Assortment Optimization under the Multinomial Logit Model in the Presence of Endogenous Consideration Sets

Yicheng Bai[1], Jacob Feldman[2], Huseyin Topaloglu[1], Laura Wagner[3]
[1]School of Operations Research and Information Engineering, Cornell Tech, New York, NY 10044, USA
[2] Olin Business School, University of Washington, St Louis, MO 63130, USA
[3] Catolica-Lisbon School of Business and Economics, Palma de Cima, 1649-023 Lisboa, Portugal
yb279@cornell.edu, jbfeldman@wustl.edu, topaloglu@orie.cornell.edu, lwagner@ucp.pt

January 27, 2021

We study assortment optimization problems under a natural variant of the multinomial logit model where the customers form consideration sets by focusing only on a certain number of products that provide the largest utilities. In particular, each customer has a consideration depth, characterizing the number of products that she includes in her consideration set. Given that we offer a certain assortment of products, the choice process of a customer with consideration depth k proceeds as follows. The customer associates random utilities with all of the products as well as the no-purchase option, and drops from consideration all alternatives whose utilities are not within the k largest utilities. Among the remaining alternatives, the customer chooses the available alternative that provides the largest utility. Under the assumption that the utilities follow Gumbel distributions with the same scale parameter, we provide a recursion to compute the choice probabilities. Considering the assortment optimization problem to find the revenue-maximizing assortment of products to offer, we show that the problem is NP-hard. We give a polynomial-time approximation scheme for the assortment optimization problem. Since customers drop products from consideration in our variant of the multinomial logit model, intuitively speaking, our variant captures choosier choice behavior than the standard multinomial logit model. Accordingly, we show that the revenue-maximizing assortment under our variant includes the revenue-maximizing assortment under the standard multinomial logit model, so choosier behavior leads to larger assortments offered to maximize the expected revenue. We conduct computational experiments on both synthetic and real datasets to demonstrate that incorporating consideration sets can significantly improve the ability of the multinomial logit model to predict customer choices.

1. Introduction

Using choice models to capture customer choice behavior has steadily become the common practice in revenue management. Discrete choice models allow us to capture the fact that customers substitute among the offered products, so if a particular product is not offered, then a portion of the customers interested in this product will substitute into a suitable available alternative, whereas another portion will leave without making a purchase. A growing body of literature indicates that incorporating such substitution possibilities into revenue management models can yield significant improvements in the revenues; see, for example, Talluri and van Ryzin (2004) and Vulcano et al. (2010). When picking a choice model to work with, there is a natural tension between using more sophisticated choice models to capture the customer choice process more accurately and using simpler choice models to make operational decisions more efficiently. Incorporating consideration sets within a choice model helps mitigate such a tension by providing an additional layer of modeling flexibility. When working with a consideration set, one postulates that each customer arrives with a, possibly random, consideration set in mind, drops from consideration all offered products that are not in the consideration set, and chooses among the remaining products according to the choice model on hand. Thus, the probabilistic structure guiding the consideration set formation, along with the choice model on hand, specify the choice process of the customers.

We consider a natural variant of the multinomial logit model that incorporates consideration sets. In our variant, each customer has a consideration depth, characterizing the number of products that she is willing to include in her consideration set. Given that we offer a certain assortment of products, a customer with consideration depth k makes a choice within the assortment as follows. The customer associates random utilities with all products, including those not offered, as well as the no-purchase option. The utilities are independent of each other and have a Gumbel distribution with the same scale parameter, as is the case in the standard multinomial logit model. The customer drops from consideration all alternatives whose utilities are not within the largest k. Among the remaining alternatives, the customer chooses the available alternative that provides the largest utility. If none of the alternatives that provide the largest k utilities are available in the offered assortment or the available alternative that provides the largest utility is the no-purchase option, then the customer leaves without making a purchase. Intuitively speaking, our variant models picky customers who are simply not willing to purchase products that are not within their k most preferred products. Such purchasing behavior is common when, for example, shopping for groceries, where a customer leaves without a purchase when she cannot find one of her top two or three cereal or yogurt varieties on the shelf.

Main Contributions: We make contributions in formulating the choice model, solving the associated assortment optimization problem and providing comparative statistics.

Multinomial Logit Model with Consideration Sets. Under the assumption that the consideration depth of a customer follows a general discrete distribution, we give a recursion to compute the choice probability of each product within an assortment (Theorem 2.1). The choice probability of a product under the standard multinomial logit model takes the form of a fraction where the numerator involves the mean utility of the product and the denominator involves the mean utilities of all available alternatives. In contrast, the choice probability of a product under our choice model involves sums of products of fractions. Thus, the choice probabilities under the two choice models are different. Our variant provides a natural approach for incorporating consideration sets into the multinomial logit model. In a general method to incorporate consideration sets, each customer has a, possibly random, consideration set C. If we offer the assortment S of products, then a customer drops from consideration the products that are not in the consideration set C, focusing only on the products in the assortment $S \cap C$, and choosing within this assortment according to the multinomial logit model. In practice, specifying a probabilistic structure on the collection of possible consideration sets and estimating the parameters of this probabilistic structure may be difficult. In our approach, we succinctly characterize the consideration set of a customer with the distribution of a single scalar, which is the consideration depth.

Assortment Optimization and Impact of Consideration Sets. We study the problem of finding the revenue-maximizing assortment of products to offer under our choice model. We show that this assortment optimization problem is NP-hard (Theorem 3.1). Before giving an approximation scheme for the problem, we compare the optimal assortment under our choice model with two edge cases. First, if all customers have consideration depth of infinity, then they do not drop any product from consideration, so our variant becomes equivalent to the standard multinomial logit model. Second, if all customers have consideration depth of one, then they purchase only their most favorite product, if available, so our variant becomes equivalent to the independent demand model. Intuitively speaking, the customers become pickier as their consideration depths decrease, so the customers under our choice model are picker than those under the standard multinomial logit model but less picky than those under the independent demand model. Accordingly, we show that the revenue-maximizing assortment under our choice model includes the one under the standard multinomial logit model, but included by the one under the independent demand model (Theorem 4.3). Thus, as the customers get pickier, we have to offer a larger assortment to maximize the expected revenue. The proof of this result uses a property of independent Gumbel random variables implying that given the top p alternatives, the conditional choice process of the customer among the remaining q alternatives is identical to the unconditional choice process.

Approximation Scheme. We provide a polynomial-time approximation scheme (PTAS) for the assortment optimization problem. Letting n be the number of products available to the retailer and RevOps be the number of operations to compute the expected revenue from any assortment, for any $\epsilon \in (0,1)$, our PTAS finds a $(1-\epsilon)$-approximate solution in $O(\mathsf{RevOps}\ (n/\epsilon)^{O(1/\epsilon^2)})$ operations (Theorem 5.1). In the multinomial logit model, each product is characterized by a preference weight that is a function of its mean utility, characterizing its attractiveness. We refer to the product of the revenue and preference weight of a product as its weight. In our approach, we group the products into product classes so that the weights of the products in each class differ at most by a factor of $1+\epsilon$. Our PTAS is based on two results. First, we show that there exists an ideal assortment that offers products in at most $O(\frac{\log(n/\epsilon)}{\epsilon})$ product classes and yields at least a $(1-\epsilon)^2$ factor of the optimal expected revenue. This ideal assortment is light in the sense that it offers a certain number of products with the smallest preference weights in each product class (Proposition 6.1). Second, identifying the ideal assortment appears to be difficult. Guessing the total weight of the products offered by the ideal assortment in each product class, we construct a collection of $O((n/\epsilon)^{O(1/\epsilon^2)})$ light assortments that is guaranteed to include the ideal assortment (Proposition 6.2). Our approach for the assortment optimization problem under our choice model is drastically different from the approaches under the standard multinomial logit model. Under the standard multinomial logit model, the expected revenue takes the form of a fraction, where both numerators and denominators can be expressed as linear function of the product offer decisions, so we can use standard fractional programming techniques. The form of the choice probability under our choice model is so different from the one under the standard multinomial logit model that traditional fractional programming techniques, as far as we can see, become ineffective.

Prediction Ability with Consideration Sets. We give computational experiments to demonstrate that incorporating consideration sets into the multinomial logit model can improve its ability to predict customer choices and pick more profitable assortments. We use randomly generated synthetic datasets, as well as a dataset on the preferences of diners among sushi varieties. In our synthetic datasets, consideration sets provide an average improvement of 5.75% in the expected revenues when compared with the standard multinomial logit model. The improvement is 0.65% when working with the sushi preferences. The improvements across our multiple datasets are consistent. We also numerically test the practical performance of our PTAS.

Related Literature: The work on incorporating consideration sets is growing. Jagabathula and Rusmevichientong (2017) study parameter estimation and pricing problems when customers build consideration sets by focusing on products whose prices are below a certain threshold. Wang and Sahin (2018) study assortment optimization problem when the customers form consideration sets by trading off the expected utility from the purchase with the search cost. Jagabathula and Vulcano (2018) work with a choice model, where the consideration set of a customer includes the last purchased product and the products in promotion. They test the predictive power of their model on real data. Feldman and Topaloglu (2018) study the assortment optimization problem under the multinomial logit model with multiple customer types, where different customer types have nested consideration sets. Their choice model is a special case of a mixture of multinomial logit models with the number of customer types increasing polynomially with the number of products. They give a fully polynomial-time approximation scheme. Wang (2019) studies the assortment optimization problem for a variant of the multinomial logit model, where the customers drop products from consideration when their mean utilities are away from the largest mean utility of an available product. Here, the choice probabilities closely resemble those under the standard multinomial logit model, so it is possible to build on the fractional programming arguments. Aouad et al. (2019) give an approximation algorithm for an assortment optimization problem where each product is included in the consideration set of customer with a fixed probability. Aouad et al. (2020) consider a choice model where customers arrive with a preference ranking of the products in mind and drop from consideration the products not in the preference ranking. The authors study the assortment optimization problem for preference rankings with special structures.

Under the standard multinomial logit model with a single customer type and no consideration sets, Gallego et al. (2004) and Talluri and van Ryzin (2004) show that the revenue-maximizing assortment includes a certain number of products with the largest revenues. Rusmevichientong et al. (2010), Wang (2012) and Sumida et al. (2020) focus on the case where there are constraints on the offered assortment. Bront et al. (2009), Mendez-Diaz et al. (2014), Rusmevichientong et al. (2014) and Desir and Goyal (2014) study the assortment optimization problem under a mixture of multinomial logit models, where there are multiple customer types, each choosing according to a multinomial logit model with different parameters. The authors show that the problem is NP-hard, give integer programming formulations and approximation schemes. The form of the choice probabilities under our choice model are remarkably different than those under the multinomial logit model with one or multiple customer types, so the approximation schemes in the papers discussed in this paragraph do not extend to our assortment optimization problem. In particular, as far as we are aware, there does not exist a straightforward way of casting our choice model as an equivalent mixture of multinomial logit models. In our literature review, we focus on variants of the multinomial logit model, but refer the reader to Davis et al. (2014), Blanchet et al. (2016), Aouad et al. (2018), Zhang et al. (2020), and Berbeglia and Joret (2020) for representative approaches under various choice models, including the preference list-based, nested logit, paired combinatorial logit and Markov chain choice models.

Organization: In Section 2, we describe our choice model and give a recursion to compute the choice probabilities of the products. In Section 3, we formulate the assortment optimization problem and show that it is NP-hard. In Section 4, we compare the optimal assortment under our choice model with that under the standard multinomial logit model. In Section 5, we overview our PTAS. In Section 6, we flesh out the details of our PTAS by establishing the existence of an ideal light assortment and constructing a collection that includes this ideal light assortment. In Section 7, we numerically test the ability of our choice model to predict customer choices. In Section 8, we check the practical performance of our PTAS. In Section 9, we conclude.

2. Multinomial Logit Model with Endogenous Consideration Sets

We formulate the multinomial logit model with endogenous consideration sets and give an expression for the choice probabilities of the products. The set of products is $N = \{1, \ldots, n\}$. The set of possible consideration depths is $M = \{1, \ldots, m\}$. The utility of product i is given by the random variable U_i, which has the Gumbel distribution with location-scale parameters $(\mu_i, 1)$. Letting $v_i = e^{\mu_i}$, we refer to v_i as the preference weight of product i. We capture the total preference weight of the products in the assortment $S \subseteq N$ by $V(S) = \sum_{i \in S} v_i$. The utility of the no-purchase option is given by the random variable U_0, which has the Gumbel distribution with location-scale parameters $(0, 1)$. Lastly, the consideration depth of a customer is given by the random variable Y taking values in M. We capture the distribution of Y by $\lambda^k = \mathbb{P}\{Y = k\}$, where we have $\sum_{k \in M} \lambda^k = 1$. The random variables $\{U_i : i \in N\}$, U_0 and Y are independent of each other.

Given that we offer the assortment S, a customer with consideration depth k makes her choice as follows. The customer associates utilities with all products, including those not offered, as well as the no-purchase option. The available alternatives are the products in the offered assortment S and the no-purchase option. Starting from the alternative with the largest utility and following the order of decreasing utilities, the customer sequentially considers the top k alternatives with the largest utilities and chooses the first available alternative. If there is no available alternative within the top k alternatives, then the customer leaves without making a purchase. Thus, a customer leaves without making a purchase when none of the k most preferred alternatives is offered or the no-purchase option is the available alternative that provides the largest utility. To derive the choice probabilities of the products, we use two properties of Gumbel random variables.

- **(Maximum of Gumbel Random Variables)** Let X and Y be independent Gumbel random variables with location-scale parameters $(\mu, 1)$ and $(\eta, 1)$. In this case, $\max\{X, Y\}$ is a Gumbel random variable with location-scale parameters $(\log(e^\mu + e^\eta), 1)$. Also, we have

$$\mathbb{P}\{X \geq Y\} = \frac{e^\mu}{e^\mu + e^\eta}.$$

- **(Independence from Top Ranked Choices)** Let $\{X_i : i \in G\}$ be a collection of independent Gumbel random variables, where X_i has the location-scale parameters $(\mu_i, 1)$. For any partition of $\{i_1, \ldots, i_p\}$ and $\{j_1, \ldots, j_q\}$ of G, we have

$$\mathbb{P}\Big\{X_{j_1} \geq \ldots \geq X_{j_q} \,\Big|\, X_{i_1} \geq \ldots \geq X_{i_p} \geq \max\{X_{j_1}, \ldots, X_{j_q}\}\Big\} \;=\; \mathbb{P}\Big\{X_{j_1} \geq \ldots \geq X_{j_q}\Big\}.$$

To interpret the second property, view X_i as the utility of alternative i. By the second property, given that the alternatives $\{i_1, \ldots, i_p\}$ are the top p alternatives in G with utilities satisfying the order $X_{i_1} \geq \ldots \geq X_{i_p}$, the conditional choice process of a customer among the remaining alternatives $\{j_1,\ldots,j_q\}$ is identical to the unconditional choice process among the alternatives $\{j_1,\ldots,j_q\}$. Jagabathula and Vulcano (2018) give a proof of the first property. We show the second property in Appendix A. In the next theorem, we use these properties to give an expression for the choice probabilities. Throughout the paper, we let $S_{-i} = S\setminus\{i\}$ and $a\vee b = \max\{a,b\}$.

Theorem 2.1 (Recursion for Choice Probabilities) *For each $k\in M$ and $S\subseteq N$, setting $B^1(\cdot,\cdot)=1$, define $B^k(S,N)$ recursively as*

$$B^k(S,N) = 1 + \sum_{i\in N\setminus S} \frac{v_i}{1+V(N_{-i})}\, B^{k-1}(S,N_{-i}). \tag{1}$$

Then, given that we offer the assortment S, a customer with consideration depth k purchases product $i\in S$ with probability $\pi_i^k(S) = \frac{v_i}{1+V(N)}\, B^k(S,N)$.

Proof: We use induction over the consideration depth to show the result for any set of products N, $S\subseteq N$ and $k\in M$. For a customer with consideration depth one to purchase product $i\in S$, the utility of product i should be the largest among all utilities, so $\pi_i^1(S) = \mathbb{P}\{U_i \geq U_0 \vee \max_{j\in N_{-i}} U_j\} = \frac{v_i}{1+V(N)}$, where the last equality follows from the first property of Gumbel random variables and the fact that $V(S)=\sum_{i\in S} e^{\mu_i}$. So, since $B^1(\cdot,\cdot)=1$, the result holds for $k=1$. Assuming that the result holds for consideration depth of $k-1$, we show that the result holds for consideration depth of k. For a customer with consideration depth k to purchase product $i\in S$, there are two possibilities. First, product i may have the largest utility among all utilities, in which case, the customer indeed purchases product i. Second, some product $j\in N\setminus S$ may have the largest utility among all utilities, in which case, product j occupies the top spot, leaving $k-1$ spots for the remaining products N_{-j} and the no-purchase option. By the second property of Gumbel random variables, the conditional on the fact that product j occupies the top spot, the choice process of the customer among the remaining products N_{-j} and the no-purchase option is identical to the unconditional choice process of the customer, so the probability that the customer purchases product i is the identical to the probability that a customer with consideration depth $k-1$ chooses product i when the set of all products is N_{-j}. Using the induction assumption with the set of products N_{-j}, the latter probability is $\frac{v_i}{1+V(N_{-j})}\, B^{k-1}(S,N_{-j})$. Collecting the two possibilities together, we get

$$\begin{aligned}\pi_i^k(S) &= \frac{v_i}{1+V(N)} + \sum_{j\in N\setminus S} \frac{v_j}{1+V(N)}\,\frac{v_i}{1+V(N_{-j})}\, B^{k-1}(S,N_{-j}) \\ &= \frac{v_i}{1+V(N)}\left(1+\sum_{j\in N\setminus S}\frac{v_j}{1+V(N_{-j})}\, B^{k-1}(S,N_{-j})\right) = \frac{v_i}{1+V(N)}\, B^k(S,N),\end{aligned}$$

where the first equality holds since, by the first property of Gumbel random variables, the probability that product ℓ has the largest utility among the utilities of all products and the no-purchase option is given by $\frac{v_\ell}{1+V(N)}$, whereas the last equality holds by the definition of $B^k(S,N)$ in (1). Thus, the result holds for consideration depth k as well. ■

Given that we offer the assortment S, a customer with consideration depth one purchases product $i \in S$ if and only if product i has the largest utility among all products and the no-purchase option. Thus, using the first property of Gumbel random variables, a customer with consideration depth one purchases product i with probability $\mathbb{P}\{U_i \geq U_0 \vee \max_{j \in N_{-i}} U_j\} = \frac{v_i}{1+V(N)}$. By the theorem above, given that we offer the assortment S, a customer with consideration depth k purchases product $i \in S$ with probability $\pi_i^k(S) = \frac{v_i}{1+V(N)} B^k(S,N)$. Thus, we can view $B^k(S,N)$ as the increase in the purchase probability of product i due to the fact that the customer has consideration depth k rather than one. Under our choice model, this increase in the purchase probability turns out to be the same for all products. A customer has consideration depth k with probability λ^k. So, if we offer the assortment S, then a customer purchases product i with probability

$$\phi_i(S) = \sum_{k \in M} \lambda^k \pi_i^k(S) = \frac{v_i}{1+V(N)} \sum_{k \in M} \lambda^k B^k(S,N). \tag{2}$$

Note that the computational effort to obtain $B^k(S,N)$ through (1) increases exponentially with k, so the computational effort to obtain $\phi_i(S)$ increases exponentially with m.

3. Assortment Optimization Problem

We formulate our assortment optimization problem and characterize its complexity. Under the standard multinomial logit model, given that we offer the assortment S, the choice probability of a product depends on the preference weights of the other products only through $V(S)$. Under our choice model, by the discussion in the previous section, the choice probability of a product depends on the preference weight of the other products through $\sum_{k \in M} \lambda^k B^k(S,N)$, ultimately making our assortment optimization problem significantly more difficult. In our assortment optimization problem, the revenue of product i is $r_i \geq 0$. Noting (2), if we offer the assortment S, then a customer chooses product i with probability $\phi_i(S)$, in which case, the expected revenue from a customer is $\sum_{i \in S} r_i \phi_i(S)$. Our goal is to find an assortment that maximizes the expected revenue. Therefore, letting $W(S) = \sum_{i \in S} r_i v_i$, using (2), we want to solve the problem

$$\max_{S \subseteq N} \left\{ \sum_{i \in N} r_i \phi_i(S) \right\} = \frac{1}{1+V(N)} \max_{S \subseteq N} \left\{ W(S) \sum_{k \in M} \lambda^k B^k(S,N) \right\}. \tag{3}$$

The assortment optimization problem above turns out to be NP-hard. To show this result, we reduce any instance of the partition problem to an instance of the feasibility version of our assortment optimization problem. In particular, the feasibility version of our assortment optimization problem is defined as follows.

Assortment Feasibility Problem: Given an expected revenue threshold K, does there exist an assortment that provides an expected revenue of K or more?

The partition problem is a well-known NP-complete problem; see Garey and Johnson (1979). The partition problem is defined as follows.

Partition Problem: Given a set of items $N = \{1, \ldots, n\}$, each item i having weight w_i, does there exist a subset of items S such that $\sum_{i \in S} w_i = \frac{1}{2} \sum_{i \in N} w_i$?

Theorem 3.1 (Computational Complexity) *The feasibility version of the assortment optimization problem in* (3) *is NP-complete.*

Proof: We use a reduction from the partition problem. Consider an instance of the partition problem with the set of items $N = \{1, \ldots, n\}$, where the weight of item i is w_i. Scaling the weight of each item by the same amount does not change the answer to the partition problem, so we normalize the weight of each item by $\sum_{i \in N} w_i$ so that $\sum_{i \in N} w_i = 1$. Corresponding to the instance of the partition problem, we construct an instance of the assortment feasibility problem as follows. The set of products is $N \cup \{n+1\}$. Letting $\Theta = \sum_{i \in N} \frac{w_i}{1+w_i}$, for $i = 1, \ldots, n$, the preference weight of product i is $v_i = \frac{2}{1-\Theta} \frac{w_i}{1+w_i}$. We set $v_{n+1} = 1$. Since $\Theta = \sum_{i \in N} \frac{w_i}{1+w_i} < \sum_{i \in N} w_i = 1$, we have $v_i > 0$ for all $i \in N$. For $i = 1, \ldots, n$, the revenue of product i is $r_i = \frac{1-\Theta}{2}(1+w_i)$. We set $r_{n+1} = 1$. Since $(1-\Theta)(1+w_i) < (1 - \frac{w_i}{1+w_i})(1+w_i) = 1$, we have $r_i < \frac{1}{2}$ for all $i \in N$, so product $n+1$ has the largest revenue. All customers have consideration depth of 2, so $m = 2$ and $(\lambda^1, \lambda^2) = (0, 1)$. Lastly, the expected revenue threshold is $K = \frac{9(1-\Theta)}{8}$. Since product $n+1$ has the largest revenue, offering product $n+1$ improves the expected revenue of any assortment. Thus, the assortment feasibility problem asks which of the remaining products in N to offer to obtain an expected revenue of $\frac{9(1-\Theta)}{8}$ or more. By (3), if we offer the subset of products $S \subseteq N$, then the expected revenue is

$$\begin{aligned} &\frac{\lambda^2}{1 + V(N) + v_{n+1}} \times W(S \cup \{n+1\}) \times B^2(S \cup \{n+1\}, N \cup \{n+1\}) \\ &\qquad = \frac{1}{1 + V(N) + v_{n+1}} \left(\sum_{i \in S} r_i \, v_i + r_{n+1} \, v_{n+1} \right) \left(1 + \sum_{i \in N \setminus S} \frac{v_i}{1 + V(N_{-i}) + v_{n+1}} \right), \end{aligned}$$

where the equality uses the the definition of $B^k(S, N)$ in (1) along with the fact that product $n+1$ is offered, $B^1(\cdot, \cdot) = 1$ and $(N \cup \{n+1\}) \setminus (S \cup \{n+1\}) = N \setminus S$.

We show that there exists a subset of products $S \subseteq N$ with expected revenue $\frac{9(1-\Theta)}{8}$ if and only if there exists a subset $S \subseteq N$ with $\sum_{i \in S} w_i = \frac{1}{2} \sum_{i \in N} w_i = \frac{1}{2}$. Noting the preference weights and revenues of the products, we have $1+V(N)+v_{n+1} = 2+\frac{2}{1-\Theta}\sum_{i\in N}\frac{w_i}{1+w_i} = 2+\frac{2\,\Theta}{1-\Theta} = \frac{2}{1-\Theta}$ and $r_i\,v_i = w_i$. Furthermore, we have $1+V(N_{-i})+v_{n+1} = 1+V(N)+v_{n+1}-v_i = \frac{2}{1-\Theta}-\frac{2}{1-\Theta}\frac{w_i}{1+w_i} = \frac{2}{1-\Theta}\frac{1}{1+w_i}$, so we get $\frac{v_i}{1+V(N_{-i})+v_{n+1}} = w_i$. Using the last three identities in the equality above, since $\sum_{i\in N} w_i = 1$, the expected revenue from the subset of products $S\subseteq N$ is

$$\frac{1}{1+V(N)+v_{n+1}}\left(\sum_{i\in S} r_i\,v_i + r_{n+1}\,v_{n+1}\right)\left(1+\sum_{i\in N\setminus S}\frac{v_i}{1+V(N_{-i})+v_{n+1}}\right)$$
$$= \frac{1-\Theta}{2}\left(1+\sum_{i\in S} w_i\right)\left(1+\sum_{i\in N\setminus S} w_i\right) = \frac{1-\Theta}{2}\left(1+\sum_{i\in S} w_i\right)\left(2-\sum_{i\in S} w_i\right).$$

Thus, there exists a subset $S\subseteq N$ with expected revenue $\frac{9(1-\Theta)}{8}$ if and only if there exists a subset $S\subseteq N$ such that $(1+\sum_{i\in S} w_i)\,(2-\sum_{i\in S} w_i)\geq \frac{9}{4}$.

The maximum of $g(x)=(1+x)(2-x)$ is at $x=\frac{1}{2}$ with $g(\frac{1}{2})=\frac{9}{4}$. So, there exists a subset $S\subseteq N$ with expected revenue $\frac{9(1-\Theta)}{8}$ if and only if there exists a subset $S\subseteq N$ with $\sum_{i\in S} w_i=\frac{1}{2}$. ■

To put Theorem 3.1 into perspective, consider two edge cases mentioned in the introduction. First, if the consideration depth of all customers is n, then a customer simply chooses the available alternative with the largest utility, so our choice model reduces to the standard multinomial logit model. Under the standard multinomial logit model, there exists a revenue-ordered optimal solution to problem (3), including a certain number of products with the largest revenues. Thus, we can obtain the optimal assortment by checking the expected revenue from each revenue-ordered assortment. Second, if the consideration depth of all customers is one, then a customer focuses only on the top one alternative with the largest utility and chooses this alternative if it is available. If all customers have a consideration depth of one, then the objective function of problem (3) reduces to $\frac{1}{1+V(N)}\,W(S)$, in which case, the optimal assortment is obtained by setting $S=N$. Therefore, the assortment optimization problem admits a polynomial-time solution in the two edge cases. By Theorem 3.1, if customers have consideration depths between the two edge cases, then problem (3) is NP-hard. To demonstrate how the revenues and preference weights of the products interact in problem (3), consider a problem instance with three products. The revenues and preference weights are given by $(r_1,r_2,r_3)=(100,12,9)$ and $(v_1,v_2,v_3)=(3,90,20)$. All customers have consideration depth of 2, so $m=2$ and $(\lambda^1,\lambda^2)=(0,1)$. In Table 1, we give the choice probability of each product under each assortment, along with the corresponding expected revenue from the assortment.

The optimal assortment is $\{1,3\}$ with an expected revenue of 20. This assortment is not revenue-ordered, as it includes the products with the largest and smallest revenue but skips the product with the second largest revenue. If we offer the assortment $\{1\}$, then the choice probability of product 1 is 0.131, but there are no other products to generate revenue, so the probability of

Assrt.	Exp. Rev.	Choice Prob. 1	2	3
$\varnothing$	0.0	0.0	0.0	0.0
$\{1\}$	13.060	0.131	0.0	0.0
$\{2\}$	11.745	0.0	0.979	0.0
$\{3\}$	7.543	0.0	0.0	0.838

Assrt.	Exp. Rev.	Choice Prob. 1	2	3
$\{1,2\}$	14.681	0.032	0.957	0.0
$\{1,3\}$	20.0	0.125	0.0	0.833
$\{2,3\}$	11.351	0.0	0.811	0.180
$\{1,2,3\}$	13.684	0.026	0.789	0.175

Table 1 Expected revenues and choice probabilities for different assortments.

no-purchase is 0.869. If we add product 3 to the assortment $\{1\}$, so the offered assortment is $\{1,3\}$, then the choice probabilities of products 1 and 3 are 0.125 and 0.833. Adding product 3, which has a moderate preference weight, does not reduce the demand for product 1 significantly, but it reduces the probability of no-purchase drastically. If we add product 2 to the assortment $\{1\}$, so the offered assortment is $\{1,2\}$, then the choice probabilities of products 1 and 2 are 0.032 and 0.957. Adding product 2, which has a large preference weight, substantially reduces the demand for product 1. Although not reported in the table, if all customers have consideration depth of 3, then the optimal assortment is $\{1\}$. Under consideration depth of 3, if we offer the assortment $\{1\}$, then the purchase probability of product 1 is 0.75. As the consideration depth increases from 2 to 3, intuitively speaking, the customers become more willing to substitute, so the purchase probability of product 1 within the assortment $\{1\}$ drastically increases from 0.131 to 0.75.

4. Impact of Consideration Sets

As discussed earlier, if all customers have consideration depth n, then our choice model becomes equivalent to the standard multinomial logit model. In our choice model, however, not all customers have consideration depth n. Intuitively speaking, smaller consideration depths translate into choosier customers. In that sense, the customers choosing according to our choice model are choosier than those picking according to the standard multinomial logit model. In this section, we show that the optimal solution to our assortment optimization problem includes all products in an optimal solution to the assortment optimization problem under the standard multinomial logit model. Therefore, to maximize the expected revenue from choosier customers, the optimal assortment to offer becomes larger. Under the standard multinomial logit model, if we offer the assortment S, then a customer chooses product $i \in S$ with probability $\frac{v_i}{1+V(S)}$, so we can find the optimal assortment by solving the problem

$$\max_{S \subseteq N} \left\{ \frac{\sum_{i \in S} r_i \, v_i}{1 + V(S)} \right\}. \tag{4}$$

Throughout this section, if there are multiple revenue-maximizing assortments, then we choose the one with the largest cardinality. Letting S^* be an optimal solution to our assortment optimization problem in (3) and $\widetilde{S}$ be an optimal solution to the assortment optimization problem under the standard multinomial logit model in (4), our goal is to show that $S^* \supseteq \widetilde{S}$.

To show this result, we will use two preliminary lemmas. In the next lemma, we give an upper bound on $B^k(S,N)$, where $B^k(S,N)$ is computed through the recursion in (1).

Lemma 4.1 (Upper Bound on Choice Bump) *For $k \in M$, $S \subseteq N$, letting $B^k(S,N)$ be computed through the recursion in* (1)*, we have*

$$\frac{1}{1+V(N)} B^k(S,N) \;\le\; \frac{1}{1+V(S)}. \tag{5}$$

Proof: We use induction over the consideration depth to show that (5) holds for any set of products N, $S \subseteq N$ and $k \in M$. For $k=1$, we have $B^1(S,N)=1$ by the boundary condition in (1), in which case, noting that $S \subseteq N$, we obtain $\frac{1}{1+V(N)} B^1(S,N) = \frac{1}{1+V(N)} \le \frac{1}{1+V(S)}$. Therefore, the result holds for $k=1$. Assuming that the result holds for consideration depth of $k-1$, we show that the result holds for consideration depth of k. For $S \subseteq N$, for any $i \in N \setminus S$, we have $S \subseteq N_{-i}$. In this case, using the induction assumption with the set of products N_{-i}, we have $\frac{1}{1+V(N_{-i})} B^{k-1}(S,N_{-i}) \le \frac{1}{1+V(S)}$ for each $i \in N \setminus S$. Thus, by the definition of $B^k(S,N)$ in (1), we get

$$B^k(S,N) = 1 + \sum_{i \in N \setminus S} \frac{v_i}{1+V(N_{-i})} B^{k-1}(S,N_{-i}) \le 1 + \sum_{i \in N \setminus S} \frac{v_i}{1+V(S)} = 1 + \frac{V(N)-V(S)}{1+V(S)} = \frac{1+V(N)}{1+V(S)}.$$

Arranging the terms in the chain of inequalities above yields $\frac{1}{1+V(N)} B^k(S,N) \le \frac{1}{1+V(S)}$, so the result holds for consideration depth of k as well. ∎

Given that the set of products is N and we offer the assortment S, let $R^k(S,N)$ be the revenue from a customer with consideration depth k. Note that $R^k(S,N)$ is a random variable.

In the next lemma, we focus on computing the expected revenue from a customer with consideration depth k conditional on the top product with the largest utility.

Lemma 4.2 (Expected Revenue Conditional on Top Product) *For any set of products N, $S \subseteq N$ and $k \in M$, the revenue from a customer with consideration depth k satisfies*

$$\mathbb{E}\Big\{ R^k(S,N) \,\Big|\, U_i \ge U_0 \vee \max_{j \in N_{-i}} U_j \Big\} \;=\; \begin{cases} r_i & \text{if } i \in S \\ \mathbb{E}\{R^{k-1}(S,N_{-i})\} & \text{if } i \notin S. \end{cases} \tag{6}$$

Proof: Given $U_i \ge U_0 \vee \max_{j \in N_{-i}} U_j$, product i occupies the top spot. So, if $i \in S$, then a customer with consideration depth k purchases product i, yielding $\mathbb{E}\{R^k(S,N) \,|\, U_i \ge U_0 \vee \max_{j \in N_{-i}} U_j\} = r_i$.

If $i \notin S$, then product i is not available, but given $U_i \ge U_0 \vee \max_{j \in N_{-i}} U_j$, it occupies the top spot. Thus, if we offer the assortment S, then a customer with consideration depth k does not purchase the product in the top spot, so we are left with $k-1$ spots for the remaining products N_{-i} and the no-purchase option. By the second property of the Gumbel random variables, conditional on the fact that product i occupies the top spot, the choice process of the customer among the remaining products N_{-i} and no-purchase option is identical to the unconditional choice process, so the probability that the customer chooses some product $j \in S$ is the identical to the probability that a customer with consideration depth $k-1$ chooses product j when the set of all products is N_{-i}, in which case, we get $\mathbb{E}\{R^k(S,N) \,|\, U_i \geq U_0 \vee \max_{j \in N_{-i}} U_j\} = \mathbb{E}\{R^{k-1}(S,N_{-i})\}$. ■

Noting the definition of $R^k(S,N)$, if we offer the assortment S, then the expected revenue from a customer is $\sum_{k \in M} \lambda^k \, \mathbb{E}\{R^k(S,N)\}$. By Theorem 2.1, given that the set of products is N and we offer the assortment S, a customer with consideration depth k purchases product i with probability $\frac{v_i}{1+V(N)} B^k(S,N)$, in which case, the expected revenue from a customer with consideration depth k is $\mathbb{E}\{R^k(S,N)\} = \frac{\sum_{i \in S} r_i v_i}{1+V(N)} B^k(S,N)$. In the next theorem, we use this observation and the previous two lemmas to show that an optimal solution to our assortment optimization problem includes all products in the one under the standard multinomial logit model.

Theorem 4.3 (Monotonicity of Optimal Assortments) *Letting S^* be an optimal solution to problem* (3) *and $\widetilde{S}$ be an optimal solution to problem* (4)*, we have $S^* \supseteq \widetilde{S}$.*

Proof: Letting $S_{+i} = S \cup \{i\}$, we use induction over the consideration depth to show that $\mathbb{E}\{R^k(S_{+\ell},N)\} \geq \mathbb{E}\{R^k(S,N)\}$ for any set of products N, $S \subseteq N$, $\ell \in \widetilde{S} \setminus S$ and $k \in M$. In this case, if there exists some $\ell \in \widetilde{S}$ with $\ell \notin S^*$, then $\sum_{k \in M} \lambda^k \, \mathbb{E}\{R^k(S^*_{+\ell},N)\} \geq \sum_{k \in M} \lambda^k \, \mathbb{E}\{R^k(S^*,N)\}$, contradicting the fact that S^* is an optimal solution to problem (3) with the largest cardinality. For $k=1$, since $B^1(\cdot,\cdot)=1$, the discussion right before the theorem yields $\mathbb{E}\{R^1(S,N)\} = \frac{\sum_{q \in S} r_q v_q}{1+V(N)} \leq \frac{r_\ell v_\ell + \sum_{q \in S} r_q v_q}{1+V(N)} = \mathbb{E}\{R^1(S_{+\ell},N)\}$, so the result holds for $k=1$. Assuming that the result holds for consideration depth of $k-1$, we show that the result holds for consideration depth of k.

Let product i be the one that provides the largest utility among all utilities. That is, we have $U_i \geq U_0 \vee \max_{j \in N_{-i}} U_j$. For notational brevity, given that the set of products is N and we offer the assortment S, let $\overline{R}_i^k(S,N)$ be the expected revenue from a customer with consideration depth k, conditional on the fact that product i provides the largest utility. In other words, we have $\overline{R}_i^k(S,N) = E\{R^k(S,N) \,\big|\, U_i \geq U_0 \vee \max_{j \in N_{-i}} U_j\}$, which corresponds to the left side of (6). By the tower property of conditional expectations, we have $\mathbb{E}\{\overline{R}_i^k(S,N)\} = \mathbb{E}\{R^k(S,N)\}$. We consider three possibilities for product i to show that $\overline{R}_i^k(S_{+\ell},N) \geq \overline{R}_i^k(S,N)$.

First, consider the case $i \in S$. Thus, we have $i \in S$ and $i \in S_{+\ell}$, in which case, by (6), $\overline{R}_i^k(S,N) = r_i = \overline{R}_i^k(S_{+\ell},N)$. Second, consider the case $i \notin S$ and $i \neq \ell$. Therefore, we have $i \notin S$ and $i \notin S_{+\ell}$, so by (6), we get $\overline{R}_i^k(S,N) = \mathbb{E}\{R^{k-1}(S,N_{-i})\} \leq \mathbb{E}\{R^{k-1}(S_{+\ell},N_{-i})\} = \overline{R}_i^k(S_{+\ell},N)$, where the inequality uses the induction assumption with the set of products N_{-i}. Third, consider the case $i \notin S$ and $i = \ell$. Thus, we have $i \notin S$ but $i \in S_{+\ell}$. Since $i \in S_{+\ell}$, by (6), we get $\overline{R}_i^k(S_{+\ell},N) = r_i = r_\ell$. Recalling that $\widetilde{S}$ is the optimal solution to problem (4), a well-known result for assortment optimization under the standard multinomial logit model, given as Lemma B.1 in Appendix B, shows that $r_\ell \geq \frac{\sum_{q\in\widetilde{S}} r_q\, v_q}{1+V(\widetilde{S})}$ for all $\ell \in \widetilde{S}$. Thus, since $\ell \in \widetilde{S}\setminus S$, we get $\overline{R}_i^k(S_{+\ell},N) = r_\ell \geq \frac{\sum_{q\in\widetilde{S}} r_q\, v_q}{1+V(\widetilde{S})} \geq \frac{\sum_{q\in S} r_q\, v_q}{1+V(S)}$, where the last inequality is by the fact that $\widetilde{S}$ is an optimal solution to problem (4). Also, since $i \notin S$, by (6), we have $\overline{R}_i^k(S,N) = \mathbb{E}\{R^{k-1}(S,N_{-i})\}$, but by the discussion right before the theorem, the last expectation is $\mathbb{E}\{R^{k-1}(S,N_{-i})\} = \frac{\sum_{q\in S} r_q\, v_q}{1+V(N_{-i})}\, B^{k-1}(S,N_{-i})$. So, we get

$$\overline{R}_i^k(S,N) \;=\; \mathbb{E}\{R^{k-1}(S,N_{-i})\} \;=\; \frac{\sum_{q\in S} r_q\, v_q}{1+V(N_{-i})}\, B^{k-1}(S,N_{-i}) \;\leq\; \frac{\sum_{q\in S} r_q\, v_q}{1+V(S)},$$

where the last inequality follows by Lemma 4.1. Noting that $\overline{R}_i^k(S_{+\ell},N) \geq \frac{\sum_{q\in S} r_q\, v_q}{1+V(S)}$, we get $\overline{R}_i^k(S,N) \leq \overline{R}_i^k(S_{+\ell},N)$. Collecting all three cases yields $\overline{R}_i^k(S,N) \leq \overline{R}_i^k(S_{+\ell},N)$.

In this case, we get $\mathbb{E}\{R^k(S,N)\} = \mathbb{E}\{\overline{R}_i^k(S,N)\} \leq \mathbb{E}\{\overline{R}_i^k(S_{+\ell},N)\} = \mathbb{E}\{R^k(S_{+\ell},N)\}$, where the two equalities use the tower property, so the result holds for consideration depth of k. ■

Customers choosing under our choice model are, intuitively speaking, choosier than those picking under the standard multinomial logit model. By Theorem 4.3, since customers under our choice model are choosier, hence harder to satisfy, the revenue-maximizing assortment becomes larger. Consider a problem instance with five products, where the revenues and preference weights are $(r_1,r_2,r_3,r_4,r_5) = (96,69,33,30,10)$ and $(v_1,v_2,v_3,v_4,v_5) = (7,16,1,0.2,12)$. Let S_κ^* be the optimal assortment when all customers have consideration depth of κ. That is, S_κ^* is an optimal solution to problem (3), when $\lambda^\kappa = 1$ and $\lambda^k = 0$ for all $k \in M_{-\kappa}$. If all customers have consideration depth of 5, so that they choose according to the standard multinomial logit model, then the optimal assortment is $S_5^* = \{1\}$. If all customers have consideration depth of 3, then the optimal assortment is $S_3^* = \{1,3,4\}$. As predicted by Theorem 4.3, we have $\{1,3,4\} = S_3^* \supseteq S_5^* = \{1\}$. A tempting conjecture is that as the consideration depth of the customers decreases, the optimal assortments are nested; that is, $S_1^* \supseteq S_2^* \supseteq S_3^* \supseteq S_4^* \supseteq S_5^*$. This conjecture is incorrect. For the problem instance in this paragraph, if all customers have consideration depth of 2, then the optimal assortment is $S_2^* = \{1,2\}$, so we do not have $S_2^* \supseteq S_3^*$. In fact, this problem instance shows that we may not even have $|S_1^*| \geq |S_2^*| \geq |S_3^*| \geq |S_4^*| \geq |S_5^*|$, as $|S_2^*| = 2$ and $|S_3^*| = 3$. Theorem 4.3 allows us to compare the optimal assortment under endogenous consideration sets only with S_n^*, where the assortment S_n^* is the optimal solution to the assortment optimization problem when all customers have consideration depth n, corresponding to the case where the customers choose under the standard multinomial logit model. However, we do not necessarily have $S_k^* \supseteq S_{k+1}^*$, when $k < n-1$.

5. Polynomial-Time Approximation Scheme

To give a PTAS for problem (3), for some accuracy parameter $\epsilon \in (0,1)$, we partition the products into product classes such that the product of the preference weight and revenue for the products in each class is within a factor of $1+\epsilon$ of each other. Once we construct the product classes, we proceed in two parts. First, we argue that there exists a $(1-\epsilon)^2$-approximate solution that offers products only in $O(\frac{\log(n/\epsilon)}{\epsilon})$ product classes. Furthermore, this assortment is light in the sense that the products that the assortment offers in each product class correspond to those with the smallest preference weights. Second, we construct a collection of $O((n/\epsilon)^{O(1/\epsilon^2)})$ light assortments such that this collection includes an assortment whose expected revenue deviates from that of the best light assortment by a factor of at most $1-\epsilon$. In this section, we explain what we precisely mean in the two parts and put the two parts together to obtain our PTAS.

Letting $w_i = r_i\, v_i$, we refer to w_i as the weight of product i. Fixing some $\epsilon \in (0,1)$, for each $g \in \mathbb{Z}$, we construct product class g as $N_g = \{i \in N : (1+\epsilon)^g \le w_i < (1+\epsilon)^{g+1}\}$. Therefore, if we round the weights of the products in product class g down to the nearest integer power of $1+\epsilon$, then the rounded weight is $(1+\epsilon)^g$. For all $i \in N_g$, let $\overline{w}_i = (1+\epsilon)^g$, which is the rounded weight of the products in product class g. Note that $\overline{w}_i \le w_i \le (1+\epsilon)\,\overline{w}_i$ by our construction, so letting $\overline{W}(S) = \sum_{i\in S} \overline{w}_i$, we have $\overline{W}(S) \le W(S) \le (1+\epsilon)\,\overline{W}(S)$. We use $S_g^{\mathsf{light}}(k) \subseteq N^g$ to denote the assortment that includes the k products with the smallest preference weights in product class g. We say that the assortment $S \subseteq N$ is light if we have $S \cap N_g \in \{S_g^{\mathsf{light}}(k) : k = 0,\ldots,|N_g|\}$ for all $g \in \mathbb{Z}$. That is, an assortment is light if the products that it offers in each product class correspond to those with the smallest preference weights in the class. We will proceed in two steps.

Part 1. Existence of an Ideal Light Assortment: We will show that there exists a light assortment $\widehat{S}$ that satisfies the following two properties.

- **(Bounded Support)** Letting $\widehat{G} = \{g \in \mathbb{Z} : \widehat{S} \cap N_g \neq \varnothing\}$, we have $|\widehat{G}| = O(\frac{\log(n/\epsilon)}{\epsilon})$, so the assortment $\widehat{S}$ offers products in $O(\frac{\log(n/\epsilon)}{\epsilon})$ product classes.
- **(Limited Degradation)** Letting S^* be an optimal solution to problem (3), we have $\overline{W}(\widehat{S}) \ge (1-\epsilon)\,\overline{W}(S^*)$ and $B^k(\widehat{S},N) \ge B^k(S^*,N)$ for all $k \in M$.

Since $\overline{W}(S) \le W(S) \le (1+\epsilon)\,\overline{W}(S)$, using the limited degradation property, $W(\widehat{S}) \ge \overline{W}(\widehat{S}) \ge (1-\epsilon)\,\overline{W}(S^*) \ge \frac{1-\epsilon}{1+\epsilon} W(S^*)$. Also noting that $B^k(\widehat{S},N) \ge B^k(S^*,N)$ for all $k \in M$, we get

$$\frac{1}{1+V(N)}\left\{W(\widehat{S})\sum_{k\in M}\lambda^k\, B^k(\widehat{S},N)\right\} \;\ge\; \frac{\frac{1-\epsilon}{1+\epsilon}}{1+V(N)}\left\{W(S^*)\sum_{k\in M}\lambda^k\, B^k(S^*,N)\right\}. \tag{7}$$

Using the fact that $\frac{1-\epsilon}{1+\epsilon} \ge (1-\epsilon)^2$, the inequality above implies that the assortment $\widehat{S}$ is a $(1-\epsilon)^2$-approximate solution to problem (3). In other words, there exists a light assortment $\widehat{S}$

that offers products only in $O(\frac{\log(n/\epsilon)}{\epsilon})$ product classes and corresponds to a $(1-\epsilon)^2$-approximate solution to problem (3). We refer to this assortment as the ideal light assortment.

Part 2. Candidate Assortments: Letting $\widehat{S}$ be the ideal light assortment in Part 1, we will construct a collection of assortments $\{A_t : t \in \mathcal{A}\}$ that satisfies the following two properties.

- **(Small and Light Collection)** There are $O((n/\epsilon)^{O(1/\epsilon^2)})$ assortments in the collection and each one of the assortments is light.
- **(Approximation to Ideal)** Letting $\widehat{G} = \{g \in \mathbb{Z} : \widehat{S} \cap N_g \neq \varnothing\}$, there exists an assortment $\widetilde{S} \in \{A_t : t \in \mathcal{A}\}$ such that

$$\overline{W}(\widetilde{S} \cap N_g) \ \leq \ \overline{W}(\widehat{S} \cap N_g) \ \leq \ \overline{W}(\widetilde{S} \cap N_g) + \frac{\epsilon}{|\widehat{G}|} \overline{W}(\widehat{S}) \quad \forall g \in \mathbb{Z}. \tag{8}$$

The number of operations to construct the collection of assortments $\{A_t : t \in \mathcal{A}\}$ will be $O((n/\epsilon)^{O(1/\epsilon^2)} \log n)$. We refer to $\{A_t : t \in \mathcal{A}\}$ as the collection of candidate assortments.

By the small and light collection property, the number of candidate assortments increases polynomially with the number of products n and it is independent of the maximum consideration depth m. By (8) in the approximation to ideal property, if $\widehat{S} \cap N_g = \varnothing$, then $\widetilde{S} \cap N_g = \varnothing$, so there exists a candidate assortment that uses no more product classes than the ideal light assortment. Also, using (8), we will be able to show that the expected revenue from this candidate assortment deviates from the expected revenue of the ideal light assortment by at most a factor of $1-\epsilon$. In Section 6.1, we show that Part 1 holds, whereas in Section 6.2, we show that Part 2 holds. In the remainder of this section, we put the two parts together to give our PTAS. In particular, using the two parts above, we get the next theorem. In this theorem, we use RevOps to denote the number of operations to compute the expected revenue from an assortment.

Theorem 5.1 (PTAS) *For any $\epsilon \in (0,1)$, we can find a $(1-\epsilon)$-approximate solution to problem (3) in $O(\mathsf{RevOps}\ (n/\epsilon)^{O(1/\epsilon^2)})$ operations.*

Proof: Let $\widehat{S}$ be the ideal light assortment in Part 1 and $\widetilde{S}$ be the assortment that satisfies the approximation to ideal property in Part 2. Note that both $\widehat{S}$ and $\widetilde{S}$ are light. We have

$$(1-\epsilon)\,\overline{W}(\widehat{S}) \overset{(a)}{=} \sum_{g \in \widehat{G}} \left\{ \overline{W}(\widehat{S} \cap N_g) - \frac{\epsilon}{|\widehat{G}|} \overline{W}(\widehat{S}) \right\} \overset{(b)}{\leq} \sum_{g \in \widehat{G}} \overline{W}(\widetilde{S} \cap N_g) \overset{(c)}{=} \sum_{g \in G} \overline{W}(\widetilde{S} \cap N_g) = \overline{W}(\widetilde{S}), \tag{9}$$

where (a) holds since $\widehat{S} \cap N_g = \varnothing$ for all $g \notin \widehat{G}$, (b) follows from (8) and (c) uses the fact that if $g \notin \widehat{G}$, then $\widehat{S} \cap N_g = \varnothing$, in which case $\widetilde{S} \cap N_g = \varnothing$ by (8).

Since $\widehat{S}$ and $\widetilde{S}$ are light, for each $g \in \mathbb{Z}$, both $\widehat{S} \cap N_g$ and $\widetilde{S} \cap N_g$ include a certain number of products with the smallest preference weights in product class g. Noting that $\widehat{S}$ and $\widetilde{S}$ satisfy

(8), we have $\overline{W}(\widetilde{S}\cap N_g) \le \overline{W}(\widehat{S}\cap N_g)$, so $\widetilde{S}\cap N_g \subseteq \widehat{S}\cap N_g$ for all $g\in\mathbb{Z}$. In this case, we obtain $\widetilde{S} = \cup_{g\in\mathbb{Z}}(\widetilde{S}\cap N_g) \subseteq \cup_{g\in\mathbb{Z}}(\widehat{S}\cap N_g) = \widehat{S}$. We use induction over the consideration depth to show that $B^k(S,N) \ge B^k(Q,N)$ for any set of products N, $S\subseteq Q\subseteq N$ and $k\in M$. Since $B^1(\cdot,\cdot)=1$, the result holds for $k=1$. Assuming that the results holds for consideration depth of $k-1$, we show that the result holds for consideration depth of k. In particular, by (1), we have

$$B^k(Q,N) = 1 + \sum_{i\in N\setminus Q} \frac{v_i}{1+V(N_{-i})}\, B^{k-1}(Q,N_{-i}) \overset{(d)}{\le} 1 + \sum_{i\in N\setminus S} \frac{v_i}{1+V(N_{-i})}\, B^{k-1}(S,N_{-i}) = B^k(S,N),$$

where (d) holds since we have $S\subseteq Q\subseteq N_{-i}$ for any $i\in N\setminus Q$, so by the induction assumption with the set of products N_{-i}, we have $B^{k-1}(Q,N_{-i}) \le B^{k-1}(S,N_{-i})$, along with the fact that $N\setminus Q\subseteq N\setminus S$, so the sum on the left side of (d) includes fewer terms than the sum on the right side. By the chain of inequalities above, the result holds for consideration depth of k. Since $\widetilde{S}\subseteq\widehat{S}$ as discussed at the beginning of this paragraph, we get $B^k(\widetilde{S},N) \ge B^k(\widehat{S},N)$ for all $k\in M$.

Thus, letting S^* be an optimal solution to problem (3), noting that $\overline{W}(S) \le W(S) \le (1+\epsilon)\,\overline{W}(S)$ for any $S\subseteq N$, by (7) and (9), the expected revenue from the assortment $\widetilde{S}$ satisfies

$$\begin{aligned}
\frac{1}{1+V(N)}\left\{W(\widetilde{S})\sum_{k\in M}\lambda^k\, B^k(\widetilde{S},N)\right\} &\ge \frac{1}{1+V(N)}\left\{\overline{W}(\widetilde{S})\sum_{k\in M}\lambda^k\, B^k(\widetilde{S},N)\right\}\\
&\overset{(e)}{\ge} \frac{1-\epsilon}{1+V(N)}\left\{\overline{W}(\widehat{S})\sum_{k\in M}\lambda^k\, B^k(\widetilde{S},N)\right\} \overset{(f)}{\ge} \frac{1-\epsilon}{1+V(N)}\left\{\overline{W}(\widehat{S})\sum_{k\in M}\lambda^k\, B^k(\widehat{S},N)\right\}\\
&\ge \frac{\frac{1-\epsilon}{1+\epsilon}}{1+V(N)}\left\{W(\widehat{S})\sum_{k\in M}\lambda^k\, B^k(\widehat{S},N)\right\} \ge \frac{\frac{(1-\epsilon)^2}{(1+\epsilon)^2}}{1+V(N)}\left\{W(S^*)\sum_{k\in M}\lambda^k\, B^k(S^*,N)\right\},
\end{aligned}$$

where (e) is by (9) and (f) uses the fact that $B^k(\widetilde{S},N)\ge B^k(\widehat{S},N)$ for all $k\in M$. Noting that $\frac{(1-\epsilon)^2}{(1+\epsilon)^2} \ge (1-\epsilon)^4 \ge 1-4\,\epsilon$, the assortment $\widetilde{S}$ is a $(1-4\,\epsilon)$-approximate solution to problem (3).

Since $\widetilde{S}\in\{A_t : t\in\mathcal{A}\}$, there exists an assortment in the collection of candidate assortments $\{A_t : t\in\mathcal{A}\}$ that is a $(1-4\,\epsilon)$-approximate solution to problem (3). By the small and light collection property in Part 2, there are $O((n/\epsilon)^{O(1/\epsilon^2)})$ assortments in the collection. Also, we can construct the collection in $O((n/\epsilon)^{O(1/\epsilon^2)})$ operations. Thus, by constructing the collection of candidate assortments and checking the expected revenue from each assortment, which takes $O(\mathsf{RevOps}\;(n/\epsilon)^{O(1/\epsilon^2)})$ operations, we get a $(1-4\,\epsilon)$-approximate solution to problem (3). Given any $\delta\in(0,1)$, by repeating the discussion in the proof with $\epsilon=\delta/4$, we get a $(1-\delta)$-approximate solution in $O(\mathsf{RevOps}\;(n/\delta)^{O(1/\delta^2)})$ operations. ■

By the discussion right after (2), the number of operations to compute the expected revenue from an assortment, which we denote by RevOps, increases exponentially with the maximum consideration depth m. Thus, the running time of our PTAS increases polynomially in the number of products n but exponentially in m. However, the running time of our PTAS depends on m only through RevOps. In particular, the number of operations to generate the collection of candidate assortments is independent of m. The running time of our PTAS could be independent of m if one is content with approximating the expected revenue from an assortment by using simulation, but naturally, using simulation would introduce an additional source of error.

6. Ideal Light Assortment and Candidate Collection

Considering Parts 1 and 2 in the previous section, we show that there exists a light assortment that satisfies Part 1 and construct the collection of candidate assortments that satisfies Part 2.

6.1 Existence of an Ideal Light Assortment

We show that there exists an ideal light assortment $\widehat{S}$ that satisfies the bounded support and limited degradation properties in Part 1. Letting S^* be an optimal solution to problem (3), we set $k^*_g = |S^* \cap N_g|$, which is the number of products offered by the optimal assortment S^* in product class g. Recalling that $S_g^{\mathsf{light}}(k)$ is the assortment that includes the k products with the smallest preference weights in product class g, we set $\widehat{S}_g = S_g^{\mathsf{light}}(k^*_g)$. Thus, the assortment $\widehat{S}_g$ offers the same number of products in product class g as the optimal assortment S^*, but it offers the products with the smallest preference weights. Lastly, we set $g^*_{\max} = \max\{g \in \mathbb{Z} : S^* \cap N_g \neq \varnothing\}$, which is the largest product class in which the optimal assortment S^* offers a product. Using $\lceil \cdot \rceil$ to denote the round up function, letting $L = \lceil \frac{\log(n/\epsilon)}{\log(1+\epsilon)} \rceil$ and $g^*_{\min} = g^*_{\max} - L + 1$, we set the ideal light assortment $\widehat{S}$ as

$$\widehat{S} = \bigcup_{g=g^*_{\min}}^{g^*_{\max}} \widehat{S}_g. \tag{10}$$

By construction, the assortment $\widehat{S}$ is light. Also, it offers products in $g^*_{\max} - g^*_{\min} + 1 = \lceil \frac{\log(n/\epsilon)}{\log(1+\epsilon)} \rceil = O(\frac{\log(n/\epsilon)}{\epsilon})$ product classes. Thus, the assortment $\widehat{S}$ satisfies the bounded support property.

In the next proposition, we show that the assortment $\widehat{S}$ that we construct as in the previous paragraph satisfies the limited degradation property.

Proposition 6.1 (Limited Degradation) *Letting S^* be an optimal solution to problem* (3) *and $\widehat{S}$ be as in* (10), *we have $\overline{W}(\widehat{S}) \geq (1-\epsilon)\,\overline{W}(S^*)$ and $B^k(\widehat{S}, N) \geq B^k(S^*, N)$ for all $k \in M$.*

Proof: Note that the assortment $\widehat{S}$ offers product only in product classes $g^*_{\min}, \ldots, g^*_{\max}$. Furthermore, the set of products that the assortment $\widehat{S}$ offers in product class g is $\widehat{S}_g$, and $\widehat{S}_g$ satisfies

$|\widehat{S}_g| = k_g^*$. Lastly, by the discussion at the beginning of Section 5, we have $\overline{w}_i = (1+\epsilon)^g$ for all $i \in N_g$. Therefore, we have $\overline{W}(\widehat{S}) = \sum_{g=g^*_{\min}}^{g^*_{\max}} \sum_{i \in \widehat{S}_g} \overline{w}_i = \sum_{g=g^*_{\min}}^{g^*_{\max}} (1+\epsilon)^g\, k_g^*$. On the other hand, the largest product class in which the optimal assortment S^* offers a product is $g^*_{\max}$, which implies that $\overline{W}(S^*) = \sum_{g=-\infty}^{g^*_{\max}} \sum_{i \in S^* \cap N_g} \overline{w}_i = \sum_{g=-\infty}^{g^*_{\max}} \sum_{i \in S^* \cap N_g} (1+\epsilon)^g = \sum_{g=-\infty}^{g^*_{\max}} (1+\epsilon)^g\, k_g^*$, where the last equality uses the fact that $k_g^* = |S^* \cap N_g|$. Thus, we get

$$\begin{aligned}
\overline{W}(\widehat{S}) \;&=\; \sum_{g=g^*_{\min}}^{g^*_{\max}} (1+\epsilon)^g\, k_g^* \;=\; \sum_{g=-\infty}^{g^*_{\max}} (1+\epsilon)^g\, k_g^* - \sum_{g=-\infty}^{g^*_{\min}-1} (1+\epsilon)^g\, k_g^* \;=\; \overline{W}(S^*) - \sum_{g=-\infty}^{g^*_{\min}-1} (1+\epsilon)^g\, k_g^* \\
&\geq\; \overline{W}(S^*) - (1+\epsilon)^{g^*_{\min}-1} \sum_{g=-\infty}^{g^*_{\min}-1} k_g^* \;\overset{(a)}{\geq}\; \overline{W}(S^*) - n\,(1+\epsilon)^{g^*_{\max}-L} \\
&\overset{(b)}{\geq}\; \overline{W}(S^*) - n\,(1+\epsilon)^{g^*_{\max}}\,(1+\epsilon)^{-\log_{1+\epsilon}(n/\epsilon)} \;=\; \overline{W}(S^*) - \epsilon\,(1+\epsilon)^{g^*_{\max}} \;\overset{(c)}{\geq}\; (1-\epsilon)\,\overline{W}(S^*),
\end{aligned}$$

where (a) holds since $g^*_{\min} = g^*_{\max} - L + 1$ and $\sum_{g=-\infty}^{g^*_{\min}-1} k_g^* \leq \sum_{g=-\infty}^{\infty} k_g^* \leq n$, (b) is by noting that $L \geq \frac{\log(n/\epsilon)}{\log(1+\epsilon)} = \log_{1+\epsilon}(n/\epsilon)$ and (c) follows as $S^* \cap N_{g^*_{\max}} \neq \varnothing$, so $\overline{W}(S^*) \geq (1+\epsilon)^{g^*_{\max}}$.

Next, we show that $B^k(\widehat{S}, N) \geq B^k(S^*, N)$ for all $k \in M$. Letting $S_g^* = S^* \cap N_g$ and noting that $k_g^* = |S^* \cap N_g|$, the assortment S_g^* includes the k_g^* products that the assortment S^* offers in product class g. On the other hand, recalling that $\widehat{S}_g = S_g^{\mathsf{light}}(k_g^*)$ by the discussion at the beginning of this section, the assortment $\widehat{S}_g$ includes the k_g^* products with the <u>*smallest*</u> preference weights in product class g. Thus, for each $i \in S_g^*$, there exists a different $j(i) \in \widehat{S}_g$ such that $v_i \geq v_{j(i)}$. In other words, we can map a product in S_g^* to a product in $\widehat{S}_g$ with a smaller preference weight and this mapping is one-to-one. In this case, a simple lemma, given as Lemma C.1 in Appendix C, shows that we have $B^k(\cup_{g=g^*_{\min}}^{g^*_{\max}} S_g^*, N) \leq B^k(\cup_{g=g^*_{\min}}^{g^*_{\max}} \widehat{S}_g, N)$ for all $k \in M$. The proof of this lemma uses the fact that $\frac{v_i}{1+V(N_{-i})}$ in (1) satisfies $\frac{v_i}{1+V(N_{-i})} = \frac{v_i}{1+V(N)-v_i} \geq \frac{v_j}{1+V(N)-v_j} = \frac{v_j}{1+V(N_{-j})}$ whenever $v_i \geq v_j$. Lastly, by the discussion at the beginning of the proof of Theorem 5.1, we have $B^k(Q, N) \leq B^k(S, N)$ for all $S \subseteq Q \subseteq N$ and $k \in M$. In this case, since $\cup_{g=g^*_{\min}}^{g^*_{\max}} S_g^* \subseteq \cup_{g \in \mathbb{Z}} S_g^* = S^* \subseteq N$, we obtain $B^k(S^*, N) \leq B^k(\cup_{g=g^*_{\min}}^{g^*_{\max}} S_g^*, N) \leq B^k(\cup_{g=g^*_{\min}}^{g^*_{\max}} \widehat{S}_g, N) = B^k(\widehat{S}, N)$, where the last equality is by (10). ■

Thus, by the proposition above, there exists an ideal light assortment $\widehat{S}$ that satisfies the bounded support and limited degradation properties in Part 1.

6.2 Constructing the Candidate Assortments

We construct the collection of candidate assortments $\{A_t : t \in \mathcal{A}\}$ that satisfies the small and light collection and approximation to ideal properties in Part 2. Considering the ideal light assortment $\widehat{S}$ defined as in (10), our approach is based on guessing two quantities. First, we guess a value of

$\widehat{\eta}$ such that $\widehat{\eta} \leq \overline{W}(\widehat{S}) \leq 2\,\widehat{\eta}$. Second, recalling that we set $L = \lceil \frac{\log(n/\epsilon)}{\log(1+\epsilon)} \rceil$ at the beginning of this section, we guess the integer values of $\widehat{\kappa} = \{\widehat{\kappa}_g : g \in \mathbb{Z}\}$ such that

$$\widehat{\kappa}_g\, \epsilon \frac{\widehat{\eta}}{L} \;\leq\; \overline{W}(\widehat{S} \cap N_g) \;<\; (\widehat{\kappa}_g + 1)\, \epsilon \frac{\widehat{\eta}}{L} \quad \forall\, g \in \mathbb{Z}. \tag{11}$$

Shortly in this section, we argue that even if we do not know the ideal light assortment $\widehat{S}$, we can come up with $O((n/\epsilon)^{O(1/\epsilon^2)})$ guesses for $(\widehat{\eta}, \widehat{\kappa})$ such that one of these guesses ends up being correct; that is, satisfying $\widehat{\eta} \leq \overline{W}(\widehat{S}) \leq 2\,\widehat{\eta}$ and (11). Before we make this argument, however, we proceed to showing how we can use these guesses for $(\widehat{\eta}, \widehat{\kappa})$ to construct a collection of candidate assortments $\{A_t : t \in \mathcal{A}\}$ that satisfies the small and light collection and approximation to ideal properties in Part 2. Let $\{(\eta, \kappa) \in \Theta\}$ be the set of our guesses for $(\widehat{\eta}, \widehat{\kappa})$, so that $|\Theta| = O((n/\epsilon)^{O(1/\epsilon^2)})$. We construct one candidate assortment for each guess $(\eta, \kappa) \in \Theta$. In particular, by the discussion at the beginning of Section 5, the assortment $S_g^{\mathsf{light}}(k)$ includes the k products with the smallest preference weights in product class g. Letting $\widetilde{k}_g(\eta, \kappa_g) = \min\{k = 0, 1, \ldots, |N_g| : \overline{W}(S_g^{\mathsf{light}}(k)) \geq \kappa_g\, \epsilon \frac{\eta}{L}\}$, we define the assortment $\widetilde{S}_g(\eta, \kappa_g) = S_g^{\mathsf{light}}(\widetilde{k}_g(\eta, \kappa_g))$, which corresponds to the smallest light assortment in product class g with a total rounded weight of at least $\kappa_g\, \epsilon \frac{\eta}{L}$. In our collection of candidate assortments, the candidate assortment corresponding to the guess $(\eta, \kappa) \in \Theta$ is given by

$$\widetilde{S}(\eta, \kappa) \;=\; \bigcup_{g \in \mathbb{Z}} \widetilde{S}_g(\eta, \kappa_g), \tag{12}$$

which is, by definition, light. Thus, the collection of assortments $\{\widetilde{S}(\eta, \kappa) : (\eta, \kappa) \in \Theta\}$ includes $O((n/\epsilon)^{O(1/\epsilon^2)} \log n)$ light assortments, so it satisfies the small and light collection property.

In the next proposition, we show that the collection of assortments $\{\widetilde{S}(\eta, \kappa) : (\eta, \kappa) \in \Theta\}$ satisfies the approximation to ideal property as well.

Proposition 6.2 (Approximation to Ideal) *Letting $\widehat{S}$ be as in (10) and defining $\widehat{G} = \{g \in \mathbb{Z} : \widehat{S} \cap N_g \neq \varnothing\}$, there exists an assortment $\widetilde{S} \in \{\widetilde{S}(\eta, \kappa) : (\eta, \kappa) \in \Theta\}$ such that*

$$\overline{W}(\widetilde{S} \cap N_g) \;\leq\; \overline{W}(\widehat{S} \cap N_g) \;\leq\; \overline{W}(\widetilde{S} \cap N_g) + \frac{\epsilon}{|\widehat{G}|}\, \overline{W}(\widehat{S}) \quad \forall\, g \in \mathbb{Z}.$$

Proof: Let $(\widehat{\eta}, \widehat{\kappa}) \in \Theta$ be the correct guess in the sense that it satisfies $\widehat{\eta} \leq \overline{W}(\widehat{S}) \leq 2\,\widehat{\eta}$ and $\widehat{\kappa}_g\, \epsilon \frac{\widehat{\eta}}{L} \leq \overline{W}(\widehat{S} \cap N_g) < (\widehat{\kappa}_g + 1)\, \epsilon \frac{\widehat{\eta}}{L}$ for all $g \in \mathbb{Z}$. By our construction of the set of guesses, there exists such $(\widehat{\eta}, \widehat{\kappa}) \in \Theta$. Since $\widetilde{S}(\widehat{\eta}, \widehat{\kappa}) = \cup_{g \in \mathbb{Z}} \widetilde{S}_g(\widehat{\eta}, \widehat{\kappa}_g)$ by (12), we have $\widetilde{S}(\widehat{\eta}, \widehat{\kappa}) \cap N_g = \widetilde{S}_g(\widehat{\eta}, \widehat{\kappa}_g)$. Noting that $\widehat{S}$ is a light assortment, $\widehat{S} \cap N_g$ is a light assortment in product class g. Furthermore, we have $\overline{W}(\widehat{S} \cap N_g) \geq \widehat{\kappa}_g\, \epsilon \frac{\widehat{\eta}}{L}$, so the total rounded weight of $\widehat{S} \cap N_g$ is at least $\widehat{\kappa}_g\, \epsilon \frac{\widehat{\eta}}{L}$. On the other hand, by its definition, $\widetilde{S}_g(\widehat{\eta}, \widehat{\kappa}_g)$, is the <u>smallest</u> light assortment in product class g with a total rounded weight of at least $\widehat{\kappa}_g\,\epsilon\frac{\widehat{\eta}}{L}$. Therefore, it follows that $\overline{W}(\widehat{S}\cap N_g)\geq \overline{W}(\widetilde{S}_g(\widehat{\eta},\widehat{\kappa}_g))\geq \widehat{\kappa}_g\,\epsilon\frac{\widehat{\eta}}{L}$. Lastly, by (10), $\widehat{S}\cap N_g$ can be nonempty only for $g\in\{g^*_{\min},\ldots,g^*_{\max}\}$. Since $g^*_{\max}-g^*_{\min}=L-1$, there are L elements in the set $\{g^*_{\min},\ldots,g^*_{\max}\}$, so $|\widehat{G}|\leq L$. In this case, we get

$$\overline{W}(\widetilde{S}_g(\widehat{\eta},\widehat{\kappa}_g)) \;\leq\; \overline{W}(\widehat{S}\cap N_g) \;\stackrel{(a)}{<}\; (\widehat{\kappa}_g+1)\,\epsilon\frac{\widehat{\eta}}{L} \;\leq\; \overline{W}(\widetilde{S}_g(\widehat{\eta},\widehat{\kappa}_g))+\epsilon\frac{\widehat{\eta}}{L} \;\stackrel{(b)}{\leq}\; \overline{W}(\widetilde{S}_g(\widehat{\eta},\widehat{\kappa}_g))+\frac{\epsilon}{|\widehat{G}|}\overline{W}(\widehat{S}),$$

where (a) and (b) hold since $\overline{W}(\widehat{S}\cap N_g)<(\widehat{\kappa}_g+1)\,\epsilon\frac{\widehat{\eta}}{L}$ and $\overline{W}(\widehat{S})\geq\widehat{\eta}$ by our choice of $(\widehat{\kappa},\widehat{\eta})$ at the beginning of the proof. Letting $\widetilde{S}=\widetilde{S}(\widehat{\eta},\widehat{\kappa})$, by (12), we have $\widetilde{S}\cap N_g=\widetilde{S}_g(\widehat{\eta},\widehat{\kappa}_g)$. So, by the chain of inequalities above, the assortment $\widetilde{S}$ satisfies the chain of inequalities in the proposition. ■

In the remainder of this section, we explain how we can come up with $O((n/\epsilon)^{O(1/\epsilon^2)})$ guesses for $(\widehat{\eta},\widehat{\kappa})$ such that one of these guesses satisfy $\widehat{\eta}\leq\overline{W}(\widehat{S})\leq 2\,\widehat{\eta}$, as well as (11).

Counting the Number of Guesses:

We guess the largest product class $\widehat{g}$ in which the assortment $\widehat{S}$ offers a product. Since there are n products, there are at most n nonempty product classes, so we have $O(n)$ guesses for $\widehat{g}$. Considering the number of guesses for $\widehat{\eta}$, by the definition of N_g, we have $\overline{w}_i=(1+\epsilon)^g$ for all $i\in N_g$. Since $\widehat{g}$ is the largest product class in which the assortment $\widehat{S}$ offers a product, we have $\overline{w}_i=(1+\epsilon)^{\widehat{g}}$ for some $i\in\widehat{S}$ and $\overline{w}_i\leq(1+\epsilon)^{\widehat{g}}$ for all $i\in\widehat{S}$, which implies that $(1+\epsilon)^{\widehat{g}}\leq\overline{W}(\widehat{S})\leq n\,(1+\epsilon)^{\widehat{g}}$. In this case, there exists some $q=0,1,\ldots,\lceil\frac{\log n}{\log 2}\rceil$ such that $2^q\,(1+\epsilon)^{\widehat{g}}\leq\overline{W}(\widehat{S})\leq 2^{q+1}\,(1+\epsilon)^{\widehat{g}}$. So, for each guess of $\widehat{g}$, our guess of $\widehat{\eta}$ has the form $2^q\,(1+\epsilon)^{\widehat{g}}$ for $q=0,1,\ldots,\lceil\frac{\log n}{\log 2}\rceil$. In this way, for each guess of $\widehat{g}$, we have $O(\log n)$ guesses for $\widehat{\eta}$ such that one of these guesses satisfies $\widehat{\eta}\leq\overline{W}(\widehat{S})\leq 2\,\widehat{\eta}$.

Considering the number of guesses for $\widehat{\kappa}$, by (10), the assortment $\widehat{S}$ offers products in at most L consecutive product classes. Thus, for each guess of $\widehat{g}$, we can set $\widehat{\kappa}_g=0$ in (11) for all $g\neq\widehat{g}-L+1,\ldots,\widehat{g}$. Also, if our guess of $(\widehat{\eta},\widehat{\kappa})$ is to satisfy $\overline{W}(\widehat{S})\leq 2\,\widehat{\eta}$ and $\widehat{\kappa}_g\,\epsilon\frac{\widehat{\eta}}{L}\leq\overline{W}(\widehat{S}\cap N_g)$ for all $g\in\mathbb{Z}$, then we need to have $\epsilon\frac{\widehat{\eta}}{L}\sum_{g\in\mathbb{Z}}\widehat{\kappa}_g\leq\sum_{g\in\mathbb{Z}}\overline{W}(\widehat{S}\cap N_g)=\overline{W}(\widehat{S})\leq 2\,\widehat{\eta}$, which implies that $\sum_{g\in\mathbb{Z}}\widehat{\kappa}_g\leq\lceil\frac{2\,L}{\epsilon}\rceil$. In this case, since we can set $\widehat{\kappa}_g=0$ for all $g\neq\widehat{g}-L+1,\ldots,\widehat{g}$, we need to have $\sum_{g=\widehat{g}-L+1}^{\widehat{g}}\widehat{\kappa}_g\leq\lceil\frac{2\,L}{\epsilon}\rceil$. Therefore, for each guess of $\widehat{g}$, the number of guesses for $\widehat{\kappa}$ is upper bounded by the number of ways to divide $\lceil\frac{2\,L}{\epsilon}\rceil$ items into L bins and the latter quantity is given by $\binom{\lceil\frac{2\,L}{\epsilon}\rceil+L-1}{L}$. Thus, for each guess of $\widehat{g}$, the number of guesses for $\widehat{\kappa}$ is at most

$$\binom{\lceil\frac{2L}{\epsilon}\rceil+L-1}{L} \;\stackrel{(a)}{\leq}\; 2^{\lceil\frac{2L}{\epsilon}\rceil+L-1} \;\leq\; 2^{\frac{3L}{\epsilon}} \;\stackrel{(b)}{\leq}\; 2^{\frac{9}{\epsilon^2}\log(n/\epsilon)} \;=\; 2^{\log(n/\epsilon)^{9/\epsilon^2}} \;=\; \left(\frac{n}{\epsilon}\right)^{O(1/\epsilon^2)},$$

here (a) uses the fact that $\binom{n}{k}\leq 2^n$ and (b) holds since $L=\lceil\frac{\log(n/\epsilon)}{\log(1+\epsilon)}\rceil\leq\lceil\frac{\log(n/\epsilon)}{\epsilon/2}\rceil\leq\frac{3}{\epsilon}\log(n/\epsilon)$, where we use the inequalities $\log(1+x)\geq x/2$ for $x\in[0,1]$ and $\log(n/\epsilon)\geq 1$ for $n\geq 3$.

For each guess of $\widehat{g}$, we have $O(\log n)$ guesses for $\widehat{\eta}$ and $(n/\epsilon)^{O(1/\epsilon^2)}$ guesses for $\widehat{\kappa}$. Since there are $O(n)$ guesses for $\widehat{g}$, we end up with $O((n/\epsilon)^{O(1/\epsilon^2)}\log n)=O((n/\epsilon)^{O(1/\epsilon^2)})$ guesses for $(\widehat{\eta},\widehat{\kappa})$.

7. Prediction Performance with Endogenous Consideration Sets

We give computational experiments to understand how much consideration sets improve the ability of the multinomial logit model to predict customer purchases and to identify profitable assortments.

7.1 Experimental Setup

We generate purchase histories from a ground choice model that does not comply with the multinomial logit model and test the ability of our choice model to predict the purchases in these histories. The ground choice model is the non-parametric choice model. In this choice model, we have C customer types. Customers of type ℓ are characterized by a preference list $(j^\ell(1),\ldots,j^\ell(n^\ell))$, where n^ℓ is the number of products in the list and $j^\ell(k)$ is the product at position k. A customer of type ℓ arrives with probability β^ℓ. She purchases her most preferred product that is available. If no product in her preference list is available, then she leaves without a purchase. Thus, the parameters of the ground choice model are the arrival probability β^ℓ and the preference list $(j^\ell(1),\ldots,j^\ell(n^\ell))$ for each customer type ℓ. We generate instances of the ground choice model as follows.

We index the products such that product 1 has the highest quality and price, whereas product n has the lowest quality and price. Customers of a particular type have a highest willingness to pay and lowest quality threshold. Thus, the preference lists are of the form $(i, i+1,\ldots,j)$, but we introduce some idiosyncratic behavior. In particular, to generate the preference list of customers of type ℓ, we sample L^ℓ from the uniform distribution over $\{1,\ldots,n\}$ and U^ℓ from the uniform distribution over $\{L^\ell,\ldots,n\}$. Considering the preference list $(L^\ell, L^\ell+1,\ldots,U^\ell)$, we drop each of the products with probability 0.1. After dropping the products, with probability 0.5, we randomly pick just one product in the preference list and flip its ordering with its successor to get the preference list of customers of type ℓ. With probability 0.5, we leave the preference list untouched.

Customers of each type arrive with equal probability of $\beta^\ell = 1/C$. Throughout, we have $n = 10$ products and $C = 100$ customer types. Once we generate an instance of the ground choice model as above, we sample the purchase histories of τ customers making choices according to the ground choice model. In particular, we capture a sampled purchase history by $\{(S_t, i_t) : t = 1,\ldots,\tau\}$, where S_t is the assortment offered to customer t and i_t is the product purchased, if any, by customer t. To come up with the assortment S_t, we include each product in the assortment with probability 0.5. We sample the product i_t within the assortment S_t according to the ground choice model. We use these past purchase histories as training data. To use as validation and testing data, we follow the same approach to generate two other purchase histories, each including 1250 customers.

We use maximum likelihood estimation to fit a multinomial logit model with endogenous consideration sets to the training dataset; see, for example, Vulcano et al. (2012). The parameters of this choice model are the preference weights $(v_1, \ldots, v_n)$ and the consideration depth distribution $(\lambda^1, \ldots, \lambda^m)$. We use cross-validation to choose the value of m. When we have endogenous consideration sets, the log-likelihood function is not concave in the parameters of the choice model. We use the fmincon routine in Matlab with multiple initial solutions to obtain a local maximum of the log-likelihood function. As a benchmark, we fit a standard multinomial logit model to the training dataset. The parameters of this choice model are only the preference weights $(v_1, \ldots, v_n)$. Throughout this section, we use CS and ST to, respectively, refer to the fitted multinomial logit model with endogenous consideration sets and the fitted standard multinomial logit model, where CS stands for consideration sets and ST stands for standard.

7.2 Predicting Customer Purchases

We use the following approach to compare the ability of CS and ST to predict the purchases of the customers. We generate an instance of the ground choice model as discussed earlier in this section. Using the ground choice model, we generate three training datasets by varying the number of customers in the purchase history over $\tau \in \{1000, 1750, 2500\}$. In this way, we obtain three levels of data availability to fit CS and ST. To each of the three training datasets, we fit CS and ST by using maximum likelihood estimation. We compare the two fitted choice models by using two performance measures. The first performance measure is the out-of-sample log-likelihoods on the testing dataset. The second performance measure is deviation between the choice probabilities under the fitted choice models and the actual ground choice model.

In Table 2, we compare the out-of-sample log-likelihoods of CS and ST. We replicated our computational experiments for 10 ground choice models that we randomly generate. Each row in the table corresponds to a different ground choice model that we work with. In the first column of the table, we give the index of the ground choice model. There are three blocks of three columns in the rest of the table. Each block corresponds to a different value for the number of customers τ in the training dataset, yielding different levels of data availability to fit the choice models. In each block, the first two columns show the out-of-sample log-likelihoods of CS and ST, whereas the third column shows the percent gap between the two log-likelihoods. In all of the tables that we give throughout this section, positive values for the percent gaps favor CS.

The results in Table 2 indicate that CS provides consistent improvements over ST in terms of out-of-sample log-likelihoods. When we have $\tau = 1000$, 1750 and 2500 customers in the training dataset, the average gaps in the out-of-sample log-likelihoods of the two fitted choice models are respectively 0.39%, 0.47% and 0.48%. As the level of data availability increases, CS provide

Grnd. Ch.	$\tau = 1000$ CS	ST	Perc. Gap	$\tau = 1750$ CS	ST	Perc. Gap	$\tau = 2500$ CS	ST	Perc. Gap
1	-2,018.96	-2,031.06	0.60	-2,014.88	-2,026.02	0.55	-2,011.29	-2,022.62	0.56
2	-2,138.26	-2,140.47	0.10	-2,136.50	-2,140.21	0.17	-2,134.17	-2,137.50	0.16
3	-2,071.94	-2,079.36	0.36	-2,066.73	-2,074.21	0.36	-2,066.35	-2,073.67	0.35
4	-2,038.86	-2,046.14	0.36	-2,041.39	-2,049.98	0.42	-2,041.41	-2,050.40	0.44
5	-2,048.19	-2,052.55	0.21	-2,036.74	-2,050.78	0.69	-2,036.21	-2,050.36	0.69
6	-2,045.22	-2,057.04	0.58	-2,036.14	-2,050.64	0.71	-2,036.84	-2,052.45	0.77
7	-2,066.56	-2,079.30	0.62	-2,061.09	-2,074.48	0.65	-2,062.92	-2,076.18	0.64
8	-2,067.29	-2,074.95	0.37	-2,067.72	-2,076.62	0.43	-2,067.04	-2,076.94	0.48
9	-2,093.47	-2,104.09	0.51	-2,092.89	-2,104.12	0.54	-2,088.47	-2,099.43	0.52
10	-2,051.99	-2,055.88	0.19	-2,049.73	-2,052.25	0.12	-2,049.27	-2,051.92	0.13
Avg.			0.39			0.47			0.48

Table 2 Comparison of out-of-sample log-likelihoods of CS and ST.

slightly more noticeable improvements over ST. We shortly demonstrate that these improvements in log-likelihoods translate into more profitable assortments. Considering our results under 10 ground choice models, when we fit CS to the training data, the maximum consideration depth of a customer comes out to be four to five. Fitting CS using maximum likelihood estimation takes about four to six minutes depending on the number of customers in the training data.

Note that ST has n parameters, whereas CS has $n+m$. Furthermore, ST is a special case of CS obtained by setting $n=m$, $\lambda_n = 1$ and $\lambda^k = 0$ for all $k = 1, \ldots, n-1$. Therefore, CS has more modeling flexibility than ST to capture the choice process of the customers, so CS always provides larger in-sample log-likelihoods than ST, but it is not guaranteed that CS provides larger out-of-sample log-likelihoods. With its larger number of parameters, CS may overfit to the training data and may yield poor out-of-sample performance, especially when we do not have enough training data; see Section 1.2 in Bishop (2006). However, in our datasets, overfitting does not seem to be a concern for CS even with the smallest data availability of $\tau = 1000$. As τ increases to 1750 and 2500, the gap between the out-of-sample log-likelihoods slightly increases.

In Table 3, we give the mean absolute errors of the choice probabilities under the fitted choice models when compared with the ground choice model. In particular, we use $\phi_i^{\mathsf{CS}}(S)$ and $\phi_i^{\mathsf{GR}}(S)$ to denote the choice probabilities of product i within assortment S, respectively, under the fitted CS and the ground choice models. Letting $\{S_t : t = 1, \ldots, 1250\}$ be the assortments in the testing dataset, we compute the mean absolute error for CS as $\frac{1}{1250}\sum_{t=1}^{1250}\sum_{i\in S_t}\frac{1}{|S_t|}|\phi_i^{\mathsf{CS}}(S) - \phi_i^{\mathsf{GR}}(S)|$. We focus on mean absolute errors rather than mean absolute percent errors, because if a product has a small purchase probability, then misestimating this purchase probability even by a small amount may increase the mean absolute percent error substantially, putting disproportionate weight on estimating small choice probabilities more accurately. We compute the mean absolute error for the

Grnd. Ch.	$\tau = 1000$ CS	ST	Perc. Gap	$\tau = 1750$ CS	ST	Perc. Gap	$\tau = 2500$ CS	ST	Perc. Gap
1	0.0595	0.0607	2.02	0.0584	0.0589	0.83	0.0583	0.0588	0.95
2	0.0518	0.0518	0.08	0.0494	0.0494	0.06	0.0493	0.0496	0.58
3	0.0529	0.0540	2.08	0.0524	0.0539	2.80	0.0516	0.0528	2.27
4	0.0581	0.0586	0.82	0.0581	0.0586	0.81	0.0581	0.0586	0.90
5	0.0556	0.0554	-0.41	0.0550	0.0547	-0.49	0.0546	0.0544	-0.32
6	0.0554	0.0565	1.96	0.0537	0.0546	1.72	0.0537	0.0545	1.46
7	0.0536	0.0546	1.98	0.0530	0.0540	1.96	0.0536	0.0551	2.78
8	0.0504	0.0525	4.04	0.0494	0.0512	3.54	0.0492	0.0508	3.16
9	0.0592	0.0599	1.14	0.0595	0.0601	0.90	0.0581	0.0588	1.22
10	0.0529	0.0540	2.07	0.0517	0.0529	2.33	0.0522	0.0530	1.48
Avg.			1.58			1.45			1.45

Table 3 Comparison of mean absolute errors in choice probabilities of CS and ST.

choice probabilities under ST similarly. The layout of Table 3 is identical to that of Table 2, except that the first two columns in each block give the mean absolute errors for CS and ST. The results in Table 3 indicate that CS provides improvements over ST in terms of mean absolute errors as well. For one ground choice model, CS lags behind ST in terms of mean absolute error, but the gaps over the other ground choice models can reach 4.04%.

7.3 Picking Profitable Assortments

The improvements in the out-of-sample log-likelihoods and the mean absolute errors provided by CS over ST can translate into significantly more profitable assortments. We use the following approach to compare the ability of CS and ST to pick profitable assortments. Recall that $\phi_i^{\mathsf{CS}}(S)$ and $\phi_i^{\mathsf{GR}}(S)$ are the choice probabilities of product i within assortment S, respectively, under the fitted CS and the ground choice models. Once we fit CS to the training dataset, we generate 100 samples of product revenues, denoted by $\{(r_{1k}, \ldots, r_{nk}) : k = 1, \ldots, 100\}$. Each product revenue is generated from the uniform distribution over $[1, 10]$. For each sample $(r_{1k}, \ldots, r_{nk})$, we solve the problem $S_k^{\mathsf{CS}} = \arg\max_{S \subseteq N} \sum_{i \in N} r_{ik}\, \phi_i^{\mathsf{CS}}(S)$, which is the optimal assortment if the choices of the customers were governed by the fitted CS model. The choices of the customers are actually governed by the ground choice model, so the actual expected revenue from the assortment S_k^{CS} is $\mathsf{Rev}_k^{\mathrm{SC}} = \sum_{i \in N} r_{ik}\, \phi_i^{\mathsf{GR}}(S_k^{\mathsf{CS}})$, characterizing the revenue performance of CS on the sample $(r_{1k}, \ldots, r_{nk})$. We compute $\mathsf{Rev}_k^{\mathsf{ST}}$ similarly to characterize the revenue performance of ST on the same sample.

In Table 4, we compare the expected revenues from the assortments picked by CS and ST. As earlier, there are three bocks in Table 4, each block corresponding to a different value for the number of customers τ in the training dataset. In each block, the first column gives the average percent gap between the expected revenues obtained by CS and ST, where the average is computed over the 100 product revenue samples. Thus, this column gives the average of the data $\{100 \times \frac{\mathsf{Rev}_k^{\mathsf{CS}} - \mathsf{Rev}_k^{\mathsf{ST}}}{\mathsf{Rev}_k^{\mathsf{CS}}} : k =$

Grnd. Ch.	$\tau = 1000$				$\tau = 1750$				$\tau = 2500$			
	Perc. Gap	Std. Err.	CS ≻ ST	ST ≻ CS	Perc. Gap	Std. Err.	CS ≻ ST	ST ≻ CS	Perc. Gap	Std. Err.	CS ≻ ST	ST ≻ CS
1	7.54	1.73	53	20	7.41	1.77	52	25	8.04	1.71	53	20
2	4.23	1.36	47	32	3.76	1.34	45	31	5.64	1.65	46	29
3	3.97	1.17	52	22	5.85	1.91	49	17	3.50	1.13	50	23
4	4.92	1.57	48	21	6.02	1.66	53	23	5.71	1.72	51	29
5	5.81	1.62	50	29	5.60	1.53	49	24	6.09	1.56	50	23
6	6.94	1.86	40	12	5.60	1.70	41	19	5.43	1.74	44	26
7	3.29	1.37	48	28	3.98	1.41	50	29	5.51	1.47	53	25
8	7.30	2.01	52	28	5.17	1.75	46	29	5.21	1.78	45	35
9	6.31	1.87	49	21	6.48	1.87	53	23	5.67	1.80	49	21
10	6.98	2.12	51	18	7.93	2.28	52	22	6.54	2.15	53	31
Avg.	5.73	1.67	49	23	5.78	1.72	49	24	5.73	1.67	49	26

Table 4 Comparison of expected revenues from the assortments obtained by using CS and ST.

$1,\ldots,100\}$. The second column gives the standard error of the percent gaps, corresponding to the standard deviation of the same data divided by the square root of 100. The last two columns show the number of samples for which we have $\mathsf{Rev}_k^{\mathsf{CS}} > \mathsf{Rev}_k^{\mathsf{ST}}$ and $\mathsf{Rev}_k^{\mathsf{ST}} < \mathsf{Rev}_k^{\mathsf{CS}}$, giving the numbers of samples for which one fitted choice model performs better than the other. The entries in these columns may not add up to 100, since we may have $\mathsf{Rev}_k^{\mathsf{CS}} = \mathsf{Rev}_k^{\mathsf{ST}}$. The results in Table 4 indicate that CS provides significant improvements over ST in terms of expected revenues. Over a total of 3000 product revenue samples, in 1474 samples, the revenue performance of CS is better than that of ST, whereas in 735 samples, the revenue performance of ST is better than that of CS. The assortments picked under the assumption that the customers choose according to the fitted CS can provide as much as 8.04% improvement in the average expected revenue.

The results in Tables 2, 3 and 4 are based one purchase history. To ensure that our results are robust, we replicated them for 10 different purchase histories sampled from each of the 10 ground choice models. We work with three levels of data availability, so we end up with 300 ground choice model-purchase history-data availability combinations. To provide high level statistics, over all of the 300 combinations, the average percent gap between the out-of-sample log-likelihoods of CS and ST is 0.41%. The average percent gaps in mean absolute errors and expected revenues obtained by using CS and ST are, respectively, 1.41% and 5.93%. Over the 300 combinations, in 292 and 258 combinations, the out-of-sample log-likelihoods and mean absolute errors of CS are, respectively, better than those of ST. In 24% of the product revenue samples, the expected revenue performance ST is better than that of CS, whereas in 52% of the samples, CS is better.

We also compared the performance of CS and ST using a dataset that includes the preferences of 5000 diners for 10 sushi varieties; see Kamishima (2018). We report the results of these computational experiments in Appendix D. In these computational experiments, the improvements of CS over ST turn out to be more modest, but they are still consistent.

Param. (θ, γ)	Avg.	Std. Dev.	90th Perc.	Max.
(40, 1)	0.37	0.51	0.74	2.53
(40, 10)	0.34	0.49	0.69	2.45
(40, 50)	0.25	0.43	0.52	2.19
Avg.	0.32	0.47	0.65	2.39

Param. (θ, γ)	Avg.	Std. Dev.	90th Perc.	Max.
(50, 1)	0.28	0.58	0.75	2.62
(50, 10)	0.27	0.58	0.74	2.63
(50, 50)	0.27	0.59	0.78	2.70
Avg.	0.27	0.58	0.75	2.65

Param. (θ, γ)	Avg.	Std. Dev.	90th Perc.	Max.
(60, 1)	0.20	0.48	0.54	2.35
(60, 10)	0.21	0.48	0.56	2.37
(60, 50)	0.30	0.51	0.75	2.45
Avg.	0.24	0.49	0.62	2.39

Table 5 Performance of our PTAS.

8. Performance of the Approximation Scheme

We test the performance of the PTAS given Section 5 by comparing the expected revenues from the assortments obtained by the PTAS with an upper bound on the optimal expected revenues.

Experimental Setup: In our computational experiments, we randomly generate a large number of problem instances. For each problem instance, we use our PTAS to obtain an approximate solution to problem (3). In Appendix E, we give an efficient approach to obtain an upper bound on the optimal expected revenue. For each problem instance, we use this approach to obtain an upper bound on the optimal expected revenue. We compare the expected revenue from the solution provided by our PTAS with the upper bound on the optimal expected revenue. We use the following approach to generate our problem instances. In all of our problem instances, we have $n = 25$ products. Each product is one of three types. A product either has high revenue and low preference weight, or low revenue and high preference weight, or low revenue and low preference weight. In this way, we eliminate the possibility of a product with high revenue and high preference weight, which is a clear candidate to include in the optimal assortment. We generate the high and low preference weights, respectively, from the uniform distribution over $[\frac{1}{\gamma}100, \frac{1}{\gamma}200]$ and $[\frac{1}{\gamma}10, \frac{1}{\gamma}20]$, where γ is a parameter that we vary. Noting that the preference weight of the no purchase option is fixed at one, a larger value of γ implies that the customers are more likely to leave without a purchase. We generate the high and low revenues, respectively, from the uniform distribution over $[150, 200]$ and $[\theta, \theta + 10]$, where θ is another parameter that we vary. The consideration depth of all customers is $m = 2$. We use $\epsilon = 0.7$ in our PTAS, yielding a performance guarantee of less than 50%, but even this value of ϵ provides solutions with remarkably high quality.

Varying (γ, θ) over $\{1, 10, 50\} \times \{40, 50, 60\}$, we obtain nine parameter configurations. In each parameter configuration, we generate 50 problem instances.

Computational Results: We give our numerical results in Table 5. In this table, the first column shows the parameter configuration using the pair (θ, γ). Recalling that we generate 50 problem instances in each parameter configuration, the second column shows the average percent gap between the expected revenue from the assortment obtained by our PTAS and the upper bound on the optimal expected revenue. In other words, letting Rev^k be the expected revenue from the assortment obtained by our PTAS for problem instance k and UB^k be the upper bound on the optimal expected revenue, the first column gives the average of the data $\{100 \times \frac{\mathsf{UB}^k - \mathsf{Rev}^k}{\mathsf{UB}^k} : k = 1, \ldots, 50\}$. The third, fourth and fifth columns, respectively, give the standard deviation, 90th percentile and maximum of the same data.

The result in Table 5 indicate that our PTAS performs remarkably well. The average optimality gap is 0.28%. Over all problem instances, the CPU times for our PTAS range between eight to ten seconds. In more than 90% of the problem instances, the optimality gap is no larger than 0.78%, even when we compare the expected revenue obtained by our PTAS with an upper bound on the optimal expected revenue, rather than the optimal expected revenue itself.

9. Conclusions

Our work opens up several areas of investigation. We gave a PTAS for the assortment optimization problem, but our complexity result does not rule out the existence of a fully polynomial-time approximation scheme. Either establishing that a fully polynomial-time approximation scheme is not possible or giving such a scheme is one research path. Also, the structure of the choice probabilities under our choice model is significantly different when compared with the standard multinomial logit and nested logit models, so we needed to develop new techniques to study the assortment optimization problem. Our PTAS naturally extends to handle the case where we have a cardinality constraint that limits the number of offered products. In particular, we generate the candidate assortments in the same way, but focus on those that satisfy the cardinality constraint. However, our PTAS does not extend to handle other types of constraints, including a knapsack constraint, where each product consumes a certain amount of space and we have a limit on the total space consumption of the offered products. Both giving a fully polynomial-time approximation scheme and dealing with constraints appear to require new lines of attack and our efforts in these directions have so far been unfruitful. Furthermore, our approach to estimate the parameters of our choice model was based on maximizing the log-likelihood function but this function is non-concave in the parameters of our choice model. One can study more sophisticated methods to estimate the parameters of our choice model. Lastly, incorporating consideration sets to enhance the ability of choice models to predict customer choices is a rich area, irrespective of whether the starting point is the multinomial logit or any other choice model.

References

Aouad, A., V. Farias, R. Levi. 2020. Assortment optimization under consider-then-choose choice models. *Management Science* (forthcoming).

Aouad, A., V. Farias, R. Levi, D. Segev. 2018. 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.

Beggs, S., S. Cardell, J. Hausman. 1981. Assessing the potential demand for electric cars. *Journal of Econometrics* **17**(1) 1–19.

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

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.

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

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. 2014. Near-optimal algorithms for capacity constrained assortment optimization. Tech. rep., Columbia University, New York, NY.

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.

Gallego, G., G. Iyengar, R. Phillips, A. Dubey. 2004. Managing flexible products on a network. Tech. rep., Columbia University, New York, NY.

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., P. Rusmevichientong. 2017. A nonparametric joint assortment and price choice model. *Management Science* **63**(9) 3128–3145.

Jagabathula, S., G. Vulcano. 2018. A partial-order-based model to estimate individual preferences using panel data. *Management Science* **64**(4) 1609–1628.

Kamishima, T. 2018. Sushi preference datasets. URL `https://www.kamishima.net/sushi/`. Last access date: January 8, 2021.

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, Paat, David B. Shmoys, Chaoxu Tong, Huseyin Topaloglu. 2014. Assortment optimization under the multinomial logit model with random choice parameters. *POM* **23**(11) 2023–2039.

Sumida, M., G. Gallego, P. Rusmevichientong, H. Topaloglu, J. M. Davis. 2020. Revenue-utility tradeoff in assortment optimization under the multinomial logit model with totally unimodular constraints. *Management Science* (to appear).

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

Vulcano, G., G. van Ryzin, W. Chaar. 2010. OM practice – Choice-based revenue management: An empirical study of estimation and optimization. *M&SOM* **12**(3) 371–392.

Vulcano, G., G. J. van Ryzin, R. Ratliff. 2012. Estimating primary demand for substitutable products from sales transaction data. *Operations Research* **60**(2) 313–334.

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

Wang, R. 2019. Customer choice with consideration set: Threhold Luce model. Tech. rep., Johns Hopkins University, Baltimore, MD.

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

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

Online Appendix: Assortment Optimization under the Multinomial Logit Model in the Presence of Endogenous Consideration Sets

Appendix A: Independence from Top Ranked Choices

We give a proof of the independence from top ranked choices property. Beggs et al. (1981) give a result that implies the same property, but they do not state the property in the form we use.

Lemma A.1 *Letting* $\{X_1,\ldots,X_q,Y_1,\ldots,Y_p\}$ *be independent Gumbel random variables with the same scale parameter of one, we have*

$$\mathbb{P}\big\{X_1 \geq \ldots \geq X_q \,\big|\, Y_1 \geq \ldots \geq Y_p \geq \max\{X_1,\ldots,X_q\}\big\} \;=\; \mathbb{P}\big\{X_1 \geq \ldots \geq X_q\big\}.$$

Proof: Let X_i have the location-scale parameters $(\mu_i,1)$ and Y_i have the location-scale parameters $(\gamma_i,1)$. If $\{Z_1,\ldots,Z_n\}$ are independent Gumbel random variables with Z_i having the location-scale parameter $(\zeta_i,1)$, then we have $\mathbb{P}\{Z_1 \geq \ldots \geq Z_n\} = \frac{e^{\mu_1}}{\sum_{i=1}^n e^{\mu_i}} \frac{e^{\mu_2}}{\sum_{i=2}^n e^{\mu_i}} \cdots \frac{e^{\mu_{n-1}}}{\sum_{i=n-1}^n e^{\mu_i}}$; see, for example, Section 3.1 in Jagabathula and Vulcano (2018). In this case, using the fact that $\max\{X_1,\ldots,X_q\}$ has the Gumbel distribution with location-scale parameters $(\log\sum_{i=1}^q e^{\mu_i},1)$ by the maximum of Gumbel random variables property in Section 2, we obtain the chain of equalities

$$\begin{aligned}
\mathbb{P}\big\{X_1 \geq \ldots \geq X_q \,|\, Y_1 \geq \ldots \geq Y_p \geq \max\{X_1,\ldots,X_q\}\big\} \;&=\; \frac{\mathbb{P}\big\{Y_1 \geq \ldots \geq Y_p \geq X_1 \geq \ldots \geq X_q\big\}}{\mathbb{P}\big\{Y_1 \geq \ldots \geq Y_p \geq \max\{X_1,\ldots,X_q\}\big\}} \\
&\overset{(a)}{=} \frac{\frac{e^{\gamma_1}}{\sum_{i=1}^p e^{\gamma_i}+\sum_{i=1}^q e^{\mu_i}} \frac{e^{\gamma_2}}{\sum_{i=2}^p e^{\gamma_i}+\sum_{i=1}^q e^{\mu_i}} \cdots \frac{e^{\gamma_p}}{e^{\gamma_p}+\sum_{i=1}^q e^{\mu_i}} \frac{e^{\mu_1}}{\sum_{i=1}^q e^{\mu_i}} \frac{e^{\mu_2}}{\sum_{i=2}^q e^{\mu_i}} \cdots \frac{e^{\mu_{q-1}}}{\sum_{i=q-1}^q e^{\mu_i}}}{\frac{e^{\gamma_1}}{\sum_{i=1}^p e^{\gamma_i}+\sum_{i=1}^q e^{\mu_i}} \frac{e^{\gamma_2}}{\sum_{i=2}^p e^{\gamma_i}+\sum_{i=1}^q e^{\mu_i}} \cdots \frac{e^{\gamma_p}}{e^{\gamma_p}+\sum_{i=1}^q e^{\mu_i}}} \\
&= \frac{e^{\mu_1}}{\sum_{i=1}^q e^{\mu_i}} \frac{e^{\mu_2}}{\sum_{i=2}^q e^{\mu_i}} \cdots \frac{e^{\mu_{q-1}}}{\sum_{i=q-1}^q e^{\mu_i}} \overset{(b)}{=} \mathbb{P}\big\{X_1 \geq \ldots \geq X_q\big\},
\end{aligned}$$

where (a) and (b) use the identity at the beginning of the proof and noting that $\max\{X_1,\ldots,X_q\}$ is Gumbel with location-scale parameters $(\log\sum_{i=1}^q e^{\mu_i},1)$ and independent of $\{Y_1,\ldots,Y_p\}$. ■

Appendix B: Product Revenues in an Optimal Assortment

In the next lemma, we give a property of an optimal solution to the assortment optimization problem under the standard multinomial logit model.

Lemma B.1 *Letting* $\widetilde{S}$ *be an optimal solution to the assortment optimization problem in* (4)*, we have* $r_\ell \geq \frac{\sum_{i\in\widetilde{S}} r_i v_i}{1+\sum_{i\in\widetilde{S}} v_i}$ *for all* $\ell \in \widetilde{S}$.

Proof: Let $\widetilde{Z}$ be the optimal objective value of problem (4), so $\widetilde{Z} = \frac{\sum_{i\in\widetilde{S}} r_i v_i}{1+\sum_{i\in\widetilde{S}} v_i}$. In this case, arranging the terms in the last equality, we obtain $\widetilde{Z} = \sum_{i\in\widetilde{S}}(r_i - \widetilde{Z})\, v_i$. To get a contradiction, assume that there exists $\ell \in \widetilde{S}$ such that $r_\ell < \widetilde{Z}$. In this case, we obtain $\widetilde{Z} = \sum_{i\in\widetilde{S}}(r_i - \widetilde{Z})\, v_i < \sum_{i\in\widetilde{S}\setminus\{\ell\}}(r_i - \widetilde{Z})\, v_i$, yielding the inequality $\widetilde{Z} < \sum_{i\in\widetilde{S}\setminus\{\ell\}}(r_i - \widetilde{Z})\, v_i$. Arranging the terms in this inequality, we get $\widetilde{Z} < \frac{\sum_{i\in\widetilde{S}\setminus\{\ell\}} r_i v_i}{1+\sum_{i\in\widetilde{S}\setminus\{\ell\}} v_i}$. Thus, the solution $\widetilde{S}\setminus\{\ell\}$ provides an objective value for problem (4) that exceeds the optimal objective value for this problem, which is a contradiction. ■

Appendix C: Comparison of Choice Probability Boosts

In the next lemma, we give a monotonicity property of $B^k(S,N)$ computed through (1). We use this lemma in the proof of Proposition 6.1 when showing the limited degradation property.

Lemma C.1 *Consider $\widetilde{S}, \widehat{S} \subseteq N$ with $|\widetilde{S}| = |\widehat{S}|$, where for each $i \in \widetilde{S}$, there exists $j(i) \in \widehat{S}$ such that $v_i \geq v_{j(i)}$ and $j(i) \neq j(\ell)$ for each $i \neq \ell$. Then, we have $B^k(\widetilde{S},N) \leq B^k(\widehat{S},N)$ for all $k \in M$.*

Proof: By the assumption in the lemma, for each $i \in N \setminus \widetilde{S}$, there must exists $k(i) \in N \setminus \widehat{S}$ such that $v_i \leq v_{k(i)}$ and $k(i) \neq k(\ell)$ for each $i \neq \ell$. For $F \subseteq N$ with $F \cap \widetilde{S} = \varnothing$, we define $K(F)$ as $K(F) = \{k(i) : i \in F\}$. We use induction over the consideration depth to show that $B^k(\widetilde{S}, N\setminus F) \leq B^k(\widehat{S}, N \setminus K(F))$ for all $F \subseteq N$ with $F \cap \widetilde{S} = \varnothing$ and $k \in M$. Since $B^1(\cdot,\cdot) = 1$, the result holds for $k = 1$. Assuming that the result holds for consideration depth of $k-1$, we show that the result holds for consideration depth of k. For $i \notin \widetilde{S}$ and $i \notin F$, letting $F_{+i} = F \cup \{i\}$, since $F \cap \widetilde{S} = \varnothing$, we get $F_{+i} \cap \widetilde{S} = \varnothing$, so $B^{k-1}(\widetilde{S}, (N\setminus F)_{-i}) = B^{k-1}(\widetilde{S}, N \setminus F_{+i}) \leq B^{k-1}(\widehat{S}, N \setminus K(F_{+i})) = B^{k-1}(\widehat{S}, (N \setminus K(F))_{-k(i)})$, where the inequality follows from the induction assumption along with the fact that $F_{+i} \cap \widetilde{S} = \varnothing$ and the second equality uses the fact that $i \notin F$ and $k(i) \neq k(\ell)$ for $i \neq \ell$, so $K(F_{+i}) = K(F) \cup \{k(i)\}$. In this case, by the definition of $B^k(S,N)$ in (1), we get

$$\begin{aligned} B^k(\widetilde{S}, N\setminus F) &= 1 + \sum_{i\in(N\setminus F)\setminus\widetilde{S}} \frac{v_i}{1+V((N\setminus F)_{-i})}\, B^{k-1}(\widetilde{S}, (N\setminus F)_{-i}) \\ &\leq 1 + \sum_{i\in(N\setminus F)\setminus\widetilde{S}} \frac{v_i}{1+V((N\setminus F)_{-i})}\, B^{k-1}(\widehat{S}, (N\setminus K(F))_{-k(i)}) \\ &= 1 + \sum_{i\in(N\setminus F)\setminus\widetilde{S}} \frac{v_i}{1+V(N)-V(F)-v_i}\, B^{k-1}(\widehat{S}, (N\setminus K(F))_{-k(i)}), \end{aligned} \quad (13)$$

where the last equality uses the fact that $i \in (N\setminus F)\setminus\widetilde{S}$, which implies that $i \notin F$, so we have $V((N\setminus F)_{-i}) = V(N\setminus F_{+i}) = V(N) - V(F_{+i}) = V(N) - V(F) - v_i$.

Using the definition of $k(i)$ at the beginning of the proof, we have $v_{k(i)} \geq v_i$. Since $v_{k(i)} \geq v_i$, we also get $V(K(F)) = \sum_{i\in F} v_{k(i)} \geq \sum_{i\in F} v_i = V(F)$. Thus, we have

$$\frac{v_i}{1+V(N)-V(F)-v_i} \leq \frac{v_{k(i)}}{1+V(N)-V(K(F))-v_{k(i)}}. \quad (14)$$

Lastly, in Lemma C.2, which we shortly give in this section, we show the identity $(N\setminus K(F))\setminus\widehat{S} = \{k(i) : i \in (N\setminus F)\setminus\widetilde{S}\}$. Thus, for any function $g : \Re \to \Re$, it follows that $\sum_{i\in(N\setminus F)\setminus\widetilde{S}} g(v_{k(i)}) =$

$\sum_{\ell \in (N\setminus K(F))\setminus \widehat{S}} g(v_\ell)$ by change of variables in the sum. In this case, noting the inequality in (14), we can continue the chain of inequalities in (13) as

$$
\begin{aligned}
1+\sum_{i\in(N\setminus F)\setminus\widetilde{S}} \frac{v_i}{1+V(N)-V(F)-v_i}\, B^{k-1}(\widehat{S},(N\setminus K(F))_{-k(i)}) \\
&\le\; 1+\sum_{i\in(N\setminus F)\setminus\widetilde{S}} \frac{v_{k(i)}}{1+V(N)-V(K(F))-v_{k(i)}}\, B^{k-1}(\widehat{S},(N\setminus K(F))_{-k(i)}) \\
&\overset{(a)}{=}\; 1+\sum_{i\in(N\setminus F)\setminus\widetilde{S}} \frac{v_{k(i)}}{1+V((N\setminus K(F))_{-k(i)})}\, B^{k-1}(\widehat{S},(N\setminus K(F))_{-k(i)}) \\
&\overset{(b)}{=}\; 1+\sum_{\ell\in(N\setminus K(F))\setminus\widehat{S}} \frac{v_\ell}{1+V((N\setminus K(F))_{-\ell})}\, B^{k-1}(\widehat{S},(N\setminus K(F))_{-\ell}) \\
&\overset{(c)}{=}\; B^{k}(\widehat{S},N\setminus K(F)),
\end{aligned}
$$

where (a) holds because $i\notin F$, so noting that $k(i)\neq k(\ell)$ for $i\neq \ell$, we get $k(i)\notin K(F)$, (b) follows since $\sum_{i\in(N\setminus F)\setminus\widetilde{S}} g(v_{k(i)})=\sum_{\ell\in(N\setminus K(F))\setminus\widehat{S}} g(v_\ell)$ and (c) follows from (1).

The chain of inequalities above, along with (13), completes the induction argument. In this case, using the inequality $B^k(\widetilde{S},N\setminus F)\le B^k(\widehat{S},N\setminus K(F))$ with $F=\varnothing$ yields the desired result. ■

We use the next lemma in the proof of Lemma C.1, where we show that the sets $(N\setminus K(F))\setminus\widehat{S}$ and $\{k(i): i\in(N\setminus F)\setminus\widetilde{S}\}$ are identical.

Lemma C.2 *Consider $\widetilde{S},\widehat{S}\subseteq N$ with $|\widetilde{S}|=|\widehat{S}|$, where for each $i\in N\setminus\widetilde{S}$, there exists $k(i)\in N\setminus\widehat{S}$ such that $v_i\le v_{k(i)}$ and $k(i)\neq k(\ell)$ for each $i\neq\ell$. Then, for any $F\subseteq N$ with $F\cap\widetilde{S}=\varnothing$, letting $K(F)=\{k(i): i\in F\}$, we have $(N\setminus\widehat{S})\setminus K(F) \;=\; \{k(i): i\in(N\setminus\widetilde{S})\setminus F\}$.*

Proof: For each $i\in N\setminus\widetilde{S}$, we have $k(i)\in N\setminus\widehat{S}$, which implies that $\{k(i): i\in N\setminus\widetilde{S}\}\subseteq N\setminus\widehat{S}$. On the other hand, using the fact that $k(i)\neq k(\ell)$ for $i\neq\ell$, we get $|\{k(i): i\in N\setminus\widetilde{S}\}|=|N\setminus\widetilde{S}|=|N\setminus\widehat{S}|$, where the last equality holds since $|\widetilde{S}|=|\widehat{S}|$. In this case, having $\{k(i): i\in N\setminus\widetilde{S}\}\subseteq N\setminus\widehat{S}$ and $|\{k(i): i\in N\setminus\widetilde{S}\}|=|N\setminus\widehat{S}|$ implies that $N\setminus\widehat{S}=\{k(i): i\in N\setminus\widetilde{S}\}$. Moreover, we have $F\subseteq N\setminus\widetilde{S}$ and $K(F)=\{k(i): i\in F\}$. In this case, having $N\setminus\widehat{S}=\{k(i): i\in N\setminus\widetilde{S}\}$ and $K(F)=\{k(i): i\in F\}$ with $k(i)\neq k(\ell)$ for $i\neq\ell$ implies that $(N\setminus\widehat{S})\setminus K(F) \;=\; \{k(i): i\in(N\setminus\widetilde{S})\setminus F\}$. ■

Appendix D: Computational Experiments on Sushi Preferences

We give computational experiments that follow an outline similar to the one in Section 7, but we populate the preference lists by using a dataset from Kamishima (2018).

Experimental Setup: We generate purchase histories of customers making choices according to a ground choice model that does not comply with the multinomial logit model. Our goal is to check the ability of our choice model to predict the purchases of the customers and to pick profitable assortments. The ground choice model that we use is the non-parametric choice model. Recall that we have C customer types in the non-parametric choice model. The probability that a customer of type ℓ arrives into the system is β^ℓ. A customer of type ℓ is characterized by the preference list $(j^\ell(1),\dots,j^\ell(n^\ell))$, where n^ℓ is the number of products in the preference list and $j^\ell(k)$ is the product at position k. To populate the preference lists in the non-parametric choice model, we use a dataset from Kamishima (2018), which includes the rankings of 10 sushi varieties declared by 5000 diners. We use $(i^\ell(1),\dots,i^\ell(10))$ to denote the ranked list declared by diner ℓ in the dataset, where $i^\ell(k)$ is the sushi variety with rank order k for diner ℓ. In our non-parametric choice model, we have one customer type for each diner, so the number of customer types is $C=5000$. To come up with the preference list for customer type ℓ in the non-parametric choice model, we randomly truncate the ranked list of diner ℓ in the dataset. In particular, sampling n^ℓ from the uniform distribution over $\{1,\dots,10\}$, the preference list for customer type ℓ is $(i^\ell(1),\dots,i^\ell(n^\ell))$. In the non-parametric choice model, customers of each type ℓ arrives with an equal probability of $\beta^\ell=1/5000$.

Once we generate an instance of the ground choice model as above, we use the same approach in Section 7.1 to sample the purchase history of τ customers whose choices are governed by the ground choice model. We use this past purchase history as training data. To use as validation and testing data, we also generate two other purchase histories, each including 1250 customers. We fit a multinomial logit model with endogenous consideration sets and a standard multinomial logit model to the training data. We refer to the two fitted choice models as CS and ST.

Predicting Customer Purchases: We use the same approach in Section 7.2 to compare the ability of CS and ST to predict the purchases of the customers. We generate an instance of the ground choice model as discussed earlier in this section. Using the ground choice model, we generate three training datasets by varying the number of customers in the purchase history over $\tau\in\{1000,1750,2500\}$. To each of the three training datasets, we fit CS and ST. In Table EC.1, we compare the out-of-sample log-likelihoods of CS and ST. We replicated our computational experiments for 10 ground choice models that we randomly generate. The ground choice models differ in the samples of $\{n^\ell:\ell=1,\dots,5000\}$ that we use to truncate the ranked lists of the diners when populating the preference lists. Each row in the table corresponds to a different ground choice model and we list the index of the ground choice model in the first column. The layout of the rest of the table is identical to that of Table 2. Our results indicate that CS improves the out-of-sample log-likelihoods of ST on a majority of the ground choice model-training dataset pairs. In Table EC.2, we compare mean absolute errors of the choice probabilities under CS and ST. The layout of this table is identical to that of Table 3. Other than two ground choice model-training

Grnd. Ch.	$\tau = 1000$ CS	ST	Perc. Gap	$\tau = 1750$ CS	ST	Perc. Gap	$\tau = 2500$ CS	ST	Perc. Gap
1	-1834.10	-1840.64	0.36	-1837.59	-1843.25	0.31	-1835.07	-1840.69	0.31
2	-1927.40	-1937.53	0.53	-1926.58	-1935.66	0.47	-1924.57	-1932.61	0.42
3	-1908.68	-1909.80	0.06	-1903.79	-1904.98	0.06	-1904.37	-1905.37	0.05
4	-1848.16	-1859.12	0.59	-1848.05	-1856.73	0.47	-1843.63	-1853.59	0.54
5	-1892.07	-1894.01	0.10	-1891.68	-1893.66	0.10	-1890.49	-1892.97	0.13
6	-1913.06	-1916.44	0.18	-1906.01	-1907.84	0.10	-1902.56	-1905.67	0.16
7	-1919.35	-1927.84	0.44	-1916.36	-1923.56	0.38	-1916.55	-1924.01	0.39
8	-1891.27	-1889.84	-0.08	-1886.30	-1889.57	0.17	-1882.90	-1885.79	0.15
9	-1889.57	-1894.29	0.25	-1889.04	-1893.78	0.25	-1883.89	-1888.98	0.27
10	-1922.98	-1923.24	0.01	-1924.98	-1924.54	-0.02	-1924.58	-1924.15	-0.02
Avg.			0.24			0.23			0.24

Table EC.1 Comparison of out-of-sample log-likelihoods of CS and ST.

Grnd. Ch.	$\tau = 1000$ CS	ST	Perc. Gap	$\tau = 1750$ CS	ST	Perc. Gap	$\tau = 2500$ CS	ST	Perc. Gap
1	0.0239	0.0241	0.57	0.0216	0.0217	0.72	0.0220	0.0227	3.31
2	0.0217	0.0236	8.52	0.0219	0.0233	6.41	0.0217	0.0228	4.97
3	0.0221	0.0234	5.70	0.0186	0.0196	5.25	0.0180	0.0191	6.25
4	0.0239	0.0236	-1.25	0.0231	0.0233	1.20	0.0230	0.0236	2.74
5	0.0213	0.0214	0.48	0.0199	0.0202	1.86	0.0198	0.0206	4.13
6	0.0266	0.0280	5.55	0.0229	0.0236	2.73	0.0193	0.0201	4.33
7	0.0195	0.0209	7.35	0.0198	0.0204	2.82	0.0191	0.0199	3.91
8	0.0252	0.0252	-0.11	0.0212	0.0221	4.05	0.0198	0.0209	5.81
9	0.0276	0.0287	3.88	0.0249	0.0257	3.27	0.0234	0.0238	1.82
10	0.0196	0.0207	5.60	0.0197	0.0204	3.63	0.0184	0.0195	5.70
Avg.			3.63			3.19			4.30

Table EC.2 Comparison of mean absolute errors in choice probabilities of CS and ST.

dataset combinations, the mean absolute errors for CS are better than those for ST. When $\tau = 2500$ so that we have the largest amount of data to fit the two choice models, CS consistently improves the out-of-sample log-likelihoods and mean absolute errors for ST.

Picking Profitable Assortments: We use the same approach in Section 7.3 to compare the ability of CS and ST to pick profitable assortments. In Table EC.3, we compare the expected revenues from the assortments obtained by using CS and ST. The layout of this table is identical to that of Table 4. Our results indicate that CS performs slight but noticeable improvements over ST in terms of picking profitable assortments. As in Section 7, our results in Tables EC.1, EC.2 and EC.3 are based on one purchase history for each ground choice model. We replicated our results for 10 different purchase histories sampled from each of the 10 ground choice models. Considering the three levels of data availability that we work with, we end up with 300 ground choice model-purchase history-data availability combinations. Over all of the 300 combinations, the average percent gaps between the out-of-sample log-likelihoods, mean absolute errors and expected revenues corresponding to CS and ST are, respectively, 0.11%, 1.27% and 0.59%. Over the 300

	$\tau = 1000$				$\tau = 1750$				$\tau = 2500$			
Grnd. Ch.	Perc. Gap	Std. Err.	CS ≻ ST	ST ≻ CS	Perc. Gap	Std. Err.	CS ≻ ST	ST ≻ CS	Perc. Gap	Std. Err.	CS ≻ ST	ST ≻ CS
1	0.11	0.32	42	34	0.33	0.32	45	32	0.42	0.29	41	25
2	0.79	0.29	53	23	0.77	0.29	51	19	0.86	0.29	48	17
3	0.50	0.25	41	19	0.45	0.25	39	20	0.53	0.25	41	19
4	0.52	0.27	49	25	0.66	0.26	43	18	0.66	0.27	45	19
5	0.68	0.28	45	27	1.00	0.28	43	19	1.00	0.28	45	20
6	1.42	0.33	52	13	1.17	0.32	46	14	1.16	0.34	48	17
7	1.03	0.29	49	16	0.79	0.27	41	17	0.72	0.27	42	18
8	-0.55	0.34	37	47	0.40	0.27	41	27	0.60	0.30	47	26
9	-0.07	0.26	34	32	-0.03	0.26	34	31	-0.21	0.26	33	35
10	0.86	0.29	44	18	0.78	0.29	44	20	0.75	0.29	42	20
Avg.	0.53	0.29	45	25	0.63	0.28	43	22	0.65	0.28	43	22

Table EC.3 Comparison of expected revenues from the assortments obtained by using CS and ST.

combinations, in 213, 194 and 287 combinations, respectively, the out-of-sample log-likelihood, mean absolute errors and expected revenues of CS are better than those of ST.

Appendix E: Upper Bound on the Optimal Expected Revenue

In this section, we give an efficient approach to compute an upper bound on the optimal objective value of problem (3). Our approach is applicable to the case when $m = 2$, which is the case that we consider in our computational experiments. When $m = 2$, noting that $B^1(S,N) = 1$ and $B^2(S,N) = 1 + \sum_{i\in N\setminus S} \frac{v_i}{1+V(N_{-i})}$ by (1), letting $\theta_i = \frac{v_i}{1+V(N_{-i})}$ and dropping the constant $\frac{1}{1+V(N)}$ in (3) for notational brevity, the objective function of problem (3) is

$$W(S)\Big\{\lambda_1 + \lambda_2\Big(1 + \sum_{i\in N\setminus S}\theta_i\Big)\Big\} \;=\; W(S)\Big\{1 + \lambda_2 \sum_{i\in N\setminus S}\theta_i\Big\}, \tag{15}$$

where we use the fact that we have $m = 2$, so $\lambda_1 + \lambda_2 = 1$. Let $\overline{\Theta} = \sum_{i\in N}\theta_i$, which is the largest value that $\sum_{i\in N\setminus S}\theta_i$ can take.

We partition the interval $[0, \overline{\Theta}]$, into K subintervals $\{[\nu_{k-1}, \nu_k] : k = 1,\ldots,K\}$, where we have $0 = \nu_0 \le \nu_1 \le \ldots \le \nu_{K-1} \le \nu_K = \overline{\Theta}$. Any partition yields an upper bound on the optimal expected revenue, but finer partitions will yield tighter upper bounds. Intuitively speaking, for the objective function in (15) to take a large value, both $W(S)$ and $\sum_{i\in N\setminus S}\theta_i$ should take large values. For each $k = 1,\ldots,K$, using the decision variables $\boldsymbol{x} = (x_1,\ldots,x_n)$, we consider the problem

$$Z_k \;=\; \max_{\boldsymbol{x}\in[0,1]^n} \left\{ \sum_{i\in N} r_i\, v_i\, x_i \;:\; \sum_{i\in N}\theta_i\,(1 - x_i) \ge \nu_{k-1} \right\}. \tag{16}$$

If we impose the constraint $\boldsymbol{x} \in \{0,1\}^n$ in the problem above, then this problem finds an assortment S that maximizes $W(S)$ while ensuring that $\sum_{i\in N\setminus S}\theta_i \ge \nu_{k-1}$. Since we impose the constraint

$\boldsymbol{x} \in [0,1]^n$ in the problem above, this problem is a continuous knapsack problem and we can solve it efficiently. In the next theorem, we show that we can obtain an upper bound on the optimal objective value of problem (3) by solving the problem above for each $k = 1, \ldots, K$.

Theorem E.1 *Noting that Z_k is the optimal objective value of problem* (16), *letting* Rev^* *be the optimal objective value of problem* (3), *we have*

$$\frac{1}{1+V(N)} \max_{k=1,\ldots,K} \Big\{ (1+\lambda_2\, \nu_k)\, Z_k \Big\} \;\geq\; \mathsf{Rev}^*.$$

Proof: Using S^* to denote an optimal solution to problem (3), let $\widehat{\boldsymbol{x}} = (\widehat{x}_1, \ldots, \widehat{x}_n) \in \{0,1\}^n$ such that $\widehat{x}_i = 1$ if and only if $i \in S^*$. Furthermore, let $\ell \in \{1, \ldots, K\}$ be such that $\sum_{i\in N} \theta_i\,(1-\widehat{x}_i) \in [\nu_{\ell-1}, \nu_\ell]$. In this case, $\widehat{\boldsymbol{x}}$ is a feasible solution to problem (16), when we solve this problem with $k=\ell$. Thus, we obtain $Z_\ell \geq \sum_{i\in N} r_i\, v_i\, \widehat{x}_i = W(S^*)$, where the last equality uses the definition of $\widehat{\boldsymbol{x}}$. Also, by the definition of ℓ, we have $\nu_\ell \geq \sum_{i\in N} \theta_i\,(1-\widehat{x}_i) = \sum_{i\in N\setminus S^*} \theta_i$. In this case, we get

$$\begin{aligned}\frac{1}{1+V(N)} \max_{k=1,\ldots,K} \Big\{ (1+\lambda_2\, \nu_k)\, Z_k \Big\} \;&\geq\; \frac{1}{1+V(N)}\,(1+\lambda_2\,\nu_\ell)\, Z_\ell \\ &\geq\; \frac{1}{1+V(N)} \Big(1+\lambda_2 \sum_{i\in N\setminus S^*} \theta_i\Big) W(S^*) \;=\; \mathsf{Rev}^*,\end{aligned}$$

where the last equality above holds since (15) implies that the expression on the left side of the equality corresponds to the optimal objective value of problem (3). ■

In our computational experiments, we divide the interval $[0, \overline{\Theta}]$ into subintervals of width 0.0001 to obtain an upper bound on the optimal expected revenue.