Assortment Optimization under the Multinomial Logit Model with Utility-Based Rank Cutoffs

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

We study assortment optimization problems under a natural variant of the multinomial logit model where the customers are willing to focus only on a certain number of products that provide the largest utilities. In particular, each customer has a rank cutoff, characterizing the number of products that she will focus on during the course of her choice process. Given that we offer a certain assortment of products, the choice process of a customer with rank cutoff k proceeds as follows. The customer associates random utilities with all of the products as well as the no-purchase option. She ignores 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 and give a polynomial-time approximation scheme. Since the customers ignore the products below their rank cutoffs 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 rank cutoffs can yield better predictions of customer choices and yield more profitable assortment recommendations. *Dated August 31, 2023*.

---

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. One class of choice models that is often used in practice is based on random utility maximization, where a customer associates random utilities with each alternative and chooses the one that provides the largest utility. A common assumption in such choice models is that the customer can potentially choose any of the available alternatives. One way of enriching the choice models that are based on random utility maximization is to incorporate a probabilistic structure that describes a subset of alternatives that the customer is willing to focus on. The customer chooses the available alternative with the largest utility among the alternatives that she is willing to focus on. The probabilistic structure guiding the subset of alternatives that the customer is willing to focus on, along with the distribution of the utilities, specify the choice process of the customers.

We consider a natural variant of the multinomial logit model that uses utility rank cutoffs to describe the subset of alternatives that the customer is willing to focus on. In our variant, each customer has a rank cutoff, characterizing the number of products that she is willing to focus on. Given that we offer a certain assortment of products, a customer with rank cutoff 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 for the standard multinomial logit model. The customer ignores all alternatives whose utilities are not within the largest k. Among the remaining ones, she 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 a purchase. 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 varieties on the shelf. Thus, the utilities determine the ranking of the products, whereas the rank cutoff determines minimum acceptable utility. We overview our technical contributions, followed by a detailed discussion to motivate our choice model.

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

Choice Probabilities for the Multinomial Logit Model with Rank Cutoffs. Under the assumption that the rank cutoff of a customer follows a general discrete distribution, we give a recursion to compute the choice probability of each product within an assortment (Theorem 3.1). 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. 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 drastically different. Our variant provides a natural approach for incorporating choosy customers into the multinomial logit model. In a general method to incorporate choosy customers, each customer has a, possibly random, set of products C to focus on. If we offer the assortment S of products, then a customer ignores all the products that are not in the 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 all possible sets to focus on and estimating the parameters of this probabilistic structure may be difficult. In our approach, the rank cutoff is a scalar random variable and it allows us to succinctly characterize the set of products that a customer is willing to focus on.

Assortment Optimization and Impact of Rank Cutoffs. 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 4.1). 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 a rank cutoff of infinity, then they do not ignore any product during the course of their choice process, so our variant becomes equivalent to the standard multinomial logit model. Second, if all customers have a rank cutoff 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 choosier as their rank cutoffs decrease, so the customers under our choice model are choosier than those under the standard multinomial logit model but less choosy 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 is included by the one under the independent demand model (Theorem 4.2). Thus, as the customers get choosier, we offer a larger assortment to maximize the expected revenue. We can show this result even though computing the revenue-maximizing assortment under our choice model is NP-hard. The proof uses a property of independent Gumbel random variables stating that given the top p alternatives, the conditional choice process of the customer among the remaining q alternatives is identical to the unconditional one.

Approximation Scheme. We provide a polynomial-time approximation scheme (PTAS) for the assortment optimization problem. Letting n be the number of available products 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 $O((n/\epsilon)^{O(1/\epsilon^2)}\ \mathsf{RevOps})$ operations (Theorem 5.1). In the multinomial logit model, each product has a preference weight that is a function of its mean utility. 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 uses two results. First, we show that there exists, what we call, an ideal assortment that offers products in at most $O(\frac{\log(n/\epsilon)}{\epsilon})$ product classes and provides a $(1-\epsilon)^2$-approximate solution. 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, 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). Letting m be the maximum rank cutoff, the number of operations to compute the expected revenue from an assortment, RevOps, increases exponentially with m, but polynomially with n. Thus, our approach is a PTAS when the maximum rank cutoff is $O(1)$. The running time of $O((n/\epsilon)^{O(1/\epsilon^2)}\,\mathsf{RevOps})$ depends on m only through RevOps. Using θ to measure the relative gap between the largest and smallest preference weights in our choice model, we also show that if we use simulation with $O(\frac{\theta\, n^2}{\epsilon^4}\log(n/\epsilon))$ samples to estimate the expected revenue of an assortment, then our PTAS yields a $(1-\epsilon)$-approximate solution with $1-\epsilon$ probability, eliminating the dependence of the running time on m.

Prediction Ability with Rank Cutoffs. We test the ability of our multinomial logit model with rank cutoffs to predict customer choices and to identify profitable assortments. We use randomly generated synthetic datasets, along with a dataset based on the preferences of diners among sushi varieties. As benchmarks, we use the standard multinomial logit model and mixture of multinomial logit models, as well as the attention and consideration model and its mixture; see Gallego and Li (2017). Our choice model consistently performs better than the standard multinomial logit model, as well as the attention and consideration model. Our choice model is competitive to the mixture benchmarks, providing consistent improvements especially when the training dataset is small to moderate. We also numerically test the practical performance of our PTAS.

Motivating and Interpreting Our Choice Model from Multiple Angles: We can easily describe our choice model as a modification of the standard multinomial logit model using rank cutoffs, but there are lots of other ways to motive our choice model and connect it to the literature.

A Preference List-Based Choice Model. Our choice model can be cast as a preference list-based choice model. Under the preference list-based choice model, each customer arrives with a random preference list of the products and no-purchase option in mind, ranking all alternatives. The customer chooses the available alternative with the highest rank in her preference list; see Jagabathula et al. (2020) and Aouad et al. (2020). The distribution of the random preference lists of the customers specifies the choice probabilities. The class of preference list-based models is rather general in the sense that it encapsulates any choice model that is based on random utility maximization. After introducing our choice model, we show that our choice model is a preference list-based choice model, where the distribution of the preference lists are jointly determined through the distribution of the utilities and rank cutoffs. In other words, we can view our multinomial logit model with rank cutoffs as a parsimonious way of parameterizing a distribution over the preference lists. There are three benefits that arise from such parsimonious parameterization. First, if there are n products in consideration, then there are $O(n!)$ possible preference lists. Estimating a distribution over such a large number of preference lists for a general preference list-based choice model is difficult. Our choice model has $O(n)$ parameters. Second, we can express the mean utility of each product in our choice model by using a small number of numerical or categorical features, such as price, weight or color. If a new product is introduced, then we can compute the mean utility of the product as a function of its features without having to refit our choice model. Third, it is provably not possible to give a PTAS for assortment optimization under a general preference list-based choice model; see Aouad et al. (2018). We can give a PTAS under our choice model.

Compatibility with Random Utility Maximization. Under the random utility maximization principle, a customer associates random utilities with the products and no-purchase option, choosing the available alternative with the largest utility. In our choice model, a customer may leave without a purchase when there exists an available product that has a utility larger than that of the no-purchase option, but this product is not within the rank cutoff of the customer. Thus, the customer does not follow utility-maximizing behavior with respect to the utilities in our model. However, it is known that all preference list-based choice models are compatible with random utility maximization, so by the discussion in the previous paragraph, our choice model is compatible with random utility maximization. Because our choice model is compatible with random utility maximization, there exists some utility distribution for the alternatives such that the choice probabilities under our choice model match the choice probabilities when a customer picks the utility-maximizing alternative under this utility distribution without using any rank cutoffs.

Why Introduce Rank Cutoffs? In the standard multinomial logit model, a customer may purchase her, for example, tenth favorite product as long as the utility of this product is larger than the utility of the no-purchase option. The understanding is that if the tenth favorite product provides larger utility than the no-purchase option, then there is no reason for the customer not to purchase the product. We introduce a layer of flexibility by allowing customers to leave without a purchase when their few top choice products are not available. Our choice model generalizes the standard multinomial logit model, as we recover the standard multinomial logit model when the rank cutoff is always equal to the number of products. We can also interpret different rank cutoffs as different customer types. Due to the rank cutoffs in our choice model, there are two reasons for a customer with rank cutoff k to leave without a purchase. First, none of her k favorite products may be available. Second, the utility of no-purchase may exceed the utilities of all available products. To simplify the dynamics of our model, we can easily eliminate either reason. In particular, if the utility of the no-purchase option is Gumbel with location parameter of negative infinity, then no-purchase happens only because of the first reason. In this case, we obtain a multinomial logit model that effectively does not associate a utility with the no-purchase option, but the rank cutoff determines when a customer does not make a purchase. All of our results, including the computational complexity of the assortment problem and PTAS, apply to this case. The flexibility provided by our choice model pays off. In our computational experiments, our choice model can predict customer choices significantly better than the standard multinomial logit model.

Product Landscape. In our choice model, the customers associate utilities with all products, including those not offered. The implication is that the customers are aware of the products that are not offered. In some settings, the customers may know the products that are not offered. Search goods, for example, have transparent features that can be evaluated without seeing them; see Nelson (1970). Theater tickets and airfare have transparent features such as proximity to the stage and departure time. When shopping for such products, the customers have some understanding of the products that are not offered. Also, there are product categories, such as those in grocery retail, for which the universe of products is small and the customers frequently shop in these product categories. In such cases, the customers may know the products that are not offered.

The assumption that the customers are aware of the not offered products occurs when working with other choice models. In the Markov chain choice model, each customer carries out a random walk over the whole universe of products until she reaches an available alternative; see Blanchet et al. (2016). Baked into such a random walk is an implicit assumption that the customers are familiar with every product, offered or not. Thus, the Markov chain choice model also assumes that the customers are aware of the not offered products, yet it has been shown to fit choice data well even when this assumption is not fully valid. A choice model built with the assumption that the customers associate utilities with the not offered products can be robust in the sense that it can still predict customer choices well when this assumption does not hold.

Organization: In Section 2, we review the related literature. In Section 3, we describe our choice model, give a recursion to compute its choice probabilities and show that it is a preference list-based choice model. In Section 4, we formulate the assortment optimization problem, show that it is NP-hard and 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 our PTAS by establishing the existence of an ideal light assortment and constructing a collection of assortments that includes this ideal light assortment. In Section 7, we test the ability of our choice model to predict customer choices. In Section 8, we conclude.

2. Related Literature

There is growing work in assortment optimization and pricing models that explicitly incorporate the set of alternatives that the customer are wiling to focus on. Jagabathula and Rusmevichientong (2017) study pricing problems when customers focus only on products whose prices are below a certain threshold. Wang and Sahin (2018) study the assortment optimization problem when customers balance the expected utility from the purchase with the search cost when choosing the products to focus on. Jagabathula and Vulcano (2018) work with a choice model where the customers only focus on the last purchased product and products on promotion. Feldman and Topaloglu (2018) study the assortment optimization problem under the multinomial logit model with multiple customer types having nested consideration sets. Wang (2019) studies the assortment optimization problem for a variant of the multinomial logit model, where the customers ignore products when their mean utilities differ substantially from the largest mean utility of an available product. Aouad et al. (2019) study the assortment optimization problem under the multinomial logit model when a customer chooses to focus on each product with a fixed probability. Gallego and Li (2017) and Aouad et al. (2020) use similar probabilistic structures to characterize the set of products that the customers are willing to focus on.

Customers may limit the set of alternatives that they focus on for two reasons. First, customers may rule out some alternatives based on a small number of desirable features so that they lessen the burden of choice. By explicitly modeling the set of alternatives that the customers are willing to focus on, we can come up with a choice model that better mimics the behavioral dynamics of the customers. Second, putting a probabilistic structure to characterize the set of alternatives that the customers are willing to focus on may enrich expressive power of a choice model already on hand. Our approach is aligned with the second reason. By incorporating rank cutoffs into the multinomial logit model, we enrich the expressive power of the multinomial logit model. Similar to us, there is work on probabilistic consideration sets that postulate a probabilistic structure on the set of alternatives that the customers are willing to focus on. Gaundry and Dagenais (1979) introduce the dogit model as a simple modification on the multinomial logit model to enrich its modeling flexibility. Swait and Ben-Akiva (1987) interpret the dogit model as a probabilistic consideration set model, where the consideration set of a customer includes either a single alternative or all available alternatives. McCarthy (1997) demonstrates that the dogit model can improve the ability of the multinomial logit model to predict customer choices.

Another probabilistic consideration set model is the independent availability model, where a customer chooses to focus on each product independently with a fixed consideration probability. Swait (1984) estimates a different consideration probability for each product, so the number of consideration probabilities is as large as the number of products. Andrews and Srinivasan (1995) use a more parsimonious model by parameterizing the utility of each product through its features and postulating that a product is included in the consideration set if its utility exceeds a random threshold sampled from the normal distribution. Bronnenberg and Vanhonacker (1996) use a similar parameterization but sample the random thresholds from the type-I extreme value distribution. Thus, both the dogit and independent availability models, as well as our multinomial logit model with rank cutoffs, improve the richness the multinomial logit model by imposing a probabilistic structure on the set of products that the customers are willing to focus on, rather than trying to better mimic the behavior dynamics of the customers.

In contrast to probabilistic consideration sets, deterministic consideration sets try to explicitly model the behavioral dynamics behind the consideration set formation, but they can still utilize stochastic components. In Hauser and Wernerfelt (1990), a customer chooses to include an additional product into her consideration set by checking whether the increase in the expected maximum utility in her consideration set exceeds the incremental cognitive cost of considering the product. Roberts and Lattin (1991) put this framework into action by parameterizing the utility of each product as a function of its features. Mehta et al. (2003) give a similar consideration set model by balancing utility with consideration cost, but they can estimate their model even when they do not know the consideration sets of the customers. Andrews and Srinivasan (1995) give a comprehensive discussion of probabilistic and deterministic consideration set models. There is also behavior science work to explain the use of consideration sets. Miller (1956) discusses the limited information processing capacity in humans to motivate consideration sets. Stigler (1961) gives stylized models to show the diminishing benefit from doing extended searches, supporting small consideration sets observed in practice. Gensch (1987) use data to validate a choice process that first eliminates some alternatives based on their attributes, then makes a final pick. Alba and Hutchinson (1987) link the size of the consideration set of a customer to her expertise. Hauser (2014) discusses the rules of thumb that the customers use to form their consideration sets.

Under the standard multinomial logit model, 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) 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. For representative work on assortment optimization under other choice models, such as the preference list-based, nested logit, paired combinatorial logit and Markov chain choice model, we refer to Davis et al. (2014), Blanchet et al. (2016), Aouad et al. (2018), Zhang et al. (2020) and Berbeglia and Joret (2020).

3. Multinomial Logit Model with Rank Cutoffs

We formulate the multinomial logit model with rank cutoffs and give an expression for the choice probabilities of the products. The set of products is $N = \{1, \ldots, n\}$. The set of possible values for rank cutoffs 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 rank cutoff 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 rank cutoff 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. The customer only considers the top k alternatives with the largest utilities and chooses the available alternative that provides the largest utility. 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. We assume that $m \le n$, so the rank cutoff does not exceed the number of products. 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
$$\mathbb{P}\{X \ge Y\} = \frac{e^\mu}{e^\mu + e^\eta}.$$
- **(Independence from Top Ranked Choices)** Let $\{X_i : i \in G\}$ be a collection of independent Gumbel random variables, where X_i has 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
$$\mathbb{P}\Big\{X_{j_1} \ge \ldots \ge X_{j_q} \,\Big|\, X_{i_1} \ge \ldots \ge X_{i_p} \ge \max\{X_{j_1}, \ldots, X_{j_q}\}\Big\} \;=\; \mathbb{P}\Big\{X_{j_1} \ge \ldots \ge X_{j_q}\Big\}.$$

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 $X_{i_1} \ge \ldots \ge X_{i_p}$, 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 $\{j_1,\ldots,j_q\}$. The first property is standard; see Section 7.2.2.3 in Talluri and van Ryzin (2004). The second property is used in other settings in different forms; see, for example, (4) in Beggs et al. (1981) and (3.12) in Hanemann (1984). For completeness, we show the second property in Appendix A in the form we use. 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 3.1 (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 rank cutoff 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 rank cutoff to show the result for any set of products N, $S \subseteq N$ and $k \in M$. A customer with rank cutoff one purchases product $i \in S$ only when the utility of product i is 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 before the theorem and the fact that $V(S) = \sum_{i \in S} e^{\mu_i}$. Thus, since $B^1(\cdot,\cdot) = 1$, the result holds for $k = 1$. Assuming that the result holds for rank cutoff of $k-1$, we show that the result holds for rank cutoff of k. A customer with rank cutoff k purchases product $i \in S$ only under 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 before the theorem, conditional on the fact that product j occupies the top spot, the choice process among the remaining products N_{-j} and the no-purchase option is identical to the unconditional choice process, so the probability that the customer purchases product i is identical to the probability that a customer with rank cutoff $k-1$ 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, 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 before the theorem, 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 rank cutoff of k as well. ■

Given that we offer the assortment S, a customer with rank cutoff 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 rank cutoff 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 rank cutoff 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 rank cutoff 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 rank cutoff k with probability λ^k. Therefore, if we offer the assortment S, then a customer purchases product $i \in S$ with probability

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

The number of operations to compute $B^k(S,N)$ through (1) increases exponentially with k, so the number of operations to compute $\phi_i(S)$ increases exponentially with m.

Relationship to the Preference List-Based Choice Model:

Our multinomial logit model with rank cutoffs is a preference list-based choice model. In the preference list-based choice model, each customer arrives into the system with a random preference list of the products and no-purchase option, choosing the available alternative with the highest rank in her preference list. The distribution of the random preference lists specifies the choice probabilities. Letting Ω be the set of all possible permutations of $N \cup \{0\} = \{0,1,\ldots,n\}$, we use $\mathsf{rank}(i,\omega)$ to denote the rank of alternative i in permutation ω. For permutation $\omega = (4,3,0,2,5,1)$, for example, we have $\mathsf{rank}(4,\omega) = 1$, $\mathsf{rank}(0,\omega) = 3$ and $\mathsf{rank}(1,\omega) = 6$. In the preference list-based choice model, we use the random variable W taking values in Ω to capture the preference list of a customer. Letting $\mathbf{1}(\cdot)$ be the indicator function, under the preference list-based choice model, if we offer the assortment S, then a customer purchases product $i \in S$ with probability

$$\psi_i(S) = \sum_{\omega \in \Omega} \mathbb{P}\{W = \omega\} \, \mathbf{1}\Big(\mathsf{rank}(i,\omega) < \mathsf{rank}(j,\omega) \quad \forall j \in S_{-i} \cup \{0\}\Big), \tag{3}$$

where use the fact that a customer purchases product i when the product has the highest rank among all of the products in the offered assortment, as well as the no-purchase option.

We show that our multinomial logit model with rank cutoffs can be viewed as a preference list-based choice model, where the distribution of the preference lists is characterized by the Gumbel distributions driving the utilities of the products and no-purchase option, as well as the distribution of the rank cutoff. This discussion will also demonstrate how the utility of the no-purchase option and the rank cutoff jointly determine the distribution of the ranking of the unique no-purchase option in the preference list-based choice model. We use the random variable Z taking values in Ω to capture the ranking of the utilities of the products and no-purchase option, each sampled from the corresponding Gumbel distribution. For example, if the utilities of the alternatives satisfy $U_4 > U_3 > U_0 > U_2 > U_5 > U_1$, then $Z = (4,3,0,2,5,1)$. We continue using the random variable Y taking values in $M = \{1,\ldots,n\}$ to capture the rank cutoff of a customer. In this case, under our multinomial logit model with rank cutoffs, if we offer the assortment S, then a customer purchases product $i \in S$ with probability

$$\phi_i(S) = \sum_{k\in M}\sum_{w\in\Omega} \mathbb{P}\{Y=k,\ Z=\omega\}\, \mathbf{1}\Big(\mathsf{rank}(i,\omega) < \mathsf{rank}(j,\omega) \quad \forall\, j \in S_{-i}\cup\{0\}\Big)\, \mathbf{1}\Big(\mathsf{rank}(i,\omega)\le k\Big), \tag{4}$$

where we use the fact that a customer with rank cutoff k purchases product i when the ranking of the utility of product i is at most k and it is highest among all available alternatives.

To express the choice probability in (4) as a special case of the one in (3), we use $\mathsf{flip}(k,\omega)$ to denote the permutation obtained by bumping up the ranking of the no-purchase option in permutation ω to k if the ranking of the no-purchase option is below k. If the ranking of the no-purchase option in permutation ω is higher than k, then permutation $\mathsf{flip}(k,\omega)$ is identical to permutation ω. For example, if $\omega = (4,3,2,0,5,1)$, then we have $\mathsf{flip}(2,\omega) = (4,0,3,2,5,1)$ and $\mathsf{flip}(5,\omega) = (4,3,2,0,5,1)$. We make two observations. First, all products with rank higher than k in permutation ω maintain their ranking in permutation $\mathsf{flip}(k,\omega)$. In other words, if $\mathsf{rank}(i,\omega) < k$, then we have $\mathsf{rank}(i,\omega) = \mathsf{rank}(i,\mathsf{flip}(k,\omega))$. Second, the ranking of the no-purchase option in permutation $\mathsf{flip}(k,\omega)$ is the minimum of k and its ranking in permutation ω. Therefore, letting $a\wedge b = \min\{a,b\}$, we have $\mathsf{rank}(0,\mathsf{flip}(k,\omega)) = \mathsf{rank}(0,\omega)\wedge k$. In this case, we obtain the chain of equalities

$$\begin{aligned}
&\mathbf{1}\Big(\mathsf{rank}(i,\omega) < \mathsf{rank}(0,\omega)\Big)\ \mathbf{1}\Big(\mathsf{rank}(i,\omega) < k+1\Big)\\
&\qquad = \mathbf{1}\Big(\mathsf{rank}(i,\mathsf{flip}(k+1,\omega)) < \mathsf{rank}(0,\omega)\Big)\ \mathbf{1}\Big(\mathsf{rank}(i,\mathsf{flip}(k+1,\omega)) < k+1\Big)\\
&\qquad = \mathbf{1}\Big(\mathsf{rank}(i,\mathsf{flip}(k+1,\omega)) < \mathsf{rank}(0,\omega)\wedge(k+1)\Big)\\
&\qquad = \mathbf{1}\Big(\mathsf{rank}(i,\mathsf{flip}(k+1,\omega)) < \mathsf{rank}(0,\mathsf{flip}(k+1,\omega))\Big),
\end{aligned}\tag{5}$$

where the first equality holds because $\mathsf{rank}(i,\omega) < k+1$ implies $\mathsf{rank}(i,\omega) = \mathsf{rank}(i,\mathsf{flip}(k+1,\omega))$ by the first observation, whereas the last equality uses the second observation.

If product i was ranked higher than product j in permutation ω, then it is still ranked higher than product j in permutation $\mathsf{flip}(k,\omega)$, so for all $i,j\in N$, $\mathsf{rank}(i,\omega) < \mathsf{rank}(j,\omega)$ if and only if

$\mathsf{rank}(i,\mathsf{flip}(k,\omega)) < \mathsf{rank}(j,\mathsf{flip}(k,\omega))$. Thus, the expression $\mathbf{1}(\mathsf{rank}(i,\omega) < \mathsf{rank}(j,\omega)\ \forall j \in S_{-i} \cup \{0\}) \times \mathbf{1}(\mathsf{rank}(i,\omega) \le k)$ in (4) is equivalently given by

$$\begin{aligned}
&\mathbf{1}\Big(\mathsf{rank}(i,\omega) < \mathsf{rank}(j,\omega)\quad \forall j \in S_{-i} \cup \{0\}\Big)\ \mathbf{1}\Big(\mathsf{rank}(i,\omega) \le k\Big) \\
&\quad = \mathbf{1}\Big(\mathsf{rank}(i,\omega) < \mathsf{rank}(j,\omega)\quad \forall j \in S_{-i}\Big)\ \mathbf{1}\Big(\mathsf{rank}(i,\omega) < \mathsf{rank}(0,\omega)\Big)\ \mathbf{1}\Big(\mathsf{rank}(i,\omega) < k+1\Big) \\
&\qquad = \mathbf{1}\Big(\mathsf{rank}(i,\omega) < \mathsf{rank}(j,\omega)\quad \forall j \in S_{-i}\Big)\ \mathbf{1}\Big(\mathsf{rank}(i,\mathsf{flip}(k+1,\omega)) < \mathsf{rank}(0,\mathsf{flip}(k+1,\omega))\Big) \\
&\qquad\quad = \mathbf{1}\Big(\mathsf{rank}(i,\mathsf{flip}(k+1,\omega)) < \mathsf{rank}(j,\mathsf{flip}(k+1,\omega))\quad \forall j \in S_{-i} \cup \{0\}\Big), \qquad (6)
\end{aligned}$$

where the second equality uses (5), whereas the third equality holds because $\mathsf{rank}(i,\omega) < \mathsf{rank}(j,\omega)$ if and only if $\mathsf{rank}(i,\mathsf{flip}(k+1,\omega)) < \mathsf{rank}(j,\mathsf{flip}(k+1,\omega))$ for all $i,j \in N$.

Using (6) in (4), under our multinomial logit model with rank cutoffs, if we offer the assortment S, then a customer purchases product $i \in S$ with probability

$$\begin{aligned}
\phi_i(S) &= \sum_{k\in M}\sum_{w\in\Omega} \mathbb{P}\{Y=k,\ Z=\omega\}\ \mathbf{1}\Big(\mathsf{rank}(i,\mathsf{flip}(k+1,\omega)) < \mathsf{rank}(j,\mathsf{flip}(k+1,\omega))\quad \forall j \in S_{-i} \cup \{0\}\Big) \\
&= \sum_{k\in M}\sum_{w\in\Omega} \mathbb{P}\{Y=k,\ Z=\omega\} \sum_{\eta\in\Omega} \left\{ \mathbf{1}\Big(\eta = \mathsf{flip}(k+1,\omega)\Big)\ \mathbf{1}\Big(\mathsf{rank}(i,\eta) < \mathsf{rank}(j,\eta)\quad \forall j \in S_{-i} \cup \{0\}\Big) \right\} \\
&= \sum_{\eta\in\Omega} \left\{ \sum_{k\in M}\sum_{w\in\Omega} \mathbb{P}\{Y=k,\ Z=\omega\}\ \mathbf{1}\Big(\eta = \mathsf{flip}(k+1,\omega)\Big) \right\} \mathbf{1}\Big(\mathsf{rank}(i,\eta) < \mathsf{rank}(j,\eta)\quad \forall j \in S_{-i} \cup \{0\}\Big),
\end{aligned}$$

where the second equality holds by noting that we have the identity $f(\omega) = \sum_{\eta\in\Omega} \mathbf{1}(\eta=\omega)\, f(\eta)$ for any function $f:\Omega\to\mathbb{R}$, whereas the third equality follows by arranging the terms.

Letting $\mathbb{P}\{W=\eta\} = \sum_{k\in M}\sum_{w\in\Omega} \mathbb{P}\{Y=k,\ Z=\omega\}\,\mathbf{1}(\eta=\mathsf{flip}(k+1,\omega))$, the last expression above is $\sum_{\eta\in\Omega} \mathbb{P}\{W=\eta\}\,\mathbf{1}(\mathsf{rank}(i,\eta) < \mathsf{rank}(j,\eta)\quad \forall j \in S_{-i} \cup \{0\})$, which has the same form as the choice probability under the preference list-based choice model in (3). Thus, our multinomial logit model with rank cutoffs is a preference list-based choice model, where an arriving customer has the preference list η with probability $\sum_{k\in M}\sum_{w\in\Omega} \mathbb{P}\{Y=k,\ Z=\omega\}\,\mathbf{1}(\eta=\mathsf{flip}(k+1,\omega))$. Based on this result, we make three observations. First, our choice model is a parameterized version of the preference list-based choice model, where the distributions of the utilities and rank cutoff determine the distribution of the preference lists. Second, the utilities of the alternatives and rank cutoff jointly determine the ranking of the unique no-purchase option in the preference list-based choice model. In particular, if the rank cutoff of a customer is k, then we bump up the utility of the no-purchase option to just below the utility of the alternative at rank k. Third, it is not possible to give a PTAS for the assortment optimization problem under a general preference list-based model; see Theorem 1 in Aouad et al. (2018). In our choice model, the distribution of the preference lists is characterized by the distribution of the utilities and rank cutoff, which, surprisingly, gives way to a choice model that admits a PTAS for its corresponding assortment optimization problem.

Implications on Rationality and Customer Choice Process:

Recalling the Gumbel utility random variables $\{U_i : i \in N \cup \{0\}\}$ at the beginning of this section, consider the case where we offer all products except for the first one. Under our choice model, if a customer with rank cutoff of one evaluates the alternatives as $U_1 > U_2 > \ldots > U_n > U_0$, then this customer leaves without a purchase because her top choice is not available and the rank cutoff of one prevents her from entertaining any alternative other than her top choice. In other words, this customer leaves without a purchase even though there exist products in the offered assortment, such as products $2, 3, \ldots, n$, with utilities larger than the utility of the no-purchase option. Thus, our choice model is *not* compatible with utility-maximizing behavior with respect to the utility random variables $\{U_i : i \in N \cup \{0\}\}$. On the other hand, by the earlier discussion, there exists an equivalent preference list-based choice model that acts like our choice model in that it produces the same choice probabilities. The equivalent preference list-based choice model is compatible with random utility maximization, but the utility random variables that make the equivalent preference list-based choice model compatible with random utility maximization are likely not the Gumbel utilities $\{U_i : i \in N \cup \{0\}\}$. In that sense, a rational customer could behave in line with the choice probabilities of our choice model, but our choice model does not reflect the behavior of a merely utility-maximizing customer with respect to the utilities $\{U_i : i \in N \cup \{0\}\}$. As a result, our choice model should not be taken as a literal model of how customers act.

Our choice model can be viewed as a preference list-based choice model in which the preference list of a customer with rank cutoff of k has length at most k. We do not explicitly model the cognitive process behind the construction of the preference lists, but the effort to construct a list of top k choices can be significantly different from constructing full preference lists including all alternatives, lending support for preference lists with small lengths. In particular, if there are n alternatives under consideration, then forming a full preference list requires $O(n^2)$ comparisons between the alternatives in the worst case, whereas forming a preference list of top k alternatives requires $O(nk)$ comparisons. Ruling out products other than top k choices may be viewed through the lens of bounded rationality; see Simon (1955), Rubinstein (1998) and Kahneman (2003). Rational customers would be better off if they were to evaluate all utilities and purchase a product as long as its utility exceeds that of the no-purchase option. However, the effort to evaluate all products or the possible disappointment due to settling for an alternative worse than one of the top k choices may lead a customer to decide that evaluating all alternatives is not worth the effort. Customers may form short preference lists for other reasons as well. For example, those who dislike decaffeinated coffee may immediately eliminate such options without making any comparisons. Similarly, customers may eliminate airfare options by time of departure. In our choice model, the distribution of the rank cutoffs controls the length of the preference lists.

4. Assortment Optimization Problem

We give 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

$$\max_{S\subseteq N}\left\{\sum_{i\in S} r_i\,\phi_i(S)\right\} \;=\; \frac{1}{1+V(N)}\max_{S\subseteq N}\left\{W(S)\sum_{k\in M}\lambda^k\,B^k(S,N)\right\}. \tag{7}$$

The assortment optimization problem above turns out to be NP-hard. To show this result, we consider the feasibility version of our assortment optimization problem.

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

Theorem 4.1 (Computational Complexity) *The feasibility version of the assortment optimization problem in* (7) *is NP-complete.*

We give the proof of the theorem in Appendix B. The proof uses a reduction from the partition problem, which is a well-known NP-complete problem; see Garey and Johnson (1979).

To put Theorem 4.1 into perspective, consider two edge cases. First, if the rank cutoff 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 (7), 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 rank cutoff of all customers is one, then a customer focuses only on the top alternative with the largest utility and chooses this alternative if it is available. If all customers have a rank cutoff of one, then the objective function of problem (7) 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

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.

a polynomial-time solution in the two edge cases. By Theorem 4.1, if customers have rank cutoffs between the two edge cases, then problem (7) is NP-hard. To demonstrate how the revenues and preference weights of the products interact in problem (7), 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 rank cutoff 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 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 rank cutoff of 3, then the optimal assortment is $\{1\}$. Under rank cutoff of 3, if we offer the assortment $\{1\}$, then the purchase probability of product 1 is 0.75. As the rank cutoff 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.

In our choice model, a customer with rank cutoff k leaves without a purchase when none of the k alternatives with the largest utilities is available or when the utility of the no-purchase option exceeds those of all other available alternatives. A variant of our choice model occurs when the utility of the no-purchase option has location parameter of negative infinity. In this case, a customer with rank cutoff k leaves without a purchase only when none of the k alternatives with the largest utilities are available. Under this variant, a customer leaves without a purchase only when she cannot find one of her most preferred products as determined by her rank cutoff, which characterizes the minimal acceptable utility for the customer. The assortment optimization problem is still NP-hard under this variant, but our PTAS can be used to obtain an approximate solution.

Impact of Rank Cutoffs:

If all customers have rank cutoff of n, then our choice model becomes equivalent to the standard multinomial logit model. In our choice model, however, not all customers have rank cutoff of n. Intuitively speaking, smaller rank cutoffs translate into choosier customers, so the customers choosing according to our choice model are choosier than those choosing according to the standard multinomial logit model. Accordingly, we can 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. Thus, to maximize the expected revenue from choosier customers, we need to offer a larger assortment. Under the standard multinomial logit model, if we offer the assortment S, then a customer chooses product $i \in S$ with probability $\frac{v_i}{1+V(S)}$, in which case, we can find the optimal assortment by solving the problem

$$\max_{S \subseteq N} \left\{ \frac{\sum_{i \in S} r_i \, v_i}{1 + V(S)} \right\}. \tag{8}$$

In the next theorem, we show that the optimal assortment in (7) under our choice model includes the optimal optimal assortment in (8) under the standard multinomial logit model.

Theorem 4.2 (Impact of Rank Cutoffs) *There exist an optimal solution S^* to problem* (7) *and an optimal solution $\widetilde{S}$ to problem* (8) *such that $S^* \supseteq \widetilde{S}$.*

We give the proof of the theorem in Appendix C. We can also argue that the optimal expected revenue under our choice model, which captures choosier customers, cannot exceed the optimal expected revenue under the standard multinomial logit model. In particular, a simple lemma that we use when giving a proof for Theorem 4.2, given as Lemma C.1 in Appendix C, shows that $\frac{1}{1+V(N)} \, B^k(S,N) \leq \frac{1}{1+V(S)}$, in which case, by Theorem 3.1, we obtain $\pi_i^k(S) = \frac{v_i}{1+V(N)} \, B^k(S,N) \leq \frac{v_i}{1+V(S)}$ for any $S \subseteq N$. Therefore, using the fact that $\sum_{k \in M} \lambda^k = 1$, the last chain of inequalities yields $\phi_i(S) = \sum_{k \in M} \lambda^k \, \pi_i^k(S) \leq \frac{v_i}{1+V(S)}$. Using S^* to denote an optimal solution to our assortment optimization problem in (7), we obtain $\max_{S \subseteq N} \sum_{i \in S} r_i \, \phi_i(S) = \sum_{i \in S^*} r_i \, \phi_i(S^*) \leq \frac{\sum_{i \in S^*} r_i \, v_i}{1+V(S^*)} \leq \max_{S \subseteq N} \frac{\sum_{i \in S} r_i \, v_i}{1+V(S)}$. Thus, the optimal expected revenue under our choice model cannot exceed the optimal total expected revenue under the standard multinomial logit model.

We can ask several questions based on Theorem 4.2. As a first question, noting that $\widetilde{S}$ is the revenue-maximizing assortment under the standard multinomial logit model, one may ask whether this assortment provides a performance guarantee for the assortment optimization problem under our choice model. The answer to this question is negative. The revenue-maximizing assortment under the standard multinomial logit model can perform arbitrarily badly for our assortment optimization problem. Consider a problem instance with two products, where the revenues and preference weights are $(r_1, r_2) = (4, 1)$ and $(v_1, v_2) = (1, \overline{v})$. All customers have rank cutoff of one. The revenue-maximizing assortment under the standard multinomial logit model is $\widetilde{S} = \{1\}$. Noting that $\pi_i^1(S) = \frac{v_i}{1+V(N)}$ by Theorem 3.1, if we offer the assortment $\widetilde{S}$, then the expected revenue that we obtain under our choice model is $\frac{4}{2+\overline{v}}$, which gets arbitrarily close to zero when $\overline{v}$ is arbitrarily large. However, if we offer the assortment $\{1, 2\}$, then the expected revenue that we obtain under our choice model is $\frac{4+\overline{v}}{2+\overline{v}}$, which gets arbitrarily close to one when $\overline{v}$ is arbitrarily large. Thus, the assortment $\widetilde{S}$ can perform arbitrarily badly for our assortment optimization problem.

By the discussion right after Theorem 4.1, under the standard multinomial logit model, there exists a revenue-maximizing assortment that is revenue-ordered. By the discussion in the previous paragraph, this assortment can perform arbitrarily badly for our assortment optimization problem. As a second question, one may ask whether we get a performance guarantee for our assortment optimization problem if we check the expected revenue of all revenue-ordered assortments for our assortment optimization problem and pick the best one. We partially answer this question. In Appendix D, we show that if the rank cutoff of a customer does not exceed two, then the best revenue-ordered assortment is a $\frac{1}{2}$-approximation for the assortment optimization problem under our choice model. Furthermore, we show that this bound is tight in the sense that the best revenue-ordered assortment can never provide a performance guarantee better than $\frac{1}{2}$. Despite our best efforts, we are not able to settle whether the best revenue-ordered assortment still provides a performance guarantee when the rank cutoffs take values of three or more.

Let S_κ^* be the optimal solution under our choice model when all customers have rank cutoff of κ. That is, S_κ^* is an optimal solution to problem (7) when $\lambda^\kappa = 1$ and $\lambda^k = 0$ for all $k \in M_{-\kappa}$. If all customers have rank cutoff of n, then our choice model reduces to the standard standard multinomial logit model. Thus, Theorem 4.2 implies that $S_\kappa^* \supseteq S_n^*$ for all $\kappa = 1, \ldots, n$. As a third question, one may ask whether the optimal assortments are nested in rank cutoffs. In other words, one may ask whether we have $S_1^* \supseteq S_2^* \supseteq \ldots \supseteq S_{n-1}^* \supseteq S_n^*$. The answer to this question is negative. 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 $(v_1, v_2, v_3, v_4, v_5) = (7, 16, 1, 0.2, 12)$. For this problem instance, we have $S_2^* = \{1, 2\}$, $S_3^* = \{1, 3, 4\}$ and $S_5^* = \{1\}$. As predicted by Theorem 4.2, $S_2^* \supseteq S_5^*$ and $S_3^* \supseteq S_5^*$, but we do not have $S_2^* \supseteq S_3^*$. Furthermore, this problem instance shows that we may not even have $|S_1^*| \geq |S_2^*| \geq \ldots \geq |S_{n-1}^*| \geq |S_n^*|$, as $|S_2^*| = 2$ and $|S_3^*| = 3$.

5. Polynomial-Time Approximation Scheme

To give a PTAS for problem (7), 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. Then, 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. Also, 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 give our PTAS. As we develop our PTAS, we summarize the notation in Table 2.

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 $S \cap N_g \in \{S_g^{\mathsf{light}}(k) : k = 0, \ldots, |N_g|\}$ 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 (7), 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

$$\frac{1}{1+V(N)}\left\{W(\widehat{S})\sum_{k\in M}\lambda^k\, B^k(\widehat{S},N)\right\} \;\ge\; \frac{\frac{1-\epsilon}{1+\epsilon}}{1+V(N)}\left\{W(S^*)\sum_{k\in M}\lambda^k\, B^k(S^*,N)\right\}. \tag{9}$$

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 (7). 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 (7). 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

$$\overline{W}(\widetilde{S} \cap N_g) \;\leq\; \overline{W}(\widehat{S} \cap N_g) \;\leq\; \overline{W}(\widetilde{S} \cap N_g) + \frac{\epsilon}{|\widehat{G}|}\,\overline{W}(\widehat{S}) \quad \forall\, g \in \mathbb{Z}. \tag{10}$$

The number of operations to construct the collection of assortments $\{A_t : t \in \mathcal{A}\}$ will be $O((n/\epsilon)^{O(1/\epsilon^2)})$. 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 rank cutoff m. By (10) 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. Using (10), we will also show that the expected revenue from this candidate assortment deviates from that of the ideal light assortment by at most a factor of $1-\epsilon$. In our PTAS, we focus on light assortments because if products i and j have the same weight but product i is lighter than product j so that $w_i = w_j$ and $v_i \leq v_j$, then we have $W(S \cup \{i\}) = W(S \cup \{j\})$ but $B^k(S \cup \{i\}, N) \geq B^k(S \cup \{j\}, N)$ for $\{i,j\} \cap S = \varnothing$, so by (7), using product i rather than j yields a larger expected revenue. In Section 6.1, we show that Part 1 holds. In Section 6.2, we show that Part 2 holds. In the remainder of this section, we put the two parts together in the next theorem to give our PTAS. In this theorem, let RevOps be 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 (7) 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

$$(1-\epsilon)\,\overline{W}(\widehat{S}) \stackrel{(a)}{=} \sum_{g \in \widehat{G}} \left\{ \overline{W}(\widehat{S} \cap N_g) - \frac{\epsilon}{|\widehat{G}|}\overline{W}(\widehat{S}) \right\} \stackrel{(b)}{\leq} \sum_{g \in \widehat{G}} \overline{W}(\widetilde{S} \cap N_g) \stackrel{(c)}{=} \sum_{g \in \mathbb{Z}} \overline{W}(\widetilde{S} \cap N_g) = \overline{W}(\widetilde{S}), \tag{11}$$

where (a) holds since $\widehat{S} \cap N_g = \varnothing$ for all $g \notin \widehat{G}$, (b) is by (10) 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 (10). Since $\widehat{S}$ and $\widetilde{S}$ are light, for each $g \in \mathbb{Z}$, both

N_g	=	Set of products in product class g. That is, $N_g = \{i \in N : (1+\epsilon)^g \leq w_i < (1+\epsilon)^{g+1}\}$.
$\overline{w}_i$	=	Rounded weight of product i. That is, we have $\overline{w}_i = (1+\epsilon)^g$ for all $i \in N_g$.
$\overline{W}(S)$	=	Total rounded weight of products in the assortment S. That is, $\overline{W}(S) = \sum_{i \in S} \overline{w}_i$.
$S_g^{\mathsf{light}}(k)$	=	Assortment that includes the k products with the smallest preference weights in product class g.
S^*	=	Optimal solution to problem (7).
$\widehat{S}$	=	Ideal light assortment.
$\widehat{G}$	=	Set of product classes used by the ideal light assortment. That is, $\widehat{G} = \{g \in \mathbb{Z} : \widehat{S} \cap N_g \neq \varnothing\}$.
$\widetilde{S}$	=	Candidate light assortment that approximates the ideal light assortment as in (10).
k_g^*	=	Number of products offered by the optimal assortment S^* in product class g.
$g^*_{\max}$	=	Largest product class used by the optimal assortment S^*. That is, $g^*_{\max} = \max\{g \in \mathbb{Z} : S^* \cap N_g \neq \varnothing\}$.
$g^*_{\min}$	=	Smallest product class used in our construction of the ideal light assortment as in (12).

Table 2 Notation used in the development of our PTAS.

$\widehat{S} \cap N_g$ and $\widetilde{S} \cap N_g$ include a certain number of products with the smallest preference weights in product class g. Because $\widehat{S}$ and $\widetilde{S}$ satisfy (10), we have $\overline{W}(\widetilde{S} \cap N_g) \leq \overline{W}(\widehat{S} \cap N_g)$, so $\widetilde{S} \cap N_g \subseteq \widehat{S} \cap N_g$ for all $g \in \mathbb{Z}$. Thus, we get $\widetilde{S} = \cup_{g \in \mathbb{Z}} (\widetilde{S} \cap N_g) \subseteq \cup_{g \in \mathbb{Z}} (\widehat{S} \cap N_g) = \widehat{S}$. We use induction over the rank cutoff to show that $B^k(S,N) \geq B^k(Q,N)$ 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 rank cutoff of $k-1$, we show that the result holds for rank cutoff of k. In particular, by (1), we have

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

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 $B^{k-1}(Q, N_{-i}) \leq B^{k-1}(S, N_{-i})$, 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 rank cutoff of k. Since $\widetilde{S} \subseteq \widehat{S}$ as discussed at the beginning of this paragraph, we get $B^k(\widetilde{S}, N) \geq B^k(\widehat{S}, N)$ for all $k \in M$.

Thus, letting S^* be an optimal solution to problem (7), noting that $\overline{W}(S) \leq W(S) \leq (1+\epsilon)\,\overline{W}(S)$ for any $S \subseteq N$, by (9) and (11), the expected revenue from the assortment $\widetilde{S}$ satisfies

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

where (e) is by (11), (f) uses the fact that $B^k(\widetilde{S}, N) \geq B^k(\widehat{S}, N)$ for all $k \in M$ and (g) is by (9). Since $\frac{(1-\epsilon)^2}{(1+\epsilon)^2} \geq (1-\epsilon)^4 \geq 1 - 4\,\epsilon$, the assortment $\widetilde{S}$ is a $(1-4\,\epsilon)$-approximate solution to problem (7).

Since $\widetilde{S} \in \{A_t : t \in \mathcal{A}\}$, the collection of candidate assortments $\{A_t : t \in \mathcal{A}\}$ includes a $(1-4\,\epsilon)$-approximate solution to problem (7). By the small and light collection property in Part 2, there are

$O((n/\epsilon)^{O(1/\epsilon^2)})$ assortments in the collection and 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 in $O(\mathsf{RevOps}\ (n/\epsilon)^{O(1/\epsilon^2)})$ operations, we get a $(1-4\,\epsilon)$-approximate solution to problem (7). 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. ∎

We can use (1)-(2) to compute the choice probability $\phi_i(S)$ for each product $i \in S$, in which case, the expected revenue from assortment S is $\sum_{i\in S} r_i\,\phi_i(S)$. By the discussion right after (2), the number of operations to compute $\phi_i(S)$ increases exponentially with the maximum rank cutoff m, so the number of operations to compute the expected revenue from an assortment, which we denote by RevOps, increases exponentially with m as well. In Appendix E, however, letting $\theta = (\frac{1}{n} + \max_{i\in N} v_i)/\min_{i\in N} v_i$ for notational brevity, we show that we can use simulation in our PTAS to estimate the expected revenue from an assortment with $O(\frac{\theta\, n^2}{\epsilon^4}\log(n/\epsilon))$ samples of product utilities to obtain a $(1-6\,\epsilon)$-approximate solution with $1-\epsilon$ probability, eliminating the exponential dependence of the running time of our PTAS on m.

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 (7), 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

$$\widehat{S} = \bigcup_{g=g_{\min}^*}^{g_{\max}^*} \widehat{S}_g. \tag{12}$$

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 Part 1.

In the next proposition, we show that the assortment $\widehat{S}$ that we construct as in (12) satisfies the limited degradation property in Part 1 as well.

Proposition 6.1 (Limited Degradation) *Letting S^* be an optimal solution to problem* (7) *and $\widehat{S}$ be as in* (12)*, 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 products 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 $\widehat{S}_g$, where we have $|\widehat{S}_g| = k^*_g$. 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 $\widehat{S}_g = S^{\mathsf{light}}_g(k^*_g)$ 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 F.1 in Appendix F, 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 (12). ■

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 (12), our approach is based on guessing two quantities. First, we guess a value of $\widehat{\eta}$ such that $\widehat{\eta} \le \overline{W}(\widehat{S}) \le 2\,\widehat{\eta}$. 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

$$\widehat{\kappa}_g\,\epsilon\frac{\widehat{\eta}}{L} \;\le\; \overline{W}(\widehat{S} \cap N_g) \;<\; (\widehat{\kappa}_g + 1)\,\epsilon\frac{\widehat{\eta}}{L} \quad \forall\, g \in \mathbb{Z}. \tag{13}$$

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 $\widehat{\eta} \le \overline{W}(\widehat{S}) \le 2\,\widehat{\eta}$ and (13). 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 $\widetilde{k}_g(\eta, \kappa_g) = \min\{k = 0, 1, \ldots, |N_g| : \overline{W}(S_g^{\mathsf{light}}(k)) \ge \kappa_g\,\epsilon\frac{\eta}{L}\}$, 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

$$\widetilde{S}(\eta, \kappa) \;=\; \bigcup_{g \in \mathbb{Z}} \widetilde{S}_g(\eta, \kappa_g), \tag{14}$$

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

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 in Part 2 as well.

Proposition 6.2 (Approximation to Ideal) *Letting $\widehat{S}$ be as in (12) 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*

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

Proof: Let $(\widehat{\eta}, \widehat{\kappa}) \in \Theta$ be the correct guess in the sense that it satisfies $\widehat{\eta} \le \overline{W}(\widehat{S}) \le 2\,\widehat{\eta}$ and $\widehat{\kappa}_g\,\epsilon\frac{\widehat{\eta}}{L} \le \overline{W}(\widehat{S} \cap N_g) < (\widehat{\kappa}_g + 1)\,\epsilon\frac{\widehat{\eta}}{L}$ 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 (14), 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 *smallest* 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 (12), $\widehat{S}\cap N_g$ 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, noting that $\overline{W}(\widehat{S}\cap N_g)<(\widehat{\kappa}_g+1)\,\epsilon\frac{\widehat{\eta}}{L}$ and $\overline{W}(\widehat{S})\geq\widehat{\eta}$ by our choice of $(\widehat{\kappa},\widehat{\eta})$ at the beginning of the proof, we get

$$\begin{aligned}\overline{W}(\widetilde{S}_g(\widehat{\eta},\widehat{\kappa}_g)) \;\leq\; \overline{W}(\widehat{S}\cap N_g) \;<\; (\widehat{\kappa}_g+1)\,\epsilon\frac{\widehat{\eta}}{L}&\\ \leq\; \overline{W}(\widetilde{S}_g(\widehat{\eta},\widehat{\kappa}_g))+\epsilon\frac{\widehat{\eta}}{L} \;\leq\; \overline{W}(\widetilde{S}_g(\widehat{\eta},\widehat{\kappa}_g))+\frac{\epsilon}{|\widehat{G}|}\,\overline{W}(\widehat{S}).&\end{aligned}$$

Letting $\widetilde{S}=\widetilde{S}(\widehat{\eta},\widehat{\kappa})$, by (14), we have $\widetilde{S}\cap N_g=\widetilde{S}_g(\widehat{\eta},\widehat{\kappa}_g)$. Thus, 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 satisfies $\widehat{\eta}\leq\overline{W}(\widehat{S})\leq 2\,\widehat{\eta}$, as well as (13).

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 the construction in (12), 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 (13) for all $g\neq\widehat{g}-L+1,\ldots,\widehat{g}$. Furthermore, 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 we need to have $\sum_{g\in\mathbb{Z}}\widehat{\kappa}_g\leq\lceil\frac{2L}{\epsilon}\rceil$. In this case, using the fact that we can set $\widehat{\kappa}_g=0$ for all $g\neq\widehat{g}-L+1,\ldots,\widehat{g}$, it follows that we need to have $\sum_{g=\widehat{g}-L+1}^{\widehat{g}}\widehat{\kappa}_g\leq\lceil\frac{2L}{\epsilon}\rceil$. Therefore, for each guess of $\widehat{g}$, the number of guesses for $\widehat{\kappa}$ is upper bounded by the number of ways to divide $\lceil\frac{2L}{\epsilon}\rceil$ items into L bins. Recall that the number of ways to divide $\lceil\frac{2L}{\epsilon}\rceil$ items into L bins is given by $\binom{\lceil \frac{2L}{\epsilon} \rceil + L - 1}{L}$. Therefore, for each guess of $\widehat{g}$, we conclude that we can upper bound the number of guesses for $\widehat{\kappa}$ by using the chain of inequalities

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

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 the rank cutoffs 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 idiosyncraticity. 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 product with probability 0.1, assuming that the customer ignores the product. After dropping, with probability 0.5, we randomly pick just one product in the list and flip its place with its successor to get the preference list of customers of type ℓ. With probability 0.5, we leave the 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 vary τ to obtain different levels of data availability. We use these past purchase histories as the training dataset. To use as the validation and testing datasets, we follow the same approach to generate two other purchase histories, each including 1250 customers. The offered assortments, arriving customers and their choices in the training, validation and testing datasets are all independent draws. In particular, we do not bootstrap from one of the datasets to construct another one.

We use maximum likelihood to fit our multinomial logit model with rank cutoffs to the training dataset; see, for example, Vulcano et al. (2012). The parameters of our choice model are the preference weights $(v_1, \ldots, v_n)$ and the rank cutoff distribution $(\lambda^1, \ldots, \lambda^m)$. We parameterize v_i as $v_i = e^{\beta_i}$ for $\beta_i \in \mathbb{R}$. We use cross validation to choose the value of the maximum rank cutoff m. The number of products is relatively small, so we compute the choice probabilities using (1). We use the fmincon routine in Matlab without derivative information with multiple initial solutions to obtain a local maximum of the log-likelihood function. We use RCO to refer to the fitted multinomial logit model with rank cutoffs. As benchmark choice models, we fit a standard multinomial logit model, a mixture of multinomial logit models, an attention and consideration model and a mixture of attention and consideration models to the training dataset.

Multinomial logit variants are widely used in the literature; see Section 7.2 in Talluri and van Ryzin (2004). The parameters of the standard multinomial logit model are $(v_1, \ldots, v_n)$, where v_i is the preference weight of product i. If we offer the assortment S of products, then a customer chooses product $i \in S$ with probability $v_i/(1 + \sum_{j \in S} v_j)$. In a mixture of multinomial logit models, we have K customer types. An arriving customer is of type k with probability θ_k. A customer of type k chooses according to the multinomial logit model with preference weights $(v_{1k}, \ldots, v_{nk})$. The parameters of the mixture are $(\theta_1, \ldots, \theta_K)$ and $\{(v_{1k}, \ldots, v_{nk}) : k = 1, \ldots, K\}$. We use maximum likelihood to fit the two variants of the multinomial logit model, parameterizing v_i as $v_i = e^{\beta_i}$ for $\beta_i \in \mathbb{R}$. We choose the value of K for the mixture by using cross validation. We use SML and MML, respectively, to refer to the standard and mixture of multinomial logit models that we fit.

In the attention and consideration model, the customers have a fixed preference ranking for the products. We index the products so that the preference ranking has the order $1, \ldots, n$. A customer includes each product i in her consideration set independently with probability γ_i. Once she forms her consideration set, the customer purchases the available product in her consideration set that has the highest ranking. If none of the products in her consideration set is available, then the customer leaves without a purchase. The parameters of the attention and consideration model are the consideration probabilities $(\gamma_1, \ldots, \gamma_n)$, along with the preference ranking of the products. This choice model is studied in Gallego and Li (2017). Jagabathula et al. (2020) also work with this choice model, referring to it as the independent consideration set model.

In a mixture of attention and consideration models, we have K customer types. The preference ranking for all customer types is the same. An arriving customer is of type k with probability θ_k. A customer of type k chooses among the offered products according to the attention and consideration model with consideration probabilities $(\gamma_{1k}, \ldots, \gamma_{nk})$. The mixture of attention and consideration models is a variant of the general consider-then-choose model studied by Jagabathula et al. (2020), where the authors also assume that there is a single preference ranking for all customers, but customers of different types pick the products to consider according to a different probabilistic structure. Jagabathula et al. (2020) give an integer program to fit variants of the attention and consideration model by using maximum likelihood estimation. We refer to the standard and mixture of attention and consideration models, respectively, as SAC and MAC. Considering our multinomial logit model with rank cutoffs and the four benchmarks, we compare five choice models.

7.2 Comparing Out-of-Sample Log-Likelihoods

To compare the out-of-sample log-likelihoods of the fitted choice models, we generate an instance of the ground choice model as discussed at the beginning of Section 7.1. 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 when fitting the five choice models. To each of the three training datasets, we fit RCO, SML, MML, SAC and MAC by using maximum likelihood estimation. Using the ground choice model, we also generate validation and testing datasets, each including 1250 customers. We use the validation dataset to choose the value of m for RCO and the value of K for MML and MAC. We compute the out-of-sample log-likelihoods of the testing dataset under the five fitted choice models. A larger out-of-sample log-likelihood indicates that the choice model in question will do a better job of predicting the purchases of the customers not in the training dataset.

In Table 3, we compare the out-of-sample log-likelihoods of the five fitted choice models. The top, middle and bottom portions of the table correspond to different values for the number of customers τ in the training dataset, yielding different levels of data availability to fit the choice models. We replicated our computational experiments for 10 ground choice models that we randomly generated. Each row in the table corresponds to a different ground choice model that we work with. In the

$\tau = 1000$

Grnd.	Out-of-Sample Log-Likelihoods					Perc. Gap with RCO			
Ch.	RCO	SML	MML	SAC	MAC	SML	MML	SAC	MAC
1	-2024.52	-2041.73	-2049.29	-2041.93	-2041.93	0.85	1.22	0.86	0.86
2	-2043.58	-2062.84	-2054.18	-2058.98	-2055.02	0.94	0.52	0.75	0.56
3	-2073.68	-2095.13	-2084.94	-2081.92	-2081.87	1.03	0.54	0.40	0.39
4	-2021.50	-2033.60	-2037.16	-2034.39	-2034.39	0.60	0.77	0.64	0.64
5	-2072.32	-2097.19	-2097.49	-2087.79	-2087.79	1.20	1.21	0.75	0.75
6	-2038.07	-2044.53	-2037.59	-2047.09	-2045.00	0.32	-0.02	0.44	0.34
7	-2117.68	-2134.45	-2136.35	-2126.15	-2131.86	0.79	0.88	0.40	0.67
8	-2059.85	-2067.13	-2066.66	-2060.40	-2069.75	0.35	0.33	0.03	0.48
9	-2056.47	-2072.01	-2073.23	-2072.59	-2070.40	0.76	0.82	0.78	0.68
10	-2111.71	-2132.01	-2121.54	-2127.90	-2135.61	0.96	0.47	0.77	1.13
Avg.						0.78	0.67	0.58	0.65

$\tau = 1750$

Grnd.	Out-of-Sample Log-Likelihoods					Perc. Gap with RCO			
Ch.	RCO	SML	MML	SAC	MAC	SML	MML	SAC	MAC
1	-2018.65	-2035.55	-2021.40	-2026.34	-2021.02	0.84	0.14	0.38	0.12
2	-2031.15	-2051.42	-2033.46	-2047.10	-2038.65	1.00	0.11	0.79	0.37
3	-2069.24	-2090.22	-2081.48	-2086.80	-2077.63	1.01	0.59	0.85	0.41
4	-2016.42	-2027.42	-2031.36	-2034.67	-2027.62	0.55	0.74	0.91	0.56
5	-2066.12	-2090.87	-2076.15	-2084.51	-2082.76	1.20	0.49	0.89	0.81
6	-2036.11	-2042.28	-2024.59	-2039.57	-2033.36	0.30	-0.57	0.17	-0.13
7	-2106.98	-2125.46	-2109.03	-2120.47	-2126.67	0.88	0.10	0.64	0.93
8	-2050.78	-2061.83	-2058.04	-2051.70	-2054.44	0.54	0.35	0.04	0.18
9	-2053.21	-2067.84	-2066.09	-2072.09	-2074.07	0.71	0.63	0.92	1.02
10	-2102.75	-2122.98	-2096.79	-2119.15	-2107.93	0.96	-0.28	0.78	0.25
Avg.						0.80	0.23	0.64	0.45

$\tau = 2500$

Grnd.	Out-of-Sample Log-Likelihoods					Perc. Gap with RCO			
Ch.	RCO	SML	MML	SAC	MAC	SML	MML	SAC	MAC
1	-2018.54	-2035.28	-2021.43	-2026.31	-2017.70	0.83	0.14	0.38	-0.04
2	-2031.27	-2051.87	-2032.35	-2043.61	-2034.63	1.01	0.05	0.61	0.17
3	-2065.05	-2086.22	-2063.80	-2081.81	-2065.91	1.03	-0.06	0.81	0.04
4	-2016.25	-2027.05	-2025.37	-2038.19	-2035.57	0.54	0.45	1.09	0.96
5	-2067.92	-2092.77	-2076.76	-2092.16	-2082.10	1.20	0.43	1.17	0.69
6	-2035.18	-2041.81	-2022.07	-2039.39	-2031.77	0.33	-0.64	0.21	-0.17
7	-2104.60	-2122.61	-2109.68	-2117.43	-2111.04	0.86	0.24	0.61	0.31
8	-2049.10	-2061.81	-2054.24	-2051.95	-2046.28	0.62	0.25	0.14	-0.14
9	-2050.94	-2064.82	-2058.89	-2069.26	-2069.67	0.68	0.39	0.89	0.91
10	-2099.23	-2117.59	-2097.13	-2118.68	-2100.61	0.87	-0.10	0.93	0.07
Avg.						0.80	0.11	0.68	0.28

Table 3 Comparison of the out-of-sample log-likelihoods of the fitted choice models.

first column of the table, we give the index of the ground choice model. The next five columns show the out-of-sample log-likelihoods obtained by the five fitted choice models, whereas the last four columns show the percent gap between the out-of-sample log-likelihoods of RCO and each of the remaining four fitted choice models. In all of the tables that we give throughout this section, we follow the convention that positive values for the percent gaps favor RCO.

The results in Table 3 indicate that RCO provides uniform improvements over SML and SAC in terms of out-of-sample log-likelihoods. Intuitively speaking, we view RCO, SML and SAC as parsimonious choice models, as they do not involve multiple customer types. It is encouraging that

RCO consistently provides the best out-of-sample log-likelihoods among the parsimonious choice models. The performance gaps between RCO and the other two choice models slightly increase as we have more data availability in the training dataset. Comparing RCO with the two choice models that involve multiple customer types, the out-of-sample log-likelihoods of RCO are larger than those of MML and MAC for a large majority of the ground choice models. It is known that MML can approximate any choice model based on random utility maximization arbitrarily well, as long as the number of customer types in the mixture is large; see McFadden and Train (2000). However, if one uses a small training dataset to fit MML with a large number of customer types, then the large number of parameters in MML may result in overfitting to the training dataset and MML may provide poor out-of-sample log-likelihoods. Aligned with this expectation, improvements of RCO over the out-of-sample log-likelihoods of MML are larger when we have the smallest training dataset with 1000 customers. As the number of customers in the training dataset increases, the gap between the out-of-sample log-likelihoods gets smaller, but RCO maintains its edge over MML for a majority of the ground choice models. Similar observations hold when we compare RCO with MAC. We shortly demonstrate that these improvements in log-likelihoods translate into more profitable assortments for RCO. Although RCO and MML are variants of the multinomial logit model, it is encouraging that RCO can yield more profitable assortments.

When we fit RCO to the training dataset, the maximum rank cutoff of a customer comes out to be four to five. Fitting RCO using maximum likelihood estimation takes about four to six minutes. When we fit SAC and MAC with 1000 customers in the training dataset, cross validation chooses the number of customer types for MAC as one in three of the 10 ground choice models, in which case, the out-of-sample log-likelihoods of SAC and MAC end up being the same. When we have 1750 or 2500 customers in the training dataset, this behavior goes away and we end up choosing the number of customer types for MAC larger than one. Similarly, the out-of-sample log-likelihoods of MAC may be smaller than those of SAC when we have a small number of customers in the training dataset, which is related to the potential overfitting issue for MAC discussed above when we have a small number of customers in the training dataset. Lastly, to identify situations where RCO performs especially well when compared with the benchmarks, we replicated our results under different approaches for generating the ground choice model. We do not report these results in detail, but if the length of the preference lists in the ground choice model is small, then the out-of-sample log-likelihoods of RCO can be dramatically larger than those of SML and MML. Shorter preference lists in the ground choice model translate into customers being more picky. While the rank cutoffs in RCO allow us to capture picky customers, SML and MML do not necessarily have a similar capability. On the other hand, there is an inherent ordering between the products in the ground choice model that we use, where product 1 is the most preferred and product n is the least preferred. Note that SAC and MAC also work under the assumption that there is a fixed preference ranking for the products. In contrast, if the preference lists in the ground choice model are random permutations of the products that do not necessarily follow a fixed order, then the out-of-sample log-likelihoods of RCO end up being dramatically larger than those of SAC and MAC.

7.3 Comparing Expected Revenue Performance

It turns out that the improvements in the out-of-sample log-likelihoods provided by RCO can translate into significantly more profitable assortments. We use the following approach to compare the ability of the fitted choice models to pick profitable assortments. Let $\phi_i^{\mathsf{RCO}}(S)$ and $\phi_i^{\mathsf{GR}}(S)$ be the choice probabilities of product i within assortment S, respectively, under the fitted RCO and ground choice models. Once we fit RCO 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 $S_k^{\mathsf{RCO}} = \arg\max_{S \subseteq N} \sum_{i \in N} r_{ik}\, \phi_i^{\mathsf{RCO}}(S)$, which is the optimal assortment if the choices of the customers were governed by the fitted RCO. The choices of the customers are actually governed by the ground choice model, so the actual expected revenue from the assortment S_k^{RCO} is $\mathsf{Rev}_k^{\mathsf{RCO}} = \sum_{i \in N} r_{ik}\, \phi_i^{\mathsf{GR}}(S_k^{\mathsf{RCO}})$, characterizing the revenue performance of RCO on the sample $(r_{1k}, \ldots, r_{nk})$. We compute $\mathsf{Rev}_k^{\mathsf{SML}}$, $\mathsf{Rev}_k^{\mathsf{MML}}$, $\mathsf{Rev}_k^{\mathsf{SAC}}$ and $\mathsf{Rev}_k^{\mathsf{MAC}}$ similarly to characterize the revenue performance of the benchmarks SML, MML, SAC and MAC on the sample $(r_{1k}, \ldots, r_{nk})$.

In Table 4, we compare the expected revenues from the assortments picked by the five fitted choice models. As earlier, the top, middle and bottom portions of the table correspond to different values for the number of customers τ in the training dataset. The first column of the table gives the index of the ground choice model. In the rest of the table, we have four blocks, each containing three columns. The four blocks compare RCO with each of the four benchmarks SML, MML, SAC and MAC. The first column in the first block gives the average percent gap between the expected revenues obtained by RCO and SML, where the average is computed over the 100 product revenue samples. Thus, this column gives the average of the data $\{100 \times \frac{\mathsf{Rev}_k^{\mathsf{RCO}} - \mathsf{Rev}_k^{\mathsf{SML}}}{\mathsf{Rev}_k^{\mathsf{RCO}}} : k = 1, \ldots, 100\}$. The second and third columns in the first block give the number of samples for which we, respectively, have $\mathsf{Rev}_k^{\mathsf{RCO}} > \mathsf{Rev}_k^{\mathsf{SML}}$ and $\mathsf{Rev}_k^{\mathsf{SML}} > \mathsf{Rev}_k^{\mathsf{RCO}}$, 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 $\mathsf{Rev}_k^{\mathsf{RCO}} = \mathsf{Rev}_k^{\mathsf{SML}}$ for some revenue samples. The remaining three blocks compare the expected revenues from the assortments picked by the fitted RCO with the expected revenues

$\tau = 1000$

Grnd. Ch.	Comp. with SML Perc. Gap	RCO ≻ SML	SML ≻ RCO	Comp. with MML Perc. Gap	RCO ≻ MML	MML ≻ RCO	Comp. with SAC Perc. Gap	RCO ≻ SAC	SAC ≻ RCO	Comp. with MAC Perc. Gap	RCO ≻ MAC	MAC ≻ RCO
1	8.33	77	10	6.14	82	10	5.86	70	28	5.86	70	28
2	8.09	78	8	6.00	72	12	3.10	70	29	2.83	64	36
3	7.96	80	11	4.09	67	21	4.04	71	28	3.49	65	34
4	7.73	74	15	7.83	82	12	4.44	71	28	4.44	71	28
5	6.66	72	22	5.76	78	17	6.17	80	19	6.17	80	19
6	6.84	66	28	5.49	69	26	5.47	76	23	5.16	69	30
7	4.98	62	33	3.61	66	33	3.34	63	36	4.65	72	28
8	9.26	77	14	4.59	71	18	4.87	67	32	1.01	60	40
9	6.44	73	22	6.17	74	25	4.57	68	32	3.22	62	38
10	5.10	63	31	2.12	53	39	2.98	56	42	1.76	60	39
Avg.	7.14	72	19	5.18	71	21	4.48	69	30	3.86	67	32

$\tau = 1750$

Grnd. Ch.	Comp. with SML Perc. Gap	RCO ≻ SML	SML ≻ RCO	Comp. with MML Perc. Gap	RCO ≻ MML	MML ≻ RCO	Comp. with SAC Perc. Gap	RCO ≻ SAC	SAC ≻ RCO	Comp. with MAC Perc. Gap	RCO ≻ MAC	MAC ≻ RCO
1	7.99	68	20	2.88	66	22	2.36	59	38	-0.02	46	53
2	7.64	78	8	3.54	67	19	3.23	71	28	1.67	56	43
3	7.90	80	12	4.13	68	18	3.46	58	41	2.44	64	34
4	7.62	76	16	3.83	70	21	4.39	81	18	4.94	80	19
5	6.94	74	20	3.09	57	32	5.54	72	26	4.24	68	28
6	6.53	62	29	0.07	47	48	2.73	61	38	0.23	49	51
7	5.43	63	27	1.99	58	32	3.66	76	24	4.39	77	22
8	8.02	75	14	0.81	55	33	3.49	69	31	0.11	43	56
9	6.86	74	21	3.43	71	23	6.17	72	27	4.68	69	30
10	7.03	71	22	-0.45	39	52	3.34	63	35	2.92	64	34
Avg.	7.20	72	19	2.33	60	30	3.84	68	31	2.56	62	37

$\tau = 2500$

Grnd. Ch.	Comp. with SML Perc. Gap	RCO ≻ SML	SML ≻ RCO	Comp. with MML Perc. Gap	RCO ≻ MML	MML ≻ RCO	Comp. with SAC Perc. Gap	RCO ≻ SAC	SAC ≻ RCO	Comp. with MAC Perc. Gap	RCO ≻ MAC	MAC ≻ RCO
1	8.10	73	21	1.98	66	29	2.45	61	34	0.14	49	51
2	8.26	82	9	2.55	62	23	4.24	78	21	3.26	67	30
3	6.77	74	15	1.29	45	35	3.13	54	46	1.19	47	50
4	7.56	78	15	5.02	73	21	5.17	78	20	4.36	78	21
5	6.93	74	20	2.54	63	30	4.79	76	23	1.36	56	44
6	6.55	64	28	0.22	45	45	3.06	65	33	1.51	57	43
7	5.56	66	23	2.65	66	23	4.09	77	23	2.55	70	30
8	7.58	73	17	0.78	53	35	4.12	72	27	0.07	48	52
9	6.70	73	22	1.69	58	35	3.56	72	28	3.90	74	25
10	7.06	72	20	0.61	49	41	3.72	65	35	1.74	61	38
Avg.	7.11	73	19	1.93	58	32	3.83	70	29	2.01	61	38

Table 4 Comparison of the expected revenues obtained by using the fitted choice models.

from the assortments picked by the fitted MML, SAC and MAC. The layout of the columns in these three blocks is identical to that of the first block.

The results in Table 4 indicate that RCO provides significant improvements over SML and SAC in terms of expected revenues. Over a total of 3000 product revenue samples, in 2172 samples, the revenue performance of RCO is better than that of SML, whereas in 573 samples, the revenue performance of SML is better than that of RCO. The product revenue samples for which the expected revenues obtained by RCO and SML match correspond to the case where the assortment S_k^{RCO} that RCO picks is revenue-ordered. Recall that the optimal assortment S_k^{SML} under SML is known to be revenue-ordered assortment. Therefore, when S_k^{RCO} happens to be revenue-ordered, there is a significant likelihood that we will have $S_k^{\mathsf{RCO}} = S_k^{\mathsf{SML}}$, but note that S_k^{RCO} and S_k^{SML} can also be two different revenue-ordered assortments. In 2072 samples, the revenue performance of RCO is better than that of SAC, whereas the situation is reversed in 893 samples. Comparing RCO with the benchmarks that involve multiple customer types, in 1892 samples, the revenue performance of RCO is better than that of MML, whereas in 830 samples, the revenue performance of MML is better than that of RCO. As the number of customers τ in the training dataset increases, the gap between the expected revenues obtained by RCO and MML gets smaller. As the number of customers in the training dataset increases, MML can exploit the fact that it can approximate any choice model that is based on random utility maximization. Nevertheless, even when we have 2500 customers in the training dataset, RCO can maintain its edge over MML. Similar observations hold when we compare RCO with MAC. In 1896 samples, the revenue performance of RCO is better than that of MAC, whereas the situation is reversed in 1074 samples.

The results in Tables 3 and 4 are based on 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, RCO improves the out-of-sample log-likelihoods of SML, MML, SAC and MAC, respectively, by 0.81%, 0.27%, 0.61% and 0.34% on average. Over the 300 combinations, in 298, 226, 296 and 233 combinations, the out-of-sample log-likelihoods of RCO are, respectively, better than those of SML, MML, SAC and MAC. Moving to the expected revenue performance, RCO improves the expected revenues obtained by SML, MML, SAC and MAC, respectively, by 7.35%, 3.04%, 3.79% and 2.44% on average. In 73%, 62%, 68% and 63% of the product revenue samples, the expected revenue performance of RCO is, respectively, better than that of SML, MML, SAC and MAC. Overall, our results indicate that adding the rank cutoffs to the multinomial logit model can noticeably enrich the behavior of the multinomial logit model. Although MML can approximate any choice model that is based on random utility maximization, RCO remains, at minimum, competitive to MML, especially under low to moderate amounts of data availability.

We also compared RCO with the four benchmark choice models 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 G. In Appendix H, we test the performance of our PTAS in Section 5 by comparing the expected revenues from the assortments obtained by the PTAS with an upper bound on the optimal expected revenues.

8. Conclusions

Our work opens up several areas of investigation. First, 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. Second, our PTAS naturally handles a cardinality constraint that limits the number of offered products. We generate the candidate assortments in the same way, but focus on those that satisfy the cardinality constraint. However, our PTAS does not 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. Third, an interesting line of investigation is to extend our work to a mixture of multinomial logit models, where there are multiple customer types, each differing not only in their rank cutoff but also in the preference weight they attach to a product. Under a mixture of multinomial logit models, even if all customers have a rank cutoff of n, it is NP-hard to approximate the assortment optimization problem within a factor that is better than $\frac{1}{n^{1-\epsilon}}$ for any $\epsilon > 0$; see Desir and Goyal (2014). On the other hand, choosing, for example, $\epsilon = \frac{1}{2}$ in our PTAS, we can obtain a $\frac{1}{2}$-approximation for our assortment optimization problem in polynomial time. Thus, the extension from all customer types assigning the same preference weight to a product to different customer types assigning different preference weights to a product dramatically changes the complexity of the problem. Giving a fully polynomial-time approximation scheme, dealing with knapsack constraints and incorporating multiple customer types that attach different preference weights to a product appear to require new lines of attack. Lastly, incorporating rank cutoffs or other means to enhance the flexibility of choice models is a rich research area, irrespective of whether the starting point is the multinomial logit or any other choice model.

Acknowledgements: We thank the area editor, associate editor and two anonymous referees for their comments that substantially improved the results and exposition in the paper.

References

Alba, J. W., J. W. Hutchinson. 1987. Dimensions of consumer expertise. *Journal of Consumer Research* **13**(4) 411–454.

Andrews, R. L., T. C. Srinivasan. 1995. Studying consideration effects in empirical choice models using scanner panel data. *Journal of Marketing Research* **32**(1) 30–41.

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.

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

Bronnenberg, B. J., W. R. Vanhonacker. 1996. Limited choice sets, local price response, and implied measures of price competition. *Journal of Marketing Research* **33**(2) 163–173.

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.

Gallego, G., A. Li. 2017. Attention, consideration then selection choice model. Tech. rep., Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong.

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.

Gaundry, M. J. I., M. G. Dagenais. 1979. The dogit model. *Transportation Research Part B: Methodological* **13**(2) 105–111.

Gensch, D. H. 1987. A two-stage disaggregate attribute choice model. *Marketing Science* **6**(3) 223–239.

Hanemann, W. M. 1984. Discrete/continuous models of consumer demand. *Econometrica* **52**(3) 541–561.

Hauser, J. R. 2014. Consideration-set heuristics. *Journal of Business Research* **67**(8) 1688–1699.

Hauser, J. R., B. Wernerfelt. 1990. An evaluation cost model of consideration sets. *Journal of Consumer Research* **16**(4) 393–408.

Jagabathula, S., D. Mitrofanov, G. Vulcano. 2020. Inferring consideration sets from sales transaction data. Tech. rep., New York University, 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.

Kahneman, D. 2003. Maps of bounded rationality: Psychology for behavioral economics. *The American Economic Review* **93**(5) 1449–1475.

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

McCarthy, P. S. 1997. The role of captivity in aggregate share models of intercity passenger travel. *Journal of Transport Economics and Policy* **31**(3) 293–308.

McFadden, Daniel, Kenneth Train. 2000. Mixed MNL models for discrete response. *Journal of Applied Economics* **15** 447–470.

Mehta, N., S. Rajiv, K. Srinivasan. 2003. Price uncertainty and consumer search: A structural model of consideration set formation. *Marketing Science* **22**(1) 58–84.

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.

Miller, G. A. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* **63**(2) 81.

Nelson, P. 1970. Information and consumer behavior. *Journal of Political Economy* **78**(2) 311–329.

Roberts, J. H., J. M. Lattin. 1991. Development and testing of a model of consideration set composition. *Journal of Marketing Research* **28**(4) 429–440.

Rubinstein, A. 1998. *Modeling Bounded Rationality*. MIT Press, Cambridge, MA.

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

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

Simon, H. A. 1955. A behavioral model of rational choice. *The Quarterly Journal of Economics* **69**(1) 99–118.

Stigler, G. J. 1961. The economics of information. *Journal of Political Economy* **69**(3) 213–225.

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).

Swait, J., M. Ben-Akiva. 1987. Incorporating random constraints in discrete models of choice set generation. *Transportation Research Part B: Methodological* **21**(2) 91–102.

Swait, J. D. 1984. Probabilistic choice set generation in transportation demand models. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA.

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. 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 for "Assortment Optimization under the Multinomial Logit Model with Utility-Based Rank Cutoffs"

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 3, 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: Computational Complexity

We give a proof for Theorem 4.1. The proof uses a reduction from the partition problem. Using $\mathbb{Q}_+$ to denote the set of positive rationals, the partition problem is defined as follows.

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

Proof of Theorem 4.1:

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 $w_i\in\mathbb{Q}_+$. Scaling the weight of each item by the same amount does not change the answer to the partition problem, so we normalize the weights so that $\sum_{i\in N} w_i = 1$. Given the particular 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 $\Theta = \sum_{i\in N} \frac{w_i}{1+w_i} < 1$, for $i = 1,\dots,n$, we set the preference weight of product i as $v_i = \frac{2}{1-\Theta}\frac{w_i}{1+w_i}$. We set $v_{n+1} = 1$. For $i = 1,\dots,n$, we set the revenue of product i as $r_i = \frac{1-\Theta}{2}(1+w_i)$. 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 rank cutoff of 2, so $m = 2$ and $(\lambda^1, \lambda^2) = (0, 1)$. Lastly, we set the expected revenue threshold as $K = \frac{9(1-\Theta)}{8}$. Since product $n+1$ has the largest revenue, offering product $n+1$ improves the expected revenue of any assortment. We establish this fact in Lemma B.1 shortly in this section. Thus, the assortment feasibility problem asks which of the remaining products in $N = \{1,\dots,n\}$ to offer to obtain an expected revenue of $\frac{9(1-\Theta)}{8}$ or more. By (7), 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 noting 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

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

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 $(1+\sum_{i\in S} w_i)\,(2-\sum_{i\in S} w_i) \ge \frac{9}{4}$.

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 $\sum_{i\in S} w_i = \frac{1}{2}$. ■

In the next lemma, we show that adding the product with the largest revenue to any assortment increases the expected revenue under our choice model.

Lemma B.1 *For any product $\ell \in N$ and assortment $S \subseteq N_{-\ell}$, if $r_\ell \geq r_i$ for all $i \in S$, then we have $\sum_{i\in S\cup\{\ell\}} r_i\,\phi_i(S\cup\{\ell\}) \geq \sum_{i\in S} r_i\,\phi_i(S)$.*

Proof: In Lemma B.2 that we give next in this section, we show that adding a product to any assortment decreases the purchase probability of any product already in the assortment, as well as the no-purchase probability under our choice model. In other words, using $\phi_0(S)$ to denote the probability that a customer does not make a purchase when we offer the assortment S, for any $j \in N$ and $S \subseteq N_{-j}$, we have $\phi_i(S\cup\{j\}) \leq \phi_i(S)$ for all $i \in S$ and $\phi_0(S\cup\{j\}) \leq \phi_0(S)$. In this case, using the fact that $\phi_\ell(S\cup\{\ell\}) = 1 - \phi_0(S\cup\{\ell\}) - \sum_{i\in S}\phi_i(S\cup\{\ell\})$, we get

$$\begin{aligned}
\sum_{i\in S\cup\{\ell\}} r_i\,\phi_i(S\cup\{\ell\}) \;&=\; r_\ell\Big[1-\phi_0(S\cup\{\ell\}) - \sum_{i\in S}\phi_i(S\cup\{\ell\})\Big] + \sum_{i\in S} r_i\,\phi_i(S\cup\{\ell\})\\
&=\; r_\ell\Big[1-\phi_0(S\cup\{\ell\})\Big] + \sum_{i\in S}(r_i - r_\ell)\,\phi_i(S\cup\{\ell\})\\
&\overset{(a)}{\geq}\; r_\ell\Big[1-\phi_0(S)\Big] + \sum_{i\in S}(r_i - r_\ell)\,\phi_i(S) \;\overset{(b)}{=}\; \sum_{i\in S} r_i\,\phi_i(S),
\end{aligned}$$

where (a) follows from Lemma B.2 as well as the fact that $r_\ell \geq r_i$ for all $i \in S$, whereas (b) holds because we have $\phi_0(S) + \sum_{i\in S}\phi_i(S) = 1$. ∎

In the next lemma, we show that adding a product to any assortment decreases the purchase probability of any product already in the assortment, as well as the no-purchase probability.

Lemma B.2 *Letting $\pi_0^k(S) = 1 - \sum_{i\in S}\pi_i^k(S)$, for any $j \in N$, $S \subseteq N_{-j}$ and $k \in M$, we have $\pi_i^k(S\cup\{j\}) \leq \pi_i^k(S)$ for all $i \in S$ and $\pi_0^k(S\cup\{j\}) \leq \pi_0^k(S)$.*

Proof: We use induction over the rank cutoff to show that $B^k(S\cup\{j\},N) \leq B^k(S,N)$ for all $j \in N$, $S \subseteq N_{-j}$ and $k \in M$. Since $B^1(\cdot,\cdot) = 1$, the result holds for $k = 1$. Assuming that the result holds for rank cutoff of $k-1$, we show that the result holds for rank cutoff of k. By the definition of $B^k(S,N)$ in (1), for any $j \in N$ and $S \subseteq N_{-j}$, we have

$$\begin{aligned}
&B^k(S\cup\{j\},N) \;=\; 1 \;+ \sum_{i\in N\setminus(S\cup\{j\})} \frac{v_i}{1+V(N_{-i})}\,B^{k-1}(S\cup\{j\},N_{-i})\\
&\overset{(a)}{\leq}\; 1 \;+ \sum_{i\in N\setminus(S\cup\{j\})} \frac{v_i}{1+V(N_{-i})}\,B^{k-1}(S,N_{-i}) \;\leq\; 1 \;+ \sum_{i\in N\setminus S} \frac{v_i}{1+V(N_{-i})}\,B^{k-1}(S,N_{-i}) \;=\; B^k(S,N),
\end{aligned}$$

where (a) holds because if we have $S \subseteq N_{-j}$ and $i \in N\setminus(S\cup\{j\})$, then we also have $S \subseteq N_{-i}\setminus\{j\}$, in which case, using the induction assumption with the set of products N_{-i}, we have the inequality $B^{k-1}(S\cup\{j\},N_{-i}) \leq B^{k-1}(S,N_{-i})$. The chain of inequalities above completes the induction argument. In this case, by Theorem 3.1, we obtain $\pi_i^k(S\cup\{j\}) =$

$\frac{v_i}{1+V(N)} B^k(S\cup\{j\},N) \le \frac{v_i}{1+V(N)} B^k(S,N) = \pi_i^k(S)$, which is the first inequality in the lemma. Next, we focus on the second inequality in the lemma. We use induction over the rank cutoff to show that $V(S\cup\{j\})\, B^k(S\cup\{j\},N) \ge V(S)\, B^k(S,N)$ for all $j\in N$, $S\subseteq N_{-j}$ and $k\in M$. Since $B^1(\cdot,\cdot)=1$ and $V(S\cup\{j\})\ge V(S)$, the result holds for $k=1$. Assuming that the result holds for rank cutoff of $k-1$, we show that the result holds for rank cutoff of k. A simple lemma, which we give later in the paper in Lemma C.1, shows that $\frac{1}{1+V(N)}\, B^k(S,N) \le \frac{1}{1+V(S)}$ for any $S\subseteq N$. In this case, by the definition of $B^k(S,N)$ in (1), for any $j\in N$ and $S\subseteq N_{-j}$, we have

$$
\begin{aligned}
V(S\cup\{j\})\, B^k(S\cup\{j\},N) \;&=\; V(S\cup\{j\}) \;+ \sum_{i\in N\setminus(S\cup\{j\})} \frac{v_i}{1+V(N_{-i})}\, V(S\cup\{j\})\, B^{k-1}(S\cup\{j\},N_{-i}) \\
&\overset{(b)}{\ge}\; V(S\cup\{j\}) \;+ \sum_{i\in N\setminus(S\cup\{j\})} \frac{v_i}{1+V(N_{-i})}\, V(S)\, B^{k-1}(S,N_{-i}) \\
&=\; V(S\cup\{j\}) + \sum_{i\in N\setminus S} \frac{v_i}{1+V(N_{-i})}\, V(S)\, B^{k-1}(S,N_{-i}) - \frac{v_j}{1+V(N_{-j})}\, V(S)\, B^{k-1}(S,N_{-j}) \\
&\overset{(c)}{\ge} V(S\cup\{j\}) + \sum_{i\in N\setminus S} \frac{v_i}{1+V(N_{-i})}\, V(S)\, B^{k-1}(S,N_{-i}) - v_j\, \frac{V(S)}{1+V(S)} \\
&\overset{(d)}{\ge} V(S) + \sum_{i\in N\setminus S} \frac{v_i}{1+V(N_{-i})}\, V(S)\, B^{k-1}(S,N_{-i}) \\
&\overset{(e)}{=}\; V(S)\, B^k(S,N),
\end{aligned}
$$

where (b) is by the induction assumption, (c) uses the fact that $\frac{1}{1+V(N_{-j})}\, B^{k-1}(S,N_{-j}) \le \frac{1}{1+V(S)}$ by Lemma C.1, (d) holds by noting that $V(S\cup\{j\}) = V(S)+v_j$ and (e) follows from (1). The chain of inequalities above completes the induction argument.

By Theorem 3.1, we have $\pi_0^k(S) = 1-\sum_{i\in S}\pi_i^k(S) = 1-\frac{\sum_{i\in S} v_i}{1+V(N)} B^k(S,N) = 1-\frac{V(S)}{1+V(N)}\, B^k(S,N)$, so noting the chain of inequalities above, the second equality in the lemma holds because we have $\pi_0^k(S\cup\{j\}) = 1-\frac{V(S\cup\{j\})}{1+V(N)}\, B^k(S\cup\{j\},N) \le 1-\frac{V(S)}{1+V(N)}\, B^k(S,N) = \pi_0^k(S)$. ■

The choice probability of product i is $\phi_i(S) = \sum_{k\in M}\lambda^k\, \pi_i^k(S)$, so $\phi_0(S) = 1-\sum_{i\in S}\phi_i(S) = \sum_{k\in M}\lambda^k - \sum_{k\in M}\sum_{i\in S}\lambda^k\, \pi_i^k(S) = \sum_{k\in M}\lambda^k\,(1-\sum_{i\in S}\pi_i^k(S)) = \sum_{k\in M}\lambda^k\, \pi_0^k(S)$. Thus, Lemma B.2 yields $\phi_i(S\cup\{j\}) \le \phi_i(S)$ and $\phi_0(S\cup\{j\}) \le \phi_0(S)$ for all $j\in N$, $S\subseteq N_{-j}$ and $i\in S$.

Appendix C: Impact of Rank Cutoffs

In this section, we give a proof for Theorem 4.2. We will use two preliminary lemmas in the proof. In the next lemma, we start by giving an upper bound on $B^k(S,N)$ that we use to compute our choice probabilities. Recall that $B^k(S,N)$ is computed through the recursion in (1).

Lemma C.1 *For any rank cutoff $k \in M$ and assortment $S \subseteq N$, letting $B^k(S,N)$ be computed through the recursion in* (1)*, we have*

$$\frac{1}{1+V(N)} B^k(S,N) \;\leq\; \frac{1}{1+V(S)}. \tag{15}$$

Proof: We use induction over the rank cutoff to show that the inequality in (15) holds for any set of products N, rank cutoff $k \in M$ and assortment $S \subseteq N$. 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 $\frac{1}{1+V(N)}\, B^1(S,N) = \frac{1}{1+V(N)} \leq \frac{1}{1+V(S)}$. Therefore, the result holds for $k=1$. Assuming that the the inequality in (15) holds for rank cutoff of $k-1$, we show that the inequality in (15) holds for rank cutoff of k as well. 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 $\frac{1}{1+V(N_{-i})} B^{k-1}(S,N_{-i}) \;\leq\; \frac{1}{1+V(S)}$ for each $i \in N \setminus S$. Therefore, using the definition of $B^k(S,N)$ in (1), we get the chain of inequalities

$$\begin{aligned} B^k(S,N) \;&=\; 1 + \sum_{i \in N\setminus S} \frac{v_i}{1+V(N_{-i})}\, B^{k-1}(S,N_{-i}) \\ &\leq\; 1 + \sum_{i \in N\setminus S} \frac{v_i}{1+V(S)} \;=\; 1 + \frac{V(N)-V(S)}{1+V(S)} = \frac{1+V(N)}{1+V(S)}. \end{aligned}$$

Arranging the terms in the chain of inequalities above yields $\frac{1}{1+V(N)} B^k(S,N) \;\leq\; \frac{1}{1+V(S)}$, so the result holds for rank cutoff 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 rank cutoff 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 rank cutoff k conditional on the top product with the largest utility.

Lemma C.2 *For any set of products N, rank cutoff $k \in M$ and assortment $S \subseteq N$, the revenue from a customer with rank cutoff k satisfies*

$$\mathbb{E}\Big\{R^k(S,N)\,\Big|\, U_i \geq U_0 \vee \max_{j \in N_{-i}} U_j\Big\} \;=\; \begin{cases} r_i & \text{if } i \in S \\ \mathbb{E}\{R^{k-1}(S,N_{-i})\} & \text{if } i \notin S. \end{cases} \tag{16}$$

Proof: Given $U_i \geq U_0 \vee \max_{j \in N_{-i}} U_j$, product i occupies the top spot. Thus, if we have $i \in S$, then a customer with rank cutoff k purchases product i, yielding $\mathbb{E}\{R^k(S,N) \,|\, U_i \geq U_0 \vee \max_{j \in N_{-i}} U_j\} = r_i$. Indeed, a customer with any rank cutoff would purchase product i as long as the rank cutoff is non-zero, so the result holds for all $i \in S$. On the other hand, if we have $i \notin S$, then product i is not available, but given $U_i \geq U_0 \vee \max_{j \in N_{-i}} U_j$, it occupies the top spot. Thus, if we offer the assortment S, then a customer with rank cutoff 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 in Section 3, 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 rank cutoff $k-1$ chooses product j when set of all products is N_{-i}, in which case, we get $\mathbb{E}\{R^k(S,N) \,|\, U_i \geq U_0 \vee \max_{j \in N_{-i}} U_j\} = \mathbb{E}\{R^{k-1}(S, N_{-i})\}$, so the result holds for all $i \notin S$. ■

We make two observations. First, noting the definition of $R^k(S,N)$, if we offer the assortment S, then the expected revenue from a customer is $\sum_{k \in M} \lambda^k \, \mathbb{E}\{R^k(S,N)\}$. Second, by Theorem 3.1, given that the set of products is N and we offer the assortment S, a customer with rank cutoff 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 rank cutoff k is $\frac{\sum_{i \in S} r_i \, v_i}{1+V(N)} \, B^k(S,N)$. By the definition of $R^k(S,N)$, if we offer the assortment S, then the expected revenue from a customer with rank cutoff k is also given by $\mathbb{E}\{R^k(S,N)\}$. Therefore, we have $\frac{\sum_{i \in S} r_i \, v_i}{1+V(N)} \, B^k(S,N) = \mathbb{E}\{R^k(S,N)\}$.

We proceed to giving a proof for Theorem 4.2. In the proof, we assume that we choose the optimal solutions to problems (7) and (8) as those with the largest cardinality.

Proof of Theorem 4.2:

Letting $S_{+i} = S \cup \{i\}$, we use induction over the rank cutoff 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 (7) 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 rank cutoff of $k-1$, we show that the result holds for rank cutoff 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 rank cutoff k, conditional on the fact that product i provides the largest utility. In other words, we have $\overline{R}_i^k(S,N) = \mathbb{E}\{R^k(S,N) \,\big|\, U_i \geq U_0 \vee \max_{j \in N_{-i}} U_j\}$, which corresponds to the left side of (16). 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 $\overline{R}_i^k(S_{+\ell}, N) \geq \overline{R}_i^k(S,N)$ for each $i \in N$.

First, consider the case $i \in S$. Thus, we have $i \in S$ and $i \in S_{+\ell}$, in which case, by (16), $\overline{R}_i^k(S,N) = r_i = \overline{R}_i^k(S_{+\ell}, N)$. Second, consider the case $i \notin S$ and $i \neq \ell$. Therefore, we have $i \notin S$ and $i \notin S_{+\ell}$, so by (16), we get $\overline{R}_i^k(S,N) = \mathbb{E}\{R^{k-1}(S,N_{-i})\} \le \mathbb{E}\{R^{k-1}(S_{+\ell},N_{-i})\} = \overline{R}_i^k(S_{+\ell},N)$, 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 (16), we get $\overline{R}_i^k(S_{+\ell},N) = r_i = r_\ell$. A well-known result for assortment optimization under the standard multinomial logit model shows that the products with revenues exceeding the optimal expected revenue are included in the optimal assortment. For completeness, we give this result in Lemma C.3 shortly in this section. In this case, recalling that $\widetilde{S}$ is the optimal assortment to problem (8), which is the assortment optimization problem under the standard multinomial logit model, we have $r_\ell \ge \frac{\sum_{q\in\widetilde{S}} r_q\, v_q}{1+V(\widetilde{S})}$ for all $\ell \in \widetilde{S}$. Therefore, since $\ell \in \widetilde{S} \setminus S$, we get $\overline{R}_i^k(S_{+\ell},N) = r_\ell \ge \frac{\sum_{q\in\widetilde{S}} r_q\, v_q}{1+V(\widetilde{S})} \ge \frac{\sum_{q\in S} r_q\, v_q}{1+V(S)}$, where the last inequality is by the fact that $\widetilde{S}$ is an optimal solution to problem (8). Furthermore, since $i \notin S$, by (16), 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 $\mathbb{E}\{R^{k-1}(S,N_{-i})\} = \frac{\sum_{q\in S} r_q\, v_q}{1+V(N_{-i})}\, B^{k-1}(S,N_{-i})$. Thus, we get

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

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

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

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

Lemma C.3 *Letting $\widetilde{S}$ be an optimal solution to the assortment optimization problem in* (8)*, for all $\ell \in \widetilde{S}$, we have*

$$r_\ell \ge \frac{\sum_{i\in\widetilde{S}} r_i\, v_i}{1+\sum_{i\in\widetilde{S}} v_i}.$$

Proof: Let $\widetilde{Z}$ be the optimal objective value of problem (8), so we have $\widetilde{Z} = \frac{\sum_{i\in\widetilde{S}} r_i\, v_i}{1+\sum_{i\in\widetilde{S}} v_i}$. 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 once more, 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 (8) that exceeds the optimal objective value for this problem, which is a contradiction. ■

Appendix D: Performance of Revenue-Ordered Assortments

We show that if $m=2$, so that the rank cutoff of a customer does not exceed two, then the best revenue-ordered assortment provides a $\frac{1}{2}$-approximation for the assortment optimization problem under our multinomial logit model with rank cutoffs. Furthermore, we give a problem instance to demonstrate that this bound is tight. In particular, the ratio between the expected revenue from the best revenue-ordered assortment and the optimal expected revenue is arbitrarily close to $\frac{1}{2}$ for this problem instance. However, we are not able to show that if $m \geq 3$, so that the rank cutoff of a customer can exceed two, then the best-revenue ordered assortment still provides a $\frac{1}{2}$-approximation. We leave the question of characterizing the performance of best revenue-ordered assortment for larger rank cutoffs as an open problem. Throughout this section, we focus on the case with $m=2$. Using the fact that $B^1(\cdot,\cdot)=1$, by (1), we get $B^2(S,N)=1+\sum_{i\in N\setminus S}\frac{v_i}{1+V(N_{-i})}$. Thus, letting $\Theta(S)=\sum_{i\in S}\frac{v_i}{1+V(N_{-i})}$ for notational brevity, we have $B^2(S,N)=1+\Theta(N\setminus S)$. In this case, dropping the multiplicative constant $\frac{1}{1+V(N)}$ that does not change the optimal solution to problem (7), letting $\mu=\lambda^2$ for notational brevity and noting that $\lambda^1+\lambda^2=1$, problem (7) with $m=2$ can equivalently be written as

$$\max_{S\subseteq N}\Big\{W(S)\Big[\lambda^1+\lambda^2\Big\{1+\Theta(N\setminus S)\Big\}\Big]\Big\} \;=\; \max_{S\subseteq N}\Big\{W(S)\,\Big[1+\mu\,\Theta(N\setminus S)\Big]\Big\}. \tag{17}$$

We use $\mathsf{Rev}(S)$ to denote the objective function of problem (17). Letting S^* be an optimal solution to problem (17), define the assortment $\overline{S}=\{i\in N: r_i\geq \frac{1}{2\,(1+V(N))}\,\mathsf{Rev}(S^*)\}$.

The assortment $\overline{S}$ is revenue-ordered, including all products with revenues exceeding $\frac{1}{2\,(1+V(N))}\,\mathsf{Rev}(S^*)$. In the next theorem, we show that $\overline{S}$ is a $\frac{1}{2}$-approximate solution to (17).

Theorem D.1 *Letting S^* be an optimal solution to problem* (17) *and defining the assortment $\overline{S}=\{i\in N: r_i\geq \frac{1}{2\,(1+V(N))}\,\mathsf{Rev}(S^*)\}$, we have*

$$\mathsf{Rev}(\overline{S})\geq \frac{1}{2}\mathsf{Rev}(S^*).$$

Proof: Let $S^*_D=S^*\setminus\overline{S}$, $\overline{S}_D=\overline{S}\setminus S^*$ and $S_U=S^*\cup\overline{S}$. Note that $\overline{S}\cup S^*_D=S_U=S^*\cup\overline{S}_D$, so $\overline{S}=(S^*\cup\overline{S}_D)\setminus S^*_D$, which implies that $W(\overline{S})=W(S^*)+W(\overline{S}_D)-W(S^*_D)$. Thus, we have

$$\begin{aligned}\mathsf{Rev}(\overline{S}) \;&=\; W(\overline{S})\,\big[1+\mu\,\Theta(N\setminus\overline{S})\big]\\ &=\; W(S^*)\,\big[1+\mu\,\Theta(N\setminus\overline{S})\big]+\big[W(\overline{S}_D)-W(S^*_D)\big]\,\big[1+\mu\,\Theta(N\setminus\overline{S})\big].\end{aligned} \tag{18}$$

By the definitions of S^*_D and $\overline{S}_D$, we have $(N\setminus\overline{S})\cup\overline{S}_D=N\setminus(S^*\cap\overline{S})=(N\setminus S^*)\cup S^*_D$, in which case, we have $N\setminus\overline{S}=((N\setminus S^*)\cup S^*_D)\setminus\overline{S}_D$. Therefore, we get $\Theta(N\setminus\overline{S})=\Theta(N\setminus S^*)+\Theta(S^*_D)-\Theta(\overline{S}_D)$.

Furthermore, by the definition of $\mathsf{Rev}(S^*)$, we have $W(S^*) = \frac{1}{1+\mu\,\Theta(N\setminus S^*)}\,\mathsf{Rev}(S^*)$. Using the last two equalities, we express the right side of (18) equivalently as

$$
\begin{aligned}
&\frac{1+\mu\,\Theta(N\setminus S^*)+\mu\,\Theta(S^*_D)-\mu\,\Theta(\overline{S}_D)}{1+\mu\,\Theta(N\setminus S^*)}\,\mathsf{Rev}(S^*)+\big[W(\overline{S}_D)-W(S^*_D)\big]\,\big[1+\mu\,\Theta(N\setminus \overline{S})\big]\\
&=\mathsf{Rev}(S^*)+\frac{\mu\,\Theta(S^*_D)-\mu\,\Theta(\overline{S}_D)}{1+\mu\,\Theta(N\setminus S^*)}\,\mathsf{Rev}(S^*)+\big[W(\overline{S}_D)-W(S^*_D)\big]\,\big[1+\mu\,\Theta(N\setminus \overline{S})\big]\\
&\overset{(a)}{\geq}\mathsf{Rev}(S^*)+\frac{\mu\,\Theta(S^*_D)-\mu\,\Theta(\overline{S}_D)}{1+\mu\,\Theta(N\setminus S^*)}\,\mathsf{Rev}(S^*)+\frac{\mathsf{Rev}(S^*)}{2\,(1+V(N))}\big[V(\overline{S}_D)-V(S^*_D)\big]\,\big[1+\mu\,\Theta(N\setminus \overline{S})\big]\\
&=\mathsf{Rev}(S^*)\left[1+\frac{\mu\,\Theta(S^*_D)}{1+\mu\,\Theta(N\setminus S^*)}-\frac{\mu\,\Theta(\overline{S}_D)}{1+\mu\,\Theta(N\setminus S^*)}+\frac{\big[V(\overline{S}_D)-V(S^*_D)\big]\,\big[1+\mu\,\Theta(N\setminus \overline{S})\big]}{2(1+V(N))}\right],
\end{aligned}
\tag{19}
$$

where (a) holds because noting that we have $r_i \geq \frac{1}{2\,(1+V(N))}\,\mathsf{Rev}(S^*)$ for all $i\in\overline{S}$ and $\overline{S}_D\subseteq\overline{S}$, we obtain $W(\overline{S}_D)=\sum_{i\in\overline{S}_D} r_i\,v_i \geq \frac{1}{2\,(1+V(N))}\,\mathsf{Rev}(S^*)\sum_{i\in\overline{S}_D} v_i = \frac{1}{2\,(1+V(N))}\,\mathsf{Rev}(S^*)\,V(\overline{S}_D)$, as well as we have $r_i < \frac{1}{2\,(1+V(N))}\,\mathsf{Rev}(S^*)$ for all $i\in N\setminus\overline{S}$ and $S^*_D = S^*\setminus\overline{S}\subseteq N\setminus\overline{S}$, so we obtain $W(S^*_D)=\sum_{i\in S^*_D} r_i\,v_i \leq \frac{1}{2\,(1+V(N))}\,\mathsf{Rev}(S^*)\sum_{i\in S^*_D} v_i = \frac{1}{2\,(1+V(N))}\,\mathsf{Rev}(S^*)\,V(S^*_D)$. The two equalities above follow by arranging the terms. The first expression in the chain of inequalities in (19) is equal to $\mathsf{Rev}(\overline{S})$, which implies that the last expression in the same chain of inequalities lower bounds $\mathsf{Rev}(\overline{S})$. We focus on each term in the square brackets on the right side of (19) and lower bound each one of these terms, in which case, we lower bound $\mathsf{Rev}(\overline{S})$ as well.

Consider the first term in the square brackets on the right side of (19). By the definition of S^*_D, we have $S^*_D\subseteq S^*$, which yields $\Theta(N\setminus S^*)\leq\Theta(N\setminus S^*_D)$. Noting also that $\mu\leq 1$, we have

$$
\begin{aligned}
\frac{\mu\,\Theta(S^*_D)}{1+\mu\,\Theta(N\setminus S^*)} \;\geq\; \frac{\mu\,\Theta(S^*_D)}{1+\Theta(N\setminus S^*_D)} \;&=\; \frac{\mu\,\Theta(S^*_D)}{1+\sum_{i\in N\setminus S^*_D}\frac{v_i}{1+V(N_{-i})}}\\
&\overset{(b)}{\geq}\; \frac{\mu\,\Theta(S^*_D)}{1+\sum_{i\in N\setminus S^*_D}\frac{v_i}{1+V(S^*_D)}} \;=\; \frac{\mu\,\Theta(S^*_D)}{1+\frac{V(N\setminus S^*_D)}{1+V(S^*_D)}} \;\overset{(c)}{=}\; \frac{1+V(S^*_D)}{1+V(N)}\,\mu\,\Theta(S^*_D),
\end{aligned}
\tag{20}
$$

where (b) holds because if $i\in N\setminus S^*_D$, then $V(N_{-i})=V(N)-v_i\geq V(N)-V(N\setminus S^*_D)=V(S^*_D)$ and (c) uses the fact that $V(S^*_D)+V(N\setminus S^*_D)=V(N)$.

Consider the second term in the square brackets on the right side of (19). By the definition of $\overline{S}_D$, we have $\overline{S}_D=\overline{S}\setminus S^*\subseteq N\setminus S^*$, so $\Theta(\overline{S}_D)\leq\Theta(N\setminus S^*)$. Thus, we get

$$
\begin{aligned}
\frac{\mu\,\Theta(\overline{S}_D)}{1+\mu\,\Theta(N\setminus S^*)} \;\leq\; \frac{\mu\,\Theta(\overline{S}_D)}{1+\mu\,\Theta(\overline{S}_D)} \;&=\; \frac{\mu\sum_{i\in\overline{S}_D}\frac{v_i}{1+V(N_{-i})}}{1+\mu\sum_{i\in\overline{S}_D}\frac{v_i}{1+V(N_{-i})}}\\
&\overset{(d)}{\leq}\; \frac{\sum_{i\in\overline{S}_D}\frac{v_i}{1+V(N\setminus\overline{S}_D)}}{1+\sum_{i\in\overline{S}_D}\frac{v_i}{1+V(N\setminus\overline{S}_D)}} \;=\; \frac{\frac{V(\overline{S}_D)}{1+V(N\setminus\overline{S}_D)}}{1+\frac{V(\overline{S}_D)}{1+V(N\setminus\overline{S}_D)}} \;\overset{(e)}{=}\; \frac{V(\overline{S}_D)}{1+V(N)},
\end{aligned}
\tag{21}
$$

where (d) holds because if $i\in\overline{S}_D$, then we have $\frac{\mu}{1+V(N_{-i})}=\frac{\mu}{1+V(N)-v_i}\leq\frac{\mu}{1+V(N)-V(\overline{S}_D)}=\frac{\mu}{1+V(N\setminus\overline{S}_D)}\leq\frac{1}{1+V(N\setminus\overline{S}_D)}$, as well as noting that $\frac{x}{1+x}$ is increasing in x, whereas (e) uses the fact that $V(\overline{S}_D)+$

$V(N \setminus \overline{S}_D) = V(N)$. Lastly, consider the third term in the square brackets on the right side of (19). We have the identity $N \setminus \overline{S} = (N \setminus (\overline{S} \cup S^*)) \cup (S^* \setminus \overline{S})$, which can be verified by drawing a Venn diagram. Thus, we get $\Theta(N \setminus \overline{S}) = \Theta(N \setminus (\overline{S} \cup S^*)) + \Theta(S^* \setminus \overline{S}) = \Theta(N \setminus S_U) + \Theta(S^*_D)$, where the last equality uses the definition of S_U and S^*_D. In this case, we obtain

$$\begin{aligned}
&\left[V(\overline{S}_D) - V(S^*_D)\right]\left[1 + \mu\,\Theta(N \setminus \overline{S})\right] \\
&\qquad = V(\overline{S}_D)\left[1 + \mu\,\Theta(N \setminus \overline{S})\right] - V(S^*_D)\left[1 + \mu\,\Theta(N \setminus S_U) + \mu\,\Theta(S^*_D)\right] \\
&\qquad\qquad \geq\ V(\overline{S}_D) - V(S^*_D)\left[1 + \mu\,\Theta(N \setminus S_U) + \mu\,\Theta(S^*_D)\right].
\end{aligned} \tag{22}$$

Thus, using the chains of inequalities in (20), (21) and (22) on the right side of (19), we will be able to lower bound the right side of (19), which is what we do next.

By the discussion at the beginning of the proof, a lower bound on the expression on the right side of (19) also yields a lower bound on $\mathsf{Rev}(\overline{S})$. Thus, by (20), (21) and (22), we get

$$\begin{aligned}
\mathsf{Rev}(\overline{S}) &\geq \mathsf{Rev}(S^*)\left[1 + \frac{1 + V(S^*_D)}{1 + V(N)}\,\mu\Theta(S^*_D) - \frac{V(\overline{S}_D)}{1 + V(N)} + \frac{V(\overline{S}_D) - V(S^*_D)\left[1 + \mu\Theta(N \setminus S_U) + \mu\Theta(S^*_D)\right]}{2\,(1 + V(N))}\right] \\
&\stackrel{(f)}{=} \mathsf{Rev}(S^*)\left[1 + \frac{\mu\,\Theta(S^*_D)}{1 + V(N)}\left[1 + \frac{V(S^*_D)}{2}\right] - \frac{V(\overline{S}_D)}{2(1 + V(N))} - \frac{V(S^*_D)}{2\,(1 + V(N))}\left[1 + \mu\,\Theta(N \setminus S_U)\right]\right] \\
&\geq \mathsf{Rev}(S^*)\left[1 - \frac{V(\overline{S}_D)}{2(1 + V(N))} - \frac{V(S^*_D)}{2\,(1 + V(N))}\left[1 + \mu\,\Theta(N \setminus S_U)\right]\right] \\
&\geq \mathsf{Rev}(S^*)\left[1 - \frac{V(\overline{S}_D) + V(S^*_D)}{2(1 + V(N))}\left[1 + \mu\,\Theta(N \setminus S_U)\right]\right] \\
&\stackrel{(g)}{\geq} \mathsf{Rev}(S^*)\left[1 - \frac{V(S_U)}{2(1 + V(N))}\left\{1 + \mu \sum_{i \in N \setminus S_U} \frac{v_i}{1 + V(N_{-i})}\right\}\right] \\
&\stackrel{(h)}{\geq} \mathsf{Rev}(S^*)\left[1 - \frac{V(S_U)}{2(1 + V(N))}\,B^2(S_U, N)\right] \stackrel{(i)}{\geq} \frac{1}{2}\mathsf{Rev}(S^*),
\end{aligned}$$

where (f) is by arranging the terms, (g) follows because $\overline{S}_D \cap S^*_D = \varnothing$ and $\overline{S}_D \cup S^*_D \subseteq \overline{S} \cup S^* = S_U$ as well as using the definition of $\Theta(N \setminus S_U)$, (h) holds because $B^2(S_U, N) = 1 + \sum_{i \in N \setminus S_U} \frac{v_i}{1 + V(N_{-i})}$ by the discussion at the beginning of this section and (i) holds because $B^2(S_U, N) \leq \frac{1 + V(N)}{1 + V(S_U)}$ by Lemma C.1. By the chain of inequalities above, we have $\mathsf{Rev}(\overline{S}) \geq \frac{1}{2}\mathsf{Rev}(S^*)$, as desired. ■

We give a problem instance to show that the best revenue-ordered assortment cannot provide a performance guarantee better than $\frac{1}{2}$. We have $n = 3$ products and all customers have rank cutoff of $m = 2$. The revenues and preference weights are $(r_1, r_2, r_3) = (1 + \epsilon + \frac{1}{\epsilon}, 1 + \epsilon, 1)$ and $(v_1, v_2, v_3) = (\frac{1}{1 + \epsilon + \epsilon^2}, \frac{1}{\epsilon^2}, \frac{1}{\epsilon})$. The revenue-ordered assortments are $\{1\}$, $\{1, 2\}$ and $\{1, 2, 3\}$. Defining $f(\epsilon) = \epsilon^{-2}/(1 + (1 + \epsilon + \epsilon^2)^{-1} + \epsilon^{-1})$ and $g(\epsilon) = \epsilon^{-1}/(1 + (1 + \epsilon + \epsilon^2)^{-1} + \epsilon^{-2})$, the expected revenues from these assortments are $\mathsf{Rev}(\{1\}) = \frac{1}{\epsilon}\,(1+f(\epsilon)+g(\epsilon))$, $\mathsf{Rev}(\{1,2\}) = (\frac{1}{\epsilon}+\frac{1+\epsilon}{\epsilon^2})\,(1+g(\epsilon))$, $\mathsf{Rev}(\{1,2,3\}) = \frac{2}{\epsilon}+\frac{1+\epsilon}{\epsilon^2}$, whereas $\mathsf{Rev}(\{1,3\}) = \frac{2}{\epsilon}\,(1+f(\epsilon))$. Simple algebra yields $\lim_{\epsilon\to 0}\frac{\mathsf{Rev}(\{1\})}{\mathsf{Rev}(\{1,3\})} = \lim_{\epsilon\to 0}\frac{\mathsf{Rev}(\{1,2\})}{\mathsf{Rev}(\{1,3\})} = \lim_{\epsilon\to 0}\frac{\mathsf{Rev}(\{1,2,3\})}{\mathsf{Rev}(\{1,3\})} = \frac{1}{2}$. Thus, as ϵ gets arbitrarily close to zero, the best revenue-ordered assortment provides at most half of the optimal expected revenue.

Appendix E: Using Simulation to Estimate Expected Revenues

In our PTAS, we construct $O((n/\epsilon)^{O(1/\epsilon^2)})$ candidate assortments, evaluate the expected revenue of each candidate assortment and pick the best one. The running time to construct the candidate assortments as in Section 6.2 is independent of m, but the running time to compute the expected revenue from an assortment, which we denote by RevOps, depends exponentially on m. In this section, we use simulation to estimate the expected revenue from each assortment so that our PTAS is guaranteed to provide a $(1-6\,\epsilon)$-approximate solution with $1-\epsilon$ probability. The number of samples used in the simulation will depend polynomially on n and will be independent of m. We estimate the choice probability of any product within any assortment as follows. Using the random variables $\{U_i : i \in N\}$ and U_0 to denote the utilities of the products and no-purchase option, we generate L samples of these random variables. Letting $T = O((n/\epsilon)^{O(1/\epsilon^2)})$ to denote the number of candidate assortments and setting $\theta = (\frac{1}{n} + \max_{i\in N} v_i)/\min_{i\in N} v_i$ to measure the relative gap between largest and smallest preference weights, we choose the number of samples as $L = \lceil \frac{\theta^2 n^2}{2\,\epsilon^2}\log(\frac{2\,T\,n^2}{\epsilon})\rceil$. Letting $\{\widehat{U}_i^\ell : i \in N\}$ and $\widehat{U}_0^\ell$ for $\ell = 1,\ldots,L$ be our samples and adopting the notation at the end of Section 3, we use $\widehat{\omega}^\ell$ to capture the permutation corresponding to the ordering of the utilities in sample ℓ. Noting (4), we estimate the probability that a customer with rank cutoff k chooses product i within assortment S as

$$\widehat{\pi}_i^k(S) \;= \frac{1}{L}\sum_{\ell=1}^{L} \mathbf{1}\Big(\mathsf{rank}(i,\widehat{\omega}^\ell) < \mathsf{rank}(j,\widehat{\omega}^\ell) \quad \forall\, j \in S_{-i}\cup\{0\}\Big)\;\mathbf{1}\Big(\mathsf{rank}(i,\widehat{\omega}^\ell) \le k\Big). \tag{23}$$

In this case, we can interpret $\sum_{k\in M}\lambda^k\,\widehat{\pi}_i^k(S)$ as a sample average approximation to the choice probability given by (2) under our multinomial logit model with rank cutoffs.

In the next lemma, we use the Hoeffding inequality to bound the gap between the choice probabilities given in (2) and (23).

Lemma E.1 *Letting $\widehat{\pi}_i^k(S)$ be computed as in (23), for any assortment $S \subseteq N$ and product $i \in S$, we have $\mathbb{P}\{|\widehat{\pi}_i^k(S) - \pi_i^k(S)| \ge \frac{\epsilon}{\theta\,n}\} \le \frac{\epsilon}{T\,n^2}$.*

Proof: We can view the expression in (23) as the average of L independent samples of Bernoulli random variables. Furthermore, each of these Bernoulli random variables has expectation $\pi_i^k(S)$. By

Hoeffding inequality, if S_L is a sum of L independent Bernoulli random variables, then we have $\mathbb{P}\{|S_L - \mathbb{E}\{S_L\}| \geq t\} \leq 2\exp(-\frac{2t^2}{L})$. Therefore, we obtain

$$\begin{aligned}
&\mathbb{P}\Big\{|\widehat{\pi}_i^k(S) - \pi_i^k(S)| \geq \frac{\epsilon}{\theta\, n}\Big\} \\
&\quad = \mathbb{P}\left\{\left|\sum_{\ell=1}^{L} \mathbf{1}\Big(\mathsf{rank}(i,\widehat{\omega}^\ell) < \mathsf{rank}(j,\widehat{\omega}^\ell) \ \ \forall\, j \in S_{-i} \cup \{0\}\Big)\, \mathbf{1}\Big(\mathsf{rank}(i,\widehat{\omega}^\ell) \leq k\Big) - L\,\pi_i^k(S)\right| \geq \frac{L\,\epsilon}{\theta\, n}\right\} \\
&\quad \leq 2\exp\left(-\frac{2\,L\,\epsilon^2}{\theta^2\, n^2}\right) \leq \frac{\epsilon}{T\, n^2},
\end{aligned}$$

where the last inequality holds because we have $L \geq \frac{\theta^2\, n^2}{2\,\epsilon^2} \log(\frac{2\,T\,n^2}{\epsilon})$ by our choice of the number of samples at the beginning of this section. ■

Let $R(S) = \sum_{k\in M}\sum_{i\in S} \lambda^k\, r_i\, \pi_i^k(S)$ be the expected revenue from assortment S. Using (23), the sample average approximation to this expected revenue is $\widehat{R}(S) = \sum_{k\in M}\sum_{i\in S}\lambda^k\, r_i\, \widehat{\pi}_i^k(S)$. In the next lemma, we use the previous lemma along with the union bound to characterize the gap between the expected revenue from any assortment and its sample average approximation.

Lemma E.2 *Letting S^* be an optimal solution to problem (7), for any assortment S, the expected revenue from the assortment satisfies $\mathbb{P}\{|\widehat{R}(S) - R(S)| \leq \epsilon\, R(S^*)\} \geq 1 - \frac{\epsilon}{T}$.*

Proof: By Lemma E.1, for each $i \in S$ and $k \in M$, $\mathbb{P}\{|\widehat{\pi}_i^k(S) - \pi_i^k(S)| \geq \frac{\epsilon}{\theta\, n}\} \leq \frac{\epsilon}{T\, n^2}$. Since $|S| \leq n$ and $|M| \leq n$, the union bound yields $\mathbb{P}\{|\widehat{\pi}_i^k(S) - \pi_i^k(S)| \geq \frac{\epsilon}{\theta\, n}$ for some $i \in S$ and $k \in M\} \leq \frac{\epsilon}{T}$, in which case, we get $\mathbb{P}\{|\widehat{\pi}_i^k(S) - \pi_i^k(S)| \leq \frac{\epsilon}{\theta\, n}$ for all $i \in S$ and $k \in M\} \geq 1 - \frac{\epsilon}{T}$. On the other hand, if we have $|\widehat{\pi}_i^k(S) - \pi_i^k(S)| \leq \frac{\epsilon}{\theta\, n}$ for all $i \in S$ and $k \in M$, then we get

$$\begin{aligned}
|\widehat{R}(S) - R(S)| &\leq \sum_{k\in M}\sum_{i\in S}\lambda^k\, r_i\, |\widehat{\pi}_i^k(S) - \pi_i^k(S)| \leq \sum_{k\in M}\sum_{i\in S}\lambda^k\, r_i\, \frac{\epsilon}{\theta\, n} \\
&\overset{(a)}{=} \epsilon \sum_{k\in M}\sum_{i\in S}\lambda^k\, r_i\, \frac{\min_{j\in N} v_j}{1 + n\max_{j\in N} v_j} \leq \epsilon \sum_{k\in M}\sum_{i\in S}\lambda^k\, r_i\, \frac{v_i}{1+V(N)} \overset{(b)}{=} \epsilon \sum_{k\in M}\sum_{i\in N}\lambda^k\, r_i\, \pi_i^k(N) \overset{(c)}{\leq} \epsilon\, R(S^*),
\end{aligned}$$

where (a) uses the definition of θ, (b) holds because $\pi_i^k(N) = \frac{v_i}{1+V(N)}$ by Theorem 3.1 and (c) holds because the assortment N provides an objective value of $\sum_{k\in M}\sum_{i\in N}\lambda^k\, r_i\, \pi_i^k(N)$ for (7). ■

By the discussion in the proof of Theorem 5.1, there exists a $(1-4\,\epsilon)$-approximate solution among the candidate assortments constructed by our PTAS. Thus, letting S^* be an optimal solution to problem (7) and using $\mathcal{C} = \{A_t : t \in \mathcal{A}\}$ to capture the collection of candidate assortments, setting $\widetilde{S} = \arg\max_{S\in\mathcal{C}} R(S)$, we have $R(\widetilde{S}) \geq (1-4\,\epsilon)\, R(S^*)$. Finding $\widetilde{S}$ requires checking the expected revenue from each candidate assortment. Instead, we check the sample average approximation to the expected revenue, so we set $\widehat{S} = \arg\max_{S\in\mathcal{C}} \widehat{R}(S)$. Since there are T candidate assortments, using the union bound in Lemma E.2, we get $\mathbb{P}\{|\widehat{R}(S) - R(S)| \leq \epsilon R(S^*)$ for all $S \in \mathcal{C}\} \geq 1-\epsilon$, in which case, with probability at least $1-\epsilon$, we have $R(\widehat{S}) \geq \widehat{R}(\widehat{S}) - \epsilon\, R(S^*) \geq \widehat{R}(\widetilde{S}) - \epsilon\, R(S^*) \geq R(\widetilde{S}) - 2\epsilon\, R(S^*) \geq (1-4\,\epsilon)\, R(S^*) - 2\epsilon\, R(S^*)$, where the second inequality holds because the definition of $\widehat{S}$ yields $\widehat{R}(\widehat{S}) \geq \widehat{R}(\widetilde{S})$. Thus, the assortment $\widehat{S}$ satisfies $R(\widehat{S}) \geq (1-6\,\epsilon)\, R(S^*)$. Using $T = O((n/\epsilon)^{O(1/\epsilon^2)})$, the number of samples is $L = O(\frac{\theta^2 n^2}{\epsilon^2} \log(\frac{T\, n^2}{\epsilon})) = O(\frac{\theta^2 n^2}{\epsilon^4} \log(n/\epsilon))$.

Appendix F: 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 F.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 rank cutoff 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 rank cutoff of $k-1$, we show that the result holds for rank cutoff 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

$$\begin{aligned} B^k(\widetilde{S}, N \setminus F) \;&=\; 1 + \sum_{i \in (N \setminus F) \setminus \widetilde{S}} \frac{v_i}{1 + V((N \setminus F)_{-i})}\, B^{k-1}(\widetilde{S}, (N \setminus F)_{-i}) \\ &\leq 1 + \sum_{i \in (N \setminus F) \setminus \widetilde{S}} \frac{v_i}{1 + V((N \setminus F)_{-i})}\, B^{k-1}(\widehat{S}, (N \setminus K(F))_{-k(i)}) \\ &= 1 + \sum_{i \in (N \setminus F) \setminus \widetilde{S}} \frac{v_i}{1 + V(N) - V(F) - v_i}\, B^{k-1}(\widehat{S}, (N \setminus K(F))_{-k(i)}), \end{aligned} \tag{24}$$

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

$$\frac{v_i}{1 + V(N) - V(F) - v_i} \;\leq\; \frac{v_{k(i)}}{1 + V(N) - V(K(F)) - v_{k(i)}}. \tag{25}$$

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 (25), we can continue the chain of inequalities in (24) as

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

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 (24), completes the induction argument. In this case, using the inequality $B^k(\widetilde{S}, N \setminus F) \leq B^k(\widehat{S}, N \setminus K(F))$ with $F = \varnothing$ yields the desired result. ■

We use the next lemma in the proof of Lemma F.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 F.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 $v_i \leq v_{k(i)}$ 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 G: 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 $(j^\ell(1),\ldots,j^\ell(n^\ell))$, 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 $(i^\ell(1),\ldots,i^\ell(10))$ 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 $\{1,\ldots,10\}$, the preference list for customer type ℓ is $(i^\ell(1),\ldots,i^\ell(n^\ell))$. In the non-parametric choice model, customers of each type ℓ arrive 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 dataset. To use as validation and testing datasets, we also generate two other purchase histories, each including 1250 customers. We generate the training, validation and testing datasets independently. We fit the five choice models used in Section 7.1, which we refer to as RCO, SML, MML, SAC and MAC.

Comparing Out-of-Sample Log-Likelihoods: We use the same approach in Section 7.2 to compare the out-of-sample log-likelihoods of the fitted choice models. 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 RCO, SML, MML, SAC and MAC. In Table EC.1, we compare the out-of-sample log-likelihoods of the fitted choice models. We replicated our computational experiments for 10 ground choice models that we randomly generate. The ground choice models differ in the samples of $\{n^\ell:\ell=1,\ldots,5000\}$ 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 3. Our results indicate that if we have a large number of customers in the training datasets to estimate the parameters of the fitted choice models, then MML can catch up with RCO. As discussed in Section 7.2, MML can approximate any choice model that is based on random utility maximization

$\tau = 1000$

Grnd. Ch.	Out-of-Sample Log-Likelihoods RCO	SML	MML	SAC	MAC	Perc. Gap with RCO SML	MML	SAC	MAC
1	-1834.09	-1840.64	-1840.64	-1872.63	-1872.63	0.36	0.36	2.10	2.10
2	-1927.39	-1937.53	-1932.15	-1968.38	-1967.61	0.53	0.25	2.13	2.09
3	-1908.67	-1909.80	-1909.80	-1950.17	-1950.17	0.06	0.06	2.17	2.17
4	-1848.16	-1859.12	-1859.12	-1904.29	-1904.29	0.59	0.59	3.04	3.04
5	-1892.07	-1894.01	-1894.01	-1919.29	-1919.29	0.10	0.10	1.44	1.44
6	-1913.07	-1916.44	-1918.32	-1964.98	-1965.32	0.18	0.27	2.71	2.73
7	-1919.34	-1927.84	-1924.59	-1969.15	-1969.15	0.44	0.27	2.59	2.59
8	-1891.25	-1889.84	-1896.00	-1907.49	-1907.49	-0.07	0.25	0.86	0.86
9	-1889.58	-1894.29	-1885.36	-1919.50	-1909.19	0.25	-0.22	1.58	1.04
10	-1922.99	-1923.24	-1927.64	-1944.06	-1947.74	0.01	0.24	1.10	1.29
Avg.						0.24	0.22	1.97	1.93

$\tau = 1750$

Grnd. Ch.	Out-of-Sample Log-Likelihoods RCO	SML	MML	SAC	MAC	Perc. Gap with RCO SML	MML	SAC	MAC
1	-1837.59	-1843.25	-1843.25	-1880.29	-1880.29	0.31	0.31	2.32	2.32
2	-1926.58	-1935.66	-1926.86	-1967.52	-1963.86	0.47	0.01	2.12	1.93
3	-1903.79	-1904.98	-1904.95	-1941.90	-1941.90	0.06	0.06	2.00	2.00
4	-1848.05	-1856.73	-1860.98	-1904.91	-1904.91	0.47	0.70	3.08	3.08
5	-1891.68	-1893.66	-1893.66	-1914.94	-1914.94	0.10	0.10	1.23	1.23
6	-1906.02	-1907.84	-1906.51	-1955.84	-1958.38	0.10	0.03	2.61	2.75
7	-1916.36	-1923.56	-1933.51	-1954.31	-1951.88	0.38	0.89	1.98	1.85
8	-1886.31	-1889.57	-1889.57	-1912.94	-1912.94	0.17	0.17	1.41	1.41
9	-1889.03	-1893.78	-1893.59	-1925.89	-1918.67	0.25	0.24	1.95	1.57
10	-1924.98	-1924.54	-1932.45	-1945.61	-1954.63	-0.02	0.39	1.07	1.54
Avg.						0.23	0.29	1.98	1.97

$\tau = 2500$

Grnd. Ch.	Out-of-Sample Log-Likelihoods RCO	SML	MML	SAC	MAC	Perc. Gap with RCO SML	MML	SAC	MAC
1	-1835.09	-1840.69	-1840.69	-1877.27	-1877.27	0.31	0.31	2.30	2.30
2	-1924.56	-1932.61	-1930.71	-1966.48	-1962.96	0.42	0.32	2.18	2.00
3	-1904.37	-1905.37	-1904.52	-1938.31	-1938.31	0.05	0.01	1.78	1.78
4	-1843.63	-1853.59	-1845.50	-1907.69	-1906.94	0.54	0.10	3.47	3.43
5	-1890.50	-1892.97	-1890.61	-1922.95	-1918.64	0.13	0.01	1.72	1.49
6	-1902.56	-1905.67	-1890.21	-1945.08	-1940.41	0.16	-0.65	2.23	1.99
7	-1916.54	-1924.01	-1924.67	-1963.19	-1962.05	0.39	0.42	2.43	2.37
8	-1882.90	-1885.79	-1875.24	-1909.40	-1909.40	0.15	-0.41	1.41	1.41
9	-1883.90	-1888.98	-1875.08	-1926.91	-1921.07	0.27	-0.47	2.28	1.97
10	-1924.58	-1924.15	-1930.16	-1958.98	-1967.08	-0.02	0.29	1.79	2.21
Avg.						0.24	-0.01	2.16	2.10

Table EC.1 Comparison of the out-of-sample log-likelihoods of the fitted choice models.

arbitrarily well, but it needs a large number of customers in the training dataset to avoid overfitting. When the number of customers in the training datasets is on the small side, RCO has an edge over MML. Overall a large majority of the ground choice models, the out-of-sample log-likelihoods of the fitted RCO are better than those of the fitted SML, SAC and MAC. Even when we have the largest number of customers in the training datasets, the out-of-sample log-likelihoods of RCO are better than those of MML for seven ground choice models.

Comparing Expected Revenue Performance: We use the same approach in Section 7.3 to compare the ability of the fitted choice models to pick profitable assortments. In Table EC.2, we compare the expected revenues from the assortments obtained by RCO, SML, MML, SAC and MAC. The layout of this table is identical to that of Table 4. Our results indicate that RCO performs slight but noticeable improvements over SML and MML. The gaps in the expected revenues obtained by RCO and MML decrease as we have more customers in the training datasets, but RCO maintains its edge over MML for a majority of the ground choice models. Comparing RCO with SAC and MAC, the expected revenues obtained by RCO are significantly better than those obtained by SAC and MAC. To ensure that our results are robust, 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, RCO improves the out-of-sample log-likelihoods of SML, MML, SAC and MAC, respectively, by 0.15%, 0.10%, 2.39% and 2.34% on average. Over the 300 combinations, in 209, 210, 300 and 300 combinations, the out-of-sample log-likelihoods of RCO are, respectively, better than those of SML, MML, SAC and MAC. Although the improvements in out-of-sample log-likelihoods provided by RCO over SML and MML are slight, they are consistent. Considering the expected revenue performance, RCO improves the expected revenues obtained by SML, MML, SAC and MAC, respectively, by 0.60%, 0.54%, 5.54% and 5.45% on average. In 77%, 78%, 86% and 85% of the product revenue samples, the expected revenue performance RCO is, respectively, at least as good as that of SML, MML, SAC and MAC.

Appendix H: Performance of the Approximation Scheme

We test the performance of the PTAS given in 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 (7). At the end of this section, 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. We classify each product along two dimensions. First, a product can have high or low revenue. Second, a product can have high or low preference weight. Thus, we have four classes of products, corresponding the two choices along two dimensions. 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

$\tau = 1000$

Grnd. Ch.	Comp. with SML Perc. Gap	RCO ≻ SML	SML ≻ RCO	Comp. with MML Perc. Gap	RCO ≻ MML	MML ≻ RCO	Comp. with SAC Perc. Gap	RCO ≻ SAC	SAC ≻ RCO	Comp. with MAC Perc. Gap	RCO ≻ MAC	MAC ≻ RCO
1	0.11	42	34	0.11	42	34	4.66	83	17	4.66	83	17
2	0.79	53	23	0.95	52	23	5.81	89	11	5.86	90	10
3	0.49	40	19	0.49	40	19	5.53	86	14	5.53	86	14
4	0.52	49	25	0.52	49	25	5.17	82	18	5.17	82	18
5	0.68	45	27	0.68	45	27	5.01	85	15	5.01	85	15
6	1.42	52	13	3.88	64	14	9.88	87	13	9.02	83	17
7	1.03	49	16	0.73	44	15	6.02	85	15	6.02	85	15
8	-0.55	37	47	-0.69	34	43	6.77	83	17	6.77	83	17
9	-0.07	34	32	-0.10	29	25	6.22	83	17	6.42	85	15
10	0.86	44	18	0.95	42	18	5.19	88	12	5.09	86	14
Avg.	0.53	45	25	0.75	44	24	6.03	85	15	5.95	85	15

$\tau = 1750$

Grnd. Ch.	Comp. with SML Perc. Gap	RCO ≻ SML	SML ≻ RCO	Comp. with MML Perc. Gap	RCO ≻ MML	MML ≻ RCO	Comp. with SAC Perc. Gap	RCO ≻ SAC	SAC ≻ RCO	Comp. with MAC Perc. Gap	RCO ≻ MAC	MAC ≻ RCO
1	0.33	45	32	0.33	45	32	5.21	87	13	5.21	87	13
2	0.77	51	19	0.17	39	19	5.64	89	11	5.46	86	14
3	0.45	39	20	0.30	37	19	5.85	88	12	5.85	88	12
4	0.66	43	18	0.65	35	18	4.96	78	22	4.96	78	22
5	1.00	43	19	1.00	43	19	4.86	88	12	4.86	88	12
6	1.17	46	14	1.05	45	12	5.19	84	16	5.28	83	17
7	0.79	41	17	0.90	43	16	5.62	79	21	5.32	79	21
8	0.40	41	27	0.40	41	27	7.53	88	12	7.53	88	12
9	-0.03	34	31	0.05	30	28	5.54	89	11	5.14	89	11
10	0.78	44	20	0.17	30	27	4.90	90	10	4.53	88	12
Avg.	0.63	43	22	0.50	39	22	5.53	86	14	5.41	85	15

$\tau = 2500$

Grnd. Ch.	Comp. with SML Perc. Gap	RCO ≻ SML	SML ≻ RCO	Comp. with MML Perc. Gap	RCO ≻ MML	MML ≻ RCO	Comp. with SAC Perc. Gap	RCO ≻ SAC	SAC ≻ RCO	Comp. with MAC Perc. Gap	RCO ≻ MAC	MAC ≻ RCO
1	0.41	41	26	0.41	41	26	5.90	88	12	5.90	88	12
2	0.86	48	17	0.37	38	17	5.12	87	13	4.71	86	14
3	0.53	41	19	0.27	35	18	5.12	91	9	5.12	91	9
4	0.66	45	19	0.20	36	20	7.26	80	20	7.09	80	20
5	1.00	45	20	1.19	43	17	5.10	85	15	5.63	90	10
6	1.16	48	17	0.65	51	19	5.88	79	21	5.40	76	24
7	0.72	42	18	1.05	41	17	6.17	86	14	5.74	84	16
8	0.60	47	26	0.21	38	21	7.76	88	12	7.76	88	12
9	-0.21	33	35	-0.53	18	32	5.90	86	14	6.81	91	9
10	0.75	42	20	0.05	30	23	4.71	85	15	4.30	84	16
Avg.	0.65	43	22	0.39	37	21	5.89	86	15	5.85	86	14

Table EC.2 Comparison of the expected revenues obtained by using the fitted choice models.

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 rank cutoff 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 EC.3. 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 second 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 EC.3 indicate that our PTAS performs remarkably well. The average optimality gap is 0.26%. In more than 90% of the problem instances, the optimality gap is no larger than 0.72%, even considering the fact that 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. Over all problem instances, the average CPU time for our PTAS is 11.97 seconds. Recall that the rank cutoffs of all customers is $m = 2$ in our test problems. We carried out a limited set of computational experiments with $m \in \{3, 4\}$ as well. In these computational experiments, the average CPU time for our PTAS is, respectively, 18.20 and 109.72 seconds, when we have $m = 3$ and $m = 4$.

Upper Bound on the Optimal Expected Revenue: We give an efficient approach to compute an upper bound on the optimal objective value of problem (7). 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 (7) for notational brevity, the objective function of problem (7) is $W(S)\big[\lambda^1 + \lambda^2 (1 + \sum_{i \in N \setminus S} \theta_i)\big]$. Noting that $\lambda^1 + \lambda^2 = 1$, problem (7) becomes

$$\max_{S \subseteq N} \left\{ W(S) \Big[1 + \lambda^2 \sum_{i \in N \setminus S} \theta_i \Big] \right\}. \tag{26}$$

To obtain an upper bound on the optimal objective value of the problem above, 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 $0 = \nu_0 \leq \nu_1 \leq \ldots \leq \nu_{K-1} \leq \nu_K = \overline{\Theta}$. 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 (26) to take a large value, both $W(S)$ and $\sum_{i \in N \setminus S} \theta_i$ should take large values. Thus, we formulate a knapsack problem to

Param. (θ, γ)	Avg.	Std. Dev.	90th Perc.	Max.
(40, 1)	0.16	0.49	0.38	2.71
(40, 10)	0.17	0.49	0.40	2.72
(40, 50)	0.18	0.48	0.37	2.59
Avg.	0.17	0.49	0.39	2.67

Param. (θ, γ)	Avg.	Std. Dev.	90th Perc.	Max.
(50, 1)	0.23	0.22	0.51	0.86
(50, 10)	0.23	0.23	0.53	0.86
(50, 50)	0.27	0.24	0.58	0.95
Avg.	0.24	0.23	0.54	0.89

Param. (θ, γ)	Avg.	Std. Dev.	90th Perc.	Max.
(60, 1)	0.34	0.25	0.69	1.15
(60, 10)	0.35	0.25	0.69	1.18
(60, 50)	0.37	0.25	0.72	1.05
Avg.	0.35	0.25	0.70	1.13

Table EC.3 Performance of our PTAS.

maximize $W(S)$ while making sure that $\sum_{i \in N \setminus S} \theta_i$ is also above a certain threshold. For each $k = 1, \ldots, K$, using the decision variables $\boldsymbol{x} = (x_1, \ldots, x_n)$, consider the problem

$$Z_k \; = \; \max_{\boldsymbol{x} \in [0,1]^n} \left\{ \sum_{i \in N} r_i \, v_i \, x_i \; : \; \sum_{i \in N} \theta_i \, (1 - x_i) \geq \nu_{k-1} \right\}. \tag{27}$$

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 $\sum_{i \in N \setminus S} \theta_i \geq \nu_{k-1}$. 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 (26) by solving the problem above for each $k = 1, \ldots, K$.

Theorem H.1 *Noting that Z_k is the optimal objective value of problem* (27)*, letting* Rev^* *be the optimal objective value of problem* (26)*, we have*

$$\max_{k=1,\ldots,K} \left\{ (1 + \lambda^2 \, \nu_k) \, Z_k \right\} \; \geq \; \mathsf{Rev}^*.$$

Proof: Using S^* to denote an optimal solution to (26), let $\widehat{\boldsymbol{x}} = (\widehat{x}_1, \ldots, \widehat{x}_n) \in \{0,1\}^n$ be 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 (27), 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

$$\max_{k=1,\ldots,K} \left\{ (1 + \lambda^2 \, \nu_k) \, Z_k \right\} \; \geq \; (1 + \lambda^2 \, \nu_\ell) \, Z_\ell \; \geq \; \Big(1 + \lambda^2 \sum_{i \in N \setminus S^*} \theta_i \Big) W(S^*) \; = \; \mathsf{Rev}^*,$$

where the second inequality holds because we have the inequalities $Z_\ell \geq W(S^*)$ and $\nu_\ell \geq \sum_{i \in N \setminus S^*} \theta_i$ as discussed earlier in the proof. ∎

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.