Assortment Optimization under Variants of the Nested Logit Model

James M. Davis
School of Operations Research and Information Engineering,
Cornell University, Ithaca, New York 14853, USA
jmd388@cornell.edu

Guillermo Gallego
Department of Industrial Engineering and Operations Research
Columbia University, New York, New York 10027, USA
gmg2@columbia.edu

Huseyin Topaloglu
School of Operations Research and Information Engineering,
Cornell University, Ithaca, New York 14853, USA
topaloglu@orie.cornell.edu

December 11, 2013

Abstract

We study a class of assortment optimization problems where customers choose among the offered products according to the nested logit model. There is a fixed revenue associated with each product. The objective is to find an assortment of products to offer so as to maximize the expected revenue per customer. We show that the problem is polynomially solvable when the nest dissimilarity parameters of the choice model are less than one and the customers always make a purchase within the selected nest. Relaxing either of these assumptions renders the problem NP-hard. To deal with the NP-hard cases, we develop parsimonious collections of candidate assortments with worst-case performance guarantees. We also formulate a convex program whose optimal objective value is an upper bound on the optimal expected revenue. Thus, we can compare the expected revenue provided by an assortment with the upper bound on the optimal expected revenue to get a feel for the optimality gap of the assortment. By using this approach, our computational experiments test the performance of the parsimonious collections of candidate assortments that we develop.

Discrete choice models have been used for nearly half a century to understand how customers select among a group of products that vary in terms of price and quality. Of particular interest is how demand for the different products changes as the offer set changes in composition, quality or price. To advance this agenda, researchers have developed discrete choice models based on axioms as in Luce (1959), resulting in the basic attraction model, and based on random utility theory as in McFadden (1974), resulting in the celebrated multinomial logit model. Important extensions include the nested attraction model of which the nested logit model, introduced by Williams (1977), is a special case. Justifications and extensions for the nested logit model are provided in McFadden (1980) and Borsch-Supan (1990). Under the nested logit model, customers first select a nest, and then, a product within the selected nest. The nested logit model was developed primarily to avoid the independence of irrelevant alternatives property suffered by the multinomial logit model; see Ben-Akiva and Lerman (1997).

In this paper, we study a class of assortment optimization problems where the choices of the customers are governed by the nested logit model. Under this model, customers first select a nest, and then, a product within the nest. We assume that there is a revenue associated with each product and the objective is to find a set of products, or an assortment, to offer that maximizes the expected revenue per customer. This assortment optimization problem is combinatorial in nature and the number of possible assortments can be very large, particularly when there are many potential products to offer. In airline and hotel revenue management settings, for example, the number of products can easily exceed 30 or 40, yielding 2^{30} or 2^{40} possible assortments. Therefore, it is important to classify when the problem is polynomially solvable. When not, it is important to find solution methods with worst-case performance guarantees. One of our main contributions is to classify the complexity of the assortment problem for nested attraction models. We do this along two dimensions. The first dimension is the magnitude of the nest dissimilarity parameters, which characterize the degree of dissimilarity of the products within a nest. The second dimension is the presence or absence of the no purchase alternative within a nest. We show that the only polynomially solvable case is when the nest dissimilarity parameters are less than one and the no purchase alternative is only available at the time of selecting a nest. If the nest dissimilarity parameters exceed one or the customers can choose a no purchase option after selecting a nest, then we show that the problem is NP-hard. For the NP-hard cases, we develop tractable methods to obtain assortments with worst-case performance guarantees.

Related Literature. Research on pricing in the context of the multinomial logit and nested logit models has been fairly active. In that setting, the problem is to choose a set of prices for the products, where the prices of all products jointly determine the probability that a customer purchases a particular product. The objective is to maximize the expected revenue per customer. For the pricing problem, Hanson and Martin (1996) notice that the expected revenue function fails to be concave in prices for the multinomial logit model, but significant progress was made by formulating the pricing problem in terms of market shares, as this results in a concave expected revenue function; see Song and Xue (2007) and Dong et al. (2009). Li and Huh (2011) extend the concavity result to the nested logit model by assuming that the price sensitivities of the products are constant within each nest and the nest dissimilarity parameters are all between zero and one. They show that the expected revenue maximization problem can be reduced to optimizing over a single variable. Gallego and Wang (2011)

relax both of the assumptions in Li and Huh (2011) and extend the analysis to more general nested attraction models. The key result is that the optimal prices add two terms to the unit costs, where the first term is the inverse of the price sensitivities of the products and the second term is a nest-dependent constant. This result implies that products with the same price sensitivity in a nest have the same markup, irrespective of their quality.

In many situations, prices of the products are fixed and are not in the control of the decision maker, at least not in the short run. For example, this situation holds in the context of revenue management, where a menu of fares is designed to allow the same capacity to be sold at different prices. This is done by differentiating the products by time-of-purchase, traveling restrictions, and the inclusion or exclusion of ancillary services such as luggage handling, mileage accrual and advance seat selection. Revenue managers must dynamically decide which set of products to offer depending on the state of the system, which includes the time-to-departure and the remaining inventory. In retailing, pricing decisions are often centralized, and hence, are fixed in advance, while assortment decisions can be made at the local level. The assortment problem is particularly important at the design stage, where several products could be built based on different design features and prices.

Assortment optimization is an active area of research, and our review of assortment optimization problems focuses on papers that use attraction-based choice models, such as multinomial or nested logit model. We refer the reader to Kok et al. (2008) and Farias et al. (2011) for assortment optimization under other choice models. The paper by van Ryzin and Mahajan (1999) considers an assortment optimization problem where the products both generate revenue and result in operational cost. The objective is to balance the revenue benefit of offering product variety with the overhead cost of carrying a large number of products. Cachon et al. (2005) extend this work to include product search costs, where a customer may find an acceptable product in one store, but still not purchase hoping that another store may carry a more desirable product. Mahajan and van Ryzin (2001) develop algorithms to set stocking levels for multiple substitutable products.

If the customers choose according to the multinomial logit model, then the assortment optimization problem can be solved efficiently as Talluri and van Ryzin (2004) show that the optimal assortment includes a certain number of products with the largest revenues. We refer to assortments that include a certain number of products with the largest revenues as nested-by-revenue assortments. The problem becomes more complicated when more general choice models are considered. Rusmevichientong, Shmoys and Topaloglu (2010) study the assortment problem under the mixed multinomial logit model, where there are multiple customer types and customers of different types choose according to different multinomial logit models. They show that the assortment optimization problem is NP-hard in the weak sense even with two customer types and provide a performance guarantee for nested-by-revenue assortments. Bront et al. (2009) show that the same problem is NP-hard in the strong sense and Mendez-Diaz et al. (2010) give a branch-and-cut algorithm to find the optimal assortment. Rusmevichientong and Topaloglu (2011) study the robust assortment problem under the multinomial logit model when some of the parameters of the choice model are not known. Rusmevichientong, Shen and Shmoys (2010) consider constraints on the size of the offered assortment when customers choose according to

the multinomial logit model. Kok and Xu (2011) consider joint assortment optimization and pricing problems under the nested logit model, where both the set of products offered and their corresponding prices are decision variables. They work with two nest structures. In the first nest structure, customers first select a brand, and then, a product type within the selected brand, whereas in the second nest structure, customers first select a product type, and then, a brand for the selected product type. The authors characterize the structure of the optimal solution, but the problem becomes difficult when the number of brands is large or the product prices are fixed.

In this paper, we use the nested logit model to capture customer choices. There are several desirable aspects of this choice model. To begin with, the nested logit model alleviates the independence of irrelevant alternatives property suffered by the multinomial logit model; see Ben-Akiva and Lerman (1997). In particular, if a product is added to the offered assortment, then the multinomial logit model predicts that the market share of each product in the offered assortment decreases by the same relative amount, which clearly should not occur when different products cannibalize on each other to different extents. Furthermore, it is possible to show that the nested logit model is compatible with a random utility-based choice model, where customers associate random utilities with the products and with the option of not making a purchase and they follow the option providing the largest utility. This feature gives some behavioral justification to the nested logit model; see Borsch-Supan (1990). Finally, the nested logit model allows correlations between the utilities of the products, capturing the fact that the way a customer evaluates a certain product may help us predict how this customer would evaluate other similar products.

Beside its desirable aspects, the standard form of the nested logit model has some limitations. This choice model works with a fixed nest structure, where customers first select a nest, and then, a product within the selected nest. For example, nests may correspond to different airlines and products within a nest may correspond to different cabin classes offered by an airline. However, not all customers follow the same nest structure. Some customers may select an airline first, and then, a cabin class offered by this airline, whereas some customers may select a cabin class first, and then, an airline that offers this cabin class. Several extensions to the nested logit model are designed to alleviate this concern; see Train (2003). In the mixed nested logit model, there are customers of multiple types and customers of different types choose according to different nested logit models, possibly with different nest structures. In paired combinatorial and cross nested logit models, a product may appear in multiple nests. In this paper, we use the standard form of the nested logit model with a fixed nest structure. Bront et al. (2009) show that the assortment problem is NP-hard under the mixed multinomial logit model and this result carries over to the mixed nested logit model. Kok and Xu (2011) characterize the optimal assortment under a mixed nested logit model with identical product prices.

Our Contributions. To our knowledge, there is no work on assortment optimization under the nested logit model that can deal with a large number of nests. We consider four cases to characterize the situations where our assortment optimization problem can be solved exactly or approximately. The first case considers the situation where the dissimilarity parameters of the nests are less than one and customers always make a purchase within the selected nest. This situation conforms to the standard

4

form of the nested logit model; see Borsch-Supan (1990). For this case, we show that it is optimal to offer a nested-by-revenue assortment within each nest, but this result does not immediately imply that the problem is polynomially solvable since there are exponentially many combinations of nested-by-revenue assortments we can choose for the different nests. We deal with this difficulty by giving a linear program that finds the best combination of nested-by-revenue assortments for each nest. Thus, the problem is tractable under the standard form of the nested logit model.

We show that the assortment optimization problem is NP-hard in all of the remaining cases. These cases correspond to nonstandard versions of the nested logit model, but we justify the practical importance of these cases. For the NP-hard cases, we give parsimonious collections of assortments such that if we focus only on these assortments, then we obtain a solution with a certain worst-case performance guarantee. In particular, the second case we consider focuses on the situation where the dissimilarity parameters can take on any value, but the customers always purchase a product within the selected nest. For this case, we show that if we focus only on nested-by-revenue assortments, then assuming that the revenues of the products in the same nest differ by at most a factor of ρ and the attractiveness parameters of the products in the same nest differ by at most a factor of κ, the expected revenue from the best nested-by-revenue assortment cannot deviate from the optimal expected revenue by more than a factor of $\min\{\rho, 2\kappa\}$. Therefore, we can expect the nested-by-revenue assortments to perform well when the revenues or the attractiveness of the products within a particular nest are not too different from each other. It is important to emphasize that this result allows the revenues or the attractiveness of the products to differ arbitrarily when the products are in different nests.

In the third case, we consider the situation where the dissimilarity parameters of the nests are less than one, but customers may leave a chosen nest without purchasing. For this case, we construct a small collection of assortments such that the best assortment within this collection provides an expected revenue that deviates from the optimal expected revenue by no more than a factor of two. Finally, the fourth case considers the most general problem instances with no restrictions on the dissimilarity parameters of the nests and the no purchase behavior. For this case, we give a collection of assortments such that the best assortment within this collection has a worst-case performance guarantee of $2\,\kappa$, where κ is as defined in the paragraph above. Furthermore, we exploit the connections of our assortment problem to the partition problem to give a pseudo-polynomial-time algorithm for the most general instances of our problem. Finally, using $\bar{\gamma}$ to denote the largest dissimilarity parameter for the nests, for any given $\delta > 1$, we give a collection of assortments that provides a worst-case performance guarantee of $\delta^{2\max\{\bar{\gamma},1\}+1}$, but the work required to obtain this collection increases as δ gets close to one. Thus, this result is akin to a polynomial-time approximation scheme when $\bar{\gamma}$ is fixed. Table 1 summarizes the four cases and indicates which sections in the paper include each one of these cases.

In addition to the worst-case performance guarantees, we formulate a convex program that yields an upper bound on the optimal expected revenue. By comparing the upper bound on the optimal expected revenue with the expected revenue provided by an assortment, we bound the optimality gap of the assortment we obtain for a particular problem instance. We use this approach in our computational experiments to test the performance of the solutions we obtain by focusing on the

		No Purchase Behavior	
		Always Make a Purchase within Selected Nest	May Leave without Purchasing Even After Selecting a Nest
Nest Dissimilarity Parameters	Between 0 and 1	Case 1: Section 3 Polynomially-solvable by offering a nested-by-revenue assortment in each nest	Case 3: Section 5 Worst-case performance guarantee of two
	Free	Case 2: Section 4 Worst-case performance guarantee of $\min\{\rho, 2\kappa\}$	Case 4: Section 6 Worst-case performance guarantee of 2κ, pseudo-polynomial-time algorithm, polynomial-time approximation scheme for fixed $\bar{\gamma}$

Table 1: Four cases considered in the paper. The ratio between the largest and the smallest product revenues in a nest is bounded by ρ, the ratio between the largest and the smallest attractiveness parameters in a nest is bounded by κ and the dissimilarity parameters are bounded by $\bar{\gamma}$.

collections of assortments mentioned above. In this way, we characterize the problem parameters that affect the solution quality and empirically demonstrate that the performance of the assortments we propose follows the trends predicted by their worst-case performance guarantees.

Organization. In Section 1, we formulate the assortment optimization problem. In Section 2, we give an alternative linear programming formulation of this problem. The number of constraints in this formulation grows exponentially with the number of products within each nest, but we build on the linear program to come up with a general approximation framework for the assortment optimization problem. The following four sections focus on the four cases mentioned above. In Section 3, we assume that the dissimilarity parameters of the nests are less than one and the customers always make a purchase among the products of the selected nest. We show that it is optimal to offer a nested-by-revenue assortment within each nest. In Section 4, we relax the assumption that the dissimilarity parameters are less than one, whereas in Section 5, we allow customers to leave a selected nest without purchasing. In these sections, we show that the assortment optimization problem becomes NP-hard under either of these relaxations and we provide worst-case performance guarantees for certain collections of assortments. In Section 6, we focus on the most general instances of the assortment optimization problem with no restrictions on the dissimilarity parameters of the nests and the no purchase behavior of the customers. Similar to the previous two sections, we give a collection of assortments with a certain worst-case performance guarantee. Also, we give a pseudo-polynomial-time algorithm and a polynomial-time approximation scheme. The performance guarantees we give in Sections 4, 5 and 6 are based on the general approximation framework we develop in Section 2. In Section 7, we formulate a convex program that provides an upper bound on the optimal expected revenue. In Section 8, we give computational experiments. In Section 9, we conclude.

1 Problem Formulation

In this section, we describe the nested logit model that we use to model the customer choice, and then, formulate our assortment optimization problem. There are m nests indexed by $M = \{1, \ldots, m\}$. Depending on the application setting, each nest may represent a different category of products, a different sales channel or a different retail store. There are n products that we can offer in

each nest. We index the products in each nest by $N = \{1, \ldots, n\}$. We use r_{ij} to denote the revenue associated with product j in nest i. Without loss of generality, we assume that the products in each nest are ordered such that $r_{i1} \geq r_{i2} \geq \ldots \geq r_{in}$ for all $i \in M$. We let v_{ij} be the preference weight of product j in nest i and v_{i0} be the preference weight of the no purchase option in nest i. We use $V_i(S_i)$ to denote the total preference weight of all available options when we offer the assortment $S_i \subset N$ in nest i. In other words, we have $V_i(S_i) = v_{i0} + \sum_{j \in S_i} v_{ij}$. Under the nested logit model, given that a customer decides to purchase a product in nest i, if we offer the assortment S_i in this nest, then the probability that the customer purchases product $j \in S_i$ is given by

$$P_{ij}(S_i) = \frac{v_{ij}}{v_{i0} + \sum_{k \in S_i} v_{ik}} = \frac{v_{ij}}{V_i(S_i)}.$$

We observe that the assumption that each nest includes the same number of products is without loss of generality because if some nest i includes fewer than n products, then we can include additional products j in this nest with preference weight $v_{ij} = 0$ and these products would never be purchased. Also, it is possible to have $v_{i0} = 0$ for some nest i or even all nests, in which case, given that a customer selects nest i, he never leaves without purchasing anything. We note that if $v_{i0} = 0$, then the expression for $P_{ij}(\emptyset)$ can evaluate to $0/0$, but this does not pose any difficulty since if we offer the empty assortment in nest i and $v_{i0} = 0$, then a customer would never decide to make a purchase in nest i and the value of $P_{ij}(\emptyset)$ becomes irrelevant.

Each nest i has a parameter $\gamma_i \geq 0$ associated with it that characterizes the degree of the dissimilarity of the products in the nest. Furthermore, we use v_0 to denote the preference weight for the option of not choosing any of the nests. If we offer the assortment $(S_1, \ldots, S_m)$ over all nests, then a customer chooses nest i with probability

$$Q_i(S_1, \ldots, S_m) = \frac{V_i(S_i)^{\gamma_i}}{v_0 + \sum_{l \in M} V_l(S_l)^{\gamma_l}}. \tag{1}$$

We note that γ_i serves the purpose of dampening or magnifying the total preference weight of the available options within nest i. We allow having $v_0 = 0$. We note that if we have $v_0 = 0$, we offer the empty assortment in all nests and $v_{i0} = 0$ as well for all $i \in M$, then the expression for $Q_i(\emptyset, \ldots, \emptyset)$ evaluates to $0/0$, but this does not create any complication since if we offer the empty assortment in all nests, then we trivially make an expected revenue of zero. Therefore, all of our development in the paper applies to the case where some or all of the preference weights v_0 and $(v_{10}, \ldots, v_{m0})$ associated with the no purchase options are equal to zero.

If we offer the assortment S_i in nest i, then we can write the expected revenue we obtain from this nest as

$$R_i(S_i) = \sum_{j \in S_i} r_{ij} P_{ij}(S_i) = \frac{\sum_{j \in S_i} r_{ij} v_{ij}}{V_i(S_i)},$$

with the interpretation that $R_i(\emptyset) = 0$. Therefore, if we offer the assortment $(S_1, \ldots, S_m)$ over all nests with $S_i \subset N$ for all $i \in M$, then we obtain an expected revenue of

$$\Pi(S_1, \ldots, S_m) = \sum_{i \in M} Q_i(S_1, \ldots, S_m) R_i(S_i).$$

Our goal is to choose an assortment $(S_1, \ldots, S_m)$ that maximizes the expected revenue over all nests, yielding the assortment optimization problem

$$Z^* = \max_{(S_1, \ldots, S_m): S_i \subset N, \, i \in M} \Pi(S_1, \ldots, S_m). \tag{2}$$

Throughout this paper, we classify the instances of the assortment optimization problem in (2) along two dimensions. The first dimension is based on the values of the dissimilarity parameters $(\gamma_1, \ldots, \gamma_m)$ of the nests. Along this dimension, we separately consider the two cases where (i) we have $\gamma_i \leq 1$ for all $i \in M$, and (ii) there are no restrictions on the dissimilarity parameters. The second dimension of classification is based on the values of the preference weights $(v_{10}, \ldots, v_{m0})$ of the no purchase options within the nests. Along this dimension, we separately consider the two cases where (i) we have $v_{i0} = 0$ for all $i \in M$, and (ii) there are no restrictions on the preference weights of the no purchase options. Since there are two cases to consider along each one of the two dimensions, we study the assortment optimization problem in (2) under four cases. It turns out that while the assortment optimization problem in (2) is polynomially solvable when $\gamma_i \leq 1$ and $v_{i0} = 0$ for all $i \in M$, lifting any one of these restrictions renders the problem NP-hard and we resort to approximation methods.

Along the first dimension, the case with $\gamma_i \leq 1$ for all $i \in M$ corresponds to the standard form of the nested logit model studied by McFadden (1974) and McFadden (1981). In these papers, the author notes that having $\gamma_i \leq 1$ for all $i \in M$ implies that the nested logit model is always compatible with a random utility-based choice model, irrespective of the values of the preference weights of the products and the preference weights of the no purchase options. In particular, McFadden (1981) postulates three assumptions that need to be satisfied for the nested logit model to be compatible with a random utility-based choice model and shows that these assumptions are always satisfied when the nest dissimilarity parameters do not exceed one. Borsch-Supan (1990) revisits the compatibility of the nested logit model with utility maximization principle and shows that even if we do not have $\gamma_i \leq 1$ for all $i \in M$, the three assumptions postulated by McFadden (1981) can be satisfied for certain values of the preference weights $\{v_{ij} : i \in M, \, j \in N\}$, $\{v_{i0} : i \in M\}$ and v_0. Therefore, the nested logit model can still be compatible with a random utility-based choice model when we have $\gamma_i \geq 1$ for some $i \in M$. Similarly, Train (2003) notes that the preference weight v_{ij} of product j in nest i has the form $e^{\bar{u}_{ij}/\gamma_i}$, where $\bar{u}_{ij}$ is the mean utility of product j in nest i captured through a linear combination of its attributes, such as price, quality and ease of use. He argues that even if the dissimilarity parameters exceed one, the nested logit model can be compatible with utility maximization principle for certain values of the mean utilities $\{\bar{u}_{ij} : i \in M, \, j \in N\}$. There are a number of empirical studies that fit the nested logit model to customer choice data and end up with estimates for the nest dissimilarity parameters exceeding one; see Train et al. (1987), Train et al. (1989), Lee (1999), Tiwari and Hasegawa (2004) and Yates and Mackay (2006) applications in telephone service and housing choice. Train et al. (1987) interpret a nest dissimilarity parameter exceeding one as a signal that substitution across nests happens more readily than substitution within a nest.

Interestingly, using dissimilarity parameters that take on values larger than one allows us to model synergistic, or halo, effects among the products within the same nest. In particular, under the nested

logit model, if we offer the assortment $(S_1, \ldots, S_m)$, then the probability that a customer purchases product $j \in S_i$ in nest i is given by

$$Q_i(S_1, \ldots, S_m) \, P_{ij}(S_i) = \frac{V_i(S_i)^{\gamma_i - 1}}{v_0 + \sum_{l \in M} V_l(S_l)^{\gamma_l}} \, v_{ij}.$$

From the expression above, we observe that when $\gamma_i \leq 1$, adding a product $k \notin S_i$ to nest i always decreases the purchase probability of product $j \in S_i$. Therefore, when the dissimilarity parameters of the nests do not exceed one, the products in a nest always act as competitors to each other and adding a new product to a nest decreases the probability of purchase for the other products in the nest. In practice, this is not always the case. For example, if the nests correspond to different car dealers, then offering a new luxury car may increase the probability of purchase for other cars in the same dealer because the newly offered luxury car may help attract a larger fraction of customers. We observe that if $\gamma_i > 1$ and we add a new product to nest i, then both $V_i(S_i)^{\gamma_i - 1}$ and $\sum_{l \in M} V_l(S_l)^{\gamma_l}$ increase in the expression above. As a result, the probability that a customer purchases product j may increase or decrease. This feature may allow us to model synergies between different products in a nest. When such synergies exist, it may even be beneficial to include loss leaders in a nest to attract traffic to this nest. Motivated by this observation, we refer to the case with $\gamma_i \leq 1$ for all $i \in M$ as the case with purely *competitive products*, whereas we refer to the case with no restrictions on the dissimilarity parameters of the nests as the case with *possibly synergistic products*.

Along the second dimension, the case with $v_{i0} = 0$ for all $i \in M$ corresponds to a situation where if a customer decides to make a purchase within a particular nest, then the customer always makes a purchase within the selected nest. In other words, the demand within a nest is fully captured without loss to the no purchase option. We refer to this situation as the case with *fully-captured nests*. On the other hand, if the preference weight of the no purchase option within a nest is strictly positive, then a customer may leave without purchasing anything in the selected nest. We refer to this situation as the case with *partially-captured nests*. There are a number of reasons to consider the case with partially-captured nests. To begin with, firms do not make their assortment offer decisions in isolation and the preference weight v_{i0} of the no purchase option in nest i can be used to capture the attractiveness of the products offered by other firms in nest i. For example, if each nest corresponds to a particular store, then the total preference weight of the products offered by all firms may serve to attract the customers to a store, but once a customer decides to make a purchase in a particular store, he may choose a product offered by another firm, which is equivalent to not making a purchase within the offered assortment. As mentioned above, the assortment problem becomes NP-hard when we allow partially-captured nests. Game theoretic assortment optimization models is beyond the scope of our paper, but this computational complexity result also indicates that finding the best response of a firm to the assortments already offered by other firms is a computationally difficult problem.

Another reason to consider the case with partially-captured nests is that there are extensions of the nested logit model that allow offering a particular alternative within multiple nests. These extensions are referred to as generalized nested logit models and Train (2003) shows that generalized nested logit models are consistent with a random utility-based choice model. Our use of the partially-captured nests corresponds to a version of generalized nested logit models where the no purchase option is offered

9

within multiple nests. When we also impose the restriction that the dissimilarity parameters of all of the nests are equal to each other, the resulting choice model is referred to as the cross nested logit model. Farias et al. (2011) use the cross nested logit model to include a no purchase option within each nest. The choice model in Farias et al. (2011) precisely corresponds to the partially-captured nests that we consider in this paper.

A final reason to consider the case with partially-captured nests is that certain products may have to be included in the offered assortment, in which case, the parameter v_{i0} can be used to capture the total preference weight of the products that have to be included in the offered assortment. Naturally, there is a revenue contribution associated with the products that have to be included in the offered assortment and the definition of the expected revenue $R_i(S_i)$ above does not keep track of this revenue contribution. However, it turns out that we can follow the same line of reasoning that we use to deal with partially-captured nests to deal with the case where certain products have to be included in the offered assortment. Thus, by building on our treatment of partially-captured nests, we can address assortment optimization problems where certain products are already included in the offered assortment and we need to decide which additional products should be offered.

In our nested logit model, the dissimilarity parameters $(\gamma_1, \ldots, \gamma_m)$ are assumed to be constants. To understand this assumption, it is useful to view the nested logit model as a random utility-based choice model. In particular, consider the case where a customer associates random utilities with the products and with the no purchase options. The random utilities have a multi-dimensional generalized extreme value distribution. The utility of product j in nest i has mean $\bar{u}_{ij}$ and unit variance. The utilities of the options in different nests are independent of each other. The customer, being a utility maximizer, follows the option with the largest utility. In this case, we can show that the probability of choosing a particular product under this random utility-based choice model has the same form specified by the nested logit model; see McFadden (1981). To obtain the nested logit model corresponding to this random utility-based choice model, we need to set the preference weight v_{ij} of product j in nest i as $e^{\bar{u}_{ij}/\gamma_i}$. The correlation structure between the random utilities of the options in nest i determines the value of the dissimilarity parameter γ_i of nest i. Thus, the means and correlation structure of the random utilities are the primitives of a random utility-based choice model that is consistent with the nested logit model. In our nested logit model, we assume that the means and correlations of the random utilities are fixed, implying that the preference weights and the dissimilarity parameters are fixed as well. It is possible to consider richer choice models where the means and correlations of the random utilities depend on the offered assortment, in which case, the preference weights and the dissimilarity parameters depend on the offered assortment as well, but this extension makes it difficult to solve the corresponding assortment optimization problem. We expand on this issue in Section 9.

2 Linear Programming Representation

The assortment optimization problem in (2) is of a combinatorial nature. In this section, we present a linear programming formulation of this problem. The linear programming formulation is not too useful directly as a computational tool since its number of constraints grows exponentially with the number of

products. However, it turns out that we can build on the linear programming formulation to develop a general approximation result for problem (2). The approximation methods that we propose throughout the paper are tightly related to this general approximation result.

To formulate problem (2) as a linear program, we first observe that this problem is equivalent to $\min\{x : x \geq \sum_{i\in M} Q_i(S_1,\ldots,S_m) R_i(S_i) \ \forall (S_1,\ldots,S_m) \text{ with } S_i \subset N, i \in M\}$. By using the definition of $Q_i(S_1,\ldots,S_m)$ in (1), we can write the constraints in this problem as

$$ v_0\,x \geq \sum_{i\in M}(V_i(S_i)^{\gamma_i} R_i(S_i) - V_i(S_i)^{\gamma_i} x) \qquad \forall (S_1,\ldots,S_m) \text{ with } S_i \subset N,\ i \in M, $$

which, in turn, are equivalent to the single constraint

$$ v_0\,x \geq \max_{(S_1,\ldots,S_m):S_i\subset N, i\in M}\left\{\sum_{i\in M} V_i(S_i)^{\gamma_i}(R_i(S_i) - x)\right\}. $$

The key observation is that the optimization problem on the right side of the constraint above decomposes by the nests and the constraint can be written as

$$ v_0\,x \geq \sum_{i\in M} \max_{S_i\subset N} V_i(S_i)^{\gamma_i}(R_i(S_i) - x). $$

Therefore, problem (2) is equivalent to

$$ \begin{aligned} \min \quad & x \\ \text{s.t.} \quad & v_0\,x \geq \sum_{i\in M} \max_{S_i\subset N} V_i(S_i)^{\gamma_i}(R_i(S_i) - x), \end{aligned} $$

where the only decision variable is x. Noting that Z^* is the optimal objective value of problem (2), the discussion so far implies that if x^* is the optimal solution to the problem above, then the optimal objective value of this problem is also x^* and we have $x^* = Z^*$. The constraint of the problem above is nonlinear, but to linearize this constraint, we can define the decision variables $y = (y_1,\ldots,y_m)$ as $y_i = \max_{S_i\subset N} V_i(S_i)^{\gamma_i}(R_i(S_i) - x)$ and write the problem as

$$ \begin{aligned} \min \quad & x & (3)\\ \text{s.t.} \quad & v_0\,x \geq \sum_{i\in M} y_i \\ & y_i \geq V_i(S_i)^{\gamma_i}(R_i(S_i) - x) \qquad \forall\, S_i \subset N,\ i \in M, \end{aligned} $$

where the decision variables are (x, y). Problem (3) is a linear program with $1 + m$ decision variables and $1 + m\,2^n$ constraints. When $v_0 = 0$, the first constraint above reads $\sum_{i\in M} y_i \leq 0$, but the second set of constraints prevent x from becoming arbitrarily small because if x becomes arbitrarily small, then the decision variables $(y_1,\ldots,y_m)$ would take arbitrarily large values and we cannot satisfy the constraint $\sum_{i\in M} y_i \leq 0$. Therefore, problem (3) continues to apply when the preference weight v_0 of the no purchase option is zero. Another useful observation is that the number of possible assortments in problem (2) is 2^{mn}, which increases exponentially in both the number of nests and the number of products in each nest. In contrast, the numbers of decision variables and constraints in problem (3)

grow linearly with the number of nests, and problem (3) can be tractable when the number of products in each nest is relatively small, irrespective of the number of nests. When the number of products in each nest is large, a possible solution approach for problem (3) is to use column generation on its dual, but due to the presence of the dissimilarity coefficient in the second set of constraints, the column generation subproblem is nonlinear and this renders column generation intractable.

Although problem (3) is difficult to solve when the number of products in each nest is large, we can build on this problem to develop a general approximation method. Assume that we identify a collection of candidate assortments $\{A_{it} : t \in \mathcal{T}_i\}$ that we may consider offering in nest i, where we have $A_{it} \subset N$ for all $t \in \mathcal{T}_i$. We are interested in finding a combination of these assortments for the different nests so that the combined assortment provides the largest possible expected revenue. In other words, we are interested in finding the assortment that provides the largest expected revenue when we consider assortments of the form $(S_1, \ldots, S_m)$ with $S_i \in \{A_{it} : t \in \mathcal{T}_i\}$ for all $i \in M$. This problem can be formulated as the linear program

$$\min \quad x \tag{4}$$
$$\text{s.t.} \quad v_0\, x \geq \sum_{i \in M} y_i$$
$$y_i \geq V_i(S_i)^{\gamma_i}(R_i(S_i) - x) \qquad \forall\, S_i \in \{A_{it} : t \in \mathcal{T}_i\},\ i \in M.$$

The number of decision variables in problem (4) is still $1+m$. The number of constraints is $1 + \sum_{i \in M} |\mathcal{T}_i|$, which can be reasonable when the collections of candidate assortments are not too large.

We now provide some observations to develop a general approximation result by building on problem (4). The constraints in problem (4) can succinctly be written as

$$v_0\, x \geq \sum_{i \in M} \max_{S_i \in \{A_{it} : t \in \mathcal{T}_i\}} V_i(S_i)^{\gamma_i}(R_i(S_i) - x).$$

We will now argue that the constraint above must be satisfied as equality at an optimal solution. Let $(\hat{x}, \hat{y})$ be an optimal solution to problem (4) and suppose there is a gap. We can then decrease the value of $\hat{x}$ without violating the constraint, thereby obtaining a strictly better solution to problem (4) than $\hat{x}$, establishing the claim. Therefore, letting $\hat{S}_i$ be the solution to the maximization problem on the right side of the constraint above with $x = \hat{x}$, it must be the case that $v_0\, \hat{x} = \sum_{i \in M} V_i(\hat{S}_i)^{\gamma_i}(R_i(\hat{S}_i) - \hat{x})$. Solving for $\hat{x}$, we obtain

$$\hat{x} = \frac{\sum_{i \in M} V_i(\hat{S}_i)^{\gamma_i} R_i(\hat{S}_i)}{v_0 + \sum_{i \in M} V_i(\hat{S}_i)^{\gamma_i}} = \sum_{i \in M} Q_i(\hat{S}_1, \ldots, \hat{S}_m)\, R_i(\hat{S}_i) = \Pi(\hat{S}_1, \ldots, \hat{S}_m).$$

Consequently, if $(\hat{x}, \hat{y})$ is an optimal solution to problem (4) and $\hat{S}_i$ is an optimal solution to the problem $\max_{S_i \in \{A_{it} : t \in \mathcal{T}_i\}} V_i(S_i)^{\gamma_i}(R_i(S_i) - \hat{x})$, then the expected revenue obtained by offering the assortment $(\hat{S}_1, \ldots, \hat{S}_m)$ is precisely $\hat{x}$.

We observe that problem (4) includes only a subset of the constraints in problem (3), which implies that problem (4) is a relaxed version of problem (3). Therefore, if we let (x^*, y^*) and $(\hat{x}, \hat{y})$ respectively be the optimal solutions to problems (3) and (4), then we have $Z^* = x^* \geq \hat{x}$. Furthermore, for some

α and β, if we can show that $(\alpha\,\hat{x}, \beta\,\hat{y})$ is a feasible solution to problem (3), then we also obtain $\alpha\,\hat{x} \geq Z^* = x^* \geq \hat{x}$, which implies that the expected revenue obtained by offering the assortment $(\hat{S}_1, \ldots, \hat{S}_m)$ as defined above deviates from the optimal expected revenue by no more than a factor of α. We collect these observations in the following theorem.

Theorem 1 *Let $(\hat{x}, \hat{y})$ be an optimal solution to problem (4), and for all $i \in M$, let $\hat{S}_i$ be an optimal solution to the problem*

$$\max_{S_i \in \{A_{it} : t \in \mathcal{T}_i\}} V_i(S_i)^{\gamma_i}(R_i(S_i) - \hat{x}). \tag{5}$$

If $(\alpha\,\hat{x}, \beta\,\hat{y})$ is a feasible solution to problem (3) for some α and β, then the expected revenue obtained by offering the assortment $(\hat{S}_1, \ldots, \hat{S}_m)$ deviates from the optimal objective value of problem (2) by no more than a factor of α. In other words, letting $\hat{Z} = \Pi(\hat{S}_1, \ldots, \hat{S}_m)$, we have $\alpha\,\hat{Z} \geq Z^* \geq \hat{Z}$.

Theorem 1 provides sufficient conditions under which we can stitch together a good assortment from the collections of candidate assortments $\{A_{it} : t \in \mathcal{T}_i\}$ for $i \in M$. The thought process we used to reach Theorem 1 will be critical throughout the paper. In particular, we will design collections of assortments such that if $(\hat{x}, \hat{y})$ is the optimal solution to problem (4) with these collections of assortments, then $(\alpha\,\hat{x}, \beta\,\hat{y})$ ends up being a feasible solution to problem (3) for some α and β. In that case, we can solve problem (4) with these collections of assortments to obtain the optimal solution $(\hat{x}, \hat{y})$. Letting $\hat{S}_i$ be the assortment that solves problem (5), Theorem 1 implies that the expected revenue from the assortment $(\hat{S}_1, \ldots, \hat{S}_m)$ deviates from the optimal expected revenue by at most a factor of α.

We use various collections of candidate assortments. One possibility is to use the assortments that include a certain number of products with the largest revenues. This class of assortments is known to be optimal when the customer choices are governed by the multinomial logit model and Rusmevichientong, Shmoys and Topaloglu (2010) give a performance guarantee for such assortments when the underlying choice model is the multinomial logit model with multiple customer types. In the next section, we show that this class of assortments is still optimal under the nested logit model as long as we only have competitive products and fully-captured nests, but in general, we may need to look beyond this class to find good solutions for the assortment optimization problem we are interested in.

3 Competitive Products and Fully-Captured Nests

In this section, we focus on instances of the assortment optimization problem in (2) with $\gamma_i \leq 1$ and $v_{i0} = 0$ for all $i \in M$. For this case, we show that there exists an optimal solution $(S_1^*, \ldots, S_m^*)$ to problem (2) such that each one of the assortments $S_1^*, \ldots, S_m^*$ is either the empty assortment or is of the form $S_i^* = \{1, 2, \ldots, j\}$ for some $j \in N$. Noting that the products in each nest are ordered such that $r_{i1} \geq r_{i2} \geq \ldots \geq r_{in}$, this result implies that an optimal assortment in each nest includes a certain number of products with the largest revenues. We call such assortments nested-by-revenue assortments. For notational brevity, we use N_{ij} to denote the nested-by-revenue assortment that includes the first j products with the largest revenues in nest i. In other words, we have $N_{ij} = \{1, 2, \ldots, j\}$ for all

$i \in M$, $j \in N$. For notational uniformity, we also let $N_{i0} = \emptyset$ and $N_+ = N \cup \{0\}$, in which case, our goal is to show that an optimal solution $(S_1^*, \dots, S_m^*)$ to problem (2) is of the form $S_i^* = N_{ij}$ for some $j \in N_+$. Throughout this section, we assume without loss of generality that $v_0 > 0$. Otherwise, since we have $v_{i0} = 0$ for all $i \in M$, it is optimal to offer only one product with the largest revenue $\max_{i \in M} r_{i1}$ over all nests and this product would be purchased with probability one.

The following proposition shows that if it is optimal to offer a nonempty assortment in a nest, then the expected revenue from this nest should at least be equal to the optimal expected revenue over all nests.

Proposition 2 *If $(S_1^*, \dots, S_m^*)$ is an optimal solution to problem (2) and $S_i^* \neq \emptyset$, then $R_i(S_i^*) \geq Z^*$.*

Proof. To get a contradiction, assume that $S_i^* \neq \emptyset$ and $R_i(S_i^*) < Z^*$. For notational convenience, let $R_l = R_l(S_l^*)$ and $q_l = V_l(S_l^*)^{\gamma_l}$ for all $l \in M$. Thus, we have $Q_i(S_1^*, \dots, S_m^*) = q_i/(v_0 + \sum_{l \in M} q_l)$ and we can write the optimal expected revenue as

$$Z^* = \Pi(S_1^*, \dots, S_m^*) = \frac{\sum_{l \in M} q_l R_l}{v_0 + \sum_{l \in M} q_l} = \frac{v_0 + \sum_{l \in M, l \neq i} q_l}{v_0 + \sum_{l \in M} q_l} \frac{\sum_{l \in M, l \neq i} q_l R_l}{v_0 + \sum_{l \in M, l \neq i} q_l} + \frac{q_i R_i}{v_0 + \sum_{l \in M} q_l}$$

$$= \frac{v_0 + \sum_{l \in M, l \neq i} q_l}{v_0 + \sum_{l \in M} q_l} \Pi(S_1^*, \dots, S_{i-1}^*, \emptyset, S_{i+1}^*, \dots, S_m^*) + \frac{q_i}{v_0 + \sum_{l \in M} q_l} R_i.$$

Noting that $v_0 > 0$ and $q_i = V_i(S_i^*)^{\gamma_i} > 0$, the equality above shows that Z^* is a nontrivial convex combination of $\Pi(S_1^*, \dots, S_{i-1}^*, \emptyset, S_{i+1}^*, \dots, S_m^*)$ and $R_i = R_i(S_i^*)$. So, having $R_i < Z^*$ implies that $\Pi(S_1^*, \dots, S_{i-1}^*, \emptyset, S_{i+1}^*, \dots, S_m^*) > Z^*$ contradicting the fact that Z^* is the optimal expected revenue. $\square$

In the following lemma, we show that the products with small revenues can be removed from the assortment without degrading the performance.

Lemma 3 *Assume that $Z = \Pi(S_1, \dots, S_m)$ for some assortment $(S_1, \dots, S_m)$ and there exists a product $j \in S_i$ that satisfies $r_{ij} < \gamma_i Z + (1 - \gamma_i) R_i(S_i)$ and $R_i(S_i) \geq Z$. Then, removing product j from S_i yields a strictly larger expected revenue than Z.*

Proof. Let $\hat{S}_i$ be the assortment constructed by removing product j from S_i. We show that the assortment $(S_1, \dots, S_{i-1}, \hat{S}_i, S_{i+1}, \dots, S_m)$ provides an expected revenue of $\hat{Z}$ satisfying $\hat{Z} > Z$. For notational convenience, let $\hat{R}_i = R_i(\hat{S}_i)$, $\hat{q}_i = V_i(\hat{S}_i)^{\gamma_i}$, $R_l = R_l(S_l)$ and $q_l = V_l(S_l)^{\gamma_l}$ for all $l \in M$. Following an argument similar to the one in the proof of Proposition 2, we can write the expected revenue from the assortment $(S_1, \dots, S_m)$ as

$$Z = \Pi(S_1, \dots, S_m) = \frac{\sum_{l \in M} q_l R_l}{v_0 + \sum_{l \in M} q_l} = \frac{v_0 + \sum_{l \in M, l \neq i} q_l + \hat{q}_i}{v_0 + \sum_{l \in M} q_l} \frac{\sum_{l \in M, l \neq i} q_l R_l + \hat{q}_i \hat{R}_i}{v_0 + \sum_{l \in M, l \neq i} q_l + \hat{q}_i} + \frac{q_i R_i - \hat{q}_i \hat{R}_i}{v_0 + \sum_{l \in M} q_l}$$

$$= \frac{v_0 + \sum_{l \in M, l \neq i} q_l + \hat{q}_i}{v_0 + \sum_{l \in M} q_l} \Pi(S_1, \dots, S_{i-1}, \hat{S}_i, S_{i+1}, \dots, S_m) + \frac{q_i - \hat{q}_i}{v_0 + \sum_{l \in M} q_l} \frac{q_i R_i - \hat{q}_i \hat{R}_i}{q_i - \hat{q}_i}.$$

14

Therefore, Z is a convex combination of $\hat{Z} = \Pi(S_1, \ldots, S_{i-1}, \hat{S}_i, S_{i+1}, \ldots, S_m)$ and $(q_i R_i - \hat{q}_i \hat{R}_i)/(q_i - \hat{q}_i)$. In this case, the desired result follows if we can show that $Z > (q_i R_i - \hat{q}_i \hat{R}_i)/(q_i - \hat{q}_i)$. In the rest of the proof, we equivalently show that $q_i R_i - \hat{q}_i \hat{R}_i < (q_i - \hat{q}_i) Z$.

Let $\alpha = V_i(\hat{S}_i)/V_i(S_i)$, so $\hat{q}_i = \alpha^{\gamma_i} q_i$. Using the fact that $v_{ij} = V_i(S_i) - V_i(\hat{S}_i)$, we can write $\hat{R}_i$ as

$$\hat{R}_i = \frac{\sum_{k \in \hat{S}_i} r_{ik} v_{ik}}{V_i(\hat{S}_i)} = \frac{\sum_{k \in S_i} r_{ik} v_{ik} - r_{ij} (V_i(S_i) - V_i(\hat{S}_i))}{\alpha V_i(S_i)} = \frac{1}{\alpha} R_i - \frac{1-\alpha}{\alpha} r_{ij}.$$

Therefore, $q_i R_i - \hat{q}_i \hat{R}_i < (q_i - \hat{q}_i) Z$ holds if and only of

$$q_i R_i - \alpha^{\gamma_i} q_i \left[\frac{1}{\alpha} R_i - \frac{1-\alpha}{\alpha} r_{ij} \right] < (q_i - \alpha^{\gamma_i} q_i) Z.$$

Arranging the terms in the expression above, we observe that $q_i R_i - \hat{q}_i \hat{R}_i < (q_i - \hat{q}_i) Z$ holds if and only if $r_{ij} < g(\alpha) Z + (1 - g(\alpha)) R_i$, where $g(\alpha) = (1 - \alpha^{\gamma_i})/(\alpha^{\gamma_i - 1} - \alpha^{\gamma_i})$. By the hypothesis, we have $r_{ij} < \gamma_i Z + (1 - \gamma_i) R_i$. So, it is enough to show that $\gamma_i Z + (1 - \gamma_i) R_i \leq g(\alpha) Z + (1 - g(\alpha)) R_i$. Since both sides of the last inequality are convex combinations of Z and R_i and $Z \leq R_i$ by the hypothesis, the inequality holds as long as $g(\alpha) \leq \gamma_i$. However, the last relationship is true because $g(\alpha)$ is increasing in α when $\gamma_i \leq 1$ and by L'Hopital's rule, $g(\alpha) \leq \lim_{\alpha \uparrow 1} g(\alpha) = \gamma_i$. $\qquad\square$

Lemma 3 gives us a mechanism to remove certain products with small revenues without reducing the expected revenue from an assortment. To see a useful implication of Proposition 2 and Lemma 3, assume that $(S_1^*, \ldots, S_m^*)$ is an optimal solution to problem (2) with $S_i^* \neq \emptyset$. We must have $R_i(S_i^*) \geq Z^*$ by Proposition 2. In this case, we must have $r_{ij} \geq \gamma_i Z^* + (1 - \gamma_i) R_i(S_i^*)$ for all $j \in S_i^*$ by Lemma 3. Otherwise, we can remove a product from S_i^* and obtain an assortment that provides a strictly larger expected revenue than Z^*. We use this observation in the following theorem to show that nested-by-revenue assortments provide an optimal solution to problem (2).

Theorem 4 *There exists an optimal solution $(S_1^*, \ldots, S_m^*)$ to problem (2) such that, for all $i \in M$, we have $S_i^* = N_{ij}$ for some $j \in N_+$.*

Proof. Assume that $(S_1^*, \ldots, S_m^*)$ is an optimal solution to problem (2) providing an expected revenue of Z^* and that there is a nest i such that S_i^* contains product j but not product k with $k < j$. Let $\hat{S}_i$ be the assortment constructed by adding product k to S_i^*. Using $\hat{Z}$ to denote the expected revenue from the assortment $(S_1^*, \ldots, S_{i-1}^*, \hat{S}_i, S_{i+1}^*, \ldots, S_m^*)$, we show that $\hat{Z} \geq Z^*$. Therefore, the assortment $(S_1^*, \ldots, S_{i-1}^*, \hat{S}_i, S_{i+1}^*, \ldots, S_m^*)$ must also be optimal. Repeating the argument until the assortments for all nests are of the form $\{1, 2, \ldots, j\}$ for some $j \in N_+$ establishes the result.

Similar to the notation in the proof of Lemma 3, let $\hat{R}_i = R_i(\hat{S}_i)$, $R_i = R_i(S_i^*)$, $\hat{q}_i = V_i(\hat{S}_i)^{\gamma_i}$ and $q_i = V_i(S_i^*)^{\gamma_i}$. The main idea is to use an argument similar to the one in the proof of Lemma 3 to write $\hat{Z} = \Pi(S_1^*, \ldots, S_{i-1}^*, \hat{S}_i, S_{i+1}^*, \ldots, S_m^*)$ as a convex combination of Z^* and $(\hat{q}_i \hat{R}_i - q_i R_i)/(\hat{q}_i - q_i)$. In this case, the desired result follows if we can show that $Z^* \leq (\hat{q}_i \hat{R}_i - q_i R_i)/(\hat{q}_i - q_i)$. We equivalently show that $(\hat{q}_i - q_i) Z^* \leq \hat{q}_i \hat{R}_i - q_i R_i$. Let $\alpha = V_i(S_i^*)/V_i(\hat{S}_i)$, so $\hat{q}_i = q_i/\alpha^{\gamma_i}$. Using the fact that

$v_{ik} = V_i(\hat{S}_i) - V_i(S_i^*)$, we write $\hat{R}_i$ as

$$\hat{R}_i = \frac{\sum_{j' \in \hat{S}_i} r_{ij'} \, v_{ij'}}{V_i(\hat{S}_i)} = \alpha \frac{\sum_{j' \in S_i^*} r_{ij'} \, v_{ij'} + r_{ik} \, (V_i(\hat{S}_i) - V_i(S_i^*))}{V_i(S_i^*)} = \alpha \, R_i + (1 - \alpha) \, r_{ik}.$$

Therefore, $(\hat{q}_i - q_i) \, Z^* \leq \hat{q}_i \hat{R}_i - q_i R_i$ holds if and only if

$$\left[\frac{q_i}{\alpha^{\gamma_i}} - q_i \right] Z^* \leq \frac{q_i}{\alpha^{\gamma_i}} \, (\alpha \, R_i + (1 - \alpha) \, r_{ik}) - q_i R_i.$$

Arranging the terms in the expression above, we observe that $(\hat{q}_i - q_i) \, Z^* \leq \hat{q}_i \hat{R}_i - q_i R_i$ holds if and only if $h(\alpha) \, Z^* + (1 - h(\alpha)) \, R_i \leq r_{ik}$, where we let $h(\alpha) = (1 - \alpha^{\gamma_i})/(1 - \alpha)$. From the discussion that follows Lemma 3, since $j \in S_i^*$, we know that $\gamma_i \, Z^* + (1 - \gamma_i) \, R_i \leq r_{ij} \leq r_{ik}$, where the last inequality follows from the fact that $k < j$. So, it is enough to show that $h(\alpha) \, Z^* + (1 - h(\alpha)) \, R_i \leq \gamma_i \, Z^* + (1 - \gamma_i) \, R_i$. Since both sides of the last inequality are convex combinations of Z^* and R_i and $Z^* \leq R_i$ by Proposition 2, the inequality holds if and only if $h(\alpha) \geq \gamma_i$. However, the last relationship is true because $h(\alpha)$ is decreasing in α and by L'Hoptial's rule, $h(\alpha) \geq \lim_{\alpha \uparrow 1} h(\alpha) = \gamma_i$. □

Theorem 4 shows that we can construct an optimal solution to problem (2) by only considering the nested-by-revenue assortments. To find the best combination of such assortments for the different nests, we can make use of problem (4). In particular, we replace the collection of candidate assortments $\{A_{it} : t \in \mathcal{T}_i\}$ in problem (4) with the nested-by-revenue assortments $\{N_{ij} : j \in N_+\}$ and solve problem (4) to find the best combination of nested-by-revenue assortments for the different nests. By Theorem 4, this best combination has to be an optimal solution to problem (2). In this way, we can find an optimal solution to problem (2) by solving a linear program with $1 + m$ decision variables and $1 + m \, (1 + n)$ constraints.

4 Possibly Synergistic Products and Fully-Captured Nests

In this section, we focus on instances of the assortment optimization problem in (2), where we do not have any restrictions on the dissimilarity parameters $(\gamma_1, \ldots, \gamma_m)$ of the nests, but we still have $v_{i0} = 0$ for all $i \in M$. We show that allowing the dissimilarity parameters for the nests to take on values larger than one changes the structure of problem (2) drastically. In particular, the result that we establish in the previous section does not necessarily hold when the dissimilarity parameters of the nests can take on arbitrary values and the nested-by-revenue assortments are no longer optimal. In the next section, we first characterize the computational complexity of the problem when we have no restrictions on the nest dissimilarity parameters. Following this result, we give a performance guarantee for the nested-by-revenue assortments. Throughout this discussion, we assume that $\gamma_i > 1$ for some $i \in M$. Otherwise, nested-by-revenue assortments are optimal by Theorem 4.

4.1 Computational Complexity

We begin by giving an example that shows why nested-by-revenue assortments are no longer optimal when the dissimilarity parameters of the nests can take on values larger than one. This example also demonstrates that nested-by-revenue assortments can perform arbitrarily badly when the revenues and

the preference weights of the products in a nest drastically differ from each other. Following this example, we establish that the assortment optimization problem in (2) is NP-hard whenever we allow $\gamma_i > 1$ for some $i \in M$.

To give an example where nested-by-revenue assortments do not perform well, we consider an instance of problem (2) with a single nest. The preference weight for the option of not choosing any of the nests is $v_0 = 1$. The dissimilarity parameter of the nest is $\gamma_1 = 2$. There are three products in the nest. Letting $\varepsilon \leq 1$ be a small positive number, the following table gives the revenues and the preference weights associated with the three products.

Product	1	2	3
Revenue	1	ε^4	0
Preference Weight	ε^2	$3/\varepsilon^2$	$1/\varepsilon$

Since there is only one nest with $\gamma_1 = 2$, the expected revenue from an assortment $S_1 \subset \{1, 2, 3\}$ is

$$\Pi(S_1) = Q_1(S_1)\,R_1(S_1) = \frac{V_1(S_1)^2}{v_0 + V_1(S_1)^2}\,\frac{\sum_{j \in S_1} r_{1j}\,v_{1j}}{V_1(S_1)} = \frac{V_1(S_1)\sum_{j \in S_1} r_{1j}\,v_{1j}}{v_0 + V_1(S_1)^2}.$$

We compute and bound the expected revenues from the three nested-by-revenue assortments as

$$\Pi(\{1\}) = \frac{\varepsilon^2\,\varepsilon^2}{1 + (\varepsilon^2)^2} \leq \varepsilon^4$$

$$\Pi(\{1,2\}) = \frac{(\varepsilon^2 + 3/\varepsilon^2)\,(\varepsilon^2 + 3\,\varepsilon^2)}{1 + (\varepsilon^2 + 3/\varepsilon^2)^2} \leq \frac{(4/\varepsilon^2)\,4\,\varepsilon^2}{9/\varepsilon^4} = \frac{16}{9}\,\varepsilon^4$$

$$\Pi(\{1,2,3\}) = \frac{(\varepsilon^2 + 3/\varepsilon^2 + 1/\varepsilon)\,(\varepsilon^2 + 3\,\varepsilon^2)}{1 + (\varepsilon^2 + 3/\varepsilon^2 + 1/\varepsilon)^2} \leq \frac{(5/\varepsilon^2)\,4\,\varepsilon^2}{9/\varepsilon^4} = \frac{20}{9}\,\varepsilon^4,$$

which implies that the expected revenue from the best nested-by-revenue assortment is no larger than $\frac{20}{9}\,\varepsilon^4$. On the other hand, the expected revenue from the assortment $\{1, 3\}$ is given by

$$\Pi(\{1,3\}) = \frac{(\varepsilon^2 + 1/\varepsilon)\,(\varepsilon^2)}{1 + (\varepsilon^2 + 1/\varepsilon)^2} \geq \frac{(1/\varepsilon)\,\varepsilon^2}{1 + (1/\varepsilon + 1/\varepsilon)^2} \geq \frac{\varepsilon}{1/\varepsilon^2 + (1/\varepsilon + 1/\varepsilon)^2} = \frac{1}{5}\,\varepsilon^3.$$

Thus, the optimal expected revenue exceeds the expected revenue from the best nested-by-revenue assortment by at least a factor of $(\varepsilon^3/5)/(\frac{20}{9}\,\varepsilon^4) = 9/(100\,\varepsilon)$. As $\varepsilon \to 0$, the performance of the best nested-by-revenue assortment becomes arbitrarily poor. The key observation in this problem instance is that the revenue associated with product 1 is quite large when compared with the other product revenues. Therefore, we would like to be able to sell product 1 with high probability. One can check that if we offer product 1 by itself, then the probability of purchase for product 1 is between $\varepsilon^4/2$ and ε^4. On the other hand, if we offer products 1 and 2 together, then the probability of purchase for product 1 is always smaller than $\varepsilon^4/2$. Therefore, if we offer product 2 next to product 1, then the probability of purchase for product 1 goes down. In contrast, if we offer products 1 and 3, then it is possible to check that the probability of purchase for product 1 always exceeds $\varepsilon^3/5$, which is larger than ε^4 for small values of ε. This observation indicates that product 3 acts as a synergistic product to product 1 and offering product 3 next to product 1 increases the probability of purchase for product 1. We also

note that even if the revenue of product 3 was not zero but slightly negative, it would still be beneficial to add this product to the offered assortment, justifying a loss leader.

It turns out that if we allow the dissimilarity parameters of the nests to take on values larger than one, then not only the nested-by-revenue assortments cease to be optimal, but problem (2) becomes NP-hard. We devote the rest of this section to showing this result. To show the result we are interested in, we focus on the following decision-theoretic formulation of the assortment optimization problem.

Assortment Feasibility. Given a profit threshold K, is there an assortment $(S_1, \ldots, S_m)$ that provides an expected revenue of K or more for problem (2)?

To establish the NP-hardness of problem (2), Theorem 5 below shows that any instance of the partition problem, which is a well-known NP-hard problem as established in Garey and Johnson (1979), can be reduced to an instance of the assortment feasibility problem. Rusmevichientong, Shmoys and Topaloglu (2010) also use a reduction from the partition problem to show the NP-hardness of an assortment optimization problem, but their choice model is the multinomial logit model with multiple customer types, rather than the nested logit model. The partition problem is described as follows.

Partition. Given integer-valued sizes $(c_1, \ldots, c_n)$ such that $\sum_{j=1}^{n} c_j = 2T$ with T integer, can we find a subset $S \subset \{1, \ldots, n\}$ such that $\sum_{j \in S} c_j = \sum_{j \in \{1, \ldots, n\} \setminus S} c_j = T$?

Theorem 5 *If we allow the dissimilarity parameters for the nests to take on values larger than one, then the assortment feasibility problem is NP-hard.*

Proof. Assume that we are given any instance of the partition problem with sizes $(c_1, \ldots, c_n)$ and $\sum_{j=1}^{n} c_j = 2T$. We define an instance of the assortment feasibility problem as follows. There is only one nest. The preference weight for the option of not choosing any of the nests is $v_0 = (1 + T)^2$. The dissimilarity parameter of the nest is $\gamma_1 = 2$. There are $n + 1$ products in the nest. The revenue associated with the first n products is given by $r_{1j} = 0$ for all $j = 1, \ldots, n$. The revenue associated with the last product is $r_{1,n+1} = 2(1 + T)$. The preference weights of the first n products are given by $v_{1j} = c_j$ for all $j = 1, \ldots, n$. The preference weight associated with the last product is $v_{1,n+1} = 1$. We set the expected revenue threshold in the assortment feasibility problem as $K = 1$.

In the rest of the proof, we show that there exists an assortment that provides an expected revenue of K or more in the assortment feasibility problem if and only if there exists a subset $S \subset \{1, \ldots, n\}$ such that $\sum_{j \in S} c_j = T$. The first observation that if we want to get a positive expected revenue in the assortment feasibility problem, then we have to offer the last product with revenue $2(1 + T)$. Therefore, the only question for the assortment feasibility problem is to choose a subset S among the products with zero revenues that makes sure that we obtain an expected revenue of $K = 1$ or more. If we offer a subset S of the first n products together with the last product, then the expected revenue is $Q_1(S \cup \{n+1\}) R_1(S \cup \{n+1\})$, which evaluates to

$$\frac{(\sum_{j \in S} c_j + 1)^2}{(1 + T)^2 + (\sum_{j \in S} c_j + 1)^2} \frac{2(1 + T)}{(\sum_{j \in S} c_j + 1)}.$$

Therefore, there exists an assortment with an expected revenue of $K = 1$ or more if and only if

$$\frac{(\sum_{j \in S} c_j + 1) \, 2 \, (1 + T)}{(1 + T)^2 + (\sum_{j \in S} c_j + 1)^2} \geq 1.$$

Arranging the terms in the expression above, the inequality above is equivalent to

$$2 \, (1 + T) \sum_{j \in S} c_j + 2 \, (1 + T) \geq 1 + 2 \, T + T^2 + 1 + 2 \sum_{j \in S} c_j + \left(\sum_{j \in S} c_j \right)^2,$$

which can equivalently be written as

$$\left(\sum_{j \in S} c_j \right)^2 - 2 \, T \sum_{j \in S} c_j + T^2 \leq 0.$$

Since the last inequality is equivalent to $(\sum_{j \in S} c_j - T)^2 \leq 0$, there exists an assortment with an expected revenue of $K = 1$ or more if and only if there exists a subset S with $(\sum_{j \in S} c_j - T)^2 \leq 0$. However, the only way for the last inequality to hold is to have $\sum_{j \in S} c_j = T$. Therefore, finding an assortment that yields an expected revenue of K or more is equivalent to finding a subset S that satisfies $\sum_{j \in S} c_j = T$ and the latter statement is precisely what the partition problem is interested in. $\qquad \square$

4.2 Performance of Nested-by-Revenue Assortments

In the previous section, we show that nested-by-revenue assortments may not perform well when we allow the dissimilarity parameters of the nests to take on values larger than one. Our goal in this section is to develop a performance bound for this class of assortments as a function of the problem data. In particular, recalling that we use N_{ij} to denote the nested-by-revenue assortment that includes the first j products with the largest revenues in nest i, we show that by focusing only on the nested-by-revenue assortments, we can construct a solution to problem (2) whose expected revenue deviates from the optimal expected revenue by no more than a factor of

$$\max_{i \in M, j=2,\dots,n} \left\{ \frac{R_i(N_{i,j-1})}{R_i(N_{ij})} \wedge \frac{R_i(N_{ij})}{R_i(N_{i,j-1})} \frac{V_i(N_{ij})^{\gamma_i}}{V_i(N_{i,j-1})^{\gamma_i}} \right\}, \tag{6}$$

where we let $a \wedge b = \min\{a, b\}$. Before we show this result, it is useful to observe the implications of this performance guarantee.

To see the effect of the first term in the minimum operator in (6), assume that the revenues of the products within a nest are balanced in the sense that the largest and the smallest product revenues within a nest differ from each other by at most a factor of ρ. Since the preference weights of the no purchase options within the nests are zero, $R_i(\cdot)$ is always smaller than the largest product revenue in nest i and is always larger than the smallest product revenue. Therefore, the ratio $R_i(N_{ij})/R_i(N_{i,j-1})$ in the expression above cannot exceed ρ. This observation implies that we expect the nested-by-revenue assortments to perform well when the revenues of the products within a particular nest are balanced. Note that the revenues of the products in different nests can still differ from each other arbitrarily. On the other hand, if we assume that the preference weights of the

products within a nest are balanced in the sense that the largest and the smallest preference weights within a nest differ from each other by at most a factor of κ, then we obtain $R_i(N_{i,j-1})/R_i(N_{ij}) = (\sum_{k=1}^{j-1} r_{ik} v_{ik} / \sum_{k=1}^{j} r_{ik} v_{ik}) (\sum_{k=1}^{j} v_{ik} / \sum_{k=1}^{j-1} v_{ik}) \leq \sum_{k=1}^{j} v_{ik} / \sum_{k=1}^{j-1} v_{ik} = V_i(N_{ij})/V_i(N_{i,j-1})$. Since the assortments N_{ij} and $N_{i,j-1}$ respectively include j and $j-1$ products, the ratio $V_i(N_{ij})/V_i(N_{i,j-1})$ is bounded from above by $(j\kappa)/(j-1)$ and the latter expression does not exceed 2κ for any $j = 2, \ldots, n$. Therefore, (6) indicates that the nested-by-revenue assortments provide a performance guarantee of 2κ as well, implying that these assortments are also expected to perform well when the preference weights of the products within a nest do not differ from each other drastically. Similar to the discussion for the product revenues, the preference weights of the products in different nests can still differ from each other arbitrarily. To see the effect of the second term in the minimum operator in (6), we observe that the expression in the curly brackets in (6) takes its largest value when the two terms of the minimum operator are equal to each other and this happens when

$$\frac{R_i(N_{i,j-1})}{R_i(N_{ij})} = \frac{V_i(N_{ij})^{\gamma_i/2}}{V_i(N_{i,j-1})^{\gamma_i/2}}.$$

Therefore, the value of the minimum operator in (6) is bounded from above by the expression on the right side above. Noting that we have $V_i(N_{ij})^{\gamma_i/2}/V_i(N_{i,j-1})^{\gamma_i/2} \leq (2\kappa)^{\gamma_i/2}$, the performance guarantee we give for nested-by-revenue assortments cannot exceed $\max_{i \in M}(2\kappa)^{\gamma_i/2}$ either. When $\gamma_i \leq 2$ for all $i \in M$, the latter performance guarantee is better than 2κ.

We make use of Theorem 1 to establish the performance guarantee that we give in (6). Assume that we solve problem (4) after replacing the collection of candidate assortments $\{A_{it} : t \in \mathcal{T}_i\}$ in the second set of constraints with the nested-by-revenue assortments $\{N_{ij} : j \in N_+\}$. Letting $(\hat{x}, \hat{y})$ be the optimal solution we obtain in this fashion and using α to denote the expression in (6), if we can show that $(\alpha\hat{x}, \alpha\hat{y})$ is a feasible solution to problem (3), then Theorem 1 implies that we can focus only on nested-by-revenue assortments and still obtain an assortment whose expected revenue deviates from the optimal expected revenue by no more than a factor of α.

To pursue this line of reasoning, we note that the second set of constraints in problem (3) can be written as $y_i \geq \max_{S_i \subset N} V_i(S_i)^{\gamma_i}(R_i(S_i) - x)$ for all $i \in M$. Using the decision variables $z_i = (z_{i1}, \ldots, z_{in}) \in [0,1]^n$, we formulate a tighter version of problem (3) as

$$
\begin{aligned}
\min \quad & x & (7)\\
\text{s.t.} \quad & v_0\, x \geq \sum_{i \in M} y_i \\
& y_i \geq \max_{z_i \in [0,1]^n} \left\{ \left(\sum_{j \in N} v_{ij}\, z_{ij} \right)^{\gamma_i} \left[\frac{\sum_{j \in N} r_{ij}\, v_{ij}\, z_{ij}}{\sum_{j \in N} v_{ij}\, z_{ij}} - x \right] \right\} \qquad \forall\, i \in M.
\end{aligned}
$$

Note that if we imposed the constraint $z_i \in \{0,1\}^n$ in the maximization problem on the right side of the second set of constraints above, then problems (3) and (7) would be equivalent to each other. The way it is formulated, problem (3) is a relaxed version of problem (7) in the sense that any feasible solution (x, y) to problem (7) is also feasible to problem (3). In the following lemma, we study the optimal solution to the maximization problem on the right side of the second set of constraints above.

Lemma 6 *There exists an optimal solution z_i^* to the problem*

$$\max_{z_i \in [0,1]^n} \left\{ \left(\sum_{j \in N} v_{ij} \, z_{ij} \right)^{\gamma_i} \left[\frac{\sum_{j \in N} r_{ij} \, v_{ij} \, z_{ij}}{\sum_{j \in N} v_{ij} \, z_{ij}} - x \right] \right\} \tag{8}$$

such that $z_{i1}^* = 1$, $z_{i2}^* = 1, \ldots, z_{i,k-1}^* = 1$, $z_{ik}^* \in [0,1]$, $z_{i,k+1}^* = 0, \ldots, z_{in}^* = 0$ for some $k = 1, \ldots, n$.

Proof. Assume that $\hat{z}_i$ is an optimal solution to problem (8) and let $C = \sum_{j \in N} v_{ij} \, \hat{z}_{ij}$. In this case, note that an optimal solution to the continuous knapsack problem

$$\max \left\{ \sum_{j \in N} r_{ij} \, v_{ij} \, z_{ij} : \sum_{j \in N} v_{ij} \, z_{ij} = C, z_i \in [0,1]^n \right\}$$

is also an optimal solution to problem (8). In the continuous knapsack problem above, the utility of item j is $r_{ij} \, v_{ij}$ and the space consumption of item j is v_{ij}. Thus, we can solve this problem by sorting the products with respect to their utility-to-space consumption ratios and filling the knapsack starting from the item with the largest utility-to-space consumption ratio. Since the utility-to-space consumption ratio of item j is r_{ij} and the products are ordered such that $r_{i1} \geq r_{i2} \geq \ldots \geq r_{in}$, there exists an optimal solution to the continuous knapsack problem above with the form given in the lemma. $\square$

Except for at most one possible fractional component, Lemma 6 shows that a nested-by-revenue assortment is optimal for the maximization problem on the right side of the second set of constraints in problem (7). Thus, it is not too surprising that if we solve problem (4) after replacing the collection of candidate assortments $\{A_{it} : t \in \mathcal{T}_i\}$ in the second set of constraints with the nested-by-revenue assortments $\{N_{ij} : j \in N_+\}$ and obtain the optimal solution $(\hat{x}, \hat{y})$, then $(\hat{x}, \hat{y})$ is almost feasible to problem (7). Since problem (3) is a relaxed version of problem (7), the solution $(\hat{x}, \hat{y})$ would be almost feasible to problem (3) as well. We make this intuitive argument precise in the following theorem and show that nested-by-revenue assortments provide the performance guarantee that we give in (6). We defer the proof of this result to the appendix.

Theorem 7 *Let $(\hat{x}, \hat{y})$ be an optimal solution to problem (4) when we solve this problem after replacing the collection of candidate assortments $\{A_{it} : t \in \mathcal{T}_i\}$ in the second set of constraints with the nested-by-revenue assortments $\{N_{ij} : j \in N_+\}$. Then, using α to denote the expression in (6), $(\alpha \, \hat{x}, \alpha \, \hat{y})$ is a feasible solution to problem (3).*

Theorem 7, along with Theorem 1, shows that if we only consider the nested-by-revenue assortments as candidate assortments, then we can construct a solution to problem (2) whose expected revenue deviates from the optimal expected revenue by at most a factor given in (6). Similar to the discussion at the end of Section 3, to find the best combination of nested-by-revenue assortments, we can solve problem (4) after replacing the collection of candidate assortments $\{A_{it} : t \in \mathcal{T}_i\}$ with the nested-by-revenue assortments $\{N_{ij} : j \in N_+\}$. This amounts to solving a linear program with $1 + m$ decision variables and $1 + m \, (1 + n)$ constraints.

5   Competitive Products and Partially-Captured Nests

In this section, we consider instances of the assortment optimization problem in (2) with $\gamma_i \leq 1$ for all $i \in M$, but the preference weights of the no purchase options within the nests can take on arbitrary values. In other words, we allow $(v_{10}, \ldots, v_{m0})$ to take on strictly positive values so that a customer can leave without purchasing anything even after this customer chooses a particular nest. For this case, nested-by-revenue assortments are no longer optimal. As a matter of fact, problem (2) turns out to be NP-hard. However, we are able to characterize a small class of assortments such that if we focus on this class of assortments, then we can construct a solution to problem (2) whose expected revenue deviates from the optimal expected revenue by no more than a factor of two.

In the following theorem, we show that the assortment optimization problem in (2) is NP-hard when we have $v_{i0} > 0$ for some $i \in M$. Our proof technique is similar to the one in Section 4.1. We consider the assortment feasibility problem as defined in Section 4.1 and show that any instance of the partition problem can be reduced to an instance of the assortment feasibility problem. However, the specifics of the reduction are more involved and we give the proof of this theorem in the appendix.

Theorem 8 *If we allow the preference weights of the no purchase options within the nests to take on strictly positive values, then the assortment feasibility problem is NP-hard.*

The difficulty with partially-captured nests is that such nests do not allow Proposition 2 to hold, which implies that it may be optimal to offer a nonempty assortment in a partially-captured nest even if the expected revenue from this nest is below the optimal expected revenue. As a result, our line of reasoning in Section 3 does not hold. It is possible to show that if we have a mixture of partially-captured and fully-captured nests, then it is still optimal to offer nested-by-revenue assortments in the fully-captured nests, but it is not clear what to do for the partially-captured ones. Motivated by the computational complexity result in Theorem 8, we turn our attention to obtaining approximate solutions. In this section, we develop a tractable approach that obtains a solution to problem (2) whose expected revenue deviates from the optimal expected revenue by at most a factor of two. To that end, we use an alternative representation of problem (3). Using the decision variables $z_i = (z_{i1}, \ldots, z_{in})$, we define $K_i(\epsilon_i)$ as the optimal objective value of the knapsack problem

$$K_i(\epsilon_i) = \max \left\{ \sum_{j \in N} r_{ij}\, v_{ij}\, z_{ij} : \sum_{j \in N} v_{ij}\, z_{ij} \leq \epsilon_i,\ z_i \in \{0,1\}^n \right\}. \tag{9}$$

In this case, noting that $V_i(S_i)$ and $R_i(S_i)$ in the second set of constraints in problem (3) are respectively given by $V_i(S_i) = v_{i0} + \sum_{j \in S_i} v_{ij}$ and $R_i(S_i) = \sum_{j \in S_i} r_{ij}\, v_{ij}/V_i(S_i)$, we consider the problem

$$\min \quad x \tag{10}$$
$$\text{s.t.} \quad v_0\, x \geq \sum_{i \in M} y_i$$
$$y_i \geq \max_{\epsilon_i \geq 0} \left\{ (v_{i0} + \epsilon_i)^{\gamma_i} \left[\frac{K_i(\epsilon_i)}{v_{i0} + \epsilon_i} - x \right] \right\} \qquad \forall\, i \in M.$$

The following lemma shows that problem (10) is equivalent to problem (3).

22

Lemma 9 *Problems (3) and (10) are equivalent to each other in the sense that an optimal solution to one problem is also an optimal solution to the other.*

Proof. Noting that the second set of constraints in problem (3) can equivalently be written as $y_i \geq \max_{S_i \subset N} V_i(S_i)^{\gamma_i}(R_i(S_i) - x)$ for all $i \in M$, the result follows if we can show that

$$\max_{S_i \subset N} V_i(S_i)^{\gamma_i}(R_i(S_i) - x) = \max_{\epsilon_i \geq 0} \left\{ (v_{i0} + \epsilon_i)^{\gamma_i} \left[\frac{K_i(\epsilon_i)}{v_{i0} + \epsilon_i} - x \right] \right\}$$

for any $x \geq 0$. Let ζ_L^* and ζ_R^* respectively be the optimal objective values of the problems on the left and right side above. First, we show that $\zeta_L^* \leq \zeta_R^*$. Assume that S_i^* is an optimal solution to the problem on the left side above and define $\epsilon_i^* = \sum_{j \in S_i^*} v_{ij}$. The solution obtained by setting $z_{ij} = 1$ for all $j \in S_i^*$ and $z_{ij} = 0$ otherwise is a feasible solution to problem (9) with $\epsilon_i = \epsilon_i^*$, which implies that $K_i(\epsilon_i^*) \geq \sum_{j \in S_i^*} r_{ij} v_{ij}$. Thus, if we evaluate the objective value of the problem on the right side above at $\epsilon_i = \epsilon_i^*$, then we obtain at least ζ_L^*, in which case, we obtain $\zeta_R^* \geq \zeta_L^*$.

Second, we show that $\zeta_L^* \geq \zeta_R^*$. Let ϵ_i^* be an optimal solution to the problem on the right side above and solve problem (9) after setting $\epsilon_i = \epsilon_i^*$. Letting z_i^* be the solution we obtain, we observe that we can assume without loss of generality that $\sum_{j \in N} v_{ij} z_{ij}^* = \epsilon_i^*$. To see this claim, if we have $\sum_{j \in N} v_{ij} z_{ij}^* < \epsilon_i^*$, then we can decrease the value of ϵ_i^* to $\hat{\epsilon}_i = \sum_{j \in N} v_{ij} z_{ij}^*$ while still preserving $K_i(\epsilon_i^*) = K_i(\hat{\epsilon}_i)$. In this case, using the fact that $\gamma_i \leq 1$ and $x \geq 0$, we obtain

$$(v_{i0} + \epsilon_i^*)^{\gamma_i} \left[\frac{K_i(\epsilon_i^*)}{v_{i0} + \epsilon_i^*} - x \right] = \frac{K_i(\epsilon_i^*)}{(v_{i0} + \epsilon_i^*)^{1-\gamma_i}} - (v_{i0} + \epsilon_i^*)^{\gamma_i} x$$

$$\leq \frac{K_i(\hat{\epsilon}_i)}{(v_{i0} + \hat{\epsilon}_i)^{1-\gamma_i}} - (v_{i0} + \hat{\epsilon}_i)^{\gamma_i} x = (v_{i0} + \hat{\epsilon}_i)^{\gamma_i} \left[\frac{K_i(\hat{\epsilon}_i)}{v_{i0} + \hat{\epsilon}_i} - x \right],$$

which shows that $\hat{\epsilon}_i$ should also be an optimal solution to the problem on the right side above, establishing the claim. By using the solution z_i^*, we define the assortment S_i^* as $S_i^* = \{j \in N : z_{ij}^* = 1\}$. Since $\sum_{j \in S_i^*} v_{ij} = \sum_{j \in N} v_{ij} z_{ij}^* = \epsilon_i^*$ and $\sum_{j \in S_i^*} r_{ij} v_{ij} = \sum_{j \in N} r_{ij} v_{ij} z_{ij}^* = K_i(\epsilon_i^*)$, the assortment S_i^* provides an objective value of ζ_R^* for the problem on the left side above and we obtain $\zeta_L^* \geq \zeta_R^*$. □

To exploit the equivalence between problems (3) and (10) in a tractable fashion, we use the continuous relaxation of the knapsack problem in (9), which is given by

$$\hat{K}_i(\epsilon_i) = \max \left\{ \sum_{j \in N} r_{ij} v_{ij} z_{ij} : \sum_{j \in N} v_{ij} z_{ij} \leq \epsilon_i, \ 0 \leq z_{ij} \leq \mathbf{1}(v_{ij} \leq \epsilon_i) \ \forall j \in N \right\}, \tag{11}$$

where we use $\mathbf{1}(\cdot)$ to denote the indicator function. Since problem (11) is a relaxation of problem (9), we have $\hat{K}_i(\epsilon_i) \geq K_i(\epsilon_i)$. The problem above is a continuous knapsack problem, where the utility of item j is $r_{ij} v_{ij}$, the space consumption of item j is v_{ij} and we can only consider the items whose space consumptions do not exceed ϵ_i. Noting that the utility-to-space consumption ratio of item j is r_{ij}, we can solve this problem by sorting the products with respect to their revenues and filling the knapsack starting from the product with the largest revenue, as long as we only consider the products whose preference weights do not exceed ϵ_i. We let $\hat{z}_i(\epsilon_i) = (\hat{z}_{i1}(\epsilon_i), \dots, \hat{z}_{in}(\epsilon_i))$ be an optimal solution to problem (11)

that we obtain in this fashion. We observe that $\hat{z}_i(\epsilon_i)$ has at most one fractional component. By using this solution, we define the assortment $\hat{S}_i(\epsilon_i)$ as $\hat{S}_i(\epsilon_i) = \{j \in N : \hat{z}_{ij}(\epsilon_i) = 1\}$, which includes only the strictly positive and integer-valued components of $\hat{z}_i(\epsilon_i)$. We use the assortments $\{\hat{S}_i(\epsilon_i) : \epsilon_i \in [0, \infty]\}$ as a collection of candidate assortments for nest i. We shortly show in this section that this collection of assortments includes no more than $1 + n^2$ assortments and each one of these $1 + n^2$ assortments can be identified in a tractable fashion.

To be able to obtain the performance guarantee of two, we augment the collection of assortments $\{\hat{S}_i(\epsilon_i) : \epsilon_i \in [0, \infty]\}$ for nest i by the collection of singleton assortments $\{\{j\} : j \in N\}$. We solve problem (4) after replacing the collection of candidate assortments $\{A_{it} : t \in \mathcal{T}_i\}$ in the second set of constraints with the collection of assortments $\{\hat{S}_i(\epsilon_i) : \epsilon_i \in [0, \infty]\} \cup \{\{j\} : j \in N\}$. Letting $(\hat{x}, \hat{y})$ be the optimal solution to problem (4) that we obtain in this fashion, if we can show that $(2\,\hat{x}, 2\,\hat{y})$ is a feasible solution to problem (3), then Theorem 1 implies that we can focus only on the assortments $\{\hat{S}_i(\epsilon_i) : \epsilon_i \in [0, \infty]\} \cup \{\{j\} : j \in N\}$ for nest i and still obtain an assortment whose expected revenue deviates from the optimal expected revenue by no more than a factor of two. We pursue this result in the following theorem, but defer the proof of this result to the appendix.

Theorem 10 *Let $(\hat{x}, \hat{y})$ be an optimal solution to problem* (4) *when we solve this problem after replacing the collection of candidate assortments $\{A_{it} : t \in \mathcal{T}_i\}$ in the second set of constraints with the assortments $\{\hat{S}_i(\epsilon_i) : \epsilon_i \in [0, \infty]\} \cup \{\{j\} : j \in N\}$. Then, $(2\,\hat{x}, 2\,\hat{y})$ is a feasible solution to problem* (3).

The equivalence between problems (3) and (10) given in Lemma 9 lays out the connection between the assortment optimization problem we are interested in and the knapsack problem, as long as we have $\gamma_i \leq 1$ for all $i \in M$. It is well-known that one can construct a solution to the knapsack problem in (9) by using its continuous relaxation in (11) and the objective value of this solution would deviate from the optimal objective value of the knapsack problem by no more than a factor of two; see Williamson and Shmoys (2011). The proof of Theorem 10 implicitly makes use of this result. Similarly, there exists a well-known fully polynomial-time approximation scheme for the knapsack problem, which can be found in Williamson and Shmoys (2011). Building on this fully polynomial-time approximation scheme, it is indeed possible to develop a fully polynomial-time approximation scheme for the assortment optimization problem we are interested in. We do not pursue the fully polynomial-time approximation scheme because this scheme can be developed by using an argument that is very similar to the preceding discussion in this section. The only difference is that instead of constructing approximate solutions to the knapsack problem in (9) by using its continuous relaxation, we construct approximate solutions by using the fully polynomial-time approximation scheme for knapsack problems.

In the remainder of this section, we argue that the collection of assortments $\{\hat{S}_i(\epsilon_i) : \epsilon_i \in [0, \infty]\}$ includes no more than $1 + n^2$ assortments and each one of these $1 + n^2$ assortments can be identified in a tractable fashion. Fix ϵ_i and consider the assortment $\hat{S}_i(\epsilon_i)$. The definition of $\hat{z}_i(\epsilon_i)$ implies that $\hat{S}_i(\epsilon_i)$ is a nested-by-revenue assortment as long as we focus only on the products whose preference weights do not exceed ϵ_i. In other words, letting k be the number of products whose preference weights do not exceed ϵ_i, $\hat{S}_i(\epsilon_i)$ is a nested-by-revenue assortment as long as we focus only on the products with the k

smallest preference weights. Given that we focus only on the products with the k smallest preference weights, we use N_{ij}^k to denote the nested-by-revenue assortment that includes the first j products with the largest revenues in nest i. Therefore, $\hat{S}_i(\epsilon_i)$ must be one of the assortments $\{N_{ij}^k : j = 0, \ldots, k\}$, where we let $N_{i0}^k = \emptyset$ for notational uniformity. Since the only possible values for k are $k = 1, \ldots, n$, it follows that $\{\hat{S}_i(\epsilon_i) : \epsilon_i \in [0, \infty]\} \subset \{N_{ij}^k : k \in N, \ j = 0, \ldots, k\}$. The latter collection of assortments includes no more than $1 + n^2$ distinct assortments, all of which can be easily be identified.

Theorem 10 shows that if we use the assortments $\{\hat{S}_i(\epsilon_i) : \epsilon_i \in [0, \infty]\} \cup \{\{j\} : j \in N\}$ as a collection of candidate assortments for nest i, then we can stitch together from these candidate assortments a solution to problem (2) whose expected revenue deviates from the optimal expected revenue by at most a factor of two. Since we have $\{\hat{S}_i(\epsilon_i) : \epsilon_i \in [0, \infty]\} \subset \{N_{ij}^k : k \in N, j = 0, \ldots, k\}$, using the assortments $\{N_{ij}^k : k \in N, \ j = 0, \ldots, k\} \cup \{\{j\} : j \in N\}$ as candidate assortments for nest i cannot degrade the performance guarantee of two. To find the best combination of these assortments, we can solve problem (4) after replacing the collection of candidate assortments $\{A_{it} : t \in \mathcal{T}_i\}$ with the assortments $\{N_{ij}^k : k \in N, \ j = 0, \ldots, k\} \cup \{\{j\} : j \in N\}$. This amounts to solving a linear program with $1 + m$ decision variables and $1 + m\,(1 + n + n^2)$ constraints.

6 Possibly Synergistic Products and Partially-Captured Nests

In this section, we consider the most general instances of the assortment optimization problem in (2), where we do not have any restrictions on the dissimilarity parameters $(\gamma_1, \ldots, \gamma_m)$ of the nests and the preference weights $(v_{10}, \ldots, v_{m0})$ of the no purchase options within the nests. We establish three results for these most general instances. First, we show that if the largest and the smallest preference weights within a nest do not differ from each other by more than a factor of κ, then the collection of assortments developed in the previous section provides a performance guarantee of 2κ. Second, using the links of our assortment problem to the partition problem, we give a pseudo-polynomial-time algorithm to obtain the optimal assortment. Third, letting $\bar{\gamma} = \max_{i \in M} \gamma_i$ for notational brevity, for any given $\delta > 1$, we construct a collection of assortments that provides a performance guarantee of $\delta^{2\bar{\gamma}+1}$, but the computational work required to obtain this collection increases as δ gets close to one. The important point about the last result is that it allows us to choose δ close to one to obtain an assortment with an arbitrarily good performance guarantee, but choosing δ close to one also increases the computational work. So, this result allows us to tradeoff computational work with performance guarantee. Throughout this section, we assume that $\bar{\gamma} > 1$. Otherwise, we can use the ideas in the previous section.

6.1 Performance Guarantee

We begin by showing that if the largest and the smallest preference weights within a nest differ from each other by at most a factor of κ, then the collection of assortments that we develop in the previous section provides a performance guarantee of $2\,\kappa$. It turns out that this result follows by synthesizing the results that are already given in Sections 4 and 5. In Section 4, we show that if we focus only on the nested-by-revenue assortments $\{N_{ij} : j \in N_+\}$, then we obtain the performance guarantee in (6) when we have possibly synergistic products and fully-captured nests. Noting the definition of N_{ij}^k in the previous

section, N_{ij}^n is identical to the nested-by-revenue assortment N_{ij}. Therefore, we have $\{N_{ij} : j \in N_+\} \subset \{N_{ij}^k : k \in N, \ j = 0, \ldots, k\}$, which implies that if we focus on the collection of assortments $\{N_{ij}^k : k \in N, \ j = 0, \ldots, k\}$, then we still obtain the performance guarantee in (6) when we have possibly synergistic products and fully-captured nests. On the other hand, in Section 5, we show that if we focus only on the collection of assortments $\{N_{ij}^k : k \in N, \ j = 0, \ldots, k\} \cup \{\{j\} : j \in N\}$, then we obtain a performance guarantee of two when we have competitive products and partially-captured nests. In this case, using M^f and M^p to respectively denote the sets of fully-captured and partially-captured nests and letting $a \vee b = \max\{a, b\}$, it is possible to synthesize these results to show that if we focus only on the collection of assortments $\{N_{ij}^k : k \in N, \ j = 0, \ldots, k\} \cup \{\{j\} : j \in N\}$, then we obtain a performance guarantee of

$$\max_{i \in M^f, j=2,\ldots,n} \left\{ \frac{V_i(N_{ij})}{V_i(N_{i,j-1})} \right\} \vee \max_{i \in M^p, j=1,\ldots,n} \left\{ \frac{V_i(N_{ij})}{V_i(N_{i,j-1})} \right\} \vee 2 \tag{12}$$

for the most general instances of the assortment optimization problem. Note that if i is a fully-captured nest, then $V_i(N_{i0}) = V_i(\emptyset) = 0$ and we do not consider the term $V_i(N_{i1})/V_i(N_{i0})$ in the expression above for fully-captured nests. To see the implication of the performance guarantee in (12), we observe that if the largest and the smallest preference weights within a nest differ from each other by at most a factor of κ, then we have $V_i(N_{ij})/V_i(N_{i,j-1}) \leq (j\kappa)/(j-1)$ for fully-captured nests and $V_i(N_{ij})/V_i(N_{i,j-1}) \leq (j+1)\kappa/j$ for partially-captured nests. Thus, the performance guarantee in (12) cannot exceed $2\,\kappa \vee 2 = 2\,\kappa$. This observation implies that if we use $\{N_{ij}^k : k \in N, \ j = 0, \ldots, k\} \cup \{\{j\} : j \in N\}$ as the collection of candidate assortments for nest i, then the best assortment that we can stitch together from these candidate assortments is expected to perform well when the preference weights within a nest do not differ too much from each other.

We can build on Theorem 1 to establish the performance guarantee in (12). Assume that we solve problem (4) after replacing the collection of candidate assortments $\{A_{it} : t \in \mathcal{T}_i\}$ in the second set of constraints with the collection $\{N_{ij}^k : k \in N, \ j = 0, \ldots, k\} \cup \{\{j\} : j \in N\}$. Letting $(\hat{x}, \hat{y})$ be the optimal solution we obtain in this fashion and using β to denote the expression in (12), if we can show that $(\beta\,\hat{x}, \beta\,\hat{y})$ is a feasible solution to problem (3), then Theorem 1 implies that we can focus only on the collection of assortments $\{N_{ij}^k : k \in N, \ j = 0, \ldots, k\} \cup \{\{j\} : j \in N\}$ and obtain an assortment whose expected revenue deviates from the optimal expected revenue by no more than a factor of β. The following theorem shows this result. Its proof is given in the appendix.

Theorem 11 *Let* $(\hat{x}, \hat{y})$ *be an optimal solution to problem* (4) *when we solve this problem after replacing the collection of candidate assortments* $\{A_{it} : t \in \mathcal{T}_i\}$ *in the second set of constraints with the assortments* $\{N_{ij}^k : k \in N, \ j = 0, \ldots, k\} \cup \{\{j\} : j \in N\}$. *Then, using* β *to denote the expression in* (12), $(\beta\,\hat{x}, \beta\,\hat{y})$ *is a feasible solution to problem* (3).

Thus, similar to the discussion at the end of Section 5, since there are $1 + n + n^2$ assortments in the collection $\{N_{ij}^k : k \in N, \ j = 0, \ldots, k\} \cup \{\{j\} : j \in N\}$, we can solve a linear program with $1 + m$ decision variables and $1 + m\,(1 + n + n^2)$ constraints to obtain an assortment whose expected revenue deviates from the optimal expected revenue by at most a factor given in (12).

6.2 Near-Optimal Assortments

In this section, we give a pseudo-polynomial-time algorithm for our assortment problem. Following this result, we develop an approximation scheme that can tradeoff performance guarantee with computational work. Both of these results are based on viewing problem (3) from a knapsack perspective. In particular, we define $G_i(\epsilon_i)$ as the optimal objective value of the knapsack problem

$$G_i(\epsilon_i) = \max_{S_i \subset N} \left\{ \sum_{j \in S_i} r_{ij} \, v_{ij} : v_{i0} + \sum_{j \in S_i} v_{ij} = \epsilon_i \right\}, \tag{13}$$

where we let $G_i(\epsilon_i) = -\infty$ when the problem on the right side above is infeasible. The second set of constraints in problem (3) are given by $y_i \geq V_i(S_i)^{\gamma_i} (\sum_{j \in S_i} r_{ij} \, v_{ij} / V_i(S_i) - x)$ for all $S_i \subset N$, $i \in M$ and problem (13) finds the largest value of $\sum_{j \in S_i} r_{ij} \, v_{ij}$ while keeping $V_i(S_i)$ constant at ϵ_i. Thus, problem (3) can equivalently be written as

$$\begin{aligned}
\min \quad & x \tag{14} \\
\text{s.t.} \quad & v_0 \, x \geq \sum_{i \in M} y_i \\
& y_i \geq \epsilon_i^{\gamma_i} \left[\frac{G_i(\epsilon_i)}{\epsilon_i} - x \right] \qquad \forall \, \epsilon_i \geq 0, \; i \in M,
\end{aligned}$$

where we treat $0/0$ as zero in the second set of constraints. Letting $v_i^L = v_{i0} + \min_{j \in N} v_{ij}$ and $v_i^U = v_{i0} + \sum_{j \in N} v_{ij}$, the definition of $G_i(\epsilon_i)$ in (13) implies that $G_i(\epsilon_i) = G_i(v_{i0})$ or $G_i(\epsilon_i) = -\infty$ for all $\epsilon_i < v_i^L$, whereas $G_i(\epsilon_i) = -\infty$ for all $\epsilon_i > v_i^U$. Therefore, it is enough to consider the values of ϵ_i with $\epsilon_i = v_{i0}$ or $\epsilon_i \in [v_i^L, v_i^U]$ in the second set of constraints in problem (14).

Problem (13) can be visualized as an optimization version of the partition problem since it tries to find an assortment S_i that maximizes the objective value among all assortments whose preference weights add up to ϵ_i. There is a well-known pseudo-polynomial-time algorithm for the partition problem; see Garey and Johnson (1979). This pseudo-polynomial-time algorithm operates under the assumption that all preference weights are integer-valued, in which case, the only values of ϵ_i that render problem (13) feasible are integer-valued. By using a dynamic program, the pseudo-polynomial-time algorithm computes $G_i(\epsilon_i)$ for all $\epsilon_i \in [v_i^L, v_i^U]$ in $O(n \, v_i^U)$ time. We give such a dynamic program in Appendix A.5. Thus, also noting that $G_i(v_{i0}) = 0$, the pseudo-polynomial-time algorithm can compute $G_i(\epsilon_i)$ for all values of ϵ_i that we need to consider in the second set of constraints in problem (14). Once we have $G_i(\epsilon_i)$ for all values of ϵ_i, we can solve problem (14) to obtain the optimal objective value of problem (3), which is, in turn, equal to the optimal expected revenue in problem (2). This approach amounts to solving a linear program with $1 + m$ decision variables and $1 + m \, (2 + v_i^U - v_i^L)$ constraints and it provides a pseudo-polynomial-time algorithm for our assortment optimization problem.

It is also possible to build on problem (14) to develop an approximation method that can tradeoff computational work with performance guarantee. This approximation method is based on constructing tractable approximations to $G_i(\epsilon_i)$. In particular, instead of considering every single possible value of ϵ_i with $\epsilon_i = v_{i0}$ or $\epsilon_i \in [v_i^L, v_i^U]$ in the second set of constraints in problem (14), we choose some

$\delta > 1$ and consider the values of ϵ_i that are close to the powers of δ. To do this, we define l_i^L as $l_i^L = \min\{l \in \mathbb{Z} : \delta^l \geq v_i^L\}$ and $l_i^U = \min\{l \in \mathbb{Z} : \delta^l \geq v_i^U\}$ so that we have $[v_i^L, v_i^U] \subset [\delta^{l_i^L - 1}, \delta^{l_i^U}]$. In this case, whenever ϵ_i lies in the interval $[\delta^{l-1}, \delta^l]$ for some $l = l_i^L, \ldots, l_i^U$, we approximate $G_i(\epsilon_i)$ by

$$\hat{G}_{il} = \max_{S_i \subset N} \left\{ \sum_{j \in S_i} r_{ij} \, v_{ij} : \delta^{l-1} \leq v_{i0} + \sum_{j \in S_i} v_{ij} \leq \delta^l \right\}. \tag{15}$$

Problem (15) is still difficult to solve exactly as it is a knapsack problem with both upper and lower bounds, but it turns out we can compute approximate solutions to this problem. In particular, Proposition 15 in the appendix shows that we can find an assortment $\hat{S}_{il}$ that satisfies $\delta^{l-1} \leq v_{i0} + \sum_{j \in \hat{S}_{il}} v_{ij} \leq \delta^l$ and $\delta \sum_{j \in \hat{S}_{il}} r_{ij} \, v_{ij} \geq \hat{G}_{il}$. In other words, the assortment $\hat{S}_{il}$ is a feasible solution to problem (15) and the objective value provided by this assortment deviates from the optimal objective value of problem (15) by no more than a factor of δ. Proposition 15 also shows that the computational work required to obtain the assortment $\hat{S}_{il}$ takes $O(\lceil \delta/(\delta-1) \rceil^2 \, n^{2 \lceil \delta/(\delta-1) \rceil})$ time, where we use $\lceil \cdot \rceil$ to denote the round up function. Thus, if we want $\hat{S}_{il}$ to be a more accurate solution to problem (15), then we need to choose δ closer to one, but choosing δ closer to one also increases the computational work to obtain the assortment $\hat{S}_{il}$. In this way, we can use $\hat{S}_{il}$ as an approximate solution problem (15) whenever ϵ_i lies in the interval $[\delta^{l-1}, \delta^l]$, while balancing computational work with accuracy.

We propose using the assortments $\{\hat{S}_{il} : l = l_i^L, \ldots, l_i^U\} \cup \{\emptyset\}$ as a collection of candidate assortments for nest i. To find the best combination of such assortments, we solve problem (4) after replacing the collection of candidate assortments $\{A_{it} : t \in \mathcal{T}_i\}$ in the second set of constraints with the collection of assortments $\{\hat{S}_{il} : l = l_i^L, \ldots, l_i^U\} \cup \{\emptyset\}$. Letting $(\hat{x}, \hat{y})$ be the optimal solution to problem (4) that we obtain in this fashion and recalling that $\bar{\gamma} = \max_{i \in M} \gamma_i$, if we can show that $(\delta^{2\bar{\gamma}+1} \hat{x}, \delta^{\bar{\gamma}+1} \hat{y})$ is a feasible solution to problem (3), then Theorem 1 implies that we can focus only on the assortments $\{\hat{S}_{il} : l = l_i^L, \ldots, l_i^U\} \cup \{\emptyset\}$ for nest i and still obtain an assortment whose expected revenue deviates from the optimal expected revenue by no more than a factor of $\delta^{2\bar{\gamma}+1}$. We establish this result in the following theorem. The proof of this theorem is given in the appendix.

Theorem 12 *Let $(\hat{x}, \hat{y})$ be an optimal solution to problem* (4) *when we solve this problem after replacing the collection of candidate assortments $\{A_{it} : t \in \mathcal{T}_i\}$ in the second set of constraints with the assortments $\{\hat{S}_{il} : l = l_i^L, \ldots, l_i^U\} \cup \{\emptyset\}$. Then, $(\delta^{2\bar{\gamma}+1} \hat{x}, \delta^{\bar{\gamma}+1} \hat{y})$ is a feasible solution to problem* (3).

Theorem 12 implies that if we use the assortments $\{\hat{S}_{il} : l = l_i^L, \ldots, l_i^U\} \cup \{\emptyset\}$ as candidate assortments for nest i, then we can combine these assortments for the different nests to obtain a solution to problem (2) whose expected revenue deviates from the optimal expected revenue by no more than a factor of $\delta^{2\bar{\gamma}+1}$. To find the best combination of these assortments, we need to solve problem (4) after replacing the collection of candidate assortments $\{A_{it} : t \in \mathcal{T}_i\}$ in the second set of constraints with the collection of assortments $\{\hat{S}_{il} : l = l_i^L, \ldots, l_i^U\} \cup \{\emptyset\}$. Since we have $\delta^{l_i^L} \geq v_i^L$ and $\delta^{l_i^U - 1} \leq v_i^U$, we have $l_i^U - l_i^L \leq 1 + \log_\delta(v_i^U/v_i^L)$. Therefore, for the most general instances of the assortment optimization problem we are interested in, we can solve a linear program with $1 + m$ decision variables and at most $1 + m \, (2 + \log_\delta(v_i^U/v_i^L))$ constraints to find an assortment whose expected revenue deviates from the optimal expected revenue by no more than a factor of $\delta^{2\bar{\gamma}+1}$.

7 Upper Bounds on Optimal Expected Revenue

In the previous three sections, we give collections of candidate assortments such that if we focus on these collections, then we can stitch together an assortment with a certain performance guarantee. The performance guarantees we give reflect the worst-case performance of a given collection of assortments, where the worst-case is taken over all possible problem instances. In this section, our goal is to develop a tractable upper bound on the optimal expected revenue that we can compute for each individual problem instance. We can then use the collections of assortments given in the previous sections to obtain the best possible assortment and compare the expected revenue from this assortment with the problem instance-specific upper bound on the optimal expected revenue. In this way, we can bound the optimality gap of the assortment we obtain for a particular problem instance.

To construct an upper bound on the optimal expected revenue, we use a tighter version of problem (3) that is similar to the one in Section 4.2. In particular, since the second set of constraints in problem (3) can be written as $y_i \geq \max_{S_i \subset N} V_i(S_i)^{\gamma_i}(R_i(S_i) - x)$ for all $i \in M$, using the decision variables $z_i = (z_{i1}, \ldots, z_{in}) \in [0,1]^n$, we formulate a tighter version of problem (3) as

$$\min \quad x \tag{16}$$
$$\text{s.t.} \quad v_0 \, x \geq \sum_{i \in M} y_i$$
$$y_i \geq \max_{z_i \in [0,1]^n} \left\{ \left(v_{i0} + \sum_{j \in N} v_{ij} \, z_{ij} \right)^{\gamma_i} \left[\frac{\sum_{j \in N} r_{ij} \, v_{ij} \, z_{ij}}{v_{i0} + \sum_{j \in N} v_{ij} \, z_{ij}} - x \right] \right\} \qquad \forall \, i \in M.$$

Problem (16) is a tighter version of problem (3) as any feasible solution to problem (16) is a feasible solution to problem (3). Therefore, if we can solve problem (16) in a tractable fashion, then the optimal objective value of this problem provides an upper bound on the optimal expected revenue.

To see how we can solve problem (16) in a tractable fashion, for a fixed value of x, we use $F_i(z_i \,|\, x)$ to denote the objective function of the maximization problem on the right side of the second set of constraints in problem (16) and let $\hat{F}_i(x) = \max_{z_i \in [0,1]^n} F_i(z_i \,|\, x)$. So, problem (16) is equivalent to

$$\min \quad x \tag{17}$$
$$\text{s.t.} \quad v_0 \, x \geq \sum_{i \in M} y_i$$
$$y_i \geq \hat{F}_i(x) \qquad \forall \, i \in M.$$

If we can show that $\hat{F}_i(x)$ is a convex function of x, then the feasible region of problem (17) ends up being convex. Thus, noting that the objective function is linear, problem (17) becomes a convex optimization problem. Furthermore, if we can obtain subgradients of $\hat{F}_i(x)$ with respect to x in a tractable fashion, then we can use a standard cutting plane method for convex optimization to solve problem (17); see Ruszczynski (2006). The following proposition follows a standard argument in nonlinear programming to show that $\hat{F}_i(x)$ is indeed a convex function of x and demonstrates how to obtain subgradients of this function with respect to x.

Proposition 13 *The function $\hat{F}_i(\cdot)$ is convex. Furthermore, if we let $\hat{z}_i(x)$ be an optimal solution to the problem $\max_{z_i \in [0,1]^n} F_i(z_i \mid x)$, then a subgradient of $\hat{F}_i(\cdot)$ at x is given by $-(v_{i0} + \sum_{j \in N} v_{ij}\, \hat{z}_{ij}(x))^{\gamma_i}$.*

Proof. By the definitions of $\hat{F}_i(x)$ and $\hat{z}_i(x)$, it follows that $\hat{F}_i(x) = F_i(\hat{z}_i(x) \mid x)$ and $\hat{F}_i(x') \geq F_i(\hat{z}_i(x) \mid x')$. Subtracting the equality from the inequality and noting that $F_i(\hat{z}_i(x) \mid x') - F_i(\hat{z}_i(x) \mid x) = -(v_{i0} + \sum_{j \in N} v_{ij}\, \hat{z}_{ij}(x))^{\gamma_i}\, (x' - x)$, we obtain

$$\hat{F}_i(x') \geq \hat{F}_i(x) - \left(v_{i0} + \sum_{j \in N} v_{ij}\, \hat{z}_{ij}(x) \right)^{\gamma_i} (x' - x),$$

which implies that $\hat{F}_i(\cdot)$ satisfies the subgradient inequality at x with a subgradient given by $-(v_{i0} + \sum_{j \in N} v_{ij}\, \hat{z}_{ij}(x))^{\gamma_i}$. In this case, by Theorem 3.2.6 in Bazaraa et al. (1993), $\hat{F}_i(x)$ is a convex function of x with a subgradient as given in the proposition. $\qquad\qquad\square$

Proposition 13 shows that we can obtain a subgradient of $\hat{F}_i(\cdot)$ at x by solving the problem

$$\max_{z_i \in [0,1]^n} F_i(z_i \mid x) = \max_{z_i \in [0,1]^n} \left\{ \left(v_{i0} + \sum_{j \in N} v_{ij}\, z_{ij} \right)^{\gamma_i} \left[\frac{\sum_{j \in N} r_{ij}\, v_{ij}\, z_{ij}}{v_{i0} + \sum_{j \in N} v_{ij}\, z_{ij}} - x \right] \right\} \tag{18}$$

at a fixed value of x. Thus, if we can solve the problem above in a tractable fashion, then we can also solve problem (17) through a standard cutting plane method. In the rest of this section, we focus on solving problem (18) at a fixed value of x.

Following the same argument in the proof of Lemma 6, we can show that there exists an optimal solution z_i^* to problem (18) that satisfies $z_{i1}^* = 1$, $z_{i2}^* = 1, \ldots, z_{i,k-1}^* = 1$, $z_{ik}^* \in [0,1]$, $z_{i,k+1}^* = 0, \ldots, z_{in}^* = 0$ for some $k = 1, \ldots, n$. In other words, if we define the vector $\eta^k = (\eta_1^k, \ldots, \eta_n^k) \in \Re^n$ such that $\eta_1^k = 1, \ldots, \eta_{k-1}^k = 1, \eta_k^k = 0, \ldots, \eta_n^k = 0$ and use e^k to denote the k-th unit vector in $\Re^n$, then an optimal solution to problem (18) is of the form $\eta^k + \rho\, e^k$ for some $\rho \in [0,1]$ and $k = 1, \ldots, n$. To find the best value for ρ, we can solve the problem $\max_{\rho \in [0,1]} F_i(\eta^k + \rho\, e^k \mid x)$, which is a scalar optimization problem. A simple check verifies that the first derivative of $F_i(\eta^k + \rho\, e^k \mid x)$ with respect to ρ vanishes at one point and the point at which the first derivative vanishes can be computed in closed-form fashion. Therefore, the optimal objective value of the problem $\max_{\rho \in [0,1]} F_i(\eta^k + \rho\, e^k \mid x)$ is attained either at one of the two end points of the interval $[0,1]$ or at the point where the first derivative of $F_i(\eta^k + \rho\, e^k \mid x)$ with respect to ρ vanishes. By checking the objective value of the problem $\max_{\rho \in [0,1]} F_i(\eta^k + \rho\, e^k \mid x)$ at these three points, we can easily solve this problem. In this case, the optimal objective value of problem (18) can be obtained by solving the problem $\max_{\rho \in [0,1]} F_i(\eta^k + \rho\, e^k \mid x)$ for all $k = 1, \ldots, n$ and picking the one that yields the largest optimal objective value.

8 Computational Experiments

In this section, we provide computational experiments to test the quality of the solutions that we obtain by focusing on various collections of candidate assortments. Our goal is to compare the performance of different collections of candidate assortments under different problem characteristics. We begin by describing the details of our experimental setup. Following this description, we give the findings from our computational experiments.

8.1 Experimental Setup

The results in Sections 4.2 and 6.1 suggest that the performance guarantee provided by the collections of assortments considered in these sections can depend on how much the revenues or the preference weights of the products in the same nest differ. Similarly, the results in Sections 4.2 and 6.2 show that the performance of the collections of assortments that we develop can depend on the magnitude of the dissimilarity parameters. In our computational experiments, we investigate how the performance of different collections of candidate assortments changes with various problem parameters.

Throughout this section, we consider test problems with possibly synergistic products and partially-captured nests, which correspond to the most general instances of our assortment problem. We build on the example in Section 4.1 to generate our test problems. Noting that $\varepsilon \leq 1$ is a small positive number, product 1 in this example has a large revenue but a small preference weight. Product 2 has a small revenue but a large preference weight. Product 3 has zero revenue and a moderate preference weight. We observe that although product 1 has a large revenue, its preference weight is too small to attract customers to its nest. In contrast, product 2 has a large preference weight and it can effectively attract customers to the nest, but if a customer is attracted to the nest, then the preference weight of product 2 is so large that the customer ends up buying product 2 with high probability. Since the revenue of product 2 is small, this outcome yields a small revenue. While product 3 has a revenue of zero, its moderate preference weight can attract customers to the nest, but its preference weight is not so large that once a customer is attracted to the nest, there is a reasonable probability that he can end up buying product 1. So, product 3 is essentially a loss leader, whose purpose is to attract customers to the nest. Recall that nested-by-revenue assortments can perform poorly in this example.

We proceed as follows to generate test problems with the same flavor as above, but with reasonably large numbers of nests and products. In all of our test problems, the number of nests is $m = 5$ and the number of products in each nest is $n = 25$. This results in problem sizes that correspond to applications arising in some revenue management and retail settings. We choose a positive parameter $\varepsilon \leq 1$ that characterizes the degree to which the revenues and the preference weights of the products in the same nest differ. We vary the parameter ε in our computational experiments. To come up with the revenues and the preference weights associated with the first $n-1$ products in nest i, we generate U_{ij} from the uniform distribution over $[0, 4]$ for all $j \in N \setminus \{n\}$. Also, we generate X_{ij} and Y_{ij} respectively from the uniform distributions over $[1, 10]$ and $[0.2, 1.8]$ for all $j \in N \setminus \{n\}$. In this case, we set the revenue of product j in nest i as $r_{ij} = \varepsilon^{U_{ij}} \times X_{ij}$, whereas we set the preference weight associated with product j in nest i as $v_{ij} = \varepsilon^{2-U_{ij}} \times Y_{ij}$. The reasoning behind our choice of revenues and preference weights is that if we multiply the revenues and the preference weights of products 1 or 2 in the example in Section 4.1, then we end up with a quantity that is of magnitude ε^2. If we multiply the revenues and the preference weights of the products we generate, then we obtain $\varepsilon^{U_{ij}} \times X_{ij} \times \varepsilon^{2-U_{ij}} \times Y_{ij}$, which is of magnitude ε^2 as well. If the generated value of U_{ij} turns out to be closer to zero, then we obtain a product that resembles product 1 in the example in Section 4.1, whereas if the generated value of U_{ij} turns out to be closer to four, then we obtain a product that resembles product 2. The role of X_{ij} and Y_{ij} is to introduce some noise in the revenues and the preference weights. To generate the revenue and

the preference weight associated with the last product n in nest i, we generate Y_{in} from the uniform distribution over $[0.2, 1.8]$ and set the revenue and the preference weight of this product respectively as $r_{in} = 0$ and $v_{in} = \varepsilon^{-1} \times Y_{in}$. We observe that the revenue and the preference weight of the last product are of the same magnitude as the revenue and the preference weight of product 3 in the example in Section 4.1. With this setup, it is possible to check that as ε gets smaller, there are larger differences between the revenues or the preference weights of the products in the same nest.

To come up with the dissimilarity parameter of nest i, we generate γ_i from the uniform distribution over $[\gamma^L, \gamma^U]$ for all $i \in M$, where γ^L and γ^U are parameters we vary. We set the preference weights of the no purchase options to $v_0 = 10$ and $v_{i0} = \varepsilon^{-4}$ for all $i \in M$. In our computational experiments, we vary ε over $\{0.6, 0.5, 0.4, 0.3\}$, whereas we vary $[\gamma^L, \gamma^U]$ over $\{[0.5, 1.5], [1.0, 2.0], [1.5, 2.5], [2.0, 3.0]\}$. In this way, we obtain 16 combinations of problem parameters. Following the approach described in the paragraph above, we generate 5,000 individual problem instances for each combination of problem parameters ε and $[\gamma^L, \gamma^U]$.

We test the performance of three different collections of assortments on each problem instance. The first collection of assortments is *nested-by-revenue assortments*. This collection for nest i corresponds to $\{N_{ij} : j \in N_+\}$. While nested-by-revenue assortments provide the performance guarantee in (6) for fully-captured nests, we are not able to give a performance guarantee for these assortments for the most general instances of our assortment problem. However, since they are intuitively appealing and easy to implement, it is useful to test their performance. We refer to the second collection of assortments as *nested-by-preference-and-revenue assortments*. This collection of assortments for nest i is given by $\{N_{ij}^k : k \in N, \ j = 0, \ldots, k\} \cup \{\{j\} : j \in N\}$. In Section 6.1, we show that this collection provides the performance guarantee given in (12) for the most general instances of our assortment problem. The third collection that we consider is referred to as *powers-of-δ assortments* and this collection for nest i is given by $\{\hat{S}_{il} : l = l_i^L, \ldots, l_i^U\} \cup \{\emptyset\}$. We work with powers-of-δ assortments in Section 6.2 and show that these assortments provide a performance guarantee of $\max_{i \in M} \delta^{2\gamma_i + 1}$. The computational work required to obtain the powers-of-δ assortments depends on the value of δ and we set $\delta = 1.25$ in all of our computational experiments. With this value of δ, we can generate all powers-of-δ assortments for a particular problem instance and find the best assortment within this class in one tenths of a second. We note that there are $1 + n$ nested-by-revenue assortments and $1 + n + n^2$ nested-by-preference-and-revenue assortments in each nest. So, both of these collections grow polynomially with n. In contrast, there are $2 + \log_\delta(v_i^U/v_i^L)$ powers-of-δ assortments in nest i and noting that $v_i^U = v_{i0} + \sum_{j \in N} v_{ij}$ and $v_i^L = v_{i0} + \min_{j \in N} v_{ij}$, the number of powers-of-δ assortments grows logarithmically with n.

8.2 Computational Results

We give our main computational results in Tables 2, 3 and 4, where Tables 2, 3 and 4 respectively show the performance of the nested-by-revenue, nested-by-preference-and-revenue and powers-of-δ assortments. In these tables, the first column shows the combination of problem parameters by using the tuple $([\gamma^L, \gamma^U], \varepsilon)$. Recall that we generate 5,000 individual problem instances for each combination of problem parameters. For each problem instance we generate, we solve problem (16) to obtain an upper

bound on the optimal expected revenue. We use UB^k to denote the upper bound on the optimal expected revenue we obtain for problem instance k. Furthermore, given a collection of candidate assortments, we find the best assortment that we can stitch together by focusing on these candidate assortments. In Table 2, the candidate assortments are the nested-by-revenue assortments, whereas in Tables 3 and 4, the candidate assortments are respectively the nested-by-preference-and-revenue and powers-of-δ assortments. For problem instance k, we use Best^k to denote the expected revenue provided by the best assortment we can find by focusing on a particular collection of candidate assortments. The second column in Tables 2, 3 and 4 shows the number of problem instances k for which we have $\mathsf{UB}^k > \mathsf{Best}^k$. These problem instances correspond to those where we are not able to establish the optimality of the best assortment we can find. The third column focuses on the problem instances for which we cannot establish the optimality of the best assortment we can find, and reports the average percent gap between UB^k and Best^k over these problem instances. In other words, using $\mathcal{K}$ to denote the set of problem instances $\{k = 1, \ldots, 5{,}000 : \mathsf{UB}^k > \mathsf{Best}^k\}$, the third column gives

$$\frac{1}{|\mathcal{K}|} \sum_{k \in \mathcal{K}} 100 \, \frac{\mathsf{UB}^k - \mathsf{Best}^k}{\mathsf{UB}^k}.$$

The third column can be interpreted as the estimate of the average optimality gap of the best assortment we can find given that we cannot verify the optimality of this assortment. This column gives only an estimate of the average optimality gap since we do not know the optimal expected revenue for our problem instances and we only have an upper bound on the optimal expected revenue. The fourth column in Tables 2, 3 and 4 gives the 99-th percentile of the gaps between UB^k and Best^k over all 5,000 problem instances. That is, the fourth column gives the 99-th percentile of the data $\{100(\mathsf{UB}^k - \mathsf{Best}^k)/\mathsf{UB}^k : k = 1, \ldots, 5{,}000\}$. The fifth column shows the largest percent gap between UB^k and Best^k over all 5,000 problem instances. The sixth column shows the average number of products per nest in the best assortment we find. Finally, the last two columns give a feel for how much the revenues and the preference weights of the products in the same nest differ. In particular, the seventh column gives the average ratio between the largest and the smallest revenues in a nest, averaged over all nests and all 5,000 problem instances. Naturally, we do not consider the loss leader product with a revenue of zero when computing the ratio between the largest and the smallest revenues in a nest. The eighth column gives the average ratio between the largest and the smallest preference weights in a nest, averaged over all nests and all 5,000 problem instances.

The results in Table 2 indicate that nested-by-revenue assortments can perform well. When the dissimilarity parameters of the nests take values over $[0.5, 1.5]$, we can verify that nested-by-revenue assortments are optimal in about half of the problem instances. Furthermore, the optimality gaps of these assortments never exceeds 0.81%. As $[\gamma^L, \gamma^U]$ shifts from $[0.5, 1.5]$ to $[2.0, 3.0]$, the average optimality gaps of nested-by-revenue assortments increase from 0.01% to 0.38%. The overall trend is that the performance of nested-by-revenue assortments can degrade as $[\gamma^L, \gamma^U]$ increases and the nest dissimilarity parameters become larger. Another trend we observe from Table 2 is that the optimality gaps of nested-by-revenue assortments tend to increase as ε gets smaller. For the problem instances with $[\gamma^L, \gamma^U] = [0.5, 1.5]$, the average optimality gaps of nested-by-revenue assortments increase from 0.01% to 0.05% as ε decreases from 0.6 to 0.3, whereas for the problem instances with $[\gamma^L, \gamma^U] = [2.0, 3.0]$, the

Prob. Param. $([\gamma^L, \gamma^U], \varepsilon)$	Sub. Cnt.	Avg. % Gap in Sub.	99-th Per. % Gap.	Larg. % Gap	Avg. Ass. Size	Avg. Rev. Rat.	Avg. Pref. Rat.
$([0.5, 1.5], 0.6)$	1,999	0.01	0.07	0.34	6.38	24.08	22.59
$([0.5, 1.5], 0.5)$	2,235	0.02	0.12	0.38	6.37	42.58	40.05
$([0.5, 1.5], 0.4)$	2,494	0.03	0.20	0.66	6.55	88.42	83.35
$([0.5, 1.5], 0.3)$	2,832	0.05	0.33	0.81	6.90	234.61	221.63
$([1.0, 2.0], 0.6)$	3,933	0.03	0.20	0.52	5.94	24.08	22.59
$([1.0, 2.0], 0.5)$	4,194	0.05	0.35	0.71	6.36	42.58	40.05
$([1.0, 2.0], 0.4)$	4,447	0.08	0.51	1.43	7.00	88.42	83.35
$([1.0, 2.0], 0.3)$	4,624	0.13	0.87	2.70	7.87	234.61	221.63
$([1.5, 2.5], 0.6)$	4,529	0.06	0.40	1.56	5.17	24.08	22.59
$([1.5, 2.5], 0.5)$	4,773	0.09	0.59	1.63	5.85	42.58	40.05
$([1.5, 2.5], 0.4)$	4,899	0.14	0.85	2.13	6.74	88.42	83.35
$([1.5, 3.0], 0.3)$	4,947	0.25	1.49	4.21	7.87	234.61	221.63
$([2.0, 3.0], 0.6)$	4,696	0.09	0.61	2.00	4.42	24.08	22.59
$([2.0, 3.0], 0.5)$	4,871	0.14	0.86	2.47	5.28	42.58	40.05
$([2.0, 3.0], 0.4)$	4,960	0.22	1.25	3.70	6.33	88.42	83.35
$([2.0, 3.0], 0.3)$	4,984	0.38	2.07	5.45	7.61	234.61	221.63

Table 2: Performance of nested-by-revenue assortments.

average optimality gaps increase from 0.09% to 0.38% as ε decreases from 0.6 to 0.3. We recall that as ε gets smaller, the difference between both the revenues and the preference weights of the products in the same nest gets larger. Thus, our results indicate that the performance nested-by-revenue assortments can degrade when we have drastic differences between the revenues and the preference weights. When we have $[\gamma^L, \gamma^U] = [2.0, 3.0]$ and $\varepsilon = 0.3$, there are problem instances where the estimated optimality gap of nested-by-revenue assortments can reach 5.45%. Nevertheless, we note that these problem instances seem unlikely to come up in practice as they involve products in the same nest with revenues differing by a factor of about 234 and preference weights differing by a factor of about 221.

Our findings in Table 3 indicate that nested-by-preference-and-revenue assortments provide small improvements over nested-by-revenue assortments when we consider the problem instances with $[\gamma^L, \gamma^U] = [0.5, 1.5]$. The performance of nested-by-revenue assortments is already quite satisfactory for these problem instances and it turns out to be difficult to improve over these assortments. As $[\gamma^L, \gamma^U]$ increases, however, nested-by-preference-and-revenue assortments can provide noticeable improvements over nested-by-revenue assortments. For the problem instances with $[\gamma^L, \gamma^U] = [2.0, 3.0]$, nested-by-preference-and-revenue assortments can decrease the largest optimality gaps of nested-by-revenue assortments by more than 2%. Finally, Table 4 shows that the performance of powers-of-δ assortments is not as good as the performance of the other two collections of assortments we consider. For almost all of the problem instances, we cannot verify the optimality of the best powers-of-δ assortment. Nevertheless, the largest optimality gaps of powers-of-δ assortments are still below 5%. Also, for the problem instances with $[\gamma^L, \gamma^U] = [2.0, 3.0]$ and $\varepsilon = 0.3$, comparing the largest optimality gaps of nested-by-revenue and powers-of-δ assortments indicates that there are problem instances where powers-of-δ assortments improve over nested-by-revenue assortments.

Our results suggest two problem parameters that can affect the performance of the collections of assortments we propose. The first parameter is the dissimilarity parameters of the nests. The second

Prob. Param. $([\gamma^L, \gamma^U], \varepsilon)$	Sub. Cnt.	Avg. % Gap in Sub.	99-th Per. % Gap.	Larg. % Gap	Avg. Ass. Size	Avg. Rev. Rat.	Avg. Pref. Rat.
$([0.5, 1.5], 0.6)$	1,999	0.01	0.06	0.20	6.38	24.08	22.59
$([0.5, 1.5], 0.5)$	2,235	0.02	0.11	0.31	6.37	42.58	40.05
$([0.5, 1.5], 0.4)$	2,494	0.03	0.18	0.48	6.55	88.42	83.35
$([0.5, 1.5], 0.3)$	2,832	0.04	0.28	0.78	6.90	234.61	221.63
$([1.0, 2.0], 0.6)$	3,933	0.03	0.18	0.51	5.94	24.08	22.59
$([1.0, 2.0], 0.5)$	4,194	0.04	0.27	0.71	6.37	42.58	40.05
$([1.0, 2.0], 0.4)$	4,447	0.07	0.39	0.92	7.02	88.42	83.35
$([1.0, 2.0], 0.3)$	4,624	0.11	0.65	1.31	7.89	234.61	221.63
$([1.5, 2.5], 0.6)$	4,529	0.05	0.34	0.79	5.18	24.08	22.59
$([1.5, 2.5], 0.5)$	4,773	0.08	0.47	1.03	5.87	42.58	40.05
$([1.5, 2.5], 0.4)$	4,899	0.12	0.68	2.13	6.78	88.42	83.35
$([1.5, 3.0], 0.3)$	4,947	0.19	1.03	2.42	7.92	234.61	221.63
$([2.0, 3.0], 0.6)$	4,696	0.08	0.51	1.24	4.43	24.08	22.59
$([2.0, 3.0], 0.5)$	4,871	0.12	0.68	1.82	5.31	42.58	40.05
$([2.0, 3.0], 0.4)$	4,960	0.18	0.88	2.20	6.38	88.42	83.35
$([2.0, 3.0], 0.3)$	4,984	0.29	1.33	3.26	7.68	234.61	221.63

Table 3: Performance of nested-by-preference-and-revenue assortments.

Prob. Param. $([\gamma^L, \gamma^U], \varepsilon)$	Sub. Cnt.	Avg. % Gap in Sub.	99-th Per. % Gap.	Larg. % Gap	Avg. Ass. Size	Avg. Rev. Rat.	Avg. Pref. Rat.
$([0.5, 1.5], 0.6)$	4,996	0.93	2.22	3.20	6.46	24.08	22.59
$([0.5, 1.5], 0.5)$	4,991	0.89	2.13	3.09	6.43	42.58	40.05
$([0.5, 1.5], 0.4)$	4,969	0.80	2.06	2.90	6.61	88.42	83.35
$([0.5, 1.5], 0.3)$	4,937	0.68	1.82	2.86	6.96	234.61	221.63
$([1.0, 2.0], 0.6)$	4,994	1.01	2.42	3.68	6.01	24.08	22.59
$([1.0, 2.0], 0.5)$	4,998	0.97	2.33	3.72	6.42	42.58	40.05
$([1.0, 2.0], 0.4)$	4,998	0.88	2.27	3.95	7.08	88.42	83.35
$([1.0, 2.0], 0.3)$	4,997	0.78	2.05	3.14	7.91	234.61	221.63
$([1.5, 2.5], 0.6)$	4,999	1.12	2.87	4.40	5.25	24.08	22.59
$([1.5, 2.5], 0.5)$	4,999	1.07	2.67	3.90	5.92	42.58	40.05
$([1.5, 2.5], 0.4)$	5,000	0.99	2.53	3.80	6.81	88.42	83.35
$([1.5, 3.0], 0.3)$	4,999	0.91	2.37	3.26	7.91	234.61	221.63
$([2.0, 3.0], 0.6)$	4,997	1.22	3.42	4.37	4.53	24.08	22.59
$([2.0, 3.0], 0.5)$	5,000	1.18	3.08	4.64	5.37	42.58	40.05
$([2.0, 3.0], 0.4)$	5,000	1.10	2.84	4.92	6.40	88.42	83.35
$([2.0, 3.0], 0.3)$	4,999	1.03	2.73	4.02	7.66	234.61	221.63

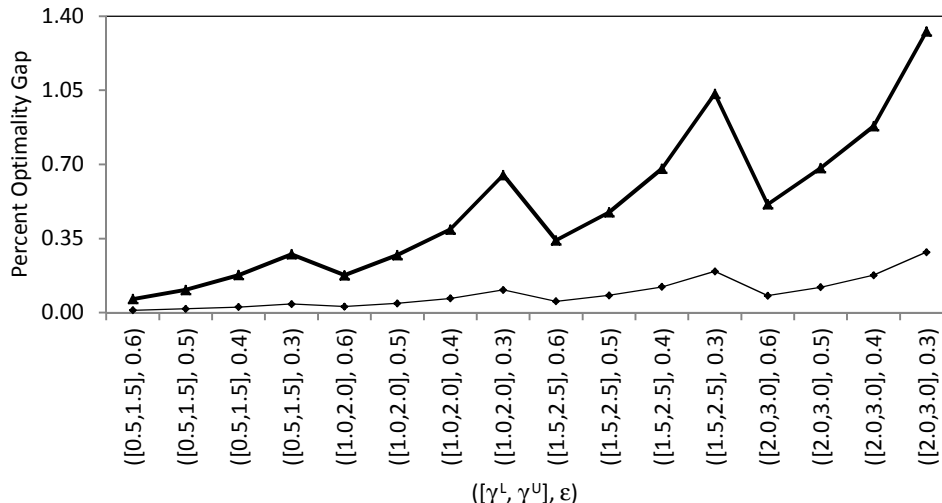Table 4: Performance of powers-of-δ assortments.

Figure 1: Optimality gaps for nested-by-preference-and-revenue assortments.

parameter is the degree to which the revenues or the preference weights of the products in the same nest differ. To make such trends clear, Figure 1 plots the optimality gaps for nested-by-preference-and-revenue assortments as a function of the nest dissimilarity parameters and the difference between the revenues and the preference weights of the products in the same nest. Our earlier results indicate that nested-by-preference-and-revenue assortments generally provide better performance than the other two collections and to conserve space, we focus only on these assortments in Figure 1. The horizontal axis in this figure shows the values of $[\gamma^L, \gamma^U]$ and ε for the 16 combinations of problem parameters. The thin data series plot the average optimality gap of nested-by-preference-and-revenue assortments, whereas the thick data series plot the 99-th percentile of the optimality gaps over the 5,000 problems instances generated for a certain combination of problem parameters. From the figure, we observe that as ε shifts from 0.6 to 0.3, keeping $[\gamma^L, \gamma^U]$ constant, the optimality gaps of nested-by-preference-and-revenue assortments get larger. Similarly, as $[\gamma^L, \gamma^U]$ shifts from $[0.5, 1.5]$ to $[2.0, 3.0]$, keeping ε constant, the optimality gaps of nested-by-preference-and-revenue assortments get larger as well.

In Tables 2, 3 and 4, we keep the width of the interval $[\gamma^L, \gamma^U]$ constant at one and systematically increase the value of γ^L. An interesting question is how our collections of assortments perform as we increase the width of the interval $[\gamma^L, \gamma^U]$. To answer this question, we fix γ^L at one and vary γ^U over $\{1.5, 2.0, 2.5, 3.0\}$. We fix ε at 0.3, which corresponds to the case with the largest optimality gaps in our earlier results. Table 5 shows our findings. The layout of this table is similar to that of Tables 2, 3 and 4. The three portions of this table focus on the performance of nested-by-revenue, nested-by-preference-and-revenue and powers-of-δ assortments. The results in Table 5 indicates that as γ^U increases and the nest dissimilarity parameters tend to take values significantly larger than one, the optimality gaps of all three collections of assortments get larger. Nevertheless, even with $\varepsilon = 0.3$, which corresponds to a case where the revenues and the preference weights of the products in the same nest differ respectively by factors of about 234 and 221, the 99-th percentile of the optimality gaps for the best collections of assortments do not exceed 1.40% and the largest optimality gaps do not exceed 3.85%.

36

Nested-by-revenue assortments

Prob. Param. $([\gamma^L, \gamma^U], \varepsilon)$	Sub. Cnt.	Avg. % Gap in Sub.	99-th Per. % Gap.	Larg. % Gap	Avg. Ass. Size	Avg. Rev. Rat.	Avg. Pref. Rat.
$([1.0, 1.5], 0.3)$	3,866	0.05	0.37	0.88	7.81	234.61	221.63
$([1.0, 2.0], 0.3)$	4,624	0.13	0.87	2.70	7.87	234.61	221.63
$([1.0, 2.5], 0.3)$	4,827	0.22	1.53	4.62	7.74	234.61	221.63
$([1.0, 3.0], 0.3)$	4,898	0.32	2.12	6.73	7.55	234.61	221.63

Nested-by-preference-and-revenue assortments

Prob. Param. $([\gamma^L, \gamma^U], \varepsilon)$	Sub. Cnt.	Avg. % Gap in Sub.	99-th Per. % Gap.	Larg. % Gap	Avg. Ass. Size	Avg. Rev. Rat.	Avg. Pref. Rat.
$([1.0, 1.5], 0.3)$	3,866	0.05	0.31	0.76	7.82	234.61	221.63
$([1.0, 2.0], 0.3)$	4,624	0.11	0.65	1.31	7.89	234.61	221.63
$([1.0, 2.5], 0.3)$	4,827	0.18	1.00	1.93	7.78	234.61	221.63
$([1.0, 3.0], 0.3)$	4,898	0.24	1.40	3.85	7.60	234.61	221.63

Powers-of-δ assortments

Prob. Param. $([\gamma^L, \gamma^U], \varepsilon)$	Sub. Cnt.	Avg. % Gap in Sub.	99-th Per. % Gap.	Larg. % Gap	Avg. Ass. Size	Avg. Rev. Rat.	Avg. Pref. Rat.
$([1.0, 1.5], 0.3)$	4,981	0.69	1.84	2.46	7.86	234.61	221.63
$([1.0, 2.0], 0.3)$	4,997	0.78	2.05	3.14	7.91	234.61	221.63
$([1.0, 2.5], 0.3)$	4,998	0.87	2.31	3.32	7.78	234.61	221.63
$([1.0, 3.0], 0.3)$	4,999	0.97	2.69	3.96	7.60	234.61	221.63

Table 5: Performance of the three different collections of assortments as a function of γ^U.

9 Conclusions

In this paper, we studied a class of assortment optimization problems under variants of the nested logit model. We showed that the problem is polynomially solvable when the dissimilarity parameters of the nests are less than one, and the customers always purchase a product within their selected nest. Relaxing either one of these assumptions renders the problem NP-hard. To deal with the NP-hard cases, we developed collections of assortments with worst-case performance guarantees. Furthermore, we formulated a tractable convex program whose optimal objective value provides an upper bound on the optimal expected revenue. In this case, we can compare the expected revenue provided by an assortment with the upper bound on the optimal expected revenue to get a feel for the optimality gap of the assortment. By following this approach, our computational experiments tested the performance of the collections of candidate assortments that we develop.

There are still interesting research questions within the context of assortment optimization under the nested logit model. In this paper, we assume that the dissimilarity parameters $(\gamma_1, \ldots, \gamma_m)$ of the nests and the preference weights $\{v_{ij} : i \in M, j \in N\}$, $\{v_{i0} : i \in M\}$ and v_0 of the products and the no purchase options are constants. As we explain at the end of Section 1, this assumption can be justified by deriving the nested logit model through a random utility-based choice model, where the means and correlation structure of the random utilities are assumed to be fixed. Naturally, we can obtain richer choice models by allowing the means and correlation structure of the random utilities to

depend on the offered assortment. For example, if a certain product is offered along with some others, then it may appear more attractive to customers, which can be modeled by allowing the mean utility of the product to depend on the whole set of offered products. Similarly, the correlation structure of the random utilities may also depend on which assortment is offered. When we allow the means or correlation structure of the random utilities to depend on the offered assortment, the preference weights or the dissimilarity parameters become dependent on the offered assortment as well.

The approach that we use in this paper has difficulties when working with such assortment-dependent preference weights or dissimilarity parameters. For example, consider a case where the preference weights of the products in nest i or the dissimilarity parameter of nest i depends on the assortment offered within this nest. In this case, one difficulty that we run into is that the performance guarantees in Sections 4.2, 5 and 6.1 are obtained by using various continuous relaxations, such as those on the right side of the second set of constraints in problems (7) and (10), but it is not clear how to construct and work with such continuous relaxations when we have assortment-dependent preference weights or dissimilarity parameters. Furthermore, the proofs of the results in these sections use the monotonicity, concavity and convexity properties of $V_i(S_i)^{\gamma_i}$ when viewed as a function of $V_i(S_i)$. For example, the proof of Theorem 10 uses the fact that $V_i(S_i)^{\gamma_i}$ and $V_i(S_i)^{1-\gamma_i}$ are both increasing in $V_i(S_i)$ when $\gamma_i \leq 1$, whereas the proof of Theorem 11 uses the fact that $V_i(S_i)^{1-\gamma_i}$ is a convex function of $V_i(S_i)$ when $\gamma_i \geq 1$. Similar monotonicity, concavity and convexity properties are used in the proofs of Theorems 4 and 7 as well. These properties naturally hold when the dissimilarity parameters are constants but it is not clear how to ensure the analogous monotonicity, concavity and convexity properties when the preference weights or the dissimilarity parameters depend on the offered assortment.

Another important research direction is to work with more general forms of the nested logit model, such as the mixed or cross nested logit models. In Section 2, we exploit the fact that the problem $\max_{(S_1,\ldots,S_m):S_i \subset N, i \in M} \sum_{i \in M} V_i(S_i)^{\gamma_i}(R_i(S_i) - x)$ decomposes by the nests. As a result, for a fixed value of the decision variable x, the second set of constraints in problem (3) separates by the nests and we can construct candidate assortments by considering each nest separately. Unfortunately, this separable structure is lost under more general forms of the nested logit model. Therefore, dealing with assortment-dependent preference weights or dissimilarity parameters and solving assortment problems under more general forms of the nested logit model remain open for further research.

References

Bazaraa, M. S., Sherali, H. D. and Shetty, C. M. (1993), *Nonlinear Programming: Theory and Algorithms*, second edn, John Wiley & Sons, Inc., New York.

Ben-Akiva, M. and Lerman, S. (1997), *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press in Transportation Studies, Cambridge, MA.

Borsch-Supan, A. (1990), 'On the compatibility of nested logit models with utility maximization', *Journal of Econometrics* **43**, 373–388.

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

Cachon, G., Terwiesch, C. and Xu, Y. (2005), 'Assortment planning in the presence of consumer search', *Manufacturing & Service Operations Management* **7**(4), 330–346.

Dong, L., Kouvelis, P. and Tian, Z. (2009), 'Dynamic pricing and inventory control of substitute products', *Manufacturing & Service Operations Management* **11**(2), 317–339.

Farias, V. F., Jagabathula, S. and Shah, D. (2011), 'A non-parametric approach to modeling choice with limited data', *Management Science* (to appear).

Frieze, A. M. and Clarke, M. R. B. (1984), 'Approximation algorithms for the m-dimensional 0-1 knapsack problem: Worst-case and probabilistic analyses', *European Journal of Operational Research* **15**, 100–109.

Gallego, G. and Wang, R. (2011), Multi-product price optimization and competition under the nested attraction model, Technical report, Columbia University.

Garey, M. and Johnson, D. (1979), *Computers and Intractability: A Guide to the Theory of NP-Completeness*, Freeman, San Francisco, CA.

Hanson, W. and Martin, K. (1996), 'Optimizing multinomial logit profit functions', *Management Science* **42**(7), 992–1003.

Kok, A. G., Fisher, M. L. and Vaidyanathan, R. (2008), Assortment planning: Review of literature and industry practice, *in* 'Retail Supply Chain Management', Kluwer Publishers.

Kok, A. G. and Xu, Y. (2011), 'Optimal and competitive assortments with endogenous pricing under hierarchical consumer choice models', *Management Science* **57**(9), 1546–1563.

Lee, B. (1999), 'Calling patterns and usage of residential toll service under self selecting tariffs', *Journal of Regulatory Economics* **16**, 45–81.

Li, H. and Huh, W. T. (2011), 'Pricing multiple products with the multinomial logit and nested models: Concavity and implications', *Manufacturing & Service Operations Management* (to appear).

Luce, R. D. (1959), *Individual Choice Behavior: A Theoretical Analysis*, Wiley, New York, NY.

Mahajan, S. and van Ryzin, G. (2001), 'Stocking retail assortments under dynamic customer substitution', *Operations Research* **49**(3), 334–351.

McFadden, D. (1974), Conditional logit analysis of qualitative choice behavior, *in* P. Zarembka, ed., '*Frontiers in Economics*', Academic Press, pp. 105–142.

McFadden, D. (1980), 'Econometric models for probabilistic choice among products', *The Journal of Business* **53**(3), S13–29.

McFadden, D. (1981), Econometric models of probabilistic choice, *in* C. F. Manski and D. McFadden, eds, '*Structural Analysis of Discrete Data with Economic Applications*', MIT Press, Cambridge, MA, pp. 198–272.

Mendez-Diaz, I., Bront, J. J. M., Vulcano, G. and Zabala, P. (2010), 'A branch-and-cut algorithm for the latent-class logit assortment problem', *Discrete Applied Mathematics* (to appear).

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

Rusmevichientong, P., Shmoys, D. B. and Topaloglu, H. (2010), Assortment optimization with mixtures of logits, Technical report, Cornell University, School of Operations Research and Information Engineering.
Available at http://legacy.orie.cornell.edu/~huseyin/publications/publications.html.

Rusmevichientong, P. and Topaloglu, H. (2011), Robust assortment optimization in revenue management under the multinomial logit choice model, Technical report, Cornell University, School of Operations Research and Information Engineering.
Available at http://legacy.orie.cornell.edu/~huseyin/publications/publications.html.

Ruszczynski, A. (2006), *Nonlinear Optimization*, Princeton University Press, Princeton, New Jersey.

Song, J.-S. and Xue, Z. (2007), Demand management and inventory control for substitutable products, Technical report, Duke University, Durham, NC.

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

Tiwari, P. and Hasegawa, H. (2004), 'Demand for housing in Tokyo: A discrete choice analysis', *Regional Studies* **38**(1), 27–42.

Train, K. (2003), *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge, UK.

Train, K. E., Ben-Akiva, M. and Atherton, T. (1989), 'Patterns and self-selecting tariffs', *The Review of Economics and Statistics* **71**(1), 62–73.

Train, K. E., McFadden, D. L. and Ben-Akiva, M. (1987), 'The demand for local telephone service: A fully discrete model of residential calling patterns and service choices', *The RAND Journal of Economics* **18**(1), 109–123.

van Ryzin, G. and Mahajan, S. (1999), 'On the relationship between inventory costs and variety benefits in retailassortments', *Management Science* **45**(11), 1496–1509.

Williams, H. C. W. L. (1977), 'On the formation of travel demand models and economic evaluation measures of user benefit', *Environment and Planning* **9**(3), 285–344.

Williamson, D. P. and Shmoys, D. B. (2011), *The Design of Approximation Algorithms*, Cambridge University Press, New York.

Yates, J. and Mackay, D. F. (2006), 'Discrete choice modeling of urban housing markets: A critical review and an application', *Urban Studies* **43**(3), 559–581.

A   Appendix: Omitted Proofs

A.1   Proof of Theorem 7

Noting that problem (3) is a relaxed version of problem (7), it is enough to show that $(\alpha\,\hat{x}, \alpha\,\hat{y})$ is a feasible solution to problem (7). We observe that since $(\hat{x}, \hat{y})$ is an optimal solution to problem (4) after replacing the collection of assortments $\{A_{it} : t \in \mathcal{T}_i\}$ in the second set of constraints with the nested-by-revenue assortments $\{N_{ij} : j \in N_+\}$, this solution satisfies the second set of constraints for nest i and the nested-by-revenue assortment $N_{i0} = \emptyset$. Noting that $V_i(\emptyset) = 0$, it follows that $\hat{y}_i \geq 0$ for all $i \in M$. In this case, the first constraint in problem (4) implies that $\hat{x} \geq 0$.

We fix an arbitrary nest i and let $\hat{z}_i$ be an optimal solution to the maximization problem on the right side of the second set of constraints in problem (7) when this maximization problem is solved at $x = \alpha\,\hat{x}$. By Lemma 6, $\hat{z}_i$ is of the form $\hat{z}_{i1} = 1, \hat{z}_{i2} = 1, \ldots, \hat{z}_{i,k-1} = 1, \hat{z}_{ik} \in [0, 1], \hat{z}_{i,k+1} = 0, \ldots, \hat{z}_{in} = 0$ for some $k = 1, \ldots, n$. We define ρ as $\rho = \hat{z}_{ik} \in [0, 1]$ and consider two cases.

Case 1. Assume that $k \geq 2$. We branch into two subcases.

Case 1.a. Noting that $(\hat{x}, \hat{y})$ is the optimal solution to problem (4) after replacing the collection of assortments $\{A_{it} : t \in \mathcal{T}_i\}$ in the second set of constraints with the nested-by-revenue assortments $\{N_{ij} : j \in N_+\}$, this solution satisfies the second set of constraints in problem (4) for nest i and the nested-by-revenue assortment $N_{ik} = \{1, 2, \ldots, k\}$. Thus, it holds that

$$\hat{y}_i \geq \Big(\sum_{j=1}^{k} v_{ij}\Big)^{\gamma_i} \left[\frac{\sum_{j=1}^{k} r_{ij}\,v_{ij}}{\sum_{j=1}^{k} v_{ij}} - \hat{x} \right].$$

For notational convenience, let $R_{ik'} = \sum_{j=1}^{k'} r_{ij}\,v_{ij}$ and $q_{ik'} = \sum_{j=1}^{k'} v_{ij}$ for all $k' = 1, \ldots, n$. Multiplying the inequality above by $\frac{R_{i,k-1} + r_{ik}\,v_{ik}\,\rho}{q_{i,k-1} + v_{ik}\,\rho}\,\frac{q_{ik}}{R_{ik}}$, we obtain

$$\frac{R_{i,k-1} + r_{ik}\,v_{ik}\,\rho}{q_{i,k-1} + v_{ik}\,\rho}\,\frac{q_{ik}}{R_{ik}}\,\hat{y}_i \geq q_{ik}^{\gamma_i} \left[\frac{R_{i,k-1} + r_{ik}\,v_{ik}\,\rho}{q_{i,k-1} + v_{ik}\,\rho} - \frac{R_{i,k-1} + r_{ik}\,v_{ik}\,\rho}{q_{i,k-1} + v_{ik}\,\rho}\,\frac{q_{ik}}{R_{ik}}\,\hat{x} \right]. \qquad (19)$$

It is simple to check that the first derivative of $\frac{R_{i,k-1} + r_{ik}\,v_{ik}\,\rho}{q_{i,k-1} + v_{ik}\,\rho}$ with respect to ρ has the same sign as $r_{ik}\,q_{i,k-1} - R_{i,k-1}$ and we have $r_{ik}\,q_{i,k-1} - R_{i,k-1} = \sum_{j=1}^{k-1}(r_{ik} - r_{ij})\,v_{ij} \leq 0$, where the last inequality is by the fact that $r_{i1} \geq r_{i2} \geq \ldots \geq r_{in}$. Thus, it follows that $\frac{R_{i,k-1} + r_{ik}\,v_{ik}\,\rho}{q_{i,k-1} + v_{ik}\,\rho}$ is decreasing in ρ so that is it bounded from above by $R_{i,k-1}/q_{i,k-1}$. Also, noting the definitions of $R_i(S_i)$, $R_{ik'}$ and $q_{ik'}$, we have $R_{ik'}/q_{ik'} = R_i(N_{ik'})$. In this case, we can bound the expression that multiplies $\hat{y}_i$ and $\hat{x}$ in (19) as

$$\frac{R_{i,k-1} + r_{ik}\,v_{ik}\,\rho}{q_{i,k-1} + v_{ik}\,\rho}\,\frac{q_{ik}}{R_{ik}} \leq \frac{R_{i,k-1}}{q_{i,k-1}}\,\frac{q_{ik}}{R_{ik}} = \frac{R_i(N_{i,k-1})}{R_i(N_{ik})}.$$

Letting $\alpha_{ik}^1 = R_i(N_{i,k-1})/R_i(N_{ik})$ for notational brevity, using the upper bound above in (19) and noting that $\hat{y}_i \geq 0$ and $\hat{x} \geq 0$, the inequality in (19) implies that

$$\alpha_{ik}^1\,\hat{y}_i \geq q_k^{\gamma_i} \left[\frac{R_{i,k-1} + r_{ik}\,v_{ik}\,\rho}{q_{i,k-1} + v_{ik}\,\rho} - \alpha_{ik}^1\,\hat{x} \right].$$

Finally, if we multiply the right side of the inequality above by $(q_{k-1} + v_{ik}\rho)^{\gamma_i}/q_k^{\gamma_i} \leq 1$, but not the left side, then the inequality is above still preserved and we have

$$\alpha_{ik}^1 \hat{y}_i \geq (q_{i,k-1} + v_{ik}\rho)^{\gamma_i} \left[\frac{R_{i,k-1} + r_{ik}\,v_{ik}\,\rho}{q_{i,k-1} + v_{ik}\,\rho} - \alpha_{ik}^1\,\hat{x} \right]. \tag{20}$$

Case 1.b. Noting that $(\hat{x}, \hat{y})$ is the optimal solution to problem (4) after replacing the collection of assortments $\{A_{it} : t \in \mathcal{T}_i\}$ in the second set of constraints with the nested-by-revenue assortments $\{N_{ij} : j \in N_+\}$, this solution also satisfies the second set of constraints in problem (4) for nest i and the nested-by-revenue assortment $N_{i,k-1} = \{1, 2, \ldots, k-1\}$. Thus, we obtain

$$\hat{y}_i \geq \left(\sum_{j=1}^{k-1} v_{ij}\right)^{\gamma_i} \left[\frac{\sum_{j=1}^{k-1} r_{ij}\,v_{ij}}{\sum_{j=1}^{k-1} v_{ij}} - \hat{x} \right] = q_{i,k-1}^{\gamma_i} \left[\frac{R_{i,k-1}}{q_{i,k-1}} - \hat{x} \right].$$

Multiplying the inequality above by $\frac{(q_{i,k-1}+v_{ik}\rho)^{\gamma_i}}{q_{i,k-1}^{\gamma_i}} \frac{R_{i,k-1}+r_{ik}\,v_{ik}\,\rho}{q_{i,k-1}+v_{ik}\,\rho} \frac{q_{i,k-1}}{R_{i,k-1}}$ and arranging the terms, we have

$$\frac{R_{i,k-1} + r_{ik}\,v_{ik}\,\rho}{(q_{i,k-1} + v_{ik}\,\rho)^{1-\gamma_i}} \frac{q_{i,k-1}^{1-\gamma_i}}{R_{i,k-1}} \hat{y}_i \geq (q_{i,k-1} + v_{ik}\,\rho)^{\gamma_i} \left[\frac{R_{i,k-1} + r_{ik}\,v_{ik}\,\rho}{q_{i,k-1} + v_{ik}\,\rho} - \frac{R_{i,k-1} + r_{ik}\,v_{ik}\,\rho}{q_{i,k-1} + v_{ik}\,\rho} \frac{q_{i,k-1}}{R_{i,k-1}} \hat{x} \right].$$

Since $\hat{x} \geq 0$, if we make the expression that multiplies $\hat{x}$ in the inequality above even larger by multiplying it by $\frac{(q_{i,k-1}+v_{ik}\rho)^{\gamma_i}}{q_{i,k-1}^{\gamma_i}} \geq 1$, then the inequality above is still preserved and we have

$$\frac{R_{i,k-1} + r_{ik}\,v_{ik}\,\rho}{(q_{i,k-1} + v_{ik}\,\rho)^{1-\gamma_i}} \frac{q_{i,k-1}^{1-\gamma_i}}{R_{i,k-1}} \hat{y}_i \geq (q_{i,k-1} + v_{ik}\,\rho)^{\gamma_i} \left[\frac{R_{i,k-1} + r_{ik}\,v_{ik}\,\rho}{q_{i,k-1} + v_{ik}\,\rho} - \frac{R_{i,k-1} + r_{ik}\,v_{ik}\,\rho}{(q_{i,k-1} + v_{ik}\,\rho)^{1-\gamma_i}} \frac{q_{i,k-1}^{1-\gamma_i}}{R_{i,k-1}} \hat{x} \right]. \tag{21}$$

It is simple to check that whenever the first derivative of $\frac{R_{i,k-1}+r_{ik}\,v_{ik}\,\rho}{(q_{i,k-1}+v_{ik}\,\rho)^{1-\gamma_i}}$ with respect to ρ vanishes, the second derivative takes a positive value. Therefore, this expression is maximized at either $\rho = 0$ or $\rho = 1$. In this case, we can bound the expression that multiplies $\hat{y}_i$ and $\hat{x}$ in (21) as

$$\frac{R_{i,k-1} + r_{ik}\,v_{ik}\,\rho}{(q_{i,k-1} + v_{ik}\,\rho)^{1-\gamma_i}} \frac{q_{i,k-1}^{1-\gamma_i}}{R_{i,k-1}} \leq 1 \vee \frac{R_{ik}}{q_{ik}^{1-\gamma_i}} \frac{q_{i,k-1}^{1-\gamma_i}}{R_{i,k-1}} = 1 \vee \frac{R_i(N_{ik})}{R_i(N_{i,k-1})} \frac{V_i(N_{ik})^{\gamma_i}}{V_i(N_{i,k-1})^{\gamma_i}},$$

where we use $a \vee b = \max\{a, b\}$. The two terms in the maximum operator on the right side of the first inequality are obtained by evaluating the expression on the left side of the inequality at $\rho = 0$ and $\rho = 1$. The equality above follows by noting that $R_{ik'}/q_{ik'} = R_i(N_{ik'})$ and $q_{ik'} = V_i(N_{ik'})$. Letting $\alpha_{ik}^2 = \frac{R_i(N_{ik})}{R_i(N_{i,k-1})} \frac{V_i(N_{ik})^{\gamma_i}}{V_i(N_{i,k-1})^{\gamma_i}}$ for notational convenience, using the upper bound above in (21) and noting that $\hat{y}_i \geq$ and $\hat{x} \geq 0$, the inequality in (21) implies that

$$(1 \vee \alpha_{ik}^2)\,\hat{y}_i \geq (q_{i,k-1} + v_{ik}\,\rho)^{\gamma_i} \left[\frac{R_{i,k-1} + r_{ik}\,v_{ik}\,\rho}{q_{i,k-1} + v_{ik}\,\rho} - (1 \vee \alpha_{ik}^2)\,\hat{x} \right]. \tag{22}$$

Putting Cases 1.a and 1.b together, we observe that $\hat{y}_i$ and $\hat{x}$ satisfy both (20) and (22), in which case, they must also satisfy

$$[\alpha_{ik}^1 \wedge (1 \vee \alpha_{ik}^2)]\,\hat{y}_i \geq (q_{i,k-1} + v_{ik}\,\rho)^{\gamma_i} \left[\frac{R_{i,k-1} + r_{ik}\,v_{ik}\,\rho}{q_{i,k-1} + v_{ik}\,\rho} - [\alpha_{ik}^1 \wedge (1 \vee \alpha_{ik}^2)]\,\hat{x} \right]. \tag{23}$$

Lemma 14 below shows that $\alpha \geq 1$ for the value of α given in (6). The proof of that lemma also shows that $\alpha_{ik}^1 = R_i(N_{ik})/R_i(N_{i,k-1}) \geq 1$. By the definitions of α, α_{ik}^1 and α_{ik}^2, we obtain $\alpha \geq \alpha_{ik}^1 \wedge \alpha_{ik}^2$ as well. In this case, we have

$$\alpha_{ik}^1 \wedge (1 \vee \alpha_{ik}^2) = (\alpha_{ik}^1 \wedge 1) \vee (\alpha_{ik}^1 \wedge \alpha_{ik}^2) = 1 \vee (\alpha_{ik}^1 \wedge \alpha_{ik}^2) \leq \alpha.$$

Thus, replacing the expression that multiplies $\hat{y}_i$ and $\hat{x}$ in (23) with an even larger expression α, the inequality is still preserved and we obtain

$$\alpha \, \hat{y}_i \geq (q_{i,k-1} + v_{ik} \, \rho)^{\gamma_i} \left[\frac{R_{i,k-1} + r_{ik} \, v_{ik} \, \rho}{q_{i,k-1} + v_{ik} \, \rho} - \alpha \, \hat{x} \right]. \tag{24}$$

Case 2. Assume that $k = 1$. Since $(\hat{x}, \hat{y})$ is the optimal solution to problem (4) after replacing the collection of assortments $\{A_{it} : t \in \mathcal{T}_i\}$ in the second set of constraints with the nested-by-revenue assortments $\{N_{ij} : j \in N_+\}$, this solution satisfies the second set of constraints in problem (4) for nest i and the nested-by-revenue assortment $N_{i1} = \{1\}$. Therefore, we have $\hat{y}_i \geq v_{i1}^{\gamma_i} \left[\frac{r_{i1} v_{i1}}{v_{i1}} - \hat{x} \right]$. Since $\alpha \geq 1$, $\hat{y}_i \geq 0$ and $\hat{x} \geq 0$, this inequality yields $\alpha \, \hat{y}_i \geq v_{i1}^{\gamma_i} \left[\frac{r_{i1} v_{i1}}{v_{i1}} - \alpha \, \hat{x} \right]$. If we multiply the right side of the last inequality by $\rho^{\gamma_i} \leq 1$, but not the left side, then the inequality is still preserved and we obtain

$$\alpha \, \hat{y}_i \geq (v_{i1} \, \rho)^{\gamma_i} \left[\frac{r_{i1} \, v_{i1} \, \rho}{v_{i1} \, \rho} - \alpha \, \hat{x} \right]. \tag{25}$$

Putting Cases 1 and 2 together, we succinctly write the inequalities in (24) and (25) as

$$\alpha \, \hat{y}_i \geq \left(\sum_{j=1}^{k-1} v_{ij} \, \hat{z}_{ij} + v_{ik} \, \hat{z}_{ik} \right)^{\gamma_i} \left[\frac{\sum_{j=1}^{k-1} r_{ij} \, v_{ij} \, \hat{z}_{ij} + r_{ik} \, v_{ik} \, \hat{z}_{ik}}{\sum_{j=1}^{k-1} v_{ij} \, \hat{z}_{ij} + v_{ik} \, \hat{z}_{ik}} - \alpha \, \hat{x} \right]$$

$$= \max_{z_i \in [0,1]^n} \left\{ \left(\sum_{j \in N} v_{ij} \, z_{ij} \right)^{\gamma_i} \left[\frac{\sum_{j \in N} r_{ij} \, v_{ij} \, z_{ij}}{\sum_{j \in N} v_{ij} \, z_{ij}} - \alpha \, \hat{x} \right] \right\},$$

where the equality follows from the definition of $\hat{z}_i$. Since the choice of nest i is arbitrary, the inequality above holds for all $i \in M$, which implies that the solution $(\alpha \, \hat{x}, \alpha \, \hat{y})$ satisfies the second set of constraints in problem (7). Since $(\hat{x}, \hat{y})$ is an optimal solution to problem (4), we have $v_0 \, \hat{x} \geq \sum_{i \in M} \hat{y}_i$, which implies that $v_0 \, \alpha \, \hat{x} \geq \sum_{i \in M} \alpha \, \hat{y}_i$. Therefore, the solution $(\alpha \, \hat{x}, \alpha \, \hat{y})$ satisfies the first constraint in problem (7) as well and we obtain the desired result. $\qquad \square$

Lemma 14 *Using α to denote the expression in (6), if we have $\gamma_i > 1$ for some $i \in M$, then $\alpha \geq 1$.*

Proof. Noting that $R_i(N_{ij}) = \sum_{k=1}^j r_{ik} \, v_{ik} / \sum_{k=1}^j v_{ik}$, $R_{ij}(N_{ij})$ is the weighted average of the revenues of the first j products in nest i. Since $r_{i1} \geq r_{i2} \geq \ldots \geq r_{in}$, it follows that $R_i(N_{i,j-1}) \geq R_i(N_{ij})$. On the other hand, since $\gamma_i > 1$, we have

$$R_i(N_{ij}) \, V_i(N_{ij})^{\gamma_i} = \frac{\sum_{k=1}^j r_{ij} \, v_{ij}}{\sum_{k=1}^j v_{ij}} \left(\sum_{k=1}^j v_{ij} \right)^{\gamma_i} \geq \frac{\sum_{k=1}^{j-1} r_{ij} \, v_{ij}}{\sum_{k=1}^{j-1} v_{ij}} \left(\sum_{k=1}^{j-1} v_{ij} \right)^{\gamma_i} = R_i(N_{i,j-1}) \, V_i(N_{i,j-1})^{\gamma_i}.$$

Therefore, both terms of the minimum operator in the expression in (6) for nest i are at least one, which implies that α is also at least one. $\qquad \square$

A.2 Proof of Theorem 8

Assume that we are given any instance of the partition problem with sizes $(c_1, \ldots, c_n)$ and $\sum_{j=1}^{n} c_j = 2T$. We define an instance of the assortment feasibility problem as follows. There are two nests. The preference weight for the option of not choosing any of the nests is $v_0 = 0$. The dissimilarity parameters of the two nests are $\gamma_1 = \gamma_2 = 1/2$. For the first nest, the preference weight of the no purchase option is $v_{10} = 2$. This nest has only one product in it. The revenue and the preference weight associated with this product are $r_{11} = 2(T+1)(T+3)$ and $v_{11} = 2(2T+1)$. For the second nest, the preference weight of the no purchase option is $v_{20} = 1$. The second nest has n products in it. The revenues of the products in the second nest are identical and they are given by $r_{2j} = (T+1)(2T+1)$ for all $j = 1, \ldots, n$. The preference weights of the products in the second nest are given by $v_{2j} = c_j$ for all $j = 1, \ldots, n$. We set the expected revenue threshold in the assortment feasibility problem as $K = (T+2)(2T+1)$.

The first observation that if we only offer the product in the first nest, then the expected revenue we generate from the first nest is

$$R_1(\{1\}) = \frac{r_{11} \, v_{11}}{v_{01} + v_{11}} = \frac{2(T+1)(T+3) \, 2(2T+1)}{2 + 2(2T+1)} = (T+3)(2T+1),$$

which is larger than the revenues of the products in the second nest. Thus, if we want to get the largest possible expected revenue, then it is always optimal to offer the product in the first nest. Therefore, the only question for the assortment feasibility problem is to choose a subset S among the products in the second nest that makes sure that we obtain an expected revenue of $K = (T+2)(2T+1)$ or more. If we offer a subset S of the products in the second nest together with the product in the first nest, then the expected revenue is $Q_1(\{1\}, S) \, R_1(\{1\}) + Q_2(\{1\}, S) \, R_2(S)$, which evaluates to

$$\frac{\sqrt{2 + 2(2T+1)}}{\sqrt{2 + 2(2T+1)} + \sqrt{1 + \sum_{j \in S} c_j}} \, \frac{2(T+1)(T+3) \, 2(2T+1)}{2 + 2(2T+1)}$$

$$+ \frac{\sqrt{1 + \sum_{j \in S} c_j}}{\sqrt{2 + 2(2T+1)} + \sqrt{1 + \sum_{j \in S} c_j}} \, \frac{(T+1)(2T+1) \sum_{j \in S} c_j}{1 + \sum_{j \in S} c_j}.$$

Thus, arranging the terms in the expression above, the assortment feasibility problem asks the question of whether there is a subset S such that

$$\frac{\dfrac{2(T+1)(T+3) \, 2(2T+1)}{\sqrt{2 + 2(2T+1)}} + (T+1)(2T+1) \dfrac{\sum_{j \in S} c_j}{\sqrt{1 + \sum_{j \in S} c_j}}}{\sqrt{2 + 2(2T+1)} + \sqrt{1 + \sum_{j \in S} c_j}} \geq (T+2)(2T+1).$$

If we cancel the terms in the first fraction in the numerator on the left side and move the denominator to the right, then the inequality above is equivalent to

$$2(T+3)(2T+1)\sqrt{T+1} + (T+1)(2T+1) \frac{\sum_{j \in S} c_j}{\sqrt{1 + \sum_{j \in S} c_j}}$$

$$\geq 2(T+2)(2T+1)\sqrt{T+1} + (T+2)(2T+1)\sqrt{1 + \sum_{j \in S} c_j}.$$

Canceling the term $2T + 1$ from both sides of the inequality above, multiplying the inequality by $\sqrt{1 + \sum_{j \in S} c_j}$ and adding and subtracting one from the term $\sum_{j \in S} c_j$, the inequality above can be written as

$$2(T+3)\sqrt{T+1}\sqrt{1 + \sum_{j \in S} c_j} + (T+1)\Big(1 + \sum_{j \in S} c_j\Big) - (T+1)$$
$$\geq 2(T+2)\sqrt{T+1}\sqrt{1 + \sum_{j \in S} c_j} + (T+2)\Big(1 + \sum_{j \in S} c_j\Big).$$

Finally, collecting all of the terms to the right, the last inequality becomes

$$\Big(1 + \sum_{j \in S} c_j\Big) - 2\sqrt{T+1}\sqrt{1 + \sum_{j \in S} c_j} + (T+1) \leq 0.$$

Since the last inequality is equivalent to $(\sqrt{1 + \sum_{j \in S} c_j} - \sqrt{T+1})^2 \leq 0$, there exists an assortment with an expected revenue of $K = (T+2)(2T+1)$ or more if and only if there exists a subset S with $(\sqrt{1 + \sum_{j \in S} c_j} - \sqrt{T+1})^2 \leq 0$. However, the only way for the last inequality to hold is to have $\sum_{j \in S} c_j = T$. Therefore, finding an assortment that yields an expected revenue of K or more is equivalent to finding a subset S that satisfies $\sum_{j \in S} c_j = T$ and the latter statement is precisely what the partition problem is interested in. $\qquad \square$

A.3   Proof of Theorem 10

Since problem (10) is equivalent to problem (3), it is enough to show that $(2\,\hat{x}, 2\,\hat{y})$ is a feasible solution to problem (10). First, note that $\hat{x} \geq 0$. To see this claim, if $\hat{x} < 0$, then the right sides of the second set of constraints in problem (4) are strictly positive for nonempty assortments so that $\hat{y}_i > 0$ for all $i \in M$. In this case, $(\hat{x}, \hat{y})$ cannot satisfy the first constraint in problem (4), establishing the claim.

We fix an arbitrary nest i and let $\hat{\epsilon}_i$ be the optimal solution to the maximization problem on the right side of the second set of constraints in problem (10) when this maximization problem is solved at $x = 2\,\hat{x}$. Finally, let $\hat{z}_i(\hat{\epsilon}_i)$ be the optimal solution to problem (11) when this continuous knapsack problem is solved at $\epsilon_i = \hat{\epsilon}_i$. We consider two cases.

Case 1. Assume that the solution $\hat{z}_i(\hat{\epsilon}_i)$ has exactly one fractional component. We denote this factional component by $k \in N$. Since $(\hat{x}, \hat{y})$ is the optimal solution to problem (4) after replacing the collection assortments $\{A_{it} : t \in \mathcal{T}_i\}$ in the second set of constraints with $\{\hat{S}_i(\epsilon_i) : \epsilon_i \in [0, \infty]\} \cup \{\{j\} : j \in N\}$, the solution $(\hat{x}, \hat{y})$ satisfies the second set of constraints in problem (4) for nest i and the assortment $\hat{S}_i(\hat{\epsilon}_i)$ and we obtain

$$\hat{y}_i \geq V_i(\hat{S}_i(\hat{\epsilon}_i))^{\gamma_i}(R_i(\hat{S}_i(\hat{\epsilon}_i)) - \hat{x})$$
$$= \frac{\sum_{j \in \hat{S}_i(\hat{\epsilon}_i)} r_{ij}\, v_{ij}}{(v_{i0} + \sum_{j \in \hat{S}_i(\hat{\epsilon}_i)} v_{ij})^{1-\gamma_i}} - \Big(v_{i0} + \sum_{j \in \hat{S}_i(\hat{\epsilon}_i)} v_{ij}\Big)^{\gamma_i} \hat{x} \geq \frac{\sum_{j \in \hat{S}_i(\hat{\epsilon}_i)} r_{ij}\, v_{ij}}{(v_{i0} + \hat{\epsilon}_i)^{1-\gamma_i}} - (v_{i0} + \hat{\epsilon}_i)^{\gamma_i}\,\hat{x}, \quad (26)$$

where the second inequality above follows by $\gamma_i \leq 1$ and noting that we have $\sum_{j \in \hat{S}_i(\hat{\epsilon}_i)} v_{ij} \leq \sum_{j \in N} v_{ij}\, \hat{z}_{ij}(\hat{\epsilon}_i) \leq \hat{\epsilon}_i$ by the definitions of $\hat{S}_i(\hat{\epsilon}_i)$ and $\hat{z}_i(\hat{\epsilon}_i)$. Similarly, the solution $(\hat{x}, \hat{y})$ satisfies the

second set of constraints in problem (4) for nest i and the singleton assortment $\{k\}$ so that

$$\hat{y}_i \geq V_i(\{k\})^{\gamma_i}(R_i(\{k\}) - \hat{x}) = \frac{r_{ik}\,v_{ik}}{(v_{i0}+v_{ik})^{1-\gamma_i}} - (v_{i0}+v_{ik})^{\gamma_i}\,\hat{x} \geq \frac{r_{ik}\,v_{ik}}{(v_{i0}+\hat{\epsilon}_i)^{1-\gamma_i}} - (v_{i0}+\hat{\epsilon}_i)^{\gamma_i}\,\hat{x}, \quad (27)$$

where the second inequality above follows from the fact that we must have $v_{ik} \leq \hat{\epsilon}_i$ for $\hat{z}_{ik}(\hat{\epsilon}_i)$ to take a fractional value. Since $\hat{S}_i(\hat{\epsilon}_i)$ includes all strictly positive and integer-valued components of $\hat{z}_i(\hat{\epsilon}_i)$ and k is the only component of $\hat{z}_i(\hat{\epsilon}_i)$ that takes a fractional value, we have $\sum_{j\in\hat{S}_i(\hat{\epsilon}_i)} r_{ij}\,v_{ij} + r_{ik}\,v_{ik} \geq \sum_{j\in N} r_{ij}\,v_{ij}\,\hat{z}_{ij}(\hat{\epsilon}_i) = \hat{K}_i(\hat{\epsilon}_i)$, where the equality follows by the definition of $\hat{z}_i(\hat{\epsilon}_i)$. Using this relationship and adding (26) and (27), we have

$$2\,\hat{y}_i \geq \frac{\sum_{j\in\hat{S}_i(\hat{\epsilon}_i)} r_{ij}\,v_{ij} + r_{ik}\,v_{ik}}{(v_{i0}+\hat{\epsilon}_i)^{1-\gamma_i}} - (v_{i0}+\hat{\epsilon}_i)^{\gamma_i}\,2\,\hat{x} \geq \frac{\hat{K}_i(\hat{\epsilon}_i)}{(v_{i0}+\hat{\epsilon}_i)^{1-\gamma_i}} - (v_{i0}+\hat{\epsilon}_i)^{\gamma_i}\,2\,\hat{x}$$

$$\geq \frac{K_i(\hat{\epsilon}_i)}{(v_{i0}+\hat{\epsilon}_i)^{1-\gamma_i}} - (v_{i0}+\hat{\epsilon}_i)^{\gamma_i}\,2\,\hat{x} = \max_{\epsilon_i \geq 0}\left\{(v_{i0}+\epsilon_i)^{\gamma_i}\left[\frac{K_i(\epsilon_i)}{v_{i0}+\epsilon_i} - 2\,\hat{x}\right]\right\}, \quad (28)$$

where the last inequality follows from the fact that problem (11) is a relaxation of problem (9) and the last equality follows from the definition of $\hat{\epsilon}_i$.

Case 2. Assume that the solution $\hat{z}_i(\hat{\epsilon}_i)$ does not have any fractional components. In this case, $\hat{S}_i(\hat{\epsilon}_i)$ includes all strictly positive components of $\hat{z}_i(\hat{\epsilon}_i)$ and we obtain $\sum_{j\in\hat{S}_i(\hat{\epsilon}_i)} r_{ij}\,v_{ij} = \sum_{j\in N} r_{ij}\,v_{ij}\,\hat{z}_{ij}(\hat{\epsilon}_i) = \hat{K}_i(\hat{\epsilon}_i)$. Using this relationship and following the same argument that we used to obtain (26) in the first case, we have

$$\hat{y}_i \geq \frac{\sum_{j\in\hat{S}_i(\hat{\epsilon}_i)} r_{ij}\,v_{ij}}{(v_{i0}+\hat{\epsilon}_i)^{1-\gamma_i}} - (v_{i0}+\hat{\epsilon}_i)^{\gamma_i}\,\hat{x} \geq \frac{K_i(\hat{\epsilon}_i)}{(v_{i0}+\hat{\epsilon}_i)^{1-\gamma_i}} - (v_{i0}+\hat{\epsilon}_i)^{\gamma_i}\,\hat{x}.$$

Multiplying the inequality above by two, we obtain

$$2\,\hat{y}_i \geq 2\,\frac{K_i(\hat{\epsilon}_i)}{(v_{i0}+\hat{\epsilon}_i)^{1-\gamma_i}} - (v_{i0}+\hat{\epsilon}_i)^{\gamma_i}\,2\,\hat{x}$$

$$\geq \frac{K_i(\hat{\epsilon}_i)}{(v_{i0}+\hat{\epsilon}_i)^{1-\gamma_i}} - (v_{i0}+\hat{\epsilon}_i)^{\gamma_i}\,2\,\hat{x} = \max_{\epsilon_i \geq 0}\left\{(v_{i0}+\epsilon_i)^{\gamma_i}\left[\frac{K_i(\epsilon_i)}{v_{i0}+\epsilon_i} - 2\,\hat{x}\right]\right\}, \quad (29)$$

where the last inequality follows from the definition of $\hat{\epsilon}_i$.

Collecting (28) and (29) in the two cases, the solution $(2\,\hat{x}, 2\,\hat{y})$ satisfies the second set of constraints for nest i in problem (10). Noting that our choice of nest i is arbitrary, the second set of constraints in problem (10) is satisfied by the solution $(2\,\hat{x}, 2\,\hat{y})$. Finally, since the solution $(\hat{x}, \hat{y})$ is optimal to problem (4), we have $v_0\,\hat{x} \geq \sum_{i\in M} \hat{y}_i$, which implies that $v_0\,2\,\hat{x} \geq \sum_{i\in M} 2\,\hat{y}_i$. Therefore, the solution $(2\,\hat{x}, 2\,\hat{y})$ satisfies the first constraint in problem (10) as well and we obtain the desired result. $\square$

A.4 Proof of Theorem 11

The proof follows from a reasoning similar to those in the proofs of Theorems 7 and 10, but there are some key new points. Following an argument similar to the one at the beginning of the proof of Theorem 10, we have $\hat{x} \geq 0$. We fix an arbitrary nest i and consider three cases.

Case 1. Assume that $\gamma_i > 1$ and $\hat{y}_i \geq 0$. We let $\hat{z}_i$ be an optimal solution to the problem

$$\max_{z_i \in [0,1]^n} \left\{ \left(v_{i0} + \sum_{j \in N} v_{ij}\, z_{ij} \right)^{\gamma_i} \left[\frac{\sum_{j \in N} r_{ij}\, v_{ij}\, z_{ij}}{v_{i0} + \sum_{j \in N} v_{ij}\, z_{ij}} - \beta\,\hat{x} \right] \right\}.$$

Using the same idea in the proof of Lemma 6, we can show that $\hat{z}_i$ is of the form $\hat{z}_{i1} = 1$, $\hat{z}_{i2} = 1, \ldots, \hat{z}_{i,k-1} = 1, \hat{z}_{ik} \in [0,1], \hat{z}_{i,k+1} = 0, \ldots, \hat{z}_{in} = 0$ for some $k = 1, \ldots, n$. We define ρ as $\rho = \hat{z}_{ik} \in [0,1]$ and branch into two subcases.

Case 1.a. Assume that $k \geq 2$. Noting that $(\hat{x}, \hat{y})$ is the optimal solution to problem (4) after replacing the collection of assortments $\{A_{it} : t \in \mathcal{T}_i\}$ in the second set of constraints with $\{N_{ij}^{k'} : k' \in N,\; j = 0, \ldots, k'\} \cup \{\{j\} : j \in N\}$, this solution satisfies the second set of constraints in problem (4) for nest i and the assortment $N_{ik}^n = \{1, 2, \ldots, k\}$. Thus, it holds that

$$\hat{y}_i \geq \left(v_{i0} + \sum_{j=1}^{k} v_{ij} \right)^{\gamma_i} \left[\frac{\sum_{j=1}^{k} r_{ij}\, v_{ij}}{v_{i0} + \sum_{j=1}^{k} v_{ij}} - \hat{x} \right].$$

For notational convenience, let $R_{ik'} = \sum_{j=1}^{k'} r_{ij}\, v_{ij}$ and $q_{ik'} = \sum_{j=1}^{k'} v_{ij}$ for all $k' = 1, \ldots, n$. The expression in the numerator of the fraction above corresponds to R_{ik} and if we replace the expression R_{ik} with the smaller expression $R_{i,k-1} + r_{ik}\, v_{ik}\, \rho$, then the inequality above still holds. Also, since $\beta \geq 1$, we can increase $\hat{y}_i$ on the left side to $\beta\,\hat{y}_i$ and the inequality still holds. Thus, the inequality above yields $\beta\,\hat{y}_i \geq (v_{i0} + q_{ik})^{\gamma_i - 1}\, (R_{i,k-1} + r_{ik}\, v_{ik}\, \rho - (v_{i0} + q_{ik})\,\hat{x})$. Since $\hat{y}_i \geq 0$, if we multiply the right side of the last inequality by $(v_{i0} + q_{i,k-1} + v_{ik}\, \rho)^{\gamma_i - 1}/(v_{i0} + q_{ik})^{\gamma_i - 1} \leq 1$, but not the left side, then the last inequality is preserved and we obtain $\beta\,\hat{y}_i \geq (v_{i0} + q_{i,k-1} + v_{ik}\, \rho)^{\gamma_i - 1}\, (R_{i,k-1} + r_{ik}\, v_{ik}\, \rho - (v_{i0} + q_{ik})\,\hat{x})$. We write this inequality as

$$\beta\,\hat{y}_i \geq (v_{i0} + q_{i,k-1} + v_{ik}\, \rho)^{\gamma_i} \left[\frac{R_{i,k-1} + r_{ik}\, v_{ik}\, \rho}{v_{i0} + q_{i,k-1} + v_{ik}\, \rho} - \frac{v_{i0} + q_{ik}}{v_{i0} + q_{i,k-1} + v_{ik}\, \rho}\,\hat{x} \right]$$

$$\geq (v_{i0} + q_{i,k-1} + v_{ik}\, \rho)^{\gamma_i} \left[\frac{R_{i,k-1} + r_{ik}\, v_{ik}\, \rho}{v_{i0} + q_{i,k-1} + v_{ik}\, \rho} - \beta\,\hat{x} \right], \quad (30)$$

where the second inequality above follows by $(v_{i0} + q_{ik})/(v_{i0} + q_{i,k-1} + v_{ik}\, \rho) \leq V_i(N_{ik}^n)/V_i(N_{i,k-1}^n) \leq \beta$.

Case 1.b. Assume that $k = 1$. In this case, if nest i is a partially-captured nest with $v_{i0} > 0$, then we can use the same approach in Case 1.a to show that the inequality in (30) is satisfied with $k = 1$. On the other hand, if nest i is a fully-captured nest with $v_{i0} = 0$, then we can use the same approach in Case 2 in the proof of Theorem 7 to show that the inequality in (30) is satisfied with $k = 1$. Thus, the inequality in (30) holds under Case 1.b as well. Putting Cases 1.a and 1.b together, we have

$$\beta\,\hat{y}_i \geq (v_{i0} + q_{i,k-1} + v_{ik}\, \rho)^{\gamma_i} \left[\frac{R_{i,k-1} + r_{ik}\, v_{ik}\, \rho}{v_{i0} + q_{i,k-1} + v_{ik}\, \rho} - \beta\,\hat{x} \right]$$

$$= \left(v_{i0} + \sum_{j \in N} v_{ij}\, \hat{z}_{ij} \right)^{\gamma_i} \left[\frac{\sum_{j \in N} r_{ij}\, v_{ij}\, \hat{z}_{ij}}{v_{i0} + \sum_{j \in N} v_{ij}\, \hat{z}_{ij}} - \beta\,\hat{x} \right] \geq V_i(S_i)^{\gamma_i} (R_i(S_i) - \beta\,\hat{x})$$

for all $S_i \subset N$, where the second inequality follows from the definition of $\hat{z}_i$. Therefore, the solution $(\beta\,\hat{x}, \beta\,\hat{y})$ satisfies the second set of constraints for nest i in problem (3).

Case 2. Assume that $\gamma_i > 1$ and $\hat{y}_i < 0$. We let $\hat{z}_i$ be an optimal solution to the problem

$$\max_{z_i \in [0,1]^n} \left\{ \sum_{j \in N} r_{ij}\, v_{ij}\, z_{ij} - \beta\,\hat{x} \left(v_{i0} + \sum_{j \in N} v_{ij}\, z_{ij} \right) - \beta\,\hat{y}_i \left(v_{i0} + \sum_{j \in N} v_{ij}\, z_{ij} \right)^{1-\gamma_i} \right\}. \tag{31}$$

Following the idea in the proof of Lemma 6, we can show that $\hat{z}_i$ is of the form $\hat{z}_{i1} = 1, \hat{z}_{i2} = 1, \ldots, \hat{z}_{i,k-1} = 1, \hat{z}_{ik} \in [0,1], \hat{z}_{i,k+1} = 0, \ldots, \hat{z}_{in} = 0$ for some $k = 1, \ldots, n$. We claim that we have either $\hat{z}_{ik} = 0$ or $\hat{z}_{ik} = 1$, implying that the optimal solution to the problem above corresponds to a nested-by-revenue assortment. To get a contradiction, we define ρ as $\rho = \hat{z}_{ik}$ and assume that $\rho \in (0, 1)$. Since setting $\hat{z}_{i1} = 1, \hat{z}_{i2} = 1, \ldots, \hat{z}_{i,k-1} = 1, \hat{z}_{ik} = \rho, \hat{z}_{i,k+1} = 0, \ldots, \hat{z}_{in} = 0$ yields an optimal solution to problem (31), if we fix all of the decision variables except for z_{ik} at their optimal values in the problem above and optimize only over the decision variable z_{ik}, then $\rho \in (0, 1)$ must be an optimal solution. However, since $\gamma_i > 1$ and $\hat{y}_i < 0$, the objective function of the problem above is strictly convex and its optimal solution must occur at either $\rho = 0$ or $\rho = 1$. Thus, the claim holds, implying that the optimal solution to problem (31) corresponds to a nested-by-revenue assortment and we denote this nested by revenue assortment by $\{1, 2, \ldots, k\}$ for some $k \in N$.

Noting that $(\hat{x}, \hat{y})$ is the optimal solution to problem (4) after replacing the collection of assortments $\{A_{it} : t \in \mathcal{T}_i\}$ in the second set of constraints with $\{N_{ij}^{k'} : k' \in N, \ j = 0, \ldots, k'\} \cup \{\{j\} : j \in N\}$, this solution satisfies the second set of constraints in problem (4) for nest i and the assortment $N_{ik}^n = \{1, 2, \ldots, k\}$. Therefore, it follows that $\hat{y}_i \geq V_i(N_{ik}^n)^{\gamma_i} (R_i(N_{ik}^n) - \hat{x})$. Multiplying both sides of this inequality with β, we obtain

$$\beta\,\hat{y}_i \geq \beta\, V_i(N_{ik}^n)^{\gamma_i} (R_i(N_{ik}^n) - \hat{x}) \geq V_i(N_{ik}^n)^{\gamma_i} (R_i(N_{ik}^n) - \beta\,\hat{x}),$$

where the second inequality uses the fact that $\beta \geq 1$. Arranging the terms in the inequality $\beta\,\hat{y}_i \geq V_i(N_{ik}^n)^{\gamma_i} (R_i(N_{ik}^n) - \beta\,\hat{x})$ by using the definitions of $V_i(S_i)$ and $R_i(S_i)$, we obtain $\sum_{j \in N_{ik}^n} r_{ij}\, v_{ij} - V_i(N_{ik}^n)\,\beta\,\hat{x} - V_i(N_{ik}^n)^{1-\gamma_i}\,\beta\,\hat{y}_i \leq 0$. Noting that the optimal solution to problem (31) corresponds to the nested-by-revenue assortment $N_{ik}^n = \{1, 2, \ldots, k\}$, the last inequality shows that the optimal objective value of problem (31) is negative. Therefore, for all $S_i \subset N$, we have

$$\sum_{j \in S_i} r_{ij}\, v_{ij} - V_i(S_i)\,\beta\,\hat{x} - V_i(S_i)^{1-\gamma_i}\,\beta\,\hat{y}_i$$

$$\leq \max_{z_i \in [0,1]^n} \left\{ \sum_{j \in N} r_{ij}\, v_{ij}\, z_{ij} - \beta\,\hat{x} \left(v_{i0} + \sum_{j \in N} v_{ij}\, z_{ij} \right) - \beta\,\hat{y}_i \left(v_{i0} + \sum_{j \in N} v_{ij}\, z_{ij} \right)^{1-\gamma_i} \right\} \leq 0.$$

Arranging the terms in the inequality $\sum_{j \in S_i} r_{ij}\, v_{ij} - V_i(S_i)\,\beta\,\hat{x} - V_i(S_i)^{1-\gamma_i}\,\beta\,\hat{y}_i \leq 0$ by using the definitions of $V_i(S_i)$ and $R_i(S_i)$, we obtain $\beta\,\hat{y}_i \geq V_i(S_i)^{\gamma_i} (R_i(S_i) - \beta\,\hat{x})$ for all $S_i \subset N$, which shows that the solution $(\beta\,\hat{x}, \beta\,\hat{y})$ satisfies the second set of constraints for nest i in problem (3).

Case 3. Assume that $\gamma_i \leq 1$. In this case, by using a line of reasoning that is similar to the one in the proof of Theorem 10, we can show that $2\,\hat{y}_i \geq V_i(S_i)^{\gamma_i} (R_i(S_i) - 2\,\hat{x})$ for all $S_i \subset N$. Multiplying

this inequality by $\beta/2 \geq 1$, we obtain $\beta \hat{y}_i \geq V_i(S_i)^{\gamma_i} (\beta R_i(S_i)/2 - \beta \hat{x}) \geq V_i(S_i)^{\gamma_i} (R_i(S_i) - \beta \hat{x})$ for all $S_i \subset N$. The last inequality shows that the solution $(\beta \hat{x}, \beta \hat{y})$ satisfies the second set of constraints for nest i in problem (3).

Collecting the three cases together, the solution $(\beta \hat{x}, \beta \hat{y})$ satisfies the second set of constraints for nest i in problem (3). Noting that our choice of nest i is arbitrary, the second set of constraints in problem (3) is satisfied by the solution $(\beta \hat{x}, \beta \hat{y})$. Finally, since the solution $(\hat{x}, \hat{y})$ is optimal to problem (4), we have $v_0 \hat{x} \geq \sum_{i \in M} \hat{y}_i$, which implies that $v_0 \beta \hat{x} \geq \sum_{i \in M} \beta \hat{y}_i$. Therefore, the solution $(\beta \hat{x}, \beta \hat{y})$ satisfies the first constraint in problem (3) as well and we obtain the desired result. $\qquad\square$

A.5 Knapsack Problems with Equality Constraints

In this section, we begin by giving a dynamic program that obtains the optimal solution to problem (13) in $O(n \, v_i^U)$ time. Following this result, we develop a tractable method to obtain approximate solutions to problem (15). To obtain the optimal solution to problem (13) through a dynamic program, we let $\zeta_i(k, b_i)$ be the optimal objective value of problem (13) when we focus only on the first k products in this problem and replace the right side of the constraint with b_i. In other words, we have

$$\zeta_i(k, b_i) = \max_{S_i \subset \{1,\ldots,k\}} \left\{ \sum_{j \in S_i} r_{ij} \, v_{ij} : v_{i0} + \sum_{j \in S_i} v_{ij} = b_i \right\}$$

with the convention that $\zeta_i(k, b_i) = -\infty$ when the problem on the right side above is infeasible. We note that $\zeta_i(n, \epsilon_i)$ corresponds to the optimal objective value of problem (13). In this case, $\zeta_i(k, b_i)$ satisfies the dynamic programming recursion

$$\zeta_i(k, b_i) = \max \left\{ r_{ik} \, v_{ik} + \zeta_i(k-1, b_i - v_{ik}), \ \zeta_i(k-1, b_i) \right\}$$

with the boundary condition that $\zeta_i(0, v_{i0}) = 0$ and $\zeta_i(0, b_i) = -\infty$ for all $b_i \in \{0, 1, \ldots, v_i^U\} \setminus \{v_{i0}\}$. We can use the dynamic programming recursion above to compute $\zeta_i(n, b_i)$ for all $b_i \in \{0, 1, \ldots, v_i^U\}$ in $O(n \, v_i^U)$ time. In this case, the values in the set $\{\zeta_i(n, b_i) : b_i = 0, 1, \ldots, v_i^U\}$ correspond to the values $\{G_i(\epsilon_i) : \epsilon_i = 0, 1, \ldots, v_i^U\}$, as desired. The dynamic program above is similar to the one that is used for solving the partition and knapsack problems; see Garey and Johnson (1979).

In the rest of this section, we focus on obtaining approximate solutions to problem (15). For notational brevity, we omit the subscripts for the nest and use the decision variables $z = (z_1, \ldots, z_n)$ to consider the problem

$$\hat{G}_l = \max \left\{ \sum_{j=1}^n r_j \, v_j \, z_j : \delta^{l-1} \leq v_0 + \sum_{j=1}^n v_j \, z_j \leq \delta^l, \ z \in \{0,1\}^n \right\}. \tag{32}$$

We are interested in finding a feasible solution to the problem above whose objective value deviates from the optimal objective value by no more than a factor of δ. To that end, we begin by classifying the products in the problem above into two categories. A product j satisfying $v_j > (\delta - 1) \, \delta^{l-1}$ is called a large product, whereas a product j satisfying $v_j \leq (\delta - 1) \, \delta^{l-1}$ is called a small product. We use N^L and N^S to respectively denote the sets of large and small products, with $N = N^L \cup N^S$. We observe

that an optimal solution to problem (32) cannot include $\lceil \delta/(\delta-1) \rceil$ or more large products. Otherwise, the constraint in problem (32) evaluates to more than $\lceil \delta/(\delta-1) \rceil (\delta-1)\delta^{l-1} \geq \delta^l$, violating its upper bound. For notational brevity, we set $q = \lceil \delta/(\delta-1) \rceil$ throughout this section so that an optimal solution to problem (32) cannot include q or more large products.

To obtain a tractable approximation to problem (32), we consider a special linear programming relaxation of this problem. In particular, we choose a subset $J^L \subset N^L$ of large products and a subset $J^S \subset N^S$ of small products and solve the problem

$$\max \quad \sum_{j=1}^{n} r_j\, v_j\, z_j \tag{33}$$

$$\text{s.t.} \quad \sum_{j=1}^{n} v_j\, z_j \leq \delta^l$$

$$z_j = 1 \qquad\qquad \forall\, j \in J^L \cup J^S$$

$$z_j = 0 \qquad\qquad \forall\, j \in N^L \setminus J^L$$

$$0 \leq z_j \leq \mathbf{1}(r_j\, v_j \leq \min_{k \in J^S} r_k\, v_k) \quad \forall\, j \in N^S \setminus J^S.$$

The problem above is a continuous knapsack problem with only an upper bound constraint and the values of some of the variables are fixed at zero or one. In particular, the decision variables corresponding to the products in J^L and J^S are set at one. The decision variables corresponding to the large products in $N^L \setminus J^L$ are set to zero. If the objective function coefficient of a small product in $N^S \setminus J^S$ is smaller than the smallest of the objective function coefficient of the small products in J^S, then the decision variable corresponding to this small product is allowed to take values between zero and one. Otherwise, the decision variable corresponding to this product is fixed at zero. Noting that the utility of product j in problem (33) is given by $r_j\, v_j$ and the utility-to-space consumption ratio of product j is given by r_j, we can solve problem (33) by using the following procedure. We put all of the products in $J^L \cup J^S$ into the knapsack and drop these products from consideration. Also, we drop the products in $N^L \setminus J^L$ from consideration immediately. We order the products in N^S with respect to their utilities. If there are any products in $N^S \setminus J^S$ whose utilities exceed the smallest of the utilities in J^S, then we drop these products from consideration as well. Considering the remaining products in $N^S \setminus J^S$, we order these products with respect to their utility-to-space consumption ratios and fill the knapsack starting from the products with the largest utility-to-space consumption ratios. Therefore, assuming that we already have the orderings of the items with respect to their utilities and utility-to-space consumption ratios, we can solve problem (33) in $O(n)$ time. It is also useful to observe that the optimal solution to problem (33) that we obtain in this fashion includes at most one fractional component. The continuous knapsack problem in (33) is inspired by Frieze and Clarke (1984), where the authors use a similar continuous knapsack problem to construct polynomial-time approximation schemes for multi-dimensional knapsack problems. We use $\hat{z}(J^L, J^S)$ to denote the optimal solution to problem (33), where our notation emphasizes the fact that the solution to this problem depends on the choice of J^L and J^S.

Using the solution $\hat{z}(J^L, J^S)$, we define the assortment $\hat{S}(J^L, J^S) = \{j \in N : \hat{z}_j(J^L, J^S) = 1\}$, including to the products that take strictly positive and integer values in the solution $\hat{z}(J^L, J^S)$. By

using the discussion above, for given J^L and J^S, we can compute $\hat{S}(J^L, J^S)$ in $O(n)$ time. We use $\wp^L$ to denote the set of subsets of N^L with cardinality not exceeding q. Similarly, we use $\wp^S$ to denote the set of subsets of N^S with cardinality not exceeding q. In this case, the next proposition shows that the collection $\{\hat{S}(J^L, J^S) : J^L \in \wp^L, \ J^S \in \wp^S\}$ includes a feasible solution to problem (32) such that the objective value provided by this feasible solution deviates from the optimal objective value of problem (32) by at most a factor of δ.

Proposition 15 *Assuming that problem* (32) *has a feasible solution, there exists an assortment in the collection* $\{\hat{S}(J^L, J^S) : J^L \in \wp^L, \ J^S \in \wp^S\}$ *such that this assortment yields a feasible solution to problem* (32) *and the objective value provided by this assortment deviates from the optimal objective value of problem* (32) *by at most a factor of δ. Furthermore, all of the assortments in the collection* $\{\hat{S}(J^L, J^S) : J^L \in \wp^L, \ J^S \in \wp^S\}$ *can be constructed in* $O(q^2 \, n^{2q+1})$ *time.*

Proof. By the definitions of $\wp^L$ and $\wp^S$, we have $|\wp^L| = |\wp^S| = O(q \, n^q)$. Therefore, there are $O(q^2 \, n^{2q})$ assortments in the collection $\{\hat{S}(J^L, J^S) : J^L \in \wp^L, \ J^S \in \wp^S\}$. Noting the discussion right before the proposition, each one of these assortments can be constructed in $O(n)$ time. Thus, all of the assortments in the collection $\{\hat{S}(J^L, J^S) : J^L \in \wp^L, \ J^S \in \wp^S\}$ can be constructed in $O(q^2 \, n^{2q+1})$ time.

Letting $\tilde{z}$ be the optimal solution to problem (32), we define the assortment $\tilde{S}$ corresponding to this solution as $\tilde{S} = \{j \in N : \tilde{z}_j = 1\}$. We let $\tilde{J}^L$ and $\tilde{J}^S$ to respectively be the large and small products in the assortment $\tilde{S}$. By the discussion that follows problem (32), we must have $|\tilde{J}^L| \leq q$, implying that $\tilde{J}^L \in \wp^L$. If we assume that $|\tilde{J}^S| \leq q$, then we have $\tilde{J}^S \in \wp^S$ as well. Thus, the assortment $\hat{S}(\tilde{J}^L, \tilde{J}^S)$ is included in the collection $\{\hat{S}(J^L, J^S) : J^L \in \wp^L, \ J^S \in \wp^S\}$. Furthermore, by the definitions of $\hat{z}(J^L, J^S)$ and $\hat{S}(J^L, J^S)$, the assortment $\hat{S}(\tilde{J}^L, \tilde{J}^S)$ includes all of the products in $\tilde{J}^L$ and $\tilde{J}^S$, which implies that $\hat{S}(\tilde{J}^L, \tilde{J}^S) \supset \tilde{J}^L \cup \tilde{J}^S = \tilde{S}$ so that $\hat{S}(\tilde{J}^L, \tilde{J}^S)$ includes all of the products in $\tilde{S}$. Thus, $\hat{S}(\tilde{J}^L, \tilde{J}^S)$ must provide an objective value for problem (32) that is at least as large as the one provided by $\tilde{S}$ and we conclude that $\hat{S}(\tilde{J}^L, \tilde{J}^S)$ is an optimal solution to problem (32). This establishes the desired result under the assumption that $|\tilde{J}^S| \leq q$. In the rest of the proof, we assume that $|\tilde{J}^S| > q$.

We let A^S be the subset of $\tilde{J}^S$ that includes the q products in $\tilde{J}^S$ with the largest utilities. In other words, we have $A^S \subset \tilde{J}^S$, $|A^S| = q$ and $r_j \, v_j \leq \min_{k \in A^S} r_k \, v_k$ for all $j \in \tilde{J}^S \setminus A^S$. Consider the solution $\hat{z}(\tilde{J}^L, A^S)$ that we obtain by solving problem (33) with $J^L = \tilde{J}^L$ and $J^S = A^S$. If this solution has a fractional component j', then by the fourth set of constraints in problem (33), this component must satisfy $r_{j'} \, v_{j'} \leq r_k \, v_k$ for all $k \in A^S$. Also, the component j' must be in $N^S \setminus A^S$ so that the product j' is a small product. In this case, noting that the optimal objective value of problem (32) is given by $\hat{G}_l$, we have $\hat{G}_l = \sum_{j \in \tilde{J}^L} r_j \, v_j + \sum_{j \in \tilde{J}^S} r_j \, v_j = \sum_{j \in \tilde{J}^L} r_j \, v_j + \sum_{j \in A^S} r_j \, v_j + \sum_{j \in \tilde{J}^S \setminus A^S} r_j \, v_j \geq \sum_{j \in A^S} r_j \, v_j \geq q \, r_{j'} \, v_{j'}$, where the first equality is by the fact that $\tilde{S} = \tilde{J}^L \cup \tilde{J}^S$ is an optimal solution to problem (32) and the second inequality is by the fact that $r_{j'} \, v_{j'} \leq r_k \, v_k$ for all $k \in A^S$ and $|A^S| = q$. The last chain of inequalities yields $r_{j'} \, v_{j'} \leq \hat{G}_l/q$.

We claim that the assortment $\hat{S}(\tilde{J}^L, A^S)$ provides a feasible solution to problem (32). First, we show this claim under the assumption that the solution $\hat{z}(\tilde{J}^L, A^S)$ consumes all of the knapsack capacity in

51

problem (33) when we solve this problem with $J^L = \tilde{J}^L$ and $J^S = A^S$. We use j' to denote the fractional component of $\hat{z}(\tilde{J}^L, A^S)$ when there is one. By the discussion in the paragraph above, j' must be a small product. Since the solution $\hat{z}(\tilde{J}^L, A^S)$ consumes all of the knapsack capacity, we have $\delta^l = \sum_{j=1}^n v_j\,\hat{z}_j(\tilde{J}^L, A^S) \leq \sum_{j\in\hat{S}(\tilde{J}^L,A^S)} v_j + v_{j'} \leq \sum_{j\in\hat{S}(\tilde{J}^L,A^S)} + (\delta-1)\,\delta^{l-1}$, where the first inequality follows from the fact that $\hat{S}(\tilde{J}^L, A^S)$ includes all components of $\hat{z}(\tilde{J}^L, A^S)$ with the exception of j' and the second inequality follows from the fact that product j' is a small product. From the last chain of inequalities, it follows that $\sum_{j\in\hat{S}(\tilde{J}^L,A^S)} v_j \geq \delta^l - (\delta-1)\,\delta^{l-1} = \delta^{l-1}$ so that the assortment $\hat{S}(\tilde{J}^L, A^S)$ satisfies the lower bound constraint in problem (32). Furthermore, noting that $\delta^l \geq \sum_{j=1}^n v_j\,\hat{z}_j(\tilde{J}^L, A^S) = \sum_{j\in\hat{S}(\tilde{J}^L,A^S)} v_j + v_{j'}\,z_{j'}(\tilde{J}^L, A^S) \geq \sum_{j\in\hat{S}(\tilde{J}^L,A^S)} v_j$, the assortment $\hat{S}(\tilde{J}^L, A^S)$ satisfies the upper bound constraint in problem (32) as well and the claim follows.

Second, we show the claim under the assumption that the solution $\hat{z}(\tilde{J}^L, A^S)$ does not consume all of the knapsack capacity in problem (33) when we solve this problem with $J^L = \tilde{J}^L$ and $J^S = A^S$. We recall that the discussion at the beginning of the third paragraph of the proof shows that $r_j\,v_j \leq \min_{k\in A^S} r_k\,v_k$ for all $j \in \tilde{J}^S \setminus A^S$. Thus, if we solve problem (33) with $J^L = \tilde{J}^L$ and $J^S = A^S$, then by the fourth set of constraints in this problem, the decision variables corresponding to the products in $\tilde{J}^S \setminus A^S$ are free to take values between zero and one. In this case, since the solution $\hat{z}(\tilde{J}^L, A^S)$ does not consume all of the knapsack capacity in problem (33) and all of the objective function coefficients are positive, the decision variables corresponding to the products in $\tilde{J}^S \setminus A^S$ must take value one. Furthermore, the decision variables corresponding to the products in $\tilde{J}^L$ and A^S are fixed at one when we solve problem (33) with $J^L = \tilde{J}^L$ and $J^S = A^S$. Therefore, if we solve problem (33) with $J^L = \tilde{J}^L$ and $J^S = A^S$, then the decision variables corresponding to the products in $\tilde{J}^L$, A^S and $\tilde{J}^S \setminus A^S$ take value one, which implies that $\delta^{l-1} \leq \sum_{j\in\tilde{S}} v_j = \sum_{j\in\tilde{J}^L} v_j + \sum_{j\in A^S} v_j + \sum_{j\in\tilde{J}^S\setminus A^S} v_j \leq \sum_{j=1}^n v_j\,z_j(\tilde{J}^L, A^S) = \sum_{j\in\hat{S}(\tilde{J}^L,A^S)} v_j$, where the first inequality follows from the fact that $\tilde{S}$ is an optimal solution to problem (32), the first equality uses the fact that $\tilde{S} = \tilde{J}^L \cup \tilde{J}^S$, the second inequality follows from the fact that the products in $\tilde{J}^L$, A^S and $\tilde{J}^S \setminus A^S$ all take value one when we solve problem (33) with $J^L = \tilde{J}^L$ and $J^S = A^S$ and the last equality follows from the fact that if the solution $\hat{z}(\tilde{J}^L, A^S)$ does not consume all of the knapsack capacity, then it cannot have any fractional components. On the other hand, since the solution $\hat{z}(\tilde{J}^L, A^S)$ does not have any fractional components and it is an optimal solution to problem (33), we obtain $\delta^l \geq \sum_{j=1}^n v_j\,z_j(\tilde{J}^L, A^S) = \sum_{j\in\hat{S}(\tilde{J}^L,A^S)} v_j$. The last two chains of inequalities show that the assortment $\hat{S}(\tilde{J}^L, A^S)$ satisfies the constraint in problem (32) and the claim holds.

We proceed to checking the objective function value provided by the assortment $\hat{S}(\tilde{J}^L, A^S)$. To that end, letting $\zeta(J^L, J^S)$ be the optimal objective value of problem (33), we begin by arguing that $\hat{G}_l \leq \zeta(\tilde{J}^L, A^S)$. To establish this inequality, we observe that by the definitions of $\tilde{J}^L$ and $\tilde{J}^S$, the products in $\tilde{J}^L$ and $\tilde{J}^S$ take value one in the optimal solution to problem (32). On the other hand, if we solve problem (33) with $J^L = \tilde{J}^L$ and $J^S = A^S$, then the products in $\tilde{J}^L$ and A^S take value one in the optimal solution. Furthermore, as discussed at the beginning of the previous paragraph, the products in $\tilde{J}^S \setminus A^S$ are free to take values between zero and one when we solve problem (33). Thus, the optimal solution to problem (32) is a feasible solution to problem (33) when we solve this problem with $J^L = \tilde{J}^L$ and $J^S = A^S$, which implies that $\hat{G}_l \leq \zeta(\tilde{J}^L, A^S)$ as desired. In this case, using j' to denote the fractional component of $\hat{z}(\tilde{J}^L, A^S)$ when there is one, we obtain $\hat{G}_l \leq \zeta(\tilde{J}^L, A^S) =$

$\sum_{j=1}^{n} r_j \, v_j \, \hat{z}_j(\tilde{J}^L, A^S) \leq \sum_{j \in \hat{S}(\tilde{J}^L, A^S)} r_j \, v_j + r_{j'} \, v_{j'} \leq \sum_{j \in \hat{S}(\tilde{J}^L, A^S)} r_j \, v_j + \hat{G}_l/q$, where the last inequality follows by noting that $r_{j'} \, v_{j'} \leq \hat{G}_l/q$, which is shown in the third paragraph of the proof. Focusing on the first and last expressions in the last chain of inequalities and noting that $q = \lceil \delta/(\delta - 1) \rceil$, we get $\sum_{j \in \hat{S}(\tilde{J}^L, A^S)} r_j \, v_j \geq ((q-1)/q)\hat{G}_l \geq \hat{G}_l/\delta$. So, the assortment $\hat{S}(\tilde{J}^L, A^S)$ corresponds to a feasible solution to problem (32), providing an objective value to this problem that deviates from the optimal objective value by no more than a factor of δ. Furthermore, noting that $|\tilde{J}^L| \leq q$ and $|A^S| = q$, we have $\hat{S}(\tilde{J}^L, A^S) \in \{\hat{S}(J^L, J^S) : J^L \in \wp^L, \ J^S \in \wp^S\}$ and the result follows. □

A.6 Proof of Theorem 12

By using the same argument at the beginning of the proof of Theorem 10, it follows that $\hat{x} \geq 0$. Fix an arbitrary nest i. Choose any assortment $S_i \subset N$ within this nest. First, we consider the case where $S_i \neq \emptyset$. Fix $l = l_i^L, \ldots, l_i^U$ such that $\delta^{l-1} \leq v_{i0} + \sum_{j \in S_i} v_{ij} \leq \delta^l$. Since $(\hat{x}, \hat{y})$ is the optimal solution to problem (4) after replacing the collection of assortments $\{A_{it} : t \in \mathcal{T}_i\}$ in the second set of constraints with the assortments $\{\hat{S}_{i\ell} : \ell = l_i^L, \ldots, l_i^U\} \cup \{\emptyset\}$, this solution satisfies the second set of constraints in problem (4) for nest i and the assortment $\hat{S}_{il}$. Therefore, we have

$$\hat{y}_i \geq V_i(\hat{S}_{il})^{\gamma_i}(R_i(\hat{S}_{il}) - \hat{x}) = V_i(\hat{S}_{il})^{\gamma_i - 1} \left(\sum_{j \in \hat{S}_{il}} r_{ij} \, v_{ij} \right) - V_i(\hat{S}_{il})^{\gamma_i} \, \hat{x}.$$

Multiplying both sides of this inequality by $\delta^{\bar{\gamma}+1}$, we obtain

$$\delta^{\bar{\gamma}+1} \, \hat{y}_i \geq \delta^{\bar{\gamma}} \, V_i(\hat{S}_{il})^{\gamma_i - 1} \left(\delta \sum_{j \in \hat{S}_{il}} r_{ij} \, v_{ij} \right) - \delta^{\bar{\gamma}+1} \, V_i(\hat{S}_{il})^{\gamma_i} \, \hat{x}. \tag{34}$$

We proceed to bound each one of the terms $\delta \sum_{j \in \hat{S}_{il}} r_{ij} \, v_{ij}$, $V_i(\hat{S}_{il})^{\gamma_i - 1}$ and $V_i(\hat{S}_{il})^{\gamma_i}$ in the inequality above. By the definition of $\hat{S}_{il}$, we have

$$\delta \sum_{j \in \hat{S}_{il}} r_{ij} \, v_{ij} \geq \hat{G}_{il} = \max_{S_i' \subset N} \left\{ \sum_{j \in S_i'} r_{ij} \, v_{ij} : \delta^{l-1} \leq v_{i0} + \sum_{j \in S_i'} v_{ij} \leq \delta^l \right\} \geq \sum_{j \in S_i} r_{ij} \, v_{ij}, \tag{35}$$

where the second inequality follows by noting that l is chosen such that $\delta^{l-1} \leq v_{i0} + \sum_{j \in S_i} v_{ij} \leq \delta^l$. The definition of $\hat{S}_{il}$ also implies that $\delta^{l-1} \leq v_{i0} + \sum_{j \in \hat{S}_{il}} v_{ij} = V_i(\hat{S}_{il}) \leq \delta^l$. In this case, if we have $\gamma_i \leq 1$, then $V_i(\hat{S}_{il})^{\gamma_i - 1} \geq (\delta^l)^{\gamma_i - 1}$. If, on the other hand, we have $\gamma_i > 1$, then $V_i(\hat{S}_{il})^{\gamma_i - 1} \geq (\delta^{l-1})^{\gamma_i - 1} = (\delta^l)^{\gamma_i - 1} \delta^{-(\gamma_i - 1)}$. So, combining the two cases, we bound $V_i(\hat{S}_{il})^{\gamma_i - 1}$ by

$$V_i(\hat{S}_{il})^{\gamma_i - 1} \geq (\delta^l)^{\gamma_i - 1} \delta^{-[\gamma_i - 1]^+},$$

where we use $[a]^+ = \max\{a, 0\}$. Noting also that $V_i(\hat{S}_{il})^{\gamma_i} \leq (\delta^l)^{\gamma_i}$, using these bounds on $V_i(\hat{S}_{il})^{\gamma_i - 1}$ and $V_i(\hat{S}_{il})^{\gamma_i}$ together with the inequality in (35) in (34), we obtain

$$\delta^{\bar{\gamma}+1} \, \hat{y}_i \geq \delta^{\bar{\gamma}} \, (\delta^l)^{\gamma_i - 1} \delta^{-[\gamma_i - 1]^+} \left(\sum_{j \in S_i} r_{ij} \, v_{ij} \right) - \delta^{\bar{\gamma}+1} \, (\delta^l)^{\gamma_i} \, \hat{x}$$

$$= \delta^{\bar{\gamma}} \, (\delta^l)^{\gamma_i - 1} \delta^{-[\gamma_i - 1]^+} \left(\sum_{j \in S_i} r_{ij} \, v_{ij} \right) - \delta^{\bar{\gamma}+\gamma_i+1} \, (\delta^{l-1})^{\gamma_i} \, \hat{x}. \tag{36}$$

Our choice of l at the beginning of the proof implies that $\delta^{l-1} \leq v_{i0} + \sum_{j \in S_i} v_{ij} = V_i(S_i) \leq \delta^l$. In this case, if we have $\gamma_i \leq 1$, then $V_i(S_i)^{\gamma_i - 1} \leq (\delta^{l-1})^{\gamma_i - 1} = (\delta^l)^{\gamma_i - 1} \delta^{1 - \gamma_i}$. If, on the other hand, we have $\gamma_i > 1$, then $V_i(S_i)^{\gamma_i - 1} \leq (\delta^l)^{\gamma_i - 1}$. Combining the two cases yields $V_i(S_i)^{\gamma_i - 1} \leq (\delta^l)^{\gamma_i - 1} \delta^{[1 - \gamma_i]^+}$ so that we can bound $(\delta^l)^{\gamma_i - 1}$ by

$$(\delta^l)^{\gamma_i - 1} \geq \delta^{-[1 - \gamma_i]^+} V_i(S_i)^{\gamma_i - 1}.$$

Furthermore, noting that $(\delta^{l-1})^{\gamma_i} \leq V_i(S_i)^{\gamma_i}$, we use these bounds on $(\delta^l)^{\gamma_i - 1}$ and $(\delta^{l-1})^{\gamma_i}$ in the two terms on the right side of (36) to obtain

$$\delta^{\bar{\gamma} + 1} \hat{y}_i \geq \delta^{\bar{\gamma}} \delta^{-[\gamma_i - 1]^+} \delta^{-[1 - \gamma_i]^+} V_i(S_i)^{\gamma_i - 1} \left(\sum_{j \in S_i} r_{ij} v_{ij} \right) - \delta^{\bar{\gamma} + \gamma_i + 1} V_i(S_i)^{\gamma_i} \hat{x}.$$

If we have $\gamma_i \leq 1$, then $\bar{\gamma} - [\gamma_i - 1]^+ - [1 - \gamma_i]^+ = \bar{\gamma} - 1 + \gamma_i \geq 0$, where we use the fact that $\bar{\gamma} > 1$. On the other hand, if we have $\gamma_i > 1$, then $\bar{\gamma} - [\gamma_i - 1]^+ - [1 - \gamma_i]^+ = \bar{\gamma} - \gamma_i + 1 \geq 0$ since $\bar{\gamma} \geq \gamma_i$. Therefore, $\delta^{\bar{\gamma}} \delta^{-[\gamma_i - 1]^+} \delta^{-[1 - \gamma_i]^+} \geq 1$. We also have $\delta^{\bar{\gamma} + \gamma_i + 1} \leq \delta^{2\bar{\gamma} + 1}$. Thus, the last inequality above yields

$$\delta^{\bar{\gamma} + 1} \hat{y}_i \geq V_i(S_i)^{\gamma_i - 1} \left(\sum_{j \in S_i} r_{ij} v_{ij} \right) - \delta^{2\bar{\gamma} + 1} V_i(S_i)^{\gamma_i} \hat{x}.$$

Since $\sum_{j \in S_i} r_{ij} v_{ij} / V_i(S_i) = R_i(S_i)$, the inequality above shows that the solution $(\delta^{2\bar{\gamma} + 1} \hat{x}, \delta^{\bar{\gamma} + 1} \hat{y})$ satisfies the second set of constraints in problem (3) for the assortment S_i and nest i, as long as $S_i \neq \emptyset$.

Second, we consider the case where $S_i = \emptyset$. The solution $(\hat{x}, \hat{y})$ satisfies the second set of constraints in problem (4) for the empty assortment within nest i, in which case, we obtain $\hat{y}_i \geq V_i(\emptyset)^{\gamma_i} (R_i(\emptyset) - \hat{x})$. If we multiply this inequality by $\delta^{\bar{\gamma} + 1}$ and note that $R_i(\emptyset) = 0$, then we have $\delta^{\bar{\gamma} + 1} \hat{y}_i \geq \delta^{\bar{\gamma} + 1} V_i(\emptyset)^{\gamma_i} R_i(\emptyset) - \delta^{\bar{\gamma} + 1} V_i(\emptyset)^{\gamma_i} \hat{x} = V_i(\emptyset)^{\gamma_i} R_i(\emptyset) - \delta^{\bar{\gamma} + 1} V_i(\emptyset)^{\gamma_i} \hat{x}$. Replacing the term $\delta^{\bar{\gamma} + 1}$ on the right side of the last inequality with an even larger term $\delta^{2\bar{\gamma} + 1}$, it follows that

$$\delta^{\bar{\gamma} + 1} \hat{y}_i \geq V_i(\emptyset)^{\gamma_i} R_i(\emptyset) - \delta^{2\bar{\gamma} + 1} V_i(\emptyset)^{\gamma_i} \hat{x}.$$

Therefore, the solution $(\delta^{2\bar{\gamma} + 1} \hat{x}, \delta^{\bar{\gamma} + 1} \hat{y})$ satisfies the second set of constraints in problem (3) for assortment $S_i = \emptyset$ and nest i. Combining the two cases above and noting that our choice of nest i and assortment S_i is arbitrary, we conclude that the solution $(\delta^{2\bar{\gamma} + 1} \hat{x}, \delta^{\bar{\gamma} + 1} \hat{y})$ satisfies the second set of constraints in problem (3).

Since the solution $(\hat{x}, \hat{y})$ is optimal to problem (4), we have $v_0 \hat{x} \geq \sum_{i \in M} \hat{y}_i$. This implies that $v_0 \delta^{2\bar{\gamma} + 1} \hat{x} \geq v_0 \delta^{\bar{\gamma} + 1} \hat{x} \geq \sum_{i \in M} \delta^{\bar{\gamma} + 1} \hat{y}_i$, in which case, the solution $(\delta^{2\bar{\gamma} + 1} \hat{x}, \delta^{\bar{\gamma} + 1} \hat{y})$ satisfies the first constraint in problem (3) as well and we obtain the desired result. $\square$