the values of $\mathbf{x} \in \mathbb{R}_+^n$ that are feasible to problem (7), which are given by the polytope

$$\mathcal{Q}(\lambda, x_0) = \left\{ \mathbf{x} \in \mathbb{R}_+^n : \sum_{i \in N} x_i = 1 - x_0, \quad 0 \leq x_i \leq \frac{1}{\lambda(n+1)} \quad \forall i \in N \right\}.$$

Note that $\mathcal{Q}(\lambda, x_0)$ is the knapsack polytope, so if $\hat{\mathbf{x}} \in \mathbb{R}_+^n$ is an extreme point of $\mathcal{Q}(\lambda, x_0)$, then $0 < \hat{x}_i < \frac{1}{\lambda(n+1)}$ for at most one $i \in N$. For all other $j \in N \setminus \{i\}$, we have $\hat{x}_j = 0$ or $\hat{x}_j = \frac{1}{\lambda(n+1)}$.

Using the polytope above, $(\lambda, x_0, \mathbf{x})$ is feasible to problem (7) if and only if $(\lambda, x_0) \in [0, 1]^2$, $x_0 \leq \frac{1}{\lambda(n+1)}$ and $\mathbf{x} \in \mathcal{Q}(\lambda, x_0)$. Also, to capture the objective function of problem (7), let

$$F(\lambda, x_0, x_i) = \frac{\lambda x_i}{x_0 + x_i} + \frac{(1 - \lambda) \left(\frac{1}{n+1} - \lambda x_i \right)}{\left(\frac{1}{n+1} - \lambda x_0 \right) + \left(\frac{1}{n+1} - \lambda x_i \right)},$$

so the objective function of problem (7) is $\sum_{i \in N} F(\lambda, x_0, x_i)$. In the next lemma, we give a lower bound on the optimal objective value of the problem $\min_{\mathbf{x} \in \mathcal{Q}(\lambda, x_0)} \sum_{i \in N} F(\lambda, x_0, x_i)$.

Lemma A.3 *For any fixed $(\lambda, x_0) \in [0, 1]^2$ such that $0 \leq x_0 \leq \frac{1}{\lambda(n+1)}$, we have the lower bound given by*

$$\min_{\mathbf{x} \in \mathcal{Q}(\lambda, x_0)} \left\{ \sum_{i \in N} F(\lambda, x_0, x_i) \right\} \geq -\frac{3}{2} + \frac{n+1}{2} \left[\lambda^2 (1 - x_0) + (1 - \lambda) (1 - \lambda(1 - x_0)) \right]. \quad (9)$$

Proof. For fixed $(\lambda, x_0) \in [0, 1]^2$ such that $0 \leq x_0 \leq \frac{1}{\lambda(n+1)}$, let \mathbf{x}^* be an optimal solution to the minimization problem in (9). Note that $F(\lambda, x_0, x_i)$ is a concave function of x_i , so there exists an optimal solution to the minimization problem in (9) that is an extreme point of $\mathcal{Q}(\lambda, x_0)$. We

proceed under the assumption that \mathbf{x}^* is an extreme point of $\mathcal{Q}(\lambda, x_0)$. In this case, by the discussion right after the definition of $\mathcal{Q}(\lambda, x_0)$, there exists at most one coordinate $i \in N$ of \mathbf{x}^* such that $0 < x_i^* < \frac{1}{\lambda(n+1)}$. All other coordinates of \mathbf{x}^* take the value zero or $\frac{1}{\lambda(n+1)}$. Let H be the number of coordinates of \mathbf{x}^* that take the value $\frac{1}{\lambda(n+1)}$. Thus, the number of coordinates that take the value zero is at least $n - H - 1$. Since $\mathbf{x}^* \in \mathcal{Q}(\lambda, x_0)$, we have $x_0 + \sum_{i \in N} x_i^* = 1$, which implies that $x_0 + H \frac{1}{\lambda(n+1)} \leq 1 \leq x_0 + (H + 1) \frac{1}{\lambda(n+1)}$, where the first inequality holds since there are at least H strictly positive coordinates of \mathbf{x}^* that take the value $\frac{1}{\lambda(n+1)}$, whereas the second inequality holds since there are at most $H + 1$ strictly positive components of \mathbf{x}^* and each of these components takes at most the value $\frac{1}{\lambda(n+1)}$. We write the last chain of inequalities as $(1 - x_0) \lambda(n + 1) - 1 \leq H \leq (1 - x_0) \lambda(n + 1)$. Letting $B = (1 - x_0) \lambda(n + 1)$ for notational brevity, we get $B - 1 \leq H \leq B$. So, the optimal objective value of the minimization problem in (9) satisfies

$$\begin{aligned} \min_{\mathbf{x} \in \mathcal{Q}(\lambda, x_0)} \left\{ \sum_{i \in N} F(\lambda, x_0, x_i) \right\} &\stackrel{(a)}{\geq} \min_{\substack{H \in \{0, \dots, n-1\}: \\ B-1 \leq H \leq B}} \left\{ H \times F\left(\lambda, x_0, \frac{1}{\lambda(n+1)}\right) + (n - H - 1) \times F(\lambda, x_0, 0) \right\} \\ &\stackrel{(b)}{\geq} \min_{\substack{H \in [0, n-1]: \\ B-1 \leq H \leq B}} \left\{ H \frac{\lambda}{\lambda x_0(n+1) + 1} + (n - H - 1) \frac{1 - \lambda}{2 - \lambda x_0(n+1)} \right\}, \quad (10) \end{aligned}$$

where (a) holds by dropping one component that is between zero and $\frac{1}{\lambda(n+1)}$, whereas (b) holds since $F(\lambda, x_0, \frac{1}{\lambda(n+1)}) = \frac{\lambda}{\lambda x_0(n+1) + 1}$ and $F(\lambda, x_0, 0) = \frac{1 - \lambda}{2 - \lambda x_0(n+1)}$ by the definition of $F(\lambda, x_0, x_i)$.

Note that we have the identity $\min_{B-1 \leq x \leq B} ax = \min\{aB, a(B-1)\} = aB - \max\{a, 0\}$. Also, using the fact that $0 \leq x_0 \leq \frac{1}{\lambda(n+1)}$, we obtain

$$\begin{aligned} &\min_{\substack{H \in [0, n-1]: \\ B-1 \leq H \leq B}} \left\{ H \frac{\lambda}{\lambda x_0(n+1) + 1} + (n - H - 1) \frac{1 - \lambda}{2 - \lambda x_0(n+1)} \right\} \\ &\stackrel{(c)}{\geq} \min_{\substack{H \in [0, n-1]: \\ B-1 \leq H \leq B}} \left\{ H \frac{\lambda}{2} + (n - H - 1) \frac{1 - \lambda}{2} \right\} \stackrel{(d)}{\geq} \min_{B-1 \leq H \leq B} \left\{ \left(\lambda - \frac{1}{2}\right) H \right\} + (n - 1) \frac{1 - \lambda}{2} \\ &\stackrel{(e)}{=} \left(\lambda - \frac{1}{2}\right) B - \max\left\{\lambda - \frac{1}{2}, 0\right\} + (n - 1) \frac{1 - \lambda}{2} \stackrel{(f)}{\geq} \left(\lambda - \frac{1}{2}\right) (1 - x_0) \lambda(n + 1) - \frac{1}{2} + (n - 1) \frac{1 - \lambda}{2} \\ &\stackrel{(g)}{=} \frac{n + 1}{2} \left[\lambda^2 (1 - x_0) + (1 - \lambda) (1 - \lambda (1 - x_0)) \right] - \frac{3}{2} + \lambda, \quad (11) \end{aligned}$$

where (c) holds since $\lambda x_0(n + 1) \leq 1$, (d) holds by arranging the terms and dropping the constraint $H \in [0, n - 1]$, (e) holds by the identity at the beginning of this paragraph, (f) holds by using the definition of B and noting that $\lambda - \frac{1}{2} \leq \frac{1}{2}$ and (g) follows by arranging the terms. Combining (10) and (11), noting that $\lambda \geq 0$, the result follows. \blacksquare

Proof of Theorem A.1. In the rest of this section, we use Lemmas A.2 and A.3 to give a proof for Theorem A.1. By the discussion right after the definition of the polytope $\mathcal{Q}(\lambda, x_0)$, $(\lambda, x_0, \mathbf{x})$ is

feasible to problem (7) if and only if $(\lambda, x_0) \in [0, 1]^2$, $x_0 \leq \frac{1}{\lambda(n+1)}$ and $\mathbf{x} \in \mathcal{Q}(\lambda, x_0)$. Thus, since the objective function of this problem is $\sum_{i \in N} F(\lambda, x_0, x_i)$, problem (7) is equivalent to

$$\begin{aligned} Y_n^* &\stackrel{(a)}{=} \min_{\substack{(\lambda, x_0) \in [0, 1]^2 : \\ x_0 \leq \frac{1}{\lambda(n+1)}}} \left\{ \min_{\mathbf{x} \in \mathcal{Q}(\lambda, x_0)} \left\{ \sum_{i \in N} F(\lambda, x_0, x_i) \right\} \right\} \\ &\stackrel{(b)}{\geq} -\frac{3}{2} + \frac{n+1}{2} \min_{(\lambda, x_0) \in [0, 1]^2} \left\{ \lambda^2(1-x_0) + (1-\lambda)(1-\lambda(1-x_0)) : x_0 \leq \frac{1}{\lambda(n+1)} \right\}, \end{aligned} \quad (12)$$

where (a) follows by sequentially optimizing over the decision variables (λ, x_0) and \mathbf{x} instead of simultaneously optimizing over the decision variables $(\lambda, x_0, \mathbf{x})$ and (b) uses Lemma A.3.

In the last minimization problem above, the constraint implies that $\lambda x_0 \leq \frac{1}{n+1}$. Checking all possible values of (λ, x_0) that satisfies $\lambda x_0 \leq \frac{1}{n+1}$, this problem is equivalent to

$$\begin{aligned} W_n^* &= \min_{(\lambda, x_0) \in [0, 1]^2} \left\{ \lambda^2(1-x_0) + (1-\lambda)(1-\lambda(1-x_0)) : x_0 \leq \frac{1}{\lambda(n+1)} \right\} \\ &= \min_{c \in [0, \frac{1}{n+1}]} \left\{ \min_{(\lambda, x_0) \in [0, 1]^2} \left\{ \lambda^2(1-x_0) + (1-\lambda)(1-\lambda(1-x_0)) : \lambda x_0 = c \right\} \right\} \\ &\stackrel{(c)}{=} \min_{c \in [0, \frac{1}{n+1}]} \left\{ \min_{\lambda \in [c, 1]} \left\{ \lambda(\lambda - c) + (1-\lambda)(1-\lambda + c) \right\} \right\}, \end{aligned} \quad (13)$$

where c holds by using the constraint $\lambda x_0 = c$ to replace x_0 with c/λ , in which case, the constraints $\lambda x_0 = c$ and $(\lambda, x_0) \in [0, 1]^2$ imply that $\lambda \in [c, 1]$.

The objective function of the inner minimization problem on the right side of (13) is convex and quadratic in λ . Checking the first order condition of the objective function, the unconstrained minimum occurs at $\lambda = \frac{c+1}{2}$, which satisfies the constraint $\lambda \in [c, 1]$. In this case, evaluating the objective function at $\lambda = \frac{c+1}{2}$, the optimal objective value of the inner minimization problem on the right side of (13) is given by $\frac{1-c^2}{2}$. Using this optimal objective value in (13), we obtain $W_n^* = \min_{c \in [0, \frac{1}{n+1}]} \frac{1-c^2}{2} = \frac{1}{2} - \frac{1}{2(n+1)^2}$, where the last equality holds since $\frac{1-c^2}{2}$ is decreasing in c over the interval $[0, \frac{1}{n+1}]$. Therefore, the optimal objective value of the minimization problem on the right side of (12) is $\frac{1}{2} - \frac{1}{2(n+1)^2}$, so $Y_n^* \geq -\frac{3}{2} + \frac{n+1}{2} \left[\frac{1}{2} - \frac{1}{2(n+1)^2} \right]$. Thus, by Lemma A.2, we get

$$\begin{aligned} Z_n^* &= Y_n^* - \frac{n}{n+1} \geq -\frac{3}{2} + \frac{n+1}{2} \left[\frac{1}{2} - \frac{1}{2(n+1)^2} \right] - \frac{n}{n+1} \\ &= -\frac{3}{2} + \frac{n}{4} + \frac{1}{4} - \frac{1}{4(n+1)} - \frac{n}{n+1} \geq \frac{n}{4} - \frac{5}{2}, \end{aligned}$$

where the last inequality uses the fact that $\frac{1}{4} - \frac{1}{4(n+1)} \geq 0$ and $\frac{n}{n+1} \leq 1$. The chain of inequalities above establishes the desired result. ■

Appendix B: Appendix: Red Bus-Blue Bus Paradox

A commuter has the option of using a red bus or a car to get to her destination on each day. She chooses each of the two options with an equality probability of $1/2$. Using 1 and 2 to capture the red bus and car options, let $\phi_i^{\text{MNL}}(S)$ be the choice probability of option i under the multinomial logit model given that the available options are S . Thus, we have $\frac{\phi_1^{\text{MNL}}(\{1,2\})}{\phi_2^{\text{MNL}}(\{1,2\})} = 1$. Consider introducing a third option of a blue bus. McFadden (1980) argues that the two busses are identical to each other and very different from the car for commuting purposes, so the commuter should choose the two buses with equal probability and the introduction of the blue bus should not change the choice probability of the car. Thus, the commuter should still choose the car with probability $1/2$ and the remaining probability should be split equally for the two buses, yielding a choice probability of $1/4$ for each bus. However, due to its independence of irrelevant alternatives property, if we use the multinomial logit model to capture the choice process of the commuter, then such choice probabilities cannot be attained. In particular, using 3 to capture the blue bus option, after introducing the blue bus, we want the choice probabilities of the two buses to be equal to each other, so $\frac{\phi_1^{\text{MNL}}(\{1,2,3\})}{\phi_3^{\text{MNL}}(\{1,2,3\})} = 1$. Furthermore, by the independence of irrelevant alternatives property, we have $\frac{\phi_1^{\text{MNL}}(\{1,2,3\})}{\phi_2^{\text{MNL}}(\{1,2,3\})} = \frac{\phi_1^{\text{MNL}}(\{1,2\})}{\phi_2^{\text{MNL}}(\{1,2\})}$, but noting that we have $\frac{\phi_1^{\text{MNL}}(\{1,2\})}{\phi_2^{\text{MNL}}(\{1,2\})} = 1$ before the introduction of the blue bus, we must have $\frac{\phi_1^{\text{MNL}}(\{1,2,3\})}{\phi_2^{\text{MNL}}(\{1,2,3\})} = 1$. Thus, we must have $\phi_1^{\text{MNL}}(\{1,2,3\}) = \phi_2^{\text{MNL}}(\{1,2,3\}) = \phi_3^{\text{MNL}}(\{1,2,3\})$, in which case, we get $\phi_1^{\text{MNL}}(\{1,2,3\}) = \phi_2^{\text{MNL}}(\{1,2,3\}) = \phi_3^{\text{MNL}}(\{1,2,3\}) = 1/3$. Thus, after the introduction of the blue bus, all options end up being chosen with an equal probability of $1/3$. In contrast, we can calibrate our mixture model to produce the anticipated choice probabilities before and after the introduction of the blue bus.

We set the parameters of our mixture model as $\theta_1 = 0$, $\theta_2 = 1$, $\theta_3 = 0$, $v_1 = M$, $v_2 = 0$, $v_3 = M$ and $\lambda = 1/2$. Let $\phi_i^{\text{Mix}}(S)$ be the choice probability of option i under our mixture model given that the available options are S . Considering the choice probabilities of the red bus and car before the introduction of the blue bus, under the parameters at the beginning of this paragraph, we have $\phi_1^{\text{Mix}}(\{1,2\}) = \frac{1}{2} \frac{M}{1+M}$ and $\phi_2^{\text{Mix}}(\{1,2\}) = \frac{1}{2}$, so the choice probabilities of the red bus and car get arbitrarily close to $1/2$ as $M \rightarrow \infty$. Thus, the choice probabilities from our mixture model match those anticipated before the introduction of the blue bus. Considering the choice probabilities of the two buses and car after the introduction of the blue bus, we have $\phi_1^{\text{Mix}}(\{1,2,3\}) = \frac{1}{2} \frac{M}{1+2M}$, $\phi_2^{\text{Mix}}(\{1,2,3\}) = \frac{1}{2}$ and $\phi_3^{\text{Mix}}(\{1,2,3\}) = \frac{1}{2} \frac{M}{1+2M}$, so the choice probabilities of the red and blue buses get arbitrarily close to $1/4$ and the choice probability of the car gets arbitrarily close to $1/2$ as $M \rightarrow \infty$. Thus, the choice probabilities from our mixture model match those anticipated after the introduction of the blue bus. Thus, our mixture choice model does not display the red bus-blue bus paradox in its classical form, but this discussion should not mean that our choice model is uniformly superior to the multinomial logit model or always resolves similar paradoxes.

Appendix C: Comparison with the Markov Chain Choice Model

We give an example to show that our mixture of independent demand and multinomial logit 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

$$P_i(S) = \gamma_i + \sum_{j \notin S} \rho_{ji} P_j(S) \quad \forall i \in N. \quad (14)$$

We can intuitively justify (14) 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 independent demand and multinomial logit models with $N = \{1, 2, 3\}$, $(\theta_1, \theta_2, \theta_3) = (0, 0, 1)$, $(v_1, v_2, v_3) = (1, 1, 1)$ and $\lambda = \frac{1}{4}$. Under this choice model, if we offer the assortment S , then a customer purchases product $i \in S$ with probability $\phi_i(S) = \lambda \theta_i + (1 - \lambda) \frac{v_i}{1 + V(S)}$. 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 independent demand and multinomial logit 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 independent demand and multinomial logit 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

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

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

all $S \subseteq N$, $i \in N$. To make this argument, by (14), note that $P_i(\{1,2,3\}) = \gamma_i$ for all $i \in N$. Therefore, 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 $\gamma_3 = \phi_3(\{1,2,3\}) = \frac{7}{16}$. By this reasoning, we fix 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 (14), 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 $\gamma_3 = \frac{7}{16}$, 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 (14), 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

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

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 (14), 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 $P_1(\{1\}) = \frac{7}{20}$, 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 D: 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 $M \neq \emptyset$. We construct two distinct feasible solutions $(\tilde{x}_0, \tilde{\mathbf{x}}, \tilde{\mathbf{y}})$ and $(\bar{x}_0, \bar{\mathbf{x}}, \bar{\mathbf{y}})$ to the Assortment LP such that $(\hat{x}_0, \hat{\mathbf{x}}, \hat{\mathbf{y}}) = \frac{1}{2}(\tilde{x}_0, \tilde{\mathbf{x}}, \tilde{\mathbf{y}}) + \frac{1}{2}(\bar{x}_0, \bar{\mathbf{x}}, \bar{\mathbf{y}})$, contradicting the fact that $(\hat{x}_0, \hat{\mathbf{x}}, \hat{\mathbf{y}})$ is a basic feasible solution. For small $\epsilon > 0$, we define the solution $(\tilde{x}_0, \tilde{\mathbf{x}}, \tilde{\mathbf{y}})$ as

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

We claim that $(\tilde{x}_0, \tilde{\mathbf{x}}, \tilde{\mathbf{y}})$ 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, $(\tilde{x}_0, \tilde{\mathbf{x}}, \tilde{\mathbf{y}})$ satisfies the first constraint. Noting that $M \neq \emptyset$, 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, $(\tilde{x}_0, \tilde{\mathbf{x}}, \tilde{\mathbf{y}})$ 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 $\tilde{y}_{ij} \leq \tilde{x}_i$ and $\tilde{y}_{ij} \leq \tilde{x}_j$. If, on the other hand, $\hat{y}_{ij} < \min\{\hat{x}_i, \hat{x}_j\}$, then $\tilde{y}_{ij} = \hat{y}_{ij} < \min\{\hat{x}_i, \hat{x}_j\} - V(M)\epsilon$ for small $\epsilon > 0$. Noting that $\tilde{x}_i \geq \hat{x}_i - V(M)\epsilon$ for all $i \in N$, we get $\tilde{y}_{ij} = \hat{y}_{ij} < \min\{\hat{x}_i, \hat{x}_j\} - V(M)\epsilon \leq \min\{\tilde{x}_i, \tilde{x}_j\}$, so $\tilde{y}_{ij} \leq \tilde{x}_i$ and $\tilde{y}_{ij} \leq \tilde{x}_j$. Thus, $(\tilde{x}_0, \tilde{\mathbf{x}}, \tilde{\mathbf{y}})$ satisfies the third and fourth constraints. Also, we have $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) \in \mathbb{R}_+^{n+n^2}$ for small $\epsilon > 0$, establishing the claim. Similarly, we define the solution $(\bar{x}_0, \bar{\mathbf{x}}, \bar{\mathbf{y}})$ as

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

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

If we have $\hat{y}_{ij} < \min\{\hat{x}_i, \hat{x}_j\}$, then $\tilde{y}_{ij} = \hat{y}_{ij} = \bar{y}_{ij}$, so $\hat{y}_{ij} = \frac{1}{2}\tilde{y}_{ij} + \frac{1}{2}\bar{y}_{ij}$, 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 $\bar{y}_{ij} = \hat{x}_0 + V(M)\epsilon$ as well. In this case, we get $\hat{y}_{ij} = \frac{1}{2}\tilde{y}_{ij} + \frac{1}{2}\bar{y}_{ij}$.

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 $\bar{y}_{ij} = \hat{x}_j - (1 + V(H))\epsilon$ as well. Thus, $\hat{y}_{ij} = \frac{1}{2}\tilde{y}_{ij} + \frac{1}{2}\bar{y}_{ij}$.

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 $\hat{y}_{ij} = \frac{1}{2}\tilde{y}_{ij} + \frac{1}{2}\bar{y}_{ij}$.

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 $\hat{y}_{ij} = \min\{\hat{x}_i, \hat{x}_j\} \leq \hat{x}_\ell = 0$. Furthermore, for $\ell \in L$, we have $\tilde{x}_\ell = \hat{x}_\ell = 0$, in which case, we get $\tilde{y}_{ij} = \min\{\tilde{x}_i, \tilde{x}_j\} \leq \tilde{x}_\ell = 0$. By the symmetric reasoning, we have $\bar{y}_{ij} = 0$ as well. In this case, it follows that $\hat{y}_{ij} = 0 = \frac{1}{2}\tilde{y}_{ij} + \frac{1}{2}\bar{y}_{ij}$. ■

Appendix E: Proof of Theorem 4.1

We give a proof for the second part of Theorem 4.1. Let $\delta > 0$ be such that $S^*(\lambda, x) = S^*(\lambda, 0)$ for all $x \in [0, \delta)$ and $S^*(\lambda, \delta) \neq S^*(\lambda, 0)$. 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 product revenues starting from these nominal values. The discussion so far in this paragraph fixes the value of δ , so that we have $S^*(\lambda, \delta) \neq S^*(\lambda, 0)$ and $S^*(\lambda, x) = S^*(\lambda, 0)$ for all $x \in [0, \delta)$. On the other hand, using $R(S)$ as defined right before Theorem 3.3, 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 $|R(S) + \delta \frac{V(S)}{1+V(S)} - R(Q) - \delta \frac{V(Q)}{1+V(Q)}| > \epsilon$. That is, if the quantities $R(S) + \delta \frac{V(S)}{1+V(S)}$ and $R(Q) + \delta \frac{V(Q)}{1+V(Q)}$ are different from each other, then they must differ by at least $\epsilon > 0$. For $V_{\min} = \min_{i \in N} v_i$, fix some $\gamma \in [0, \delta)$ such that

$$0 < \left(1 + \frac{\lambda}{1-\lambda} \frac{1+V(N)}{V_{\min}} \sum_{i \in N} \theta_i\right) (\delta - \gamma) \leq \epsilon. \quad (15)$$

The discussion in the latter part of this paragraph fixes the value of γ . Since $S^*(\lambda, x) = S^*(\lambda, 0)$ for all $x \in [0, \delta)$, noting that $\gamma \in [0, \delta)$, we get $S^*(\lambda, \gamma) = S^*(\lambda, 0)$.

In the proof, for the values of δ and γ that we fix in the previous paragraph, we will show that $S^*(\lambda, \delta) \supseteq S^*(\lambda, \gamma)$. In this case, since $S^*(\lambda, \gamma) = S^*(\lambda, 0)$, we get $S^*(\lambda, \delta) \supseteq S^*(\lambda, 0)$, which is the

desired result. Throughout the proof, we let $R_x(S) = \frac{\sum_{i \in S} (r_i + x) v_i}{1 + V(S)}$, $\Theta_x(S) = \sum_{i \in S} (r_i + x) \theta_i$ and $A_x(S) = \frac{\sum_{i \in S} (r_i + x) v_i}{V(S)}$. In this case, letting $S_\delta = S^*(\lambda, \delta)$ and $S_\gamma = S^*(\lambda, \gamma)$ for notational brevity, S_δ is an optimal solution to the problem $\max_{S \subseteq N} \lambda \Theta_\delta(S) + (1 - \lambda) R_\delta(S)$, whereas S_γ is an optimal solution to the problem $\max_{S \subseteq N} \lambda \Theta_\gamma(S) + (1 - \lambda) R_\gamma(S)$.

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

Case 1: Assume that $R_\delta(S_\delta) - \delta \geq R_\gamma(S_\gamma) - \gamma$. Let $K = S_\delta \setminus S_\gamma$. Thus, we have $K \subseteq S_\delta$ and $S_\gamma \cap K = \emptyset$. First, we proceed under the assumption that $K \neq \emptyset$. By definition of S_γ , we have $\lambda \Theta_\gamma(S_\gamma) + (1 - \lambda) R_\gamma(S_\gamma) \geq \lambda \Theta_\gamma(S_\gamma \cup K) + (1 - \lambda) R_\gamma(S_\gamma \cup K)$, which we equivalently write as $(1 - \lambda) (R_\gamma(S_\gamma) - R_\gamma(S_\gamma \cup K)) \geq \lambda \sum_{i \in K} (r_i + \gamma) \theta_i \geq 0$, so we get $R_\gamma(S_\gamma) \geq R_\gamma(S_\gamma \cup K)$. By the same argument at the beginning of the proof of Lemma 3.4, $R_\gamma(S_\gamma \cup K)$ is a convex combination of $R_\gamma(S_\gamma)$ and $A_\gamma(K)$, so since $R_\gamma(S_\gamma) \geq R_\gamma(S_\gamma \cup K)$, we must have $A_\gamma(K) \leq R_\gamma(S_\gamma \cup K) \leq R_\gamma(S_\gamma)$. Also, by the definition of $A_x(S)$, we have $A_x(S) = A(S) + x$ for $S \neq \emptyset$. By the assumption in Case 1, we have $R_\delta(S_\delta) - \delta \geq R_\gamma(S_\gamma) - \gamma$. Thus, using the chain of inequalities $A_\gamma(K) \leq R_\gamma(S_\gamma \cup K) \leq R_\gamma(S_\gamma)$ just established in this paragraph, it follows that $R_\delta(S_\delta) - \delta \geq R_\gamma(S_\gamma) - \gamma \geq A_\gamma(K) - \gamma = A(K)$, which yields $R_\delta(S_\delta) \geq A(K) + \delta = A_\delta(K)$. In this case, using the fact that $R_\delta(S_\delta)$ can be expressed as a convex combination of $R_\delta(S_\delta \setminus K)$ and $A_\delta(K)$, having $R_\delta(S_\delta) \geq A_\delta(K)$ implies that the chain of inequalities $R_\delta(S_\delta \setminus K) \geq R_\delta(S_\delta) \geq A_\delta(K)$ must hold.

Noting that S_δ is an optimal solution to the problem $\max_{S \subseteq N} \lambda \Theta_\delta(S) + (1 - \lambda) R_\delta(S)$ and S_γ is an optimal solution to the problem $\max_{S \subseteq N} \lambda \Theta_\gamma(S) + (1 - \lambda) R_\gamma(S)$, by Lemma 3.4, we have

$$\lambda \Theta_\gamma(K) \leq (1 - \lambda) V(K) \frac{R_\gamma(S_\gamma) - A_\gamma(K)}{1 + V(S_\gamma \cup K)} \quad \text{and} \quad \lambda \Theta_\delta(K) \geq (1 - \lambda) V(K) \frac{R_\delta(S_\delta \setminus K) - A_\delta(K)}{1 + V(S_\delta)}$$

To use Lemma 3.4 to get the two inequalities above, we need to have $S_\gamma \cap K = \emptyset$ and $K \subseteq S_\delta$, but these two conditions hold by the fact that $K = S_\delta \setminus S_\gamma$.

By the definition of $\Theta_x(S)$, we have $\Theta_\delta(K) - \Theta_\gamma(K) = (\delta - \gamma) \sum_{i \in K} \theta_i$. Adding the two inequalities above, arranging the terms and using the fact that $A_x(S) = A(S) + x$, we get

$$\begin{aligned} R_\gamma(S_\gamma) - A(K) - \gamma &\geq \frac{1 + V(S_\gamma \cup K)}{1 + V(S_\delta)} \left(R_\delta(S_\delta \setminus K) - A(K) - \delta \right) - \frac{\lambda}{1 - \lambda} \frac{1 + V(S_\gamma \cup K)}{V(K)} (\delta - \gamma) \sum_{i \in K} \theta_i \\ &\stackrel{(a)}{\geq} R_\delta(S_\delta) - A(K) - \delta - \frac{\lambda}{1 - \lambda} \frac{1 + V(N)}{V_{\min}} (\delta - \gamma) \sum_{i \in N} \theta_i \\ &\stackrel{(b)}{\geq} R_\delta(S_\delta) - A(K) - \delta - (\epsilon - (\delta - \gamma)), \end{aligned} \tag{16}$$

where (a) holds since $S_\gamma \cup K \supseteq S_\delta$ by the definition of K , so $\frac{1 + V(S_\gamma \cup K)}{1 + V(S_\delta)} \geq 1$, as well as using the inequality $R_\delta(S_\delta \setminus K) \geq R_\delta(S_\delta) \geq A_\delta(K)$ established earlier in the proof, whereas

(b) holds by (15). So, (16) yields $R_\gamma(S_\gamma) \geq R_\delta(S_\delta) - \epsilon$. By the definition of $R_x(S)$, for $x \geq y$, we have $R_x(S) \geq R_y(S)$. Thus, having $R_\gamma(S_\gamma) \geq R_\delta(S_\delta) - \epsilon$ implies that $R_\delta(S_\gamma) \geq R_\delta(S_\delta) - \epsilon$. Also, noting the definition of $R_x(S)$, we have $R_x(S) = R(S) + x \frac{V(S)}{1+V(S)}$. In this case, we get $R_\gamma(S_\gamma) + \delta - \gamma = R(S_\gamma) - \gamma \frac{1}{1+V(S_\gamma)} + \delta = R_\delta(S_\gamma) + (\delta - \gamma) \frac{1}{1+V(S_\gamma)} > R_\delta(S_\gamma)$. Since $R_\delta(S_\delta) \geq R_\gamma(S_\gamma) + \delta - \gamma$ by the assumption in Case 1, it follows that $R_\delta(S_\delta) \geq R_\gamma(S_\gamma) + \delta - \gamma > R_\delta(S_\gamma)$. By the discussion so far in this paragraph, we have $R_\delta(S_\gamma) \geq R_\delta(S_\delta) - \epsilon$ and $R_\delta(S_\delta) > R_\delta(S_\gamma)$, which imply that $|R_\delta(S_\delta) - R_\delta(S_\gamma)| \leq \epsilon$ and $R_\delta(S_\delta) \neq R_\delta(S_\gamma)$. So, the assortments $S_\gamma, S_\delta \subseteq N$ satisfy $|R(S_\delta) + \delta \frac{V(S_\delta)}{1+V(S_\delta)} - R(S_\gamma) - \delta \frac{V(S_\gamma)}{1+V(S_\gamma)}| \leq \epsilon$, while $R(S_\delta) + \delta \frac{V(S_\delta)}{1+V(S_\delta)}$ and $R(S_\gamma) + \delta \frac{V(S_\gamma)}{1+V(S_\gamma)}$ being distinct from each other, which contradicts the definition of ϵ at the beginning of the proof. Therefore, it follows that we cannot have $K \neq \emptyset$ in Case 1.

Second, we proceed under the assumption that $K = \emptyset$. Since $K = S_\delta \setminus S_\gamma$, we get $S_\gamma \supseteq S_\delta$. By our choice of δ and γ at the beginning of the proof, $S_\delta \neq S^*(\lambda, 0)$, but $S_\gamma = S^*(\lambda, 0)$, so we cannot have $S_\delta = S_\gamma$. Therefore, S_δ is a strict subset of S_γ . By the definitions of S_δ and S_γ , we have the inequalities $\lambda \Theta_\delta(S_\delta) + (1 - \lambda) R_\delta(S_\delta) \geq \lambda \Theta_\delta(S_\gamma) + (1 - \lambda) R_\delta(S_\gamma)$ and $\lambda \Theta_\gamma(S_\gamma) + (1 - \lambda) R_\gamma(S_\gamma) \geq \lambda \Theta_\gamma(S_\delta) + (1 - \lambda) R_\gamma(S_\delta)$. Adding the two inequalities, we get

$$\lambda \left[\Theta_\delta(S_\delta) - \Theta_\gamma(S_\delta) \right] + (1 - \lambda) \left[R_\delta(S_\delta) - R_\gamma(S_\delta) \right] \geq \lambda \left[\Theta_\delta(S_\gamma) - \Theta_\gamma(S_\gamma) \right] + (1 - \lambda) \left[R_\delta(S_\gamma) - R_\gamma(S_\gamma) \right].$$

Since S_δ is a strict subset of S_γ , we have $\sum_{i \in S_\delta} \theta_i \leq \sum_{i \in S_\gamma} \theta_i$ and $\frac{V(S_\delta)}{1+V(S_\delta)} \leq \frac{V(S_\gamma)}{1+V(S_\gamma)}$, where at least one of the inequalities is strict. We have $\Theta_x(S) - \Theta_y(S) = (x - y) \sum_{i \in S} \theta_i$ and $R_x(S) - R_y(S) = (x - y) \frac{V(S)}{1+V(S)}$, so we write the inequality above as $(\delta - \gamma) \left(\lambda \sum_{i \in S_\delta} \theta_i + (1 - \lambda) \frac{V(S_\delta)}{1+V(S_\delta)} \right) \geq (\delta - \gamma) \left(\lambda \sum_{i \in S_\gamma} \theta_i + (1 - \lambda) \frac{V(S_\gamma)}{1+V(S_\gamma)} \right)$, contradicting $\sum_{i \in S_\delta} \theta_i \leq \sum_{i \in S_\gamma} \theta_i$ and $\frac{V(S_\delta)}{1+V(S_\delta)} \leq \frac{V(S_\gamma)}{1+V(S_\gamma)}$ with one inequality being strict. So, we cannot have $K = \emptyset$ in Case 1 either.

Case 2: Assume that $R_\delta(S_\delta) - \delta < R_\gamma(S_\gamma) - \gamma$. Let $K = S_\gamma \setminus S_\delta$. If $K = \emptyset$, then we get $S_\gamma \subseteq S_\delta$, which is the desired result. To get a contradiction, assume that $K \neq \emptyset$. Since $K = S_\gamma \setminus S_\delta$, we have $K \subseteq S_\gamma$ and $K \cap S_\delta = \emptyset$. We claim that $R_\delta(S_\delta \cup K) < R_\delta(S_\delta)$. If the claim did not hold, then we would get $\lambda \Theta_\delta(S_\delta \cup K) + (1 - \lambda) R_\delta(S_\delta \cup K) \geq \lambda \Theta_\delta(S_\delta) + (1 - \lambda) R_\delta(S_\delta)$, which cannot hold since S_δ is an optimal solution to the problem $\max_{S \subseteq N} \lambda \Theta_\delta(S) + (1 - \lambda) R_\delta(S)$ with the largest cardinality. Thus, the claim holds and we have $R_\delta(S_\delta \cup K) < R_\delta(S_\delta)$. By the same argument at the beginning of the proof of Lemma 3.4, $R_\delta(S_\delta \cup K)$ is a convex combination of $R_\delta(S_\delta)$ and $A_\delta(K)$, so since $R_\delta(S_\delta \cup K) < R_\delta(S_\delta)$, we get $A_\delta(K) \leq R_\delta(S_\delta \cup K) < R_\delta(S_\delta)$. Noting that $A_x(S) = A(S) + x$ for $S \neq \emptyset$, subtracting δ from the last chain of inequalities yields $A(K) = A_\delta(K) - \delta < R_\delta(S_\delta) - \delta$, but since $R_\delta(S_\delta) - \delta < R_\gamma(S_\gamma) - \gamma$ by the assumption in Case 2, we have $A(K) < R_\gamma(S_\gamma) - \gamma$.

The inequality at the end of the previous paragraph yields $A_\gamma(K) = A(K) + \gamma < R_\gamma(S_\gamma)$. Also, $R_\gamma(S_\gamma)$ is a convex combination of $R_\gamma(S_\gamma \setminus K)$ and $A_\gamma(K)$, in which case, having $A_\gamma(K) < R_\gamma(S_\gamma)$

implies that $A_\gamma(K) < R_\gamma(S_\gamma) \leq R_\gamma(S_\gamma \setminus K)$. Noting that $K \cap S_\delta = \emptyset$ and $K \subseteq S_\gamma$ by the definition of K , Lemma 3.4 implies that

$$\lambda \Theta_\delta(K) \leq (1 - \lambda) V(K) \frac{R_\delta(S_\delta) - A_\delta(K)}{1 + V(S_\delta \cup K)} \quad \text{and} \quad \lambda \Theta_\gamma(K) \geq (1 - \lambda) V(K) \frac{R_\gamma(S_\gamma \setminus K) - A_\gamma(K)}{1 + V(S_\gamma)}.$$

Since $\delta > \gamma$, we have $\Theta_\delta(K) \geq \Theta_\gamma(K)$, in which case, the two inequalities imply that $\frac{R_\delta(S_\delta) - A_\delta(K)}{1 + V(S_\delta \cup K)} \geq \frac{R_\gamma(S_\gamma \setminus K) - A_\gamma(K)}{1 + V(S_\gamma)}$. Just before the two inequalities above, we established that $A_\gamma(K) < R_\gamma(S_\gamma) \leq R_\gamma(S_\gamma \setminus K)$. Furthermore, since $K = S_\gamma \setminus S_\delta$, we have $S_\delta \cup K \supseteq S_\gamma$, so $1 + V(S_\delta \cup K) \geq 1 + V(S_\gamma)$, in which case, we obtain $R_\delta(S_\delta) - A_\delta(K) \geq \frac{1 + V(S_\delta \cup K)}{1 + V(S_\gamma)} (R_\gamma(S_\gamma \setminus K) - A_\gamma(K)) \geq R_\gamma(S_\gamma \setminus K) - A_\gamma(K) \geq R_\gamma(S_\gamma) - A_\gamma(K)$. Using the fact that $A_x(S) = A(S) + x$, focusing on the first and last expressions in this chain of inequalities, we obtain $R_\delta(S_\delta) - \delta \geq R_\gamma(S_\gamma) - \gamma$, which contradicts the assumption that $R_\delta(S_\delta) - \delta < R_\gamma(S_\gamma) - \gamma$ in Case 2.

Appendix F: Proof of Theorem 5.2

In this section, we give our FPTAS. Throughout this section, for notational brevity, we let $\Theta(S) = \sum_{i \in S} r_i \theta_i$ and $W(S) = \sum_{i \in S} r_i v_i$ and $V(S) = \sum_{i \in S} v_i$. For a fixed accuracy parameter $\alpha > 0$, we construct the grid points $\text{Grid} = \{(1 + \alpha)^k : k = \lfloor \frac{\log \underline{v}}{\log(1 + \alpha)} \rfloor, \dots, \lceil \frac{\log \bar{v}}{\log(1 + \alpha)} \rceil\}$. We write the three constraints in problem (2) as $\sum_{i \in S} \frac{n}{\alpha p} r_i \theta_i \geq \frac{n}{\alpha}$, $\sum_{i \in S} \frac{n}{\alpha q} r_i v_i \geq \frac{n}{\alpha}$ and $\sum_{i \in S} \frac{n}{\alpha s} v_i \leq \frac{n}{\alpha}$. Replacing these three constraints with $\sum_{i \in S} \lceil \frac{n}{\alpha p} r_i \theta_i \rceil \geq \lfloor \frac{n}{\alpha} \rfloor$, $\sum_{i \in S} \lceil \frac{n}{\alpha q} r_i v_i \rceil \geq \lfloor \frac{n}{\alpha} \rfloor$ and $\sum_{i \in S} \lfloor \frac{n}{\alpha s} v_i \rfloor \leq \lceil \frac{n}{\alpha} \rceil$, we obtain an approximate version of problem (2). In particular, letting $\sigma_i(p) = \lceil \frac{n}{\alpha p} r_i \theta_i \rceil$, $\kappa_i(q) = \lceil \frac{n}{\alpha q} r_i v_i \rceil$ and $\gamma_i(s) = \lfloor \frac{n}{\alpha s} v_i \rfloor$, this approximate version is given by

$$\widehat{G}(p, q, s) = \min_{S \subseteq N} \left\{ \sum_{i \in S} c_i : \sum_{i \in S} \sigma_i(p) \geq \lfloor \frac{n}{\alpha} \rfloor, \sum_{i \in S} \kappa_i(q) \geq \lfloor \frac{n}{\alpha} \rfloor, \sum_{i \in S} \gamma_i(s) \leq \lceil \frac{n}{\alpha} \rceil \right\}. \quad (17)$$

Letting S^* be optimal to the Capacitated Mixture problem, by the construction of Grid , there exists $(\widehat{p}, \widehat{q}, \widehat{s}) \in \text{Grid}^3$ such that $\widehat{p} \leq \Theta(S^*) \leq (1 + \alpha)\widehat{p}$, $\widehat{q} \leq W(S^*) \leq (1 + \alpha)\widehat{q}$ and $\frac{1}{1 + \alpha}\widehat{s} \leq V(S^*) \leq \widehat{s}$.

In the next lemma, we show that S^* is feasible to problem (17) when we solve this problem with any (p, q, s) such that $p \leq \Theta(S^*) \leq (1 + \alpha)p$, $q \leq W(S^*) \leq (1 + \alpha)q$ and $\frac{1}{1 + \alpha}s \leq V(S^*) \leq s$.

Lemma F.1 *Using S^* to denote an optimal solution to the Capacitated Mixture problem, let $(\widehat{p}, \widehat{q}, \widehat{s})$ be such that $\widehat{p} \leq \Theta(S^*) \leq (1 + \alpha)\widehat{p}$, $\widehat{q} \leq W(S^*) \leq (1 + \alpha)\widehat{q}$ and $\frac{1}{1 + \alpha}\widehat{s} \leq V(S^*) \leq \widehat{s}$. Then, S^* is a feasible solution to problem (17) with $(p, q, s) = (\widehat{p}, \widehat{q}, \widehat{s})$.*

Proof. By our choice of \widehat{p} , we have $\sum_{i \in S^*} r_i \theta_i = \Theta(S^*) \geq \widehat{p}$. Multiplying this chain of inequalities with $\frac{n}{\alpha \widehat{p}}$, we obtain $\sum_{i \in S^*} \frac{n}{\alpha \widehat{p}} r_i \theta_i \geq \frac{n}{\alpha}$, which implies that $\sum_{i \in S^*} \lceil \frac{n}{\alpha \widehat{p}} r_i \theta_i \rceil \geq \lfloor \frac{n}{\alpha} \rfloor$, so noting the definition of $\sigma_i(p)$, we get $\sum_{i \in S^*} \sigma_i(\widehat{p}) \geq \lfloor \frac{n}{\alpha} \rfloor$. Thus, S^* satisfies the first constraint in problem

(17) when we solve this problem with $(p, q, s) = (\hat{p}, \hat{q}, \hat{s})$. We can use precisely the same argument to show that S^* satisfies the second constraint in problem (17) when we solve this problem with $(p, q, s) = (\hat{p}, \hat{q}, \hat{s})$. Lastly, by our choice of \hat{s} , we have $\sum_{i \in S^*} v_i = V(S^*) \leq \hat{s}$. Multiplying this chain of inequalities with $\frac{n}{\alpha \hat{s}}$, we get $\sum_{i \in S^*} \frac{n}{\alpha \hat{s}} v_i \leq \frac{n}{\alpha}$, which implies that $\sum_{i \in S^*} \lfloor \frac{n}{\alpha \hat{s}} v_i \rfloor \leq \lceil \frac{n}{\alpha} \rceil$. In this case, using the definition of $\gamma_i(s)$, we get $\sum_{i \in S^*} \gamma_i(\hat{s}) \leq \lceil \frac{n}{\alpha} \rceil$, so S^* satisfies the third constraint in problem (17) when we solve this problem with $(p, q, s) = (\hat{p}, \hat{q}, \hat{s})$. ■

By the next lemma, if we choose (p, q, s) with $p \leq \Theta(S^*) \leq (1 + \alpha)p$, $q \leq W(S^*) \leq (1 + \alpha)q$ and $\frac{1}{1 + \alpha} s \leq V(S^*) \leq s$, then (17) yields a performance guarantee for the Capacitated Mixture problem.

Lemma F.2 *Using S^* to denote an optimal solution to the Capacitated Mixture problem with the expected revenue z^* , let $(\hat{p}, \hat{q}, \hat{s})$ be such that $\hat{p} \leq \Theta(S^*) \leq (1 + \alpha)\hat{p}$, $\hat{q} \leq W(S^*) \leq (1 + \alpha)\hat{q}$ and $\frac{1}{1 + \alpha} \hat{s} \leq V(S^*) \leq \hat{s}$. Then, letting \hat{S} be an optimal solution to (17) with $(p, q, s) = (\hat{p}, \hat{q}, \hat{s})$, \hat{S} is a feasible solution to the Capacitated Mixture problem with an expected revenue of at least $\frac{(1 - 2\alpha)^2}{(1 + \alpha)^2} z^*$.*

Proof. By Lemma F.1, S^* is a feasible solution to problem (17) when we solve this problem with $(p, q, s) = (\hat{p}, \hat{q}, \hat{s})$, but \hat{S} is an optimal solution to problem (17) with $(p, q, s) = (\hat{p}, \hat{q}, \hat{s})$. Therefore, we have $\sum_{i \in \hat{S}} c_i \leq \sum_{i \in S^*} c_i \leq C$, where the last inequality follows since S^* is an optimal solution to the Capacitated Mixture problem, so it satisfies the constraint in this problem. Thus, \hat{S} is a feasible solution to the Capacitated Mixture problem. To compute the expected revenue of the assortment \hat{S} , using the fact that $x \geq \lceil x \rceil - 1$ and $\lfloor x \rfloor \geq x - 1$, we have the chain of inequalities

$$\begin{aligned} \sum_{i \in \hat{S}} r_i \theta_i &= \frac{\alpha \hat{p}}{n} \sum_{i \in \hat{S}} \frac{n}{\alpha \hat{p}} r_i \theta_i \geq \frac{\alpha \hat{p}}{n} \sum_{i \in \hat{S}} \left(\left\lfloor \frac{n}{\alpha \hat{p}} r_i \theta_i \right\rfloor - 1 \right) \stackrel{(a)}{\geq} \frac{\alpha \hat{p}}{n} \left(\sum_{i \in \hat{S}} \left\lfloor \frac{n}{\alpha \hat{p}} r_i \theta_i \right\rfloor - n \right) \\ &\stackrel{(b)}{=} \frac{\alpha \hat{p}}{n} \left(\sum_{i \in \hat{S}} \sigma_i(\hat{p}) - n \right) \stackrel{(c)}{\geq} \frac{\alpha \hat{p}}{n} \left(\left\lfloor \frac{n}{\alpha} \right\rfloor - n \right) \geq \frac{\alpha \hat{p}}{n} \left(\frac{n}{\alpha} - 1 - n \right) = \hat{p} \left(1 - \frac{\alpha}{n} - \alpha \right) \geq \hat{p} (1 - 2\alpha), \end{aligned}$$

where (a) holds since $|\hat{S}| \leq n$, (b) follows from the definition of $\sigma_i(p)$ and (c) holds since \hat{S} is a feasible solution to problem (17) with $(p, q, s) = (\hat{p}, \hat{q}, \hat{s})$.

We can use precisely the same argument in the chain of inequalities above to show that $\sum_{i \in \hat{S}} r_i v_i \geq \hat{q} (1 - 2\alpha)$. Furthermore, we have the chain of inequalities

$$\begin{aligned} \sum_{i \in \hat{S}} v_i &= \frac{\alpha \hat{s}}{n} \sum_{i \in \hat{S}} \frac{n}{\alpha \hat{s}} v_i \leq \frac{\alpha \hat{s}}{n} \sum_{i \in \hat{S}} \left(\left\lfloor \frac{n}{\alpha \hat{s}} v_i \right\rfloor + 1 \right) \leq \frac{\alpha \hat{s}}{n} \left(\sum_{i \in \hat{S}} \left\lfloor \frac{n}{\alpha \hat{s}} v_i \right\rfloor + n \right) \\ &\stackrel{(d)}{=} \frac{\alpha \hat{s}}{n} \left(\sum_{i \in \hat{S}} \gamma_i(\hat{s}) + n \right) \stackrel{(e)}{\leq} \frac{\alpha \hat{s}}{n} \left(\left\lfloor \frac{n}{\alpha} \right\rfloor + n \right) \leq \frac{\alpha \hat{s}}{n} \left(\frac{n}{\alpha} + 1 + n \right) = \hat{s} \left(1 + \frac{\alpha}{n} + \alpha \right) \leq \hat{s} (1 + 2\alpha), \end{aligned}$$

where (d) uses the definition of $\gamma_i(s)$ and (e) holds since \hat{S} is a feasible solution to problem (17) with $(p, q, s) = (\hat{p}, \hat{q}, \hat{s})$. By the discussion so far in the proof, we have $\sum_{i \in \hat{S}} r_i \theta_i \geq (1 - 2\alpha)\hat{p}$,

$\sum_{i \in \widehat{S}} r_i v_i \geq (1 - 2\alpha) \widehat{q}$ and $\sum_{i \in \widehat{S}} v_i \leq (1 + 2\alpha) \widehat{s}$. In this case, we can compute the expected revenue provided by the assortment \widehat{S} as

$$\begin{aligned} \lambda \sum_{i \in \widehat{S}} r_i \theta_i + (1 - \lambda) \frac{\sum_{i \in \widehat{S}} r_i v_i}{1 + \sum_{i \in \widehat{S}} v_i} &\geq \lambda (1 - 2\alpha) \widehat{p} + (1 - \lambda) \frac{(1 - 2\alpha) \widehat{q}}{1 + (1 + 2\alpha) \widehat{s}} \\ &\stackrel{(f)}{\geq} \lambda \frac{1 - 2\alpha}{1 + \alpha} \Theta(S^*) + (1 - \lambda) \frac{\frac{1 - 2\alpha}{1 + \alpha} W(S^*)}{1 + (1 + 2\alpha)(1 + \alpha)V(S^*)} \\ &\geq \frac{1 - 2\alpha}{(1 + 2\alpha)(1 + \alpha)^2} \left(\lambda \Theta(S^*) + (1 - \lambda) \frac{W(S^*)}{1 + V(S^*)} \right) = \frac{1 - 2\alpha}{(1 + 2\alpha)(1 + \alpha)^2} z^* \stackrel{(g)}{\geq} \frac{(1 - 2\alpha)^2}{(1 + \alpha)^2} z^* \end{aligned}$$

where (f) holds since we have $\Theta(S^*) \leq (1 + \alpha) \widehat{p}$, $W(S^*) \leq (1 + \alpha) \widehat{q}$ and $V(S^*) \geq \frac{1}{1 + \alpha} \widehat{s}$, whereas (g) holds since $\frac{1}{1 + 2\alpha} \geq 1 - 2\alpha$. So, the expected revenue of the assortment \widehat{S} is at least $\frac{(1 - 2\alpha)^2}{(1 + \alpha)^2} z^*$. ■

Thus, the assortment \widehat{S} in Lemma F.2 is a $\frac{(1 - 2\alpha)^2}{(1 + \alpha)^2}$ -approximate solution to the **Capacitated Mixture** problem. Problem (17) is a variant of the knapsack problem with three capacity dimensions. The disutility of item i is c_i . The capacity consumptions of item i in the three dimensions are $\sigma_i(p)$, $\kappa_i(q)$ and $\gamma_i(s)$. By the definitions of $\sigma_i(p)$, $\kappa_i(q)$ and $\gamma_i(s)$, the capacity consumptions are integers. The goal is to find the items to put into the knapsack so that we minimize the total disutility of the items in the knapsack, while making sure that the total capacity consumptions in the first two dimensions do not fall below $\lfloor \frac{n}{\alpha} \rfloor$, whereas the capacity consumption in the third dimension does not exceed $\lceil \frac{n}{\alpha} \rceil$. Using a state variable to keep the capacity consumptions in the three dimensions, for fixed (p, q, s) , we can solve problem (17) using the dynamic program

$$J_i(u, w, t; p, q, s) = \min \left\{ c_i + J_{i+1}(u + \sigma_i(t), w + \kappa_i(s), t + \gamma_i(s); p, q, s), J_{i+1}(u, w, t; p, q, s) \right\},$$

with the boundary condition that we have $J_{n+1}(u, w, t; p, q, s) = 0$ if $u \geq \lfloor \frac{n}{\alpha} \rfloor$, $w \geq \lfloor \frac{n}{\alpha} \rfloor$ and $t \leq \lceil \frac{n}{\alpha} \rceil$, whereas we have $J_{n+1}(u, w, t; p, q, s) = +\infty$ otherwise.

The two terms in the min operator above correspond to including and not including product i in a solution S to problem (17). The boundary condition above ensures that we find a solution S to problem (17) that satisfies $\sum_{i \in S} \sigma_i(p) \geq \lfloor \frac{n}{\alpha} \rfloor$, $\sum_{i \in S} \kappa_i(q) \geq \lfloor \frac{n}{\alpha} \rfloor$ and $\sum_{i \in S} \gamma_i(s) \leq \lceil \frac{n}{\alpha} \rceil$. Since $\sigma_i(p) \geq 0$, $\kappa_i(q) \geq 0$ and $\gamma_i(s) \geq 0$ for all $i \in N$, as we move from one decision epoch to the next, all components of the state variable can only increase. If any of the first two components of the state variable exceeds $\lfloor \frac{n}{\alpha} \rfloor$ at any decision epoch, then we can stop keeping track of the exact value of this component of the state variable, since the boundary condition of the dynamic program depends on whether the each of the first two components of the state variable exceeds $\lfloor \frac{n}{\alpha} \rfloor$. Thus, we have $O(\frac{n}{\alpha})$ possible values for each of the first components of the state variable.

Similarly, if the third component of the state variable exceeds $\lceil \frac{n}{\alpha} \rceil$ at any decision epoch, then the boundary condition immediately implies that the value function takes the value infinity. Thus,

we have $O(\frac{n}{\alpha})$ possible values for the third component of the state variable as well. Thus, there are $O(\frac{n^3}{\alpha^3})$ possible states. Noting that there are n decision epochs, we can solve the dynamic program in $O(\frac{n^4}{\alpha^3})$ operations, giving an optimal solution to problem (17) for a fixed (p, q, s) . Noting that $\text{Grid} = \{(1 + \alpha)^k : k = \lfloor \frac{\log \underline{\nu}}{\log(1 + \alpha)} \rfloor, \dots, \lceil \frac{\log \bar{\nu}}{\log(1 + \alpha)} \rceil\}$, there are $O(\frac{1}{\log(1 + \alpha)} \log(\frac{\bar{\nu}}{\underline{\nu}})) = O(\frac{\log(\bar{\nu}/\underline{\nu})}{\alpha})$ points in Grid . Thus, we can obtain an optimal solution to problem (17) for all $(p, q, s) \in \text{Grid}^3$ in $O(\frac{n^4}{\alpha^6} \log^3(\bar{\nu}/\underline{\nu}))$ operations. Using these observations, we give a proof for Theorem 5.2.

Proof of Theorem 5.2:

Given $\epsilon \in (0, 1)$, choose the accuracy parameter in the grid as $\alpha = \epsilon/6$. By the discussion in the previous paragraph, we can obtain an optimal solution to problem (17) for all $(p, q, s) \in \text{Grid}^3$ in $O(\frac{n^4}{\alpha^6} \log^3(\bar{\nu}/\underline{\nu}))$ operations. Since $\alpha = \epsilon/6 < 1/6$, we have $\frac{(1-2\alpha)^2}{(1+\alpha)^2} \geq ((1-2\alpha)(1-\alpha))^2 \geq (1-3\alpha)^2 \geq 1-6\alpha = 1-\epsilon$, so by Lemma F.2, the expected revenue from one of these solutions is at least $(1-\epsilon)z^*$. Thus, if we check the expected revenue provided by the optimal solution to problem (17) for each $(p, q, s) \in \text{Grid}^3$ 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 (17). Thus, using the definitions of $\underline{\nu}$ and $\bar{\nu}$, noting the discussion at the end of the previous paragraph, in $O(\frac{n^4}{\alpha^6} \log^3(\bar{\nu}/\underline{\nu})) = O(\frac{n^4}{\epsilon^6} \log^3(\frac{(R_{\max} \vee 1)(nV_{\max} \vee 1)}{(R_{\min} \wedge 1)(V_{\min} \wedge \Theta_{\min})}))$ operations, we can obtain a solution to the Capacitated Mixture problem with an expected revenue of at least $(1-\epsilon)z^*$. ■

Appendix G: Constructing an Upper Bound and Testing the Heuristic

We discuss computing an upper bound on the optimal objective value of the Capacitated Mixture problem and numerically test the performance of the heuristic for this problem.

Upper Bound on the Optimal Expected Revenue:

Consider $K + 1$ points $\underline{\nu} = \tilde{p}^1 < \dots < \tilde{p}^{K+1} = \bar{\nu}$, $L + 1$ points $\underline{q} = \tilde{q}^1 < \dots < \tilde{q}^{L+1} = \bar{\nu}$ and $M + 1$ points $\underline{s} = \tilde{s}^1 < \dots < \tilde{s}^{M+1} = \bar{\nu}$. Let z_{UB}^* be the optimal objective value of the problem

$$\max_{k=1, \dots, K, \ell=1, \dots, L, m=1, \dots, M} \left\{ \lambda \tilde{p}^{k+1} + (1-\lambda) \frac{\tilde{q}^{\ell+1}}{1 + \tilde{s}^m} : \tilde{G}(\tilde{p}^k, \tilde{q}^\ell, \tilde{s}^{m+1}) \leq C \right\}, \quad (18)$$

where $\tilde{G}(p, q, s)$ is the optimal objective value of the LP in (3). We proceed to showing that z_{UB}^* is an upper bound on the optimal objective value of the Capacitated Mixture problem. Letting S^* be an optimal solution to the Capacitated Mixture problem, define $\kappa \in \{1, \dots, K\}$, $\rho \in \{1, \dots, L\}$ and $\mu \in \{1, \dots, M\}$ such that $\tilde{p}^\kappa \leq \sum_{i \in S^*} r_i \theta_i \leq \tilde{p}^{\kappa+1}$, $\tilde{q}^\rho \leq \sum_{i \in S^*} r_i v_i \leq \tilde{q}^{\rho+1}$ and $\tilde{s}^\mu \leq \sum_{i \in S^*} v_i \leq \tilde{s}^{\mu+1}$. Using the assortment S^* , we define the solution $\mathbf{x}^* \in \{0, 1\}^n$ to problem (3) such that $\mathbf{x}^* = 1$ if and only if $i \in S^*$. Observe that \mathbf{x}^* is feasible to problem (3) with $(p, q, s) = (\tilde{p}^\kappa, \tilde{q}^\rho, \tilde{s}^{\mu+1})$. In particular,

by our choice of κ, ρ and μ , we get $\sum_{i \in N} r_i \theta_i x_i^* = \sum_{i \in S^*} r_i \theta_i \geq \tilde{p}^\kappa$, $\sum_{i \in N} r_i v_i x_i^* = \sum_{i \in S^*} r_i v_i \geq \tilde{q}^\rho$ and $\sum_{i \in N} v_i x_i^* = \sum_{i \in S^*} v_i \leq \tilde{s}^{\mu+1}$. In this case, it follows that $\tilde{G}(\tilde{p}^\kappa, \tilde{q}^\rho, \tilde{s}^{\mu+1}) \leq \sum_{i \in N} c_i x_i^* = \sum_{i \in S^*} c_i \leq C$, where the first inequality holds because \mathbf{x}^* is only a feasible solution to problem (3) when we solve this problem with $(p, q, s) = (\tilde{p}^\kappa, \tilde{q}^\rho, \tilde{s}^{\mu+1})$, but $\tilde{G}(\tilde{p}^\kappa, \tilde{q}^\rho, \tilde{s}^{\mu+1})$ is the optimal objective value of this problem, whereas the last inequality holds because S^* is an optimal solution to the **Capacitated Mixture** problem, so it satisfies the capacity constraint $\sum_{i \in S^*} c_i \leq C$. Thus, we have $\tilde{G}(\tilde{p}^\kappa, \tilde{q}^\rho, \tilde{s}^{\mu+1}) \leq C$, which implies that (κ, ρ, μ) is a feasible solution to problem (18). Since the optimal objective value of problem (18) is z_{UB}^* , we get

$$z_{\text{UB}}^* \geq \lambda \tilde{p}^{\kappa+1} + (1 - \lambda) \frac{\tilde{q}^{\rho+1}}{1 + \tilde{s}^\mu} \geq \lambda \sum_{i \in S^*} r_i \theta_i + (1 - \lambda) \frac{\sum_{i \in S^*} r_i v_i}{1 + \sum_{i \in S^*} v_i},$$

where we use the fact that $\sum_{i \in S^*} r_i \theta_i \leq \tilde{p}^{\kappa+1}$, $\sum_{i \in S^*} r_i v_i \leq \tilde{q}^{\rho+1}$ and $\tilde{s}^\mu \leq \sum_{i \in S^*} v_i$ by our choice of (κ, ρ, μ) . Thus, z_{UB}^* upper bounds the optimal objective value of the **Capacitated Mixture** problem.

Testing the Performance of the Heuristic:

We give a brief numerical study to test the performance of the heuristic for the **Capacitated Mixture** problem. We describe our experimental setup and give our results.

Experimental Setup. In our numerical experiments, we randomly generate a large number of test problems. For each test problem, we use the heuristic to obtain a solution and use the approach described in this section to compute an upper bound on the optimal expected revenue. We check the gap between the upper bound on the optimal expected revenue and the expected revenue from the solution obtained by the heuristic. We generate our test problems as follows. We sample the revenue of each product from the uniform distribution over $[1, 10]$. To come up with the purchase probability of each product in the independent demand segment, we sample ζ_i from the uniform distribution over $[0, 1]$ and set $\theta_i = \zeta_i / \sum_{j \in N} \zeta_j$. To come up with the preference weight of each product in the multinomial logit segment, we sample ξ_i from the uniform distribution over $[1, 10]$ and set $v_i = \frac{1 - P_0}{P_0} \frac{\xi_i}{\sum_{j \in N} \xi_j}$, where P_0 is a parameter that we vary. Thus, if we offer all products, then a customer in the multinomial logit segment leaves without making a purchase with probability $\frac{1}{1 + \sum_{i \in N} v_i} = \frac{1}{1 + (1 - P_0)/P_0} = P_0$. We sample the relative market size λ of the independent demand segment from the uniform distribution over $[0, 1]$. The capacity consumption of all products is one, so we have a constraint on the number of products that we offer. To come up with the capacity availability, once we generate the product revenues and choice model parameters, we solve the **Capacitated Mixture** problem without any constraints. Letting S^* be an optimal solution, we set the capacity availability as $C = \beta |S^*|$, where β is another parameter that we vary.

Recalling that n is the number of products, noting that P_0 controls the likelihood that a customer leaves without a purchase and β controls the tightness of the capacity, varying $n \in \{50, 100\}$,

Param. (n, P_0, β)	Perc. Opt. Gap			Params. (n, P_0, β)	Perc. Opt. Gap		
	Avg.	95th	Max		Avg.	95th	Max
(50, 0.05, 0.5)	0.29%	0.43%	0.63%	(100, 0.05, 0.5)	0.29%	0.42%	0.44%
(50, 0.05, 0.7)	0.23%	0.41%	0.43%	(100, 0.05, 0.7)	0.24%	0.37%	0.43%
(50, 0.05, 0.9)	0.19%	0.33%	0.37%	(100, 0.05, 0.9)	0.21%	0.32%	0.38%
(50, 0.1, 0.5)	0.18%	0.28%	0.30%	(100, 0.1, 0.5)	0.18%	0.29%	0.40%
(50, 0.1, 0.7)	0.14%	0.24%	0.34%	(100, 0.1, 0.7)	0.14%	0.24%	0.32%
(50, 0.1, 0.9)	0.12%	0.21%	0.25%	(100, 0.1, 0.9)	0.13%	0.22%	0.26%
(50, 0.3, 0.5)	0.06%	0.12%	0.40%	(100, 0.3, 0.5)	0.08%	0.16%	0.22%
(50, 0.3, 0.7)	0.06%	0.13%	0.26%	(100, 0.3, 0.7)	0.07%	0.18%	0.20%
(50, 0.3, 0.9)	0.05%	0.17%	0.23%	(100, 0.3, 0.9)	0.06%	0.13%	0.14%
Average	0.15%			Average	0.16%		

Table EC.2 Optimality gap of the heuristic for the Capacitated Mixture problem.

$P_0 \in \{0.05, 0.1, 0.3\}$ and $\beta \in \{0.5, 0.7, 0.9\}$, we obtain 18 parameter combinations. In each parameter combination, we randomly generate 100 problem instances by using the approach described in the previous paragraph. For the heuristic, we construct the grid points with $\alpha = 0.25$. For the upper bound, we split the interval $[\underline{\nu}, \bar{\nu}]$ into 50 equally spaced subintervals.

Numerical Results. We give our numerical results in Table EC.2. In this table, the first column shows the parameter configuration by using the triplet (n, P_0, β) . The second column shows the average percent gap between the upper bound on the optimal expected revenue and the expected revenue of the solution from the heuristic, where the average is computed over the 100 problem instances in a parameter combination. In particular, for problem instance k , letting Rev^k be the expected revenue of the solution from the heuristic and UB^k be upper bound on the optimal expected revenue, the first column gives the average of the data $\{100 \times \frac{\text{UB}^k - \text{Rev}^k}{\text{UB}^k} : k = 1, \dots, 100\}$. In the second and third columns, we give the 95th percentile and maximum of the same data. Our results indicate that the heuristic performs remarkably well. Over all problem instances, the average optimality gap is 0.15%. The maximum optimality gap does not exceed 0.63%. The average running time for the heuristic over all problem instances is less than a second. The optimality gaps tend to get slightly larger when β gets smaller so that the capacities get tighter, but even for the test problems with $\beta = 0.5$, the average optimality gap is 0.18%.

Appendix H: Proof of Theorem 6.1

The first part of Theorem 6.1 focuses on recovering a primal optimal solution to the Choice-Based LP by using the same to the Compact LP, whereas the second part of Theorem 6.1 focuses on recovering a dual optimal solution to the Choice-Based LP by using the same to the Compact LP. In this section, we show each of these two parts separately.

H.1 Recovering a Primal Optimal Solution

We give a proof for the first part of Theorem 6.1. 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 . In the next lemma, we establish a useful property of the basic optimal solutions to the Compact LP. This lemma plays a critical role in the proof of the first part of Theorem 6.1 and holds under 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 . We give the proof of the lemma after the proof of the first part of Theorem 6.1.

Lemma H.1 (Extreme Point Optimal Solutions) *Let $(x_0^*, \mathbf{x}^*, \mathbf{y}^*)$ be a basic optimal solution to the Compact LP. Then, we have $y_{ij}^* = \min\{x_i^*, x_j^*\}$ for all $i, j \in N$.*

We can generate counterexamples to show that we may not have $y_{ij}^* = \min\{x_i^*, x_j^*\}$ when we use basic optimal solutions without the largest value for the decision variable x_0 . 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. Using the lemma above, we give a proof for the first part of Theorem 6.1.

Proof of the First Part of Theorem 6.1:

Using $(x_0^*, \mathbf{x}^*, \mathbf{y}^*)$ to denote a basic optimal solution to the Compact LP, let the solution $\hat{\mathbf{w}}$ for the Choice-Based LP be provided by the Recovery formula. We establish two claims.

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 $i = 1, \dots, n$. Thus, using the Recovery formula, we get

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

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

Second, letting $\Lambda_i^* = (\lambda \theta_i + (1 - \lambda) v_i) x_i^* + \lambda \theta_i \sum_{j \in N} v_j y_{ij}^*$ for notational brevity, we show that $\sum_{S \subseteq N} \mathbf{1}(i \in S) (\lambda \theta_i + (1 - \lambda) \frac{v_i}{1 + V(S)}) \hat{w}(S) = \Lambda_i^*$. By the definition of $\hat{\mathbf{w}}$, we have

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

where the last equality uses the fact that $S_i = \{1, \dots, i\}$ and $V(S) = \sum_{i \in S} v_i$. By Lemma H.1, we have $y_{ij}^* = \min\{x_i^*, x_j^*\}$ for all $i, j \in N$. Thus, since we index the products such that

$x_1^* \geq x_2^* \geq \dots \geq x_n^*$, we have $y_{ij}^* = x_i^*$ for $i \geq j$ 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 chain of equalities above, for each $i \in N$, we get

$$\begin{aligned} \sum_{k \in N} \mathbf{1}(k \geq i) \widehat{w}(S_k) &= \sum_{k \in N} \mathbf{1}(k \geq i) (x_k^* - x_{k+1}^*) + \sum_{k \in N} \sum_{\ell \in N} \mathbf{1}(k \geq i) \mathbf{1}(\ell \leq k) (x_k^* - x_{k+1}^*) v_\ell \\ &\stackrel{(d)}{=} \sum_{k \in N} \mathbf{1}(k \geq i) (x_k^* - x_{k+1}^*) + \sum_{\ell \in N} v_\ell \sum_{k \in N} \mathbf{1}(k \geq i \vee \ell) (x_k^* - x_{k+1}^*) \\ &\stackrel{(e)}{=} x_i^* + \sum_{\ell \in N} v_\ell x_{i \vee \ell}^* = x_i^* + \sum_{\ell \in N} v_\ell y_{i\ell}^*, \end{aligned} \quad (19)$$

where (d) holds since $\mathbf{1}(k \geq i) \mathbf{1}(\ell \leq k) = 1$ if and only if $\mathbf{1}(k \geq i \vee \ell)$ 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 $\sum_{k \in N} \mathbf{1}(k \geq i) \frac{1}{1+V(S_k)} \widehat{w}(S_k) = \sum_{k \in N} \mathbf{1}(k \geq i) (x_k^* - x_{k+1}^*) = x_i^*$. In this case, noting that $i \in S_k$ if and only if $k \geq i$, we obtain

$$\begin{aligned} \sum_{S \subseteq N} \mathbf{1}(i \in S) \left(\lambda \theta_i + (1-\lambda) \frac{v_i}{1+V(S)} \right) \widehat{w}(S) &= \sum_{k \in N} \mathbf{1}(i \in S_k) \left(\lambda \theta_i + (1-\lambda) \frac{v_i}{1+V(S_k)} \right) \widehat{w}(S_k) \\ &= \sum_{k \in N} \mathbf{1}(k \geq i) \left(\lambda \theta_i + (1-\lambda) \frac{v_i}{1+V(S_k)} \right) \widehat{w}(S_k) \\ &= \lambda \theta_i \sum_{k \in N} \mathbf{1}(k \geq i) \widehat{w}(S_k) + (1-\lambda) v_i \sum_{k \in N} \mathbf{1}(k \geq i) \frac{1}{1+V(S_k)} \widehat{w}(S_k) \\ &\stackrel{(f)}{=} \lambda \theta_i \left(x_i^* + \sum_{\ell \in N} v_\ell y_{i\ell}^* \right) + (1-\lambda) v_i x_i^* \stackrel{(g)}{=} \Lambda_i^*, \end{aligned}$$

where (f) follows from (19) and (g) holds by the definition of Λ_i^* . Thus, by the two claims, we have $\sum_{S \subseteq N} \widehat{w}(S) = 1$ and $\sum_{S \subseteq N} \mathbf{1}(i \in S) \left(\lambda \theta_i + (1-\lambda) \frac{v_i}{1+V(S)} \right) \widehat{w}(S) = \Lambda_i^*$.

Using the these two claims, we show that $\widehat{\mathbf{w}}$ is feasible to the Choice-Based LP. Since $\sum_{S \subseteq N} \widehat{w}(S) = 1$, $\widehat{\mathbf{w}}$ satisfies the second constraint in the Choice-Based LP. Also, we have

$$\begin{aligned} T \sum_{S \subseteq N} \sum_{i \in S} a_{qi} \left(\lambda \theta_i + (1-\lambda) \frac{v_i}{1+V(S)} \right) \widehat{w}(S) &= T \sum_{i \in N} a_{qi} \sum_{S \subseteq N} \mathbf{1}(i \in S) \left(\lambda \theta_i + (1-\lambda) \frac{v_i}{1+V(S)} \right) \widehat{w}(S) \\ &\stackrel{(h)}{=} T \sum_{i \in N} a_{qi} \Lambda_i^* \stackrel{(i)}{=} T \sum_{i \in N} a_{qi} \left(\left[\lambda \theta_i + (1-\lambda) v_i \right] x_i^* + \lambda \theta_i \sum_{j \in N} v_j y_{ij}^* \right) \stackrel{(j)}{\leq} c_q, \end{aligned}$$

where (h) follows from the second claim established above, (i) uses the definition of Λ_i and (j) holds because $(x_0^*, \mathbf{x}^*, \mathbf{y}^*)$ is a feasible solution to the Compact LP.

By the discussion in the previous paragraph $\widehat{\mathbf{w}}$ is a feasible solution to the Choice-Based LP. Let z_{Choice}^* be the optimal objective value of the Choice-Based LP and z_{Compact}^* be the optimal objective

value of the Compact LP. Using the fact that $(x_0^*, \mathbf{x}^*, \mathbf{y}^*)$ is an optimal solution to the Compact LP, we have the chain of inequalities

$$\begin{aligned} z_{\text{Compact}}^* &= T \sum_{i \in N} r_i \left(\left[\lambda \theta_i + (1 - \lambda) v_i \right] x_i^* + \lambda \theta_i \sum_{j \in N} v_j y_{ij}^* \right) \stackrel{(k)}{=} T \sum_{i \in N} r_i \Lambda_i^* \\ &\stackrel{(l)}{=} T \sum_{i \in N} \sum_{S \subseteq N} r_i \mathbf{1}(i \in S) \left(\lambda \theta_i + (1 - \lambda) \frac{v_i}{1 + V(S)} \right) \widehat{w}(S) \stackrel{(m)}{\leq} z_{\text{Choice}}^*, \end{aligned} \quad (20)$$

where (k) uses the definition of Λ_i , (l) follows from the second claim shown earlier and (m) holds because \widehat{w} is a feasible solution to the Choice-Based LP.

By the chain of inequalities above, we have $z_{\text{Compact}}^* \leq z_{\text{Choice}}^*$. On the other hand, let \mathbf{w}^* be an optimal solution to the Choice-Based LP. We define the solution $(\widehat{x}_0, \widehat{\mathbf{x}}, \widehat{\mathbf{y}})$ to the Compact LP as

$$\begin{aligned} \widehat{x}_0 &= \sum_{S \subseteq N} \frac{1}{1 + V(S)} w^*(S), & \widehat{x}_i &= \sum_{S \subseteq N} \frac{\mathbf{1}(i \in S)}{1 + V(S)} w^*(S), \\ \widehat{y}_{ij} &= \sum_{S \subseteq N} \frac{\mathbf{1}(i \in \widehat{S}, j \in \widehat{S})}{1 + V(S)} w^*(S). \end{aligned}$$

In this case, we can use precisely the same argument in (1) to show that $(\widehat{x}_0, \widehat{\mathbf{x}}, \widehat{\mathbf{y}})$ is a feasible solution to the Compact LP and it provides an objective value of z_{Choice}^* for the Compact LP.

Since $(\widehat{x}_0, \widehat{\mathbf{x}}, \widehat{\mathbf{y}})$ is a feasible solution to the Compact LP, providing an objective value of z_{Choice}^* for the Compact LP, the optimal objective value of the Compact LP satisfies $z_{\text{Compact}}^* \geq z_{\text{Choice}}^*$, which implies that the inequality in (20) must hold as equality. In particular, noting that (m) in (20) must hold as an equality, it follows that \widehat{w} is a feasible solution to the Choice-Based LP and it provides an objective value of z_{Choice}^* for the Choice-Based LP. Therefore, \widehat{w} must be an optimal solution to the Choice-Based LP, which is the desired result. \blacksquare

We turn our attention to the proof of Lemma H.1, which played a critical role in our proof of the first part of Theorem 6.1 that we gave above.

Proof of Lemma H.1:

The proof of Lemma H.1 uses three auxiliary results. By the discussion at the end of Section 6, 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 . 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 ϵx_0 to the objective function of the Compact LP. If ϵ is small enough, then solving the Compact LP with the additional term in the objective function 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 work with a version of the Compact LP with the additional term ϵx_0 in the objective function, which is given by

$$z_{\text{Compact}}^\epsilon = \max_{(x_0, \mathbf{x}, \mathbf{y}) \in \mathbb{R} \times \mathbb{R}_+^{n+n^2}} \left\{ T \sum_{i \in N} r_i \left((\lambda \theta_i + (1 - \lambda) v_i) x_i + \lambda \theta_i \sum_{j \in N} v_j y_{ij} \right) + \epsilon x_0 \quad : \quad (21) \right.$$

$$T \sum_{i \in N} a_{qi} \left((\lambda \theta_i + (1 - \lambda) v_i) x_i + \lambda \theta_i \sum_{j \in N} v_j y_{ij} \right) \leq c_q \quad \forall q \in M,$$

$$x_0 + \sum_{i \in N} v_i x_i = 1,$$

$$x_i \leq x_0 \quad \forall i \in N,$$

$$\left. y_{ij} \leq x_i \quad \forall i, j \in N, \quad y_{ij} \leq x_j \quad \forall i, j \in N \right\}.$$

If ϵ is small enough, then a basic optimal solution to the problem above is also a basic optimal solution to the Compact LP. So, it is enough to show that if $(x_0^*, \mathbf{x}^*, \mathbf{y}^*)$ is a basic optimal solution to problem (21), then we have $y_{ij}^* = \min\{x_i^*, x_j^*\}$ for all $i, j \in N$. For notational brevity, we let $\mathcal{P} = \{(x_0, \mathbf{x}, \mathbf{y}) \in \mathbb{R} \times \mathbb{R}_+^{n+n^2} : x_0 + \sum_{i \in N} v_i x_i = 1, x_i \leq x_0 \quad \forall i \in N, y_{ij} \leq \min\{x_i, x_j\} \quad \forall i, j \in N\}$, denoting the polytope captured by the last four constraints in the LP above. The proof of Lemma H.1 uses three auxiliary results, given in three lemmas.

In our first auxiliary lemma, we consider a slightly modified version the Assortment LP, where we add the additional term ϵx_0 to the objective function. In particular, consider the LP

$$\max_{(x_0, \mathbf{x}, \mathbf{y}) \in \mathbb{R} \times \mathbb{R}_+^{n+n^2}} \left\{ \sum_{i \in N} r_i \left((\lambda \theta_i + (1 - \lambda) v_i) x_i + \lambda \theta_i \sum_{j \in N} v_j y_{ij} \right) + \epsilon x_0 \quad : \quad (x_0, \mathbf{x}, \mathbf{y}) \in \mathcal{P} \right\}. \quad (22)$$

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 H.2 *For a basic optimal solution $(x_0^*, \mathbf{x}^*, \mathbf{y}^*)$ to problem (22), let $S^* = \{i \in N : x_i^* > 0\}$. Then, S^* is an optimal solution to the problem*

$$\max_{S \subseteq N} \left\{ \sum_{i \in S} r_i \left(\lambda \theta_i + (1 - \lambda) \frac{v_i}{1 + V(S)} \right) + \frac{\epsilon}{1 + V(S)} \right\}. \quad (23)$$

Lemma H.2 is an analogue of Theorem 3.2. We skip its proof since it uses the same argument in the proof of Theorem 3.2. Also, problems (22) and (23) have the same optimal objective values.

In our second auxiliary lemma, we study an alternative representation of the dual of problem (21), obtained by viewing the dual as taking place in two stages. In particular, associating 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 $\boldsymbol{\sigma} = \{\sigma_{ij} : i, j \in N\}$ with the five sets of constraints in problem (21), we can write the dual of this problem as

$$\begin{aligned} \min_{(\boldsymbol{\mu}, \pi, \boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{\sigma}) \in \mathbb{R}_+^m \times \mathbb{R} \times \mathbb{R}_+^{n+2n^2}} \left\{ \sum_{q \in M} c_q \mu_q + \pi \quad : \quad \pi = \sum_{i \in N} \alpha_i + \epsilon, \right. \\ \left. v_i \pi + \alpha_i - \sum_{j \in N} \eta_{ij} - \sum_{j \in N} \sigma_{ji} \geq T(\lambda \theta_i + (1 - \lambda) v_i) \left(r_i - \sum_{q \in M} a_{qi} \mu_q \right) \quad \forall i \in N, \right. \\ \left. \eta_{ij} + \sigma_{ij} \geq T \lambda \theta_i v_j \left(r_i - \sum_{q \in M} a_{qi} \mu_q \right) \quad \forall i, j \in N \right\}. \end{aligned} \quad (24)$$

We can solve this problem two stages. In the inner problem, we find the optimal values $(\pi, \boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{\sigma})$ for fixed value of $\boldsymbol{\mu}$. In the outer problem, we find the optimal value of $\boldsymbol{\mu}$.

We use $F^\epsilon(\boldsymbol{\mu})$ to denote the optimal objective value of the inner problem as a function of the fixed value of $\boldsymbol{\mu}$. In particular, we have

$$\begin{aligned} F^\epsilon(\boldsymbol{\mu}) = \min_{(\pi, \boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{\sigma}) \in \mathbb{R} \times \mathbb{R}_+^{n+2n^2}} \left\{ \pi \quad : \quad \pi = \sum_{i \in N} \alpha_i + \epsilon, \right. \\ \left. v_i \pi + \alpha_i - \sum_{j \in N} \eta_{ij} - \sum_{j \in N} \sigma_{ji} \geq T(\lambda \theta_i + (1 - \lambda) v_i) \left(r_i - \sum_{q \in M} a_{qi} \mu_q \right) \quad \forall i \in N, \right. \\ \left. \eta_{ij} + \sigma_{ij} \geq T \lambda \theta_i v_j \left(r_i - \sum_{q \in M} a_{qi} \mu_q \right) \quad \forall i, j \in N \right\}, \end{aligned} \quad (25)$$

in which case, problem (24) is equivalent to $\min_{\boldsymbol{\mu} \in \mathbb{R}_+^m} \sum_{q \in M} c_q \mu_q + F^\epsilon(\boldsymbol{\mu})$. In the next lemma, we give an alternative representation of $F^\epsilon(\boldsymbol{\mu})$.

Lemma H.3 *Noting that $F^\epsilon(\boldsymbol{\mu})$ is the optimal objective value of problem (25) for fixed $\boldsymbol{\mu} \in \mathbb{R}_+^m$, we can compute $F^\epsilon(\boldsymbol{\mu})$ alternatively as*

$$\max_{S \subseteq N} \left\{ T \sum_{i \in S} \left(r_i - \sum_{q \in M} a_{qi} \mu_q \right) \left(\lambda \theta_i + (1 - \lambda) \frac{v_i}{1 + V(S)} \right) + \frac{\epsilon}{1 + V(S)} \right\}.$$

Proof. Associating the dual variables x_0 , $\boldsymbol{x} = \{x_i : i \in N\}$ and $\boldsymbol{y} = \{y_{ij} : i, j \in N\}$ with the three sets of constraints in problem (25), the dual of problem (25) is identical to problem (22) after replacing each occurrence of r_i with $r_i - \sum_{q \in M} a_{qi} \mu_q$. Thus, we can obtain $F^\epsilon(\boldsymbol{\mu})$ by solving problem (22) after replacing r_i with $T(r_i - \sum_{q \in M} a_{qi} \mu_q)$. On the other hand, by Lemma H.2, we can obtain the optimal objective value of problem (22) by solving problem (23). Thus, we can obtain $F^\epsilon(\boldsymbol{\mu})$ by solving problem (23) after replacing each occurrence of r_i with $T(r_i - \sum_{q \in M} a_{qi} \mu_q)$, which is equivalent to solving the problem given in the lemma. \blacksquare

In our third auxiliary lemma, we use complementary slackness to give two useful properties that are satisfied by an optimal primal-dual solution pair for problem (21).

Lemma H.4 Using $(x_0^*, \mathbf{x}^*, \mathbf{y}^*)$ and $(\boldsymbol{\mu}^*, \boldsymbol{\pi}^*, \boldsymbol{\alpha}^*, \boldsymbol{\eta}^*, \boldsymbol{\sigma}^*)$ to denote a basic optimal primal-dual solution pair for problem (21), let $S^* = \{i \in N : x_i^* > 0\}$. Then, we have $\pi^* = \sum_{i \in S^*} \alpha_i^* + \epsilon$ and $\sum_{i \in S^*} \sum_{j \in N} (\eta_{ij}^* + \sigma_{ji}^*) = \sum_{i \in S^*} \sum_{j \in S^*} (\eta_{ij}^* + \sigma_{ij}^*)$.

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 (21), 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 (21), we have $\alpha_i^* = 0$ for all $i \notin S^*$, in which case, by the first constraint in problem (24), 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 (21), 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 (21), 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

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

Proof of Lemma H.1. We use our auxiliary results to give a proof for Lemma H.1. Let $(x_0^*, \mathbf{x}^*, \mathbf{y}^*)$ and $(\boldsymbol{\mu}^*, \boldsymbol{\pi}^*, \boldsymbol{\alpha}^*, \boldsymbol{\eta}^*, \boldsymbol{\sigma}^*)$ be a basic optimal primal-dual solution pair for problem (21). We will 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 (21), 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 (24), 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 $\eta_{ij}^* + \sigma_{ij}^* \leq T \lambda \theta_i v_j (r_i - \sum_{q \in M} a_{qi} \mu_q^*)^+$, 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 (24), this constraint holds as equality for all $i \in S^*$. Adding over all $i \in S^*$ yields

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

where (a) follows from Lemma H.4, (b) holds since $\eta_{ij}^* + \sigma_{ij}^* \leq T \lambda \theta_i v_j (r_i - \sum_{q \in M} a_{qi} \mu_q^*)^+$ 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 H.4. 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

$$\pi^* - \frac{\epsilon}{1 + V(S^*)} \leq T \sum_{i \in S^*} \left(r_i - \sum_{q \in M} a_{qi} \mu_q^* \right)^+ \left(\lambda \theta_i + (1 - \lambda) \frac{v_i}{1 + V(S^*)} \right). \quad (26)$$

To show the result by contradiction, assume that there exist $i, j \in N$ 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 $i, j \in S^*$.

Noting that $i, j \in S^*$ are such that $y_{ij}^* < \min\{x_i^*, x_j^*\}$, using complementary slackness on the last two constraints in problem (21), it follows that $\eta_{ij}^* = 0$ and $\sigma_{ij}^* = 0$, in which case, by the last constraint in problem (24), 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.

The optimal objective value of problem (21) is $z_{\text{Compact}}^\epsilon$ and $(\boldsymbol{\mu}^*, \pi^*, \boldsymbol{\alpha}^*, \boldsymbol{\eta}^*, \boldsymbol{\sigma}^*)$ is an optimal solution to its dual given in problem (24). Therefore, we get

$$\begin{aligned} z_{\text{Compact}}^\epsilon &= \sum_{q \in M} c_q \mu_q^* + \pi^* \\ &\stackrel{(d)}{\leq} \sum_{q \in M} c_q \mu_q^* + T \sum_{i \in S^*} \left(r_i - \sum_{q \in M} a_{qi} \mu_q^* \right)^+ \left(\lambda \theta_i + (1 - \lambda) \frac{v_i}{1 + V(S^*)} \right) + \frac{\epsilon}{1 + V(S^*)} \\ &\stackrel{(e)}{=} \sum_{q \in M} c_q \mu_q^* + T \sum_{i \in S^* \setminus N^*} \left(r_i - \sum_{q \in M} a_{qi} \mu_q^* \right)^+ \left(\lambda \theta_i + (1 - \lambda) \frac{v_i}{1 + V(S^*)} \right) + \frac{\epsilon}{1 + V(S^*)} \\ &\stackrel{(f)}{<} \sum_{q \in M} c_q \mu_q^* + T \sum_{i \in S^* \setminus N^*} \left(r_i - \sum_{q \in M} a_{qi} \mu_q^* \right)^+ \left(\lambda \theta_i + (1 - \lambda) \frac{v_i}{1 + V(S^* \setminus N^*)} \right) + \frac{\epsilon}{1 + V(S^* \setminus N^*)} \\ &\leq \sum_{q \in M} c_q \mu_q^* + \max_{S \subseteq N} \left\{ T \sum_{i \in S} \left(r_i - \sum_{q \in M} a_{qi} \mu_q^* \right)^+ \left(\lambda \theta_i + (1 - \lambda) \frac{v_i}{1 + V(S)} \right) + \frac{\epsilon}{1 + V(S)} \right\} \\ &\stackrel{(g)}{=} \sum_{q \in M} c_q \mu_q^* + F^\epsilon(\boldsymbol{\mu}^*), \end{aligned} \quad (27)$$

where (d) uses (26), (e) holds since $r_i - \sum_{q \in M} a_{qi} \mu_q^* \leq 0$ for all $i \in N^*$, (f) holds since $N^* \neq \emptyset$, so $V(S^* \setminus N^*) < V(S^*)$ and (g) holds by Lemma H.3.

By the discussion right after problem (25), note that $\sum_{q \in M} c_q \mu_q^* + F^\epsilon(\boldsymbol{\mu}^*)$ corresponds to the optimal objective value of the dual of problem (21). Therefore, the right side of the chain of inequalities in (27) is also given by z_{Compact}^* , in which case, observing the strictly inequality in (f) in the chain of inequalities above, we get a contradiction. \blacksquare

H.2 Recovering a Dual Optimal Solution

We give a proof for the second part of Theorem 6.1. Associating the dual variables $\boldsymbol{\mu} = \{\mu_q : q \in M\}$ and π with the two sets of constraints in the Choice-Based LP, the dual of this problem is

$$\min_{(\boldsymbol{\mu}, \pi) \in \mathbb{R}_+^m \times \mathbb{R}} \left\{ \sum_{q \in M} c_q \mu_q + \pi : \right. \\ \left. \pi \geq T \sum_{i \in N} \left(r_i - \sum_{q \in M} a_{qi} \mu_q \right) \left(\lambda \theta_i + (1 - \lambda) \frac{v_i}{1 + V(S)} \right) \quad \forall S \subseteq N \right\}. \quad (28)$$

The dual of the Compact LP is given by setting $\epsilon = 0$ in (24). We will show that if $(\boldsymbol{\mu}^*, \pi^*, \boldsymbol{\alpha}^*, \boldsymbol{\eta}^* \boldsymbol{\sigma}^*)$ is an optimal solution to problem (24) with $\epsilon = 0$, then $(\boldsymbol{\mu}^*, \pi^*)$ is optimal to problem (28).

Letting z_{Compact}^* be the optimal objective value of the Compact LP, noting that the dual of the Compact LP is given by problem (24) with $\epsilon = 0$ and using the fact that an optimal solution to problem (24) with $\epsilon = 0$ is $(\boldsymbol{\mu}^*, \pi^*, \boldsymbol{\alpha}^*, \boldsymbol{\eta}^* \boldsymbol{\sigma}^*)$, we have $z_{\text{Compact}}^* = \sum_{q \in M} c_q \mu_q^* + \pi^*$. Furthermore, by the discussion right after problem (25), problem (24) is equivalent to $\min_{\boldsymbol{\mu} \in \mathbb{R}_+^m} \sum_{q \in M} c_q \mu_q + F^\epsilon(\boldsymbol{\mu})$, which implies that the dual of the Compact LP is equivalent to $\min_{\boldsymbol{\mu} \in \mathbb{R}_+^m} \sum_{q \in M} c_q \mu_q + F^0(\boldsymbol{\mu})$. Thus, we also have $z_{\text{Compact}}^* = \sum_{q \in M} c_q \mu_q^* + F^0(\boldsymbol{\mu}^*)$, yielding $\pi^* = F^0(\boldsymbol{\mu}^*)$.

On the other hand, letting z_{Choice}^* be the optimal objective value of the Choice-Based LP, since problem (28) is the dual of the Choice-Based LP, the optimal objective value of problem (28) is also z_{Choice}^* . By Lemma H.3, we can write the constraints in problem (28) equivalently as $\pi \geq F^0(\boldsymbol{\mu})$. So, problem (28) is equivalent to $\min_{(\boldsymbol{\mu}, \pi) \in \mathbb{R}_+^m \times \mathbb{R}} \{ \sum_{q \in M} c_q \mu_q + \pi : \pi \geq F^0(\boldsymbol{\mu}) \}$. By the discussion in the previous paragraph, we have $\pi^* = F^0(\boldsymbol{\mu}^*)$. Thus, $(\pi^*, \boldsymbol{\mu}^*)$ is feasible to problem (28). Since problem (28) is a minimization problem, its optimal objective value satisfies $z_{\text{Choice}}^* \leq \sum_{q \in M} c_q \mu_q^* + \pi^*$.

So far, we have $z_{\text{Compact}}^* = \sum_{q \in M} c_q \mu_q^* + \pi^* \geq z_{\text{Choice}}^*$. In our proof for the first part of Theorem 6.1, by (20), we have $z_{\text{Compact}}^* \leq z_{\text{Choice}}^*$. Thus, we get $z_{\text{Choice}}^* = \sum_{q \in M} c_q \mu_q^* + \pi^*$, so $(\boldsymbol{\mu}^*, \pi^*)$ is feasible to problem (28) and its objective value is equal to the optimal objective value of problem (28). ■

Appendix I: Computational Benefits of the Compact Formulation

In this section, we check the computational benefits from using the Compact LP in conjunction with Theorem 6.1 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. For each instance, we solve the Compact LP and use the Recovery formula to obtain an optimal solution to the

Choice-Based LP. We also solve the Choice-Based LP directly by using column generation. We use the following approach to generate our problem instances. The set of products is $N = \{1, \dots, n\}$ with $n = 100$ and the set of resources is $M = \{1, \dots, m\}$, 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. Thus, if we offer all products, then a customer in the multinomial logit segment 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 $\theta_i = \gamma_i / \sum_{j \in N} \gamma_j$. The purchase probability of product i within assortment S is $\phi_i(S) = \lambda \theta_i + (1 - \lambda) \frac{v_i}{1 + \sum_{j \in S} v_j}$, where λ 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 $a_{q_i} = 1$ with probability $1/5$ and $a_{q_i} = 0$ 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 $T \sum_{i \in S^*} a_{q_i} \phi_i(S^*)$. We set the capacity of resource q as $c_q = \kappa T \sum_{i \in S^*} a_{q_i} \phi_i(S^*)$, where κ is a last parameter that we vary.

Computational Results:

Varying $(m, P_0, \lambda, \kappa) \in \{25, 50\} \times \{0.1, 0.2\} \times \{0.25, 0.75\} \times \{0.6, 0.8\}$, 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 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 6.1 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 EC.3. 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

Param. (m, P_0, λ, κ)	COG Secs.	CLP Secs.	Secs. Ratio	1% Gp.	1% Gp.
				COG Secs.	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, λ, κ)	COG Secs.	CLP Secs.	Secs. Ratio	1% Gp.	1% Gp.
				COG Secs.	1% 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 EC.3 Running times for solving the Choice-Based LP through two methods.

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 CLP 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 the Recovery formula.

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 EC.4. The first column shows the problem parameters. The interpretation of the problem parameters P_0 , λ , and κ is the same as the one presented earlier in this section. The second column shows the optimality gap for COG after 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. Thus, the benefits from the Compact LP are even more pronounced when we work with larger test problems.

Appendix J: Detailed Description of the Expectation-Maximization Algorithm

In Section 7.1, we give a closed-form expression for the conditional expectations that we need to compute in the expectation step. Furthermore, we argue that the optimization problems that we

Param. (P_0, λ, κ)	COG % Opt. Gap.	CLP Secs.	Param. (P_0, λ, κ)	COG % Opt. Gap.	CLP Secs.
(0.1, 0.25, 0.6)	4.93	676.13	(0.2, 0.25, 0.6)	7.17	408.59
(0.1, 0.25, 0.8)	5.27	2139.15	(0.2, 0.25, 0.8)	4.71	1059.42
(0.1, 0.75, 0.6)	14.27	909.10	(0.2, 0.75, 0.6)	14.10	950.74
(0.1, 0.75, 0.8)	9.51	2420.80	(0.2, 0.75, 0.8)	8.53	2458.98
Average	8.49	1536.29	Average	8.63	1219.43

Table EC.4 Optimality gaps and running times for the two methods for solving the Choice-Based LP with $n = 500$ products and $m = 100$ resources.

need to solve in the maximization are convex problems and they decompose by the two segments. Below is a step-by-step description of our expectation-maximization algorithm.

Expectation-Maximization Algorithm:

The input is $\{(S_t, i_t) : t = 1, \dots, \tau\}$, where S_t is the assortment offered to customer t and i_t is the product purchased by customer t . If customer t left without a purchase, then $i_t = 0$.

Step 1. Choose the initial parameter estimates $(\lambda^1, \boldsymbol{\theta}^1, \mathbf{v}^1)$, such that $\lambda^1 \in [0, 1]$, $\boldsymbol{\theta}^1 \in [0, 1]^n$, $\sum_{i \in N} \theta_i^1 \leq 1$ and $\mathbf{v}^1 \in [0, B]^n$. Initialize the iteration counter by setting $\ell = 1$.

Step 2. (Compute the Conditional Expectations) For each customer $t = 1, \dots, \tau$ in the purchase history, compute $\bar{z}_t^\ell \in [0, 1]$ as

$$\bar{z}_t^\ell = \begin{cases} \frac{\lambda^\ell \theta_{i_t}^\ell}{\lambda^\ell \theta_{i_t}^\ell + (1 - \lambda^\ell) \frac{v_{i_t}^\ell}{1 + \sum_{j \in S_t} v_j^\ell}} & \text{if } i_t \neq 0 \\ \frac{\lambda^\ell (1 - \sum_{i \in S_t} \theta_i^\ell)}{\lambda^\ell (1 - \sum_{i \in S_t} \theta_i^\ell) + (1 - \lambda^\ell) \frac{1}{1 + \sum_{j \in S_t} v_j^\ell}} & \text{if } i_t = 0. \end{cases} \quad (29)$$

Step 3. (Maximize the Likelihood) Compute the parameters $(\lambda^{\ell+1}, \boldsymbol{\theta}^{\ell+1}, \mathbf{v}^{\ell+1})$ at the next iteration by solving the problems

$$\begin{aligned} \lambda^{\ell+1} &= \arg \max_{\lambda \in [0, 1]} \left\{ \sum_{t=1}^{\tau} \left\{ \bar{z}_t^\ell \log \lambda + (1 - \bar{z}_t^\ell) \log(1 - \lambda) \right\} \right\} \\ \boldsymbol{\theta}^{\ell+1} &= \arg \max_{\boldsymbol{\theta} \in [0, 1]^n} \left\{ \sum_{t=1}^{\tau} \bar{z}_t^\ell \left\{ \sum_{i \in S_t} \mathbf{1}(i = i_t) \log \theta_i + \mathbf{1}(i_t = 0) \log \left(1 - \sum_{i \in S_t} \theta_i \right) \right\} : \sum_{i \in N} \theta_i \leq 1 \right\} \\ \mathbf{v}^{\ell+1} &= \arg \max_{\mathbf{v} \in [0, B]^n} \left\{ \sum_{t=1}^{\tau} (1 - \bar{z}_t^\ell) \left\{ \sum_{i \in S_t} \mathbf{1}(i = i_t) \log v_i - \log \left(1 + \sum_{j \in S_t} v_j \right) \right\} \right\}. \end{aligned}$$

Step 4. For fixed $\epsilon > 0$, if $\max\{|\lambda^{\ell+1} - \lambda^\ell|, \|\boldsymbol{\theta}^{\ell+1} - \boldsymbol{\theta}^\ell\|, \|\mathbf{v}^{\ell+1} - \mathbf{v}^\ell\|\} \leq \epsilon$, then stop. Otherwise, increase the iteration counter ℓ by one and go to Step 2.

Recall that we use the random variable Z with support $\{0, 1\}$ to capture the market segment of a generic customer, whereas we use the random variable $P(S)$ with support $S \cup \{0\}$ to capture

the choice of a generic customer within the assortment S . In Step 2, given that the customers choose according to our mixture model with parameters $(\lambda^\ell, \boldsymbol{\theta}^\ell, \mathbf{v}^\ell)$, we compute the expectation $\bar{z}_t^\ell = \mathbb{E}\{Z | P(S_t) = i_t\}$. In Step 3, given the complete purchase history $\mathcal{C}^\ell = \{(\bar{z}_t^\ell, S_t, i_t) : t = 1, \dots, \tau\}$, we find a maximizer of the likelihood function $L(\lambda, \boldsymbol{\theta}, \mathbf{v}; \mathcal{C}^\ell)$ given in (4). As discussed in Section 7.1, the three optimization problems in this step can be formulated as convex programs. In Step 4, we stop the algorithm when the estimates of the parameters do not change significantly from one iteration to the next. The norm $\|\cdot\|$ stands for the Euclidean norm. In the remainder of this section, we show that the sequence of parameter estimates generated by our expectation-maximization algorithm monotonically improves the likelihood function built by using the purchase history $\mathcal{H} = \{(S_t, i_t) : t = 1, \dots, \tau\}$ that is available for estimation. Thus, we can also stop the algorithm when the improvement in the likelihood function diminishes.

Monotonicity of the Likelihood Function:

Consider the likelihood function built by using the purchase history $\mathcal{H} = \{(S_t, i_t) : t = 1, \dots, \tau\}$ available to estimate the parameters of our mixture model. This likelihood function is given by

$$\bar{L}(\lambda, \boldsymbol{\theta}, \mathbf{v}; \mathcal{H}) = \prod_{t=1}^{\tau} \left\{ \prod_{i \in S_t} \left(\lambda \theta_i + (1 - \lambda) \frac{v_i}{1 + V(S_t)} \right)^{\mathbf{1}(i_t=i)} \times \left(\lambda \left(1 - \sum_{i \in S_t} \theta_i \right) + (1 - \lambda) \frac{1}{1 + V(S_t)} \right)^{\mathbf{1}(i_t=0)} \right\}. \quad (30)$$

The likelihood function above is known as the incomplete likelihood function, whereas the likelihood function in (4) is known as the complete likelihood function.

In the next proposition, we argue that the iterates of our expectation-maximization algorithm monotonically improve the likelihood function in (30).

Proposition J.1 *Letting the sequence $\{(\lambda^\ell, \boldsymbol{\theta}^\ell, \mathbf{v}^\ell) : \ell = 1, 2, \dots\}$ be generated by the expectation-maximization algorithm, we have $\bar{L}(\lambda^{\ell+1}, \boldsymbol{\theta}^{\ell+1}, \mathbf{v}^{\ell+1}; \mathcal{H}) \geq \bar{L}(\lambda^\ell, \boldsymbol{\theta}^\ell, \mathbf{v}^\ell; \mathcal{H})$ for all $\ell = 1, 2, \dots$*

Proof. The set of parameter values $\mathcal{P} = \{(\lambda, \boldsymbol{\theta}, \mathbf{v}) \in [0, 1]^{1+n} \times [0, B]^n : \sum_{i \in N} \theta_i \leq 1\}$ for our mixture model and the incomplete likelihood function in (30) satisfy properties (5)-(7) in Wu (1983). Noting the definition of \bar{z}_t^ℓ in (29), \bar{z}_t^ℓ is continuous in the parameters $(\lambda^\ell, \boldsymbol{\theta}^\ell, \mathbf{v}^\ell)$. Also, the likelihood function in (4) is continuous in z_t . Therefore, the likelihood function in (4) satisfies the continuity property (10) in Wu (1983) as well. In this case, the result holds by Theorem 2 in Wu (1983). ■

The data that we have available to estimate the parameters of our mixture model is $\mathcal{H} = \{(S_t, i_t) : t = 1, \dots, \tau\}$. In particular, we do not have access to $\{z_t : t = 1, \dots, \tau\}$, which capture

the type of each customer. Therefore, to estimate the parameters of our mixture model, we need to maximize the likelihood function in (30). Our expectation-maximization algorithm uses the likelihood function in (4), but the sequence of parameters that is generated by our expectation-maximization algorithm does monotonically improve the likelihood function in (30), which is indeed the likelihood function that we would like to maximize to estimate the parameters of our mixture model. Thus, we can view our expectation-maximization algorithm as an alternative approach to find a local maximizer of the likelihood function in (30).

Appendix K: Performance Guarantee for Pricing

To minimize notational overhead, we consider the case where we have $c_i = 0$ for all $i \in N$. Our discussion naturally extends to the case where the products have non-zero marginal costs. Using ν to denote the common value of the price sensitivity parameters $\{\eta_i : i \in N\}$ and $\{\beta_i : i \in N\}$, we want to solve the problem $\max_{\mathbf{p} \in \mathbb{R}_+^n} \sum_{i \in N} \left(\lambda \alpha_i \frac{e^{\mu_i - \nu p_i}}{1 + e^{\mu_i - \nu p_i}} + (1 - \lambda) \frac{e^{\gamma_i - \nu p_i}}{1 + \sum_{j \in N} e^{\gamma_j - \nu p_j}} \right) p_i$. In this problem, we make the change of variables $w_i = e^{\gamma_i - \nu p_i}$, in which case, we have $p_i = \frac{1}{\nu} (\gamma_i - \log w_i)$. Since $p_i \geq 0$, we obtain $w_i \leq e^{\gamma_i}$. Letting $\sigma_i = e^{\mu_i} / e^{\gamma_i}$ for notational brevity, using the decision variables $\mathbf{w} = \{w_i : i \in N\}$, the problem given in this paragraph is equivalent to

$$\frac{1}{\nu} \max_{\mathbf{w} \in \mathbb{R}_+^n} \left\{ \sum_{i \in N} \left(\lambda \alpha_i \frac{\sigma_i w_i}{1 + \sigma_i w_i} + (1 - \lambda) \frac{w_i}{1 + \sum_{j \in N} w_j} \right) (\gamma_i - \log w_i) \right\}. \quad (31)$$

In the rest of our discussion, we drop the constant multiplier $\frac{1}{\nu}$ above. In the next lemma, we give bounds on the decision variables in an optimal solution to problem (31).

Lemma K.1 *Letting $U_i = e^{\gamma_i}$ and $L_i = e^{\gamma_i - 1} / \exp(\sum_{j \in N} e^{\mu_j} + \sum_{j \in N} e^{\gamma_j})$, there exists an optimal solution \mathbf{w}^* to problem (31) that satisfies $L_i \leq w_i^* \leq U_i$ for all $i \in N$.*

The lemma above follows by checking the derivatives of the objective function of problem (31). We defer the proof to the end of this section. For fixed $t \in \mathbb{R}_+$, consider the problem

$$G(t) = \max_{\mathbf{w} \in \mathbb{R}_+^n} \left\{ \sum_{i \in N} \left(\lambda \alpha_i \frac{\sigma_i w_i}{1 + \sigma_i w_i} + (1 - \lambda) \frac{w_i}{1 + t} \right) (\gamma_i - \log w_i) : \sum_{i \in N} w_i \leq t \right\}. \quad (32)$$

Noting Lemma K.1, there exists an optimal solution \mathbf{w}^* to problem (31) that satisfies $\sum_{i \in N} w_i^* \in [\sum_{i \in N} L_i, \sum_{i \in N} U_i]$, so it is enough to consider the values of $t \in [\sum_{i \in N} L_i, \sum_{i \in N} U_i]$ in (32). Letting $t^* = \arg \max_{t \in [\sum_{i \in N} L_i, \sum_{i \in N} U_i]} G(t)$, we can show that if we solve problem (32) with $t = t^*$, then we obtain an optimal solution to the pricing problem in (31). For fixed $t \in \mathbb{R}_+$, we can solve problem (32) by using a dynamic program with a continuous state space. In this dynamic program, the decision epochs correspond to the products. In the decision epoch for product i , the action is to choose the value of w_i for product i and the state variable is the value of $\sum_{j=1}^{i-1} w_j$

accumulated over the decision epochs $1, \dots, i-1$. If we take action w_i in decision epoch i , then the reward is $(\lambda \alpha_i \frac{\sigma_i w_i}{1+\sigma_i w_i} + (1-\lambda) \frac{w_i}{1+t})(\gamma_i - \log w_i)$. Lastly, if the value of the state variable after the last decision epoch exceeds t , then we obtain a reward of $-\infty$, ensuring that we find a solution \mathbf{w} that satisfies $\sum_{i \in N} w_i \leq t$. Using such a dynamic program to obtain an optimal solution to problem (32) poses two sources of difficulty. First, to solve this dynamic program, we need to discretize the state space, so we should characterize the error from discretizing the state space. Second, discretizing the state space in the dynamic program allows us to approximately compute $G(t)$ at a fixed t . To solve the problem $\max_{t \in [\sum_{i \in N} L_i, \sum_{i \in N} U_i]} G(t)$, we need to discretize the interval $[\sum_{i \in N} L_i, \sum_{i \in N} U_i]$, so we should also characterize the error from discretizing this interval.

We proceed to giving an approach to discretize the state space in the dynamic program and the interval $[\sum_{i \in N} L_i, \sum_{i \in N} U_i]$ in the problem $\max_{t \in [\sum_{i \in N} L_i, \sum_{i \in N} U_i]} G(t)$, so that we obtain a solution to the pricing problem (31) with a performance guarantee that we can quantify. Our approach resembles the construction of our FPTAS in Section 5. For a fixed accuracy parameter $\alpha > 0$, we consider the grid points that are integer powers of $1 + \alpha$, which are given by $\text{Grid} = \{(1 + \alpha)^k : k \in \mathbb{Z}\}$. Noting Lemma K.1, letting $\overline{\text{Grid}}_i = (\text{Grid} \cap (L_i, U_i)) \cup \{L_i, U_i\}$, we consider the values of $w_i \in \overline{\text{Grid}}_i$ in problem (32). Thus, $\overline{\text{Grid}}_i$ includes all points of Grid in the interval (L_i, U_i) as well as the two points L_i and U_i . In this case, writing the constraint in (32) as $\sum_{i \in N} \frac{n}{\alpha t} w_i \leq \frac{n}{\alpha}$, we consider an approximate version of problem (32) given by

$$\widehat{G}(t) = \max_{\mathbf{w} \in \mathbb{R}_+^n} \left\{ \sum_{i \in N} \left(\lambda \alpha_i \frac{\sigma_i w_i}{1 + \sigma_i w_i} + (1 - \lambda) \frac{w_i}{1 + t} \right) (\gamma_i - \log w_i) : \sum_{i \in N} \left\lfloor \frac{n}{\alpha t} w_i \right\rfloor \leq \left\lfloor \frac{n}{\alpha} \right\rfloor, w_i \in \overline{\text{Grid}}_i \quad \forall i \in N \right\}. \quad (33)$$

By the definition of Grid , for any $w_i \in [L_i, U_i]$, there exists some $\widehat{w}_i \in \overline{\text{Grid}}_i$ such that $\widehat{w}_i \leq w_i \leq (1 + \alpha) \widehat{w}_i$. In the next lemma, we show that we can construct a feasible solution to problem (33) by using a feasible solution to problem (32).

Lemma K.2 *Using \mathbf{w}^* to denote an optimal solution to problem (31), let \widehat{t} and $\widetilde{\mathbf{w}}$ be such that $\sum_{i \in N} w_i^* \leq \widehat{t} \leq (1 + \alpha) \sum_{i \in N} w_i^*$, $\widetilde{w}_i \leq w_i^* \leq (1 + \alpha) \widetilde{w}_i$ and $\widetilde{w}_i \in \overline{\text{Grid}}_i$ for all $i \in N$. Then, $\widetilde{\mathbf{w}}$ is a feasible solution to problem (33) when we solve this problem with $t = \widehat{t}$.*

Proof. By our choice of \widehat{t} and $\widetilde{\mathbf{w}}$, we have $\sum_{i \in N} \widetilde{w}_i \leq \sum_{i \in N} w_i^* \leq \widehat{t}$, which yields $\sum_{i \in N} \frac{n}{\alpha \widehat{t}} \widetilde{w}_i \leq \frac{n}{\alpha}$. In this case, we get $\sum_{i \in N} \left\lfloor \frac{n}{\alpha \widehat{t}} \widetilde{w}_i \right\rfloor \leq \left\lfloor \frac{n}{\alpha} \right\rfloor$, so $\widetilde{\mathbf{w}}$ satisfies the constraint in problem (33). Noting that $\widetilde{w}_i \in \overline{\text{Grid}}_i$ for all $i \in N$, $\widetilde{\mathbf{w}}$ is a feasible solution to problem (33). \blacksquare

Note that \widetilde{w}_i in the lemma above is obtained by rounding w_i^* down to the nearest point in $\overline{\text{Grid}}_i$. In the next lemma, we show that we can use problem (33) to obtain a solution with a performance guarantee for problem (31). Recall that problem (31) is equivalent to our pricing problem.

Lemma K.3 Using \mathbf{w}^* to denote an optimal solution to problem (31) with the expected revenue z^* , let \hat{t} be such that $\sum_{i \in N} w_i^* \leq \hat{t} \leq (1 + \alpha) \sum_{i \in N} w_i^*$. Then, letting $\hat{\mathbf{w}}$ be an optimal solution to problem (33) with $t = \hat{t}$, $\hat{\mathbf{w}}$ provides an expected revenue of at least $\frac{1}{(1+2\alpha)(1+\alpha)^2} z^*$.

Proof. To compute the expected revenue provided by the solution $\hat{\mathbf{w}}$, using the fact that $x \leq \lfloor x \rfloor + 1$ and $\lceil x \rceil \leq x + 1$, we get the chain of inequalities

$$\begin{aligned} \sum_{i \in N} \hat{w}_i &= \frac{\alpha \hat{t}}{n} \sum_{i \in N} \frac{n}{\alpha \hat{t}} \hat{w}_i \leq \frac{\alpha \hat{t}}{n} \sum_{i \in N} \left(\left\lfloor \frac{n}{\alpha \hat{t}} \hat{w}_i \right\rfloor + 1 \right) \stackrel{(a)}{=} \frac{\alpha \hat{t}}{n} \left(\sum_{i \in N} \left\lfloor \frac{n}{\alpha \hat{t}} \hat{w}_i \right\rfloor + n \right) \\ &\stackrel{(b)}{\leq} \frac{\alpha \hat{t}}{n} \left(\left\lceil \frac{n}{\alpha} \right\rceil + n \right) \leq \frac{\alpha \hat{t}}{n} \left(\frac{n}{\alpha} + 1 + n \right) = \hat{t} \left(1 + \frac{\alpha}{n} + \alpha \right) \leq \hat{t} (1 + 2\alpha), \end{aligned} \quad (34)$$

where (a) holds since $|N| = n$ and (b) holds since $\hat{\mathbf{w}}$ is a feasible solution to problem (33) when we solve this problem with $t = \hat{t}$.

Let $\tilde{\mathbf{w}}$ be such that $\tilde{w}_i \leq w_i^* \leq (1 + \alpha) \tilde{w}_i$ and $\tilde{w}_i \in \overline{\text{Grid}}_i$ for all $i \in N$. By Lemma K.2, $\tilde{\mathbf{w}}$ is feasible to problem (33) when we solve this problem with $t = \hat{t}$. In this case, we get

$$\begin{aligned} \sum_{i \in N} \left(\lambda \alpha_i \frac{\sigma_i \hat{w}_i}{1 + \sigma_i \hat{w}_i} + (1 - \lambda) \frac{\hat{w}_i}{1 + \sum_{j \in N} \hat{w}_j} \right) (\gamma_i - \log \hat{w}_i) \\ \stackrel{(c)}{\geq} \sum_{i \in N} \left(\lambda \alpha_i \frac{\sigma_i \hat{w}_i}{1 + \sigma_i \hat{w}_i} + (1 - \lambda) \frac{\hat{w}_i}{1 + (1 + 2\alpha) \hat{t}} \right) (\gamma_i - \log \hat{w}_i) \\ \geq \frac{1}{1 + 2\alpha} \sum_{i \in N} \left(\lambda \alpha_i \frac{\sigma_i \hat{w}_i}{1 + \sigma_i \hat{w}_i} + (1 - \lambda) \frac{\hat{w}_i}{1 + \hat{t}} \right) (\gamma_i - \log \hat{w}_i) \\ \stackrel{(d)}{\geq} \frac{1}{1 + 2\alpha} \sum_{i \in N} \left(\lambda \alpha_i \frac{\sigma_i \tilde{w}_i}{1 + \sigma_i \tilde{w}_i} + (1 - \lambda) \frac{\tilde{w}_i}{1 + \hat{t}} \right) (\gamma_i - \log \tilde{w}_i), \end{aligned} \quad (35)$$

where (c) holds by (34) and $\log \hat{w}_i \leq \log U_i = \gamma_i$ since $\hat{w}_i \in \overline{\text{Grid}}_i$, whereas (d) holds since $\hat{\mathbf{w}}$ is an optimal solution to problem (33) with $t = \hat{t}$ but $\tilde{\mathbf{w}}$ is only feasible to problem (33) with $t = \hat{t}$.

We lower bound each term in the summation on the right side above. In particular, noting that $(1 + \alpha) \tilde{w}_i \geq w_i^* \geq \tilde{w}_i$, we obtain

$$\begin{aligned} \left(\lambda \alpha_i \frac{\sigma_i \tilde{w}_i}{1 + \sigma_i \tilde{w}_i} + (1 - \lambda) \frac{\tilde{w}_i}{1 + \hat{t}} \right) (\gamma_i - \log \tilde{w}_i) &\geq \left(\lambda \alpha_i \frac{\sigma_i \frac{w_i^*}{1 + \alpha}}{1 + \sigma_i \frac{w_i^*}{1 + \alpha}} + (1 - \lambda) \frac{\frac{w_i^*}{1 + \alpha}}{1 + \hat{t}} \right) (\gamma_i - \log w_i^*) \\ &\stackrel{(e)}{\geq} \frac{1}{1 + \alpha} \left(\lambda \alpha_i \frac{\sigma_i w_i^*}{1 + \sigma_i w_i^*} + (1 - \lambda) \frac{w_i^*}{1 + \hat{t}} \right) (\gamma_i - \log w_i^*) \\ &\stackrel{(f)}{\geq} \frac{1}{(1 + \alpha)^2} \left(\lambda \alpha_i \frac{\sigma_i w_i^*}{1 + \sigma_i w_i^*} + (1 - \lambda) \frac{w_i^*}{1 + \sum_{j \in N} w_j^*} \right) (\gamma_i - \log w_i^*), \end{aligned}$$

where (e) holds by writing the first fraction in the parenthesis on the left side of the inequality as $\frac{\sigma_i w_i^*}{1 + \alpha + \sigma_i w_i^*}$ and noting that $\frac{\sigma_i w_i^*}{1 + \alpha + \sigma_i w_i^*} \geq \frac{1}{1 + \alpha} \frac{\sigma_i w_i^*}{1 + \sigma_i w_i^*}$, whereas (f) holds because we have

$\hat{t} \leq (1 + \alpha) \sum_{i \in N} w_i^*$ by our choice of \hat{t} . Using the last chain of inequalities above on the right side of (35), it follows that

$$\begin{aligned} & \sum_{i \in N} \left(\lambda \alpha_i \frac{\sigma_i \hat{w}_i}{1 + \sigma_i \hat{w}_i} + (1 - \lambda) \frac{\hat{w}_i}{1 + \sum_{j \in N} \hat{w}_j} \right) (\gamma_i - \log \hat{w}_i) \\ & \geq \frac{1}{(1 + 2\alpha)(1 + \alpha)^2} \sum_{i \in N} \left(\lambda \alpha_i \frac{\sigma_i w_i^*}{1 + \sigma_i w_i^*} + (1 - \lambda) \frac{w_i^*}{1 + \sum_{j \in N} w_j^*} \right) (\gamma_i - \log w_i^*). \end{aligned}$$

The desired result follows by noting that the sums on the left and right side above, respectively, correspond to the objective values of problem (31) evaluated at $\hat{\mathbf{w}}$ and \mathbf{w}^* . ■

Similar to the development of our FPTAS, for fixed t , we can find an optimal solution to problem (33) by using the dynamic program

$$J_i(q; t) = \max_{w_i \in \overline{\text{Grid}}_i} \left\{ \left(\lambda \alpha_i \frac{\sigma_i w_i}{1 + \sigma_i w_i} + (1 - \lambda) \frac{w_i}{1 + t} \right) (\gamma_i - \log w_i) + J_{i+1} \left(q + \left\lfloor \frac{n}{\alpha t} w_i \right\rfloor; t \right) \right\}. \quad (36)$$

The boundary condition is that $J_{n+1}(q; t) = 0$ if $q \leq \lceil \frac{n}{\alpha} \rceil$, whereas $J_{n+1}(q; t) = -\infty$ otherwise, so we find a solution \mathbf{w} to problem (33) that satisfies $\sum_{i \in N} \lfloor \frac{n}{\alpha t} w_i \rfloor \leq \lceil \frac{n}{\alpha} \rceil$.

Noting that $\lfloor \frac{n}{\alpha t} w_i \rfloor$ is an integer, the state variable in the dynamic program takes on integer values. Furthermore, as we move from one decision epoch to the next, the state variable in the dynamic program can only increase. If the state variable exceeds $\lceil \frac{n}{\alpha} \rceil$ at any decision epoch, then we can immediately conclude that the value function takes the value $-\infty$, since the boundary condition of the dynamic program depends on whether the state variable exceeds $\lceil \frac{n}{\alpha} \rceil$. In this case, noting that the state variable takes on integer values, we have $O(\frac{n}{\alpha})$ possible values for the state variable. Since there are $|\overline{\text{Grid}}_i|$ possible actions in each decision epoch, for fixed t , we can solve the dynamic program in (36) in $O\left(\frac{n}{\alpha} \sum_{i \in N} |\overline{\text{Grid}}_i|\right)$ operations. Thus, for fixed t , we can obtain an optimal solution to problem (33) in the same number of operations.

As discussed just after (32), it is enough to focus on $t \in [\sum_{i \in N} L_i, \sum_{i \in N} U_i]$. Letting \mathbf{w}^* be an optimal solution to problem (31), by the definition of Grid , there exists $\hat{t} \in \text{Grid} \cap [\sum_{i \in N} L_i, \sum_{i \in N} U_i]$ such that $\sum_{i \in N} w_i^* \leq \hat{t} \leq (1 + \alpha) \sum_{i \in N} w_i^*$. By Lemma K.3, if we solve problem (33) with $t = \hat{t}$, then we get a $\frac{1}{(1+2\alpha)(1+\alpha)^2}$ -approximate solution to the pricing problem. So, if we solve problem (33) with $t = \hat{t}$ for each $\hat{t} \in \text{Grid} \cap [\sum_{i \in N} L_i, \sum_{i \in N} U_i]$ and pick the solution that provides the largest expected revenue, then we get a $\frac{1}{(1+2\alpha)(1+\alpha)^2}$ -approximate solution to the pricing problem. We have $O\left(\frac{\log(\sum_{i \in N} U_i / \sum_{i \in N} L_i)}{\log(1+\alpha)}\right) = O\left(\frac{1}{\alpha} \log\left(\frac{\sum_{i \in N} U_i}{\sum_{i \in N} L_i}\right)\right)$ points in $\text{Grid} \cap [\sum_{i \in N} L_i, \sum_{i \in N} U_i]$. Thus, noting the number of operations to solve the dynamic program in (36), we can get a $\frac{1}{(1+2\alpha)(1+\alpha)^2}$ -approximate solution to the pricing problem in $O\left(\frac{n}{\alpha^2} \log\left(\frac{\sum_{i \in N} U_i}{\sum_{i \in N} L_i}\right) \sum_{i \in N} |\overline{\text{Grid}}_i|\right)$ operations.

Given $\epsilon \in (0, 1)$, we choose the accuracy parameter in Grid as $\alpha = \epsilon/4$. Since $\alpha = \epsilon/4 \in (0, 1)$ and $\frac{1}{(1+2\alpha)(1+\alpha)^2} \geq (1 - 4\alpha) = 1 - \epsilon$ for any $\alpha \in (0, 1)$, we can use the approach outlined above to obtain

a solution to the pricing problem in (31) such that the expected revenue provided by this solution deviates from the optimal expected revenue by at most a factor of $1 - \epsilon$ and we can obtain this solution in $O\left(\frac{n}{\epsilon^2} \log\left(\frac{\sum_{i \in N} U_i}{\sum_{i \in N} L_i}\right) \sum_{i \in N} |\overline{\text{Grid}}_i|\right)$ operations. By the definition of U_i and L_i in Lemma K.1, we have $\frac{U_i}{L_i} = O(\exp(\sum_{i \in N} e^{\mu_i} + \sum_{i \in N} e^{\gamma_i}))$ and $\frac{\sum_{i \in N} U_i}{\sum_{i \in N} L_i} = O(\exp(\sum_{i \in N} e^{\mu_i} + \sum_{i \in N} e^{\gamma_i}))$, in which case, the first equality yields $|\overline{\text{Grid}}_i| = O\left(\frac{\log(U_i/L_i)}{\log(1+\alpha)}\right) = O\left(\frac{1}{\epsilon} (\sum_{i \in N} e^{\mu_i} + \sum_{i \in N} e^{\gamma_i})\right)$. So, we can obtain a $(1 - \epsilon)$ -approximate solution to the pricing problem in $O\left(\frac{n}{\epsilon^2} \log\left(\frac{\sum_{i \in N} U_i}{\sum_{i \in N} L_i}\right) \sum_{i \in N} |\overline{\text{Grid}}_i|\right) = O\left(\frac{n^2}{\epsilon^3} (\sum_{i \in N} e^{\mu_i} + \sum_{i \in N} e^{\gamma_i})^2\right)$ operations. If e^{μ_i} and e^{γ_i} are uniformly bounded by a fixed constant, then this running time is polynomial in the size of the input and in $1/\epsilon$. If e^{μ_i} and e^{γ_i} are parts of the problem input, then this running time is pseudo-polynomial in the size of the input and in $1/\epsilon$. In either case, the discussion in this section shows how we can focus on a finite number of solutions that scale polynomially in the number of products and $1/\epsilon$ to obtain a $(1 - \epsilon)$ -approximate solution to the pricing problem. In the rest of this section, we give a proof for Lemma K.1.

Proof of Lemma K.1:

Noting that $w_i = e^{\gamma_i - \nu p_i}$ and $p_i \geq 0$, we have $w_i \leq e^{\gamma_i}$, so $w_i \leq U_i$. In the rest of the proof, we focus on showing that there exists an optimal solution \mathbf{w}^* that satisfies $w_i^* \geq L_i$ for all $i \in N$. For notational brevity, let $g_i(w_i) = \frac{w_i}{1 + \sigma_i w_i} (\gamma_i - \log w_i)$ and $h(\mathbf{w}) = \frac{\sum_{i \in N} w_i (\gamma_i - \log w_i)}{1 + \sum_{i \in N} w_i}$, in which case, the objective function of problem (31) is $\psi(\mathbf{w}) = \lambda \sum_{i \in N} \alpha_i \sigma_i g_i(w_i) + (1 - \lambda) h(\mathbf{w})$. We show that if $w_i < L_i$ for some $i \in N$, then $\frac{\partial \psi(\mathbf{w})}{\partial w_i} > 0$, which implies that we can increase the value of $\psi(\mathbf{w})$ by increasing w_i infinitesimally. Thus, we cannot have $w_i^* < L_i$ in an optimal solution \mathbf{w}^* to problem (31) and the desired result follows. First, we claim that if $w_i < L_i$, then $\frac{dg_i(w_i)}{dw_i} > 0$. In particular, differentiating $g_i(\cdot)$, we have $\frac{dg_i(w_i)}{dw_i} = \frac{1}{1 + \sigma_i w_i} \left(\frac{\gamma_i - \log w_i}{1 + \sigma_i w_i} - 1 \right)$. Since $w_i < L_i$, noting the definition L_i , we get $w_i < e^{\gamma_i - 1} / \exp(\sum_{j \in N} e^{\mu_j}) = e^{\gamma_i - 1} / \exp(\sum_{j \in N} \sigma_j e^{\gamma_j})$, where the equality uses the definition of σ_i . Thus, we have $\log w_i < \gamma_i - 1 - \sum_{j \in N} \sigma_j e^{\gamma_j} \leq \gamma_i - 1 - \sigma_i e^{\gamma_i}$. In this case, we obtain

$$\frac{dg_i(w_i)}{dw_i} = \frac{1}{1 + \sigma_i w_i} \left(\frac{\gamma_i - \log w_i}{1 + \sigma_i w_i} - 1 \right) \stackrel{(a)}{>} \frac{1}{1 + \sigma_i w_i} \left(\frac{1 + \sigma_i e^{\gamma_i}}{1 + \sigma_i w_i} - 1 \right) \stackrel{(b)}{\geq} 0,$$

where (a) holds since $\log w_i < \gamma_i - 1 - \sigma_i e^{\gamma_i}$ and (b) holds noting the fact that $w_i \leq e^{\gamma_i}$. Thus, the chain of inequalities above establishes that the first claim holds.

Second, we claim that if $w_i < L_i$, then $\frac{\partial h(\mathbf{w})}{\partial w_i} > 0$. Since $w_i < L_i$, by the definition of L_i , we get $w_i < e^{\gamma_i - 1} / \exp(\sum_{j \in N} e^{\gamma_j})$. Differentiating $h(\cdot)$, we also have

$$\frac{\partial h(\mathbf{w})}{\partial w_i} = \frac{(\gamma_i - 1 - \log w_i) \left(1 + \sum_{j \in N} w_j\right) - \sum_{j \in N} w_j (\gamma_j - \log w_j)}{\left(1 + \sum_{j \in N} w_j\right)^2}. \quad (37)$$

Since $w_i < e^{\gamma_i - 1} / \exp(\sum_{j \in N} e^{\gamma_j})$, we have $\gamma_i - 1 - \log w_i > \sum_{j \in N} e^{\gamma_j}$. Thus, we get the inequality $(\gamma_i - 1 - \log w_i) (1 + \sum_{j \in N} w_j) > (\sum_{j \in N} e^{\gamma_j}) (1 + \sum_{j \in N} w_j) > \sum_{j \in N} e^{\gamma_j}$. On the other hand, the

maximum value of the function $f(x) = x(\gamma_j - \log x)$ over the interval $[0, \infty)$ is $e^{\gamma_j - 1}$, so have $w_j(\gamma_j - \log w_j) < e^{\gamma_j - 1}$. Thus, noting the last inequality in this paragraph, it follows that

$$(\gamma_i - 1 - \log w_i) \left(1 + \sum_{j \in N} w_j\right) - \sum_{j \in N} w_j(\gamma_j - \log w_j) > \sum_{j \in N} e^{\gamma_j} - \sum_{j \in N} e^{\gamma_j - 1} > 0,$$

in which case, by (37), we have $\frac{\partial h(\mathbf{w})}{\partial w_i} > 0$, so the second claim holds. By the two claims, we get if $w_i < L_i$ for some $i \in N$, then $\frac{\partial \psi(\mathbf{w})}{\partial w_i} > 0$. ■

Appendix L: Computational Experiments under Censored Demands

In this section, we modify our expectation-maximization algorithm to handle demand censorship and give computational experiments under censored demands.

L.1 Expectation-Maximization Algorithm

Demand censorship refers to the fact that if we did not have sales over a period of time, then we may not necessarily know whether no customer arrivals occurred or the arriving customers did not purchase anything. We divide the data collection horizon into small enough time periods that there is at most one customer arrival at each time period. We use $\mathcal{H} = \{(S_t, i_t) : t = 1, \dots, \tau\}$ to capture the purchase history observed over the data collection horizon, where τ is the number of time periods in the data collection horizon, S_t is the assortment of products offered during time period t and i_t is the product purchased, if any, during time period t . If there was no purchase during time period t , then $i_t = 0$. The parameters that we need to estimate are $(\beta, \lambda, \boldsymbol{\theta}, \mathbf{v})$, where the parameters $(\lambda, \boldsymbol{\theta}, \mathbf{v})$ of our mixture model are as in Section 7 and β is the probability that we have a customer arrival during a time period. We make use of the following observation to construct our expectation-maximization algorithm. In addition to the purchase history $\mathcal{H} = \{(S_t, i_t) : t = 1, \dots, \tau\}$, if we knew whether there was an arrival during each time period, as well as the segment of the arriving customer, if any, then we could estimate the arrival probability β separately from the parameters $(\lambda, \boldsymbol{\theta}, \mathbf{v})$ of our mixture model. Moreover, we could separately fit independent demand and multinomial logit models to the purchases from the two segments.

For the moment, assume that we do have access to whether there is a customer arrival during each time period and the segment of the customer arriving during each time period. For $(a_t, y_t) \in \{0, 1\}^2$ for each $t = 1, \dots, \tau$, we use $\mathcal{C}\{(a_t, y_t, S_t, i_t) : t = 1, \dots, \tau\}$ to capture the complete purchase history, where $a_t = 1$ if and only if we had a customer arrival during time period t and $y_t = 1$ if and only if we had a customer arrival during time period t and this customer was in the independent demand segment. Note that $a_t \geq y_t$, so $a_t - y_t \in \{0, 1\}$ and $a_t - y_t = 1$ if and only if we had a customer arrival during time period t and this customer was in the multinomial logit segment. If we had

access to the complete purchase history $\mathcal{C} = \{(a_t, y_t, S_t, i_t) : t = 1, \dots, \tau\}$, then we could estimate the parameters $(\beta, \lambda, \boldsymbol{\theta}, \mathbf{v})$ by maximizing the likelihood function

$$L^{\text{Cens}}(\beta, \lambda, \boldsymbol{\theta}, \mathbf{v}; \mathcal{C}) = \prod_{t=1}^{\tau} \left\{ (1 - \beta)^{1-a_t} \left(\beta \lambda \prod_{i \in S_t} \theta_i^{\mathbf{1}(i_t=i)} \left(1 - \sum_{i \in S_t} \theta_i \right)^{\mathbf{1}(i_t=0)} \right)^{y_t} \times \left(\beta (1 - \lambda) \frac{\prod_{i \in S_t} v_i^{\mathbf{1}(i_t=i)} \mathbf{1}(i_t=0)}{1 + V(S_t)} \right)^{a_t - y_t} \right\}. \quad (38)$$

In the likelihood expression above, if $1 - a_t = 1$, then we do not have an arrival during time period t , so the likelihood of this event is $1 - \beta$. If $y_t = 1$, then we have a customer arrival during time period t and this customer is in the independent demand segment, so the likelihood of this event is $\beta \lambda$. Also, if the customer arriving during time period t is in the independent demand segment, then she purchases product $i \in S_t$ with probability θ_i , whereas she leaves without a purchase with probability $1 - \sum_{i \in S_t} \theta_i$. If, however, $a_t - y_t = 1$, then we have a customer arrival during time period t and this customer is in the multinomial logit segment, so the likelihood of this event is $\beta (1 - \lambda)$. Also, if the customer arriving during time period t is in the multinomial logit segment, then she purchases product $i \in S_t$ with probability $\frac{v_i}{1 + V(S_t)}$, whereas she leaves without a purchase with probability $\frac{1}{1 + V(S_t)}$, resulting in the likelihood function in (38). We will need three random variables. We use the random variable A with support $\{0, 1\}$ to capture whether there is a customer arrival during a generic time period, where $A = 1$ if and only if there is a customer arrival. We use the random variable Y with support $\{0, 1\}$ to capture whether there is a customer arrival during a generic time period and the segment of the arriving customer, where $Y = 1$ if and only if there is a customer arrival and this customer is in the independent demand segment. Lastly, we use the random variable $P(S)$ with support $S \cup \{0\}$ to capture the product sold during a generic time period, where $P(S) = i$ if and only if there is a sale for product $i \in S$. If $P(S) = 0$, then we may have a customer arrival and this customer left without a purchase or no customer arrival at all. If the parameters of our mixture model are $(\beta, \lambda, \boldsymbol{\theta}, \mathbf{v})$, then A is Bernoulli with parameter β , Y is Bernoulli with parameter $\beta \lambda$ and $P(S)$ takes value $i \in S$ with probability $\beta \lambda \theta_i + \beta (1 - \lambda) \frac{v_i}{1 + V(S)}$.

Description of the Expectation-Maximization Algorithm:

At iteration ℓ of our expectation-maximization algorithm, we have the current parameter estimates $(\beta^\ell, \lambda^\ell, \boldsymbol{\theta}^\ell, \mathbf{v}^\ell)$. Letting $\{(\bar{a}_t^\ell, \bar{y}_t^\ell) : t = 1, \dots, \tau\}$ be the estimates of $\{(a_t^\ell, y_t^\ell) : t = 1, \dots, \tau\}$ at iteration ℓ , we compute \bar{a}_t^ℓ as the expectation of A conditional on the fact that the arrival and choice process during time period t is governed by our mixture model with parameters $(\beta^\ell, \lambda^\ell, \boldsymbol{\theta}^\ell, \mathbf{v}^\ell)$ and the product purchased, if any, during this time period is i_t . We compute \bar{y}_t^ℓ as the expectation of Y conditional on the same information. In this way, we obtain the estimated complete

purchase history $\mathcal{C}^\ell = \{(\bar{a}_t^\ell, \bar{y}_t^\ell, S_t, i_t) : t = 1, \dots, \tau\}$ at iteration ℓ . The parameters of our mixture model take values in the set $\mathcal{Q} = \{(\beta, \lambda, \boldsymbol{\theta}, \mathbf{v}) \in [0, 1]^{2+n} \times [0, B]^n : \sum_{i \in N} \theta_i \leq 1\}$. We maximize the likelihood function $L^{\text{Cens}}(\beta, \lambda, \boldsymbol{\theta}, \mathbf{v}; \mathcal{C}^\ell)$ subject to the constraint that $(\beta, \lambda, \boldsymbol{\theta}, \mathbf{v}) \in \mathcal{Q}$ to obtain the parameter estimates $(\beta^{\ell+1}, \lambda^{\ell+1}, \boldsymbol{\theta}^{\ell+1}, \mathbf{v}^{\ell+1})$ at the next iteration. Thus, during the course of our expectation-maximization algorithm, given that the arrival and choice process is governed by our mixture model with some parameters $(\beta, \lambda, \boldsymbol{\theta}, \mathbf{v})$, we need to compute conditional expectations of the form $\mathbb{E}\{A | P(S) = i\}$ and $\mathbb{E}\{Y | P(S) = i\}$. Noting that the support of the random variables A and Y are $\{0, 1\}$, the last two conditional expectations are equal to $\mathbb{P}\{A = 1 | P(S) = i\}$ and $\mathbb{P}\{Y = 1 | P(S) = i\}$, in which case, we can compute them as

$$\mathbb{P}\{A = 1 | P(S) = i\} = \frac{\mathbb{P}\{A = 1, P(S) = i\}}{\mathbb{P}\{P(S) = i\}} = \begin{cases} 1 & \text{if } i \neq 0 \\ \frac{\beta \left(\lambda (1 - \sum_{i \in S} \theta_i) + (1 - \lambda) \frac{1}{1+V(S)} \right)}{(1 - \beta) + \beta \left(\lambda (1 - \sum_{i \in S} \theta_i) + (1 - \lambda) \frac{1}{1+V(S)} \right)} & \text{if } i = 0, \end{cases}$$

$$\mathbb{P}\{Y = 1 | P(S) = i\} = \frac{\mathbb{P}\{Y = 1, P(S) = i\}}{\mathbb{P}\{P(S) = i\}} = \begin{cases} \frac{\beta \lambda \theta_i}{\beta \left(\lambda \theta_i + (1 - \lambda) \frac{v_i}{1+V(S)} \right)} & \text{if } i \neq 0 \\ \frac{\beta \lambda (1 - \sum_{i \in S} \theta_i)}{(1 - \beta) + \beta \left(\lambda (1 - \sum_{i \in S} \theta_i) + (1 - \lambda) \frac{1}{1+V(S)} \right)} & \text{if } i = 0. \end{cases}$$

In the first case of the first equality, having $P(S) = i$ with $i \neq 0$ means that we have a purchase for product i , so a customer must have arrived. Thus, if $i \neq 0$, then $\mathbb{P}\{A = 1, P(S) = i\} = \mathbb{P}\{P(S) = i\}$. In the second case, having $P(S) = 0$ means that we do not have a purchase, which happens when either a customer does not arrive or a customer arrives and she leaves without a purchase. In the first case of the second equality, having $Y = 1$ and $P(S) = i$ with $i \neq 0$ means that a customer arrives, she is in the independent demand segment and she purchases product i , which happens with probability $\beta \lambda \theta_i$. Also, having $P(S) = i$ with $i \neq 0$ means that we have a purchase for product i , which happens when there is a customer arrival and she purchases product i regardless of her segment. We can interpret the second case of the second equality similarly.

During the course of our expectation-maximization algorithm, for some complete purchase history \mathcal{C} , we also need to maximize the likelihood function $L^{\text{Cens}}(\beta, \lambda, \boldsymbol{\theta}, \mathbf{v}; \mathcal{C})$ subject to the constraint that $(\beta, \lambda, \boldsymbol{\theta}, \mathbf{v}) \in \mathcal{Q}$. We can express the logarithm of the last likelihood function as $\log L^{\text{Cens}}(\beta, \lambda, \boldsymbol{\theta}, \mathbf{v}; \mathcal{C}) = L_1^{\text{Cens}}(\beta; \mathcal{C}) + L_2^{\text{Cens}}(\lambda; \mathcal{C}) + L_3^{\text{Cens}}(\boldsymbol{\theta}; \mathcal{C}) + L_4^{\text{Cens}}(\mathbf{v}; \mathcal{C})$, where $L_2^{\text{Cens}}(\lambda; \mathcal{C})$, $L_3^{\text{Cens}}(\boldsymbol{\theta}; \mathcal{C})$ and $L_4^{\text{Cens}}(\mathbf{v}; \mathcal{C})$ have, respectively, the same form as $L_1(\lambda; \mathcal{C})$, $L_2(\boldsymbol{\theta}; \mathcal{C})$ and $L_3(\mathbf{v}; \mathcal{C})$ in Section 7.1 after replacing each occurrence of z_t with y_t and each occurrence of $1 - z_t$ with $a_t - y_t$. Lastly, $L_1^{\text{Cens}}(\beta; \mathcal{C})$ has the form $L_1^{\text{Cens}}(\beta; \mathcal{C}) = \sum_{t=1}^{\tau} \{a_t \log \beta + (1 - a_t) \log(1 - \beta)\}$. Putting

the discussion in the last two paragraphs together, below is a step-by-step description of our expectation-maximization algorithm under censored demands.

Expectation-Maximization Algorithm Step-by-Step under Censored Demands:

The input is $\{(S_t, i_t) : t = 1, \dots, \tau\}$, where S_t is the assortment offered and i_t is the product purchased during time period t . If there was no purchase during time period t , then $i_t = 0$.

Step 1. Choose the initial parameter estimates $(\beta^1, \lambda^1, \boldsymbol{\theta}^1, \mathbf{v}^1)$, such that $\beta^1 \in [0, 1]$, $\lambda^1 \in [0, 1]$, $\boldsymbol{\theta}^1 \in [0, 1]^n$, $\sum_{i \in N} \theta_i^1 \leq 1$ and $\mathbf{v}^1 \in [0, B]^n$. Initialize the iteration counter by setting $\ell = 1$.

Step 2. (Compute the Conditional Expectations) For each time period $t = 1, \dots, \tau$ in the purchase history, compute $\bar{a}_t^\ell \in [0, 1]$ and $\bar{y}_t^\ell \in [0, 1]$ as

$$\bar{a}_t^\ell = \begin{cases} 1 & \text{if } i_t \neq 0 \\ \frac{\beta^\ell \left(\lambda^\ell (1 - \sum_{i \in S_t} \theta_i^\ell) + (1 - \lambda^\ell) \frac{1}{1 + \sum_{j \in S_t} v_j^\ell} \right)}{(1 - \beta^\ell) + \beta^\ell \left(\lambda^\ell (1 - \sum_{i \in S_t} \theta_i^\ell) + (1 - \lambda^\ell) \frac{1}{1 + \sum_{j \in S_t} v_j^\ell} \right)} & \text{if } i_t = 0, \end{cases} \quad (39)$$

$$\bar{y}_t^\ell = \begin{cases} \frac{\beta^\ell \lambda^\ell \theta_{i_t}^\ell}{\beta^\ell \left(\lambda^\ell \theta_{i_t}^\ell + (1 - \lambda^\ell) \frac{v_{i_t}^\ell}{1 + \sum_{j \in S_t} v_j^\ell} \right)} & \text{if } i_t \neq 0 \\ \frac{\beta^\ell \lambda^\ell (1 - \sum_{i \in S_t} \theta_i^\ell)}{(1 - \beta^\ell) + \beta^\ell \left(\lambda^\ell (1 - \sum_{i \in S_t} \theta_i^\ell) + (1 - \lambda^\ell) \frac{1}{1 + \sum_{j \in S_t} v_j^\ell} \right)} & \text{if } i_t = 0. \end{cases} \quad (40)$$

Step 3. (Maximize the Likelihood) Compute the parameters $(\beta^{\ell+1}, \lambda^{\ell+1}, \boldsymbol{\theta}^{\ell+1}, \mathbf{v}^{\ell+1})$ at the next iteration by solving the problems

$$\begin{aligned} \beta^{\ell+1} &= \arg \max_{\beta \in [0, 1]} \left\{ \sum_{t=1}^{\tau} \left\{ \bar{a}_t^\ell \log \beta + (1 - \bar{a}_t^\ell) \log(1 - \beta) \right\} \right\} \\ \lambda^{\ell+1} &= \arg \max_{\lambda \in [0, 1]} \left\{ \sum_{t=1}^{\tau} \left\{ \bar{y}_t^\ell \log \lambda + (\bar{a}_t^\ell - \bar{y}_t^\ell) \log(1 - \lambda) \right\} \right\} \\ \boldsymbol{\theta}^{\ell+1} &= \arg \max_{\boldsymbol{\theta} \in [0, 1]^n} \left\{ \sum_{t=1}^{\tau} \bar{y}_t^\ell \left\{ \sum_{i \in S_t} \mathbf{1}(i = i_t) \log \theta_i + \mathbf{1}(i_t = 0) \log \left(1 - \sum_{i \in S_t} \theta_i \right) \right\} : \sum_{i \in N} \theta_i \leq 1 \right\} \\ \mathbf{v}^{\ell+1} &= \arg \max_{\mathbf{v} \in [0, B]^n} \left\{ \sum_{t=1}^{\tau} (\bar{a}_t^\ell - \bar{y}_t^\ell) \left\{ \sum_{i \in S_t} \mathbf{1}(i = i_t) \log v_i - \log \left(1 + \sum_{j \in S_t} v_j \right) \right\} \right\}. \end{aligned}$$

Step 4. For fixed $\epsilon > 0$, if $\max\{|\beta^{\ell+1} - \beta^\ell|, |\lambda^{\ell+1} - \lambda^\ell|, \|\boldsymbol{\theta}^{\ell+1} - \boldsymbol{\theta}^\ell\|, \|\mathbf{v}^{\ell+1} - \mathbf{v}^\ell\|\} \leq \epsilon$, then stop. Otherwise, increase the iteration counter ℓ by one and go to Step 2.

By the discussion in Section 7.1, the last three optimization problems in Step 3 have concave objective functions and linear constraints. The first optimization problem in Step 3 share the

same characteristics. Next, we discuss the monotonicity of the likelihood function at the iterates generated by our expectation-maximization algorithm.

Monotonicity of the Likelihood Function:

Using the past purchase history $\mathcal{H} = \{(S_t, i_t) : t = 1, \dots, \tau\}$ that we have available to estimate the parameters of our mixture model, we have the likelihood function

$$\begin{aligned} \bar{L}^{\text{Cens}}(\beta, \lambda, \boldsymbol{\theta}, \mathbf{v}; \mathcal{H}) = & \prod_{t=1}^{\tau} \left\{ \prod_{i \in S_t} \left(\beta \lambda \theta_i + \beta (1 - \lambda) \frac{v_i}{1 + V(S_t)} \right)^{\mathbf{1}(i_t=i)} \right. \\ & \left. \times \left((1 - \beta) + \beta \lambda \left(1 - \sum_{i \in S_t} \theta_i \right) + \beta (1 - \lambda) \frac{1}{1 + V(S_t)} \right)^{\mathbf{1}(i_t=0)} \right\}, \quad (41) \end{aligned}$$

where we use the fact that if we have no purchase during time period t so that $i_t = 0$, then either there was no customer arrival during time period t , which happens with probability $1 - \beta$, or there was a customer arrival during time period t , but this customer left without a purchase, which happens with probability $\beta \lambda (1 - \sum_{i \in S_t} \theta_i) + \beta (1 - \lambda) \frac{1}{1 + V(S_t)}$. In the next proposition, we argue that the sequence of parameter estimates generated by our expectation-maximization algorithm monotonically increases the likelihood function in (41).

Proposition L.1 *Letting the sequence $\{(\beta^\ell, \lambda^\ell, \boldsymbol{\theta}^\ell, \mathbf{v}^\ell) : \ell = 1, 2, \dots\}$ be generated by the expectation-maximization algorithm, we have $\bar{L}^{\text{Cens}}(\beta^{\ell+1}, \lambda^{\ell+1}, \boldsymbol{\theta}^{\ell+1}, \mathbf{v}^{\ell+1}; \mathcal{H}) \geq \bar{L}^{\text{Cens}}(\beta^\ell, \lambda^\ell, \boldsymbol{\theta}^\ell, \mathbf{v}^\ell; \mathcal{H})$ for all $\ell = 1, 2, \dots$*

Proof. The result follows by verifying the conditions of Theorem 2 in Wu (1983). The likelihood function in (41) and the set \mathcal{Q} of possible values for the parameters of our mixture model satisfy properties (5)-(7) in Wu (1983). By (39)-(40), \bar{a}_t^ℓ and \bar{y}_t^ℓ are continuous in the parameters $(\beta^\ell, \lambda^\ell, \boldsymbol{\theta}^\ell, \mathbf{v}^\ell)$. Furthermore, the likelihood function in (38) is continuous in a_t and y_t . Thus, the likelihood function in (38) satisfies the continuity property (10) in Wu (1983). Therefore, the result follows from Theorem 2 in Wu (1983). ■

Thus, the monotonicity of the likelihood function at the iterates of our expectation-maximization algorithm follows by verifying the conditions of a classical result in Wu (1983).

L.2 Experimental Results

We give computational experiments under censored demands to check the ability of our mixture model to predict customer choices and to identify profitable assortments.

Experimental Setup. Our experimental setup closely follows the one in Section 7.2. We use the dataset from Kamishima (2018) exactly as in that section to come up with the ground choice

model. Recall that the parameter ψ controls the length of the preference lists in the ground choice model. Once we come up with the ground choice model, we generate the purchase histories of customers making purchases according to the ground choice model. In the past purchase history $\{(S_t, i_t) : t = 1, \dots, \tau\}$, τ corresponds to the number of time periods, S_t corresponds to the assortment offered during time period t and i_t is the product purchased, if any, during time period t . To generate the past purchase history, we include each product in the assortment S_t with probability 0.5. During each time period t , no customer arrival occurs with probability 0.1. If a customer arrival does occur, then we sample the product purchased by the customer according to the ground choice model. During each time period t , we may not have a purchase because a customer did not arrive or the arriving customer left without making a purchase. We fit our mixture choice model, multinomial logit model and independent demand model to the past purchase histories that we generate. The Python code in Berbeglia et al. (2021) to fit exponential and Markov chain choice models is not designed to work with censored demands and we do not provide comparisons with the exponential and Markov chain choice models. We vary ψ over $\{0.5, 0.6, \dots, 1.0\}$ to obtain different lengths for the preference lists in the ground choice model. We vary τ over $\{1250, 2500, 5000\}$ to capture different levels of data availability in the training data.

Comparing Out-of-Sample Log-Likelihoods. We use MIX to refer to our fitted mixture choice model, MNL to refer to the fitted pure multinomial logit model and IDM to refer to the fitted pure independent demand model. In Table EC.5, we compare the out-of-sample log-likelihoods of MIX, MNL and IDM. We use the same approach in Section 7.3 to compare the out-of-sample log-likelihoods under censored demands. The top, middle and bottom blocks of the table correspond to the values of $\tau \in \{1250, 2500, 5000\}$. Recall that we repeat our computational experiments 50 times, each replication involving a different ground choice model, as well as training and testing data. The first column gives the value of the parameter ψ controlling the length of the preference lists. The second, third and fourth columns give the average out-of-sample log-likelihoods obtained by MIX, MNL and IDM. The next three columns compare the performance of MIX and MNL, where the fifth column gives the percent gap between the average out-of-sample log-likelihoods of MIX and MNL, the sixth column gives the number of replications in which the out-of-sample log-likelihood of MIX exceeds that of MNL and the seventh column gives the number of replications in which the outcome is reversed. The last three columns compare the performance of MIX and IDM similarly. Out-of-sample log-likelihoods of MIX are noticeably better than those of MNL and IDM. These gaps in log-likelihoods transfer to gaps in expected revenues, as we demonstrate shortly.

Comparing Expected Revenues. In Table EC.6, we compare the expected revenue performance of MIX, MNL and IDM. We use the same approach in Section 7.3 to compare the expected revenues.

$\tau = 1250$									
ψ	Out-of-Sample Log-Likelihood			MIX vs. MNL			MIX vs. IDM		
	MIX	MNL	IDM	Perc. Gap	MIX Btr.	MNL Btr.	Perc. Gap	MIX Btr.	IDM Btr.
0.5	-3,709	-3,712	-3,846	0.09	40	10	3.57	50	0
0.6	-3,646	-3,652	-3,741	0.15	42	8	2.52	50	0
0.7	-3,589	-3,597	-3,655	0.23	44	6	1.82	50	0
0.8	-3,532	-3,542	-3,574	0.29	45	5	1.18	50	0
0.9	-3,486	-3,498	-3,515	0.33	49	1	0.83	50	0
1.0	-3,443	-3,456	-3,463	0.39	49	1	0.58	48	2
Avg.	-3568	-3576	-3632	0.24	44.8	5.2	1.79	49.7	0.3

$\tau = 2500$									
ψ	Out-of-Sample Log-Likelihood			MIX vs. MNL			MIX vs. IDM		
	MIX	MNL	IDM	Perc. Gap	MIX Btr.	MNL Btr.	Perc. Gap	MIX Btr.	IDM Btr.
0.5	-3,704	-3,710	-3,844	0.15	44	6	3.63	50	0
0.6	-3,641	-3,649	-3,738	0.22	47	3	2.61	50	0
0.7	-3,584	-3,594	-3,653	0.28	47	3	1.89	50	0
0.8	-3,527	-3,540	-3,572	0.36	49	1	1.26	50	0
0.9	-3,482	-3,496	-3,514	0.40	50	0	0.91	50	0
1.0	-3,437	-3,453	-3,460	0.47	50	0	0.67	50	0
Avg.	-3562	-3574	-3630	0.31	47.8	2.2	1.86	50.0	0.0

$\tau = 5000$									
ψ	Out-of-Sample Log-Likelihood			MIX vs. MNL			MIX vs. IDM		
	MIX	MNL	IDM	Perc. Gap	MIX Btr.	MNL Btr.	Perc. Gap	MIX Btr.	IDM Btr.
0.5	-3,701	-3,708	-3,842	0.19	47	3	3.67	50	0
0.6	-3,638	-3,648	-3,737	0.26	49	1	2.65	50	0
0.7	-3,581	-3,593	-3,652	0.33	49	1	1.95	50	0
0.8	-3,524	-3,538	-3,571	0.41	50	0	1.32	50	0
0.9	-3,480	-3,495	-3,513	0.43	50	0	0.96	50	0
1.0	-3,435	-3,452	-3,460	0.50	50	0	0.71	50	0
Avg.	-3560	-3572	-3629	0.35	49.2	0.8	1.91	50.0	0.0

Table EC.5 Out-of-sample log-likelihoods of the fitted choice models under censored demands.

The top, middle and bottom blocks of the table correspond to the values of $\tau \in \{1250, 2500, 5000\}$. The first column gives the value of the parameter ψ controlling the length of the preference lists in the ground choice model. The second, third and fourth columns give the average expected revenues obtained by the assortments that are computed under the assumption that the choices of the customers are governed by the fitted MIX, MNL and IDM. Recall that the averages are computed over the 50 replications. The next three columns compare the performance of MIX and MNL, where the fifth column gives the percent gap between the average expected revenues of MIX and MNL, the sixth column gives the number of replications in which the expected revenue of MIX exceeds that of MNL and the seventh column gives the number of replications in which the outcome is reversed. The last three columns compare the performance of MIX and IDM similarly. If customers choose under the independent demand model, then it is always optimal to offer all products. Thus, the expected revenue performance of the assortments obtained by IDM does not change from one replication to the next. Our results in the table indicate that the assortments obtained by using the fitted MIX perform noticeably better than those obtained by using the fitted MNL and IDM. The noticeably superior performance of MIX over MNL and IDM is largely consistent across the 50 replications under each value of the parameter ψ .

$\tau = 1250$									
ψ	Expected Revenue			MIX vs. MNL			MIX vs. IDM		
	MIX	MNL	IDM	Perc. Gap	MIX Btr.	MNL Btr.	Perc. Gap	MIX Btr.	IDM Btr.
0.5	5.821	5.803	5.503	0.32	33	17	5.77	50	0
0.6	5.713	5.655	5.503	1.02	48	2	3.80	50	0
0.7	5.641	5.535	5.503	1.92	50	0	2.50	50	0
0.8	5.587	5.441	5.503	2.68	50	0	1.51	49	1
0.9	5.551	5.362	5.503	3.52	50	0	0.85	48	2
1.0	5.527	5.292	5.503	4.45	50	0	0.43	47	3
Avg.	5.640	5.515	5.503	2.27	46.8	3.2	2.48	49.0	1.0

$\tau = 2500$									
ψ	Expected Revenue			MIX vs. MNL			MIX vs. IDM		
	MIX	MNL	IDM	Perc. Gap	MIX Btr.	MNL Btr.	Perc. Gap	MIX Btr.	IDM Btr.
0.5	5.842	5.803	5.503	0.67	46	4	6.15	50	0
0.6	5.737	5.657	5.503	1.43	50	0	4.25	50	0
0.7	5.658	5.537	5.503	2.19	50	0	2.81	50	0
0.8	5.599	5.441	5.503	2.90	50	0	1.72	50	0
0.9	5.561	5.364	5.503	3.67	50	0	1.04	50	0
1.0	5.536	5.294	5.503	4.57	50	0	0.59	50	0
Avg.	5.655	5.516	5.503	2.53	49.3	0.7	2.76	50.0	0.0

$\tau = 5000$									
ψ	Expected Revenue			MIX vs. MNL			MIX vs. IDM		
	MIX	MNL	IDM	Perc. Gap	MIX Btr.	MNL Btr.	Perc. Gap	MIX Btr.	IDM Btr.
0.5	5.861	5.805	5.503	0.96	50	0	6.50	50	0
0.6	5.755	5.658	5.503	1.71	50	0	4.56	50	0
0.7	5.670	5.539	5.503	2.36	50	0	3.02	50	0
0.8	5.608	5.443	5.503	3.04	50	0	1.90	50	0
0.9	5.567	5.365	5.503	3.76	50	0	1.15	50	0
1.0	5.540	5.295	5.503	4.62	50	0	0.66	50	0
Avg.	5.667	5.518	5.503	2.70	50.0	0.0	2.96	50.0	0.0

Table EC.6 Expected revenues obtained by the fitted choice models under censored demands.