Approximation Methods for Pricing Problems under the Nested Logit Model with Price Bounds

W. Zachary Rayfield
School of Operations Research and Information Engineering,
Cornell University, Ithaca, New York 14853, USA
wzr2@cornell.edu

Paat Rusmevichientong
Marshall School of Business,
University of Southern California, Los Angeles, California 90089, USA
rusmevic@marshall.usc.edu

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

September 5, 2013

Abstract

We consider two variants of a pricing problem under the nested logit model. In the first variant, the set of products offered to customers is fixed and we want to determine the prices of the products. In the second variant, we jointly determine the set of offered products and their corresponding prices. In both variants, the price of each product has to be chosen within given upper and lower bounds specific to the product, each customer chooses among the offered products according to the nested logit model and the objective is to maximize the expected revenue from each customer. We give approximation methods for both variants. For any $\rho > 0$, our approximation methods obtain a solution with an expected revenue deviating from the optimal expected revenue by no more than a factor of $1 + \rho$. To obtain such a solution, our approximation methods solve a linear program whose size grows at rate $1/\rho$. In addition to our approximation methods, we develop a linear program that we can use to obtain an upper bound on the optimal expected revenue. In our computational experiments, we compare the expected revenues from the solutions obtained by our approximation methods with the upper bounds on the optimal expected revenues and show that we can obtain high quality solutions quite fast.

1 Introduction

When faced with product variety, most customers make their purchase decisions by comparing the offered products through attributes such as price, richness of features and durability. In this type of a situation, the demand for a certain product is determined not only by its own attributes but also by the attributes of other products, creating interactions among the demands for different products. Discrete choice models are particularly suitable to study such demand interactions, as they model the demand for a certain product as a function of the attributes of all products offered to customers. However, optimization models that try to find the right set of products to offer or the right prices to charge may quickly become intractable when one works with complex discrete choice models and tries to incorporate operational constraints.

In this paper, we consider pricing problems where the interactions between the demands for the different products are captured through the nested logit model and there are bounds on the prices that can be charged for the products. We consider two problem variants. In the first variant, the set of products offered to customers is fixed and we want to determine the prices for these products. In the second variant, we jointly determine the products that should be offered to customers and their corresponding prices. Once the products to be offered and their prices are determined, customers choose among the offered products according to the nested logit model. In both variants, the objective is to maximize the expected revenue obtained from each customer. We give approximation methods for both variants of the problem. In particular, for any $\rho > 0$, our approximation methods obtain a solution with an expected revenue deviating from the optimal by at most a factor of $1 + \rho$. To obtain this solution, the approximation methods solve linear programs whose sizes grow linearly with $1/\log(1 + \rho)$. Noting that $1/\log(1 + \rho)$ grows at the same rate as $1/\rho$ for small values of ρ, the computational work for our approximation methods grows polynomially with the approximation factor. Our approximation methods give a performance guarantee over all problem instances, but we also develop a linear program that we can use to quickly obtain an upper bound on the optimal expected revenue for an individual problem instance. In our computational experiments, we compare the expected revenues from the solutions obtained by our approximation methods with the upper bounds on the optimal expected revenues and demonstrate that our approximation methods can quickly obtain solutions whose expected revenues differ from the optimal by less than a percent. Thus, our approximation methods have favorable theoretical performance guarantees and they are useful to obtain high quality solutions in practice.

Main Results and Contributions. The first problem variant we consider is a pricing problem where customers choose according to the nested logit model and there are bounds on the prices of the offered products. For the first variant, assuming that there are m nests in the nested logit model and each nest includes n products to offer, we show that for any $\rho > 0$, we can solve a linear program with $O(m)$ decision variables and $O(mn + mn\log(n\sigma)/\log(1 + \rho))$ constraints to obtain a set of prices with an expected revenue deviating from the optimal expected revenue by at most a factor of $1+\rho$. In this result, σ depends on the deviation between the upper and lower price bounds of the products. The second problem variant we consider is a joint assortment offering and pricing problem, where we need to choose the products to offer and their corresponding prices. For this variant, we establish a useful property for the optimal subsets of products to offer. In particular, ordering the products according to their price upper bounds, we show that it is optimal to offer a certain number of products with the largest price upper bounds. Using this result, we show that for any $\rho > 0$, we can solve a linear program with $O(m)$ decision variables and $O(mn^2 + mn^2 \log(n\sigma)/\log(1+\rho))$ constraints to find a set of products to offer and their corresponding prices such that the expected revenue obtained by this solution deviates from the optimal expected revenue by at most a factor of $1+\rho$. Comparing our results for the two variants, we observe that the extra computational burden of jointly finding a set of products to offer and pricing the offered products boils down to increasing the number of constraints in the linear program by a factor of n.

Pricing under the nested logit model has recently received attention, starting with the work of Li and Huh (2011) and Gallego and Wang (2011). Li and Huh (2011) consider pricing problems without upper or lower bound constraints on the prices. Assuming that the products in the same nest share the same price sensitivity parameter and the so called dissimilarity parameters of the nested logit model are less than one, they cleanly show that the pricing problem can be reduced to the problem of maximizing a scalar function. This scalar function turns out to be unimodal so that maximizing it is tractable. Gallego and Wang (2011) also study pricing problems under the nested logit model without price bounds, but they allow the products in the same nest to have different price sensitivities and the dissimilarity parameters of the nested logit model to take on arbitrary values. Surprisingly, their elegant argument shows that the optimal prices can still be found by maximizing a scalar function, but this scalar function is not unimodal in general and evaluating this scalar function at any point requires solving a separate high dimensional optimization problem involving implicitly defined functions. Our paper fills a number of gaps in this area. The earlier work shows that the problem of finding the optimal prices can be reduced to maximizing a scalar function, but this function is not unimodal and maximizing it can be intractable for two reasons. First, a natural approach to maximizing this scalar function is to evaluate it at a finite number of grid points and pick the best solution, but it is not clear how to place these grid points to obtain a performance guarantee. Second, given that computing the scalar function at any point requires solving a nontrivial optimization problem, it is computationally prohibitive to simply follow a brute force approach and use a large number of grid points. Thus, while the earlier work shows how to reduce the pricing problem to a problem of maximizing a scalar function, as far as we can see, it does not yet yield a computationally viable and theoretically sound algorithm to compute near optimal prices in general. Our work provides practical algorithms that deliver a desired performance guarantee of $1+\rho$ for any $\rho > 0$. To obtain our approximation methods, we transform the pricing problem into a knapsack problem with a separable and concave objective function, which ultimately allows us to use different arguments from Li and Huh (2011) and Gallego and Wang (2011).

Beside providing computationally viable algorithms to find prices with a certain performance guarantee, a unique feature of our work is that it allows imposing bounds on the prices that can be chosen by the decision maker. Such price bounds do not appear in the earlier pricing work under the nested logit model and there are a number of theoretical and practical reasons for studying such bounds. On the theoretical side, if we impose price bounds, then even in the simplest case when the price sensitivities of all products are equal to each other, the scalar functions in the works of Li and Huh (2011) and Gallego and Wang (2011) are no longer unimodal. In such cases, we emphasize that the lack of unimodality is purely due to the presence of the price bounds, as the work of Li and Huh (2011) shows that the scalar functions that they work are indeed unimodal when the price sensitivities of the products are equal to each other. Thus, price bounds can significantly complicate the structural properties of the pricing problem. Furthermore, naive approaches for satisfying price bound constraints may yield poor results. For example, a first cut approach for dealing with price bounds is to use the work of Li and Huh (2011) or Gallego and Wang (2011) to find the optimal prices for the products under the assumption that there are no price bounds. If these unconstrained prices are outside the price bound constraints, then we can round them up or down to their corresponding lower or upper bounds. This naive approach does not perform well and we can come up with problem instances where this naive approach can result in revenue losses of over 20%, when compared with approaches that explicitly incorporate price bounds.

There are also practical reasons for studying price bounds. Customers may have expectations for sensible price ranges and it is useful to incorporate these price ranges explicitly into the pricing model. Furthermore, lack of data may prevent us from fitting an accurate choice model to capture customer choices, in which case, we can guide the model by limiting the range of possible prices through price bounds. When we solve the pricing model without price bounds, we essentially rely on the choice model to find a set of reasonable prices for the products, but depending on the parameters of the choice model, the prices may not come out to be practical. Thus, incorporating price bounds into the pricing problem is a nontrivial task from a theoretical perspective and it has important practical implications. It is also worth mentioning that if there are no price bounds, then finding the right set of products to offer is not an issue as Gallego and Wang (2011) show that it is always optimal to offer all products at some finite price level. This result does not hold in the presence of price bounds and our second variant, which jointly determines the set of products to offer and their corresponding prices, becomes particularly useful.

Our approximation methods allow us to obtain prices with a certain performance guarantee. In addition to these approximation methods, we give a simple approach to compute an upper bound on the optimal expected revenue. This upper bound is obtained by solving a linear program and we can progressively refine the upper bound by increasing the number of constraints in the linear program. By comparing the expected revenue from the solution obtained by our approximation methods with the upper bound on the optimal expected revenue, we can bound the optimality gap of the solutions obtained by our approximation methods for each individual problem instance. Admittedly, our approximation methods provide a performance guarantee of $1 + \rho$ for a given $\rho > 0$, but this is the worst case performance guarantee over all problem instances and it turns out that we can use the linear program to obtain a tighter performance guarantee for an individual problem instance. The linear program we use to obtain an upper bound on the optimal expected revenue can be useful even if we do not work with our approximation methods to obtain a good solution to the pricing problem. In particular, we can use an arbitrary heuristic or an approximation method to obtain a set of prices and check the gap between the expected revenue obtained by charging these prices and the upper bound on the optimal expected revenue. If the gap turns out to be small, then there is no need to look for better prices.

Related Literature. There is a long history on building discrete choice models to capture customer preferences. Some of these models are based on axioms describing a sensible behavior of customer choice, as in the basic attraction model of Luce (1959). On the other hand, some others use a utility maximization principle, where an arriving customer associates random utilities with the products and chooses the product providing the largest utility. Such a utility based approach is followed by McFadden (1974), resulting in the celebrated multinomial logit model. The nested logit model, which plays a central role in this paper, goes back to the work of Williams (1977). Extensions for the nested logit model are provided by McFadden (1980) and Borsch-Supan (1990). An important feature of the nested logit model is that it avoids the independence of irrelevant alternatives property suffered by the multinomial logit model. The discussion of this property can be found in Ben-Akiva and Lerman (1994).

There is a body of work on assortment optimization problems under various discrete choice models. In the assortment optimization setting, the prices of the products are fixed and we choose the set of products to offer given that customers choose among the offered products according to a particular choice model. Talluri and van Ryzin (2004) study assortment problems when customers choose under the multinomial logit model and show that the optimal assortment includes a certain number of products with the largest revenues. As a result, the optimal assortment can efficiently be found by checking the performance of every assortment that includes a certain number of products with the largest revenues. Rusmevichientong et al. (2010) consider the same problem with a constraint on the number of products in the offered assortment and show that the problem can be solved in a tractable fashion. Wang (2012*a*) extends this work to more general versions of the multinomial logit model. In Bront et al. (2009), Mendez-Diaz et al. (2010) and Rusmevichientong et al. (2013), there are multiple types of customers, each choosing according to the multinomial logit model with different parameters. The authors show that the assortment problem becomes NP-hard in weak and strong sense, propose approximation methods and study integer programming formulations. Jagabathula et al. (2011) work on how to obtain good assortments with only limited computations of the expected revenue from different assortments. The work mentioned so far in this paragraph uses the multinomial logit model, but there are extensions to the nested logit model. Rusmevichientong et al. (2009) develop an approximation scheme for assortment problems when customers choose under the nested logit model and there is a shelf space constraint for the offered assortment. Davis et al. (2011) study the same problem without the shelf space constraint and give a tractable method to obtain the optimal assortment under the nested logit model. Gallego and Topaloglu (2012) show that it is tractable to obtain the optimal assortment when customers choose according to the nested logit model and there is a cardinality constraint on the number of products offered in each nest. They extend their result to the situation where each product can be offered at a finite number of price levels and one needs to jointly choose the assortment of products to offer and their corresponding price levels. Their approach does not work when the set of products to be offered is fixed and not under the control of the decision maker.

Pricing problems within the context of different discrete choice models is also an active research area. Under the multinomial logit model, Hanson and Martin (1996) note that the expected revenue function is not concave in prices. However, Song and Xue (2007) and Dong et al. (2009) make progress by formulating the problem in terms of market shares, as this formulation yields a concave expected revenue function. 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 dissimilarity parameters are less than one. Gallego and Wang (2011) relax both of the assumptions in Li and Huh (2011) and extend the analysis to more general forms of the nested logit model. Wang (2012*b*) considers a joint assortment and price optimization problem to choose the offered products and their prices. The author imposes cardinality constraints on the offered assortment, but the customer choices are captured by using the multinomial logit model, which is more restrictive than the nested logit model. Recently, there is interest in modeling large scale revenue management problems by incorporating the fact that customers make a choice depending on the assortment of available itinerary products and their prices. The main approach in these models is to formulate deterministic approximations under the assumption that customer arrivals and choices are deterministic. Such deterministic approximations have a large number of decision variables and they are usually solved by using column generation. The assortment and pricing problems described in this and the paragraph above become instrumental when solving the column generation subproblems. Deterministic approximations for large scale revenue management can be found in Gallego et al. (2004), Liu and van Ryzin (2008), Kunnumkal and Topaloglu (2008), Zhang and Adelman (2009), Zhang and Lu (2011) and Meissner et al. (2012).

Organization. In Section 2, we formulate the first variant of the problem, where the set of products to be offered is fixed and we choose the prices for these products. In Section 3, we show that this problem can be visualized as finding the fixed point of a scalar function. In Section 4, we develop an approximation framework by using the fixed point representation and computing a scalar function at a finite number of grid points. In Section 5, we show how to construct an appropriate grid with a performance guarantee and give our approximation method. In Section 6, we extend the work in the earlier sections to the second variant of the problem, where we jointly choose the products to offer and their corresponding prices. In Section 7, we show how to obtain an upper bound on the optimal expected revenue and give computational experiments to compare the performance of our approximation methods with the upper bounds on the optimal expected revenues. In Section 8, we conclude. In Appendices A and B, we give the proofs that are omitted in the paper. In Appendix C, we give a glossary to collect the crucial pieces of notation used throughout the paper.

2 Problem Formulation

In this section, we describe the nested logit model and formulate the pricing problem. There are m nests indexed by $M = \{1, \ldots, m\}$. Depending on the application setting, nests may correspond to different retail stores, different product categories or different sales channels. In each nest there are n products and we index the products by $N = \{1, \ldots, n\}$. We use p_{ij} to denote the price of product j in nest i. The price of product j in nest i has to satisfy the price bound constraint $p_{ij} \in [l_{ij}, u_{ij}]$, for the upper and lower bound parameters $l_{ij}, u_{ij} \in [0, \infty)$. We use w_{ij} to denote the preference weight of product j in nest i. Under the nested logit model, if we choose the price of product j in nest i as p_{ij}, then the preference weight of this product is $w_{ij} = \exp(\alpha_{ij} - \beta_{ij}\, p_{ij})$, where $\alpha_{ij} \in (-\infty, \infty)$ and $\beta_{ij} \in [0, \infty)$ are parameters capturing the effect of the price on the preference weight. Since there is a one to one correspondence between the price and preference weight of a product, throughout the paper, we assume that we choose the preference weight of a product, in which case, there is a price corresponding to the chosen preference weight. In particular, if we choose the preference weight of product j in nest i as w_{ij}, then the corresponding price of this product is $p_{ij} = (\alpha_{ij} - \log w_{ij})/\beta_{ij}$, which is obtained by setting $w_{ij} = \exp(\alpha_{ij} - \beta_{ij}\, p_{ij})$ and solving for p_{ij}. For brevity, we let $\kappa_{ij} = \alpha_{ij}/\beta_{ij}$ and $\eta_{ij} = 1/\beta_{ij}$ and write the relationship between price and preference weight as $p_{ij} = \kappa_{ij} - \eta_{ij} \log w_{ij}$. Noting the upper and lower bound constraint on prices, the preference weight of product j in nest i has to satisfy the constraint $w_{ij} \in [L_{ij}, U_{ij}]$ with $L_{ij} = \exp(\alpha_{ij} - \beta_{ij}\, u_{ij})$ and $U_{ij} = \exp(\alpha_{ij} - \beta_{ij}\, l_{ij})$. We use $\boldsymbol{w}_i = (w_{i1}, \ldots, w_{in})$ to denote the vector of preference weights of the products in nest i. Under the nested logit model, if we choose the preference weights of the products in nest i as $\boldsymbol{w}_i$ and a customer decides to make a purchase in this nest, then this customer purchases product j in nest i with probability $w_{ij}/\sum_{k\in N} w_{ik}$. Thus, if we choose the preference weights of the products in nest i as $\boldsymbol{w}_i$ and a customer decides to make a purchase in this nest, then we obtain an expected revenue of

$$R_i(\boldsymbol{w}_i) = \sum_{j\in N} \frac{w_{ij}}{\sum_{k\in N} w_{ik}} \left(\kappa_{ij} - \eta_{ij} \log w_{ij}\right) = \frac{\sum_{j\in N} w_{ij}\left(\kappa_{ij} - \eta_{ij} \log w_{ij}\right)}{\sum_{j\in N} w_{ij}},$$

where the term $w_{ij}/\sum_{k\in N} w_{ik}$ on the left side above is the probability that a customer purchases product j in nest i given this customer decides to make a purchase in this nest, whereas the term $\kappa_{ij} - \eta_{ij} \log w_{ij}$ captures the revenue associated with product j in nest i.

Each nest i has a parameter $\gamma_i \in (0, 1]$, characterizing the degree of dissimilarity between the products in this nest. In this case, if we choose the preference weights of the products in all nests as $(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m)$, then a customer decides to make a purchase in nest i with probability

$$Q_i(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m) = \frac{\left(\sum_{j\in N} w_{ij}\right)^{\gamma_i}}{1 + \sum_{l\in M}\left(\sum_{j\in N} w_{lj}\right)^{\gamma_l}}.$$

Depending on the interpretation of a nest as a retail store, a product category or a sales channel, the expression above computes the probability that a customer chooses a particular retail store, product category or sales channel as a function of the preference weights of all products. With probability $1 - \sum_{i \in M} Q_i(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m)$, a customer leaves without making a purchase. McFadden (1984) demonstrates that the choice probabilities above can be derived from a utility maximization principle, where a customer associates a random utility with each product and purchases the product that provides the largest utility. Thus, if we choose the preference weights as $(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m)$ over all nests, then we obtain an expected revenue of

$$\begin{aligned}\Pi(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m) &= \sum_{i \in M} Q_i(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m)\, R_i(\boldsymbol{w}_i) \\ &= \frac{1}{1 + \sum_{i \in M} \big(\sum_{j \in N} w_{ij}\big)^{\gamma_i}} \sum_{i \in M} \Big(\sum_{j \in N} w_{ij}\Big)^{\gamma_i} \; \frac{\sum_{j \in N} w_{ij}\,(\kappa_{ij} - \eta_{ij} \log w_{ij})}{\sum_{j \in N} w_{ij}}, \end{aligned} \tag{1}$$

where the second equality is by the definitions of $R_i(\boldsymbol{w}_i)$ and $Q_i(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m)$. Our goal is to choose the preference weights to maximize the expected revenue, yielding the problem

$$Z^* = \max \big\{ \Pi(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m) : \boldsymbol{w}_i \in [\boldsymbol{L}_i, \boldsymbol{U}_i] \;\; \forall\, i \in M \big\}, \tag{2}$$

where we use $\boldsymbol{L}_i$ and $\boldsymbol{U}_i$ to respectively denote the vectors $(L_{i1}, \ldots, L_{in})$ and $(U_{i1}, \ldots, U_{in})$ and interpret the constraint $\boldsymbol{w}_i \in [\boldsymbol{L}_i, \boldsymbol{U}_i]$ componentwise as $w_{ij} \in [L_{ij}, U_{ij}]$ for all $j \in N$.

We close this section with a remark on our formulation of the nested logit model. In our formulation of the nested logit model, if a customer chooses a particular nest, then this customer must purchase one of the products offered in this nest. We can extend our model to allow the possibility that a customer may leave without purchasing anything even after choosing a particular nest. To make this extension, we use w_{i0} to denote the preference weight of the no purchase option in nest i, in which case, if a customer decides to make a purchase in nest i, then this customer leaves the nest without purchasing anything with probability $w_{i0}/(w_{i0} + \sum_{j \in N} w_{ij})$. The preference weight w_{i0} is a constant, not depending on the prices of any of the products. It turns out that our results continue to hold when we allow customers to leave a nest without purchasing anything. We come back to this extension at appropriate places in the paper.

3 Fixed Point Representation

In this section, we show that problem (2) can alternatively be represented as the problem of computing the fixed point of an appropriately defined scalar function. This alternative fixed point representation allows us to work with a single decision variable for each nest i, rather than n decision variables $\boldsymbol{w}_i = (w_{i1}, \ldots, w_{in})$ and it becomes crucial when developing our approximation methods. To that end, assume that we compute the value of z that satisfies

$$z = \sum_{i \in M} \max_{\boldsymbol{w}_i \in [\boldsymbol{L}_i, \boldsymbol{U}_i]} \left\{ \Big(\sum_{j \in N} w_{ij}\Big)^{\gamma_i} \; \frac{\sum_{j \in N} w_{ij}\,(\kappa_{ij} - \eta_{ij} \log w_{ij})}{\sum_{j \in N} w_{ij}} - \Big(\sum_{j \in N} w_{ij}\Big)^{\gamma_i} z \right\}. \tag{3}$$

Viewing the right side of (3) as a function of z, finding the value of z satisfying (3) is equivalent to computing the fixed point of this scalar function. There always exists such a unique value of z

since the left side above is strictly increasing and the right side above is decreasing in z. Letting $\hat{z}$ be the value of z satisfying (3), we claim that $\hat{z}$ is the optimal objective value of problem (2). To see this claim, note that if $(\boldsymbol{w}_1^*, \ldots, \boldsymbol{w}_m^*)$ is an optimal solution to problem (2), then we have

$$\hat{z} \geq \sum_{i \in M} \left\{ \Big(\sum_{j \in N} w_{ij}^* \Big)^{\gamma_i} \frac{\sum_{j \in N} w_{ij}^* \left(\kappa_{ij} - \eta_{ij} \log w_{ij}^* \right)}{\sum_{j \in N} w_{ij}^*} - \Big(\sum_{j \in N} w_{ij}^* \Big)^{\gamma_i} \hat{z} \right\},$$

where we use the fact that $\hat{z}$ is the value of z satisfying (3) and $\boldsymbol{w}_i^*$ is a feasible but not necessarily an optimal solution to the maximization problem on the right side of (3) when we solve this problem with $z = \hat{z}$. In the inequality above, if we collect all terms that involve $\hat{z}$ on the left side of the inequality, solve for $\hat{z}$ and use the definition of $\Pi(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m)$ in (1), then it follows that $\hat{z} \geq \Pi(\boldsymbol{w}_1^*, \ldots, \boldsymbol{w}_m^*) = Z^*$. On the other hand, if we let $\hat{\boldsymbol{w}}_i$ be an optimal solution to the maximization problem on the right side of (3) when we solve this problem with $z = \hat{z}$, then we observe that $(\hat{\boldsymbol{w}}_1, \ldots, \hat{\boldsymbol{w}}_m)$ is a feasible solution to problem (2). Furthermore, since $\hat{z}$ is the value of z that satisfies (3), the definition of $\hat{\boldsymbol{w}}_i$ implies that

$$\hat{z} = \sum_{i \in M} \left\{ \Big(\sum_{j \in N} \hat{w}_{ij} \Big)^{\gamma_i} \frac{\sum_{j \in N} \hat{w}_{ij} \left(\kappa_{ij} - \eta_{ij} \log \hat{w}_{ij} \right)}{\sum_{j \in N} \hat{w}_{ij}} - \Big(\sum_{j \in N} \hat{w}_{ij} \Big)^{\gamma_i} \hat{z} \right\}. \tag{4}$$

If we solve for $\hat{z}$ in the equality above and use the definition of $\Pi(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m)$ in (1) once more, then we get $\hat{z} = \Pi(\hat{\boldsymbol{w}}_1, \ldots, \hat{\boldsymbol{w}}_m) \leq Z^*$, where the last inequality uses the fact that $(\hat{\boldsymbol{w}}_1, \ldots, \hat{\boldsymbol{w}}_m)$ is a feasible but not necessarily an optimal solution to problem (2). So, we obtain $\hat{z} = Z^*$, establishing the claim. Thus, we can obtain the optimal objective value of problem (2) by finding the value of z that satisfies (3). Furthermore, if we use $\hat{z}$ to denote such a value of z and $\hat{\boldsymbol{w}}_i$ to denote an optimal solution to the maximization problem on the right side of (3) when this problem is solved with $z = \hat{z}$, then the discussion in this paragraph establishes that $(\hat{\boldsymbol{w}}_1, \ldots, \hat{\boldsymbol{w}}_m)$ is an optimal solution to problem (2). Since the left and right sides of (3) are respectively increasing and decreasing in z, we can find the value of z satisfying (3) by using bisection search. However, one drawback of using bisection search is that we need to solve the maximization problem on the right side of (3) for each value of z visited during the course of the search. This maximization problem involves a high dimensional objective function. Also, it is not difficult to generate counterexamples to show that this objective is not necessarily concave.

To get around the necessity of dealing with high dimensional and nonconcave objective functions, we give an alternative approach for finding the value of z satisfying (3). We define $g_i(y_i)$ as the optimal objective value of the nonlinear knapsack problem

$$g_i(y_i) = \max \left\{ \sum_{j \in N} w_{ij} \left(\kappa_{ij} - \eta_{ij} \log w_{ij} \right) : \sum_{j \in N} w_{ij} \leq y_i, \; w_{ij} \in [L_{ij}, U_{ij}] \;\; \forall j \in N \right\}. \tag{5}$$

We make a number of observations regarding problem (5). We can verify that the objective function of this problem is concave. Also, if we do not have the first constraint in the problem above, then by using the first order condition for the objective function of this problem, we can check that the optimal value of the decision variable w_{ij} is given by $\min\{\max\{\exp(\kappa_{ij}/\eta_{ij} - 1), L_{ij}\}, U_{ij}\}$ for all $j \in N$. Thus, letting $\bar{U}_i = \sum_{j\in N} \min\{\max\{\exp(\kappa_{ij}/\eta_{ij} - 1), L_{ij}\}, U_{ij}\}$, if we have $y_i > \bar{U}_i$, then the first constraint in problem (5) is not tight at the optimal solution. On the other hand, letting $\bar{L}_i = \sum_{j\in N} L_{ij}$, if we have $y_i < \bar{L}_i$, then problem (5) is infeasible. Finally, if we have $y_i \in [\bar{L}_i, \bar{U}_i]$, then it follows that the first constraint in problem (5) is always tight at the optimal solution. Thus, intuitively speaking, the interesting values for y_i take values in the interval $[\bar{L}_i, \bar{U}_i]$. In this case, noting that problem (5) finds the maximum value of the numerator of the fraction in (3) while keeping the denominator of this fraction below y_i, instead of finding the value of z satisfying (3), we propose finding the value of z that satisfies

$$z = \sum_{i\in M} \max_{y_i\in[\bar{L}_i,\bar{U}_i]} \left\{ y_i^{\gamma_i} \frac{g_i(y_i)}{y_i} - y_i^{\gamma_i} z \right\}. \tag{6}$$

The value of z satisfying (6) is unique since the left side above is strictly increasing and the right side above is decreasing in z. The maximization problem on the right side above involves a scalar decision variable and the computation of $g_i(y_i)$ requires solving a convex optimization problem. In the next proposition, we show that (6) can be used to find the value of z satisfying (3).

Proposition 1 (Fixed Point Representation)**.** *The value of z that satisfies* (3) *and* (6) *are the same, corresponding to the optimal objective value of problem* (2)*.*

Proof. The value of z that satisfies (3) or (6) has to be positive. Otherwise, the left sides of these expressions evaluate to a negative number, but the right sides evaluate to a positive number. In this case, comparing (3) and (6), if we can show that

$$\max_{\boldsymbol{w}_i\in[\boldsymbol{L}_i,\boldsymbol{U}_i]} \left\{ \Big(\sum_{j\in N} w_{ij}\Big)^{\gamma_i} \frac{\sum_{j\in N} w_{ij}\,(\kappa_{ij} - \eta_{ij}\log w_{ij})}{\sum_{j\in N} w_{ij}} - \Big(\sum_{j\in N} w_{ij}\Big)^{\gamma_i} z \right\} = \max_{y_i\in[\bar{L}_i,\bar{U}_i]} \left\{ y_i^{\gamma_i} \frac{g_i(y_i)}{y_i} - y_i^{\gamma_i} z \right\}$$

for any $z > 0$, then the value of z that satisfies (3) and (6) are the same. The equality above can be established by showing that we can use the optimal solution to one of the problems above to construct a feasible solution to the other. We defer the details to Appendix A. □

The proposition above provides a tempting approach for solving problem (2). In particular, we can find the value of z that satisfies (6) by using bisection search. We observe that the maximization problem on the right side of (6) involves a scalar decision variable and the computation of $g_i(\cdot)$ requires solving a convex optimization problem. Thus, the optimization problems that we solve during the course of the bisection search may be tractable. We use $\hat{z}$ to denote the value of z that satisfies (6) and $\hat{y}_i$ to denote an optimal solution to the maximization problem on the right side of (6) when we solve this problem with $z = \hat{z}$. In this case, we can solve problem (5) with $y_i = \hat{y}_i$ to obtain an optimal solution $\hat{\boldsymbol{w}}_i$. Once we solve problem (5) with $y_i = \hat{y}_i$ for all $i \in M$, it follows that $(\hat{\boldsymbol{w}}_1, \ldots, \hat{\boldsymbol{w}}_m)$ is an optimal solution to problem (2).

It is possible to check that all of the discussion in this section holds when we use the no purchase preference weight w_{i0} to allow a customer to leave nest i without making a purchase. To make this extension, all we need to do is to replace $\sum_{j\in N} w_{ij}$ in (3) and (5) with $w_{i0} + \sum_{j\in N} w_{ij}$.

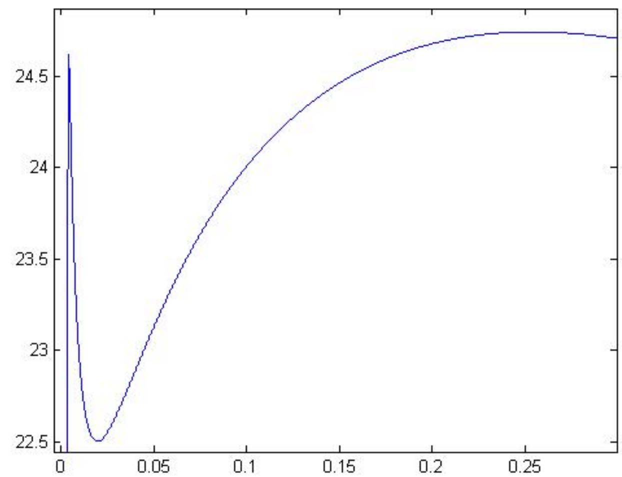

Figure 1: The function $y_1^{\gamma_1} (g_1(y_1)/y_1) - y_1^{\gamma_1} z$ as a function of y_1.

4 Approximation Framework

As mentioned at the end of the previous section, the maximization problem on the right side of (6) involves a scalar decision variable and it is tempting to try to solve problem (2) by finding the value of z satisfying (6). Unfortunately, it turns out that the objective function of this maximization problem is not unimodal and it can be intractable to solve the maximization problem on the right side of (6). To give an example where the objective function of the maximization problem on the right side of (6) is not unimodal, consider a case with a single nest and seven products. The problem parameters are given by $\gamma_1 = 0.4$, $(\alpha_{11}, \ldots, \alpha_{17}) = (2.1, 1.0, 1.7, 1.4, 1.0, 12.0, 13.0)$, $(\beta_{11}, \ldots, \beta_{17}) = (0.07, 0.07, 0.07, 0.07, 0.07, 0.07, 0.07)$, $(l_{11}, \ldots, l_{17}) = (30, 30, 30, 30, 30, 251, 330)$ and $(u_{11}, \ldots, u_{17}) = (200, 200, 200, 200, 200, 368, 383)$. For this problem instance, Figure 1 plots the objective function of the maximization problem on the right side of (6) as a function of y_1, fixing z at 24.74 and shows that this objective function is not necessarily unimodal. We note that the value of z that we use in this figure is sensible as the optimal objective value of problem (2) is close to 24.74 for this problem instance. So, we do not have unimodality even with sensible values of z. Interestingly, Gallego and Wang (2011) consider the case where there are no lower or upper bounds on the prices. The authors show that if the dissimilarity parameters of the nests satisfy $\gamma_i \geq 1 - \min_{j \in N} \beta_{ij} / \max_{j \in N} \beta_{ij}$ for all $i \in M$, then the objective function of the maximization problem on the right side of (6) is always unimodal. In the example above, we indeed have $\gamma_i \geq 1 - \min_{j \in N} \beta_{ij} / \max_{j \in N} \beta_{ij}$ for all $i \in M$, indicating that this example satisfies the condition in Gallego and Wang (2011). However, due to the presence of the lower and upper bounds on the prices, we lose the unimodality property.

The objective function of the maximization problem on the right side of (6) is not necessarily unimodal, but since this objective function is scalar, a possible strategy is to construct a grid over the interval $[\bar{L}_i, \bar{U}_i]$ and check the values of the objective function only at the grid points. To pursue this line of thought, we use $\{\tilde{y}_i^t : t = 1, \ldots, T_i\}$ to denote a collection of grid points such that $\tilde{y}_i^t \leq \tilde{y}_i^{t+1}$ for all $t = 1, \ldots, T_i - 1$. Furthermore, the collection of grid points should satisfy $\tilde{y}_i^1 = \bar{L}_i$ and $\tilde{y}_i^{T_i} = \bar{U}_i$ to make sure that the grid points cover the interval $[\bar{L}_i, \bar{U}_i]$. In this case, instead of considering all values of y_i over the interval $[\bar{L}_i, \bar{U}_i]$ as we do in (6), we can focus only on the grid points and find the value of z that satisfies

$$z = \sum_{i \in M} \max_{y_i \in \{\tilde{y}_i^t \,:\, t=1,\ldots,T_i\}} \left\{ y_i^{\gamma_i} \frac{g_i(y_i)}{y_i} - y_i^{\gamma_i} z \right\}. \tag{7}$$

The important question is that what properties the grid should possess so that the solution obtained by limiting our attention only to the grid points has a quantifiable performance guarantee. In the next theorem, we show that if the optimal objective value $g_i(y_i)$ of the knapsack problem in (5) does not change too much at the successive grid points, then we can build on the value of z satisfying (7) to construct a solution to problem (2) with a certain performance guarantee.

Theorem 2 (Requirements for a Good Grid)**.** *For some $\rho \geq 0$, assume that the collection of grid points $\{\tilde{y}_i^t : t = 1, \ldots, T_i\}$ satisfy $g_i(\tilde{y}_i^{t+1}) \leq (1+\rho)\, g_i(\tilde{y}_i^t)$ for all $t = 1, \ldots, T_i - 1$, $i \in M$. If $\hat{z}$ denotes the value of z that satisfies* (7) *and Z^* denotes the optimal objective value of problem* (2)*, then we have $(1+\rho)\,\hat{z} \geq Z^*$.*

Proof. To get a contradiction, assume that $(1+\rho)\,\hat{z} < Z^*$. For all $i \in M$, we let y_i^* be an optimal solution to the maximization problem on the right side of (6) when this problem is solved with $z = Z^*$. Furthermore, we let $t_i \in \{1, \ldots, T_i - 1\}$ be such that $y_i^* \in [\tilde{y}_t^{t_i}, \tilde{y}_i^{t_i+1}]$. We have

$$\frac{1}{1+\rho} Z^* > \hat{z} \geq \sum_{i \in M} \left\{ \left(\tilde{y}_i^{t_i}\right)^{\gamma_i} \frac{g_i(\tilde{y}_i^{t_i})}{\tilde{y}_i^{t_i}} - \left(\tilde{y}_i^{t_i}\right)^{\gamma_i} \hat{z} \right\} \geq \sum_{i \in M} \left\{ \frac{1}{1+\rho} \left(\tilde{y}_i^{t_i}\right)^{\gamma_i} \frac{g_i(y_i^*)}{\tilde{y}_i^{t_i}} - \left(\tilde{y}_i^{t_i}\right)^{\gamma_i} \hat{z} \right\},$$

where the second inequality follows from the fact that $\hat{z}$ corresponds to the value of z that satisfies (7) and $\tilde{y}_i^{t_i}$ is a feasible but not necessarily an optimal solution to the maximization problem on the right side of (7) when this problem is solved with $z = \hat{z}$. To see that the third inequality holds, we observe that $g_i(\cdot)$ is increasing, in which case, since $y_i^* \in [\tilde{y}_i^{t_i}, \tilde{y}_i^{t_i+1}]$, we obtain $g_i(y_i^*) \leq g_i(\tilde{y}_i^{t_i+1}) \leq (1+\rho)\, g_i(\tilde{y}_i^{t_i})$. In this case, noting that $\gamma_i \leq 1$ and $\tilde{y}_i^{t_i} \leq y_i^*$ so that $(\tilde{y}_i^{t_i})^{1-\gamma_i} \leq (y_i^*)^{1-\gamma_i}$, we continue the chain of inequalities above as

$$\begin{aligned} \sum_{i \in M} \left\{ \frac{1}{1+\rho} \left(\tilde{y}_i^{t_i}\right)^{\gamma_i} \frac{g_i(y_i^*)}{\tilde{y}_i^{t_i}} - \left(\tilde{y}_i^{t_i}\right)^{\gamma_i} \hat{z} \right\} &\geq \sum_{i \in M} \left\{ \frac{1}{1+\rho} \left(y_i^*\right)^{\gamma_i} \frac{g_i(y_i^*)}{y_i^*} - \left(y_i^*\right)^{\gamma_i} \hat{z} \right\} \\ &\geq \frac{1}{1+\rho} \sum_{i \in M} \left\{ \left(y_i^*\right)^{\gamma_i} \frac{g_i(y_i^*)}{y_i^*} - \left(y_i^*\right)^{\gamma_i} Z^* \right\}, \end{aligned}$$

where the second inequality uses the assumption that $(1+\rho)\,\hat{z} < Z^*$. By using the last two displayed chains of inequalities and noting the definition of y_i^*, it follows that

$$Z^* > \sum_{i \in M} \left\{ \left(y_i^*\right)^{\gamma_i} \frac{g_i(y_i^*)}{y_i^*} - \left(y_i^*\right)^{\gamma_i} Z^* \right\} = \sum_{i \in M} \max_{y_i \in [\bar{L}_i, \bar{U}_i]} \left\{ y_i^{\gamma_i} \frac{g_i(y_i)}{y_i} - y_i^{\gamma_i} Z^* \right\}.$$

By Proposition 1, Z^* corresponds to the value of z that satisfies (6), but the last chain of inequalities above shows that Z^* does not satisfy (6), which is a contradiction. $\square$

When we work with grid points that satisfy the assumption of Theorem 2, this theorem allows us to obtain a $(1+\rho)$-approximate solution to problem (2) in the following fashion. We find the value of z that satisfies (7) and use $\hat{z}$ to denote this value. We let $\hat{y}_i$ be an optimal solution to the maximization problem on the right side of (7) when this problem is solved with $z = \hat{z}$. For all $i \in M$, we solve problem (5) with $y_i = \hat{y}_i$ and use $\hat{\boldsymbol{w}}_i$ to denote an optimal solution to this problem. In this case, it is possible to show that the solution $(\hat{\boldsymbol{w}}_1, \ldots, \hat{\boldsymbol{w}}_m)$ provides an expected revenue that deviates from the optimal expected revenue by at most a factor of $1+\rho$, satisfying $(1+\rho)\,\Pi(\hat{\boldsymbol{w}}_1, \ldots, \hat{\boldsymbol{w}}_m) \geq Z^*$. To see this result, we note that since $\hat{z}$ is the value of z that satisfies (7) and $\hat{y}_i$ is an optimal solution to the maximization problem on the right side of (7) when this problem is solved with $z = \hat{z}$, we have

$$\hat{z} = \sum_{i \in M} \left\{ \hat{y}_i^{\gamma_i} \, \frac{g_i(\hat{y}_i)}{\hat{y}_i} - \hat{y}_i^{\gamma_i} \, \hat{z} \right\}. \tag{8}$$

Also, since $\hat{y}_i \in [\bar{L}_i, \bar{U}_i]$, the discussion right after the formulation of problem (5) shows that the first constraint in this problem must be tight at the optimal solution when this problem is solved with $y_i = \hat{y}_i$. Therefore, noting that $\hat{\boldsymbol{w}}_i$ is an optimal solution to problem (5) when we solve this problem with $y_i = \hat{y}_i$, we obtain $\hat{y}_i = \sum_{j \in N} \hat{w}_{ij}$ and $g_i(\hat{y}_i) = \sum_{j \in N} \hat{w}_{ij}\,(\kappa_{ij} - \eta_{ij} \log \hat{w}_{ij})$ for all $i \in M$. Replacing $\hat{y}_i$ and $g_i(\hat{y}_i)$ in (8) by their equivalents given by the last two equalities, we observe that $\hat{z}$ and $(\hat{\boldsymbol{w}}_1, \ldots, \hat{\boldsymbol{w}}_m)$ satisfy the equality in (4). So, if we collect all terms that involve $\hat{z}$ on the left side of (4), solve for $\hat{z}$ and use the definition of $\Pi(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m)$, then we get $\hat{z} = \Pi(\hat{\boldsymbol{w}}_1, \ldots, \hat{\boldsymbol{w}}_m)$. When the grid points satisfy the assumption of Theorem 2, we also have $(1+\rho)\,\hat{z} \geq Z^*$. So, we obtain $(1+\rho)\,\Pi(\hat{\boldsymbol{w}}_1, \ldots, \hat{\boldsymbol{w}}_m) \geq Z^*$, showing that the expected revenue from the solution $(\hat{\boldsymbol{w}}_1, \ldots, \hat{\boldsymbol{w}}_m)$ deviates from the optimal by at most a factor of $1+\rho$.

The preceding discussion, along with Theorem 2, gives a framework for obtaining approximate solutions to problem (2) with a performance guarantee. The crucial point is that the collection of grid points $\{\tilde{y}_i^t : t = 1, \ldots, T_i\}$ has to satisfy the assumption of Theorem 2. Also, the number of grid points in this collection should be reasonably small to be able to solve the maximization problem on the right side of (7) quickly. In the next section, we show that it is indeed possible to construct a reasonably small collection of grid points that satisfies the assumption of Theorem 2. Before doing so, however, we make a brief remark on how to find the value of z that satisfies (7). Thus far, we propose bisection search as a possible method to obtain this value of z. One shortcoming of bisection search is that it may not terminate in finite time. To get around the fact that bisection search may not terminate in finite time, we demonstrate that it is possible to obtain the value of z satisfying (7) by solving a linear program.

To formulate the linear program, we note that the left side of the equality in (7) is increasing in z, whereas the right side is decreasing. Therefore, the value of z that satisfies (7) corresponds to the smallest value of z such that the left side of the equality in (7) is still greater than or equal to the right side. This observation immediately implies that finding the value of z satisfying (7) is equivalent to solving the problem

$$\min \left\{ z : z \geq \sum_{i \in M} \max_{y_i \in \{\tilde{y}_i^t \,:\, t=1,\ldots,T_i\}} \left\{ y_i^{\gamma_i} \, \frac{g_i(y_i)}{y_i} - y_i^{\gamma_i} z \right\} \right\}.$$

If we define the additional decision variables $(x_1, \ldots, x_m)$ so that x_i represents the optimal objective value of the maximization problem in the ith term of the sum on the right side of the constraint above, then the problem above can be written as

$$\min \left\{ z : z \geq \sum_{i \in M} x_i, \;\; x_i \geq y_i^{\gamma_i} \, \frac{g_i(y_i)}{y_i} - y_i^{\gamma_i} \, z \quad \forall \, y_i \in \{\tilde{y}_i^t : t = 1, \ldots, T_i\}, \;\; i \in M \right\}, \tag{9}$$

where the decision variables are z and $(x_1, \ldots, x_m)$. The problem above is a linear program with $1 + m$ decision variables and $1 + \sum_{i \in M} T_i$ constraints. So, as long as the number of grid points is not too large, we can solve a tractable linear program to obtain the value of z satisfying (7).

5 Grid Construction

In this section, our goal is to show how we can construct a reasonably small collection of grid points $\{\tilde{y}_i^t : t = 1, \ldots, T_i\}$ that satisfies the assumption of Theorem 2. By noting the discussion that follows Theorem 2 in the previous section, such a collection of grid points allows us to obtain a solution to problem (2) with a given approximation guarantee. To construct the collection of grid points, we begin by giving a number of fundamental properties of the knapsack problem in (5). After we give these properties, we proceed to showing how we can build on these properties to construct the collection of grid points.

5.1 Properties of Knapsack Problems

The first property that we have for problem (5) is that the optimal values of the decision variables in this problem are monotonically increasing in y_i as long as $y_i \in [\bar{L}_i, \bar{U}_i]$. To see this property, we associate the Lagrange multiplier λ_i with the first constraint in problem (5) and write the Lagrangian as $L_i(\boldsymbol{w}_i, \lambda_i) = \sum_{j \in N} w_{ij} \, (\kappa_{ij} - \eta_{ij} \log w_{ij} - \lambda_i) + \lambda_i \, y_i$, which is a concave function of $\boldsymbol{w}_i$. Maximizing the Lagrangian $L_i(\boldsymbol{w}_i, \lambda_i)$ subject to the constraints that $w_{ij} \in [L_{ij}, U_{ij}]$ for all $j \in N$, the optimal solution to problem (5) can be obtained by setting

$$w_{ij} = \min \left\{ \max \left\{ \exp \left(\frac{\kappa_{ij}}{\eta_{ij}} - 1 - \frac{\lambda_i}{\eta_{ij}} \right), L_{ij} \right\}, U_{ij} \right\} \tag{10}$$

for all $j \in N$. We observe that the expression on the right side above is decreasing in λ_i, showing that the optimal value of the decision variable w_{ij} is decreasing in the optimal value of the Lagrange multiplier. On the other hand, since we have $y_i \in [\bar{L}_i, \bar{U}_i]$, by the discussion that follows the formulation of problem (5), the first constraint in this problem must be tight at the optimal solution. Therefore, noting (10), the optimal value of the Lagrange multiplier λ_i satisfies the equality $\sum_{j\in N} \min\{\max\{\exp(\kappa_{ij}/\eta_{ij} - 1 - \lambda_i/\eta_{ij}), L_{ij}\}, U_{ij}\} = y_i$. The expression on the left side of this equality is decreasing in λ_i, which implies that the optimal value of the Lagrange multiplier is decreasing in the right side of the first constraint in problem (5). To sum up, if we use $\lambda_i^*(y_i)$ to denote the optimal value of the Lagrange multiplier for the first constraint in problem (5) as a function of the right side of this constraint, then $\lambda_i^*(y_i)$ satisfies

$$\sum_{j\in N} \min\left\{ \max\left\{ \exp\left(\frac{\kappa_{ij}}{\eta_{ij}} - 1 - \frac{\lambda_i^*(y_i)}{\eta_{ij}}\right), L_{ij} \right\}, U_{ij} \right\} = y_i. \tag{11}$$

Furthermore, $\lambda_i^*(y_i)$ is decreasing in y_i. Since the optimal value of the decision variable w_{ij} in problem (5) is decreasing in the optimal value of the Lagrange multiplier and $\lambda_i^*(y_i)$ is decreasing in y_i, it follows that the optimal value of the decision variable w_{ij} in problem (5) is increasing in y_i, as desired. Therefore, we can let ζ_{ij} and ξ_{ij} be such that

$$\exp\left(\frac{\kappa_{ij}}{\eta_{ij}} - 1 - \frac{\lambda_i^*(\zeta_{ij})}{\eta_{ij}}\right) = L_{ij} \quad \text{and} \quad \exp\left(\frac{\kappa_{ij}}{\eta_{ij}} - 1 - \frac{\lambda_i^*(\xi_{ij})}{\eta_{ij}}\right) = U_{ij}, \tag{12}$$

in which case, (10) implies that if $y_i = \zeta_{ij}$, then we have $w_{ij} = L_{ij}$ in the optimal solution to problem (5), whereas if $y_i = \xi_{ij}$, then we have $w_{ij} = U_{ij}$. Also, since the optimal value of the decision variable w_{ij} is increasing in y_i, the optimal value of the decision variable w_{ij} in problem (5) satisfies $w_{ij} = L_{ij}$ for all $y_i \leq \zeta_{ij}$, whereas $w_{ij} = U_{ij}$ for all $y_i \geq \xi_{ij}$. In this way, ζ_{ij} and ξ_{ij} correspond to the two threshold values of the right side of the first constraint in problem (5) such that if $y_i \leq \zeta_{ij}$, then the optimal value of the decision variable w_{ij} is always L_{ij}, whereas if $y_i \geq \xi_{ij}$, then the optimal value of the decision variable w_{ij} is always U_{ij}.

We note that there may not exist a value of ζ_{ij} or ξ_{ij} satisfying (12). If this is the case, then we set $\zeta_{ij} = -\infty$ or $\xi_{ij} = \infty$. Building on the discussion above, we obtain the next lemma.

Lemma 3 (Intervals). *For any $j \in N$, there exists an interval $[\zeta_{ij}, \xi_{ij}]$ such that the optimal value of the decision variable w_{ij} in problem* (5) *satisfies $w_{ij} = L_{ij}$ when we have $y_i \leq \zeta_{ij}$, whereas $w_{ij} = U_{ij}$ when we have $y_i \geq \xi_{ij}$. Furthermore, if $y_i \in [\zeta_{ij}, \xi_{ij}]$, then we can drop the constraint $w_{ij} \in [L_{ij}, U_{ij}]$ in problem* (5) *without changing the optimal solution to this problem.*

Proof. We let ζ_{ij} and ξ_{ij} be as defined in (12), in which case, the first part follows from the discussion right before the lemma. To show the second part, we let $\boldsymbol{w}_i^*$ be the optimal solution to problem (5) and $\lambda_i^*(y_i)$ be the corresponding Lagrange multiplier for the first constraint. Since $y_i \in [\zeta_{ij}, \xi_{ij}]$ and $\lambda_i^*(y_i)$ is decreasing in y_i, (12) implies that $\exp(\kappa_{ij}/\eta_{ij} - 1 - \lambda_i^*(y_i)/\eta_{ij}) \geq L_{ij}$ and $\exp(\kappa_{ij}/\eta_{ij} - 1 - \lambda_i^*(y_i)/\eta_{ij}) \leq U_{ij}$. By the last two inequalities and (10), the optimal value of the decision variable w_{ij} in problem (5) is $w_{ij}^* = \min\{\max\{\exp(\kappa_{ij}/\eta_{ij} - 1 - \lambda_i^*(y_i)/\eta_{ij}), L_{ij}\}, U_{ij}\} = \exp(\kappa_{ij}/\eta_{ij} - 1 - \lambda_i^*(y_i)/\eta_{ij})$. Also, the last two inequalities imply that the max and min operators for product j can be dropped from the sum in (11) without disturbing the equality, showing that

$\lambda_i^*(y_i)$ is still the optimal value of the Lagrange multiplier for the first constraint in problem (5) when we drop the constraint $w_{ij} \in [L_{ij}, U_{ij}]$. In this case, we let $\hat{\boldsymbol{w}}_i$ be the optimal solution to problem (5) when we drop the constraint $w_{ij} \in [L_{ij}, U_{ij}]$, together with the corresponding Lagrange multiplier $\lambda_i^*(y_i)$ for the first constraint. When we drop the constraint $w_{ij} \in [L_{ij}, U_{ij}]$, setting $L_{ij} = -\infty$ and $U_{ij} = \infty$ in (10) implies that the optimal value of the decision variable w_{ij} is given by $\hat{w}_{ij} = \exp(\kappa_{ij}/\eta_{ij} - 1 - \lambda_i^*(y_i)/\eta_{ij})$. Thus, it follows that $w_{ij}^* = \hat{w}_{ij}$, as desired. $\square$

The second property that we have for problem (5) is that we can partition the extended real line $[-\infty, \infty]$ into a number of intervals $\{[\nu_i^k, \nu_i^{k+1}] : k = 1, \ldots, K_i\}$ such that if we solve problem (5) for any $y_i \in [\nu_i^k, \nu_i^{k+1}]$, then we can immediately fix the values of some of the decision variables at their upper or lower bounds and not impose the upper and lower bound constraints at all on the remaining decision variables. To see this property, we note that if we plot the $2n$ points in the set $\{\zeta_{ij} : j \in N\} \cup \{\xi_{ij} : j \in N\}$ on the extended real line $[-\infty, \infty]$, then they partition the extended real line into at most $2n + 1$ intervals. We denote these intervals by $\{[\nu_i^k, \nu_i^{k+1}] : k = 1, \ldots, K_i\}$ with $\nu_i^1 = -\infty$ and $\nu_i^{K_i+1} = \infty$. Since the intervals $\{[\nu_i^k, \nu_i^{k+1}] : k = 1, \ldots, K_i\}$ are obtained by partitioning the real line with the points $\{\zeta_{ij} : j \in N\} \cup \{\xi_{ij} : j \in N\}$, it follows that for any $k = 1, \ldots, K_i$ and $j \in N$, we must have $[\nu_i^k, \nu_i^{k+1}] \subset [\zeta_{ij}, \xi_{ij}]$, or $[\nu_i^k, \nu_i^{k+1}] \subset [-\infty, \zeta_{ij}]$, or $[\nu_i^k, \nu_i^{k+1}] \subset [\xi_{ij}, \infty]$. In this case, we define the sets of products $\mathcal{L}_i^k$, $\mathcal{U}_i^k$ and $\mathcal{F}_i^k$ as

$$\mathcal{L}_i^k = \{j \in N : [\nu_i^k, \nu_i^{k+1}] \subset [-\infty, \zeta_{ij}]\} \qquad \mathcal{U}_i^k = \{j \in N : [\nu_i^k, \nu_i^{k+1}] \subset [\xi_{ij}, \infty]\}$$
$$\mathcal{F}_i^k = \{j \in N : [\nu_i^k, \nu_i^{k+1}] \subset [\zeta_{ij}, \xi_{ij}]\}.$$

Consider problem (5) with a value of y_i satisfying $y_i \in [\nu_i^k, \nu_i^{k+1}]$ for some $k = 1, \ldots, K_i$. If product j is in the set $\mathcal{L}_i^k$, then we have $[\nu_i^k, \nu_i^{k+1}] \subset [-\infty, \zeta_{ij}]$. Since $y_i \in [\nu_i^k, \nu_i^{k+1}]$, we obtain $y_i \le \zeta_{ij}$, in which case, Lemma 3 implies that the optimal value of the decision variable w_{ij} in problem (5) is L_{ij}. By following the same reasoning, if product j is in the set $\mathcal{U}_i^k$, then the optimal value of the decision variable w_{ij} in problem (5) is U_{ij}. Finally, if product j is in the set $\mathcal{F}_i^k$, then we have $[\nu_i^k, \nu_i^{k+1}] \subset [\zeta_{ij}, \xi_{ij}]$, but since $y_i \in [\nu_i^k, \nu_i^{k+1}]$, we obtain $y_i \in [\zeta_{ij}, \xi_{ij}]$, in which case, by Lemma 3, we can drop the constraint $w_{ij} \in [L_{ij}, U_{ij}]$ in problem (5) without changing the optimal solution. Therefore, whenever we solve problem (5) with a value of $y_i \in [\nu_i^k, \nu_i^{k+1}]$, we can fix the values of the decision variables in the sets $\mathcal{L}_i^k$ and $\mathcal{U}_i^k$ respectively at their lower and upper bounds and not impose the upper and lower bound constraints on the decision variables in the set $\mathcal{F}_i^k$. The observations in this paragraph yield the next lemma.

Lemma 4 (Partition)**.** *There exist intervals $\{[\nu_i^k, \nu_i^{k+1}] : k = 1, \ldots, K_i\}$ partitioning $[\bar{L}_i, \bar{U}_i]$ such that for any $y_i \in [\nu_i^k, \nu_i^{k+1}]$, the optimal solution to problem* (5) *can be obtained by solving*

$$\max \left\{ \sum_{j \in N} w_{ij} \left(\kappa_{ij} - \eta_{ij} \log w_{ij} \right) : \sum_{j \in N} w_{ij} \le y_i, \; w_{ij} = L_{ij} \quad \forall\, j \in \mathcal{L}_i^k, \; w_{ij} = U_{ij} \quad \forall\, j \in \mathcal{U}_i^k \right\} \tag{13}$$

for some subsets of products $\mathcal{L}_i^k$, $\mathcal{U}_i^k \subset N$ that depend on the interval k containing y_i but not on the specific value of y_i. Furthermore, we have $K_i = O(n)$.

Proof. Constructing the intervals $\{[\nu_i^k, \nu_i^{k+1}] : k = 1, \ldots, K_i\}$ as defined in the discussion right before the lemma, the first part follows by this discussion, as long as we take the intersection of each one of these intervals with $[\bar{L}_i, \bar{U}_i]$. To see that the second part holds, since the points $\{\zeta_{ij} : j \in N\} \cup \{\xi_{ij} : j \in N\}$ partition the extended real line into at most $2n+1$ intervals and these intervals correspond to $\{[\nu_i^k, \nu_i^{k+1}] : k = 1, \ldots, K_i\}$, K_i is at most $2n + 1 = O(n)$. □

Lemma 4 becomes useful when constructing a collection of grid points that satisfies the assumption of Theorem 2. We focus on this task in the next section.

5.2 Properties of Grid Points

In this section, we turn our attention to constructing a collection of grid points $\{\tilde{y}_i^t : t = 1, \ldots, T_i\}$ that satisfies the assumption of Theorem 2. To that end, we choose a fixed value of $\rho > 0$ and consider the grid points that are obtained by

$$\tilde{Y}_i^{kq} = \sum_{j \in \mathcal{L}_i^k} L_{ij} + \sum_{j \in \mathcal{U}_i^k} U_{ij} + (1+\rho)^q \tag{14}$$

for $k = 1, \ldots, K_i$ and $q = \ldots, -1, 0, 1, \ldots$. In the expression above, $\mathcal{L}_i^k, \mathcal{U}_i^k$ and K_i are such that they satisfy Lemma 4. In problem (5), once we fix the decision variables in $\mathcal{L}_i^k$ at their lower bounds and the decision variables in $\mathcal{U}_i^k$ at their upper bounds, the sum of the remaining decision variables is at least $\sum_{j \in N \setminus (\mathcal{L}_i^k \cup \mathcal{U}_i^k)} L_{ij}$ and at most $\sum_{j \in N \setminus (\mathcal{L}_i^k \cup \mathcal{U}_i^k)} U_{ij}$. Therefore, we choose the possible values for q in (14) such that the smallest value of $(1+\rho)^q$ does not stay above $\sum_{j \in N \setminus (\mathcal{L}_i^k \cup \mathcal{U}_i^k)} L_{ij}$ and the largest value of $(1+\rho)^q$ does not stay below $\sum_{j \in N \setminus (\mathcal{L}_i^k \cup \mathcal{U}_i^k)} U_{ij}$. If $\mathcal{L}_i^k \cup \mathcal{U}_i^k = N$, then using a single value of $q = -\infty$ suffices. Otherwise, using $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ to denote the round down and round up functions, we can choose the smallest value of q as $q_i^L = \lfloor \log(\min_{j \in N} L_{ij}) / \log(1+\rho) \rfloor$ and the largest value of q as $q_i^U = \lceil \log(n \max_{j \in N} U_{ij}) / \log(1+\rho) \rceil$. In this case, letting $\sigma_i = \max_{j \in N} U_{ij} / \min_{j \in N} L_{ij}$, we have $q_i^U - q_i^L = O(\log(n\sigma_i) / \log(1+\rho))$.

To construct a collection of grid points that satisfies the assumption of Theorem 2, we augment the set of points $\{\tilde{Y}_i^{kq} : k = 1, \ldots, K_i,\ q = q_i^L, \ldots, q_i^U\}$ defined above with the set of points $\{\nu_i^k : k = 1, \ldots, K_i + 1\}$, where the last set of points are obtained from the set of intervals $\{[\nu_i^k, \nu_i^{k+1}] : k = 1, \ldots, K_i\}$ given in Lemma 4. We obtain our set of grid points $\{\tilde{y}_i^t : t = 1, \ldots, T_i\}$ by ordering the points in $\{\tilde{Y}_i^{kq} : k = 1, \ldots, K_i,\ q = q_i^L, \ldots, q_i^U\} \cup \{\nu_i^k : k = 1, \ldots, K_i + 1\}$ in increasing order and dropping the ones that are not included in the interval $[\bar{L}_i, \bar{U}_i]$. Also, we add the two points $\bar{L}_i$ and $\bar{U}_i$ into the collection of grid points to ensure that the smallest and the largest one of the grid points $\{\tilde{y}_i^t : t = 1, \ldots, T_i\}$ are respectively equal to $\bar{L}_i$ and $\bar{U}_i$. Thus, the collection of grid points constructed in this fashion satisfies $\tilde{y}_i^t \leq \tilde{y}_i^{t+1}$ for all $t = 1, \ldots, T_i - 1$, $\tilde{y}_i^1 = \bar{L}_i$ and $\tilde{y}_i^{T_i} = \bar{U}_i$. Since $|K_i| = O(n)$ and $q_i^U - q_i^L = O(\log(n\,\sigma_i)/\log(1+\rho))$, the number of grid points in the collection is $|T_i| = O(n + n \log(n\,\sigma_i)/\log(1+\rho))$.

There are two useful properties for the grid points $\{\tilde{y}_i^t : t = 1, \ldots, T_i\}$ constructed by using the approach above. The first property is that if $\tilde{y}_i^t$ and $\tilde{y}_i^{t+1}$ are two consecutive grid points, then they satisfy $\tilde{y}_i^t, \tilde{y}_i^{t+1} \in [\nu_i^k, \nu_i^{k+1}]$ for some $k = 1, \ldots, K_i$. To see this property, if this property does not hold, then we have $\tilde{y}_i^t \le \nu_i^k \le \tilde{y}_i^{t+1}$ for some $k = 1, \ldots, K_i$ with one of the two inequalities holding as a strict inequality. If this chain of inequalities holds, then since ν_i^k is a grid point itself, we get a contradiction to the fact that $\tilde{y}_i^t$ and $\tilde{y}_i^{t+1}$ are two consecutive grid points, establishing the first property. The second property is that if $\tilde{y}_i^t$ and $\tilde{y}_i^{t+1}$ are two consecutive grid points satisfying $\tilde{y}_i^t, \tilde{y}_i^{t+1} \in [\nu_i^k, \nu_i^{k+1}]$ for some $k = 1, \ldots, K_i$, then we have $\tilde{Y}_i^{kq} \le \tilde{y}_i^t \le \tilde{y}_i^{t+1} \le \tilde{Y}_i^{k,q+1}$ for some $q = q_i^L, \ldots, q_i^U - 1$. The idea behind the second property is similar to the one used for the first property. In particular, if the second property does not hold, then either we have $\tilde{y}_i^t \le \tilde{Y}_i^{kq} \le \tilde{y}_i^{t+1}$ for some $q = q_i^L, \ldots, q_i^U - 1$ or we have $\tilde{y}_i^t \le \tilde{Y}_i^{k,q+1} \le \tilde{y}_i^{t+1}$ for some $q = q_i^L, \ldots, q_i^U - 1$ with one of the last four inequalities holding as a strict inequality. If either one of the last two chains of inequalities holds, then since $\tilde{Y}_i^{kq}$ and $\tilde{Y}_i^{k,q+1}$ are grid points themselves, we get a contradiction to the fact that $\tilde{y}_i^t$ and $\tilde{y}_i^{t+1}$ are two consecutive grid points, establishing the second property. In the next theorem, we use these properties along with Lemma 4 to show that the collection of grid points $\{\tilde{y}_i^t : t = 1, \ldots, T_i\}$ satisfies the assumption of Theorem 2.

Theorem 5 (Grid Construction)**.** *Assume that the collection of grid points* $\{\tilde{y}_i^t : t = 1, \ldots, T_i\}$ *are obtained by ordering the points in* $\{\tilde{Y}_i^{kq} : k = 1, \ldots, K_i,\ q = q_i^L, \ldots, q_i^U\} \cup \{\nu_i^k : k = 1, \ldots, K_i + 1\}$ *in increasing order. In this case, we have* $g_i(\tilde{y}_i^{t+1}) \le (1+\rho)\, g_i(\tilde{y}_i^t)$ *for all* $t = 1, \ldots, T_i - 1$.

Proof. If $\tilde{y}_i^t$ and $\tilde{y}_i^{t+1}$ are two consecutive grid points, then the first property right before the statement of the theorem implies that there exists $k = 1, \ldots, K_i$ such that $\tilde{y}_i^t, \tilde{y}_i^{t+1} \in [\nu_i^k, \nu_i^{k+1}]$, in which case, by the second property, it follows that there exists $q = q_i^L, \ldots, q_i^U - 1$ such that $\tilde{Y}_i^{kq} \le \tilde{y}_i^t \le \tilde{y}_i^{t+1} \le \tilde{Y}_i^{k,q+1}$. Subtracting $\sum_{j \in \mathcal{L}_i^k} L_{ij} + \sum_{j \in \mathcal{U}_i^k} U_{ij}$ from the last chain of inequalities and noting the definition of $\tilde{Y}_i^{kq}$ in (14), we obtain

$$(1+\rho)^q \le \tilde{y}_i^t - \sum_{j \in \mathcal{L}_i^k} L_{ij} - \sum_{j \in \mathcal{U}_i^k} U_{ij} \le \tilde{y}_i^{t+1} - \sum_{j \in \mathcal{L}_i^k} L_{ij} - \sum_{j \in \mathcal{U}_i^k} U_{ij} \le (1+\rho)^{q+1}.$$

Using the chain of inequalities above, it follows that $\tilde{y}_i^{t+1} - \sum_{j \in \mathcal{L}_i^k} L_{ij} - \sum_{j \in \mathcal{U}_i^k} U_{ij} \le (1+\rho)^{q+1} \le (1+\rho)\,(\tilde{y}_i^t - \sum_{j \in \mathcal{L}_i^k} L_{ij} - \sum_{j \in \mathcal{U}_i^k} U_{ij})$. Since $\tilde{y}_i^t, \tilde{y}_i^{t+1} \in [\nu_i^k, \nu_i^{k+1}]$, Lemma 4 implies that the optimal solution to problem (5) with $y_i = \tilde{y}_i^t$ or $y_i = \tilde{y}_i^{t+1}$ can be obtained by solving problem (13) respectively with $y_i = \tilde{y}_i^t$ or $y_i = \tilde{y}_i^{t+1}$. We let $\boldsymbol{w}_i^*$ be the optimal solution to problem (13) when we solve this problem with $y_i = \tilde{y}_i^{t+1}$. Note that $w_{ij}^* = L_{ij}$ for all $j \in \mathcal{L}_i^k$ and $w_{ij}^* = U_{ij}$ for all $j \in \mathcal{U}_i^k$. We define the solution $\hat{\boldsymbol{w}}_i$ as $\hat{w}_{ij} = w_{ij}^*/(1+\rho)$ for all $j \in N \setminus (\mathcal{L}_i^k \cup \mathcal{U}_i^k)$, $\hat{w}_{ij} = L_{ij}$ for all $j \in \mathcal{L}_i^k$ and $\hat{w}_{ij} = U_{ij}$ for all $j \in \mathcal{U}_i^k$. In this case, we have

$$\sum_{j \in N \setminus (\mathcal{L}_i^k \cup \mathcal{U}_i^k)} \hat{w}_{ij} = \frac{1}{1+\rho} \sum_{j \in N \setminus (\mathcal{L}_i^k \cup \mathcal{U}_i^k)} w_{ij}^* \le \frac{1}{1+\rho} \left\{ \tilde{y}_i^{t+1} - \sum_{j \in \mathcal{L}_i^k} L_{ij} - \sum_{j \in \mathcal{U}_i^k} U_{ij} \right\} \le \tilde{y}_i^t - \sum_{j \in \mathcal{L}_i^k} L_{ij} - \sum_{j \in \mathcal{U}_i^k} U_{ij},$$

where the first inequality is by the fact that $\boldsymbol{w}_i^*$ is a feasible solution to problem (13) when we solve this problem with $y_i = \tilde{y}_i^{t+1}$ and the second inequality follows from the fact that $\tilde{y}_i^{t+1} - \sum_{j \in \mathcal{L}_i^k} L_{ij} - \sum_{j \in \mathcal{U}_i^k} U_{ij} \le (1+\rho)\,(\tilde{y}_i^t - \sum_{j \in \mathcal{L}_i^k} L_{ij} - \sum_{j \in \mathcal{U}_i^k} U_{ij})$, which is shown at the beginning of the proof. Therefore, the chain of equalities above shows that $\hat{\boldsymbol{w}}_i$ is a feasible solution to problem (13) when we solve this problem with $y_i = \tilde{y}_i^t$, in which case, we obtain

$$\begin{aligned} g_i(\tilde{y}_i^t) \geq \sum_{j\in N} \hat{w}_{ij}\,(\kappa_{ij} - \eta_{ij}\log \hat{w}_{ij}) = &\sum_{j\in N\setminus(\mathcal{L}_i^k\cup\mathcal{U}_i^k)} \frac{w_{ij}^*}{1+\rho}\,(\kappa_{ij} - \eta_{ij}\log w_{ij}^* + \eta_{ij}\log(1+\rho)) \\ &+ \sum_{j\in\mathcal{L}_i^k} (\kappa_{ij} - \eta_{ij}\log L_{ij})\,L_{ij} + \sum_{j\in\mathcal{U}_i^k} (\kappa_{ij} - \eta_{ij}\log U_{ij})\,U_{ij} \\ \geq \frac{1}{1+\rho}\Bigg\{ \sum_{j\in N\setminus(\mathcal{L}_i^k\cup\mathcal{U}_i^k)} & w_{ij}^*\,(\kappa_{ij} - \eta_{ij}\log w_{ij}^*) + \sum_{j\in\mathcal{L}_i^k} (\kappa_{ij} - \eta_{ij}\log L_{ij})\,L_{ij} + \sum_{j\in\mathcal{U}_i^k} (\kappa_{ij} - \eta_{ij}\log U_{ij})\,U_{ij} \Bigg\} \\ = \frac{1}{1+\rho}\, g_i(\tilde{y}_i^{t+1}), & \end{aligned}$$

where the first inequality uses the fact that $\hat{\boldsymbol{w}}_i$ is a feasible solution to problem (13) when solved with $y_i = \tilde{y}_i^t$ and this problem yields the optimal solution to problem (5) with $y_i = \tilde{y}_i^t$ and the second equality follows from the fact that $\boldsymbol{w}_i^*$ is the optimal solution to problem (13) when solved with $y_i = \tilde{y}_i^{t+1}$. The chain of inequalities above establishes the desired result. □

Therefore, the theorem above shows that the grid that we construct by using the set of points $\{\tilde{Y}_i^{kq} : k = 1, \ldots, K_i,\ q = q_i^L, \ldots, q_i^U\} \cup \{\nu_i^k : k = 1, \ldots, K_i + 1\}$ satisfies the assumption of Theorem 2. It is also useful to note that all of the discussion in this section continues to hold when we use the no purchase preference weight w_{i0} to allow a customer to leave nest i without purchasing anything. To accommodate this extension, all we need to do is to add w_{i0} to the left side of (11) when defining $\lambda_i^*(y_i)$ and add w_{i0} to the right side of (14) when defining $\tilde{Y}_i^{kq}$.

5.3 Approximation Method

In this section, we put together all of our results so far to give the following algorithm that finds a $(1+\rho)$-approximate solution to problem (2).

Step 1. For all $i \in M$, $j \in N$, we compute ζ_{ij} and ξ_{ij} such that $\exp(\kappa_{ij}/\eta_{ij} - 1 - \lambda_i^*(\zeta_{ij})/\eta_{ij}) = L_{ij}$ and $\exp(\kappa_{ij}/\eta_{ij} - 1 - \lambda_i^*(\xi_{ij})/\eta_{ij}) = U_{ij}$, where $\lambda_i^*(y_i)$ is as defined in (11). For each $i \in M$, the collection of points $\{\zeta_{ij} : j \in N\} \cup \{\xi_{ij} : j \in N\}$ partition the extended real line into $O(n)$ intervals. We denote these intervals by $\{[\nu_i^k, \nu_i^{k+1}] : k = 1, \ldots, K_i\}$.

Step 2. For all $i \in M$, $k = 1, \ldots, K_i$, we compute the sets $\mathcal{L}_i^k = \{j \in N : [\nu_i^k, \nu_i^{k+1}] \subset [-\infty, \zeta_{ij}]\}$ and $\mathcal{U}_i^k = \{j \in N : [\nu_i^k, \nu_i^{k+1}] \subset [\xi_{ij}, \infty]\}$. We choose a fixed value of $\rho > 0$ and compute the points $\tilde{Y}_i^{kq} = \sum_{j\in\mathcal{L}_i^k} L_{ij} + \sum_{j\in\mathcal{U}_i^k} U_{ij} + (1+\rho)^q$ for all $i \in M$, $k = 1, \ldots, K_i$, $q = q_i^L, \ldots, q_i^U$, where $q_i^L = \lfloor \log(\min_{j\in N} L_{ij}) / \log(1+\rho) \rfloor$ and $q_i^U = \lceil \log(n \max_{j\in N} U_{ij}) / \log(1+\rho) \rceil$.

Step 3. For each $i \in M$, we order the points in the set $\{\tilde{Y}_i^{kq} : k = 1, \ldots, K_i,\ q = q_i^L, \ldots, q_i^U\} \cup \{\nu_i^k : k = 1, \ldots, K_i + 1\}$ in increasing order to obtain a collection of grid points. We drop the points that are outside the interval $[\bar{L}_i, \bar{U}_i]$ and add the points $\bar{L}_i$ and $\bar{U}_i$ so that the smallest and the largest one of the grid points are respectively equal to $\bar{L}_i$ and $\bar{U}_i$. We use $\{\tilde{y}_i^t : t = 1, \ldots, T_i\}$ to denote the collection of grid points obtained in this fashion.

STEP 4. By using the grid points $\{\tilde{y}_i^t : t = 1, \ldots, T_i\}$ for all $i \in M$, we solve the linear program in (9) and use $\hat{z}$ to denote its optimal objective value. For all $i \in M$, we solve the maximization problem on the right side of (7) with $z = \hat{z}$ and use $\hat{y}_i$ to denote its optimal solution.

STEP 5. For all $i \in M$, we solve the knapsack problem in (5) with $y_i = \hat{y}_i$ and use $\hat{\boldsymbol{w}}_i$ to denote its optimal solution. We return $(\hat{\boldsymbol{w}}_1, \ldots, \hat{\boldsymbol{w}}_m)$ as a $(1 + \rho)$-approximate solution to problem (2).

In Steps 1, 2 and 3, we compute the collection of grid points $\{\tilde{y}_i^t : t = 1, \ldots, T_i\}$. Noting Theorem 5, it follows that this collection of grid points satisfies the assumption of Theorem 2. In Step 4, the value of $\hat{z}$ that we compute by solving the linear program in (9) corresponds to the value of z satisfying (7). In Step 5, we compute the solution $(\hat{\boldsymbol{w}}_1, \ldots, \hat{\boldsymbol{w}}_m)$ to problem (2). By the discussion that follows the proof of Theorem 2, the expected revenue provided by the solution $(\hat{\boldsymbol{w}}_1, \ldots, \hat{\boldsymbol{w}}_m)$ deviates from the optimal expected revenue by at most a factor of $1 + \rho$, satisfying $(1 + \rho)\, \Pi(\hat{\boldsymbol{w}}_1, \ldots, \hat{\boldsymbol{w}}_m) \geq Z^*$. Letting $\sigma = \max\{\max_{j \in N} U_{ij} / \min_{j \in N} L_{ij} : i \in M\}$, we observe that there are $O(n + n \log(n\sigma)/\log(1 + \rho))$ points in the collection of grid points $\{\tilde{y}_i^t : t = 1, \ldots, T_i\}$. Therefore, noting the linear program in (9), the main computational effort in obtaining a $(1 + \rho)$-approximate solution to problem (2) involves solving a linear program with $1 + m$ decision variables and $O(mn + mn \log(n\sigma)/\log(1 + \rho))$ constraints.

6 Joint Assortment Offering and Pricing

Our development so far assumes that the choice of the products offered to customers are beyond our control and all n products have to be offered in all m nests. In this section, we consider a model that jointly decides which set of products to offer in each nest, along with the prices of the offered products. Similar to our problem formulation in Section 2, we assume that the price of each product has to satisfy the constraint $p_{ij} \in [l_{ij}, u_{ij}]$ with $l_{ij}, u_{ij} \in [0, \infty)$. If we set the price of product j in nest i as p_{ij}, then its preference weight is given by $w_{ij} = \exp(\alpha_{ij} - \beta_{ij}\, p_{ij})$. We continue viewing the preference weights as decision variables, so that the preference weight w_{ij} of product j in nest i has to satisfy the constraint $w_{ij} \in [L_{ij}, U_{ij}]$ with $L_{ij} = \exp(\alpha_{ij} - \beta_{ij}\, u_{ij})$ and $U_{ij} = \exp(\alpha_{ij} - \beta_{ij}\, l_{ij})$. In this case, using $S_i \subset N$ to denote the set of products that we offer in nest i, if we offer the assortments, or subsets of products, $(S_1, \ldots, S_m)$ over all nests and choose the preference weights over all nests as $(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m)$, then we obtain an expected revenue of

$$\Theta(S_1, \ldots, S_m, \boldsymbol{w}_1, \ldots, \boldsymbol{w}_m) = \frac{1}{1 + \sum_{i \in M} \left(\sum_{j \in S_i} w_{ij}\right)^{\gamma_i}} \sum_{i \in M} \Big(\sum_{j \in S_i} w_{ij}\Big)^{\gamma_i} \frac{\sum_{j \in S_i} w_{ij}\, (\kappa_{ij} - \eta_{ij} \log w_{ij})}{\sum_{j \in S_i} w_{ij}}. \quad (15)$$

The definition of the expected revenue function $\Theta(S_1, \ldots, S_m, \boldsymbol{w}_1, \ldots, \boldsymbol{w}_m)$ is similar to the definition of $\Pi(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m)$ in (1) and it can be derived by using an argument similar to the one in Section 2, but the expected revenue function above only uses the preference weights of the products in the offered assortment. The second fraction above evaluates to 0/0 when $S_i = \emptyset$ and we treat 0/0 as zero throughout this section. Our goal is to solve the problem

$$\zeta^* = \max\big\{\Theta(S_1,\ldots,S_m,\boldsymbol{w}_1,\ldots,\boldsymbol{w}_m) : S_i \subset N \;\; \forall\, i \in M,\; \boldsymbol{w}_i \in [\boldsymbol{L}_i, \boldsymbol{U}_i] \;\; \forall\, i \in M\big\}. \tag{16}$$

The idea that we use to solve the joint assortment offering and pricing problem above is similar to the one used for solving our earlier pricing problem. We view the problem above as computing the fixed point of an appropriately defined scalar function and this visualization allows us to relate our problem to a knapsack problem. However, one crucial difference is that we need to characterize the structure of the subsets of products to be offered in the optimal solution to problem (16).

Similar to the expected revenue function $\Pi(\boldsymbol{w}_1,\ldots,\boldsymbol{w}_m)$ in (1), the expected revenue function $\Theta(S_1,\ldots,S_m,\boldsymbol{w}_1,\ldots,\boldsymbol{w}_m)$ above assumes that if a customer decides to make a purchase in a particular nest, then this customer must purchase one of the products offered in this nest. To extend our model to allow a customer to leave a nest without purchasing anything, one possibility is to use w_{i0} to denote the preference weight of the no purchase option in nest i, which is a constant that does not depend on the prices of any of the products. In this case, if we offer the assortment of products S_i in nest i and a customer chooses this nest, then this customer leaves nest i without purchasing anything with probability $w_{i0}/(w_{i0} + \sum_{j \in S_i} w_{ij})$. If we use this approach to model the possibility that a customer can leave a nest without making a purchase, then Davis et al. (2011) show that the problem of finding the optimal assortment of products to offer in each nest is NP-hard even when the prices of the products are fixed. To model the no purchase option in a nest in a more tractable fashion, letting $\mathbf{1}(\cdot)$ be the indicator function, we propose associating the preference weight $\mathbf{1}(S_i \neq \emptyset)\, w_{i0}$ with the no purchase option in nest i. This way of modeling the no purchase option is likely to be more appealing from a practical perspective since if we offer the empty assortment in a nest, then customers are not attracted to this nest at all. All of our results continue to hold when we use $\mathbf{1}(S_i \neq \emptyset)\, w_{i0}$ to capture the preference weight of the no purchase option in nest i. We elaborate on this extension at appropriate places in the paper.

6.1 Fixed Point Representation

We begin by giving a fixed point representation of problem (16). Our discussion closely follows the one for our earlier pricing problem. So, while we present our discussion in full, we omit the proofs whenever they resemble the earlier ones. Assume that we compute the value of z satisfying

$$z = \sum_{i \in M} \max_{S_i \subset N, \boldsymbol{w}_i \in [\boldsymbol{L}_i, \boldsymbol{U}_i]} \left\{ \Big(\sum_{j \in S_i} w_{ij} \Big)^{\gamma_i} \frac{\sum_{j \in S_i} w_{ij}\,(\kappa_{ij} - \eta_{ij} \log w_{ij})}{\sum_{j \in S_i} w_{ij}} - \Big(\sum_{j \in S_i} w_{ij} \Big)^{\gamma_i} z \right\}. \tag{17}$$

Following the same argument at the beginning of Section 3, one can check that if the value of z satisfying (17) is given by $\hat{z}$, then we have $\hat{z} = \zeta^*$, where ζ^* is the optimal objective value of problem (16). Furthermore, if the value of z satisfying (17) is $\hat{z}$ and we use $(\hat{S}_i, \hat{\boldsymbol{w}}_i)$ to denote an optimal solution to the maximization problem on the right side of (17) when we solve this problem with $z = \hat{z}$, then it follows that $(\hat{S}_1, \ldots, \hat{S}_m, \hat{\boldsymbol{w}}_1, \ldots, \hat{\boldsymbol{w}}_m)$ is an optimal solution to problem (16). One crucial difficulty associated with solving the maximization problem on the right side of (17) is that the decision variable S_i in this problem can take 2^n possible values, which can be too many to enumerate when we have a reasonably large number of products. However, it turns out that we can limit the number of possible values for S_i in the optimal solution to only $1 + n$. In this case, we can enumerate all possible $1 + n$ values for the decision variable S_i.

To limit the possible values for the decision variable S_i, we assume that the products in each nest are indexed in the order of decreasing price upper bounds so that $u_{i1} \geq u_{i2} \geq \ldots \geq u_{in}$. We use N_{ij} to denote the subset of products that includes the first j products with the largest price upper bounds in nest i. That is, $N_{ij} = \{1, \ldots, j\}$. We refer to such a subset as a nested by price bound assortment. Using the convention that $N_{i0} = \emptyset$, we let $\mathcal{N}_i = \{N_{ij} : j \in N \cup \{0\}\}$ to capture all nested by price bound assortments in nest i. In the next theorem, we show that a nested by price bound assortment solves the maximization problem on the right side of (17).

Theorem 6 (Optimal Assortments)**.** *For any $z > 0$, there exists an assortment $S_i^* \in \mathcal{N}_i$ that solves the maximization problem on the right side of* (17)*.*

Proof. For notational brevity, we let $R_i(S_i, \boldsymbol{w}_i) = \sum_{j \in S_i} w_{ij}\,(\kappa_{ij} - \eta_{ij} \log w_{ij}) / \sum_{j \in S_i} w_{ij}$ and $W_i(S_i, \boldsymbol{w}_i) = \sum_{j \in S_i} w_{ij}$, in which case, we can write the objective function of the maximization problem on the right side of (17) as $W_i(S_i, \boldsymbol{w}_i)^{\gamma_i} (R_i(S_i, \boldsymbol{w}_i) - z)$. To get a contradiction, we let $(S_i^*, \boldsymbol{w}_i^*)$ be an optimal solution to the maximization problem on the right side of (17) and assume that there exist products k and l such that $k < l$, $k \notin S_i^*$ and $l \in S_i^*$. We show that if we add product k into the assortment S_i^* with price u_{ik}, then we obtain a better solution for the maximization problem on the right side of (17), establishing the desired result. In particular, consider the solution $(\hat{S}_i, \hat{\boldsymbol{w}}_i)$ obtained by setting $\hat{S}_i = S_i^* \cup \{k\}$, $\hat{w}_{ik} = \exp(\alpha_{ik} - \beta_{ik}\, u_{ik})$ and $\hat{w}_{ij} = w_{ij}^*$ for all $j \in N \setminus \{k\}$, which is equivalent to setting the price of product k as u_{ik} and not changing the prices of the other products in the solution $\boldsymbol{w}_i^*$. In this case, we have

$$W_i(\hat{S}_i, \hat{\boldsymbol{w}}_i)^{\gamma_i} (R_i(\hat{S}_i, \hat{\boldsymbol{w}}_i) - z) = \frac{\sum_{j \in \hat{S}_i} \hat{w}_{ij}\,(\kappa_{ij} - \eta_{ij} \log \hat{w}_{ij} - z)}{W_i(\hat{S}_i, \hat{\boldsymbol{w}}_i)^{1-\gamma_i}}$$
$$= \frac{\sum_{j \in S_i^*} w_{ij}^*\,(\kappa_{ij} - \eta_{ij} \log w_{ij}^* - z) + \hat{w}_{ik}\,(\kappa_{ik} - \eta_{ik} \log \hat{w}_{ik} - z)}{W_i(\hat{S}_i, \hat{\boldsymbol{w}}_i)^{1-\gamma_i}}, \quad (18)$$

where the first equality follows by using the definitions of $R_i(S_i, \boldsymbol{w}_i)$ and $W_i(S_i, \boldsymbol{w}_i)$ and rearranging the terms and the second equality uses the fact that $\hat{S}_i = S_i^* \cup \{k\}$ and the products in S_i^* have the same preference weights in solutions $\boldsymbol{w}_i^*$ and $\hat{\boldsymbol{w}}_i$.

We proceed to lower bounding the last fraction above. It is possible to show that if product l is included in the optimal solution to the maximization problem on the right side of (17), then the preference weight of product l must satisfy $\kappa_{il} - \eta_{il} \log w_{il}^* \geq (1 - \gamma_i)\, R_i(S_i^*, \boldsymbol{w}_i^*) + \gamma_i\, z$. We defer the proof of this fact to Lemma 8 in Appendix A. Since $(S_i^*, \boldsymbol{w}_i^*)$ is feasible to the maximization problem on the right side of (17), we have $w_{il}^* \geq L_{il} = \exp(\alpha_{il} - \beta_{il}\, u_{il})$. Taking logarithms in this inequality and noting $\kappa_{ij} = \alpha_{ij}/\beta_{ij}$ and $\eta_{ij} = 1/\beta_{ij}$, we get $\kappa_{il} - \eta_{il} \log w_{il}^* \leq u_{il}$. In this case, noting $k < l$ so that $u_{ik} \geq u_{il}$, we obtain $\kappa_{ik} - \eta_{ik} \log \hat{w}_{ik} = \kappa_{ik} - \eta_{ik} \log(\exp(\alpha_{ik} - \beta_{ik}\, u_{ik})) = u_{ik} \geq u_{il} \geq \kappa_{il} - \eta_{il} \log w_{il}^* \geq (1-\gamma_i)\, R_i(S_i^*, \boldsymbol{w}_i^*) + \gamma_i\, z$. To lower bound to numerator of the last fraction in (18), we use the last chain of inequalities to get $\kappa_{ik} - \eta_{ik} \log \hat{w}_{ik} - z \geq (1-\gamma_i)(R_i(S_i^*, \boldsymbol{w}_i^*) - z)$. Thus, we can lower bound the numerator of the right side of (18) by the expression

$$\begin{aligned}\sum_{j \in S_i^*} w_{ij}^* \, (\kappa_{ij} - \eta_{ij} \log w_{ij}^* - z) + \hat{w}_{ik}\, (1-\gamma_i)(R_i(S_i^*, \boldsymbol{w}_i^*) - z)& \\ = (R_i(S_i^*, \boldsymbol{w}_i^*) - z)\, (W_i(S_i^*, \boldsymbol{w}_i^*) + \hat{w}_{ik}\, (1-\gamma_i)),&\end{aligned}$$

where the equality follows by using the definitions of $R_i(S_i, \boldsymbol{w}_i)$ and $W_i(S_i, \boldsymbol{w}_i)$. To upper bound the denominator of the last fraction in (18), we note that $u^{1-\gamma_i}$ is a concave function of u, satisfying the subgradient inequality $\hat{u}^{1-\gamma_i} \leq (u^*)^{1-\gamma_i} + (1-\gamma_i)\,(u^*)^{-\gamma_i}\,(\hat{u} - u^*)$ for two points $\hat{u}$ and u^*. Thus, we get $W_i(\hat{S}_i, \hat{\boldsymbol{w}}_i)^{1-\gamma_i} \leq W_i(S_i^*, \boldsymbol{w}_i^*)^{1-\gamma_i} + (1-\gamma_i)\, W_i(S_i^*, \boldsymbol{w}_i^*)^{-\gamma_i}\, (W_i(\hat{S}_i, \hat{\boldsymbol{w}}_i) - W_i(S_i^*, \boldsymbol{w}_i^*))$. Using these lower and upper bounds in (18), it follows that

$$\begin{aligned}&\frac{\sum_{j \in S_i^*} w_{ij}^* \, (\kappa_{ij} - \eta_{ij} \log w_{ij}^* - z) + \hat{w}_{ik}\, (\kappa_{ik} - \eta_{ik} \log \hat{w}_{ik} - z)}{W_i(\hat{S}_i, \hat{\boldsymbol{w}}_i)^{1-\gamma_i}} \\ &\qquad \geq \frac{(R_i(S_i^*, \boldsymbol{w}_i^*) - z)\, (W_i(S_i^*, \boldsymbol{w}_i^*) + \hat{w}_{ik}\, (1-\gamma_i))}{W_i(S_i^*, \boldsymbol{w}_i^*)^{1-\gamma_i} + (1-\gamma_i)\, W_i(S_i^*, \boldsymbol{w}_i^*)^{-\gamma_i}\, (W_i(\hat{S}_i, \hat{\boldsymbol{w}}_i) - W_i(S_i^*, \boldsymbol{w}_i^*))}. \qquad (19)\end{aligned}$$

Noting that $W_i(\hat{S}_i, \hat{\boldsymbol{w}}_i) - W_i(S_i^*, \boldsymbol{w}_i^*) = \hat{w}_{ik}$ and factoring out $W_i(S_i^*, \boldsymbol{w}_i^*)^{-\gamma_i}$ in the denominator of the last fraction in (19), the last fraction above is equal to $W_i(S_i^*, \boldsymbol{w}_i^*)^{\gamma_i}(R_i(S_i^*, \boldsymbol{w}_i^*) - z)$. Thus, (18) and (19) show that $W_i(\hat{S}_i, \boldsymbol{w}_i^*)^{\gamma_i}(R_i(\hat{S}_i, \hat{\boldsymbol{w}}_i) - z) \geq W_i(S_i^*, \boldsymbol{w}_i^*)^{\gamma_i}(R_i(S_i^*, \boldsymbol{w}_i^*) - z)$, establishing that the solution $(\hat{S}_i, \hat{\boldsymbol{w}}_i)$ provides an objective value for the maximization problem on the right side of (17) that is at least as large as the one provided by the solution $(S_i^*, \boldsymbol{w}_i^*)$. □

The theorem above shows that we can replace the constraint $S_i \subset N$ on the right side of (17) with $S_i \in \mathcal{N}_i$. Noting that $|\mathcal{N}_i| = O(n)$, we can deal with the decision variable S_i in the maximization problem on the right side of (17) simply by enumerating all of its possible values in a brute force fashion. To deal with the high dimensionality of the decision variable $\boldsymbol{w}_i$, we define $g_i(S_i, y_i)$ as the optimal objective value of the knapsack problem

$$g_i(S_i, y_i) = \max \left\{ \sum_{j \in S_i} w_{ij}\, (\kappa_{ij} - \eta_{ij} \log w_{ij}) : \sum_{j \in S_i} w_{ij} \leq y_i, \;\; w_{ij} \in [L_{ij}, U_{ij}] \;\;\; \forall j \in N \right\}, \qquad (20)$$

which is the analogue of problem (5), but we focus only on the products in S_i. Similar to the discussion that follows problem (5), letting $\bar{U}_i(S_i) = \sum_{j \in S_i} \min\{\max\{\exp(\kappa_{ij}/\eta_{ij} - 1), L_{ij}\}, U_{ij}\}$, if we have $y_i > \bar{U}_i(S_i)$, then the first constraint in the problem above is not tight at the optimal solution. On the other hand, letting $\bar{L}_i(S_i) = \sum_{j \in S_i} L_{ij}$, if we have $y_i < \bar{L}_i(S_i)$, then the problem above is infeasible. Finally, if $y_i \in [\bar{L}_i(S_i), \bar{U}_i(S_i)]$, then the first constraint above is tight at the optimal solution. Therefore, the solution to the problem above can potentially change only as y_i takes values in the interval $[\bar{L}_i(S_i), \bar{U}_i(S_i)]$. So, instead of finding the value of z satisfying (17), we propose finding the value of z satisfying

$$z = \sum_{i \in M} \max_{S_i \in \mathcal{N}_i,\, y_i \in [\bar{L}_i(S_i), \bar{U}_i(S_i)]} \left\{ y_i^{\gamma_i} \frac{g_i(S_i, y_i)}{y_i} - y_i^{\gamma_i} z \right\}. \tag{21}$$

By following the outline of the proof of Proposition 1, it is possible to show that the values of z that satisfy (17) and (21) are identical to each other and this common value corresponds to the optimal objective value of problem (16).

The decision variable S_i on the right side of (21) does not create a complication since $|\mathcal{N}_i| = O(n)$ and we can simply check each possible value of this decision variable one by one. However, the decision variable y_i on the right side of (21) can be problematic since the objective function of the maximization problem is not necessarily a unimodal function of y_i for a fixed S_i. As described in the next section, we deal with this complication by constructing a grid.

We can make several extensions to the results in this section. Although our focus is particularly on pricing problems with price bounds, we can extend the structural property in Theorem 6 to cover the case where we do not have price bounds on some of the products. In particular, we can show that if a product does not have a price upper bound, then it is always optimal to offer this product in the maximization problem on the right side of (17). One way to see this result is to assume that product k in nest i does not have a price upper bound and this product is not offered in the optimal solution $(S_i^*, \boldsymbol{w}_i^*)$, in which case, it is possible to check that we get a contradiction. To see the contradiction, we note that the optimal solution $(S_i^*, \boldsymbol{w}_i^*)$ satisfies $\sum_{j \in S_i^*} w_{ij}^* \, (\kappa_{ij} - \eta_{ij} \log w_{ij}^*) / \sum_{j \in S_i^*} w_{ij}^* - z \geq 0$. Otherwise, the optimal objective value of the maximization problem on the right side of (17) is negative, but we can set $S_i = \emptyset$ and obtain a better objective value of zero. Also, the expression $\sum_{j \in S_i^*} w_{ij}^* \, (\kappa_{ij} - \eta_{ij} \log w_{ij}^*) / \sum_{j \in S_i^*} w_{ij}^*$ is a weighted average of the prices of the products in the assortment S_i^*. Thus, we can make this expression larger if we add product k in nest i to the assortment S_i^* at a price that is larger than the prices of all of the products in S_i^*. In other words, we can choose the price of product k in nest i as p_{ik}^* such that the corresponding preference weight w_{ik}^* satisfies $\kappa_{ik} - \eta_{ik} \log w_{ik}^* \geq \kappa_{ij} - \eta_{ij} \log w_{ij}^*$ for all $j \in S_i^*$, in which case, we obtain $\sum_{j \in S_i^* \cup \{k\}} w_{ij}^* \, (\kappa_{ij} - \eta_{ij} \log w_{ij}^*) / \sum_{j \in S_i^* \cup \{k\}} w_{ij}^* - z \geq \sum_{j \in S_i^*} w_{ij}^* \, (\kappa_{ij} - \eta_{ij} \log w_{ij}^*) / \sum_{j \in S_i^*} w_{ij}^* - z \geq 0$. Since we also have $\sum_{j \in S_i^* \cup \{k\}} w_{ij}^* \geq \sum_{j \in S_i^*} w_{ij}$, the last two inequalities indicate that adding product k in nest i to the assortment S_i^* improves the optimal objective value of the maximization problem on the right side of (17), which is a contradiction. Thus, it is optimal to offer all of the products without price upper bounds. For the remaining products with price upper bounds, we can follow the proof of Theorem 6 to show that it is enough to consider nested by price bound assortments for these products.

In certain application settings, it may be desirable to limit the number of products offered in each nest. If we have a constraint on the number of products offered in each nest, then it is not necessarily true that a nested by price bound assortment is optimal for the maximization problem on the right side of (17). However, by using a significantly more involved argument, we can show that if we have a constraint on the number of products offered in each nest, then the optimal value of the decision variable S_i is one of $O(n^2)$ possible values and we can identify each one of these $O(n^2)$ possible values in a tractable fashion. Therefore, we can check all possible $O(n^2)$ values for the decision variable S_i when solving the maximization problem on the right side of (17). We defer the details of this argument to Appendix B.

It is possible to check that all of the discussion in this section holds when we use the no purchase preference weight $\mathbf{1}(S_i \neq \emptyset)\, w_{i0}$ to capture the possibility that a customer may leave nest i without making a purchase. To make this extension, all we need to do is to replace $\sum_{j \in S_i} w_{ij}$ in (17) and (20) with $\mathbf{1}(S_i \neq \emptyset)\, w_{i0} + \sum_{j \in S_i} w_{ij}$ and note that the optimal objective values of the maximization problem on the right side of (17) and the knapsack problem in (20) are zero when $S_i = \emptyset$. The proof of Lemma 8 in Appendix A, which is used in the proof of Theorem 6, requires that $\sum_{j \in S_i} w_{ij} = 0$ when $S_i = \emptyset$, but we still have $\mathbf{1}(S_i \neq \emptyset)\, w_{i0} + \sum_{j \in N} w_{ij} = 0$ when $S_i = \emptyset$. Thus, Theorem 6 continues to hold when customers can leave a nest without making a purchase.

6.2 Approximation Framework and Grid Construction

In this section, we construct a grid to deal with the nonunimodal nature of the objective function of the maximization problem on the right side of (21) and show that we can obtain solutions with a certain performance guarantee by using this grid. For each $S_i \in \mathcal{N}_i$, we propose constructing a separate grid $\{\tilde{y}_i^t(S_i) : t = 1, \ldots, T_i(S_i)\}$. These grid points are in increasing order such that $\tilde{y}_i^t(S_i) \leq \tilde{y}_i^{t+1}(S_i)$ for all $t = 1, \ldots, T_i(S_i) - 1$. Also, we ensure that the smallest and the largest one of the grid points satisfy $\tilde{y}_i^1(S_i) = \bar{L}_i(S_i)$ and $\tilde{y}_i^{T_i(S_i)}(S_i) = \bar{U}_i(S_i)$ so that the grid points cover the interval $[\bar{L}_i(S_i), \bar{U}_i(S_i)]$. In this case, instead of finding the value of z satisfying (21), we propose checking the values of y_i only at the grid points and finding the value of z satisfying

$$z = \sum_{i \in M} \max_{S_i \in \mathcal{N}_i,\, y_i \in \{\tilde{y}_i^t(S_i) \,:\, t = 1, \ldots, T_i(S_i)\}} \left\{ y_i^{\gamma_i} \, \frac{g_i(S_i, y_i)}{y_i} - y_i^{\gamma_i} z \right\}. \tag{22}$$

There are $\sum_{S_i \in \mathcal{N}_i} T_i(S_i)$ possible values for the decision variable (S_i, y_i) in the maximization problem on the right side above. Thus, solving this maximization problem is not too difficult when the number of grid points is not large. The next theorem gives a sufficient condition under which we can use the value of z satisfying (22) to obtain a good solution for problem (16).

Theorem 7 (Requirements for a Good Grid)**.** *For some $\rho \geq 0$, assume that the collection of grid points $\{\tilde{y}_i^t(S_i) : t = 1, \ldots, T_i(S_i)\}$ satisfy $g_i(S_i, \tilde{y}_i^{t+1}(S_i)) \leq (1 + \rho)\, g_i(S_i, \tilde{y}_i^t(S_i))$ for all $t = 1, \ldots, T_i(S_i) - 1$, $S_i \in \mathcal{N}_i$. If $\hat{z}$ denotes the value of z that satisfies* (22) *and ζ^* denotes the optimal objective value of problem* (16)*, then we have $(1 + \rho)\, \hat{z} \geq \zeta^*$.*

The theorem above is analogous to Theorem 2 and it can be shown by following the same reasoning in the proof of Theorem 2. By building on this theorem, we can construct an approximate solution to problem (16) with a certain performance guarantee. In particular, we find the value of z satisfying (22) and denote this value by $\hat{z}$. We let $(\hat{S}_i, \hat{y}_i)$ be an optimal solution to the maximization problem on the right side of (22) when this problem is solved with $z = \hat{z}$. For all $i \in M$, we solve the knapsack problem in (20) with $(S_i, y_i) = (\hat{S}_i, \hat{y}_i)$ and let $\hat{\boldsymbol{w}}_i$ be an optimal solution to this knapsack problem. In this case, it is possible to show that the solution $(\hat{S}_1, \ldots, \hat{S}_m, \hat{\boldsymbol{w}}_1, \ldots, \hat{\boldsymbol{w}}_m)$ is a $(1+\rho)$-approximate solution to problem (16). To see this result, since $\hat{z}$ is the value of z satisfying (22) and $(\hat{S}_i, \hat{y}_i)$ is an optimal solution to the maximization problem on the right side of (22) when this problem is solved with $z = \hat{z}$, we have

$$\hat{z} = \sum_{i \in M} \left\{ \hat{y}_i^{\gamma_i} \, \frac{g_i(\hat{S}_i, \hat{y}_i)}{\hat{y}_i} - \hat{y}_i^{\gamma_i} \, \hat{z} \right\}. \tag{23}$$

Furthermore, since $\hat{\boldsymbol{w}}_i$ is an optimal solution to problem (20) when this problem is solved with $(S_i, y_i) = (\hat{S}_i, \hat{y}_i)$, we have $g_i(\hat{S}_i, \hat{y}_i) = \sum_{j \in \hat{S}_i} \hat{w}_{ij} \, (\kappa_{ij} - \eta_{ij} \log \hat{w}_{ij})$. Also, by the discussion that follows the formulation of problem (20), since $\hat{y}_i \in [\bar{L}_i(\hat{S}_i), \bar{U}_i(\hat{S}_i)]$, the first constraint in problem (20) is tight at the optimal solution, yielding $\sum_{j \in \hat{S}_i} \hat{w}_{ij} = \hat{y}_i$. In this case, using the last two equalities in (23), we obtain

$$\hat{z} = \sum_{i \in M} \left\{ \Big(\sum_{j \in \hat{S}_i} \hat{w}_{ij} \Big)^{\gamma_i} \, \frac{\sum_{j \in \hat{S}_i} \hat{w}_{ij} \, (\kappa_{ij} - \eta_{ij} \log \hat{w}_{ij})}{\sum_{j \in \hat{S}_i} \hat{w}_{ij}} - \Big(\sum_{j \in \hat{S}_i} \hat{w}_{ij} \Big)^{\gamma_i} \hat{z} \right\}.$$

If we collect all terms that involve $\hat{z}$ in the equality above on the left side, solve for $\hat{z}$ and use the definition of $\Theta(S_1, \ldots, S_m, \boldsymbol{w}_1, \ldots, \boldsymbol{w}_m)$ in (15), then we obtain $\hat{z} = \Theta(\hat{S}_1, \ldots, \hat{S}_m, \hat{\boldsymbol{w}}_1, \ldots, \hat{\boldsymbol{w}}_m)$. As long as the grid points satisfy the assumption of Theorem 7, we also have $(1+\rho)\, \hat{z} \geq \zeta^*$, in which case, we obtain $(1+\rho)\, \Theta(\hat{S}_1, \ldots, \hat{S}_m, \hat{\boldsymbol{w}}_1, \ldots, \hat{\boldsymbol{w}}_m) = (1+\rho)\, \hat{z} \geq \zeta^*$. Therefore, it follows that if the collection of grid points satisfies the assumption of Theorem 7, then the expected revenue provided by the solution $(\hat{S}_1, \ldots, \hat{S}_m, \hat{\boldsymbol{w}}_1, \ldots, \hat{\boldsymbol{w}}_m)$ deviates from the optimal expected revenue ζ^* by no more than a factor of $1+\rho$, as desired.

The key question is how we can construct a collection of grid points $\{\tilde{y}_i^t(S_i) : t = 1, \ldots, T_i(S_i)\}$ that satisfies $g_i(S_i, \tilde{y}_i^{t+1}(S_i)) \leq (1+\rho)\, g_i(S_i, \tilde{y}_i^t(S_i))$ for all $t = 1, \ldots, T_i(S_i) - 1$ so that the assumption of Theorem 7 is satisfied. It turns out that the answer to this question is already given in Section 5. In particular, the only difference between problems (5) and (20) is that the former problem focuses on the full set of products N, whereas the latter problem focuses on the products that are in S_i. Therefore, for a fixed set of products S_i, we can repeat the same argument in Section 5, but restrict our attention only to the products in the set S_i to construct the collection of grid points $\{\tilde{y}_i^t(S_i) : t = 1, \ldots, T_i(S_i)\}$ that satisfies $g_i(S_i, \tilde{y}_i^{t+1}(S_i)) \leq (1+\rho)\, g_i(S_i, \tilde{y}_i^t(S_i))$ for all $t = 1, \ldots, T_i(S_i)$. In this case, the number of grid points in this collection is $T_i(S_i) = O(n + n \log(n\, \sigma_i(S_i)) / \log(1+\rho)) = O(n + n \log(n\, \sigma)/\log(1+\rho))$, where we let $\sigma_i(S_i) = \max_{j \in S_i} U_{ij} / \min_{j \in S_i} L_{ij}$ and $\sigma = \max\{\max_{j \in N} U_{ij} / \min_{j \in N} L_{ij} : i \in M\}$.

Finally, we note that we can find the value of z satisfying (22) by solving a linear program similar to the one in (9). The only difference is that the second set of constraints in this linear program has to be replaced with $x_i \geq y_i^{\gamma_i}\, g_i(S_i, y_i)/y_i - y_i^{\gamma_i}\, z$ for all $S_i \in \mathcal{N}_i$, $y_i \in \{\tilde{y}_i^t(S_i) : t = 1, \ldots, T_i(S_i)\}$, $i \in M$. Noting that $|\mathcal{N}_i| = O(n)$ and $T_i(S_i) = O(n + n\log(n\sigma)/\log(1+\rho))$, this linear program has $1+m$ decision variables and $O(mn^2 + mn^2\log(n\sigma)/\log(1+\rho))$ constraints. The optimal objective value of the linear program provides the value of z that satisfies (22). Once we have the value of z that satisfies (22), we can follow the approach described right after Theorem 7 to find a $(1+\rho)$-approximate solution to problem (16).

7 Computational Experiments

In this section, we test the quality of the solutions obtained by the approximation method that we propose in this paper. For economy of space, we present computational results for the first problem variant where the set of products offered to customers is fixed and we determine the prices for these products. The qualitative findings from our computational experiments do not change when we consider the second problem variant, where we jointly determine the products that should be offered to customers and their corresponding prices.

7.1 Experimental Setup

Throughout this section, we refer to our approximation method as APP. In particular, APP uses the algorithm in Section 5.3 to find a $(1+\rho)$-approximate solution to problem (2). In our computational experiments, we set $\rho = 0.005$ so that APP obtains a solution to problem (2) whose expected revenue deviates from the optimal expected revenue by at most a factor of 1.005, corresponding to a worst case optimality gap of 0.5%. We emphasize that APP ensures a performance guarantee of $1+\rho$, but this performance guarantee is in worst case sense and the solution obtained by APP for a particular problem instance can perform significantly better than what is predicted by the worst case performance guarantee of $1+\rho$. So, a natural question is whether we can come up with a more refined approach to assess the performance of the solution obtained by APP for a particular problem instance. It turns out that we can solve a linear program to obtain an upper bound on the optimal expected revenue in problem (2). To formulate this linear program, we let $\{\bar{y}_i^t : t = 1, \ldots, \tau_i\}$ be any collection of grid points such that $\bar{y}_i^t \leq \bar{y}_i^{t+1}$ for all $t = 1, \ldots, \tau_i - 1$. Also, we assume that $\bar{y}_i^1 = \bar{L}_i$ and $\bar{y}_i^{\tau_i} = \bar{U}_i$ so that the grid points cover the interval $[\bar{L}_i, \bar{U}_i]$. In this case, it is possible to show that the optimal objective value of the linear program

$$\min \left\{ z : z \geq \sum_{i \in M} x_i,\ x_i \geq \left(\bar{y}_i^t\right)^{\gamma_i} \frac{g_i(\bar{y}_i^{t+1})}{\bar{y}_i^t} - \left(\bar{y}_i^t\right)^{\gamma_i} z \ \ \forall\, t = 1, \ldots, \tau_i - 1,\ i \in M \right\} \tag{24}$$

provides an upper bound on the optimal expected revenue Z^* in problem (2). In the linear program above, the decision variables are z and $(x_1, \ldots, x_m)$. Theorem 9 in Appendix A shows that the optimal objective value of the linear program above is indeed an upper bound on the optimal expected revenue Z^*. It is worthwhile to note that the optimal objective value of problem (24) is always an upper bound on the optimal expected revenue, irrespective of the number and placement of the grid points. However, if the grid points satisfy $g_i(\bar{y}_i^{t+1}) \leq (1+\rho)\, g_i(\bar{y}_i^t)$ for all $t = 1, \ldots, \tau_i - 1$ and $i \in M$, then Theorem 9 also shows that the upper bound provided the linear program above deviates from the optimal expected revenue Z^* by at most a factor of $1 + \rho$. Thus, if we choose the grid points in the linear program above carefully, then this linear program approximates the optimal expected revenue with a factor of $1+\rho$ accuracy. For example, we can plug the grid points given in Theorem 5 into problem (24) to approximate the optimal expected revenue with a factor of $1+\rho$ accuracy. Once we solve problem (24) with a particular set of grid points, we can compare the optimal objective value of this linear program with the expected revenue from the solution obtained by APP to get a feel for the optimality gap of the solution obtained by APP.

In our computational experiments, we generate a large number of problem instances. For each problem instance, we compute the solution obtained by APP. Also, we solve the linear program in (24) to obtain an upper bound on the optimal expected revenue. By comparing the expected revenue from the solution obtained by APP with the upper bound on the optimal expected revenue, we assess the optimality gap of APP. We use the following strategy to generate our problem instances. In all of our test problems, the number of nests is equal to the number of products in each nest so that $m = n$. To come up with the dissimilarity parameters of the nests, we generate γ_i from the uniform distribution over $[\gamma^L, \gamma^U]$ for all $i \in M$. We use $[\gamma^L, \gamma^U] = [0.05, 0.35]$, $[\gamma^L, \gamma^U] = [0.35, 0.65]$ or $[\gamma^L, \gamma^U] = [0.65, 1]$. For all $i \in M$, $j \in N$, we generate α_{ij} from the uniform distribution over $[-2, 2]$ and β_{ij} from the uniform distribution over $[0.5, 1.5]$. To come up with the bounds on the prices, after generating the parameters γ_i, α_{ij} and β_{ij} for all $i \in M$, $j \in N$, we solve problem (2) under the assumption that there are no bounds on the prices of the products. We use p^*_{ij} to denote the optimal price of product j in nest i when there are no price bounds. In this case, we generate the bounds l_{ij} and u_{ij} on the price of product j in nest i such that we either have $p^*_{ij} < l_{ij}$ or $p^*_{ij} > u_{ij}$. In this way, if we solve problem (2) without any price bounds, then the unconstrained price of product j in nest i does not lie in the interval $[l_{ij}, u_{ij}]$. Our hope is that this approach allows us to generate problem instances where the price bounds are binding at the optimal solution. To be specific, after computing p^*_{ij} for all $i \in M$, $j \in N$, we set either $[l_{ij}, u_{ij}] = [p^*_{ij} + \Delta,\, 1.75 \times p^*_{ij} + \Delta]$ or $[l_{ij}, u_{ij}] = [0.25 \times p^*_{ij} - \Delta,\, p^*_{ij} - \Delta]$, each case occurring with equal probability. If one of the end points of the interval $[0.25 \times p^*_{ij} - \Delta,\, p^*_{ij} - \Delta]$ turns out to be negative, then we round it up to zero. When we generate the price bounds in this fashion, the unconstrained price p^*_{ij} of product j in nest i violates one of the price bounds l_{ij} or u_{ij} by about Δ. Furthermore, the width of the interval $[l_{ij}, u_{ij}]$ is about 75% of the unconstrained price of product j in nest i. Thus, products that tend to have larger prices also tend to have wider price bound intervals.

In our computational experiments, we vary the common value of m and n over $\{5, 10, 15\}$, corresponding to three different numbers of nests and numbers of products in each nest. We can view the common value of m and n as the scale of the problem instance, measuring the number of decision variables. We vary $[\gamma^L, \gamma^U]$ over $\{[0.05, 0.35], [0.35, 0.65], [0.65, 1]\}$, yielding low, medium and high levels of dissimilarity parameters. Finally, we vary Δ over $\{1, 2, 3\}$, corresponding to three different levels of violation of the price upper and lower bounds when we solve problem (2) without any price bounds. Varying three parameters over three levels, we obtain 27 parameter combinations. For each parameter combination, we generate 100 individual problem instances by using the approach described in the paragraph above. For each individual problem instance, we compute the solution obtained by APP. Also, we solve the linear program in (24) to obtain an upper bound on the optimal expected revenue. The grid points $\{\bar{y}_i^t : t = 1, \ldots, \tau_i\}$ that we use in this linear program are identical to the grid points $\{\tilde{y}_i^t : t = 1, \ldots, T_i\}$ that we use for APP. Thus, by Theorem 5, the grid points $\{\bar{y}_i^t : t = 1, \ldots, \tau_i\}$ satisfy $g_i(\bar{y}_i^{t+1}) \leq (1+\rho)\, g_i(\bar{y}_i^t)$ for all $t = 1, \ldots, \tau_i - 1$, in which case, Theorem 9 implies that the upper bound provided by the linear program in (24) approximates the optimal expected revenue with a factor of $1 + \rho$ accuracy. By comparing the expected revenue from the solution obtained by APP with the upper bound on the optimal expected revenue, we assess the optimality gap of the solution obtained by APP.

7.2 Computational Results

We give our main computational results in Table 1. In this table, the first three columns show the parameter combination for our test problems by using the tuple $(m, [\gamma^L, \gamma^U], \Delta)$, where the first component gives the common value for the number of nests and the number of products in each nest, the second component corresponds to the interval over which we generate the dissimilarity parameters and the third component characterizes how much the unconstrained prices violate the price bounds. The fourth and fifth columns respectively show the average lower and upper price bounds when we generate our test problems by using the approach described in the previous section. The average is computed over all products in all nests and over all problem instances in a particular parameter combination. We recall that we generate 100 individual problem instances in each parameter combination. Our goal in these two columns is to give a feel for the magnitude of the prices and their bounds. The sixth column shows the percent gap between the upper bound on the optimal expected revenue and the expected revenue from the solution obtained by APP, averaged over all problem instances in a particular parameter combination. In particular, for problem instance k, we let UB^k be the upper bound on the optimal expected revenue provided by the optimal objective value of the linear program in (24) and RAPP^k be the expected revenue from the solution obtained by APP. In this case, the sixth column shows $\frac{1}{100} \sum_{k=1}^{100} 100 \times (\text{UB}^k - \text{RAPP}^k)/\text{UB}^k$. The seventh column shows the maximum percent gap between the upper bound on the optimal expected revenue and the expected revenue from the solution obtained by APP over all problem instances in a parameter combination. That is, the seventh column shows $\max\{100 \times (\text{UB}^k - \text{RAPP}^k)/\text{UB}^k : k = 1, \ldots, 100\}$. The eighth column shows the average CPU seconds for APP to obtain a solution for one problem instance. Finally, the ninth column shows the average number of points in the grid used by APP, where the average is computed over all nests and over all problem instances in a parameter combination.

Our results indicate that the solutions obtained by APP perform remarkably well. Over all of our test problems, the average optimality gap of these solutions is no larger than 0.117%, which

	Param. Comb.		Avg. Price Bounds		% Gap with Upp. Bnd.		CPU	No. of Grid
m	$[\gamma^L, \gamma^U]$	Δ	Low.	Upp.	Avg.	Max.	Secs.	Points
5	$[0.05, 0.35]$	1	6.83	15.09	0.289%	0.430%	14.79	3,550
5	$[0.05, 0.35]$	2	7.05	15.08	0.254%	0.402%	14.88	3,550
5	$[0.05, 0.35]$	3	7.64	16.04	0.232%	0.412%	15.71	3,694
5	$[0.35, 0.65]$	1	1.92	3.85	0.114%	0.213%	1.67	500
5	$[0.35, 0.65]$	2	2.32	3.74	0.032%	0.142%	1.44	396
5	$[0.35, 0.65]$	3	2.92	4.07	0.022%	0.200%	1.39	367
5	$[0.65, 1.00]$	1	2.03	4.17	0.153%	0.214%	1.94	583
5	$[0.65, 1.00]$	2	2.50	4.16	0.029%	0.126%	1.59	430
5	$[0.65, 1.00]$	3	3.01	4.27	0.008%	0.101%	1.57	409
10	$[0.05, 0.35]$	1	8.38	18.59	0.224%	0.375%	41.17	4,762
10	$[0.05, 0.35]$	2	8.11	17.66	0.185%	0.367%	40.03	4,668
10	$[0.05, 0.35]$	3	8.87	18.92	0.173%	0.370%	43.95	5,072
10	$[0.35, 0.65]$	1	2.22	4.66	0.133%	0.180%	4.74	701
10	$[0.35, 0.65]$	2	2.71	4.63	0.030%	0.062%	3.83	515
10	$[0.35, 0.65]$	3	3.21	4.73	0.007%	0.052%	3.51	444
10	$[0.65, 1.00]$	1	2.60	5.78	0.191%	0.258%	6.80	996
10	$[0.65, 1.00]$	2	3.15	5.91	0.123%	0.203%	5.89	821
10	$[0.65, 1.00]$	3	3.61	5.82	0.014%	0.060%	4.72	596
15	$[0.05, 0.35]$	1	8.77	19.26	0.186%	0.342%	64.73	4,926
15	$[0.05, 0.35]$	2	8.93	19.40	0.157%	0.335%	67.03	5,116
15	$[0.05, 0.35]$	3	9.37	19.97	0.150%	0.305%	72.82	5,544
15	$[0.35, 0.65]$	1	2.48	5.38	0.148%	0.210%	8.88	866
15	$[0.35, 0.65]$	2	2.97	5.41	0.045%	0.086%	7.45	676
15	$[0.35, 0.65]$	3	3.48	5.40	0.009%	0.038%	6.29	528
15	$[0.65, 1.00]$	1	3.21	7.08	0.125%	0.209%	15.43	1,425
15	$[0.65, 1.00]$	2	3.59	7.11	0.083%	0.183%	14.21	1,293
15	$[0.65, 1.00]$	3	4.08	7.11	0.041%	0.075%	12.96	1,140
Average					0.117%			

Table 1: Performance of APP.

is significantly better than the worst case optimality gap of 0.5% that we ensure by choosing $\rho = 0.005$. The optimality gaps are particularly small when $[\gamma^L, \gamma^U]$ is close to one so that the dissimilarity parameters of the nests tend to be close to one. For example, if we focus only on the problem instances with $[\gamma^L, \gamma^U] = [0.65, 1]$, then the average optimality gap comes out to be 0.085%, whereas the average optimality gap comes out to be 0.206% when we focus only on the problem instances with $[\gamma^L, \gamma^U] = [0.05, 0.35]$. If the dissimilarity parameters are all equal to one, then the objective function of the maximization problem on the right side of (6) is $g_i(y_i) - y_i\, z$, in which case, noting that $g_i(y_i)$ is a concave function of y_i, the objective value of this maximization problem is a concave function of y_i as well. Thus, intuitively speaking, the objective function of the maximization problem on the right side of (6) behaves well when we have $\gamma_i = 1$, avoiding the pathological cases such as the one shown in Figure 1. When $[\gamma^L, \gamma^U] = [0.65, 1]$ so that the dissimilarity parameters of the nests take on values closer to one, the performance of APP also turns out to be substantially better than what is predicted by the worst case performance guarantee of 0.5%. Nevertheless, even when $[\gamma^L, \gamma^U] = [0.05, 0.35]$ so that the dissimilarity parameters can be far from one, APP can effectively find solutions with the desired performance guarantee. Over all of our test problems, the maximum optimality gap of the solutions obtained by APP is 0.43%. Similar to

Param. Comb. $(5, [0.65, 1], 1)$

ρ	CPU Secs.
0.01	0.93
0.005	2.13
0.001	8.78
0.0005	19.40

Param. Comb. $(15, [0.05, 0.35], 3)$

ρ	CPU Secs.
0.01	12.81
0.005	60.76
0.001	124.58
0.0005	398.21

Table 2: CPU seconds for APP as a function of the performance guarantee ρ.

our observations for the average optimality gaps, the parameter combinations for which we obtain maximum optimality gaps that are close to 0.5% correspond to the parameter combinations with $[\gamma^L, \gamma^U] = [0.05, 0.35]$, yielding dissimilarity parameters further from one.

The CPU seconds for APP are reasonable for practical implementation. In our largest problem instances with $m = 15$, noting that $n = m$, there are a total of 225 products and we can obtain solutions for these problem instances in about one minute. Furthermore, the CPU seconds for APP scale in a graceful fashion. For example, if we double m, then the total number of products increases by a factor of four and the CPU seconds increase by no more than a factor of four. In Table 2, we show the CPU seconds for APP as a function of the performance guarantee ρ. The left portion of the table focuses on a problem instance with $m = n = 5$, $[\gamma^L, \gamma^U] = [0.65, 1]$ and $\Delta = 1$, whereas the right portion focuses on a problem instance with $m = n = 15$, $[\gamma^L, \gamma^U] = [0.05, 0.35]$ and $\Delta = 3$. The parameter combination for the second problem instance corresponds to the parameter combination with the largest CPU seconds in Table 1. In each portion of Table 2, the first column shows the performance guarantee ρ and the second column shows the CPU seconds for APP. The results indicate that we can obtain a solution with a worst case optimality gap of 0.05% in about six minutes, even for a problem instance with 225 products. If we are content with a worst case optimality gap of 1%, then we can obtain solutions in about 10 seconds.

It is useful to note that naive approaches for finding solutions to problem (2) can yield poor results. For example, a first cut approach for finding a solution to problem (2) is to solve this problem under the assumption that there are no bounds on the prices of the products. If the unconstrained prices obtained in this fashion are outside the price bound constraints, then we can round them up or down to their corresponding lower or upper bounds. In Table 3, we show the performance of this approach for the test problems in our experimental setup. The first three columns in this table show the parameter combination for our test problems. The fourth column shows the average percent gap between the upper bound on the optimal expected revenue and the expected revenue from the solution that we obtain by rounding the unconstrained prices up or down to the price bounds, whereas the fifth column shows the maximum percent gap between the upper bound on the optimal expected revenue and the expected revenue obtained by rounding the unconstrained prices. The average and maximum percent gaps are computed over the same 100 problem instances in Table 1. The results in Table 3 indicate that rounding the unconstrained prices up or down to the price bounds can perform poorly. There are parameter combinations where this

Param. Comb.			% Gap with Upp. Bnd.		Param. Comb.			% Gap with Upp. Bnd.		Param. Comb.			% Gap with Upp. Bnd.	
m	$[\gamma^L, \gamma^U]$	Δ	Avg.	Max.	m	$[\gamma^L, \gamma^U]$	Δ	Avg.	Max.	m	$[\gamma^L, \gamma^U]$	Δ	Avg.	Max.
5	[0.05, 0.35]	1	0.658%	3.798%	10	[0.05, 0.35]	1	1.373%	7.347%	15	[0.05, 0.35]	1	1.483%	3.613%
5	[0.05, 0.35]	2	0.645%	4.285%	10	[0.05, 0.35]	2	1.587%	6.777%	15	[0.05, 0.35]	2	2.069%	4.575%
5	[0.05, 0.35]	3	1.154%	7.765%	10	[0.05, 0.35]	3	2.569%	7.360%	15	[0.05, 0.35]	3	3.439%	9.326%
5	[0.35, 0.65]	1	0.114%	0.213%	10	[0.35, 0.65]	1	0.133%	0.180%	15	[0.35, 0.65]	1	0.148%	0.210%
5	[0.35, 0.65]	2	1.930%	14.749%	10	[0.35, 0.65]	2	1.126%	15.492%	15	[0.35, 0.65]	2	0.050%	0.512%
5	[0.35, 0.65]	3	8.267%	32.248%	10	[0.35, 0.65]	3	17.466%	42.090%	15	[0.35, 0.65]	3	14.519%	34.727%
5	[0.65, 1.00]	1	0.153%	0.214%	10	[0.65, 1.00]	1	0.213%	0.845%	15	[0.65, 1.00]	1	0.630%	38.811%
5	[0.65, 1.00]	2	6.928%	37.986%	10	[0.65, 1.00]	2	0.445%	8.797%	15	[0.65, 1.00]	2	0.337%	2.133%
5	[0.65, 1.00]	3	39.667%	83.798%	10	[0.65, 1.00]	3	10.869%	64.471%	15	[0.65, 1.00]	3	2.402%	20.070%
Average			6.613%		Average			3.978%		Average			2.768%	

Table 3: Performance of the prices obtained by rounding the unconstrained prices up or down to the price bounds.

approach results in average optimality gaps of about 40%. Over all parameter combinations, the average optimality gap of this approach is over 4%. The dramatically high maximum optimality gaps also indicate that we can generate test problems where the unconstrained prices give essentially no indication of the optimal prices under price bounds.

If the parameters $\{\beta_{ij} : j \in N\}$ are identical to each other within each nest i and there are no bounds on the prices, then Li and Huh (2011) show that the optimal solution to problem (2) can be obtained by finding the maximizer of a scalar concave function. The computational complexity of APP remains the same when we work with test problems where the parameters $\{\beta_{ij} : j \in N\}$ are identical to each other within each nest i and there are bounds on the prices. Although APP does not enjoy further efficiencies, it is worthwhile to check whether this approximation method provides noticeably better performance when working with such test problems. Table 4 shows the performance of APP on test problems where the parameters $\{\beta_{ij} : j \in N\}$ are identical to each other within each nest i and there are bounds on the prices. We generate these test problems by using the same approach described in the previous section. The only difference is that we generate $\bar{\beta}_i$ for each nest i from the uniform distribution over $[0.5, 1.5]$ and set $\beta_{ij} = \bar{\beta}_i$ for all $j \in N$. In this way, we ensure that the parameters $\{\beta_{ij} : j \in N\}$ are identical to each other within each nest i. The first three columns in Table 4 show the parameter combination for our test problems. The fourth column shows the percent gap between the upper bound on the optimal expected revenue and the expected revenue from the solution obtained by APP, averaged over all problem instances in a particular parameter configuration. The fifth column shows the average percent gap between the upper bound on the optimal expected revenue and the expected revenue from the solution obtained by rounding the unconstrained prices up or down to the price bounds. Comparing the results in Table 4 to those in Table 1 indicates that the performance of APP does not change noticeably when we work with test problems where the parameters $\{\beta_{ij} : j \in N\}$ are identical to each other within each nest i. Although we do not report the maximum optimality gaps in Table 4, the maximum optimality gap for APP over all test problems is 0.403%. We do not report the average price bounds, CPU seconds and numbers of grid points either for economy of space, but these values have the

Param. Comb. m	$[\gamma^L, \gamma^U]$	Δ	Avg. % Gap with Upp. Bnd. APP	Rnd.
5	$[0.05, 0.35]$	1	0.252%	0.604%
5	$[0.05, 0.35]$	2	0.242%	0.596%
5	$[0.05, 0.35]$	3	0.236%	1.238%
5	$[0.35, 0.65]$	1	0.101%	0.101%
5	$[0.35, 0.65]$	2	0.042%	7.371%
5	$[0.35, 0.65]$	3	0.028%	11.375%
5	$[0.65, 1.00]$	1	0.151%	0.152%
5	$[0.65, 1.00]$	2	0.032%	18.595%
5	$[0.65, 1.00]$	3	0.010%	52.444%
Average			0.122%	10.275%

Param. Comb. m	$[\gamma^L, \gamma^U]$	Δ	Avg. % Gap with Upp. Bnd. APP	Rnd.
10	$[0.05, 0.35]$	1	0.184%	0.619%
10	$[0.05, 0.35]$	2	0.166%	0.726%
10	$[0.05, 0.35]$	3	0.168%	2.494%
10	$[0.35, 0.65]$	1	0.135%	0.135%
10	$[0.35, 0.65]$	2	0.038%	8.420%
10	$[0.35, 0.65]$	3	0.010%	31.322%
10	$[0.65, 1.00]$	1	0.210%	0.220%
10	$[0.65, 1.00]$	2	0.143%	3.725%
10	$[0.65, 1.00]$	3	0.027%	29.164%
Average			0.120%	8.536%

Param. Comb. m	$[\gamma^L, \gamma^U]$	Δ	Avg. % Gap with Upp. Bnd. APP	Rnd.
15	$[0.05, 0.35]$	1	0.151%	0.687%
15	$[0.05, 0.35]$	2	0.142%	0.749%
15	$[0.05, 0.35]$	3	0.147%	3.910%
15	$[0.35, 0.65]$	1	0.165%	0.165%
15	$[0.35, 0.65]$	2	0.058%	4.441%
15	$[0.35, 0.65]$	3	0.012%	37.764%
15	$[0.65, 1.00]$	1	0.176%	4.105%
15	$[0.65, 1.00]$	2	0.131%	1.958%
15	$[0.65, 1.00]$	3	0.049%	13.065%
Average			0.115%	7.427%

Table 4: Computational results for the test problems where the parameters $\{\beta_{ij} : j \in N\}$ are identical within each nest i.

same magnitudes as those in Table 1. In particular, the largest CPU seconds for APP over all test problems is about 40 seconds. Finally, the prices obtained by rounding the unconstrained prices up or down to the price bounds can provide poor performance, yielding average optimality gaps of about 50% for certain parameter combinations.

At appropriate places in the paper, we describe how we can use the no purchase preference weight w_{i0} to allow a customer to leave nest i without making a purchase. Table 5 shows the performance of APP when customers can leave a nest without making a purchase. The test problems that we consider in this table are generated by using the same approach described in the previous section. The only difference is that to come up with the no purchase preference weights, we generate w_{i0} from the uniform distribution over $[0, 1]$ for all $i \in M$. These no purchase preference weights are comparable to the value of one that we use for the no purchase preference weight in the definition of $Q_i(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m)$. The structure of Table 5 is identical to that of Table 4. The results in Table 5 indicate that APP continues to perform remarkably well when customers can leave a nest without making a purchase. Over all test problems, the average optimality gap of APP is 0.04%, which is significantly better than the worst case optimality gap of 0.5% that we ensure by choosing $\rho = 0.005$. In contrast, the prices obtained by rounding the unconstrained prices up or down to the price bounds can have average optimality gaps exceeding 40%.

8 Conclusions

We developed approximation methods for pricing problems where customers choose under the nested logit model and there are bounds on the prices that can be charged for the products. We considered two problem variants. In the first variant, the set of products offered to customers is fixed and we want to determine the prices for these products. In the second variant, we jointly determine the products to be offered and their corresponding prices. For both problem variants, given any $\rho > 0$, we showed how to obtain a solution whose expected revenue deviates from the

Param. Comb.			Avg. % Gap with Upp. Bnd.		Param. Comb.			Avg. % Gap with Upp. Bnd.		Param. Comb.			Avg. % Gap with Upp. Bnd.	
m	$[\gamma^L, \gamma^U]$	Δ	APP	Rnd.	m	$[\gamma^L, \gamma^U]$	Δ	APP	Rnd.	m	$[\gamma^L, \gamma^U]$	Δ	APP	Rnd.
5	$[0.05, 0.35]$	1	0.074%	0.074%	10	$[0.05, 0.35]$	1	0.101%	0.101%	15	$[0.05, 0.35]$	1	0.102%	0.102%
5	$[0.05, 0.35]$	2	0.070%	0.125%	10	$[0.05, 0.35]$	2	0.049%	0.049%	15	$[0.05, 0.35]$	2	0.040%	0.062%
5	$[0.05, 0.35]$	3	0.006%	0.435%	10	$[0.05, 0.35]$	3	0.002%	0.304%	15	$[0.05, 0.35]$	3	0.002%	0.143%
5	$[0.35, 0.65]$	1	0.046%	0.046%	10	$[0.35, 0.65]$	1	0.055%	0.055%	15	$[0.35, 0.65]$	1	0.092%	0.092%
5	$[0.35, 0.65]$	2	0.020%	2.034%	10	$[0.35, 0.65]$	2	0.006%	7.616%	15	$[0.35, 0.65]$	2	0.006%	3.775%
5	$[0.35, 0.65]$	3	0.004%	4.369%	10	$[0.35, 0.65]$	3	0.000%	12.640%	15	$[0.35, 0.65]$	3	0.000%	22.574%
5	$[0.65, 1.00]$	1	0.057%	1.064%	10	$[0.65, 1.00]$	1	0.154%	6.638%	15	$[0.65, 1.00]$	1	0.129%	5.474%
5	$[0.65, 1.00]$	2	0.018%	28.855%	10	$[0.65, 1.00]$	2	0.015%	7.283%	15	$[0.65, 1.00]$	2	0.030%	10.702%
5	$[0.65, 1.00]$	3	0.004%	43.697%	10	$[0.65, 1.00]$	3	0.000%	44.135%	15	$[0.65, 1.00]$	3	0.000%	41.336%
Average			0.033%	8.966%	Average			0.042%	8.758%	Average			0.045%	9.362%

Table 5: Computational results for the test problems where customers can leave a nest without making a purchase.

optimal expected revenue by no more than a factor of $1 + \rho$. To obtain this solution, we solved a linear program and the number of constraints in this linear program grew at rate $1/\rho$. Our computational experiments demonstrated that our approximation methods can obtain solutions to practical problems within reasonable computation time.

There are a number of research directions to pursue. In this paper, we showed how to address upper and lower bound constraints on the prices and it is tempting to see whether one can work with other types of price constraints. Furthermore, when we need to jointly determine the products to be offered to customers and their corresponding prices, certain settings may call for imposing constraints on the set of offered products. We were able to impose constraints on the number of products offered in each nest and it would be interesting to explore other types of constraints. For example, products may have dependencies on each other and it may be possible to offer a particular product only when certain other products are also offered. Finally, our approach can be used to obtain the optimal prices under the multinomial logit model and we can investigate generalizations based on this observation. The multinomial logit model corresponds to a special case of the nested logit model with $\gamma_i = 1$ for all $i \in M$. If $\gamma_i = 1$ for all $i \in M$, then the objective function of the maximization problem on the right side of (6) becomes $g_i(y_i) - y_i\, z$. Since $g_i(y_i)$ is the optimal objective value of problem (5) as a function of the right side of the first constraint, it is concave in y_i, in which case, the objective function of the maximization problem on the right of (6) is also concave in y_i. So, we can find the value of z satisfying (6) by using bisection search. During the course of the search, we solve the maximization problem on the right side of (6), but this problem can be solved efficiently since its objective function is concave. Thus, we can obtain the optimal prices when $\gamma_i = 1$ for all $i \in M$. It would be useful to characterize the performance of this approach as the dissimilarity parameters $(\gamma_1, \ldots, \gamma_m)$ deviate from one.

References

Ben-Akiva, M. and Lerman, S. (1994), *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.

Davis, J., Gallego, G. and Topaloglu, H. (2011), Assortment optimization under variants of the nested logit model, Technical report, Cornell University, School of Operations Research and Information Engineering.
Available at `http://legacy.orie.cornell.edu/~huseyin/publications/publications.html`.

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

Gallego, G., Iyengar, G., Phillips, R. and Dubey, A. (2004), Managing flexible products on a network, Computational Optimization Research Center Technical Report TR-2004-01, Columbia University.

Gallego, G. and Topaloglu, H. (2012), Constrained assortment optimization for the nested logit model, Technical report, Cornell University, School of Operations Research and Information Engineering.
Available at `http://legacy.orie.cornell.edu/~huseyin/publications/publications.html`.

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

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

Jagabathula, S., Farias, V. and Shah, D. (2011), Assortment optimization under general choice, *in* 'INFORMS Conference, Charlotte, NC'.

Kunnumkal, S. and Topaloglu, H. (2008), 'A refined deterministic linear program for the network revenue management problem with customer choice behavior', *Naval Research Logistics Quarterly* **55**(6), 563–580.

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).

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

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

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.

Meissner, J., Strauss, A. and Talluri, K. (2012), 'An enhanced concave program relaxation for choice network revenue management', *Production and Operations Management* (to appear).

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. (2009), 'A PTAS for capacitated sum-of-ratios optimization', *Operations Research Letters* **37**(4), 230–238.

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., Tong, C. and Topaloglu, H. (2013), Assortment optimization under the multinomial logit model with random choice parameters, Technical report, Cornell University, School of Operations Research and Information Engineering. Available at `http://legacy.orie.cornell.edu/~huseyin/publications/publications.html`.

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.

Wang, R. (2012*a*), Assortment management under the generalized attraction model with a capacity constraint, Technical report, HP Labs, Palo Alto, CA.

Wang, R. (2012*b*), 'Capacitated assortment and price optimization under the multinomial logit model', *Operations Research Letters* (to appear).

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.

Zhang, D. and Adelman, D. (2009), 'An approximate dynamic programming approach to network revenue management with customer choice', *Transportation Science* **42**(3), 381–394.

Zhang, D. and Lu, Z. (2011), 'Assessing the value of dynamic pricing in network revenue management', *INFORMS Journal on Computing* (to appear).

A Appendix: Omitted Results

A.1 Proof of Proposition 1

In this section, we complete the proof of Proposition 1. As mentioned in the proof of Proposition 1 in the main text, the result follows if we can show that

$$\max_{\boldsymbol{w}_i \in [\boldsymbol{L}_i, \boldsymbol{U}_i]} \left\{ \Big(\sum_{j\in N} w_{ij} \Big)^{\gamma_i} \frac{\sum_{j\in N} w_{ij}\,(\kappa_{ij} - \eta_{ij} \log w_{ij})}{\sum_{j \in N} w_{ij}} - \Big(\sum_{j\in N} w_{ij} \Big)^{\gamma_i} z \right\} = \max_{y_i \in [\bar{L}_i, \bar{U}_i]} \left\{ y_i^{\gamma_i} \frac{g_i(y_i)}{y_i} - y_i^{\gamma_i} z \right\}.$$

We 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 $\zeta_L^* \leq \zeta_R^*$. We let $\boldsymbol{w}_i^*$ be an optimal solution to the problem on the left side above. Since $\boldsymbol{w}_i^* \in [\boldsymbol{L}_i, \boldsymbol{U}_i]$, we have $\sum_{j\in N} w_{ij}^* \geq \sum_{j \in N} L_{ij} = \bar{L}_i$. We proceed under the assumption that we also have $\sum_{j\in N} w_{ij}^* \leq \bar{U}_i$ and we carefully address this assumption later on. The solution $\boldsymbol{w}_i^*$ is feasible to problem (5) when this problem is solved with $y_i = \sum_{j\in N} w_{ij}^*$. Thus, letting $\hat{y}_i = \sum_{j\in N} w_{ij}^*$, we have $g_i(\hat{y}_i) \geq \sum_{j\in N} w_{ij}^*\,(\kappa_{ij} - \eta_{ij} \log w_{ij}^*)$. Furthermore, since $\sum_{j \in N} w_{ij}^* \in [\bar{L}_i, \bar{U}_i]$, the solution $\hat{y}_i$ is feasible to the problem on the right side above. In this case, noting the last inequality and the fact that $\hat{y}_i = \sum_{j\in N} w_{ij}^*$, the solution $\hat{y}_i$ is feasible to the problem on the right side above providing an objective value for this problem that is larger than the one provided by the solution $\boldsymbol{w}_i^*$ for the problem on the left side. So, we get $\zeta_R^* \geq \zeta_L^*$. To address the assumption that $\sum_{j\in N} w_{ij}^* \leq \bar{U}_i$, assume on the contrary that $\sum_{j \in N} w_{ij}^* > \bar{U}_i$. In this case, if we solve problem (5) with $y_i = \sum_{j\in N} w_{ij}^* > \bar{U}_i$ and use $\hat{\boldsymbol{w}}_i$ to denote an optimal solution, then the discussion right after problem (5) implies that the first constraint in this problem is not tight at the optimal solution, yielding $\sum_{j\in N} \hat{w}_{ij} < y_i = \sum_{j\in N} w_{ij}^*$. Furthermore, if we solve problem (5) with $y_i = \sum_{j \in N} w_{ij}^*$, then $\boldsymbol{w}_i^*$ is a feasible solution to this problem, indicating that we have $\sum_{j\in N} w_{ij}^*\,(\kappa_{ij} - \eta_{ij}\log w_{ij}^*) \leq \sum_{j\in N} \hat{w}_{ij}\,(\kappa_{ij} - \eta_{ij} \log \hat{w}_{ij})$. Therefore, we obtain

$$\begin{aligned} &\Big(\sum_{j\in N} w_{ij}^* \Big)^{\gamma_i} \frac{\sum_{j\in N} w_{ij}^*\,(\kappa_{ij} - \eta_{ij} \log w_{ij}^*)}{\sum_{j\in N} w_{ij}^*} - \Big(\sum_{j \in N} w_{ij}^* \Big)^{\gamma_i} z \\ &\qquad < \Big(\sum_{j\in N} \hat{w}_{ij} \Big)^{\gamma_i} \frac{\sum_{j\in N} \hat{w}_{ij}\,(\kappa_{ij} - \eta_{ij} \log \hat{w}_{ij})}{\sum_{j\in N} \hat{w}_{ij}} - \Big(\sum_{j\in N} \hat{w}_{ij} \Big)^{\gamma_i} z, \end{aligned}$$

where we use the fact that $\sum_{j\in N} w_{ij}^*\,(\kappa_{ij} - \eta_{ij} \log w_{ij}^*) \leq \sum_{j \in N} \hat{w}_{ij}\,(\kappa_{ij} - \eta_{ij}\log \hat{w}_{ij})$, $\gamma_i \leq 1$, $\sum_{j\in N} \hat{w}_{ij} < \sum_{j\in N} w_{ij}^*$ and $z > 0$. The inequality above contradicts the fact that $\boldsymbol{w}_i^*$ is an optimal solution to the problem on the left side above. So, we must have $\sum_{j \in N} w_{ij}^* \leq \bar{U}_i$.

Second, we show that $\zeta_L^* \geq \zeta_R^*$. Using y_i^* to denote an optimal solution to the problem on the right side above, we let $\hat{\boldsymbol{w}}_i$ be an optimal solution to problem (5) when this problem is solved with $y_i = y_i^*$, in which case, $g_i(y_i^*) = \sum_{j \in N} \hat{w}_{ij}\,(\kappa_{ij} - \eta_{ij} \log \hat{w}_{ij})$. Furthermore, since $y_i^* \in [\bar{L}_i, \bar{U}_i]$, the discussion right after problem (5) indicates that the first constraint in problem (5) has to be satisfied as equality, yielding $\sum_{j\in N} \hat{w}_{ij} = y_i^*$. Thus, the last two equalities imply that $\hat{\boldsymbol{w}}_i$ is a feasible solution to the problem on the left side above yielding the same objective value provided by the solution y_i^* for the problem on the right side. So, we get $\zeta_L^* \geq \zeta_R^*$.

A.2 Lemma 8

Lemma 8 (Optimal Assortments)**.** *Letting* $(S_i^*, \boldsymbol{w}_i^*)$ *be an optimal solution to the maximization problem on the right side of* (17)*, if* $l \in S_i^*$*, then we have* $\kappa_{il} - \eta_{il} \log w_{il}^* \geq (1-\gamma_i)\, R_i(S_i^*, \boldsymbol{w}_i^*) + \gamma_i\, z$*.*

Proof. To get a contradiction, assume that $\kappa_{il} - \eta_{il} \log w_{il}^* < (1-\gamma_i) R_i(S_i^*, \boldsymbol{w}_i^*) + \gamma_i\, z$. We let $\hat{S}_i$ be the assortment obtained by taking product l out of S_i^*. In this case, we get

$$\begin{aligned} W_i(\hat{S}_i, \boldsymbol{w}_i^*)^{\gamma_i} (R_i(\hat{S}_i, \boldsymbol{w}_i^*) - z) &= \frac{\sum_{j \in \hat{S}_i} w_{ij}^* \, (\kappa_{ij} - \eta_{ij} \log w_{ij}^* - z)}{W_i(\hat{S}_i, \boldsymbol{w}_i^*)^{1-\gamma_i}} \\ &= \frac{\sum_{j \in S_i^*} w_{ij}^* \, (\kappa_{ij} - \eta_{ij} \log w_{ij}^* - z) - w_{il}^* \, (\kappa_{il} - \eta_{il} \log w_{il}^* - z)}{W_i(\hat{S}_i, \boldsymbol{w}_i^*)^{1-\gamma_i}}, \end{aligned} \tag{25}$$

where the first equality uses the definitions of $R_i(S_i, \boldsymbol{w}_i)$ and $W_i(S_i, \boldsymbol{w}_i)$ and the second equality uses the fact that $S_i^* = \hat{S}_i \cup \{l\}$. To lower bound to numerator of the last fraction in (25), we note that $\kappa_{il} - \eta_{il} \log w_{il}^* - z < (1-\gamma_i)(R_i(S_i^*, \boldsymbol{w}_i^*) - z)$. In this case, we can lower bound the numerator of the last fraction in (25) by

$$\begin{aligned} &\sum_{j \in S_i^*} w_{ij}^* \, (\kappa_{ij} - \eta_{ij} \log w_{ij}^* - z) - (1-\gamma_i)\, w_{il}^* \, (R_i(S_i^*, \boldsymbol{w}_i^*) - z) \\ &\qquad\qquad = (R_i(S_i^*, \boldsymbol{w}_i^*) - z)\, (W_i(S_i^*, \boldsymbol{w}_i^*) - (1-\gamma_i)\, w_{il}^*) \geq 0, \end{aligned}$$

where the equality follows by using the definitions of $R_i(S_i, \boldsymbol{w}_i)$ and $W_i(S_i, \boldsymbol{w}_i)$ and rearranging the terms. To see that the inequality above holds, we observe that we must have $R_i(S_i^*, \boldsymbol{w}_i^*) - z \geq 0$, otherwise the optimal objective value of the maximization problem on the right side of (17) is negative and we can set $S_i^* = \emptyset$ to get a better objective value of zero. Furthermore, $W_i(S_i^*, \boldsymbol{w}_i^*) \geq w_{il}^*$ and $\gamma_i \leq 1$ so that $W_i(S_i^*, \boldsymbol{w}_i^*) - (1-\gamma_i)\, w_{il}^* \geq 0$, in which case, the inequality above indeed holds. We can upper bound the denominator of the last fraction in (25) by observing the fact that $W_i(\hat{S}_i, \boldsymbol{w}_i^*)^{1-\gamma_i} \leq W_i(S_i^*, \boldsymbol{w}_i^*)^{1-\gamma_i} + (1-\gamma_i)\, W_i(S_i^*, \boldsymbol{w}_i^*)^{-\gamma_i}\, (W_i(\hat{S}_i, \boldsymbol{w}_i^*) - W_i(S_i^*, \boldsymbol{w}_i^*))$, which follows by recalling $u^{1-\gamma_i}$ is a concave function of u and using the subgradient inequality. Therefore, we can lower bound the last fraction in (25) as

$$\begin{aligned} &\frac{\sum_{j \in S_i^*} w_{ij}^* \, (\kappa_{ij} - \eta_{ij} \log w_{ij}^* - z) - w_{il}^* \, (\kappa_{il} - \eta_{il} \log w_{il}^* - z)}{W_i(\hat{S}_i, \boldsymbol{w}_i^*)^{1-\gamma_i}} \\ &\qquad > \frac{(R_i(S_i^*, \boldsymbol{w}_i^*) - z)\, (W_i(S_i^*, \boldsymbol{w}_i^*) - (1-\gamma_i)\, w_{il}^*)}{W_i(S_i^*, \boldsymbol{w}_i^*)^{1-\gamma_i} + (1-\gamma_i)\, W_i(S_i^*, \boldsymbol{w}_i^*)^{-\gamma_i}\, (W_i(\hat{S}_i, \boldsymbol{w}_i^*) - W_i(S_i^*, \boldsymbol{w}_i^*))}. \end{aligned} \tag{26}$$

Noting that $W_i(\hat{S}_i, \boldsymbol{w}_i^*) - W_i(S_i^*, \boldsymbol{w}_i^*) = -w_{il}^*$ and rearranging the terms in the last fraction, we observe that the last fraction is equal to $W_i(S_i^*, \boldsymbol{w}_i^*)^{\gamma_i} (R_i(S_i^*, \boldsymbol{w}_i^*) - z)$. Thus, (25) and (26) show that $W_i(\hat{S}_i, \boldsymbol{w}_i^*)^{\gamma_i} (R_i(\hat{S}_i, \boldsymbol{w}_i^*) - z) > W_i(S_i^*, \boldsymbol{w}_i^*)^{\gamma_i} (R_i(S_i^*, \boldsymbol{w}_i^*) - z)$, contradicting the fact that $(S_i^*, \boldsymbol{w}_i^*)$ is an optimal solution to the maximization problem on the right side of (17). So, our claim holds and we must have $\kappa_{il} - \eta_{il} \log w_{il}^* \geq (1-\gamma_i) R_i(S_i^*, \boldsymbol{w}_i^*) + \gamma_i\, z$. □

A.3 Theorem 9

Theorem 9 (Upper Bound)**.** *Letting Z^* and $\hat{z}$ respectively be the optimal objective values of problems* (2) *and* (24)*, we have $\hat{z} \geq Z^*$. Furthermore, for some $\rho \geq 0$, if the grid points in problem* (24) *satisfy $g_i(\bar{y}_i^{t+1}) \leq (1+\rho)\, g_i(\bar{y}_i^t)$ for all $t = 1, \ldots, \tau_i - 1$, $i \in M$, then we have $(1+\rho)\, Z^* \geq \hat{z}$.*

Proof. First, we show that $\hat{z} \geq Z^*$. To get a contradiction, assume that $\hat{z} < Z^*$. Noting that $\hat{z}$ is the optimal objective value of problem (24), we use $\hat{z}$ and $(\hat{x}_1, \ldots, \hat{x}_m)$ to denote an optimal solution to this problem. By Proposition 1, Z^* corresponds to the value of z that satisfies (6), in which case, if we let x_i^* be the optimal objective value of the maximization problem on the right side of (6) when this problem is solved with $z = Z^*$, then we get $Z^* = \sum_{i \in M} x_i^*$. For all $i \in M$, we let y_i^* be an optimal solution to the maximization problem on the right side of (6) when this problem is solved with $z = Z^*$. We let $t_i \in \{1, \ldots, \tau_i - 1\}$ be such that $y_i^* \in [\bar{y}_t^{t_i}, \bar{y}_i^{t_i+1}]$, where the points $\{\bar{y}_i^t : t = 1, \ldots, \tau_i\}$ are the collection of grid points in problem (24). Since $g_i(\cdot)$ is increasing and $y_i^* \leq \bar{y}_i^{t_i+1}$, we have $g_i(y_i^*) \leq g_i(\bar{y}_i^{t_i+1})$. Also, since $\gamma_i \leq 1$ and $y_i^* \geq \bar{y}_i^{t_i}$, we have $(y_i^*)^{1-\gamma_i} \geq (\bar{y}_i^{t_i})^{1-\gamma_i}$. In this case, using the last two observations, we obtain

$$\hat{x}_i \geq (\bar{y}_i^{t_i})^{\gamma_i} \frac{g_i(\bar{y}_i^{t_i+1})}{\bar{y}_i^{t_i}} - (\bar{y}_i^{t_i})^{\gamma_i} \hat{z} \geq (y_i^*)^{\gamma_i} \frac{g_i(y_i^*)}{y_i^*} - (y_i^*)^{\gamma_i}\, \hat{z} \geq (y_i^*)^{\gamma_i} \frac{g_i(y_i^*)}{y_i^*} - (y_i^*)^{\gamma_i}\, Z^* = x_i^*,$$

where the first inequality is by the fact that $\hat{z}$ and $(\hat{x}_1, \ldots, \hat{x}_m)$ form a feasible solution to problem (24), the third inequality follows from the fact that $\hat{z} < Z^*$ and the equality follows from the definitions of x_i^* and y_i^*. Since $\hat{z}$ and $(\hat{x}_1, \ldots, \hat{x}_m)$ form a feasible solution to problem (24), we have $\hat{z} \geq \sum_{i \in M} \hat{x}_i$, whereas we have $Z^* = \sum_{i \in M} x_i^*$ by the discussion at the beginning of the proof. In this case, adding the chain of inequalities above over all $i \in M$, we obtain $\hat{z} \geq \sum_{i \in M} \hat{x}_i \geq \sum_{i \in M} x_i^* = Z^*$, which contradicts the fact that $\hat{z} < Z^*$. Therefore, we must have $\hat{z} \geq Z^*$.

Second, we show that $(1+\rho)\, Z^* \geq \hat{z}$. If we let x_i^* be as defined in the paragraph above, then we can follow the same line of reasoning that we follow above to see that $Z^* = \sum_{i \in M} x_i^*$. For any $t = 1, \ldots, \tau_i - 1$, we note that $\bar{y}_i^t$ is a feasible but not necessarily an optimal solution to the maximization problem on the right side of (6) when this problem is solved with $z = Z^*$. Therefore, it follows that $x_i^* \geq (\bar{y}_i^t)^{\gamma_i}\, g_i(\bar{y}_i^t)/\bar{y}_i^t - (\bar{y}_i^t)^{\gamma_i}\, Z^*$ for all $t = 1, \ldots, \tau_i - 1$. If we multiply the last inequality by $1 + \rho$, then we obtain

$$(1+\rho)\, x_i^* \geq (\bar{y}_i^t)^{\gamma_i} \frac{(1+\rho)\, g_i(\bar{y}_i^t)}{\bar{y}_i^t} - (\bar{y}_i^t)^{\gamma_i}\, (1+\rho)\, Z^* \geq (\bar{y}_i^t)^{\gamma_i} \frac{g_i(\bar{y}_i^{t+1})}{\bar{y}_i^t} - (\bar{y}_i^t)^{\gamma_i}\, (1+\rho)\, Z^*,$$

where the second inequality follows by noting the fact that $g_i(\bar{y}_i^{t+1}) \leq (1+\rho)\, g_i(\bar{y}_i^t)$ for all $t = 1, \ldots, \tau_i - 1$. Focusing on the first and last expressions in the chain of inequalities above and noting that $Z^* = \sum_{i \in M} x_i^*$, the solution $(1+\rho)\, Z^*$ and $((1+\rho)\, x_1^*, \ldots, (1+\rho)\, x_m^*)$ is feasible to problem (24). Thus, the objective value provided by this solution for problem (24) is at least as large as the optimal objective value. Since the solution $(1+\rho)\, Z^*$ and $((1+\rho)\, x_1^*, \ldots, (1+\rho)\, x_m^*)$ provides an objective value of $(1+\rho)\, Z^*$ for problem (24), we get $(1+\rho)\, Z^* \geq \hat{z}$. □

B Appendix: Optimal Assortments with Cardinality Constraints

In this section, we show that if there is a constraint on the number of products that we can offer in each nest, then the optimal value of the decision variable S_i in the maximization problem on the right side of (17) is one of $O(n^2)$ possible values, irrespective of the value of z. To show this result, we let $R_i(S_i, \boldsymbol{w}_i) = \sum_{j \in S_i} w_{ij} (\kappa_{ij} - \eta_{ij} \log w_{ij}) / \sum_{j \in S_i} w_{ij}$ and $W_i(S_i, \boldsymbol{w}_i) = \sum_{j \in S_i} w_{ij}$ for notational brevity. In this case, assuming that we can offer at most c_i products in nest i, we write the maximization problem on the right side of (17) as

$$\max_{\substack{S_i \subset N, \boldsymbol{w}_i \in [\boldsymbol{L}_i, \boldsymbol{U}_i] \\ |S_i| \leq c_i}} \Big\{ W_i(S_i, \boldsymbol{w}_i)^{\gamma_i} \left(R_i(S_i, \boldsymbol{w}_i) - z \right) \Big\}. \tag{27}$$

Our goal is to show that the optimal value of the decision variable S_i in the problem above is one of $O(n^2)$ possible values, irrespective of the value of z. In the next lemma, we give an alternative characterization of the optimal solution to the problem above. The proof of this lemma uses an argument that closely follows the proof of Lemma 3 in Gallego and Topaloglu (2012). For completeness, we also provide a proof but defer it to the end of this section.

Lemma 10 (Optimal Assortments)**.** *As a function of z, we use $(S_i^*(z), \boldsymbol{w}_i^*(z))$ to denote an optimal solution to problem* (27) *and let $u_i^*(z) = \max\{\gamma_i\, z + (1 - \gamma_i)\, R_i(S_i^*(z), \boldsymbol{w}_i^*(z)), z\}$. In this case, if $(\hat{S}_i, \hat{\boldsymbol{w}}_i)$ is an optimal solution to the problem*

$$\max_{\substack{S_i \subset N, \boldsymbol{w}_i \in [\boldsymbol{L}_i, \boldsymbol{U}_i] \\ |S_i| \leq c_i}} \Big\{ W_i(S_i, \boldsymbol{w}_i) \left(R_i(S_i, \boldsymbol{w}_i) - u_i^*(z) \right) \Big\}, \tag{28}$$

then $(\hat{S}_i, \hat{\boldsymbol{w}}_i)$ is also an optimal solution to problem (27)*.*

The lemma above shows that we can obtain an optimal solution to problem (27) alternatively by solving problem (28). To characterize an optimal solution to problem (27), as a function of u, we use $(\hat{S}_i(u), \hat{\boldsymbol{w}}_i(u))$ to denote an optimal solution to the problem

$$\max_{\substack{S_i \subset N, \boldsymbol{w}_i \in [\boldsymbol{L}_i, \boldsymbol{U}_i] \\ |S_i| \leq c_i}} \Big\{ W_i(S_i, \boldsymbol{w}_i) \left(R_i(S_i, \boldsymbol{w}_i) - u \right) \Big\}. \tag{29}$$

In this case, the crucial observation is that the set of solutions $\{(\hat{S}_i(u), \hat{\boldsymbol{w}}_i(u)) : u \in [0, \infty)\}$ always includes an optimal solution to problem (27), irrespective of the value of z. To see this result, we note that problem (29) with $u = u_i^*(z)$ is identical to problem (28). Thus, the solution $(\hat{S}_i(u_i^*(z)), \hat{\boldsymbol{w}}_i(u_i^*(z))$ is an optimal solution to problem (28), but if $(\hat{S}_i(u_i^*(z)), \hat{\boldsymbol{w}}_i(u_i^*(z)))$ is an optimal solution to problem (28), then Lemma 10 implies that $(\hat{S}_i(u_i^*(z)), \hat{\boldsymbol{w}}_i(u_i^*(z)))$ is an optimal solution to problem (27) as well. Since $(\hat{S}_i(u_i^*(z)), \hat{\boldsymbol{w}}_i(u_i^*(z))) \in \{(\hat{S}_i(u), \hat{\boldsymbol{w}}_i(u)) : u \in [0, \infty)\}$, it follows that the set of solutions $\{(\hat{S}_i(u), \hat{\boldsymbol{w}}_i(u)) : u \in [0, \infty)\}$ includes an optimal solution to problem (27), irrespective of the value of z. The preceding discussion shows that the set $\{\hat{S}_i(u) : u \in [0, \infty)\}$ includes an optimal value of the decision variable S_i in problem (27), irrespective of the value of z. Using this observation, we obtain the next lemma.

Lemma 11 (Generation of Optimal Assortments)**.** *For any $u \geq 0$, assume that the set $\{S_i^q : q = 1, \ldots, Q_i\}$ includes an optimal value of the decision variable S_i in problem* (29)*. In this case, for any $z \geq 0$, the set $\{S_i^q : q = 1, \ldots, Q_i\}$ also includes an optimal value of the decision variable S_i in problem* (27)*.*

Proof. We fix an arbitrary value of $z \geq 0$. For the fixed value of z, we let $(S_i^*(z), \boldsymbol{w}_i^*(z))$ be an optimal solution to problem (27) and $u_i^*(z) = \max\{\gamma_i z + (1 - \gamma_i) R_i(S_i^*(z), \boldsymbol{w}_i^*(z)), z\}$. We consider solving problem (29) with $u = u_i^*(z)$. Since $\{S_i^q : q = 1, \ldots, Q_i\}$ includes an optimal value of the decision variable S_i in problem (29) when this problem is solved with any $u \geq 0$, there exists an optimal solution $(\hat{S}_i, \hat{\boldsymbol{w}}_i)$ to problem (28) such that $\hat{S}_i \in \{S_i^q : q = 1, \ldots, Q_i\}$. In this case, Lemma 10 implies that $(\hat{S}_i, \hat{\boldsymbol{w}}_i)$ is also an optimal solution to problem (27). Noting that $\hat{S}_i \in \{S_i^q : q = 1, \ldots, Q_i\}$ and the choice of z is arbitrary, it follows that for any $z \geq 0$, the set $\{S_i^q : q = 1, \ldots, Q_i\}$ includes an optimal value of the decision variable S_i in problem (27). $\square$

By the lemma above, if we can come up with a set $\{S_i^q : q = 1, \ldots, Q_i\}$ that includes an optimal value of the decision variable S_i in problem (29) for any $u \geq 0$, then this set also includes an optimal value of the decision variable S_i in problem (27) for any $z \geq 0$. Furthermore, if $Q_i = O(n^2)$, then the optimal value of the decision variable S_i in problem (27) is always one of $O(n^2)$ possible values, as desired. To come up with a set $\{S_i^q : q = 1, \ldots, Q_i\}$ that includes an optimal value of the decision variable S_i in problem (29) for any $u \geq 0$, we use the definitions of $R_i(S_i, \boldsymbol{w}_i)$ and $W_i(S_i, \boldsymbol{w}_i)$ to write this problem as

$$\max_{\substack{S_i \subset N,\, \boldsymbol{w}_i \in [\boldsymbol{L}_i, \boldsymbol{U}_i] \\ |S_i| \leq c_i}} \left\{ \sum_{j \in S_i} w_{ij} \left(\kappa_{ij} - \eta_{ij} \log w_{ij} - u \right) \right\}. \tag{30}$$

Using the first order condition for the objective function of the problem above, we can check that the optimal value of the decision variable w_{ij} is given by $\min\{\max\{\exp((\kappa_{ij} - u)/\eta_{ij} - 1), L_{ij}\}, U_{ij}\}$, which implies that this decision variable takes value L_{ij} in the optimal solution to the problem above when $u \geq \kappa_{ij} - \eta_{ij} - \eta_{ij} \log L_{ij}$, takes value U_{ij} when $u \leq \kappa_{ij} - \eta_{ij} - \eta_{ij} \log U_{ij}$ and takes value $\exp((\kappa_{ij} - u)/\eta_{ij} - 1)$ when $\kappa_{ij} - \eta_{ij} - \eta_{ij} \log U_{ij} \leq u \leq \kappa_{ij} - \eta_{ij} - \eta_{ij} \log L_{ij}$. Thus, letting $\tilde{L}_{ij} = \kappa_{ij} - \eta_{ij} - \eta_{ij} \log U_{ij}$ and $\tilde{U}_{ij} = \kappa_{ij} - \eta_{ij} - \eta_{ij} \log L_{ij}$, the decision variable w_{ij} respectively takes values L_{ij}, U_{ij} and $\exp((\kappa_{ij} - u)/\eta_{ij} - 1)$ when $u \geq \tilde{U}_{ij}$, $u \leq \tilde{L}_{ij}$ and $\tilde{L}_{ij} \leq u \leq \tilde{U}_{ij}$.

Thus, noting the objective function of problem (30) and the discussion in the paragraph above, if we have $u \in [\tilde{U}_{ij}, \infty)$, then the decision variable w_{ij} makes an objective function contribution of $L_{ij} (\kappa_{ij} - \eta_{ij} \log L_{ij} - u)$ in the optimal solution to problem (30). Similarly, if we have $u \in [0, \tilde{L}_{ij}]$, then the decision variable w_{ij} makes an objective function contribution of $U_{ij} (\kappa_{ij} - \eta_{ij} \log U_{ij} - u)$. Finally, if we have $u \in [\tilde{L}_{ij}, \tilde{U}_{ij}]$, then the decision variable w_{ij} makes an objective function contribution of $\exp((\kappa_{ij} - u)/\eta_{ij} - 1) (\kappa_{ij} - \eta_{ij} \log(\exp((\kappa_{ij} - u)/\eta_{ij} - 1)) - u)$, which can be simplified to $\eta_{ij} \exp((\kappa_{ij} - u)/\eta_{ij} - 1)$ through a straightforward algebraic manipulation. In this case, to capture the objective function contribution of the decision variable w_{ij} in the optimal solution to problem (30), we define the function $f_{ij}(\cdot)$ as follows. If

$u \in [\tilde{U}_{ij}, \infty)$, then we have $f_{ij}(u) = L_{ij}\,(\kappa_{ij} - \eta_{ij}\log L_{ij} - u)$, whereas if $u \in [0, \tilde{L}_{ij}]$, then we have $f_{ij}(u) = U_{ij}\,(\kappa_{ij} - \eta_{ij}\log U_{ij} - u)$. Finally, if $u \in [\tilde{L}_{ij}, \tilde{U}_{ij}]$, then we have $f_{ij}(u) = \eta_{ij}\exp((\kappa_{ij} - u)/\eta_{ij} - 1)$. We observe that $f_{ij}(\cdot)$ is a linear function over the intervals $[0, \tilde{L}_{ij}]$ and $[\tilde{U}_{ij}, \infty)$, whereas it is an exponential function over the interval $[\tilde{L}_{ij}, \tilde{U}_{ij}]$. Replacing $w_{ij}\,(\kappa_{ij} - \eta_{ij}\log w_{ij} - u)$ in problem (30) with the objective function contribution of w_{ij} in the optimal solution, problem (30) is equivalent to the problem

$$\max_{S_i \subset N,\, |S_i| \leq c_i} \left\{ \sum_{j \in S_i} f_{ij}(u) \right\}. \tag{31}$$

We can solve the problem above by ordering the values of $\{f_{ij}(u) : j \in N\}$ in decreasing order and choosing c_i products with the largest corresponding values of $\{f_{ij}(u) : j \in N\}$, as long as these values are positive. Thus, letting $f_{i0}(u) = 0$ for notational brevity, the optimal value of the decision variable S_i in the problem above depends only on the ordering between the values of $\{f_{ij}(u) : j \in N \cup \{0\}\}$. We proceed to showing that as u takes values over $[0, \infty)$, we obtain only $O(n^2)$ different orderings between the values of $\{f_{ij}(u) : j \in N \cup \{0\}\}$.

Consider two functions $f_{ij}(\cdot)$ and $f_{ik}(\cdot)$ for $j, k \in N \cup \{0\}$. If we plot $f_{ij}(u)$ and $f_{ik}(u)$ as a function of u on the two dimensional plane, then a simple counting argument shows that these two functions can have at most seven intersection points. We shortly give the details of this counting argument. Therefore, the functions $\{f_{ij}(\cdot) : j \in N \cup \{0\}\}$ have at most $7n(n+1)/2$ intersection points. The crucial observation is that the intersection points of the functions $\{f_{ij}(\cdot) : j \in N\}$ computed in this fashion correspond to the only values of u where the ordering between the values of $\{f_{ij}(u) : j \in N \cup \{0\}\}$ changes. Therefore, there are at most $7n(n+1)/2 = O(n^2)$ different orderings between the values of $\{f_{ij}(u) : j \in N \cup \{0\}\}$ as u takes values over $[0, \infty)$. In Figure 2, we show a possible case with $n = 3$. The bold lines show the functions $\{f_{ij}(\cdot) : j \in N \cup \{0\}\}$ and the white circles show the intersection points of these functions. Each one of the intervals between the intersection points is associated with a particular ordering. As long as u takes values between two consecutive intersection points, the ordering between the values of $\{f_{ij}(u) : j \in N\}$ does not change. In Figure 2, for example, for any $u \in [u_a, u_b]$, we have $f_{i2}(u) \geq f_{i1}(u) \geq f_{i3}(u) \geq f_{i0}(u)$, whereas for any $u \in [u_b, u_c]$, we have $f_{i2}(u) \geq f_{i3}(u) \geq f_{i1}(u) \geq f_{i0}(u)$.

The discussion in the paragraph above shows that as u takes values over $[0, \infty)$, there are $O(n^2)$ possible orderings between the values of $\{f_{ij}(u) : j \in N \cup \{0\}\}$. Since the optimal value of the decision variable in problem (31) depends only on the ordering between the values of $\{f_{ij}(u) : j \in N \cup \{0\}\}$, as u takes values over $[0, \infty)$, there are $O(n^2)$ possible values for the decision variable S_i in an optimal solution to problem (31). Noting that problem (31) is equivalent to problem (29), as u takes values over $[0, \infty)$, there are $O(n^2)$ possible values for the decision variable S_i in an optimal solution to problem (29). In other words, there exists a set $\{S_i^q : q = 1, \ldots, Q_i\}$ with $Q_i = O(n^2)$ such that for any $u \geq 0$, this set includes an optimal value of the decision variable S_i in problem (29). In this case, Lemma 11 implies that the set $\{S_i^q : q = 1, \ldots, Q_i\}$ with $Q_i = O(n^2)$ includes an optimal solution to problem (27) for any $z \geq 0$, establishing the desired result. In the

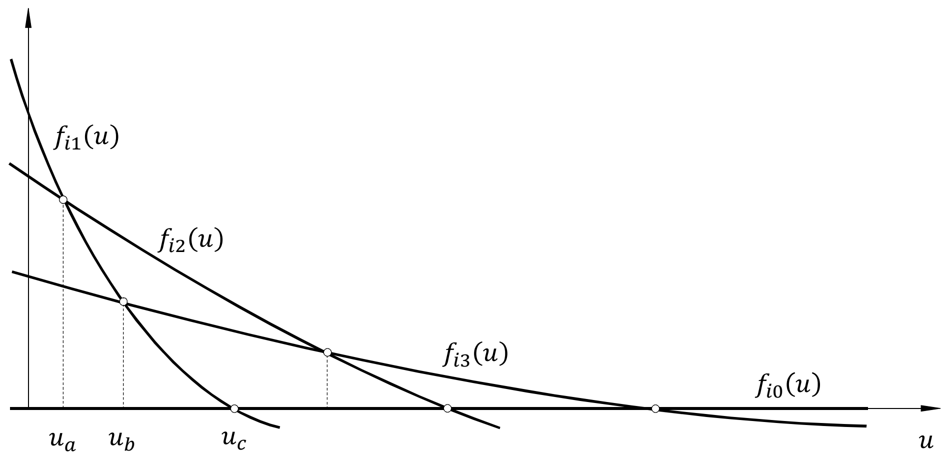

Figure 2: The functions $\{f_{ij}(\cdot) : j \in N \cup \{0\}\}$ for a possible case with $n = 3$. For clarity, any two of the functions $\{f_{ij}(\cdot) : j \in N \cup \{0\}\}$ in the figure have one point of intersection, but they can have more than one point of intersection in general.

remainder of this section, we show that two functions $f_{ij}(\cdot)$ and $f_{ik}(\cdot)$ for $j, k \in N \cup \{0\}$ can have at most seven intersection points and provide a proof for Lemma 10.

B.1 Intersection Points

To see that two functions $f_{ij}(\cdot)$ and $f_{ik}(\cdot)$ can have at most seven intersection points, we let $\tilde{L}_{ij}$ and $\tilde{U}_{ij}$ be such that the function $f_{ij}(\cdot)$ is linear outside the interval $[\tilde{L}_{ij}, \tilde{U}_{ij}]$ and exponential in the interval $[\tilde{L}_{ij}, \tilde{U}_{ij}]$. We note that if $j = 0$, then we can set $\tilde{L}_{ij} = \tilde{U}_{ij} = 0$ so that $f_{i0}(u) = 0$ is a linear function of u for all $u \in [0, \infty)$. Similarly, we let $\tilde{L}_{ik}$ and $\tilde{U}_{ik}$ be such that the function $f_{ik}(\cdot)$ is linear outside the interval $[\tilde{L}_{ik}, \tilde{U}_{ik}]$ and exponential in the interval $[\tilde{L}_{ik}, \tilde{U}_{ik}]$. If $k = 0$, then we can set $\tilde{L}_{ik} = \tilde{U}_{ik} = 0$. For brevity, we only consider the case where the intervals $[\tilde{L}_{ij}, \tilde{U}_{ij}]$ and $[\tilde{L}_{ik}, \tilde{U}_{ik}]$ satisfy $\tilde{L}_{ij} \leq \tilde{L}_{ik} \leq \tilde{U}_{ij} \leq \tilde{U}_{ik}$. The analyses of the other cases are identical. Over the interval $[0, \tilde{L}_{ij}]$, both of the functions $f_{ij}(\cdot)$ and $f_{ik}(\cdot)$ are linear. Two linear functions can have at most one intersection point. Over the interval $[\tilde{L}_{ij}, \tilde{L}_{ik}]$, the function $f_{ij}(\cdot)$ is exponential and the function $f_{ik}(\cdot)$ is linear. An exponential and a linear function can have at most two intersection points. Over the interval $[\tilde{L}_{ik}, \tilde{U}_{ij}]$, both of the functions $f_{ij}(\cdot)$ and $f_{ik}(\cdot)$ are exponential. These two exponential functions can have at most one intersection point given by $\frac{1}{1/\eta_{ij} - 1/\eta_{ik}} (\log \eta_{ij} + \kappa_{ij}/\eta_{ij} - log\, \eta_{ik} - \kappa_{ik}/\eta_{ik})$. Using a similar argument, we can check that the functions $f_{ij}(\cdot)$ and $f_{ik}(\cdot)$ can have at most two intersection points over the interval $[\tilde{U}_{ij}, \tilde{U}_{ik}]$ and at most one intersection point over the interval $[\tilde{U}_{ik}, \infty)$. Thus, we obtain a total of seven intersection points. We note that the analysis in this paragraph is not necessarily tight. In particular, the functions $f_{ij}(\cdot)$ and $f_{ik}(\cdot)$ may intersect at fewer than seven points and it may be possible to eliminate some of the potential intersection points further. Nevertheless, a constant upper bound of seven on the number of intersection points is sufficient to obtain the desired result.

B.2 Proof of Lemma 10

In this section, we give a proof for Lemma 10. We fix an arbitrary value of $z \geq 0$. For the fixed value of z, we let $(S_i^*, \boldsymbol{w}_i^*) = (S_i^*(z), \boldsymbol{w}_i(z))$, $R_i^* = R_i(S_i^*, \boldsymbol{w}_i^*)$, $W_i^* = W_i(S_i^*, \boldsymbol{w}_i^*)$, $\hat{R}_i = R_i(\hat{S}_i, \hat{\boldsymbol{w}}_i)$ and $\hat{W}_i = W_i(\hat{S}_i, \hat{\boldsymbol{w}}_i)$ for notational brevity. We show that $\hat{W}_i^{\gamma_i}(\hat{R}_i - z) \geq (W_i^*)^{\gamma_i}(R_i^* - z)$, establishing that the solution $(\hat{S}_i, \hat{\boldsymbol{w}}_i)$ provides an objective value for problem (27) that is at least as large as the objective value provided by the optimal solution. Therefore, $(\hat{S}_i, \hat{\boldsymbol{w}}_i)$ is an optimal solution to problem (27) and the desired result follows. We consider two cases. First, we assume that $z \geq R_i^*$. In this case, we have $u_i^*(z) = z$ by the definition of $u_i^*(z)$. Since $(\hat{S}_i, \hat{\boldsymbol{w}}_i)$ is an optimal solution to problem (28), we observe that $\hat{W}_i(\hat{R}_i - u_i^*(z)) \geq 0$. Otherwise, we can set $S_i = \emptyset$ and obtain a better objective value of zero for problem (28). Since $z \geq R_i^*$, we obtain $(\hat{W}_i)^{\gamma_i}(\hat{R}_i - z) = (\hat{W}_i)^{\gamma_i}(\hat{R}_i - u_i^*(z)) \geq 0 \geq (W_i^*)^{\gamma_i}(R_i^* - z)$, as desired.

Second, we assume that $z < R_i^*$. In this case, we have $u_i^*(z) = \gamma_i z + (1-\gamma_i) R_i^*$ by the definition of $u_i^*(z)$. We observe that $W_i^* = W_i(S_i^*, \boldsymbol{w}_i^*) > 0$. Otherwise, we have either $w_{ij}^* = 0$ for all $j \in S_i^*$ or $S_i^* = \emptyset$, both of which imply that $R_i^* = R_i(S_i^*, \boldsymbol{w}_i^*) = 0$, contradicting the fact that $R_i^* > z \geq 0$. On the other hand, $(\hat{S}_i, \hat{\boldsymbol{w}}_i)$ is an optimal solution to problem (28), whereas $(S_i^*, \boldsymbol{w}_i^*)$ is a feasible but not necessarily an optimal solution to this problem. Therefore, it follows that $\hat{W}_i(\hat{R}_i - u_i^*(z)) \geq W_i^*(R_i^* - u_i^*(z))$. Using the definition of $u_i^*(z)$, we write this inequality as $\hat{W}_i(\hat{R}_i - z) - (1-\gamma_i)\hat{W}_i(R_i^* - z) \geq \gamma_i W_i^*(R_i^* - z)$ and rearranging the terms we obtain

$$\hat{W}_i(\hat{R}_i - z) \geq (\gamma_i W_i^* + (1-\gamma_i)\hat{W}_i)(R_i^* - z).$$

Since $R_i^* > z$ and $W_i^* > 0$, the right side of the inequality above is strictly positive, which implies that $\hat{W}_i > 0$. Since $\gamma_i \leq 1$, x^{γ_i} is a concave function of x satisfying the subgradient inequality $(x^*)^{\gamma_i} \leq \hat{x}^{\gamma_i} + \gamma_i \hat{x}^{\gamma_i - 1}(x^* - \hat{x}) = \hat{x}^{\gamma_i - 1}(\gamma_i x^* + (1-\gamma_i)\hat{x})$ for all $x^*, \hat{x} > 0$. Using this inequality with $x^* = W_i^*$ and $\hat{x} = \hat{W}_i$, we obtain $(W_i^*)^{\gamma_i} \leq \hat{W}_i^{\gamma_i - 1}(\gamma_i W_i^* + (1-\gamma_i)\hat{W}_i)$. Thus, we can use $(W_i^*)^{\gamma_i}/\hat{W}_i^{\gamma_i - 1}$ to lower bound the expression $\gamma_i W_i^* + (1-\gamma_i)\hat{W}_i$ on the right side of the inequality above, in which case, we obtain $\hat{W}_i(\hat{R}_i - z) \geq (W_i^*)^{\gamma_i}(R_i^* - z)/\hat{W}_i^{\gamma_i - 1}$. Rearranging the terms in the last inequality, we have $\hat{W}_i^{\gamma_i}(\hat{R}_i - z) \geq (W_i^*)^{\gamma_i}(R_i^* - z)$, as desired.

C Appendix: Glossary of Notation

In this section, we provide a list of notation used throughout the paper. We accompany each piece of notation with its description and the place where it is introduced in the paper.

l_{ij}, u_{ij}	Lower and upper bounds on the price of product j in nest i.	Section 2
α_{ij}, β_{ij}	Parameters connecting the price of product j in nest i to the preference weight of this product. Using p_{ij} and w_{ij} to respectively denote the price and preference weight of product j in nest i, we have $w_{ij} = \exp(\alpha_{ij} - \beta_{ij}\, p_{ij})$.	Section 2
L_{ij}, U_{ij}	Lower and upper bounds on the preference weight of product j in nest i. We have $L_{ij} = \exp(\alpha_{ij} - \beta_{ij}\, u_{ij})$ and $U_{ij} = \exp(\alpha_{ij} - \beta_{ij}\, l_{ij})$.	Section 2

κ_{ij}, η_{ij}	Parameters connecting the preference weight of product j in nest i to the price of this product. Using w_{ij} and p_{ij} to respectively denote the preference weight and price of product j in nest i, we have $p_{ij} = \kappa_{ij} - \eta_{ij} \log w_{ij}$.	Section 2
$\boldsymbol{L}_i, \boldsymbol{U}_i$	Vectors $\boldsymbol{L}_i = (L_{i1}, \ldots, L_{in})$ and $\boldsymbol{U}_i = (U_{i1}, \ldots, U_{in})$ of lower and upper bounds on the preference weights of the products in nest i.	Section 2
Z^*	Optimal objective value of problem (2).	Section 2
$g_i(y_i)$	Optimal objective value of problem (5) as a function of the right side of the first constraint in this problem.	Section 3
$\bar{L}_i, \bar{U}_i$	Smallest and largest values of the right side of the first constraint in problem (5) that ensure that the first constraint is tight at the optimal solution.	Section 3
$\tilde{y}_i^t$	Each one of the grid points $\{\tilde{y}_i^t : t = 1, \ldots, T_i\}$ that we use for nest i in problem (7).	Section 4
$\lambda_i^*(y_i)$	Optimal value of the Lagrange multiplier associated with the first constraint in problem (5) as a function of the right side of this constraint.	Section 5.1
ζ_{ij}, ξ_{ij}	Two threshold values of the right side of the first constraint in problem (5) such that if $y_i \leq \zeta_{ij}$, then the optimal value of the decision variable w_{ij} in this problem is L_{ij}, whereas if $y_i \geq \xi_{ij}$, then the optimal value of the decision variable w_{ij} is U_{ij}.	Section 5.1
$[\nu_i^k, \nu_i^{k+1}]$	Set of possible values for the right side of the first constraint in problem (5) such that if $y_i \in [\nu_i^k, \nu_i^{k+1}]$, then we can fix the values of a certain set of decision variables at their upper or lower bounds and not impose the upper and lower bound constraints at all on the remaining decision variables.	Section 5.1
K_i	Largest index for the set of intervals $\{[\nu_i^k, \nu_i^{k+1}] : k = 1, \ldots, K_i\}$.	Section 5.1
$\mathcal{L}_i^k, \mathcal{U}_i^k$	Two sets of decision variables in problem (5) such that if $y_i \in [\nu_i^k, \nu_i^{k+1}]$, then we can fix the values of the decision variables in $\mathcal{L}_i^k$ at their lower bounds and the values of the decision variables in $\mathcal{U}_i^k$ at their upper bounds.	Section 5.1
$\tilde{Y}_i^{kq}$	A grid point that we obtain by using (14).	Section 5.2
q_i^L, q_i^U	Smallest and largest indices for the set of grid points $\{\tilde{Y}_i^{kq} : q = q_i^L, \ldots, q_i^U\}$.	Section 5.2
ζ^*	Optimal objective value of problem (16).	Section 6
$g_i(S_i, y_i)$	Optimal objective value of problem (20) as a function of the products included in this problem and the right side of the first constraint.	Section 6.1
$\bar{L}_i(S_i)$, $\bar{U}_i(S_i)$	Smallest and largest values of the right side of the first constraint in problem (20) that ensure that the first constraint is tight at the optimal solution.	Section 6.1
$\tilde{y}_i^t(S_i)$	Each one of the grid points $\{\tilde{y}_i^t(S_i) : t = 1, \ldots, T_i(S_i)\}$ that we use for nest i in problem (22).	Section 6.2