PRICING AND ASSORTMENT PROBLEMS UNDER CORRELATED PRODUCT EVALUATIONS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

William Z. Rayfield

August 2014

PRICING AND ASSORTMENT PROBLEMS UNDER CORRELATED PRODUCT EVALUATIONS

William Z. Rayfield, Ph.D.
Cornell University 2014

In many retail settings, 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. This thesis focuses on settings where a retailer offers heterogeneous products that can be designated into a number of different subgroups such that product evaluations, i.e., the utilities that a customer assigns to products within a subgroup, are correlated. We focus on positive correlation among product evaluations within subgroups, meaning that products within a subgroup are closer substitutes to each other than products from another subgroup. We use the Nested Logit (NL) model as a tool to capture correlated product evaluations. In the Nested Logit model, a customer follows a two-stage choice process where she first selects a product subgroup, followed by a specific product within that group.

The first part of the thesis focuses on a pricing problem where a firm offers a product for sales through several distinct sales channels and the utility derived from purchasing the product in each channel depends on its price. In each sales channel, there are also competing products not under our firm's control which must be taken into consideration. We provide sequential methods for obtaining maximizers of the revenue function and provide structural properties pertaining to the markups on optimal prices.

The second part of the thesis focuses on pricing problems in which the prices

of offered products are subject to bound constraints and the utility derived from each product depends on its price. We give approximation methods that allow the user to specify a performance guarantee a priori, where the final solutions are obtained by solving linear programs whose sizes scale gracefully with the specified performance guarantee. In addition, we develop a linear program that we can use to quickly obtain upper bounds on optimal expected revenues.

The final part of the thesis focuses on an assortment offering problem in which operational costs are incorporated in the form of inventory considerations. We provide theoretical and numerical results concerning the effect of correlation between product evaluations within a subgroup on optimal assortment sizes, optimal profits and optimal stocking levels. We give structural properties of optimal assortments within each subgroup, and provide efficient dynamic-programming based solution approaches for finding near-optimal assortments.

BIOGRAPHICAL SKETCH

Zach was born in Columbia, Maryland and grew up in nearby Ellicott City, Maryland. Zach has been involved in many academic pursuits since the time he entered pre-school, but always found plenty of time for extra-curricular activities, including playing soccer and basketball, obsessively reading car magazines and Calvin and Hobbes comic strips, and getting up extremely early on Sunday mornings to watch Formula One racing. Zach was heavily involved in music during his school-age years, studying piano performance and music composition at the Peabody Preparatory of the Johns Hopkins University.

After high school, Zach moved 20 minutes away from home to attend the University of Maryland Baltimore County, where he obtained his Bachelor's degree in Mathematics. During his time at UMBC he continued to study piano and write music. He also spent massive amounts of time playing games in the Halo and Call of Duty series, spent a disproportionate amount of the money he earned as a teaching assistant at Chipotle, and collected more music than it may be healthy to mention in these pages.

Upon his graduation from UMBC, Zach had no immediate desire to enter the workforce, but decided that it would be a good idea to study something that he could potentially be able to explain to his confused family members in English sentences, and joined the School of Operations Research and Information Engineering at Cornell University as a PhD student. Life in Ithaca is a graduate student was filled with many incredible experiences and Zach had the opportunity to cross paths with some very memorable people.

After graduation, Zach will be realizing a dream he had 20 years ago by taking a one-way trip to Disney World. He will be working in Disney's Revenue Management and Analytics division as a consultant.

This document is dedicated to my sister Lael.

ACKNOWLEDGEMENTS

First, I would like to thank my advisor Prof. Huseyin Topaloglu for his time, patience, advice and his constant willingness to help. I would also like to thank my committee members, Prof. Shane Henderson and Prof. Michael Todd, for the support and interest they have shown from the time I began my studies at Cornell. Thanks also to Prof. Mark Lewis for his support and guidance.

Thank you to my parents for their unwavering support and belief in me. I owe much of my success to you. Thank you to Sarah for her patience and understanding. Thank you to my sister Lael for being one of my inspirations to accomplish my goals. Thank to you Chris Sims and all of my friends from UMBC, for being a part of my life for nearly 10 years.

Thank you to Prof. Muruhan Rathinam for showing me the art of mathematical proofs, and for convincing me that research was in my future. Thank you to Dr. Freeman Hrabowski and the Meyerhoff Program staff at UMBC for their continued support and interest in my work.

Finally, thank you to the numerous friends I have made at Cornell, which include but are not limited to - Gabriel, Rolf, Martin, Jake, Dima, Matt, Brad, Jim, Baldur, Katherine, Misha, Alireza, Basit, Nabil, Yin, Coalton, Anise, and Phil.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

This thesis centers on problems in operations management settings involving customer choice. When a customer walks into a brick-and-mortar store and browses the products on the shelves, she makes the choice of what to buy. When a search for a flight is made online, the customer is shown a listing of available options, and she makes the choice. When booking a reservation at a restaurant with an online booking system, the customer views the available times, and then makes a choice. In many retail applications, customers do not finalize their decisions before entering a store (or a virtual storefront), but rather, their decision depends heavily on what they *see* when they arrive: the product offering decisions made by the retailer. The growing presence of online retailing has given customers an avenue to view a wide variety of products or services instantaneously. In light of this, capturing the substitution patterns of customers has become increasingly important when making product planning decisions.

A customer choice model is described by the following attributes: a customer arrives and views an available assortment of products S. Upon viewing the assortment, the customer either chooses one of the available products to purchase, or elects to leave without making a purchase. Associated with each product j in the set S is a probability that the customer will purchase product j. In practice, this probability depends on the assortment S as well as the prices $\mathbf{p} = (p_j)_{j \in S}$ of all the products in the assortment. As such, the purchase probability of product j is a function $q_j(\mathbf{p}, S)$. The managerial decisions studied in this dissertation take both the prices and the offered assortment into account.

One of the most prominent customer choice models is the *Multinomial Logit*

(MNL) model. The MNL model is derived from an underlying probabilistic model of customer utility, which assumes that a customer associates a random utility with each offered product, as well as the no-purchase option, and chooses the option that provides the largest realized utility. The random utilities follow a Gumbel distribution, and customer utilities are assumed to be independent from product to product The MNL model's popularity stems from its clean and attractive expressions for the customer purchase probabilities. Its major drawback is that it exhibits the independence of irrelevant alternatives (IIA) property: the ratio of probabilities of choosing any two alternatives is independent of the presence or attributes of a third alternative. Another way of saying this is that adding a new alternative to the assortment has an equal effect on the purchase probabilities of all other alternatives in the set relative to one another. If the assortment contains alternatives that can be grouped such that alternatives within a group are more similar than alternatives outside the group, the MNL model becomes unrealistic, because adding a new alternative should reduce the probability of choosing similar alternatives more than dissimilar alternatives.

In the MNL model, the IIA property results from the assumption that a customer's random utilities are assumed to be independent from product to product. The *Nested Logit (NL) model* circumvents the IIA property by relaxing the assumption of independence among customer evaluations from product to product, allowing correlation between a customer's utilities for products within designated nests (subsets of the offer set). In particular, the set of potential products is partitioned into m *nests* (subsets of the assortment), where there is correlation between the random utilities of the options in each nest. The NL model is much more appropriate for capturing settings where a retailer offers a heterogeneous collection of products that can be designated into a number of different product types or subgroups, where the products within a type are closer substitutes to each other than products from another type. Both the delineation of products into subgroups and the customer's perception of how similar products within types are can impact customer substitution patterns in ways that cannot be adequately captured by assuming that all products belong to a single category. In the NL model, the purchase probability $q_j(\mathbf{p}, S)$ of a product j depends not only on the product prices and the assortment, but also on the specific degrees of product evaluation correlation assigned to each of the m nests.

This thesis considers problems in which the product evaluations of customers are allowed to be correlated, where variants of the NL model are used to capture customer choice behavior. In these problems, three key decisions for the retailer are taken into consideration: the assortment of products offered, the prices of the offered products, and the amount of each offered product stocked.

We first consider a problem where a firm offers a product for sale via multiple sales channels, where it is assumed that competing products are present in each channel. We assume that products offered together within a particular sales channel are viewed as closer substitutes to one another than products in a different sales channel, motivating use of the NL model to capture customer choice behavior. The firm must choose the optimal price at which to offer its product in each sales channel. The work on this chapter focuses on sequential methods for finding revenue-maximizing prices and structural properties of optimal prices, which are shown to differ from properties observed in pricing problems under the MNL model.

Next, we study a more general pricing problem, where a firm offers a number of heterogeneous products than can be designated into subgroups based on their attributes. The NL model is again used to capture customer choice behavior. Our goal in this chapter is to set the prices of all the products to maximize revenue, subject to upper and lower bounds on each price. There are a number of reasons for incorporating price bounds in practice, such as ensuring that prices are competitive with the market. We also consider a variant of the same problem in which the firm must choose the assortment of products to offer in addition to their prices. The work in this chapter is algorithmic in nature, providing efficient linear-programming based approximation methods for finding near-optimal prices that allow the user to specify performance guarantees in advance. We also prove structural properties of optimal assortments for the joint assortment and pricing problem.

Finally, we consider an assortment offering problem, where we again have a firm which offers a number of different heterogeneous products that can be designated in subgroups, and the firm must choose the optimal subset of its products to offer to customers. In addition, the firm must also choose the amount of inventory of each offered product to stock. Unlike in the previous chapters, the incorporation of operational costs in the form of inventory considerations creates tradeoffs in this model. Offering larger assortments of products will attract more customers to our firm's products, but at the cost of a higher level of fragmentation of demand amongst different products, leading to higher safety stocks. Thus, a balance must be struck between the costs and benefits of product variety. The main focus of this work is examining relationships between product evaluation correlation and optimal product assortment variety. We show that fluctuations in the level of evaluation correlation between products in subgroups can impact optimal assortment variety and optimal inventory levels in ways that mirror trends observed in practice. In addition, we provide dynamic-programming solution methods for obtaining near-optimal assortments in reasonable computation time.

CHAPTER 2

PRICING A SINGLE PRODUCT SOLD VIA MULTIPLE SALES CHANNELS

2.1 Introduction

In this chapter, we are concerned with the problem of maximizing expected profits when a firm has a product that can be sold through multiple sales channels and customers follow a nested choice selection process, where they first select one of the available sales channels through which they can purchase products and then select a specific product within that sales channel. Of particular interest are the settings where a product is sold through sales channels in which there are other competing products present. Examples of such settings are found in online retailing, where a customer can utilize a number of different aggregator websites (e.g., expedia.com, kayak.com) where a product or service (e.g., a flight or hotel room) belonging to a particular firm is listed alongside competing products or services offered by other firms.

In our setting, the sales channels can either take the form of either individual retail stores or outside agencies which offer the products of competing firms together. We consider an oligopolistic pricing problem viewed from the perspective of one firm, in which each nest contains exactly one of our firm's products, but also contains alternative options that we do not control. These alternative options are assumed to be fixed before the time that our firm makes its pricing decisions, i.e., they act as no-purchase options from the perspective of our firm. In reality, they may represent products from other firms whose values are fixed ahead of time. One particular interpretation of this scenario is that our firm has m versions of a particular product and is utilizing m distinct agencies acting as intermediaries between the firm and the customer, through which it is selling these versions. The central motivating questions of this work are whether a procedure exists for finding the individual prices to set in each sales channel to maximize expected revenue, whether the markup on optimal prices is consistent across sales channels, and how optimal prices differ depending on the competing options present in each sales channel.

We utilize the nested logit (NL) model to capture the customer's selection process. The Nested Logit (NL) model, of which the Multinomial Logit (MNL) model is a special case, is a random utility maximization model which is among the most popular models to study purchase behavior of customers who must choose amongst multiple substitutable products. The MNL model has been widely used as a model of customer choice, but it suffers from limitations and may behave poorly when alternatives are correlated. This is illustrated by the independence of irrelevant alternatives property; see [26]. The NL model relaxes the assumption of independence between all the alternatives, designating the products into subgroups and capturing the level of similarity among alternatives in the same subgroup through correlation on utility components. In our setting, the NL model allows us to accurately reflect the fact that different sales channels are perceived as different product subgroups by customers and the resulting implications on customer substitution behavior. It also allows us to capture the fact that the level of perceived similarity among the products in a particular sales channel may differ from channel to channel. This is especially important when one takes into account the fact that firms have products grouped together with products offered by competing firms.

We formulate our problem as a nonlinear optimization problem, where the objective function is our firm's expected profit and the decision variables are the prices. The resulting problem is a price optimization problem under the NL model where the price in each sales channel can be viewed as the price of a distinct product. Our problem is difficult as the objective function is neither concave nor quasi-concave in its decision variables, the complexity of the purchase probability expressions under the NL model creating structural challenges.

The contributions of this chapter are as follows: we study the problem under the setting where there are two sales channels. We first show that revenue can be efficiently maximized over one price with respect to its complement price. We show that doing so can be used to define a "response function" which takes a price as an input and maximizes over the complement price with respect to the input. Despite the fact that the objective function of our original problem lacks desirable structure in its joint decision variables, we show that these response functions are unimodal and that global maximizers of the response functions can be used to construct a global maximizer of our original problem. Secondly, we provide a sequential method for maximizing our objective function using structural properties of the response functions. We also extend these results to the setting where the number of sales channels is arbitrary, showing that an analogous sequential procedure exists to find a local maximizer of our objective function, though we cannot guarantee in this case that such a maximizer is global. Lastly, we show that optimal prices have a markup (i.e., price minus cost) that corresponds to the channel in which they are placed, and that more attractive competing options within a channel lead to a lower optimal price within that channel. This differs from price optimization problems under the MNL model, where prices have been shown to have a constant markup (see [4] and [19]).

This chapter is related to several papers which explore price optimization problems under the MNL model. [16] shows that the profit function under the MNL model is not concave with respect to the prices. [36] and [11] show that the profit function is concave with respect to the purchase probabilities or market shares of the individual products and show how the optimal solution can be obtained using this transformation. [4] and [19] examine price optimization under the MNL model with identical price sensitivity coefficients, and show that the optimal markup is constant for all products; [3] shows that the profit function is unimodal and the optimal solution can be obtained by solving first-order equations.

The rest of the chapter is organized as follows. In Section 2.2, we introduce the two-channel price optimization problem. In Section 2.3, we examine structural properties that show how one price can be optimized if the second price is fixed. In Section 2.4, we develop a sequential approach for obtaining the optimal solution. Section 2.5 extends the problem to an arbitrary number of sales channels and adapts the sequential approach for finding a local maximizer of the revenue function in the general case. Section 2.6 gives structural properties of optimal prices. In Section 2.7, we conclude.

2.2 The Two-Channel Problem

We have a single product that is being sold through two different sales channels. We use $i \in \{1, 2\}$ to index a sales channel. Each channel includes one copy of our firm's product, but also contains alternative options that our firm does not control. These alternative options, which represent competing products from other firms as well as the option to purchase nothing, are assumed to be fixed before the time that our firm makes its pricing decisions. Since the customer must act within the confines of the agency he or she has chosen to utilize, our firm has no control over these alternative purchase options, only the price of the product assigned to that particular sales channel. At the top level, the customer either selects one of the two sales channels or chooses to leave the system and purchase nothing. At the second level, if the customer has chosen a sales channel, she then selects a purchase option within that channel.

In our analysis, we will use "price i" to refer to the price of our firm's product that is set for sales channel i. Let (p_1, p_2) denote our firm's price vector. Let n_i be the number of competing products in each channel $i = 1, 2$, and let $\mathbf{r}^i = (r_1^i, \ldots, r_{n_i}^i)$ be the vector of competing prices. We use use α_i and β_i to denote the base utility and price sensitivity associated with channel i, which result in the deterministic component of the random utility a customer associates with a product purchased through channel i at price p as $\alpha_i - \beta_i p$. Equivalently, we can say that this is a customer's expected utility for a product purchased through channel i at price p. We assume that there is one no-purchase option at the top level carrying a normalized expected utility of zero. Each sales channel also has a parameter $\gamma_i \in (0, 1]$ characterizing the degree of dissimilarity between the available purchase options in that channel. Under the Nested Logit (NL) model, as shown in [7], the probability that a customer makes a purchase from channel i is given by

$$\lambda_i(p_1, p_2) = \frac{(e^{\alpha_i - \beta_i p_i} + \sum_{k=1}^{n_i} e^{\alpha_i - \beta_i r_k^i})^{\gamma_i}}{1 + \sum_{l=1,2}(e^{\alpha_l - \beta_l p_l} + \sum_{k=1}^{n_l} e^{\alpha_l - \beta_l r_k^l})^{\gamma_l}}$$

for $i = 1, 2$. Given that the customer chooses channel i, the probability that a customer purchases our firm's product within channel i is given by

$$\bar{\lambda}_i(p_i) = \frac{e^{\alpha_i - \beta_i p_i}}{e^{\alpha_i - \beta_i p_i} + \sum_{k=1}^{n_i} e^{\alpha_i - \beta_i x_k^i}}.$$

Thus, the probability that a customer purchases our firm's product in sales channel i is given by

$$\Lambda_i(p_1, p_2) = \lambda_i(p_1, p_2)\bar{\lambda}_i(p_i).$$

The probability that a customer leaves the system without selecting a sales channel is $1 - \sum_{i=1,2} \Lambda_i(p_1, p_2)$, while the probability that a customer selects a sales channel i but chooses not to purchase our firm's product is given by $\lambda_i(p_1, p_2)(1 - \bar{\lambda}_i(p_i))$.

Let δ_i be the fraction of revenue that is retained from sales through channel i, i.e., $1 - \delta_i$ is the fraction of sales that channel i keeps for utilizing its service. If we offer the prices $\mathbf{p}$, the expected revenue obtained through selling the product via the two sales channels is given by

$$\Pi(p_1, p_2) = \sum_{i=1,2} \delta_i p_i \Lambda_i(p_1, p_2).$$

Our firm's objective is to choose the two prices to maximize the expected revenue. In other words, we wish to solve the problem

$$\max_{p_1, p_2 \geq 0} \Pi(p_1, p_2).$$

We note that the revenue function is not quasiconcave in the prices (p_1, p_2), even under the simpler MNL model as shown in [16]. Under the special case of the MNL model, [36] and [11] express the revenue function using the individual purchase probabilities as decision variables and show that it is concave under this transformation, but concavity does not hold for our problem. To circumvent these challenges, we will work directly with prices as decision variables, and will establish structural properties that do not require joint concavity in the prices, but nevertheless allow us to find global maximizers of the revenue function. We will start by showing how revenue can be maximized over a single price decision, and use this to characterize a response function that computes the maximizer of the revenue function over one price with respect to its complement price.

2.3 Optimization Over a Single Price

This section is concerned with maximizing the revenue function over a single price decision variable. While the revenue function is not quasiconcave in the prices (p_1, p_2), we will show that it is possible to maximize the revenue function over one price while treating the second price as fixed. This is accomplished via a first-order analysis of the revenue function. Using the subscript $-i$ to denote 1 if $i = 2$ and 2 if $i = 1$, we note that the first partial derivatives of the customer selection probabilities with respect to prices are given by

$$\frac{\partial \lambda_i}{\partial p_i}(p_1, p_2) = -\beta_i \gamma_i \bar{\lambda}_i(p_i) \lambda_i(p_1, p_2)(1 - \lambda_i(p_1, p_2)),\ i = 1, 2 \tag{2.1}$$

$$\frac{\partial \lambda_i}{\partial p_{-i}}(p_1, p_2) = \beta_{-i} \gamma_{-i} \bar{\lambda}_{-i}(p_{-i}) \lambda_{-i}(p_1, p_2) \lambda_i(p_1, p_2),\ i = 1, 2 \tag{2.2}$$

$$\frac{d\bar{\lambda}_i}{dp_i}(p_i) = -\beta_i \bar{\lambda}_i(p_i)(1 - \bar{\lambda}_i(p_i)),\ i = 1, 2. \tag{2.3}$$

Using the above, it is easily shown that the purchase probability Λ_i of the option corresponding to price i is decreasing in its own price p_i and increasing in the price p_{-i}. To express the first partial derivatives of the revenue function $\Pi(\cdot, \cdot)$, we note that

$$\begin{aligned} \frac{\partial \Pi}{\partial p_i}(p_1, p_2) &= \delta_i \Lambda_i(p_1, p_2) + \delta_i p_i \frac{\partial \Lambda_i}{\partial p_i}(p_1, p_2) + \delta_{-i} p_{-i} \frac{\partial \Lambda_{-i}}{\partial p_i}(p_1, p_2) \\ &= \delta_i \Lambda_i(p_1, p_2) + \delta_i p_i \Lambda_i(p_1, p_2) \beta_i \gamma_i \{-\bar{\lambda}_i(p_i) + \Lambda_i(p_1, p_2) - 1/\gamma_i + (1/\gamma_i)\bar{\lambda}_i(p_i)\} \\ &\qquad + \delta_{-i} p_{-i} \gamma_i \beta_i \Lambda_i(p_1, p_2) \Lambda_{-i}(p_1, p_2). \end{aligned}$$

Letting

$$C_i(p_i) = \bar{\lambda}_i(p_i) \delta_i p_i + (1 - \bar{\lambda}_i(p_i)) \frac{1}{\gamma_i} \delta_i p_i,$$

we can simplify the terms in the above equation to arrive at the expression

$$\frac{\partial \Pi}{\partial p_i}(p_1, p_2) = \Lambda_i(p_1, p_2) \Big(\delta_i + \gamma_i \beta_i (\Pi(p_1, p_2) - C_i(p_i)) \Big),\ i = 1, 2. \tag{2.4}$$

We note that $C_i(p_i)$ is a convex combination of $\delta_i p_i$ and $\frac{1}{\gamma_i}\delta_i p_i$, and since $\gamma_i \in (0,1]$, this implies that that $C_i(p_i) \geq \delta_i p_i$ for all p_i. We also note that $C_i'(p_i) > 0$ for $p_i > 0$ and $C_i'(0) = 0$, where $C_i'(\cdot)$ denotes the first derivative of $C_i(\cdot)$.

The revenue function is not necessarily quasiconcave in (p_1, p_2). However, Π is unimodal in any single price p_i when the second price p_{-i} is fixed. Furthermore, the global maximum of the function $\Pi(\cdot, p_{-i})$ can be found by solving its first-order equation. We prove this in the following lemma:

Lemma 1. *For i=1,2, the first-order equation* $\frac{\partial \Pi}{\partial p_i}(p_i, p_{-i}) = 0$ *has exactly one solution in* $p_i \geq 0$*, and this solution is a global maximizer of* $\Pi(\cdot, p_{-i})$ *on* $\mathbb{R}_+$.

Proof. Consider the first partial derivative of the revenue function with respect to p_i, as a function of p_i:

$$\frac{\partial \Pi}{\partial p_i}(p_i, p_{-i}) = \Lambda_i(p_i, p_{-i})\bigg(\delta_i + \gamma_i\beta_i(\Pi(p_i, p_{-i}) - C_i(p_i))\bigg).$$

We note that $\frac{\partial \Pi}{\partial p_i}(0, p_{-i}) > 0$, since $\lambda_i(p_1, p_2) > 0$ for all (p_1, p_2) and $C_i(0) = 0$. In addition, since $\Lambda_i(p_1, p_2) < 1$ for all (p_1, p_2), we have $\Pi(p_1, p_2) \leq \max\{\delta_1 p_1, \delta_2 p_2\}$. Since $C_i(\cdot)$ is strictly increasing for $p_i > 0$ and $C_i(p_i) \geq \delta_i p_i$, we have

$$\lim_{p_i \to \infty} \{\Pi(p_i, p_{-i}) - C_i(p_i)\} = -\infty,$$

which implies that the first-order equation in $\frac{\partial \Pi}{\partial p_i}(p_i, p_{-i}) = 0$ has at least one solution in p_i. Now consider the second derivative with respect to p_i,

$$\begin{aligned}
\frac{\partial^2 \Pi}{\partial p_i^2}(p_i, p_{-i}) = \frac{\partial \Lambda_i}{\partial p_i}(p_i, p_{-i})\bigg(\delta_i + \gamma_i\beta_i(\Pi(p_i, p_{-i}) - C_i(p_i))\bigg) \\
+ \Lambda_i(p_i, p_{-i})\gamma_i\beta_i\bigg(\frac{\partial \Pi}{\partial p_i}(p_i, p_{-i}) - C_i'(p_i)\bigg) \\
= \frac{\frac{\partial \Lambda_i}{\partial p_i}(p_i, p_{-i})}{\Lambda_i(p_i, p_{-i})}\frac{\partial \Pi}{\partial p_i}(p_i, p_{-i}) + \Lambda_i(p_i, p_{-i})\gamma_i\beta_i\bigg(\frac{\partial \Pi}{\partial p_i}(p_i, p_{-i}) - C_i'(p_i)\bigg),
\end{aligned}$$

which is negative at any p_i that satisfies $\frac{\partial \Pi}{\partial p_i}((p_i, p_{-i})) = 0$. Thus, the function $\Pi(\cdot, p_{-i})$ is strictly quasi-concave. Therefore, the first-order equation has only one solution p_i, and this solution must be a global maximizer of $\Pi(\cdot, p_{-i})$. □

We define the response function $p_i^* : \mathbb{R} \to \mathbb{R}$ by letting $p_i^*(p_{-i})$ be the solution to the first order equation $\frac{\partial \Pi}{\partial p_i}(p_i, p_{-i}) = 0$ as a function of p_{-i}. By Lemma 1, $p_i^*(p_{-i}) = \text{argmax}_{p_i} \Pi(p_i, p_{-i})$. In addition, from the first order equation (2.4), we know that, given p_{-i}, $p_i^*(p_{-i})$ is the unique price satisfying

$$C_i(p_i^*(p_{-i})) = \Pi(p_i^*(p_{-i}), p_{-i}) + \frac{\delta_i}{\gamma_i \beta_i}. \tag{2.5}$$

Thus, the value of $p_i^*(p_{-i})$ is intrinsically related to the value of the revenue function at $p_i^*(p_{-i})$. This suggests that we can maximize revenue by *maximizing the value of the function* $p_i^*(\cdot)$, as opposed to searching for a maximizer of the revenue function itself. We now show that this is in fact the case, and provide a method for obtaining the maximizers of these functions in later sections.

Lemma 2. *If $\bar{p}_{-i}$ is a global maximizer of $p_i^*(\cdot)$, then $(p_i^*(\bar{p}_{-i}), \bar{p}_{-i})$ is a global maximizer of* $\Pi(\cdot, \cdot)$.

Proof. Suppose that there exists a price vector $(\hat{p}_1, \hat{p}_2)$ such that $\Pi(\hat{p}_1, \hat{p}_2) > \Pi(p_i^*(\bar{p}_{-i}), \bar{p}_{-i})$. Then by the definition of $p_i^*(\cdot)$,

$$\Pi(p_i^*(\hat{p}_{-i}), \hat{p}_{-i}) \geq \Pi(\hat{p}_1, \hat{p}_2) > \Pi(p_i^*(\bar{p}_{-i}), \bar{p}_{-i}).$$

We know that $\frac{\partial \Pi}{\partial p_i}(p_i^*(\hat{p}_{-i}), \hat{p}_{-i}) = 0$ and $\frac{\partial \Pi}{\partial p_i}(p_i^*(\bar{p}_{-i}), \bar{p}_{-i}) = 0$. Since $\Lambda_i(\cdot, \cdot) > 0$ for all finite prices, we must have $\delta_i + \gamma_i\beta_i(\Pi(p_i^*(\hat{p}_{-i}), \hat{p}_{-i}) - C_i(p_i^*(\hat{p}_{-i}))) = \delta_i + \gamma_i\beta_i(\Pi(p_i^*(\bar{p}_{-i}), \bar{p}_{-i}) - C_i(p_i^*(\bar{p}_{-i})) = 0$. This implies that $C_i(p_i^*(\hat{p}_{-i})) - C_i(p_i^*(\bar{p}_{-i})) = \gamma_i\beta_i(\Pi(p_i^*(\hat{p}_{-i}), \hat{p}_{-i}) - \Pi(p_i^*(\bar{p}_{-i}), \bar{p}_{-i})) > 0$. But since $\bar{p}_{-i}$ is a global maximizer of $p_i^*(\cdot)$ and $C_i(\cdot)$ is increasing, this is a contradiction. □

Our next objective is to show that the response function $p_i^*(\cdot)$ in fact admits a global maximizer. We will later devise an approach for finding this maximizer and using it to construct a global maximizer of the expected revenue function.

Using the Implicit Function Theorem and equation (2.5), we verify that the first derivative of $p_i^*(\cdot)$, which we denote $\nabla p_i^*(\cdot)$, exists and is continuous. Again using the Implicit Function Theorem, $\nabla p_i^*(\cdot)$ is determined by the implicit relationship

$$\nabla p_i^*(p_{-i}) = \frac{\frac{\partial \Pi}{\partial p_{-i}}(p_i^*(p_{-i}), p_{-i})}{C_i'(p_i^*(p_{-i}))} \; \forall p_{-i}, i = 1, 2. \tag{2.6}$$

We note that $\nabla p_i^*(p_{-i}) = 0$ if and only if $\frac{\partial \Pi}{\partial p_{-i}}(p_i^*(p_{-i}), p_{-i}) = 0$, since $C_i'(\cdot) > 0$. By the definition of the function $p_i^*(\cdot)$, $\frac{\partial \Pi}{\partial p_{-i}}(p_i^*(p_{-i}), p_{-i}) = 0$ if and only if $p_{-i} = p_{-i}^*(p_i^*(p_{-i}))$, i.e., p_{-i} is the maximizer of the revenue function $\Pi(\cdot, \cdot)$ with respect to the price $p_i^*(p_{-i})$. Thus, if a price p_{-i} satisfies the first-order condition $\nabla p_i^*(p_{-i}) = 0$ for $p_i^*(\cdot)$, expected revenue cannot be improved from the price $(p_i^*(p_{-i}), p_{-i})$ by maximizing with respect to either price variable, as any such operations will simply return to the same price vector. This is illustrated in Figure 2.1, which shows that $p_1^*(\cdot)$ is maximized at the maximum value of $p_2^*(\cdot)$ and vice-versa.

A natural next question is, if such a point exists, whether or not it defines a global maximizer of $\Pi(\cdot, \cdot)$. We answer this by examining the second derivative of $p_i^*(\cdot)$, which we denote $\nabla^2 p_i^*(\cdot)$. To implicitly express the second derivative, we note that

$$\begin{aligned} \nabla p_i^*(p_{-i}) &= \frac{\Lambda_{-i}(p_i^*(p_{-i}), p_{-i})\big(\delta_{-i} + \gamma_{-i}\beta_{-i}(\Pi(p_i^*(p_{-i}), p_{-i}) - C_{-i}(p_{-i}))\big)}{C_i'(p_i^*(p_{-i}))} \\ &= \frac{\Lambda_{-i}(p_i^*(p_{-i}), p_{-i})\big(\delta_{-i} + \gamma_{-i}\beta_{-i}(C_i(p_i^*(p_{-i})) - \frac{\delta_i}{\gamma_i\beta_i} - C_{-i}(p_{-i}))\big)}{C_i'(p_i^*(p_{-i}))}, \end{aligned}$$

which follows from (2.4) and (2.5). Continuing the differentiation from the above

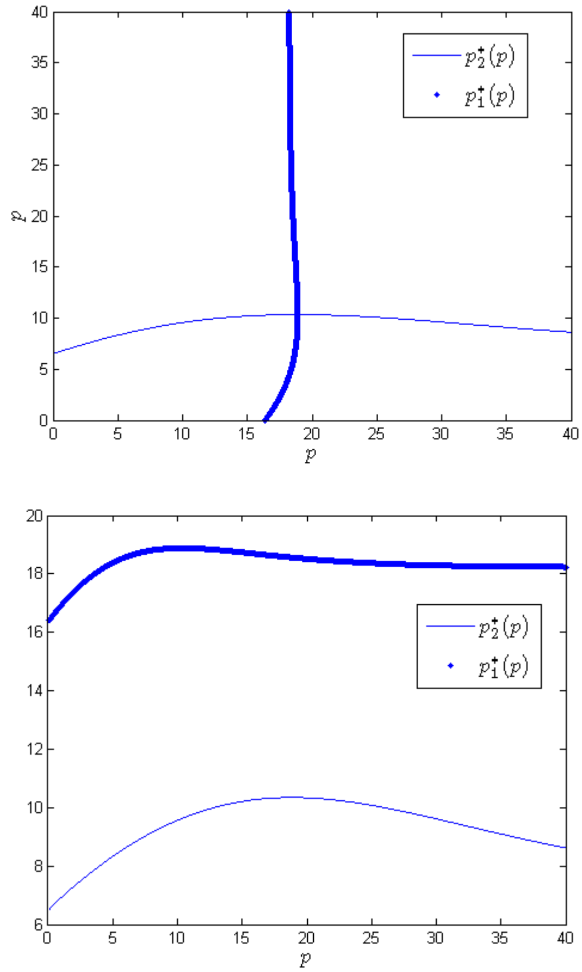

Figure 2.1: Example values of $p_1^*(\cdot)$ and $p_2^*(\cdot)$.

equation, we have

$$\begin{aligned}
&\nabla^2 p_i^*(p_{-i}) \\
&\qquad = \frac{\partial \Lambda_{-i}}{\partial p_{-i}}(p_i^*(p_{-i}), p_{-i}) \frac{\delta_{-i} + \gamma_{-i}\beta_{-i}(C_i(p_i^*(p_{-i})) - \frac{\delta_i}{\gamma_i\beta_i} - C_{-i}(p_{-i}))}{C_i'(p_i^*(p_{-i}))} \\
&\qquad\quad + \Lambda_{-i}(p_i^*(p_{-i}), p_{-i}) \cdot \left\{ \gamma_{-i}\beta_{-i} \frac{C_i'(p_i^*(p_{-i}))\nabla p_i^*(p_{-i}) - C_{-i}'(p_{-i})}{C_i'(p_i^*(p_{-i}))} \right. \\
&\qquad\quad \left. - C_i''(p_i^*(p_{-i}))\nabla p_i^*(p_{-i}) \frac{\delta_{-i} + \gamma_{-i}\beta_{-i}(C_i(p_i^*(p_{-i})) - \frac{\delta_i}{\gamma_i\beta_i} - C_{-i}(p_{-i}))}{C_i'(p_i^*(p_{-i}))^2} \right\}.
\end{aligned}$$

Using this expression, we can show that the unimodality of the revenue function over a single price extends to unimodality of the response function. This is characterized in the following proposition:

Proposition 1. *The function $p_i^*(\cdot)$ is strictly quasiconcave on $\mathbb{R}_+$.*

Proof. Suppose that $\nabla p_i^*(p_{-i}) = 0$. Then

$$
\begin{aligned}
&\nabla^2 p_i^*(p_{-i}) \\
&\quad = \frac{\partial \Lambda_{-i}}{\partial p_{-i}}(p_i^*(p_{-i}), p_{-i}) \frac{\nabla p_i^*(p_{-i})}{\Lambda_{-i}(p_i^*(p_{-i}), p_{-i})} - \Lambda_{-i}(p_i^*(p_{-i}), p_{-i})\gamma_{-i}\beta_{-i}\frac{C'_{-i}(p_{-i})}{C'_i(p_i^*(p_{-i}))} \\
&\quad = -\Lambda_{-i}(p_i^*(p_{-i}), p_{-i})\gamma_{-i}\beta_{-i}\frac{C'_{-i}(p_{-i})}{C'_i(p_i^*(p_{-i}))}.
\end{aligned}
$$

We now claim that $p_{-i} > 0$. To see this, note that since $\nabla p_i^*(p_{-i}) = 0$, we must have $p_{-i} = p_{-i}^*(p_i^*(p_{-i}))$. Therefore, p_{-i} satisfies the relationship $C_{-i}(p_{-i}) = \Pi(p_i^*(p_{-i}), p_{-i}) + \frac{\delta_{-i}}{\gamma_{-i}\beta_{-i}}$, so $C_{-i}(p_{-i})$ must be positive, implying that p_{-i} is positive. Thus, $C'_{-i}(p_{-i}) > 0$ and we have $\nabla^2 p_i^*(p_{-i}) < 0$. Since p_{-i} is an arbitrary price satisfying $\nabla p_i^*(p_{-i}) = 0$, this implies that $p_i^*(\cdot)$ is strictly quasiconcave. □

Since $p_i^*(\cdot)$ is strictly quasiconcave, any price p_{-i} that satisfies the first-order condition for $p_i^*(\cdot)$ must be a global maximizer of $p_i^*(\cdot)$. Thus, by Lemma 2, if we can find a price satisfying the first order equation $\nabla p_i^*(p_{-i}) = 0$ for either $i = 1, 2$, this price defines a global maximizer $(p_i^*(p_{-i}), p_{-i})$ of $\Pi(\cdot, \cdot)$. In the following section, we will describe a sequential approach to find this global maximizer.

2.4 A Sequential Approach To Maximizing Revenue

In this section, we present a sequential method for maximizing $\Pi(p_1, p_2)$. This method involves successively maximizing Π with respect to p_1 and subsequently maximizing Π with respect to p_2 in sequence. We will use properties of the function $p_i^*(\cdot)$ to prove that maximizing over individual prices successively in this manner leads to finding a critical point of the function $p_i^*(\cdot)$ for $i = 1, 2$, which defines a global maximizer of $\Pi(\cdot, \cdot)$.

Theorem 1. *Let* (p_1^0, p_2^0) *be an arbitrary price vector. For all* $k = 1, 2, \ldots,$ *let* $p_2^k = p_2^*(p_1^{k-1})$ *and* $p_1^k = p_1^*(p_2^k)$*. Then the sequence* $\{(p_1^k, p_2^k)\}_{k=1,2,\ldots}$ *converges to a global maximizer of* $\Pi(\cdot, \cdot)$*.*

Proof. We know that the function $p_i^*(\cdot)$ is defined for $i = 1, 2$. Since $p_2^k = p_2^*(p_1^{k-1})$ and $p_2^{k+1} = p_2^*(p_1^k)$, it follows that $\Pi(p_1^{k-1}, p_2^k) = C_1(p_2^k) - \frac{\delta_2}{\gamma_2 \beta_2}$ and $\Pi(p_1^k, p_2^{k+1}) = C_2(p_2^{k+1}) - \frac{\delta_2}{\gamma_2 \beta_2}$. Since the operation $p_i^*(\cdot)$ improves revenue, we have $\Pi(p_1^{k-1}, p_2^k) \leq \Pi(p_1^k, p_2^k) \leq \Pi(p_1^k, p_2^{k+1})$, so $C_2(p_2^k) - \frac{\delta_2}{\gamma_2 \beta_2} \leq C_2(p_2^{k+1}) - \frac{\delta_2}{\gamma_2 \beta_2}$. Since $C_2(\cdot)$ is increasing, the sequence $\{p_2^k\}_{k \geq 1}$ is increasing; we can similarly prove that $\{p_1^k\}_{k \geq 1}$ is increasing as well. Let $\bar{p}_1 = \sup_{k \geq 1}\{p_1^k\}$ and $\bar{p}_2 = \sup_{k \geq 1}\{p_2^k\}$. Since $\Pi(\cdot, \cdot)$ has a global maximizer, the functions $p_1^*(\cdot)$ and $p_2^*(\cdot)$ are bounded. This follows from equation (2.5) and the fact that $\Pi(\cdot, \cdot)$ is bounded. Therefore, $\{p_1^k\}$ and $\{p_2^k\}$ are bounded and we have $\lim\{p_1^k\} = \bar{p}_1 < \infty$ and $\lim\{p_2^k\} = \bar{p}_2 < \infty$ by the monotone convergence theorem. In addition, $\{\Pi(p_1^k, p_2^k)\}$ is bounded, monotone and convergent, and we have $\lim_{k \to \infty} \Pi(p_1^k, p_2^k) = \Pi(\bar{p}_1, \bar{p}_2)$. But also,

$$\lim_{k \to \infty} \Pi(p_1^k, p_2^k) = \lim_{k \to \infty} \Pi(p_1^k, p_2^{k+1}) = \lim_{k \to \infty} \left\{ C_2(p_2^{k+1}) - \frac{\delta_2}{\gamma_2 \beta_2} \right\} = C_2(\bar{p}_2) - \frac{\delta_2}{\gamma_2 \beta_2},$$

where the last equality follows from the fact that $C_2(\cdot)$ is increasing. Thus, we have $C_2(\bar{p}_2) = \Pi(\bar{p}_1, \bar{p}_2) + \frac{\delta_2}{\gamma_2 \beta_2}$, which implies that $\bar{p}_2 = p_2^*(\bar{p}_1)$. Similarly, we can show that $\bar{p}_1 = p_1^*(\bar{p}_2)$. Thus, $\nabla p_2^*(\bar{p}_1) = 0$ and it follows that $(\bar{p}_1, \bar{p}_2)$ is a global maximizer of $\Pi(\cdot, \cdot)$. □

Thus, despite the fact that $\Pi(p_1, p_2)$ is not quasi-concave in (p_1, p_2) in general, there exists an iterative method for finding a global maximizer of Π. We note that this sequential method differs from a tâtonnement scheme (e.g.,, in [1]), in which both prices are simultaneously optimized with respect to their complement prices in the previous iteration, i.e., $p_1^{k+1} = \mathrm{argmax}_{p_1} \Pi(p_1, p_2^k)$, $p_2^{k+1} = \mathrm{argmax}_{p_2} \Pi(p_1^k, p_2)$. Defining the sequence as in Theorem 1 is key to maintaining increasing revenue values and monotonicity of the sequences $\{p_1^k\}$ and $\{p_2^k\}$.

In the next section, we extend this iterative procedure to an arbitrary number of sales channels and show that it can be used to obtain a local maximizer of the revenue function.

2.5 Pricing with m Channels

Assume now that there are an arbitrary number of sales channels m, and let $M = \{1, \ldots, m\}$. We continue to use $i \in M$ to index sales channels. Let $\mathbf{p} = (p_1, \ldots, p_m)$ denote our firm's vector of prices to be determined across all channels. We continue to use n_i to denote the number of competing products in each channel i and x_k^i to denote the price of competing product k in channel i. Under the NL model, a customer purchases our firm's product from channel $i \in M$ with probability $\Lambda_i(\mathbf{p}) = \lambda_i(\mathbf{p})\bar{\lambda}_i(p_i)$, where

$$\lambda_i(\mathbf{p}) = \frac{(e^{\alpha_i - \beta_i p_i} + \sum_{k=1}^{n_i} e^{\alpha_i - \beta_i x_k^i})^{\gamma_i}}{1 + \sum_{l \in M}(e^{\alpha_l - \beta_l p_l} + \sum_{k=1}^{n_l} e^{\alpha_l - \beta_l x_k^l})^{\gamma_l}},$$

and $\bar{\lambda}_i(p_i)$ is the same as defined for the two-price problem. Our objective is to maximize the expected revenue

$$\Pi(\mathbf{p}) = \sum_{i \in M} \delta_i p_i \Lambda_i(\mathbf{p})$$

over $\mathbb{R}^m_+$. It is easily verified that the analogue of equation (2.4) is true in general, i.e.,

$$\frac{\partial \Pi}{\partial p_i}(\mathbf{p}) = \Lambda_i(\mathbf{p})\bigg(\delta_i + \gamma_i\beta_i(\Pi(\mathbf{p}) - C_i(p_i))\bigg),\ i \in M. \tag{2.7}$$

Analogous to the two-price setting, it is possible to maximize the function $\Pi(\cdot)$ over a single price. Namely, letting $\mathbf{p}_{-i}$ denote $(p_1, \ldots, p_{i-1}, p_{i+1}, \ldots, p_m)$, the first-order equation $\frac{\partial \Pi}{\partial p_i}(p_i, \mathbf{p}_{-i}) = 0$ has exactly one solution, which is a global maximizer of $\Pi(\cdot, \mathbf{p}_{-i})$. The proof is the same as that of Lemma 1. As such, we define the response functions $p_i^* : \mathbb{R}^{m-1} \to \mathbb{R}$ by letting $p_i^*(\mathbf{p}_{-i})$ be the solution to the first order equation $\frac{\partial \Pi}{\partial p_i}(p_i, \mathbf{p}_{-i}) = 0$, i.e.,

$$C_i(p_i^*(\mathbf{p}_{-i})) = \Pi(p_i^*(\mathbf{p}_{-i}), \mathbf{p}_{-i}) + \frac{\delta_i}{\gamma_i\beta_i}.$$

In addition, the proof of Lemma 2 can be adapted to show that if $\bar{\mathbf{p}}_{-i}$ is a local maximizer of $p_i^*(\cdot)$, then $(p_i^*(\bar{\mathbf{p}}_{-i}), \bar{\mathbf{p}}_{-i})$ is a local maximizer of $\Pi(\cdot)$.

Let $\nabla p_i^*(\cdot)$ denote the gradient of $p_i^*(\cdot)$ and let $\nabla^2 p_i^*(\cdot)$ denote its Hessian matrix. The first and second partial derivatives of $p_i^*(\cdot)$ are given by

$$\frac{\partial p_i^*}{\partial p_j}(\mathbf{p}_{-i}) = \Lambda_j(p_i^*(\mathbf{p}_{-i}), \mathbf{p}_{-i})\frac{\delta_j + \gamma_j\beta_j[C_i(p_i^*(\mathbf{p}_{-i})) - \frac{\delta_i}{\alpha_i} - C_j(p_j)]}{C_i'(p_i^*(\mathbf{p}_{-i}))},\ j \neq i,$$

and

$$\begin{aligned}\frac{\partial^2 p_i^*}{\partial p_j \partial p_k}(\mathbf{p}_{-i}) = {}& \frac{\frac{\partial \Lambda_j}{\partial p_k}(p_i^*(\mathbf{p}_{-i}), \mathbf{p}_{-i})}{\Lambda_j(p_i^*(\mathbf{p}_{-i}), \mathbf{p}_{-i})}\frac{\partial p_i^*}{\partial p_j}(\mathbf{p}_{-i}) \\ & + \Lambda_j(p_i^*(\mathbf{p}_{-i}), \mathbf{p}_{-i}) \cdot \Bigg\{\gamma_j\beta_j\frac{\alpha_j C'(p_i^*(\mathbf{p}_{-i}))\frac{\partial p_i^*}{\partial p_k}(\mathbf{p}_{-i}) - C_j'(p_j)\mathbf{1}(k = j)}{C_i'(p_i^*(\mathbf{p}_{-i}))} \\ & - C_i''(p_i^*(\mathbf{p}_{-i}))\frac{\partial p_i^*}{\partial p_k}(\mathbf{p}_{-i})\frac{\delta_j + \gamma_j\beta_j(C_i(p_i^*(\mathbf{p}_{-i})) - \frac{\delta_i}{\alpha_i} - C_j(p_j))}{(C_i'(p_i^*(\mathbf{p}_{-i})))^2}\Bigg\},\ j, k \neq i,\end{aligned}$$

respectively. We observe that if the gradient $\nabla p_i^*(\mathbf{p}_{-i})$ is zero, then $\frac{\partial p_i^*}{\partial p_j}(\mathbf{p}_{-i}) = 0$

for all $j \neq i$, so the entries of the Hessian at $\mathbf{p}_{-i}$ are given by

$$\frac{\partial^2 p_i^*}{\partial p_j \partial p_k}(\mathbf{p}_{-i}) = \begin{cases} -\Lambda_j(p_i^*(\mathbf{p}_{-i}), \mathbf{p}_{-i})\gamma_j\beta_j \frac{C_j'(p_j)}{C_i'(p_i^*(\mathbf{p}_{-i}))}, & k = j \\ 0, & k \neq j. \end{cases}$$

Thus, the Hessian at $\mathbf{p}_{-i}$ is diagonal with negative entries, and is therefore negative definite at any vector $\mathbf{p}_{-i}$ satisfying $\nabla p_i^*(\mathbf{p}_{-i}) = \mathbf{0}$. While this is not strong enough to prove that $p_i^*(\cdot)$ is a quasiconcave function as in the two-price setting, it does imply that any $\mathbf{p}_{-i}$ satisfying $\nabla p_i^*(\mathbf{p}_{-i}) = \mathbf{0}$ is a local maximizer of $p_i^*(\cdot)$. We will use this fact to formulate an analogue of our approach from the two-price setting to find a local maximizer of $\Pi(\cdot)$.

Define the families of operators $\{\mathbf{T}_i\}_{i \in M} : \mathbb{R}^m \to \mathbb{R}^m$ and$\{\bar{\mathbf{T}}_i\}_{i \in M \cup \{0\}} : \mathbb{R}^m \to \mathbb{R}^m$ as follows: $\bar{\mathbf{T}}_0(\mathbf{p}) = \mathbf{p}$, $\mathbf{T}_i(\mathbf{p}) = \bar{\mathbf{T}}_{i-1}(p_i^*(\mathbf{p}_{-i}), \mathbf{p}_{-i})$ and $\bar{\mathbf{T}}_i(\mathbf{p}) = \lim_{k \to \infty} \mathbf{T}_i^k(\mathbf{p})$, where

$$\mathbf{T}_i^k(\mathbf{p}) = \underbrace{\mathbf{T}_i(\mathbf{T}_i(\cdots(\mathbf{T}_i}_{k}(\mathbf{p})))).$$

Our sequential procedure for the m-channel problem involves evaluating $\mathbf{T}_m(\mathbf{p}^0)$ from any initial price vector $\mathbf{p}$. We observe that if $m = 2$, this corresponds exactly to the sequential procedure that we formulated for the 2-channel problem. The following theorem demonstrates that $\bar{\mathbf{T}}_i(\mathbf{p})$ is defined for all $i \in M$ and that $\bar{\mathbf{T}}_m(\mathbf{p})$ is a local maximizer of the revenue function.

Theorem 2. *For all $\boldsymbol{p} \in \mathbb{R}_+^m$, $\{\boldsymbol{T}_i^k(\boldsymbol{p})\}_{k \geq 1}$ converges for all $i \in M$, and $\bar{\boldsymbol{T}}_m(\boldsymbol{p})$ is a local maximizer of* $\Pi(\cdot)$.

Proof. The proof is by induction. For notational brevity, let $T_{ij}^k(\mathbf{p})$ denote the jth component of $\mathbf{T}_i^k(\mathbf{p})$ and $\bar{T}_{ij}(\mathbf{p})$ denote the jth component of $\bar{\mathbf{T}}_i(\mathbf{p})$; also let $\bar{\mathbf{T}}_{i,-j}(\mathbf{p})$ denote the vector of all components of $\bar{\mathbf{T}}_i(\mathbf{p})$ besides the jth component.

Base case: We note that $\mathbf{T}_1(\mathbf{p}) = (p_1^*(\mathbf{p}_{-1}), \mathbf{p}_{-1})$. Thus, since $\bar{\mathbf{T}}_0$ is the identity, $\mathbf{T}_1(\mathbf{T}_1(\mathbf{p})) = (p_1^*(\mathbf{p}_{-1}), \mathbf{p}_{-1})$ as well, and it follows that $\bar{\mathbf{T}}_1(\mathbf{p}) = (p_1^*(\mathbf{p}_{-1}), \mathbf{p}_{-1})$. We have

$$T_2(\mathbf{p}) = \bar{\mathbf{T}}_1(p_2^*(\mathbf{p}_{-2}), \mathbf{p}_{-2}) = (p_1^*(p_2^*(\mathbf{p}_{-2})), p_2^*(\mathbf{p}_{-2}), p_3, \ldots, p_m).$$

From Theorem 1, we know that for any $\mathbf{p}$, $\lim_{k\to\infty} \mathbf{T}_2^k(\mathbf{p}) = \bar{\mathbf{T}}_2(\mathbf{p})$ exists and that $\bar{T}_{22}(\mathbf{p}) = p_2^*(\bar{T}_{21}(\mathbf{p}), p_3, \ldots, p_m)$ and $\bar{T}_{21}(\mathbf{p}) = p_1^*(\bar{T}_{22}(\mathbf{p}), p_3, \ldots, p_m)$. We can show this by considering $\Pi(\cdot, p_3, \ldots, p_m)$ as a two-price revenue function and following the sequential approach over the first two prices from Theorem 1, which is exactly the procedure involved in finding $\lim_{k\to\infty} \mathbf{T}_2^k(\mathbf{p})$.

Inductive step: Now suppose that for some $2 \leq i \leq m-1$, $\bar{\mathbf{T}}_i(\mathbf{p})$ exists and $\bar{T}_{ij}(\mathbf{p}) = p_j^*(\bar{\mathbf{T}}_{i,-j}(\mathbf{p}))$ for all $j = 1, \ldots, i$. Then for all $k \geq 1$,

$$T_{i+1,i+1}^k(\mathbf{p}) = p_{i+1}^*(T_{i+1,1}^{k-1}(\mathbf{p}), \ldots, T_{i+1,i}^{k-1}(\mathbf{p}), p_{i+2}, \ldots, p_m)$$

and

$$T_{i+1,j}^k(\mathbf{p}) = \bar{T}_{ij}(T_{i+1,1}^{k-1}(\mathbf{p}), \ldots, T_{i+1,i}^{k-1}(\mathbf{p}), T_{i+1,i+1}^k(\mathbf{p}), p_{i+2}, \ldots, p_m)\, \forall j = 1, \ldots, i.$$

Thus, since $\Pi(T_{i+1,1}^{k-1}(\mathbf{p}), \ldots, T_{i+1,i}^{k-1}(\mathbf{p}), T_{i+1,i+1}^k(\mathbf{p}), p_{i+2}, \ldots, p_m) = C_i(T_{i+1,i+1}^k(\mathbf{p})) - \frac{\delta_i}{\gamma_i\beta_i}$ and $\Pi(\mathbf{T}_i^k(\mathbf{p})) = C_j(T_{i+1,j}^k(\mathbf{p})) - \frac{\delta_j}{\gamma_j\beta_j}\, \forall j = 1, \ldots, i-1$ are increasing in k, both $\{T_{i+1,i+1}^k(\mathbf{p})\}$ and $\{T_{i+1,j}^k(\mathbf{p})\}$, $j = 1, \ldots, i-1$ are increasing in k, bounded and convergent. Thus, $\bar{\mathbf{T}}_{i+1}(\mathbf{p})$ exists. We have

$$\begin{aligned}\Pi(\bar{\mathbf{T}}_{i+1}(\mathbf{p})) &= \lim_{k\to\infty} \Pi(\mathbf{T}_{i+1}^k(\mathbf{p})) \\ &= \lim_{k\to\infty} \Pi(T_{i+1,1}^{k-1}(\mathbf{p}), \ldots, T_{i+1,i}^{k-1}(\mathbf{p}), T_{i+1,i+1}^k(\mathbf{p}), p_{i+2}, \ldots, p_m) \\ &= \lim_{k\to\infty} \left\{ C_i(T_{i+1,i+1}^k(\mathbf{p})) - \frac{\delta_i}{\gamma_i\beta_i} \right\} = C_i(\bar{T}_{i+1,i+1}(\mathbf{p})) - \frac{\delta_i}{\gamma_i\beta_i},\end{aligned}$$

so $\bar{T}_{i+1,i+1}(\mathbf{p}) = p_{i+1}^*(\bar{\mathbf{T}}_{i+1,-(i+1)}(\mathbf{p}))$. We can similarly show that $\bar{T}_{i+1,j}(\mathbf{p}) = p_j^*(\bar{\mathbf{T}}_{i+1,-j}(\mathbf{p}))$ for all $j = 1, \ldots, i$.

Thus, by induction, for all $i \in M$, $\{\mathbf{T}_i^k(\mathbf{p})\}_{k\geq 1}$ converges and $\bar{T}_{ij}(\mathbf{p}) = p_j^*(\bar{\mathbf{T}}_{i,-j}(\mathbf{p}))$ for all $j = 1, \ldots, i$. So $\bar{\mathbf{T}}_m(\mathbf{p})$ exists and we have $\bar{T}_{mj}(\mathbf{p}) = p_j^*(\bar{\mathbf{T}}_{m,-j}(\mathbf{p}))$ for all $j \in M$. This implies that $\frac{\partial p_i^*}{\partial p_j}(\bar{\mathbf{T}}_{m,-i}(\mathbf{p})) = \frac{\frac{\partial \Pi}{\partial p_j}(p_i^*(\bar{\mathbf{T}}_{m,-i}(\mathbf{p})), \bar{\mathbf{T}}_{m,-i}(\mathbf{p}))}{C_i'(p_i^*(\bar{\mathbf{T}}_{m,-i}(\mathbf{p})))} = \frac{\frac{\partial \Pi}{\partial p_j}(\bar{\mathbf{T}}_m(\mathbf{p}))}{C_i'(p_i^*(\bar{\mathbf{T}}_{m,-i}(\mathbf{p})))} = 0$ for all $j \in M$. Thus, $\bar{\mathbf{T}}_{m,-i}(\mathbf{p})$ is a local maximizer of $p_i^*(\cdot)$, which implies that $\bar{\mathbf{T}}_m(\mathbf{p})$ is a local maximizer of $\Pi(\cdot)$. □

2.6 Properties of Optimal Prices

If a price vector $\mathbf{p}$ is globally optimal, then we must have $p_i^*(\mathbf{p}_{-i}) = p_i$ for all $i \in M$. Thus, we have $C_i(p_i) = \Pi(\mathbf{p}) + \frac{\delta_i}{\gamma_i \beta_i}$ for all $i \in M$. [4] and [19] show that in price optimization problems under the MNL model, all optimal prices have a constant *markup*, i.e., price minus cost. In our case, we express the markup of price i as $\delta_i p_i$, as the percentage δ_i encompasses the costs incurred from offering our firm's product in channel i. If $\delta_i = 1$, the markup of price i is simply its price, and in the MNL model when all costs are zero, a constant price among all products is optimal. In our case, we note that

$$C_i(p_i) = \frac{e^{\alpha_i - \beta_i p_i}}{A_i + e^{\alpha_i - \beta_i p_i}} \delta_i p_i + \frac{A_i}{A_i + e^{\alpha_i - \beta_i p_i}} \frac{1}{\gamma_i} \delta_i p_i,$$

where $A_i = \sum_{k=1}^{n_i} e^{\alpha_i - \beta_i r_k^i}$ is the aggregate attractiveness of the competitor prices and the no-purchase option within channel i. Therefore, if Π^* denotes the optimal revenue, every optimal price p_i must satisfy

$$\delta_i p_i = \frac{\gamma_i (A_i + e^{\alpha_i - \beta_i p_i})}{A_i + \gamma_i e^{\alpha_i - \beta_i p_i}} \Big(\Pi^* + \frac{\delta_i}{\gamma_i \beta_i} \Big). \tag{2.8}$$

We make two key observations from equation (2.8). The first is that optimal markups are not constant. We note that even if $A_i = 0$ (i.e., there are no competing purchase options within channel i), the optimal channel markup is equal to the optimal revenue plus the constant $\frac{\delta_i}{\gamma_i \beta_i}$, which is unique to the channel i. If $A_i > 0$, then the optimal channel markup is a channel-specific function of the optimal revenue. Second, we note that the right-hand side of equation (2.8) is a decreasing function in p_i, and thus equation (2.8) has a unique solution. But in addition, the right-hand side of (2.8) is decreasing in A_i for any fixed p_i and the optimal revenue Π^* is decreasing with A_i, where the latter can be shown by noting that all of the purchase probabilities $\Lambda_j(\mathbf{p})_{j \in M}$ are decreasing in A_i for any fixed $\mathbf{p}$. Thus, as A_i gets larger, the value of p_i where the left and right-hand sides of equation (2.8) intersects gets smaller. This implies that more attractive competing products within a channel lead to a lower channel price.

2.7 Conclusions

We formulated a pricing problem where a firm prices a single product through multiple sales channels and competing products are present in each channel. We used the nested logit model to capture the customer's selection process. In the setting where there are two sales channels, we showed that while the expected revenue function is not quasi concave in the prices, maximizing prices individually with respect to one another yields useful structural properties. We used this knowledge to formulate a sequential procedure that can be used to find a global maximizer of the expected revenue function by maximizing over individual prices successively. We extended this result to a setting with an arbitrary number of sales channels and showed that an analogous procedure can be used to find a local maximum of the revenue function. In both settings, we show that the markups of optimal prices are determined by channel-specific parameters and that the optimal price in a particular channel is decreasing in the aggregate attractiveness of the competing products present in that channel.

It still remains to be shown whether a tractable method exists for solving more general problems under the nested logit model, e.g., by relaxing the assumption that our firm only controls one product in each sales channel or product subgroup or by incorporating constraints on the prices of products our firm controls. These questions are addressed later in the thesis.

CHAPTER 3

APPROXIMATION METHODS FOR MULTI-PRODUCT PRICING PROBLEMS WITH PRICE BOUNDS

3.1 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 chapter, 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.

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 [24] and [14]. [24] considers 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, the authors 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. [14] also studies 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. This chapter 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, which is the expected revenue as the function of an adjusted markup parameter. However, 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 [24] and [14].

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 [24] and [14] 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 [24] 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 [24] or [14] 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 [14] 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.

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 [26]. 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 [28], resulting in the celebrated multinomial logit model. The nested logit model, which plays a central role in this chapter, goes back to the work of [42]. Extensions for the nested logit model are provided by [29] and [7]. An important feature of the nested logit model is that it avoids the independence of irrelevant alternatives property suffered by the multinomial logit model. A discussion of this property can be found in [6].

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. [37] studies 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. [34] considers 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. [40] extends this work to more general versions of the multinomial logit model. In [8], [31] and [35], 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 the weak and strong sense, propose approximation methods and study integer programming formulations. [20] 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. [33] 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. [10] studies the same problem without the shelf space constraint and give a tractable method to obtain the optimal assortment under the nested logit model. [13] 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, [16] note that the expected revenue function is not concave in prices. However, [36] and [11] make progress by formulating the problem in terms of market shares, as this formulation yields a concave expected revenue function. [24] 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. [14] relax both of the assumptions in [24] and extend the analysis to more general forms of the nested logit model. [41] 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 has been 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 problems can be found in [12], [25], [23], [43], [44] and [30].

In Section 3.2, 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 3.3, we show that this problem can be visualized as finding the fixed point of a scalar function. In Section 3.4, 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 3.5, we show how to construct an appropriate grid with a performance guarantee and give our approximation method. In Section 3.6, 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 3.7, 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 3.8, we conclude.

3.2 Problem Formulation

In this section, we describe the nested logit model and formulate the pricing problem. There are m nests indexed by $M = \{1, \dots, m\}$. 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 chapter, 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

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

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

$$\begin{aligned}\Pi(\boldsymbol{w}_1,\ldots,\boldsymbol{w}_m) &= \sum_{i\in M} Q_i(\boldsymbol{w}_1,\ldots,\boldsymbol{w}_m)\,R_i(\boldsymbol{w}_i)\\ &= \frac{\sum_{i\in M}\Big(\sum_{j\in N} w_{ij}\Big)^{\gamma_i}\frac{\sum_{j\in N} w_{ij}\,(\kappa_{ij}-\eta_{ij}\log w_{ij})}{\sum_{j\in N} w_{ij}}}{1+\sum_{i\in M}\big(\sum_{j\in N} w_{ij}\big)^{\gamma_i}},\end{aligned} \tag{3.1}$$

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

$$Z^* = \max\Big\{\Pi(\boldsymbol{w}_1,\ldots,\boldsymbol{w}_m) : \boldsymbol{w}_i \in [\boldsymbol{L}_i,\boldsymbol{U}_i] \;\; \forall\, i\in M\Big\}, \tag{3.2}$$

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 $j \in N$.

3.3 Fixed Point Representation

In this section, we show that problem (3.2) 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

$$z = \sum_{i \in M} \max_{\boldsymbol{w}_i \in [\boldsymbol{L}_i, \boldsymbol{U}_i]} \left\{ \Big(\sum_{j \in N} w_{ij} \Big)^{\gamma_i} \frac{\sum_{j \in N} w_{ij} \left(\kappa_{ij} - \eta_{ij} \log w_{ij} \right)}{\sum_{j \in N} w_{ij}} - \Big(\sum_{j \in N} w_{ij} \Big)^{\gamma_i} z \right\}. \tag{3.3}$$

Viewing the right side of (3.3) as a function of z, finding the value of z satisfying (3.3) 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 (3.3), we claim that $\hat{z}$ is the optimal objective value of problem (3.2). To see this claim, note that if $(\boldsymbol{w}_1^*, \ldots, \boldsymbol{w}_m^*)$ is an optimal solution to problem (3.2), 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 (3.3) and $\boldsymbol{w}_i^*$ is a feasible but not necessarily an optimal solution to the maximization problem on the right side of

(3.3) 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 (3.1), 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 (3.3) 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 (3.2). Furthermore, since $\hat{z}$ is the value of z that satisfies (3.3), the definition of $\hat{\boldsymbol{w}}_i$ implies that

$$\hat{z} = \sum_{i \in M} \left\{ \Big(\sum_{j \in N} \hat{w}_{ij} \Big)^{\gamma_i} \, \frac{\sum_{j \in N} \hat{w}_{ij} \, (\kappa_{ij} - \eta_{ij} \log \hat{w}_{ij})}{\sum_{j \in N} \hat{w}_{ij}} - \Big(\sum_{j \in N} \hat{w}_{ij} \Big)^{\gamma_i} \hat{z} \right\}. \tag{3.4}$$

If we solve for $\hat{z}$ in the equality above and use the definition of $\Pi(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m)$ in (3.1) 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 (3.2). So, we obtain $\hat{z} = Z^*$, establishing the claim. Thus, we can obtain the optimal objective value of problem (3.2) by finding the value of z that satisfies (3.3). 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 (3.3) 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 (3.2). Since the left and right sides of (3.3) are respectively increasing and decreasing in z, we can find the value of z satisfying (3.3) 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 (3.3) 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 (3.3). We define $g_i(y_i)$ as the optimal objective value of the nonlinear knapsack problem

$$g_i(y_i) = \max\left\{ \sum_{j \in N} w_{ij} \left(\kappa_{ij} - \eta_{ij} \log w_{ij}\right) : \sum_{j \in N} w_{ij} \leq y_i,\ w_{ij} \in [L_{ij}, U_{ij}] \ \ \forall j \in N \right\}. \tag{3.5}$$

We make a number of observations regarding problem (3.5). 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 (3.5) 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 (3.5) is infeasible. Finally, if we have $y_i \in [\bar{L}_i, \bar{U}_i]$, then it follows that the first constraint in problem (3.5) 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 (3.5) finds the maximum value of the numerator of the fraction in (3.3) while keeping the denominator of this fraction below y_i, instead of finding the value of z satisfying (3.3), we propose finding the value of z that satisfies

$$z = \sum_{i \in M} \max_{y_i \in [\bar{L}_i, \bar{U}_i]} \left\{ y_i^{\gamma_i} \, \frac{g_i(y_i)}{y_i} - y_i^{\gamma_i} z \right\}. \tag{3.6}$$

The value of z satisfying (3.6) 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 (3.6) can be used to find the value of z satisfying (3.3).

Proposition 2. *The value of* z *that satisfies* (3.3) *and* (3.6) *are the same, corresponding to the optimal objective value of problem* (3.2).

Proof. The value of z that satisfies (3.3) or (3.6) 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 (3.3) and (3.6), 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 (3.3) and (3.6) 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 the appendix. □

The proposition above provides a tempting approach for solving problem (3.2). In particular, we can find the value of z that satisfies (3.6) by using bisection search. We observe that the maximization problem on the right side of (3.6) 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 (3.6) and $\hat{y}_i$ to denote an optimal solution to the maximization problem on the right side of (3.6) when we solve this problem with $z = \hat{z}$. In this case, we can solve problem (3.5) with $y_i = \hat{y}_i$ to obtain an optimal solution $\hat{\boldsymbol{w}}_i$. Once we solve problem (3.5) 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 (3.2).

3.4 Approximation Framework

As mentioned at the end of the previous section, the maximization problem on the right side of (3.6) involves a scalar decision variable and it is tempting to try to solve problem (3.2) by finding the value of z satisfying (3.6). 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 (3.6). To give an example where the objective function of the maximization problem on the right side of (3.6) 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 3.1 plots the objective function of the maximization problem on the right side of (3.6) 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 (3.2) is close to 24.74 for this problem instance. So, we do not have unimodality even with sensible values of z. Interestingly, [14] considers 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 (3.6) 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 [14]. However, due to the presence of the lower and upper bounds on the prices, we lose the unimodality property.

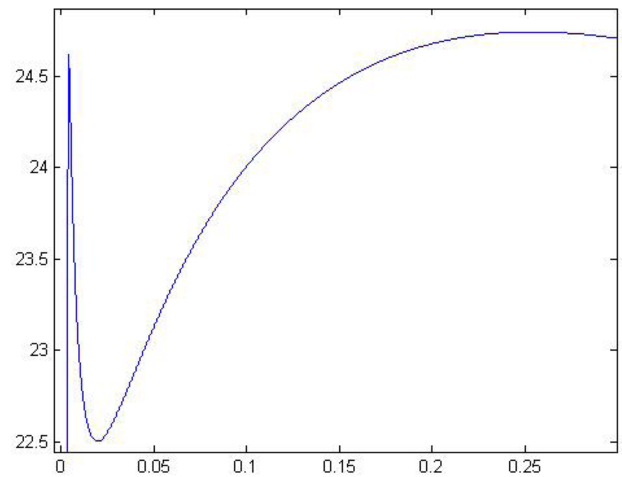

Figure 3.1: The function $y_1^{\gamma_1}\,(g_1(y_1)/y_1) - y_1^{\gamma_1} z$ as a function of y_1.

The objective function of the maximization problem on the right side of (3.6) 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 (3.6), we can focus only on the grid points and find the value of z that satisfies

$$z = \sum_{i \in M} \max_{y_i \in \{\tilde{y}_i^t \,:\, t=1,\ldots,T_i\}} \left\{ y_i^{\gamma_i} \, \frac{g_i(y_i)}{y_i} - y_i^{\gamma_i} z \right\}. \tag{3.7}$$

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 (3.5) does not change too much at the successive grid points, then we can build on the value of z satisfying (3.7) to construct a solution to problem (3.2) with a certain performance guarantee.

Theorem 3. *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* (3.7) *and Z^* denotes the optimal objective value of problem* (3.2)*, 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 (3.6) 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 (3.7) and $\tilde{y}_i^{t_i}$ is a feasible but not necessarily an optimal solution to the maximization problem on the right side of (3.7) 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

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

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 2, Z^* corresponds to the value of z that satisfies (3.6), but the last chain of inequalities above shows that Z^* does not satisfy (3.6), which is a contradiction. □

When we work with grid points that satisfy the assumption of Theorem 3, this theorem allows us to obtain a $(1+\rho)$-approximate solution to problem (3.2) in the following fashion. We find the value of z that satisfies (3.7) 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 (3.7) when this problem is solved with $z = \hat{z}$. For all $i \in M$, we solve problem (3.5) 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 (3.7) and $\hat{y}_i$ is an optimal solution to the maximization problem on the right side of (3.7) when this problem is solved with $z = \hat{z}$, we have

$$\hat{z} = \sum_{i \in M} \left\{ \hat{y}_i^{\gamma_i} \frac{g_i(\hat{y}_i)}{\hat{y}_i} - \hat{y}_i^{\gamma_i} \hat{z} \right\}. \tag{3.8}$$

Also, since $\hat{y}_i \in [\bar{L}_i, \bar{U}_i]$, the discussion right after the formulation of problem (3.5) 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 (3.5) 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 (3.8) 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 (3.4). So, if we collect all terms that involve $\hat{z}$ on the left side of (3.4), 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 3, 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 3, gives a framework for obtaining approximate solutions to problem (3.2) 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 3. 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 (3.7) 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 3. Before doing so, however, we make a brief remark on how to find the value of z that satisfies (3.7). 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 (3.7) by solving a linear program.

To formulate the linear program, we note that the left side of the equality in (3.7) is increasing in z, whereas the right side is decreasing. Therefore, the value of z that satisfies (3.7) corresponds to the smallest value of z such that the left side of the equality in (3.7) is still greater than or equal to the right side. This observation immediately implies that finding the value of z satisfying (3.7) 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

$$\min \left\{ z : z \geq \sum_{i \in M} x_i,\ x_i \geq y_i^{\gamma_i} \frac{g_i(y_i)}{y_i} - y_i^{\gamma_i} z \ \ \forall\, y_i \in \{\tilde{y}_i^t : t = 1, \ldots, T_i\},\ i \in M \right\}, \tag{3.9}$$

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 (3.7).

3.5 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 3. By noting the discussion that follows Theorem 3 in the previous section, such a collection of grid points allows us to obtain a solution to problem (3.2) 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 (3.5). After we give these properties, we proceed to showing how we can build on these properties to construct the collection of grid points.

3.5.1 Properties of Knapsack Problems

The first property that we have for problem (3.5) 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 (3.5) 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 (3.5) can be obtained by setting

$$w_{ij} = \min\left\{ \max\left\{ \exp\left(\frac{\kappa_{ij}}{\eta_{ij}} - 1 - \frac{\lambda_i}{\eta_{ij}} \right), L_{ij} \right\}, U_{ij} \right\} \tag{3.10}$$

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 (3.5), the first constraint in this problem must be tight at the optimal solution. Therefore, noting (3.10), 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 (3.5). 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 (3.5) as a function of the right side of this constraint, then $\lambda_i^*(y_i)$ satisfies

$$\sum_{j \in N} \min\left\{ \max\left\{ \exp\left(\frac{\kappa_{ij}}{\eta_{ij}} - 1 - \frac{\lambda_i^*(y_i)}{\eta_{ij}} \right), L_{ij} \right\}, U_{ij} \right\} = y_i. \tag{3.11}$$

Furthermore, $\lambda_i^*(y_i)$ is decreasing in y_i. Since the optimal value of the decision variable w_{ij} in problem (3.5) 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 (3.5) is increasing in y_i, as desired. Therefore, we can let ζ_{ij} and ξ_{ij} be such that

$$\exp\left(\frac{\kappa_{ij}}{\eta_{ij}} - 1 - \frac{\lambda_i^*(\zeta_{ij})}{\eta_{ij}}\right) = L_{ij} \quad \text{and} \quad \exp\left(\frac{\kappa_{ij}}{\eta_{ij}} - 1 - \frac{\lambda_i^*(\xi_{ij})}{\eta_{ij}}\right) = U_{ij}, \qquad (3.12)$$

in which case, (3.10) implies that if $y_i = \zeta_{ij}$, then we have $w_{ij} = L_{ij}$ in the optimal solution to problem (3.5), 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 (3.5) 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 (3.5) 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 (3.12). 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. *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* (3.5) *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* (3.5) *without changing the optimal solution to this problem.*

Proof. We let ζ_{ij} and ξ_{ij} be as defined in (3.12), 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 (3.5) 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, (3.12) 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 (3.10), the optimal value of the decision variable w_{ij} in problem (3.5) 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 (3.11) 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 (3.5) 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 (3.5) 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 (3.10) 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. □

The second property that we have for problem (3.5) 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 (3.5) 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 (3.5) 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 $y_i \leq \zeta_{ij}$, in which case, Lemma 3 implies that the optimal value of the decision variable w_{ij} in problem (3.5) 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 (3.5) 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 (3.5) without changing the optimal solution. Therefore, whenever we solve problem (3.5) 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. *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* (3.5) *can be obtained by solving*

$$\max \left\{ \sum_{j \in N} w_{ij} \left(\kappa_{ij} - \eta_{ij} \log w_{ij} \right) : \right.$$
$$\left. \sum_{j \in N} w_{ij} \leq y_i, \;\; w_{ij} = L_{ij} \;\; \forall\, j \in \mathcal{L}_i^k, \;\; w_{ij} = U_{ij} \;\; \forall\, j \in \mathcal{U}_i^k \right\} \quad (3.13)$$

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 3. We focus on this task in the next section.

3.5.2 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 3. To that end, we choose a fixed value of $\rho > 0$ and consider the grid points that are obtained by

$$\tilde{Y}_i^{kq} = \sum_{j \in \mathcal{L}_i^k} L_{ij} + \sum_{j \in \mathcal{U}_i^k} U_{ij} + (1 + \rho)^q \tag{3.14}$$

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 (3.5), 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 (3.14) 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 3, 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 $\tilde{y}_i^t \leq \nu_i^k \leq \tilde{y}_i^{t+1}$ 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 $\tilde{Y}_i^{kq} \leq \tilde{y}_i^t \leq \tilde{y}_i^{t+1} \leq \tilde{Y}_i^{k,q+1}$ 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 $\tilde{y}_i^t \leq \tilde{Y}_i^{kq} \leq \tilde{y}_i^{t+1}$ for some $q = q_i^L, \ldots, q_i^U - 1$ or we have $\tilde{y}_i^t \leq \tilde{Y}_i^{k,q+1} \leq \tilde{y}_i^{t+1}$ 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 3.

Theorem 4. *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* $g_i(\tilde{y}_i^{t+1}) \leq (1+\rho)\, g_i(\tilde{y}_i^t)$ *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 $\tilde{Y}_i^{kq} \leq \tilde{y}_i^t \leq \tilde{y}_i^{t+1} \leq \tilde{Y}_i^{k,q+1}$. 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 (3.14), we obtain

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

Using the chain of inequalities above, it follows that $\tilde{y}_i^{t+1} - \sum_{j \in \mathcal{L}_i^k} L_{ij} - \sum_{j \in \mathcal{U}_i^k} U_{ij} \leq (1+\rho)^{q+1} \leq (1+\rho)\,(\tilde{y}_i^t - \sum_{j \in \mathcal{L}_i^k} L_{ij} - \sum_{j \in \mathcal{U}_i^k} U_{ij})$. 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 (3.5) with $y_i = \tilde{y}_i^t$ or $y_i = \tilde{y}_i^{t+1}$ can be obtained by solving problem (3.13) 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 (3.13) 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

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

where the first inequality is by the fact that $\boldsymbol{w}_i^*$ is a feasible solution to problem (3.13) when we solve this problem with $y_i = \tilde{y}_i^{t+1}$ and the second inequality follows from the fact that $\tilde{y}_i^{t+1} - \sum_{j \in \mathcal{L}_i^k} L_{ij} - \sum_{j \in \mathcal{U}_i^k} U_{ij} \leq (1+\rho)\,(\tilde{y}_i^t - \sum_{j \in \mathcal{L}_i^k} L_{ij} - \sum_{j \in \mathcal{U}_i^k} U_{ij})$, 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 (3.13) when we solve this problem with $y_i = \tilde{y}_i^t$, in which case, we obtain

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

where the first inequality uses the fact that $\hat{\boldsymbol{w}}_i$ is a feasible solution to problem (3.13) when solved with $y_i = \tilde{y}_i^t$ and this problem yields the optimal solution to problem (3.5) 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 (3.13) when solved with $y_i = \tilde{y}_i^{t+1}$. The chain of inequalities above establishes the desired result. $\square$

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 3. 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 (3.11) when defining $\lambda_i^*(y_i)$ and add w_{i0} to the right side of (3.14) when defining $\tilde{Y}_i^{kq}$.

3.5.3 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 (3.2).

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 (3.11). 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 (3.9) 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 (3.7) 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 (3.5) 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 (3.2).

In Steps 1, 2 and 3, we compute the collection of grid points $\{\tilde{y}_i^t : t = 1, \ldots, T_i\}$. Noting Theorem 4, it follows that this collection of grid points satisfies the assumption of Theorem 3. In Step 4, the value of $\hat{z}$ that we compute by solving the linear program in (3.9) corresponds to the value of z satisfying (3.7). In Step 5, we compute the solution $(\hat{\boldsymbol{w}}_1, \ldots, \hat{\boldsymbol{w}}_m)$ to problem (3.2). By the discussion that follows the proof of Theorem 3, 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 (3.9), the main computational effort in obtaining a $(1+\rho)$-approximate solution to problem (3.2) involves solving a linear program with $1+m$ decision variables and $O(mn + mn \log(n\sigma)/\log(1+\rho))$ constraints.

3.6 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 3.2, 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

$$\begin{aligned}&\Theta(S_1, \ldots, S_m, \boldsymbol{w}_1, \ldots, \boldsymbol{w}_m)\\&= \frac{1}{1 + \sum_{i \in M} \big(\sum_{j \in S_i} w_{ij} \big)^{\gamma_i}} \sum_{i \in M} \Big(\sum_{j \in S_i} w_{ij} \Big)^{\gamma_i} \frac{\sum_{j \in S_i} w_{ij}\, (\kappa_{ij} - \eta_{ij} \log w_{ij})}{\sum_{j \in S_i} w_{ij}}. \quad (3.15)\end{aligned}$$

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 (3.1) and it can be derived by using an argument similar to the one in Section 3.2, 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

$$\begin{aligned}\zeta^* = \max \big\{ &\Theta(S_1, \ldots, S_m, \boldsymbol{w}_1, \ldots, \boldsymbol{w}_m) :\\ &S_i \subset N \;\; \forall\, i \in M,\; \boldsymbol{w}_i \in [\boldsymbol{L}_i, \boldsymbol{U}_i] \;\; \forall\, i \in M \big\}. \quad (3.16)\end{aligned}$$

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 (3.16).

3.6.1 Fixed Point Representation

We begin by giving a fixed point representation of problem (3.16). 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

$$z = \sum_{i \in M} \max_{S_i \subset N, \boldsymbol{w}_i \in [\boldsymbol{L}_i, \boldsymbol{U}_i]} \left\{ \Big(\sum_{j \in S_i} w_{ij} \Big)^{\gamma_i} \frac{\sum_{j \in S_i} w_{ij} \left(\kappa_{ij} - \eta_{ij} \log w_{ij} \right)}{\sum_{j \in S_i} w_{ij}} - \Big(\sum_{j \in S_i} w_{ij} \Big)^{\gamma_i} z \right\}. \tag{3.17}$$

Following the same argument at the beginning of Section 3.3, one can check that if the value of z satisfying (3.17) is given by $\hat{z}$, then we have $\hat{z} = \zeta^*$, where ζ^* is the optimal objective value of problem (3.16). Furthermore, if the value of z satisfying (3.17) 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 (3.17) 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 (3.16). One crucial difficulty associated with solving the maximization problem on the right side of (3.17) 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 (3.17).

Theorem 5. *For any $z > 0$, there exists an assortment $S_i^* \in \mathcal{N}_i$ that solves the maximization problem on the right side of* (3.17)*.*

Proof. For 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 (3.17) 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 (3.17) 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 (3.17), 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

$$\begin{aligned} W_i(\hat{S}_i, \hat{\boldsymbol{w}}_i)^{\gamma_i} (R_i(\hat{S}_i, \hat{\boldsymbol{w}}_i) - z) &= \frac{\sum_{j \in \hat{S}_i} \hat{w}_{ij} \left(\kappa_{ij} - \eta_{ij} \log \hat{w}_{ij} - z\right)}{W_i(\hat{S}_i, \hat{\boldsymbol{w}}_i)^{1-\gamma_i}} \\ &= \frac{\sum_{j \in S_i^*} w_{ij}^* \left(\kappa_{ij} - \eta_{ij} \log w_{ij}^* - z\right) + \hat{w}_{ik} \left(\kappa_{ik} - \eta_{ik} \log \hat{w}_{ik} - z\right)}{W_i(\hat{S}_i, \hat{\boldsymbol{w}}_i)^{1-\gamma_i}}, \quad (3.18) \end{aligned}$$

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 (3.17), 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 9 in the appendix. Since $(S_i^*, \boldsymbol{w}_i^*)$ is feasible to the maximization problem on the right side of (3.17), 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 (3.18), 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 (3.18) by the expression

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

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 (3.18), 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 (3.18), it follows that

$$\frac{\sum_{j\in S_i^*} w_{ij}^*\,(\kappa_{ij}-\eta_{ij}\log w_{ij}^* - z) + \hat{w}_{ik}\,(\kappa_{ik}-\eta_{ik}\log \hat{w}_{ik} - z)}{W_i(\hat{S}_i,\hat{\boldsymbol{w}}_i)^{1-\gamma_i}}$$
$$\geq \frac{(R_i(S_i^*,\boldsymbol{w}_i^*)-z)\,(W_i(S_i^*,\boldsymbol{w}_i^*)+\hat{w}_{ik}\,(1-\gamma_i))}{W_i(S_i^*,\boldsymbol{w}_i^*)^{1-\gamma_i} + (1-\gamma_i)\,W_i(S_i^*,\boldsymbol{w}_i^*)^{-\gamma_i}\,(W_i(\hat{S}_i,\hat{\boldsymbol{w}}_i)-W_i(S_i^*,\boldsymbol{w}_i^*))}. \quad (3.19)$$

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 (3.19), 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, (3.18) and (3.19) 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 (3.17) 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 (3.17) 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 (3.17) 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

$$g_i(S_i,y_i) = \max\left\{\sum_{j\in S_i} w_{ij}\,(\kappa_{ij}-\eta_{ij}\log w_{ij}) : \sum_{j\in S_i} w_{ij} \leq y_i,\ w_{ij}\in[L_{ij},U_{ij}]\ \ \forall j\in N\right\}, \quad (3.20)$$

which is the analogue of problem (3.5), but we focus only on the products in S_i. Similar to the discussion that follows problem (3.5), 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 (3.17), we propose finding the value of z satisfying

$$z = \sum_{i \in M} \max_{S_i \in \mathcal{N}_i,\, y_i \in [\bar{L}_i(S_i), \bar{U}_i(S_i)]} \left\{ y_i^{\gamma_i} \frac{g_i(S_i, y_i)}{y_i} - y_i^{\gamma_i} z \right\}. \tag{3.21}$$

By following the outline of the proof of Proposition 2, it is possible to show that the values of z that satisfy (3.17) and (3.21) are identical to each other and this common value corresponds to the optimal objective value of problem (3.16).

The decision variable S_i on the right side of (3.21) 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 (3.21) 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.

3.6.2 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 (3.21) 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 (3.21), we propose checking the values of y_i only at the grid points and finding the value of z satisfying

$$z = \sum_{i \in M} \max_{S_i \in \mathcal{N}_i,\, y_i \in \{\tilde{y}_i^t(S_i)\,:\, t=1,\ldots,T_i(S_i)\}} \left\{ y_i^{\gamma_i} \frac{g_i(S_i, y_i)}{y_i} - y_i^{\gamma_i} z \right\}. \tag{3.22}$$

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 (3.22) to obtain a good solution for problem (3.16).

Theorem 6. *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* (3.22) *and ζ^* denotes the optimal objective value of problem* (3.16), *then we have $(1+\rho)\,\hat{z} \geq \zeta^*$.*

The theorem above is analogous to Theorem 3 and it can be shown by following the same reasoning in the proof of Theorem 3. By building on this theorem, we can construct an approximate solution to problem (3.16) with a certain performance guarantee. In particular, we find the value of z satisfying (3.22) 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 (3.22) when this problem is solved with $z = \hat{z}$. For all $i \in M$, we solve the knapsack problem in (3.20) 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 (3.16). To see this result, since $\hat{z}$ is the value of z satisfying (3.22) and $(\hat{S}_i, \hat{y}_i)$ is an optimal solution to the maximization problem on the right side of (3.22) when this problem is solved with $z = \hat{z}$, we have

$$\hat{z} = \sum_{i \in M} \left\{ \hat{y}_i^{\gamma_i} \, \frac{g_i(\hat{S}_i, \hat{y}_i)}{\hat{y}_i} - \hat{y}_i^{\gamma_i} \hat{z} \right\}. \tag{3.23}$$

Furthermore, since $\hat{\boldsymbol{w}}_i$ is an optimal solution to problem (3.20) 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 (3.20), since $\hat{y}_i \in [\bar{L}_i(\hat{S}_i), \bar{U}_i(\hat{S}_i)]$, the first constraint in problem (3.20) 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 (3.23), 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 (3.15), 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 6, 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 6, 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 6 is satisfied. It turns out that the answer to this question is already given in Section 3.5. In particular, the only difference between problems (3.5) and (3.20) 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 3.5, 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 (3.22) by solving a linear program similar to the one in (3.9). 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 (3.22). Once we have the value of z that satisfies (3.22), we can follow the approach described right after Theorem 6 to find a $(1 + \rho)$-approximate solution to problem (3.16).

3.7 Computational Experiments

In this section, we test the quality of the solutions obtained by the approximation method that we propose in this chapter. 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.

3.7.1 Experimental Setup

Throughout this section, we refer to our approximation method as APP. In particular, APP uses the algorithm in Section 3.5.3 to find a $(1+\rho)$-approximate solution to problem (3.2). In our computational experiments, we set $\rho = 0.005$ so that APP obtains a solution to problem (3.2) 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 (3.2). 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

$$\min \left\{ z : z \geq \sum_{i \in M} x_i,\ x_i \geq \left(\bar{y}_i^t\right)^{\gamma_i} \frac{g_i(\bar{y}_i^{t+1})}{\bar{y}_i^t} - \left(\bar{y}_i^t\right)^{\gamma_i} z \ \ \forall\, t = 1, \ldots, \tau_i - 1,\ i \in M \right\} \tag{3.24}$$

provides an upper bound on the optimal expected revenue Z^* in problem (3.2). In the linear program above, the decision variables are z and $(x_1, \ldots, x_m)$. Theorem 10 in the appendix 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 (3.24) 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 10 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 4 into problem (3.24) to approximate the optimal expected revenue with a factor of $1+\rho$ accuracy. Once we solve problem (3.24) 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 (3.24) 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 (3.2) 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 (3.2) 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 (3.2) 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 (3.24) 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 4, 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 10 implies that the upper bound provided by the linear program in (3.24) 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.

3.7.2 Computational Results

We give our main computational results in Table 3.1. 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 (3.24) 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 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

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 3.1: Performance of APP.

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 (3.6) 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 (3.6) behaves well when we have $\gamma_i = 1$, avoiding the pathological cases such as the one shown in Figure 3.1. 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 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 3.2, 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 3.1. In each portion of Table 3.2, 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.

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

It is useful to note that naive approaches for finding solutions to problem (3.2) can yield poor results. For example, a first cut approach for finding a solution to problem (3.2) 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 3.3, 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 3.1. The results in Table 3.3 indicate that rounding the unconstrained prices up or down to the price bounds can perform poorly. There are parameter combinations where this 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

Param. Comb.			% Gap with Upp. Bnd.	
m	$[\gamma^L, \gamma^U]$	Δ	Avg.	Max.
5	$[0.05, 0.35]$	1	0.658%	3.798%
5	$[0.05, 0.35]$	2	0.645%	4.285%
5	$[0.05, 0.35]$	3	1.154%	7.765%
5	$[0.35, 0.65]$	1	0.114%	0.213%
5	$[0.35, 0.65]$	2	1.930%	14.749%
5	$[0.35, 0.65]$	3	8.267%	32.248%
5	$[0.65, 1.00]$	1	0.153%	0.214%
5	$[0.65, 1.00]$	2	6.928%	37.986%
5	$[0.65, 1.00]$	3	39.667%	83.798%
Average			6.613%	

Param. Comb.			% Gap with Upp. Bnd.	
m	$[\gamma^L, \gamma^U]$	Δ	Avg.	Max.
10	$[0.05, 0.35]$	1	1.373%	7.347%
10	$[0.05, 0.35]$	2	1.587%	6.777%
10	$[0.05, 0.35]$	3	2.569%	7.360%
10	$[0.35, 0.65]$	1	0.133%	0.180%
10	$[0.35, 0.65]$	2	1.126%	15.492%
10	$[0.35, 0.65]$	3	17.466%	42.090%
10	$[0.65, 1.00]$	1	0.213%	0.845%
10	$[0.65, 1.00]$	2	0.445%	8.797%
10	$[0.65, 1.00]$	3	10.869%	64.471%
Average			3.978%	

Param. Comb.			% Gap with Upp. Bnd.	
m	$[\gamma^L, \gamma^U]$	Δ	Avg.	Max.
15	$[0.05, 0.35]$	1	1.483%	3.613%
15	$[0.05, 0.35]$	2	2.069%	4.575%
15	$[0.05, 0.35]$	3	3.439%	9.326%
15	$[0.35, 0.65]$	1	0.148%	0.210%
15	$[0.35, 0.65]$	2	0.050%	0.512%
15	$[0.35, 0.65]$	3	14.519%	34.727%
15	$[0.65, 1.00]$	1	0.630%	38.811%
15	$[0.65, 1.00]$	2	0.337%	2.133%
15	$[0.65, 1.00]$	3	2.402%	20.070%
Average			2.768%	

Table 3.3: Performance of the prices obtained by rounding the unconstrained prices up or down to the price bounds.

the unconstrained prices give essentially no indication of the optimal prices under price bounds.

3.8 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 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.

CHAPTER 4

A JOINT ASSORTMENT AND STOCKING PROBLEM: CORRELATION-VARIETY RELATIONSHIPS AND APPROXIMATION METHODS

4.1 Introduction

Making effective assortment planning decisions in the face of customer choice requires a strong understanding of customer substitution patterns as well as the benefits and costs of product variety. Of particular interest are settings where a retailer offers a heterogeneous collection of products that can be designated into a number of different *product types* or subgroups, where the products within a type are closer substitutes to each other than products from another type. Both the delineation of products into subgroups and the customers' perception of how similar products within types are can impact customer substation patterns in ways that cannot be adequately captured by assuming that all products belong to a single category. An important question is how the level of similarity of products within a particular type (or multiple types) affects optimal assortment planning and stocking decisions. A retailer offering a single category of highly similar products (in a high-end specialty store, for example) may see benefit by reducing its product line to include only the most popular product variant in this category. On the other hand, massive big-box retailers carrying many different variants of many different product types tend to maintain large levels of product variety and stock large quantities of their offered products. An interesting question is whether observed retail practices of this nature can be supported by analysis of a joint stocking and assortment problem under customer choice.

We study a joint stocking and product offer problem in order to address the questions outlined above. We have access to a set of products that are designated into product types, with individual degrees of product similarity within each type, to satisfy the demand from customers that arrive over a finite selling horizon. Products within a particular type have the same selling price and unit cost. We need to decide which sets of products to offer and how many units of each product to stock. Product demand in our model arises from a stochastic choice process in which individual purchase decisions are made according to a random utility maximization model. The objective is to determine a set of products to offer over the entire selling horizon and how many units of each offered product to stock so as to maximize the expected profit. Our model is static in the sense that customers make their choice from the given assortment without knowledge of product availability and customer do not substitute in the event of a stock-out.

Our chosen random utility maximization model is the nested logit (NL) model. The NL model allows us to accurately reflect that products are designated into subgroups and the resulting implications on customer substitution behavior. Under this model, customers follow a hierarchical choice process, choosing first among subgroups and then a product in the chosen subgroup. A related advantageous feature of the NL model is that its purchase probability expressions contain parameters corresponding to the degree of similarity among products within a particular subgroup, allowing us to examine how varying these parameters impacts the key decisions in our problem.

We formulate our problem as a combinatorial optimization problem, where the objective function is the retailer's expected profit and the decision variables correspond to the subset of products within each type to offer and the stocking quantity for each product. This optimization problem is difficult to work with for a number of reasons. First, the number of decision variables is on the order of the number of subsets of products, which grows exponentially with the number of products. Second, the objective function is not separable by the products and is not concave. Lastly, though a previously-used tactic when dealing with assortment problems under the NL model is to separate the problem by product subgroups, our problem is not separable in this manner.

The contributions of our study are threefold. First, we circumvent the above difficulties by exploiting the structure of the NL model and proving a convenient structural property of optimal assortment offering decisions, using a nonlinear programming relaxation of the original problem. Using this property along with additional structural properties of the nonlinear program, we are able to show that there exist optimal assortments within every product type that correspond to the most popular products in each type. As a result, in each product type, we only need to consider a collection of possible assortments whose size grows linearly with the number of products in each type. To our knowledge, we are the first to provide such a property for a general problem with an arbitrary number of product subgroup designations.

Secondly, we make use of these structural properties of optimal assortments and the nonlinear programming relaxation to examine how the perception of product similarity levels within product types impacts optimal expected profit, assortment offering and product stocking decisions. In particular, we use the relaxation to show that a decrease in product similarity within any one of the offered product types leads to an increase in optimal expected profit. This is a result that we achieve solely through use of the relaxation and its proof does not require any specific characterization of optimal assortments. In addition, we show that as the selection of products within a type becomes increasingly similar, it is optimal to offer very small assortment of that product type. In our numerical experiments, we investigate whether the reverse is true - that optimal offered product variety increases as the selection of products within types become increasingly dissimilar. We also prove that if the assortment of offered products is fixed, a decrease in similarity among products of a particular type leads to higher inventory levels of product stocked within that type. Our numerical experiments suggest that decreasing product similarity also leads to an increase in optimal stocking levels as well as optimal assortment variety.

Lastly, we devise efficient solution approaches for our problem that generate well-performing approximate solutions. We show that our problem can be expressed as a dynamic program with a three-dimensional state variable. We show that we can obtain approximate solutions to our original problem in a tractable fashion via a state space discretization technique, and show that solving an approximate dynamic program provides upper and lower bounds on the optimal expected profit. In addition, we examine a suitably-calibrated deterministic approximation to our original problem, whose analogous dynamic programming formulation has only two state variables and provides a constant-factor approximation to our original problem. Using a similar discretization technique, we are able to use this approximate problem to produce high-quality solutions with a further reduction in computational effort. Our numerical experiments show that solutions generated by the approximate problem tend to perform significantly better than the worst-case performance guarantee.

The aims of this chapter are related to issues examined in consumer psychology regarding product assortment variety. [17] showed that while higher product variety within categories can entice customers to buy from those categories, increasing product variety within a category can provide diminishing returns for the retailer, as many customers who were initially lured in by the product variety would nevertheless decide not to make a purchase, especially when dealing with product categories in which the products were perceived as being very similar or unfamiliar to the customers. [32] show that in unfamiliar product categories, congruency between a customer's shopping goals and the retailer's assortment can lead to lower perceptions of product variety but increased satisfaction with the retailer's assortment, suggesting that customers who choose to buy from a specialty product category with which they are not very familiar may view the products as being highly similar but, should they decide to make a purchase, will gravitate toward a high-utility product within that category.

The NL model, as described by [29] and [7], is a generalization of the multinomial logit (MNL) choice model, providing closed-form expressions for a customer's purchase probability of any offered product, and has been widely used to model customer choice behavior (see [5], [9]). Our work is most strongly related to two other papers on assortment planning and stocking decisions under customer choice. In the first, [39] studies a static assortment and stocking problem where all products belong to a single homogeneous product category, using the MNL model to capture customer substitution behavior. They obtain the optimal solution, showing that it consists of the most popular products from the set of products that can potentially be offered. In order to obtain this result, the authors assume that all of the products have the same unit price to cost ratios. We are able to extend the work of [39] in two crucial areas: the first being that we are allowed to consider the case where we have heterogeneous products that can be designated into an arbitrary number of subgroups with varying levels of subgroup similarity, and the second being that we allow each subgroup to have its own associated price to cost ratio.

In the second, [22] examine a joint assortment planning and pricing problem utilizing a generalized cost function. The authors allow heterogenous products, using the NL model to capture customer choice behavior. The authors consider two arrangements of the customer choice hierarchy. Under the first, a customer selects one of two products brands and then selects a particular product within that brand. Under the second, a customer first chooses a product type and then chooses one of the two brands' variants of that product type. Under the brand-primary choice model, the authors show that it is optimal to offer some subset of the most popular products offered by each brand, and show that this is not necessarily the case under the type-primary choice model. Under the two choice structures, the optimal assortment can be found in $O(n^2)$ and $O(2^n)$ operations respectively, where n is the number of product types. Our work studies a similar assortment optimization problem and incorporates costs in the form of product stocking considerations, but uses exogenous prices. We extend the choice structure to include an arbitrary number of subgroups and products within those subgroups, i.e., as opposed to requiring that there be only two product brands to choose from, there can be an arbitrary number of different brands offered. We show that we can obtain an analogous structural property under this more general framework despite the fact that prices are exogenous and devise efficient approaches for generating well-performing assortments even when the number of product types or the number of subgroup designations is large.

There are a number of other related papers that study product offering and stocking decisions under customer substitution behavior. [19] apply the MNL model to study the joint assortment planning and pricing problem under assortment-based substitution. They show that the optimal assortment for a risk-averse retailer is composed of the variants with the highest quality markups, with price markups being equal to quality markups. [4] apply the MNL model to study a static joint assortment planning and pricing problem in a single product category. They find that the optimal solution is such that all products have equal price to cost ratios. [3] optimize inventory levels and prices for multiple products in a given assortment. [27] develop heuristics for joint assortment, inventory, and price optimization, while [15], [18], and [2] utilize locational choice models to study assortment planning and stocking decisions. [38] studies a joint assortment planning and stocking problem under the MNL model where different assortments may be offered over the course of the selling horizon and gives tractable methods to obtain near-optimal policies. Other papers that consider assortment optimization problems utilizing the NL model with exogenous prices include [10], who studies an unconstrained revenue maximization problem where a single assortment is shown to customers, and [13], which considers a variant of the same problem with capacity constraints on the offered assortment. Both of these papers make use of a structural property that allows the problem to be separated by product subgroups, which is not possible in our problem.

The rest of the chapter is organized as follows. Section 4.2 formulates our joint stocking and assortment planning model. Section 4.3 uses a nonlinear programming relaxation of the original problem to show a structural property of optimal assortments. Section 4.4 gives properties of the optimal solution and examines effects of varying product similarity within particular product types on optimal product variety, stocking decisions and expected profit. Section 4.5 gives a dynamic programming approach for generating approximate solutions and upper and lower bounds on the optimal expected profit. Section 4.6 presents an alternative problem which allows approximate solutions to our original problem to be obtained with reduced computational effort. Section 4.7 presents computational experiments.

4.2 Problem Formulation

Customers arrive at a rate of λ per unit time. For brevity, we normalize the length of the selling horizon to one. The firm's products are designated into m product types, indexed by $M = \{1, \ldots, m\}$. There are n variants of each product type, indexed by $N = \{1, \ldots, n\}$. Each variant of a product type $i \in M$ is assumed to have an identical purchasing cost c_i and revenue p_i, with $p_i \geq c_i$, to reflect the fact that products within a type are assumed to be similar to some degree. In the category/brand setting, for instance, we would assume that versions of the same type of product being offered by different brands should be relatively similar in price and cost.

We let x_{ij} be the number of units of variant j of type i that we stock at the beginning of the selling horizon. The demand for individual variants is influenced by choosing the set of products that we offer over the selling horizon. If we offer the assortment of products $\mathcal{S} = (S_1, \ldots, S_m)$, where $S_i \subseteq N$ is the assortment of variants of type i offered, then the probability that the customer chooses variant j of type i is denoted by $q_{ij}(\mathcal{S})$.

In order to accurately reflect the fact that our firm offers similar products that are grouped together based on type, we assume that $q_{ij}(\mathcal{S})$ is determined according to the nested logit (NL) model. Under the NL, each customer considers the assortment $\mathcal{S}$ of products offered by the store. The customer may choose to purchase a product amongst the set $\mathcal{S}$, or leave without making a purchase. Each customer assigns a random utility $U_{ij} = \mu_{ij} + \epsilon_{ij}$ to each variant $j \in S_i$, where μ_{ij} is a deterministic utility component assigned to each variant $j \in N$ and ϵ_{ijt} has a generalized extreme value (GEV) distribution $\forall i \in M$, $j \in N$. Similarly, each customer assigns a random utility $U_0 = \mu_0 + \epsilon_0$ to the option of purchasing nothing. The utilities of variants of the same type i are assumed to be correlated. We assume that all variants of a particular product type have distinct deterministic utilities. (While it is possible to show that all of the following results also hold under the case where some variants of the same product type have the same deterministic utilities, we will concentrate on the case where utilities are distinct to simplify our analysis.) Without loss of generality, we assume that the variants of each product type are ordered such that $\mu_{i1} > \mu_{i2} > \ldots > \mu_{in}$ for all $i \in M$. Each product type has a parameter $\gamma_i \in (0,1]$ associated with it that characterizes the degree of dissimilarity of products within the type. Equivalently, $1 - \gamma_i$ measures the correlation between the random utilities $(U_{ij})_{j \in N}$ of products within type i.

Let $V_i(S_i) = (\sum_{k \in S_i} e^{\mu_{ik}/\gamma_i})^{\gamma_i}$. Under the NL model, as shown in [7], each customer makes a purchase within product type i with probability $Q_i(\mathcal{S}) = \frac{V_i(S_i)}{v_0 + \sum_{l \in M} V_l(S_l)}$, where $v_0 = e^{\mu_0}$. Given that the customer chooses product type i, he chooses variant j with probability

$$q_{j|i}(S_i) = \begin{cases} \frac{e^{\mu_{ij}/\gamma_i}}{V_i(S_i)^{1/\gamma_i}}, & j \in S_i \\ 0, & \text{otherwise.} \end{cases}$$

Putting these expressions together, we have

$$q_{ij}(\mathcal{S}) = Q_i(\mathcal{S}) q_{j|i}(S_i).$$

We note that $\sum_{j \in N} q_{ij}(\mathcal{S}) = Q_i(\mathcal{S}) \, \forall i \in M$. With probability

$$Q_0(\mathcal{S}) = 1 - \sum_{i \in M} \sum_{j \in N} q_{ij}(\mathcal{S}) = \frac{v_0}{v_0 + \sum_{l \in M} V_l(S_l)},$$

each customer leaves without purchasing anything. Our assumption that there are the same number of variants of each product type is without loss of generality, because if some type i contains fewer than n variants, then we can include additional variants j in this nest with $\mu_{ij} = -\infty$ and these products would never be purchased.

We choose a single assortment of products $\mathcal{S}$ to offer over the entire selling horizon. Our static policy assumes that if a customer chooses a product in the set $\mathcal{S}$ for which we do not have any stock, the customer leaves the system without purchasing anything. One interpretation of this assumption is that a customer's initial decision is a store-visit decision. Customers makes the decision to visit our store based on knowledge of the set $\mathcal{S}$, i.e., the products that they know are being offered, but customers do not know the inventory status of particular products until they arrive at the store. If a customer shows up to the store and his most preferred product is not in stock, the customer will choose to go elsewhere and look for this product rather than making a different choice from among our store. As such, we can interpret the no-purchase option as encompassing all possible purchasing decisions that are outside the control of our firm.

We assume that each customer assigns utilities to the products in the subset $\mathcal{S}$ based on independent samples of the nested logit model described above, that is, realizations of the random variables ϵ_{ij}, $j \in N$, $i \in M$ are independent from customer to customer. Since customers are heterogeneous and independent, the observation of one customer's choice reveals no additional information about the choice of subsequent customers. This results in the following model of aggregate demand: the number of customers that request variant j of type i over the selling horizon has a normal distribution with mean $\lambda q_{ij}(\mathcal{S})$ and standard deviation $\alpha\lambda^r q_{ij}(\mathcal{S})^r$, where $\alpha > 0$ and $r \in [0, 1)$. This distribution is chosen to ensure that our problem framework satisfies economies of scale, in the sense that operational efficiency increases as the customer volume λ gets larger. Particularly, for a fixed assortment $\mathcal{S}$, we note that the coefficient of variation of the demand for variant j of type i is $\frac{\alpha}{\lambda^{1-r} q_{ij}(\mathcal{S})^{1-r}}$, which is decreasing in the customer volume λ, implying that inventory costs associated with safety stock should decrease as λ gets larger. A special case of this model is when the total number of customers that arrive over the selling horizon is Poisson with mean λ. In this case, the number of customers that request product j in nest i is also Poisson with mean $\lambda q_{ij}(\mathcal{S})$, and the normal approximation to the Poisson distribution yields $\alpha = 1$ and $r = 1/2$.

Using $\mathcal{N}(\mu, \sigma)$ to denote a normal random variable with mean μ and standard deviation σ, we can maximize the expected profit by solving the problem

$$\max_{\mathcal{S},\mathbf{x}} \sum_{i\in M}\sum_{j\in N} p_i \mathbb{E}\big[\min\big\{\mathcal{N}(\lambda q_{ij}(\mathcal{S}), \alpha\lambda^r q_{ij}(\mathcal{S})^r), x_{ij}\big\}\big] - \sum_{i\in M}\sum_{j\in N} c_i x_{ij} \tag{4.1}$$

$$\text{s.t. } \mathcal{S} = (S_1, \ldots, S_m),\ S_i \subseteq N\ \forall i \in M \tag{4.2}$$

$$x_{ij} \geq 0 \ \ \forall j \in N, \forall i \in M, \tag{4.3}$$

where the first term in the objective is the expected revenue obtained from purchases by customers and the second term is the cost of stocking the products. We note that given a fixed assortment $\mathcal{S}$ of products, the demand distribution is exogenous and it is well-known that the optimal stocking decisions for problem (4.1)-(4.3) are given by

$$x^*_{ij}(\mathcal{S}) = \lambda q_{ij}(\mathcal{S}) + \alpha\lambda^r q_{ij}(\mathcal{S})^r \Phi^{-1}(1 - c_i/p_i), \tag{4.4}$$

where $\Phi^{-1}(\cdot)$ is the inverse of the standard normal distribution function. We can collapse these expressions back into the objective function, resulting in

$$p_i\mathbb{E}\big[\min\big\{\mathcal{N}(\lambda q_{ij}(\mathcal{S}),\alpha\lambda^r q_{ij}(\mathcal{S})^r),x^*_{ij}(\mathcal{S})\big\}\big]-c_i x^*_{ij}(\mathcal{S})$$
$$=(p_i-c_i)\lambda q_{ij}(\mathcal{S})-p_i\alpha\lambda^r\phi(\Phi^{-1}(1-c_i/p_i))q_{ij}(\mathcal{S})^r\;\forall i\in M,\,j\in N,$$

where $\phi(\cdot)$ denotes the standard normal density and we use the fact that $\mathbb{E}[\min\{Z,\Phi^{-1}(1-c_i/p_i)\}]=\Phi^{-1}(1-c_i/p_i)c_i/p_i-\phi(\Phi^{-1}(1-c_i/p_i))$ for a standard normal random variable Z; see [21]. For notational brevity, let $\theta_i=p_i\alpha\lambda^r\phi(\Phi^{-1}(1-c_i/p_i))$. Then problem (4.1)-(4.3) is equivalent to

$$\max_{\mathcal{S}}\sum_{i\in M}\sum_{j\in N}(p_i-c_i)\lambda q_{ij}(\mathcal{S})-\sum_{i\in M}\sum_{j\in N}\theta_i q_{ij}(\mathcal{S})^r \tag{4.5}$$

$$\text{s.t. } \mathcal{S}=(S_1,\dots,S_m),\;S_i\subseteq N\;\forall i\in M. \tag{4.6}$$

We note that the problem of choosing optimal stocking levels is now embedded into the objective function, so that the only decision of note is choosing the optimal assortment of products to offer; the corresponding optimal stock levels can then be computed using (4.4). However, problem (4.5)-(4.6) is difficult for a number of reasons. The number of possible choices for each subset S_i is 2^n, meaning that there are 2^{nm} possible choices for the assortment $\mathcal{S}$. In addition, interactions between the assortment decisions S_i for different product types i in the terms $q_{ij}(\mathcal{S})$ and $q_{ij}(\mathcal{S})^r$ prevent the problem from being separable by the product types $i\in M$, which prevents separation-based approaches for assortment problems such as those studied in [10] and [13].

In the next section, we give a structural property of optimal assortments for problem (4.5)-(4.6) using a nonlinear programming relaxation of this problem. This nonlinear programming relaxation will be influential in allowing us to gain insights on how optimal solutions to our original problem behave with changes in the utility correlation parameters in later sections.

4.3 Optimal Assortments

Recall that within each product type, variants are ordered by their deterministic utility components μ_{ij}, such that variant 1 has the highest deterministic utility and variant n has the lowest. If we suppose that the entire set of variants N is offered within a nest i, then our expressions for the purchase probabilities $q_{ij}(\mathcal{S})$ imply that we will have $q_{i1}(\mathcal{S}) > q_{i2}(\mathcal{S}) \ldots > q_{in}(\mathcal{S})$. Thus, if all variants of type i are shown to customers, variant 1 will have the highest mean demand and variant n the lowest mean demand. We can simplify this concept by simply saying that the variants in each nest are ordered by their expected popularity. Define the popular set

$$\mathcal{P} = \big\{\emptyset, \{1\}, \{1,2\}, \ldots, \{1, \ldots, n\}\big\}$$

to be the collection of all assortments containing the k most popular variants for $k = 0, \ldots, n$. We assume that all variants of a particular product type have distinct deterministic utilities, i.e., $\mu_{i1} > \ldots > \mu_{in} \, \forall i \in M$. We will show that optimal assortments have the property that assortments of individual product types must fall in the popular set $\mathcal{P}$. (While it is also possible to show that there exist optimal assortments in the popular set when some product variants can have the same deterministic utility, we will focus on the case where utilities are distinct to simplify our analysis.)

To achieve this result, we will examine a relaxation of problem (4.5)-(4.6). In the relaxation, we let Y_i represent the probability that a customer chooses a product from type i, y_{ij} the probability that a customer chooses variant j of product type i and Y_0 the probability that the customer buys nothing. These decision variables must satisfy the constraints $\sum_{j \in N} y_{ij} = Y_i \, \forall i \in M$ and $Y_0 + \sum_{i \in M} Y_i = 1$. We note that the probability that a customer purchases variant j of product type i is $q_{ij}(\mathcal{S}) = \frac{V_i(S_i)Q_0(\mathcal{S})}{v_0}\frac{e^{\mu_{ij}/\gamma_i}}{V_i(S_i)^{1/\gamma_i}}$ whenever $j \in S_i$ and zero otherwise. Thus, when $j \in S_i$, $q_{ij}(\mathcal{S})^{\gamma_i}e^{-\mu_{ij}} = \frac{V_i(S_i)^{\gamma_i}Q_0(\mathcal{S})^{\gamma_i}}{v_0^{\gamma_i}}\frac{1}{V_i(S_i)}$, so that $v_0V_i(S_i)^{1-\gamma_i}(\frac{Q_0(\mathcal{S})}{v_0})^{1-\gamma_i}q_{ij}(\mathcal{S})^{\gamma_i}e^{-\mu_{ij}} = Q_0(\mathcal{S})$ when $j \in S_i$ and zero otherwise. Therefore, noting that $V_i(S_i)^{1-\gamma_i}(\frac{Q_0(\mathcal{S})}{v_0})^{1-\gamma_i} = Q_i(\mathcal{S})^{1-\gamma_i}$, we always have $v_0Q_i(\mathcal{S})^{1-\gamma_i}q_{ij}(\mathcal{S})^{\gamma_i}e^{-\mu_{ij}} \leq Q_0(\mathcal{S})$ for all $\mathcal{S}$, $j \in S_i$, $i \in M$ and it is reasonable to impose the constraint $v_0Y_i^{1-\gamma_i}y_{ij}^{\gamma_i}e^{-\mu_{ij}} \leq Y_0$ for all $i \in M$, $j \in N$. We propose maximizing our expected profit by solving the problem

$$\max_{(\mathbf{y},\mathbf{Y},Y_0)\in\mathcal{F}} \sum_{i\in M}\sum_{j\in N}(p_i - c_i)\lambda y_{ij} - \sum_{i\in M}\sum_{j\in N}\theta_i y_{ij}^r, \tag{4.7}$$

where $\mathcal{F}$ is the feasible region defined by the constraints

$$\begin{aligned} Y_0 + \sum_{i\in M} Y_i &= 1 && \\ \sum_{j\in N} y_{ij} &= Y_i && \forall i \in M \\ v_0Y_i^{1-\gamma_i}y_{ij}^{\gamma_i}e^{-\mu_{ij}} &\leq Y_0 && \forall j \in N, i \in M \\ y_{ij} &\geq 0 && \forall j \in N, i \in M. \end{aligned}$$

We emphasize that this is a non-convex optimization problem. While the constraints become linear if $\gamma_i = 1$ for all i and are consistent with the linear representation of the MNL model studied in [38], this problem lacks desirable structure for general values of γ_i. Despite the fact that we cannot easily solve this problem, we can use it to derive structural properties of optimal assortments.

It is clear that problem (4.7) is a relaxation of problem (4.5)-(4.6), since any assortment $\mathcal{S}$ defines a solution that satisfies all the constraints, as shown above. However, it is possible to construct feasible solutions to this problem that do not correspond to actual purchase probabilities under the nested logit model; correspondence can only happen if the last constraint is satisfied with equality for all

j such that y_{ij} is positive and its left side is zero otherwise, for all $i \in M$. We can equivalently state that the following condition is necessary for a solution to correspond to purchase probabilities:

Condition 1. $v_0 Y_i^{1-\gamma_i} y_{ij}^{\gamma_i} e^{-\mu_{ij}} = Y_0 \boldsymbol{1}(y_{ij} > 0)\, \forall i \in M, j \in N.$

Remarkably, it turns out that optimal solutions to the relaxation (4.7) actually satisfy this necessary condition. To show this, we will make use of a linearization technique. Let $(\mathbf{y}^*, \mathbf{Y}^*, Y_0^*)$ be an optimal solution to problem (4.7) and let $S_i^* = \{j \in N : y_{ij}^* > 0\}$. Since $(\cdot)^{\gamma_i}$ is a concave function, we have

$$\begin{aligned} Y_i^{1-\gamma_i} y_{ij}^{\gamma_i} = Y_i (y_{ij}/Y_i)^{\gamma_i} &\le Y_i \left\{ (y_{ij}^*/Y_i^*)^{\gamma_i} + \gamma_i (y_{ij}^*/Y_i^*)^{\gamma_i - 1} (y_{ij}/Y_i - y_{ij}^*/Y_i^*) \right\} \\ &= (1-\gamma_i)(y_{ij}^*/Y_i^*)^{\gamma_i} Y_i + \gamma_i (y_{ij}^*/Y_i^*)^{\gamma_i - 1} y_{ij}, \end{aligned}$$

where in the inequality, we have replaced $(y_{ij}/Y_i)^{\gamma_i}$ with the linear approximation given by the derivative of $(\cdot)^{\gamma_i}$ at the point y_{ij}^*/Y_i^* (an application of the subgradient inequality). Letting $\mathcal{H}(\mathbf{y}^*, \mathbf{Y}^*, Y_0^*)$ be the region defined by the linear constraints

$$\begin{aligned} Y_0 + \sum_{i \in M} Y_i &= 1 \\ \sum_{j \in N} y_{ij} &= Y_i && \forall i \in M \\ v_0 e^{-\mu_{ij}} \gamma_i (y_{ij}^*/Y_i^*)^{\gamma_i - 1} y_{ij} + v_0 (1-\gamma_i) e^{-\mu_{ij}} (y_{ij}^*/Y_i^*)^{\gamma_i} Y_i &\le Y_0 && \forall j \in S_i^*, \forall i \in M \\ y_{ij} &= 0 && \forall j \in N \setminus S_i^*, \forall i \in M \\ y_{ij} &\ge 0 && \forall j \in N, i \in M. \end{aligned}$$

it follows that $\mathcal{H}(\mathbf{y}^*, \mathbf{Y}^*, Y_0^*) \subseteq \mathcal{F}$. It also follows that $(\mathbf{y}^*, \mathbf{Y}^*, Y_0^*) \in \mathcal{H}(\mathbf{y}^*, \mathbf{Y}^*, Y_0^*)$, since the first constraint is satisfied with equality when $\mathbf{y} = \mathbf{y}^*$. Thus, $(\mathbf{y}^*, \mathbf{Y}^*, Y_0^*)$ must be optimal for the problem

$$\max_{(\mathbf{y}, \mathbf{Y}, Y_0) \in \mathcal{H}(\mathbf{y}^*, \mathbf{Y}^*, Y_0^*)} \sum_{i \in M} \sum_{j \in N} (p_i - c_i) \lambda y_{ij} - \sum_{i \in M} \sum_{j \in N} \theta_i y_{ij}^r, \tag{4.8}$$

This problem is useful because it involves maximizing a strictly convex objective function over a set of linear constraints. As such, optimal solutions to this problem must be extreme points of the feasible region. The following lemma makes use of this fact to obtain the desired result.

Lemma 5. *Any optimal solution* $(\boldsymbol{y}^*, \boldsymbol{Y}^*, Y_0^*)$ *to problem* (4.7) *must satisfy Condition 1.*

Proof. Note that $\mathcal{H}(\mathbf{y}^*, \mathbf{Y}^*, Y_0^*)$ is a polyhedron. Since the objective function of problem 4.8 is strictly convex in $\mathbf{y}$, $(\mathbf{y}^*, \mathbf{Y}^*, Y_0^*)$ must be an extreme point of $\mathcal{H}(\mathbf{y}^*, \mathbf{Y}^*, Y_0^*)$. Consider the optimization problem in (4.8). Letting $S_i^* = \{j \in N : y_{ij}^* > 0\}$, we can disregard the variables y_{ij} for $j \notin S_i^*$, as they have no effect on the objective function value. Introducing slack variables s_{ij}, the constraints of $\mathcal{H}(\mathbf{y}^*, \mathbf{Y}^*, Y_0^*)$ can now be expressed as

$$v_0\Big\{e^{-\mu_{ij}}\gamma_i(y_{ij}^*/Y_i^*)^{\gamma_i-1}y_{ij} + (1-\gamma_i)e^{-\mu_{ij}}(y_{ij}^*/Y_i^*)^{\gamma_i}Y_i\Big\} - Y_0 + s_{ij} = 0$$

$$\sum_{j\in S_i^*} y_{ij} - Y_i = 0$$

$$\sum_{i\in M} Y_i + Y_0 = 1$$

$$y_{ij}, s_{ij} \geq 0$$

for all $j \in S_i^*$, $i \in M$. The variables Y_i and Y_0 are superfluous, so we can eliminate them and disregard the second and third constraints, provided we simply re-express Y_i and Y_0 in the first constraint in terms of the variables $\mathbf{y}$. Among the $\mathbf{y}$ variables, we have $\sum_{i\in M}|S_i^*|$ variables and $\sum_{i\in M}|S_i^*|$ constraints. We know that $(\mathbf{y}^*, \mathbf{s}^*)$ must be an extreme point of this polyhedron, where $\mathbf{s}^*$ is the vector of slack variables corresponding to $\mathbf{y}^*$, taking only the indices $j \in S_i^*$ into consideration. We recall that $y_{ij}^* > 0\ \forall j \in S_i^*$. Thus, the solution $(\mathbf{y}^*, \mathbf{s}^*)$ must contain at least

$\sum_{i \in M} |S_i^*|$ positive variables. But this implies that none of the slack variables $\mathbf{s}^*$ can be positive, because there are only $\sum_{i \in M} |S_i^*|$ constraints in the region above. Thus, we must have $v_0(1-\gamma_i)e^{-\mu_{ij}}(y_{ij}^*/Y_i^*)^{\gamma_i}Y_i^* + v_0 e^{-\mu_{ij}}\gamma_i(y_{ij}^*/Y_i^*)^{\gamma_i - 1}y_{ij}^* - Y_0^* = 0$ for all $j \in S_i^*$, $\forall i \in M$. Since the left side of this equality is equal to $v_0(Y_i^*)^{1-\gamma_i}(y_{ij}^*)^{\gamma_i}e^{-\mu_{ij}} - Y_0^*$, we conclude that $(\mathbf{y}^*, \mathbf{Y}^*, Y_0^*)$ must satisfy Condition 1. □

We can now show that optimal solutions to the relaxation (4.7) correspond to assortments in the popular set. To see this, suppose that we have an optimal solution $(\mathbf{y}^*, \mathbf{Y}^*, Y_0^*)$ to problem (4.7). Note that since $(\mathbf{y}^*, \mathbf{Y}^*, Y_0^*)$ satisfies Condition 1, we must have

$$v_0(Y_i^*)^{1-\gamma_i}(y_{ij}^*)^{\gamma_i}e^{-\mu_{ij}} = Y_0^* \, \forall j \in S_i^*$$

and $y_{ij}^* = 0 \, \forall j \in N \setminus S_i^*$. Now suppose that $S_i^* = \{j \in N : y_{ij}^* > 0\} \notin \mathcal{P}$ for some $i \in M$. Then $\exists \, s, t \in N$, $s < t$ such that $\mu_{is} > \mu_{it}$, $s \notin S_i^*$ and $t \in S_i^*$. Thus, $y_{is}^* = 0$ and

$$v_0(Y_i^*)^{1-\gamma_i}(y_{it}^*)^{\gamma_i}e^{-\mu_{it}} = Y_0^*.$$

Now suppose that we construct a new solution $\hat{\mathbf{y}}$ by swapping the variables corresponding to variants s and t, i.e., letting $\hat{y}_{is} = y_{it}^*$, $\hat{y}_{it} = 0$ and $\hat{y}_{ij} = y_{ij}^* \, \forall j \neq s, t$. This results in $Y_0^* = 1 - \sum_{l \in M} \sum_{k \in N} y_{lk}^* = 1 - \sum_{l \in M} \sum_{k \in N} \hat{y}_{lk}$ and $Y_i^* = \sum_{k \in N} y_{ik}^* = \sum_{k \in N} \hat{y}_{ik} \, \forall i \in M$, so $(\hat{\mathbf{y}}, \mathbf{Y}^*, Y_0^*)$ is feasible for problem (4.7) and has the same objective value as $(\mathbf{y}^*, \mathbf{Y}^*, Y_0^*)$ (implying that it is optimal for problem (4.7).) However,

$$v_0(Y_i^*)^{1-\gamma_i}(\hat{y}_{is})^{\gamma_i}e^{-\mu_{is}} = Y_0^* e^{\mu_{it} - \mu_{is}} < Y_0^*,$$

implying that $(\hat{\mathbf{y}}, \mathbf{Y}^*, Y_0^*)$ fails to satisfy Condition 1. This contradicts Lemma 5. Thus, $S_i^* = \{j \in N : y_{ij}^* > 0\}$ must lie in $\mathcal{P}$ for all $i \in M$. Since we have only shown thus far that optimal solutions to problem (4.7) meet a necessary condition for correspondence to purchase probabilities, it still remains to be shown that we can construct an optimal solution to our original assortment problem using an optimal solution to the relaxation. This is done in the proof of the following theorem, which characterizes the main result of this section.

Theorem 7. *There exists an optimal assortment* $\mathcal{S}^* = (S_1^*, \ldots, S_m^*)$ *to problem* (4.5)-(4.6) *such that* $S_i^* \in \mathcal{P}$ *for all* $i \in M$.

Proof. By the preceding analysis, we know that there exists an optimal solution $(\mathbf{y}^*, \mathbf{Y}^*, Y_0^*)$ to problem (4.7) such that $S_i^* = \{j \in N : y_{ij}^* > 0\} \in \mathcal{P}$ for all $i \in M$. Let $\mathcal{S}^* = (S_1^*, \ldots, S_m^*)$. Since $(\mathbf{y}^*, \mathbf{Y}^*, Y_0^*)$ must satisfy Condition 1, we have

$$V_i(S_i^*) = \left(\sum_{j: y_{ij}^* > 0} e^{\mu_{ij}/\gamma_i} \right)^{\gamma_i} = \left(\sum_{j: y_{ij}^* > 0} y_{ij}^* \frac{v_0^{1/\gamma_i} (Y_i^*)^{1/\gamma_i - 1}}{(Y_0^*)^{1/\gamma_i}} \right)^{\gamma_i} = \frac{v_0 Y_i^*}{Y_0^*},$$

giving

$$\begin{aligned} q_{ij}(\mathcal{S}^*) = \frac{e^{\mu_{ij}/\gamma_i}}{(v_0 + \sum_{l \in M} V_l(S_l^*)) V_i(S_i^*)^{1/\gamma_i - 1}} &= \frac{e^{\mu_{ij}/\gamma_i}}{(v_0 + \sum_{l \in M} \frac{v_0 Y_l^*}{Y_0^*})(\frac{v_0 Y_i^*}{Y_0^*})^{1/\gamma_i - 1}} \\ &= e^{\mu_{ij}/\gamma_i} \left(\frac{Y_0^*}{v_0} \right)^{1/\gamma_i} (Y_i^*)^{1 - 1/\gamma_i} = y_{ij}^*, \end{aligned}$$

where the second-to-last equality follows from the fact that $1 + \sum_{l \in M} \frac{Y_l^*}{Y_0^*} = \frac{1}{Y_0^*}$ and the last equality follows from the fact that $(\mathbf{y}^*, \mathbf{Y}^*, Y_0^*)$ satisfies Condition 1. Thus, the assortment $\mathcal{S}^*$ provides the same objective value as the solution $(\mathbf{y}^*, \mathbf{Y}^*, Y_0^*)$. Since $(\mathbf{y}^*, \mathbf{Y}^*, Y_0^*)$ is optimal for the relaxation (4.7), $\mathcal{S}^*$ must be optimal for problem (4.5)-(4.6). □

Theorem 7 implies that there is always an optimal assortment of each product type $i \in M$ which is in the popular set $\mathcal{P}$. A similar structure was shown to be optimal for a single product category (type) under the MNL model by van [39]; [22] established the optimality of this structure for a problem under the NL model with two nests (product types in our case). Our result generalizes this structure to an arbitrary number of product types.

When trying to construct an assortment $\mathcal{S}$ that is optimal for problem (4.5)-(4.6), we need only consider assortments of each product type that are in the popular set $\mathcal{P}$. We note that there are only $n+1$ assortments in $\mathcal{P}$, as opposed to 2^n possible assortments in total, which greatly reduces the number of candidate assortments that need to be considered by the firm when n is large. This property allows us to solve our original problem via a dynamic program, as we will show in Section 4.5, as well as obtain upper and lower bounds on the optimal expected profit.

4.4 Properties of the Optimal Solution

In this section, we examine the effects of utility correlation among variants of a product type, captured in the form of the parameters γ_i, on the stocking and assortment offering decisions made by our firm. Throughout this section, in order to emphasize the dependency of the relevant expressions in our problem on the parameters $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_m)$, we will use $V_i(S_i, \gamma_i)$ to denote $V_i(S_i)$ and $x^*_{ij}(\mathcal{S}, \boldsymbol{\gamma})$ to denote $x^*_{ij}(\mathcal{S})$; $Q_i(\mathcal{S}, \boldsymbol{\gamma})$ and $q_{j|i}(S_i, \gamma_i)$ are defined similarly.

To examine the effects of utility correlation on our expected profit, we revisit our nonlinear programming relaxation in problem (4.7). We know from theorem 1 that the optimal objective values of problem (4.7) and problem (4.5)-(4.6) are equivalent. Consider the feasible region of problem (4.7). Emphasizing its dependency on the correlation parameters, let $\mathcal{F}(\boldsymbol{\gamma})$ denote the feasible region of (4.7).

We can express the constraints of $\mathcal{F}(\boldsymbol{\gamma})$ as

$$
\begin{aligned}
Y_0 + \sum_{i \in M} Y_i = 1 & & \\
\sum_{j \in N} y_{ij} = Y_i & & \forall i \in M \\
f_{ij}(\mathbf{y}, \mathbf{Y}, \gamma_i) \leq Y_0 & & \forall j \in N, i \in M \\
y_{ij} \geq 0 & & \forall j \in N, i \in M.
\end{aligned}
$$

where $f_{ij}(\mathbf{y}, \mathbf{Y}, \gamma_i) = v_0 Y_i^{1-\gamma_i} y_{ij}^{\gamma_i} e^{-\mu_{ij}} \, \forall i \in M, j \in N$. We observe that for any fixed feasible $(\mathbf{y}, \mathbf{Y})$, the derivative of f_{ij} with respect to γ_i is given by

$$
\frac{\partial f_{ij}}{\partial \gamma_i}(\mathbf{y}, \mathbf{Y}, \gamma_i) = v_0 Y_i^{1-\gamma_i} y_{ij}^{\gamma_i} e^{-\mu_{ij}} (\log y_{ij} - \log Y_i) \leq 0.
$$

This implies that, for fixed values of $(\mathbf{y}, \mathbf{Y})$, increasing the value of γ_i decreases the value of the left side of the constraint. Thus, if $(\mathbf{y}, \mathbf{Y}, Y_0) \in \mathcal{F}(\boldsymbol{\gamma}_{-i}, \gamma_i')$ and $\bar{\gamma}_i > \gamma_i'$, then $f_{ij}(\mathbf{y}, \mathbf{Y}, \bar{\gamma}_i) \leq f_{ij}(\mathbf{y}, \mathbf{Y}, \gamma_i') \leq Y_0 \, \forall j \in N$, implying that $(\mathbf{y}, \mathbf{Y}, Y_0) \in \mathcal{F}(\boldsymbol{\gamma}_{-i}, \bar{\gamma}_i)$. Therefore $\mathcal{F}(\boldsymbol{\gamma}_{-i}, \gamma_i') \subseteq \mathcal{F}(\boldsymbol{\gamma}_{-i}, \bar{\gamma}_i)$, and any solution that is feasible to problem (4.7) with parameters $(\boldsymbol{\gamma}_{-i}, \gamma_i')$ must be feasible for problem (4.7) with parameters $(\boldsymbol{\gamma}_{-i}, \bar{\gamma}_i)$. As a result, we have the following theorem:

Theorem 8. *Let* $\Pi^*(\boldsymbol{\gamma})$ *denote the optimal objective value of problem* (4.5)-(4.6) *with correlation parameters* $\boldsymbol{\gamma}$. *For all* $i \in M$, *if* $\bar{\gamma}_i > \gamma_i'$, *then* $\Pi^*(\boldsymbol{\gamma}_{-i}, \bar{\gamma}_i) \geq \Pi^*(\boldsymbol{\gamma}_{-i}, \gamma_i')$ *for all* $\boldsymbol{\gamma}_{-i} \in (0, 1]^{m-1}$.

Proof. Follows from equivalence of the optimal objective values of problem (4.5)-(4.6) and problem (4.7). □

Thus, as the utilities of products within types become less correlated, our firm's expected profits increase. We emphasize that, to our knowledge, we are the first to demonstrate this property for an assortment optimization problem under the

NL model. We also emphasize that the use of the nonlinear programming representation in (4.7) of the original problem is instrumental in achieving this result. In addition, since our proof relies only on properties of the feasible region of the nonlinear program, we could replace the objective function in (4.7) with any other objective function that can be expressed in terms of the decision variables $(\mathbf{y}, \mathbf{Y}, Y_0)$ (i.e., the purchase probabilities under the NL model) and show that decreased correlation among the utilities of products in any particular type leads to an improvement in optimal performance.

The following proposition characterizes how customers evaluate products within a particular type i as the degree of correlation between utilities of variants of that product type becomes very high. This will allow us to obtain some insights on how optimal stocking levels change with utility correlation under a fixed assortment, as well as prove an asymptotic property of optimal assortments. The proof is deferred to the appendix.

Proposition 3. *For any fixed assortment* $S_i \subseteq N$, $\lim_{\gamma_i \to 0} V_i(S_i, \gamma_i) = e^{\max_{k \in S_i} \mu_{ik}}$ *and*

$\lim_{\gamma_i \to 0} q_{j|i}(S_i, \gamma_i) = \mathbf{1}(j = \text{argmax}_{k \in S_i} \mu_{ik})$.

Consider the case when the assortment of products $\mathcal{S}$ is fixed. We have shown in Proposition 3 that as products within a type become increasingly correlated, customers who choose product type i will gravitate toward the highest-utility variant of that type (and will select product type i based solely on the attractiveness of the highest-utility variant, ignoring other variants of the type.) If products within a type become less correlated, we would expect the opposite behavior : customers would start to take the utilities of additional variants of product type i into account. We might expect that this wold increase the probability that a customer buys a product from type i. In fact, this can be shown mathematically by noting that

$$\frac{\partial}{\partial \gamma_i} V_i(S_i, \gamma_i) = V_i(S_i, \gamma_i) \left[\log(\sum_{k \in S_i} e^{\mu_{ik}/\gamma_i}) - \frac{\sum_{k \in S_i} \mu_{ik} e^{\mu_{ik}/\gamma_i}}{\gamma_i \sum_{k \in S_i} e^{\mu_{ik}/\gamma_i}} \right] \geq 0$$

since

$$\log(\sum_{k \in S_i} e^{\mu_{ik}/\gamma_i}) \geq \frac{\max_{k \in S_i} \mu_{ik}}{\gamma_i} \geq \frac{\sum_{k \in S_i} \mu_{ik} e^{\mu_{ik}/\gamma_i}}{\gamma_i \sum_{k \in S_i} e^{\mu_{ik}/\gamma_i}}.$$

This implies that $\frac{\partial}{\partial \gamma_i} Q_i(\mathcal{S}, \boldsymbol{\gamma}) \geq 0$. Raising γ_i increases the aggregate attractiveness of product i, but also increases the chance that a customer will choose a variant of product type i that is not the highest-utility variant. Naturally, we would expect that this necessitates carrying more total inventory of products in type i. Meanwhile, the fact that customers find more potential value in lower-utility variants of product type i will lure some customers who would otherwise choose other product types to consider type i. This naturally should necessitate carrying less inventory of products in other types. The following result formalizes these notions: as products within a type i become less correlated, the optimal total stock of type i increases while the optimal total stock of all other types $l \neq i$ decreases. The proof is deferred to the appendix.

Lemma 6. *For all* $i \in M$, *if* $\bar{\gamma}_i > \gamma'_i$, *then* $\sum_{j \in N} x^*_{ij}(\mathcal{S}, \bar{\gamma}_i, \boldsymbol{\gamma}_{-i}) \geq \sum_{j \in N} x^*_{ij}(\mathcal{S}, \gamma'_i, \boldsymbol{\gamma}_{-i})$ *and* $\sum_{j \in N} x^*_{lj}(\mathcal{S}, \bar{\gamma}_i, \boldsymbol{\gamma}_{-i}) \leq \sum_{j \in N} x^*_{lj}(\mathcal{S}, \gamma'_i, \boldsymbol{\gamma}_{-i}) \, \forall l \neq i$, *for all* $\boldsymbol{\gamma}_{-i} \in (0, 1]^{m-1}$.

Lastly, the following asymptotic result characterizes optimal assortments of a product type i as γ_i tends to zero (i.e.,, as the utilities of variants of product type i become increasingly correlated.) The proofs of Proposition 3, Lemma 6 and Lemma 7 are deferred to the appendix.

Lemma 7. *As γ_i approaches zero, there exists an optimal assortment $S_i^* \in \{\emptyset, \{1\}\}$ of product type i for problem* (4.5)-(4.6).

Lemma 7 implies that, as the utilities of variants of a type i become increasingly correlated, it becomes optimal to either offer no products of type i, or only the most popular product of type i. This lines up with our observations that as customers will gravitate toward the highest-utility variant of product type i as γ_i becomes very small, and the presence or absence of this highest-utility variant will become the only factor which influences a customer's decision to buy from that product type. As such, when choosing amongst a number of extremely similar product variants of one type that can potentially be offered, our decision boils down to whether or not the most popular variant should offered, and all other variants of the product type can be disregarded. One might expect the opposite to be true, i.e., that as γ_i increases, assortment variety in product type i will become a larger factor. We anticipate that Lemma 7 indicates a general trend in optimal assortments with changes in γ_i values, namely that as γ_i increases, it usually becomes optimal to offer larger assortments of products of type i. We leave further investigation of effects of utility correlation on optimal assortment sizes to our numerical experiments.

4.5 Dynamic Program

We can exploit the fact that optimal assortments are popularity-ordered in each product type to devise a solution approach for problem (4.5)-(4.6). Plugging in the exact expressions for $q_{ij}(\mathcal{S})$, the objective function in (4.5) is equal to

$$\lambda \frac{\sum_{i \in M} (p_i - c_i) V_i(S_i)}{v_0 + \sum_{i \in M} V_i(S_i)} - \frac{\sum_{i \in M} \theta_i (\sum_{j \in S_i} e^{r\mu_{ij}/\gamma_i}) V_i(S_i)^{r(1-1/\gamma_i)}}{(v_0 + \sum_{i \in M} V_i(S_i))^r}, \tag{4.9}$$

where we use the convention $0/0 := 0$. Let $\Theta_i(S_i) = \theta_i(\sum_{j\in S_i} e^{r\mu_{ij}/\gamma_i})V_i(S_i)^{r(1-1/\gamma_i)}$. We can maximize this function over assortments $\mathcal{S}$ via a dynamic program with three state variables, where the decision epochs are the product types $i \in M$ and the action variable in each decision epoch is the assortment $S_i \in \mathcal{P}$; we only need to consider assortments in the popular set for each product type. The profit (4.9) is computed after the final decision epoch m, with the first, second and third state variables representing the sums $\sum_{i\in M}(p_i - c_i)V_i(S_i)$, $\sum_{i\in M} V_i(S_i)$ and $\sum_{i\in M} \Theta_i(S_i)$, respectively, corresponding to the assortment decisions made in each of the previous epochs. The optimality equation is given by

$$J_i(z_1, z_2, z_3) = \max_{S_i\in\mathcal{P}} J_{i+1}\big(z_1 + (p_i - c_i)V_i(S_i), z_2 + V_i(S_i), z_3 + \Theta_i(S_i)\big) \tag{4.10}$$

for all $i \in M$,where the boundary condition is

$$J_{m+1}(z_1, z_2, z_3) = \lambda\frac{z_1}{v_0 + z_2} - \frac{z_3}{(v_0 + z_2)^r}.$$

We can evaluate $J_1(0, 0, 0)$ to obtain the optimal objective value of problem (4.5)-(4.6). We note that the size of the state space grows exponentially with m, which creates computational challenges as the number of product types increases. To mitigate this difficulty, we will consider a modified version of the state space by setting a fixed integer $T \geq 1$ and dividing each dimension of the state space into T intervals of equal length. Let $\mathcal{T} = \{0, \dots, T\}$. Noting that the maximum values that z_1, z_2 and z_3 can take are $\sum_{i\in M}(p_i - c_i)V_i(N)$, $\sum_{i\in M} V_i(N)$ and $\sum_{i\in M} \Theta_i(N)$, respectively, let $\Delta_1 = \frac{1}{T}\sum_{i\in M}(p_i - c_i)V_i(N)$, $\Delta_2 = \frac{1}{T}\sum_{i\in M} V_i(N)$, $\Delta_3 = \frac{1}{T}\sum_{i\in M} \Theta_i(N)$. The modified state space that we consider is

$$\mathcal{Z} = \Big\{(t_1\Delta_1, t_2\Delta_2, t_3\Delta_3) : (t_1, t_2, t_3) \in \mathcal{T}^3\Big\},$$

which contains $(T + 1)^3$ elements.

We can show that there exists a solution lying in $\mathcal{Z}$ that provides an upper bound on the optimal objective value of problem (4.5)-(4.6). To see this claim, suppose that we have an optimal solution $\mathcal{S}^* = (S_1^*, \ldots, S_m^*)$ to problem (4.5)-(4.6). Let

$$t_1^* = \left\lceil \frac{\sum_{i\in M}(p_i - c_i)V_i(S_i^*)}{\Delta_1} \right\rceil, t_2^* = \left\lfloor \frac{\sum_{i\in M} V_i(S_i^*)}{\Delta_2} \right\rfloor, t_3^* = \left\lfloor \frac{\sum_{i\in M} \Theta_i(S_i^*)}{\Delta_3} \right\rfloor.$$

Then $(t_1^*\Delta_1, t_2^*\Delta_2, t_3^*\Delta_3) \in \mathcal{Z}$ and

$$\begin{aligned}\lambda \frac{t_1^*\Delta_1}{v_0 + t_2^*\Delta_2} - \frac{t_3^*\Delta_3}{(v_0 + t_2^*\Delta_2)^r} &= \frac{\lambda t_1^*\Delta_1 - (v_0 + t_2^*\Delta_2)^{1-r} t_3^*\Delta_3}{v_0 + t_2^*\Delta_2} \\ &\geq \frac{\lambda \sum_{i\in M}(p_i - c_i)V_i(S_i^*) - (v_0 + \sum_{i\in M} V_i(S_i^*))^{1-r} \sum_{i\in M} \Theta_i(S_i^*)}{v_0 + \sum_{i\in M} V_i(S_i^*)} \\ &= \sum_{i\in M}\sum_{j\in N}(p_i - c_i)\lambda q_{ij}(\mathcal{S}^*) - \sum_{i\in M}\sum_{j\in N} \theta_i q_{ij}(\mathcal{S}^*)^r.\end{aligned}$$

Using this idea, we can define an approximation to our original dynamic program that provides an upper bound on the optimal objective value of problem (4.5)-(4.6). Let

$$\tilde{J}_i(z_1, z_2, z_3) = \max_{S_i \in \mathcal{P}} \tilde{J}_{i+1}\left(\left\lceil \frac{z_1 + (p_i - c_i)V_i(S_i)}{\Delta_1} \right\rceil \Delta_1, \left\lfloor \frac{z_2 + V_i(S_i)}{\Delta_2} \right\rfloor \Delta_2, \left\lfloor \frac{z_3 + \Theta_i(S_i)}{\Delta_3} \right\rfloor \Delta_3 \right) \quad (4.11)$$

for all $(z_1, z_2, z_3) \in \mathcal{Z}$, $i \in M$, with $\tilde{J}_{m+1}(z_1, z_2, z_3) = J_{m+1}(z_1, z_2, z_3)$. Then $\tilde{J}_1(0,0,0) \geq J_1(0,0,0)$. We can compute $\tilde{J}_1(0,0,0)$ by evaluating the value functions $\tilde{J}_i(z_1, z_2, z_3)$ for all $i = 1, \ldots, m+1$ and $(z_1, z_2, z_3) \in \mathcal{Z}$, which requires $O(mnT^3)$ operations. We can construct an assortment $(S_1, \ldots, S_m)$ that provides a lower bound on the optimal objective value of problem (4.5)-(4.6) by emplying the greedy policy with respect to the value functions $\tilde{J}_i(z_1, z_2, z_3)$, i.e., choosing the assortment S_i that achieves the maximum in (4.11) for all $i \in M$ starting from $\tilde{J}_1(0,0,0)$.

4.6 An Approximate Model

High-quality solutions to problem (4.5)-(4.6) can also be computed with less computational effort through use of an approximate problem, where we assume that for any assortment $\mathcal{S}$, with probability $q_{ij}(\mathcal{S})$ all λ customers will by product j in type i, for all $j \in N, i \in M$. In other words, we assume that either exactly one of the products that we offer will be chosen by the entire arriving customer population, or that all customers will choose to buy nothing. The optimal stocking strategy under such a situation is to stock λ units of every product whose purchase probability $q_{ij}(\mathcal{S})$ is at least as large as its cost-to-price ratio. The approximate problem that we formulate corresponds precisely to the assortment optimization problem resulting from this scenario, but with price and cost parameters that are adjusted from their original values in order to provide a good approximation to problem (4.5)-(4.6). The primary advantage of this approximate problem is that its analogous dynamic programming formulation has a smaller state space than that of the dynamic program studied in the previous section.

To formulate the approximate problem, we let $h_i(q) = (p_i - c_i)\lambda q - \theta_i q^r$, so that the objective function of problem (4.5)-(4.6) is $\sum_{i \in M} \sum_{j \in N} h_i(q_{ij}(\mathcal{S}))$. The function $h_i(\cdot)$ is convex and satisfies $h_i(q) = 0$ at $q = 0$ and at $q = \eta_i$, where $\eta_i = (\theta_i/((p_i - c_i)\lambda))^{\frac{1}{1-r}}$. For $q \in [0, \eta_i]$, we have $h_i(q) \leq 0$. We observe that if $\mathcal{S}^* = (S_1^*, \ldots, S_m^*)$ is an optimal solution for problem (4.5)-(4.6), then $h_i(q_{ij}(\mathcal{S}^*)) \geq 0 \, \forall i \in M, j \in N$. To see this, suppose that $h_i(q_{ij}(\mathcal{S}^*)) < 0$ for some $j \in S_i^*$. Then we could remove this product from the assortment. This would increase the purchase probabilities q_{lk} of all products such that $f_l(q_{lk}(\mathcal{S}^*)) \geq 0$, which would in turn cause $h_l(q_{lk})$ to increase, by the convexity of $h_l(\cdot)$. Meanwhile, since $h_i(0) = 0$ but $h_i(q_{ij}(\mathcal{S}^*)) < 0$, the new solution would provide a strictly better objective value than $\mathcal{S}^*$, contradicting the optimality of $\mathcal{S}^*$. Since an optimal solution to problem (4.5)-(4.6) will never have $h_i(q_{ij}(\mathcal{S}^*)) < 0$, problem (4.5)-(4.6) is equivalent to maximizing $\sum_{i \in M} \sum_{j \in N} [h_i(q_{ij}(\mathcal{S}))]^+$, where $[\cdot]^+$ denotes the positive-part function. We also note that the derivative of $h_i(\cdot)$ at η_i is $(1 - r)(p_i - c_i)\lambda$. Since $h_i(\cdot)$ is convex and $h_i(\eta_i) = 0$, we can lower bound $[h_i(q)]^+$ by $(1-r)(p_i - c_i)\lambda[q - \eta_i]^+$ (See Figure 4.1.) This suggests maximizing the function

$$\max_{\mathcal{S}} \sum_{i \in M} \sum_{j \in N} (1-r)(p_i - c_i)\lambda[q_{ij}(\mathcal{S}) - \eta_i]^+$$

as an approximation to problem (4.5)-(4.6). This problem corresponds exactly to the assortment optimization problem under the scenario described at the beginning of this section, where the adjusted price and cost parameters are $(1-r)(p_i - c_i)$ and $(1-r)(p_i - c_i)\eta_i$, respectively, meaning that η_i represents the margin in this approximation.

Note that if $\bar{\mathcal{S}} = (\bar{S}_1, \ldots, \bar{S}_m)$ maximizes the objective function above and $q_{ij}(\bar{\mathcal{S}}) < \eta_i$ for some $j \in \bar{S}_i$, then we could remove this product from the assortment and strictly increase our objective value (using a similar argument as in the paragraph above), implying that $q_{ij}(\bar{\mathcal{S}}) \geq \eta_i$ for any optimal assortment $\bar{\mathcal{S}}$. Thus, we can disregard the positive-part operator when solving this problem. Since the constants $(1-r)$ and λ are also superfluous, we arrive at the approximate problem

$$\max_{\mathcal{S}} \sum_{i \in M} \sum_{j \in N} (p_i - c_i) q_{ij}(\mathcal{S}) - \sum_{i \in M} (p_i - c_i)\eta_i |S_i| \tag{4.12}$$

$$\text{s.t. } \mathcal{S} = (S_1, \ldots, S_m),\ S_i \subseteq N\ \forall i \in M. \tag{4.13}$$

If there is some $i \in M$ such that $\eta_i > 1$, then since $q^r \geq q$ for all $q \in [0,1]$, $h_i(q_{ij}(\mathcal{S}))$ will be non-positive for all $j \in N$, and it will be optimal to offer no products in nest i. As such, we will assume without loss of generality that $\eta_i \leq 1$ for all $i \in M$.

Problem (4.12)-(4.13) retains important structural properties of problem (4.5)-(4.6), namely that optimal assortments belong to the popular set $\mathcal{P}$. Therefore, as in the original problem, we need only consider $n+1$ different assortments of each product type in order to construct an optimal solution.

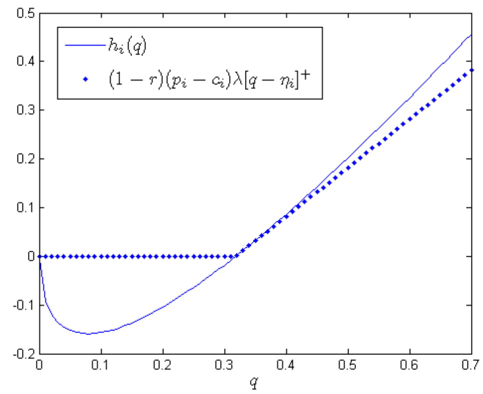

Figure 4.1: The functions $h_i(\cdot)$ and $(1-r)(p_i-c_i)\lambda[\cdot-\eta_i]^+$.

Theorem 9. *There exists an optimal assortment* $\bar{\mathcal{S}} = (\bar{S}_1, \ldots, \bar{S}_m)$ *for problem* (4.12)-(4.13) *such that* $\bar{S}_i \in \mathcal{P}$ *for all* $i \in M$.

Proof. Suppose that $\hat{\mathcal{S}} = (\hat{S}_1, \ldots, \hat{S}_m)$ is optimal for problem (4.12)-(4.13). Let $k_i = |\hat{S}_i|$ for all $i \in M$. Then we have $\sum_{i\in M}\sum_{j\in N}(p_i-c_i)q_{ij}(\hat{\mathcal{S}}) - \sum_{i\in M}\eta_i|\hat{S}_i| = \frac{\sum_{i\in M}(p_i-c_i)V_i(\hat{S}_i)}{v_0+\sum_{l\in M}V_l(\hat{S}_l)} - \sum_{i\in M}\eta_i k_i$. Now let $\bar{S}_i = \{1,\ldots,k_i\}$ for all $i \in M$. We know that $V_i(\bar{S}_i) \geq V_i(\hat{S}_i)$ for all $i \in M$, since $\mu_{i1} > \ldots > \mu_{in}$ and $\hat{S}_i$ contains k_i elements. Since the function $\frac{\sum_{i\in M}(p_i-c_i)y_i}{v_0+\sum_{i\in M}y_i}$ is increasing in $y_i \geq 0$ for each $i \in M$, it follows that $\frac{\sum_{i\in M}(p_i-c_i)V_i(\bar{S}_i)}{v_0+\sum_{l\in M}V_l(\bar{S}_l)} - \sum_{i\in M}\eta_i k_i \geq \frac{\sum_{i\in M}(p_i-c_i)V_i(\hat{S}_i)}{v_0+\sum_{l\in M}V_l(\hat{S}_l)} - \sum_{i\in M}\eta_i k_i$. Thus, $\bar{\mathcal{S}} = (\bar{S}_1, \ldots, \bar{S}_m)$ must be optimal for problem (4.12)-(4.13). □

Noting that the objective function of problem (4.12)-(4.13) is $\frac{\sum_{i\in M}(p_i-c_i)V_i(S_i)}{v_0+\sum_{i\in M}V_i(S_i)} - \sum_{i\in M}(p_i-c_i)\eta_i|S_i|$ and using Lemma 9, we can solve problem (4.12)-(4.13) using a dynamic program with optimality equation

$$G_i(z_1,z_2) = \max_{S_i\in\mathcal{P}}\Big\{-(p_i-c_i)\eta_i|S_i| + G_{i+1}\big(z_1+(p_i-c_i)V_i(S_i), z_2+V_i(S_i)\big)\Big\}, \tag{4.14}$$

where the boundary condition is $G_{m+1}(z_1,z_2) = \frac{z_1}{v_0+z_2}$. We can evaluate $G_1(0,0)$ to obtain the optimal objective value of problem (4.12)-(4.13). This dynamic program has two states, which is one less than the dynamic program used to obtain the optimal objective value of our original problem. As such, we can use this dynamic program to obtain approximate solutions with reduced computational effort. We can make use of a similar state space modification technique as the one described in the previous subsection to come up with an upper bound on the optimal objective value of problem (4.12)-(4.13) and obtain assortments that provide lower bounds on the optimal objective value of problem (4.5)-(4.6). In particular, if we specify a number T of grid points in each dimension of the modified state space and carry over the notation used in the previous section, we accomplish this by evaluating the value functions

$$\begin{aligned}&\tilde{G}_i(z_1, z_2)\\&= \max_{S_i \in \mathcal{P}} \left\{ -(p_i - c_i)\eta_i|S_i| + \tilde{G}_{i+1}\left(\left\lceil \frac{z_1 + (p_i - c_i)V_i(S_i)}{\Delta_1} \right\rceil \Delta_1, \left\lfloor \frac{z_2 + V_i(S_i)}{\Delta_2} \right\rfloor \Delta_2 \right) \right\},\end{aligned}$$

for all $(z_1, z_2) \in \{(t_1\Delta_1, t_2\Delta_2) : (t_1, t_2) \in \mathcal{T}^2\}$, $i \in M$, with $\tilde{G}_{m+1}(z_1, z_2) = G_{m+1}(z_1, z_2)$. The work required using this approach is $O(mnT^2)$.

Problem (4.12)-(4.13) can obtain quite accurate solutions to problem (4.5)-(4.6), depending on the value of the distribution parameter r. The accuracy of solutions is formalized in the following lemma, which shows that if we solve problem (4.12)-(4.13) and use this solution as a solution to problem (4.5)-(4.6), the loss in expected profit is at most $(r \cdot 100)\%$. We emphasize that this is an upper bound on the profit loss, and that our numerical experiments indicate that the loss in profit tends to be significantly less than this worst-case guarantee.

Lemma 8. *Let $\mathcal{S}^*$ be an optimal solution to problem* (4.5)-(4.6) *and $\bar{\mathcal{S}}$ an optimal solution to problem* (4.12)-(4.13). *Then* $\frac{1}{1-r} \sum_{i \in M} \sum_{j \in N} h_i(q_{ij}(\bar{\mathcal{S}})) \geq \sum_{i \in M} \sum_{j \in N} h_i(q_{ij}(\mathcal{S}^*))$.

Proof. The derivative is $h_i(\cdot)$ is given by $h_i'(q) = (p_i - c_i)\lambda - \frac{r\theta_i}{q^{1-r}} \leq (p_i - c_i)\lambda - r\theta_i \, \forall q \in [0,1]$. This implies that $((p_i - c_i)\lambda - r\theta_i)[q - \eta_i]^+ \geq [h_i(q)]^+ \, \forall q \in [0,1]$, since $h_i(q) \leq 0$ for $q \in [0, \eta_i]$. Thus, we have

$$
\begin{aligned}
\sum_{i \in M} \sum_{j \in N} h_i(q_{ij}(\mathcal{S}^*)) &= \max_{\mathcal{S}} \sum_{i \in M} \sum_{j \in N} h_i(q_{ij}(\mathcal{S})) = \max_{\mathcal{S}} \sum_{i \in M} \sum_{j \in N} [h_i(q_{ij}(\mathcal{S}))]^+ \\
&\leq \max_{\mathcal{S}} \sum_{i \in M} \sum_{j \in N} ((p_i - c_i)\lambda - r\theta_i)[q_{ij}(\mathcal{S}) - \eta_i]^+ \\
&= \max_{\mathcal{S}} \sum_{i \in M} \sum_{j \in N} \frac{1 - r\eta_i^{1-r}}{1-r} (p_i - c_i)\lambda(1-r)[q_{ij}(\mathcal{S}) - \eta_i]^+ \\
&\leq \frac{1}{1-r} \max_{\mathcal{S}} \sum_{i \in M} \sum_{j \in N} (p_i - c_i)\lambda(1-r)[q_{ij}(\mathcal{S}) - \eta_i]^+ \\
&= \frac{1}{1-r} \sum_{i \in M} \sum_{j \in N} (p_i - c_i)\lambda(1-r)[q_{ij}(\bar{\mathcal{S}}) - \eta_i]^+ \leq \frac{1}{1-r} \sum_{i \in M} \sum_{j \in N} [h_i(q_{ij}(\bar{\mathcal{S}}))]^+ \\
&= \frac{1}{1-r} \sum_{i \in M} \sum_{j \in N} h_i(q_{ij}(\bar{\mathcal{S}})),
\end{aligned}
$$

where the second inequality follows from $\eta_i \leq 1$ and the last equality follows from the fact that $q_{ij}(\bar{\mathcal{S}}) \geq \eta_i \, \forall j \in \bar{S}_i$, $\forall i \in M$. □

4.7 Numerical Experiments

In this section, we test the performance of the assortment and stocking decisions obtained by solving approximations to problem (4.5)-(4.6) and examine how these assortment and stocking decisions behave with changes in the utility correlation parameters γ_i. We make use of the three-state dynamic program from Section 4.5 to obtain upper and lower bounds on the optimal objective value of problem (4.5)-(4.6) and assortments and stocking decisions corresponding to lower bounds. In order to address our central questions, we examine how the assortment and inventory decisions obtained by the policy corresponding to the three state dynamic program change with the values of the parameters γ_i when all other problem parameters are fixed. In addition, we use the two-state dynamic program from Section 4.6 to obtain lower bounds on the optimal objective value and assortments and stocking decisions corresponding to lower bounds, and compare the performance of these assortments to the upper bound obtained by the three-state dynamic program.

4.7.1 Experimental Setup

In our test problems, we generate a number of problem instances. For each problem instance, we compute the value functions $\tilde{J}_i(\cdot,\cdot,\cdot)$ for all $i \in M$ and record the upper bound $\tilde{J}_1(0,0,0)$ on the optimal objective value of problem (4.5)-(4.6). We also compute the assortment $\mathcal{S}^3 = (S_1^3, \ldots, S_m^3)$ corresponding to the greedy policy with respect to the value functions $\tilde{J}_i(\cdot,\cdot,\cdot)$ and compute $\sum_{i\in M}\sum_{j\in N} h_i(q_{ij}(\mathcal{S}^3))$ to obtain a lower bound on the optimal objective value of problem (4.5)-(4.6) corresponding to the three-state policy (3-SP). In addition, we compute the value functions $\tilde{G}_i(\cdot,\cdot)$ for all $i \in M$ and the assortment $\mathcal{S}^2 = (S_1^2, \ldots, S_m^2)$ corresponding to the greedy policy with respect to these value functions. We compute $\sum_{i\in M}\sum_{j\in N} h_i(q_{ij}(\mathcal{S}^2))$ to obtain a lower bound on the optimal objective value of problem (4.5)-(4.6) corresponding to the two-state policy (2-SP).

In each problem instance, we have $m = 15$ product types and $n = 4$ variants of each product type, from which we need to come up with assortments to offer to customers. Letting W_{ij} be a sample from the uniform distribution over $[0,1]$, we set the deterministic utility component of variant j of type i as $\mu_{ij} = \log(1+9W_{ij})$, so that $e^{\mu_{ij}}$ is uniformly distributed over the interval $[1,10]$. To generate the no-purchase preference, we set $v_0 = \sum_{i\in M}\sum_{j\in N} e^{\mu_{ij}}/9$, so that the probability of non-purchase is 0.1 when all products are offered ($S_i = N$) and $\gamma_i = 1$ for all i.

To come up with the prices and costs of the product types, we sample the price p_i from the uniform distribution over $[200, 500]$, the margin δ_i from the uniform distribution over $[0.3, 0.7]$ and set $c_i = \delta_i p_i$. We use $r = 0.5$ and $\alpha = 1$ in all of our experiments. We vary the customer volume λ over the values $1000, 2000$ and 4000.

For each value of λ, we generate 60 individual problem instances. For each problem instance, we generate a set of correlation parameters $(\gamma_1, \ldots, \gamma_m)$ from the uniform distribution over $[\gamma^L, \gamma^H]$ for $[\gamma^L, \gamma^H] = [0.05, 0.15]$, $[0.15, 0.25]$, $\ldots, [0.75, 0.85]$, $[0.85, 1]$. For each generated $(\gamma_1, \ldots, \gamma_m)$ vector for each problem instance, we compute the approximate value functions $\tilde{J}_i(\cdot, \cdot, \cdot)$ and $\tilde{G}_i(\cdot, \cdot)$ using a fixed size of $T = 200$ intervals and record the assortments and stocking decisions obtained by 3-SP. For each $[\gamma^L, \gamma^H]$ range, we test the average performance of 3-SP and 2-SP against the upper bound obtained by 3-SP over all problem instances and compute the average assortment sizes and average inventory levels given by 3-SP over all problem instances.

4.7.2 Computational Results

Table 1 compares the expected profit bounds obtained by the dynamic programming approximations with $\lambda = 1000$.. The first two columns in this table show the range of values $[\gamma^L, \gamma^H]$ from which the parameters γ_i were generated. The third column shows the average lower bound on the optimal expected profit obtained by 3-SP, where the average is taken over all 60 problem instances that we generate. That is, using $\mathcal{S}_k^3 = (S_{1k}^3, \ldots, S_{nk}^3)$ to denote the assortment generated by 3-SP for problem instance k, the third column gives $\frac{1}{60} \sum_{k=1}^{60} \sum_{i \in M} \sum_{j \in M} h_i(q_{ij}(\mathcal{S}_k^3))$ for each $[\gamma^L, \gamma^H]$ range. The fourth column gives the average lower bound on the optimal expected profit obtained by 2-SP in a similar manner. The fifth column gives the average upper bound obtained by 3-SP, i.e., $\frac{1}{60}\sum_{k=1}^{60}\tilde{J}_{1k}$, where $\tilde{J}_{1k}$ is the value of $\tilde{J}_1(0,0,0)$ obtained in problem instance k. The sixth and seventh columns show $\frac{1}{60}\sum_{k=1}^{60} 100\frac{\tilde{J}_{1k}-\sum_{i\in M}\sum_{j\in N} h_i(q_{ij}(\mathcal{S}_k^3))}{\tilde{J}_{1k}}$ and $\frac{1}{60}\sum_{k=1}^{60} 100\frac{\tilde{J}_{1k}-\sum_{i\in M}\sum_{j\in N} h_i(q_{ij}(\mathcal{S}_k^2))}{\tilde{J}_{1k}}$, respectively, which are the average percentage gaps between the upper bound and the lower bounds obtained by 3-SP and 2-SP. Tables 2 and 3 follow the same format as Table 1 for $\lambda = 2000$ and $\lambda = 4000$ respectively. The columns of Table 4 from column three onward show $\frac{1}{60}\sum_{k=1}^{60}|S_{ik}^3|$, the average number of products of type i offered by 3-SP, for each $[\gamma^L, \gamma^H]$ range, for all $i = 1, \ldots, 15$. Similarly, the columns of Table 5 from three onward show $\frac{1}{60}\sum_{k=1}^{60}\sum_{j\in N} x_{ij}^*(\mathcal{S}_k^3)$, the average inventory stocked of product type i by 3-SP, for each $[\gamma^L, \gamma^H]$ range, for all $i = 1, \ldots, 15$.

The results in Tables 1,2, and 3 show that 3-SP can achieve lower-to-upper bound gaps of under 2 percent, suggesting that assortments computed by 3-SP perform very well relative to true optimal assortments on average for sufficiently high values of T. We note that the upper and lower bounds on the optimal expected profit obtained by 3-SP are both increasing as γ^L and γ^H increase, as one would expect given that Theorem 8 indicates that the optimal objective value of problem (4.5)-(4.6) is nondecreasing in γ_i. We also note that 2-SP generally performs better for smaller values of γ_i.

Table 4 also indicates that the averages size of assortments offered by 3-SP are increasing as γ^L and γ^H increase. While between zero and one product is offered on average when γ_i values lie in the interval $[0.05, 0.15]$, upwards of three out of a potential four products are offered on average when γ_i values are generated from $[0.85, 1]$. The results are consistent not only with Lemma 7, which indicates that either $\emptyset$ or $\{1\}$ is an optimal assortment for product type i as γ_i becomes small, but also our hypothesis in Section 4.4 that optimal assortment sizes increase with

γ_i. These properties are also illustrated in Figure 2, which shows a plot of the average total number of products offered by 3-SP as a function of γ^H.

Table 5 indicates that the average inventory levels of products stocked in each product type by 3-SP also tend to become larger as γ^L and γ^H increase. While we know from Lemma 6 that total inventory stocked in a product type i increases as γ_i increases for a fixed assortment $\mathcal{S}$, it was not clear whether this result also applied to optimal assortments, i.e., accounting for the fact that the optimal assortment is not fixed and is also changing with the values of γ_i. The results in Table 3 show that, while this may not be true, there is a general upward trend observable in the sizes of assortments offered by 3-SP as the utilities of products in all types become increasingly less correlated with one another. This is also illustrated in Figure 2, which shows a plot of the average total number of products stocked by 3-SP as a function of γ^H.

Our results also indicate that 2-SP generates assortments that can perform well relative to upper bounds on optimal expected profits. In particular, the assortments generated by 2-SP perform significantly better than the worst-case guarantee of a 50% profit loss from solving problem (4.12)-(4.13) on average, attaining a worst observed average percentage gap of 12.63% in our experiments. Examining Tables 2 and 3, we note that while the performance of 3-SP appears relatively insensitive to changes in λ, the performance of 2-SP is much more sensitive, attaining a worst observed average percentage gap of only 4.19% when $\lambda = 4000$. This suggests that the number of grid points T in each dimension is the most significant driver of performance in 3-SP, and that 2-SP may be more useful for problems with a large customer volume due to its reduced computational effort and relatively good performance.

γ^L	γ^H	Avg. Expected Profit Bounds ($\times 10^5$)			Avg. Gap with Upper Bound	
		Lower (3-SP)	Lower (2-SP)	Upper	3-SP	2-SP
0.05	0.15	1.272	1.241	1.287	1.17 %	3.56 %
0.15	0.25	1.277	1.243	1.301	1.83 %	4.40 %
0.25	0.35	1.291	1.244	1.316	1.99 %	5.46 %
0.35	0.45	1.310	1.242	1.335	1.92 %	6.87 %
0.45	0.55	1.332	1.243	1.357	1.87 %	8.30 %
0.55	0.65	1.355	1.245	1.379	1.76 %	9.64 %
0.65	0.75	1.379	1.247	1.401	1.64 %	10.89 %
0.75	0.85	1.402	1.251	1.423	1.55 %	11.93 %
0.85	1.00	1.429	1.263	1.449	1.41 %	12.63 %

Table 4.1: Performance of 3-SP and 2-SP ($\lambda = 1000$)

γ^L	γ^H	Avg. Expected Profit Bounds ($\times 10^5$)			Avg. Gap with Upper Bound	
		Lower (3-SP)	Lower (2-SP)	Upper	3-SP	2-SP
0.05	0.15	2.867	2.854	2.895	1.01 %	1.41 %
0.15	0.25	2.881	2.869	2.929	1.65 %	2.03 %
0.25	0.35	2.917	2.869	2.971	1.82 %	3.43 %
0.35	0.45	2.967	2.876	3.018	1.70 %	4.60 %
0.45	0.55	3.021	2.888	3.073	1.69 %	5.80 %
0.55	0.65	3.077	2.916	3.125	1.55 %	6.48 %
0.65	0.75	3.134	2.945	3.179	1.41 %	7.14 %
0.75	0.85	3.187	2.987	3.228	1.30 %	7.27 %
0.85	1.00	3.249	3.033	3.286	1.16 %	7.48 %

Table 4.2: Performance of 3-SP and 2-SP ($\lambda = 2000$)

4.8 Conclusions

We formulated a joint assortment offering and stocking problem where a firm offers a number of heterogeneous products that can be designated into subgroups, where customers' product evaluations are correlated. Through analysis of our problem, we demonstrated relationships between product evaluation correlation and optimal product variety, optimal inventory levels and optimal profits. The relationships demonstrated align with trends observed in practice. In addition, we provided structural properties of optimal assortments and developed dynamic-programming based solution methods to obtain upper bounds on the optimal expected profit and near-optimal solutions. Our numerical experiments demonstrated that our approx-

γ^L	γ^H	Avg. Expected Profit Bounds ($\times 10^5$)			Avg. Gap with Upper Bound	
		Lower (3-SP)	Lower (2-SP)	Upper	3-SP	2-SP
0.05	0.15	5.171	5.170	5.217	0.92 %	0.93 %
0.15	0.25	5.196	5.182	5.283	1.65 %	1.90 %
0.25	0.35	5.270	5.202	5.360	1.68 %	2.89 %
0.35	0.45	5.359	5.248	5.450	1.68 %	3.65 %
0.45	0.55	5.462	5.323	5.548	1.56 %	4.03 %
0.55	0.65	5.567	5.403	5.646	1.42 %	4.19 %
0.65	0.75	5.673	5.507	5.746	1.29 %	4.10 %
0.75	0.85	5.763	5.591	5.833	1.21 %	4.06 %
0.85	1.00	5.875	5.703	5.936	1.05 %	3.84 %

Table 4.3: Performance of 3-SP and 2-SP ($\lambda = 4000$)

imation methods can obtain well-performing solutions in reasonable computation time.

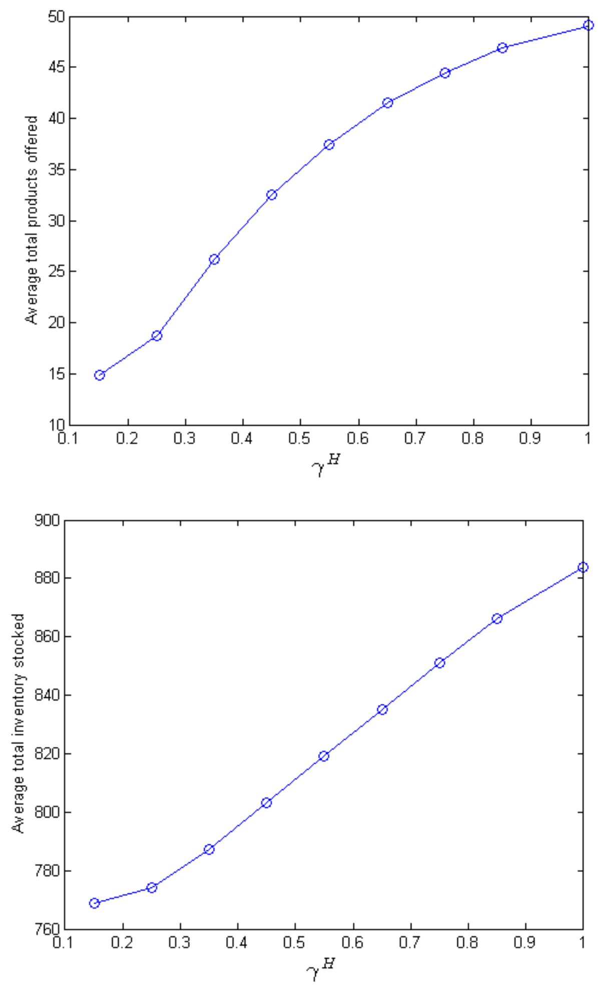

Figure 4.2: Average total products offered (top) and total inventory (bottom) stocked by 3-SP as a function of γ^H ($\lambda = 1000$)

γ^L	γ^H	Avg. # of Products Offered in Type i (3-SP)														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0.05	0.15	0.97	0.98	1.00	0.98	1.00	0.99	0.97	0.99	1.00	0.97	0.99	1.01	1.02	1.01	1.02
0.15	0.25	1.18	1.15	1.25	1.22	1.18	1.21	1.20	1.27	1.22	1.26	1.28	1.30	1.25	1.40	1.35
0.25	0.35	1.45	1.43	1.57	1.65	1.69	1.72	1.80	1.88	1.74	1.83	1.82	1.91	1.88	1.89	1.89
0.35	0.45	1.79	1.75	1.89	1.96	1.98	2.07	2.16	2.25	2.30	2.35	2.29	2.31	2.48	2.43	2.47
0.45	0.55	2.14	2.07	2.20	2.27	2.28	2.35	2.50	2.55	2.59	2.69	2.68	2.72	2.81	2.71	2.85
0.55	0.65	2.38	2.28	2.47	2.60	2.59	2.59	2.71	2.83	2.89	2.84	3.02	3.02	3.10	3.07	3.13
0.65	0.75	2.78	2.64	2.82	2.80	2.79	2.83	2.98	2.98	3.03	2.98	3.03	3.15	3.27	3.18	3.23
0.75	0.85	2.99	2.87	2.91	3.01	3.08	3.09	3.05	3.15	3.25	3.17	3.19	3.14	3.30	3.33	3.36
0.85	1.00	3.00	3.05	3.12	3.15	3.31	3.20	3.31	3.24	3.30	3.30	3.30	3.37	3.51	3.42	3.49

Table 4.4: Average assortment sizes offered by 3-SP by product type ($\lambda = 1000$)

γ^L	γ^H	Avg. Total Inventory of Products of Type i (3-SP)														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0.05	0.15	51.8	51.3	51.0	50.6	51.4	52.1	50.8	50.2	52.0	50.5	51.0	51.7	52.0	51.7	50.6
0.15	0.25	52.3	51.3	51.3	51.0	51.3	52.1	51.6	50.8	51.8	51.3	51.4	51.9	52.2	52.4	51.6
0.25	0.35	51.8	51.1	51.5	51.8	52.2	53.0	52.9	51.9	52.8	52.3	52.7	53.3	54.0	53.1	52.8
0.35	0.45	52.4	51.4	51.8	52.4	52.6	53.9	53.6	53.0	54.4	54.1	54.0	54.2	55.8	55.0	54.4
0.45	0.55	53.7	51.8	52.7	53.3	53.3	54.6	55.2	53.9	55.7	55.0	55.6	55.3	57.3	55.8	55.9
0.55	0.65	54.3	52.1	53.8	54.7	54.8	55.3	56.1	54.7	56.8	55.7	56.9	56.7	58.8	57.1	57.5
0.65	0.75	56.4	54.2	55.6	55.5	55.3	56.8	57.0	55.7	57.5	56.5	57.0	57.7	59.9	58.0	58.1
0.75	0.85	57.7	55.2	56.0	56.8	57.2	57.8	57.9	57.0	59.1	57.9	58.0	57.5	60.2	58.5	59.4
0.85	1.00	58.2	56.2	57.5	57.5	59.1	59.2	59.5	57.7	59.6	58.7	59.3	59.5	61.8	59.6	60.5

Table 4.5: Average inventory stocked by 3-SP by product type ($\lambda = 1000$)

CHAPTER 5

CONCLUSIONS

In this thesis, we considered three problems arising from scenarios in which customers choose among a number of substitutable products that can be designated into subgroups, such that products within a subgroup are more similar to one another than to products within a different subgroup. In each problem, we used the Nested Logit (NL) model to capture the customer's choice process, in which product evaluations for different products are allowed to be correlated. The central decisions for the retailer considered in this thesis were the prices of the offered products and the subset of products to offer, with some consideration also given to the amount of each offered product to stock in the fourth chapter.

In Chapter 2, we studied a pricing problem that arose from a scenario in which a firm offers a product through multiple sales channels where it is grouped together with competing products, and must choose the price to offer in each channel. We provided structural properties of the objective function and developed a sequential procedure for obtaining optimal prices. We also showed that the structure of optimal prices differs from that which is obtained when using the simpler Multinomial Logit (MNL) model, a special case of the NL model, highlighting the impact of taking product evaluation correlation into account when making pricing decisions. In Chapter 3, we considered a more general pricing problem under the NL model, where a firm controls multiple products categories with multiple products in each category, with the added constraint that each product's price must satisfy a lower and upper bound. We also considered an extension of this problem where the optimal assortment to offer must be chosen in addition to the prices of the offered products. We provided efficient approximation methods that allow the user to specific performance guarantees in advance, where the final solutions are obtained via solving linear programs. Through computational experiments, we show that our methods can vastly outperform naive heuristics for solving constrained pricing problems under the NL model. In Chapter 4, we studied a joint assortment offering and stocking problem, motivated by a desire to explain relationships between product evaluation correlation and offered product variety that tend to appear in practice. We provided structural properties of optimal assortments for our problem, and used these properties to establish results concerning relationships between evaluation correlation and optimal product variety, inventory levels and profit that aligned with trends observed in practice. In addition, while our original problem is difficult, we developed multiple strategies for efficiently obtaining near-optimal solutions and upper bounds on optimal expected profits.

There are several extensions of the work in this thesis that may be pursued in future research. In the pricing setting of Chapters 2 and 3, it would be interesting to consider more general constraints on prices, such as polyhedral constraints. An example of this is price laddering. Generally, high-end products should be priced above lower-end products in order to stay consistent with quality expectations, but low-end products with low customer price sensitivities may be priced higher as a result of optimizing prices without constraints. Price ladder constraints ensure that lower-end products must be priced below their high-end counterparts.

Considering extensions to Chapter 4, while optimal product variety within a product type does not always increase with independence among product evaluations within that type, our numerical experiments suggest that this tends to be true in practice, and it may be possible to identify certain conditions under which this is the case. The question of whether a constant-factor approximation algorithm for our original problem exists remains open. While we developed an approximate problem that provides a 2-approximation to the original problem, we still needed to use a dynamic programming-based method to solve this approximate problem. In addition, it may be interesting to consider instances of our problem where there are constraints on the offered assortment, e.g., cardinality or capacity constraints.

Another extension in both pricing and assortment problems is to remove the assumption that a customer can purchase only one item, allowing multiple purchases by one customer. This inherently changes the way that utilities are assigned to items. For instance, a customer may come into a grocery store looking for a specific combination of ingredients in order to make a recipe, but may elect to leave without buying anything if one of those ingredients is missing from the store's offered assortment. With some modifications, there may be room within the framework of the MNL and NL models to capture these behaviors.

APPENDIX A

OMITTED RESULTS

A.1 Proof of Proposition 2

In this section, we complete the proof of Proposition 2. As mentioned in the proof of Proposition 2 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 (3.5) 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 (3.5) 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 (3.5) 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 (3.5) 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

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

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 (3.5) 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 (3.5) indicates that the first constraint in problem (3.5) 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 9

Lemma 9. *Letting $(S^*_i, \boldsymbol{w}^*_i)$ be an optimal solution to the maximization problem on the right side of* (3.17)*, 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

$$W_i(\hat{S}_i, \boldsymbol{w}^*_i)^{\gamma_i} (R_i(\hat{S}_i, \boldsymbol{w}^*_i) - z) = \frac{\sum_{j \in \hat{S}_i} w^*_{ij}\, (\kappa_{ij} - \eta_{ij} \log w^*_{ij} - z)}{W_i(\hat{S}_i, \boldsymbol{w}^*_i)^{1-\gamma_i}}$$
$$= \frac{\sum_{j \in S^*_i} w^*_{ij}\, (\kappa_{ij} - \eta_{ij} \log w^*_{ij} - z) - w^*_{il}\, (\kappa_{il} - \eta_{il} \log w^*_{il} - z)}{W_i(\hat{S}_i, \boldsymbol{w}^*_i)^{1-\gamma_i}}, \quad \text{(A.1)}$$

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 (A.1), 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 (A.1) by

$$\sum_{j \in S^*_i} w^*_{ij}\, (\kappa_{ij} - \eta_{ij} \log w^*_{ij} - z) - (1-\gamma_i)\, w^*_{il}\, (R_i(S^*_i, \boldsymbol{w}^*_i) - z)$$
$$= (R_i(S^*_i, \boldsymbol{w}^*_i) - z)\, (W_i(S^*_i, \boldsymbol{w}^*_i) - (1-\gamma_i)\, w^*_{il}) \geq 0,$$

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 (3.17) 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 (A.1) 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

(A.1) as

$$\frac{\sum_{j \in S_i^*} w_{ij}^* \left(\kappa_{ij} - \eta_{ij} \log w_{ij}^* - z\right) - w_{il}^* \left(\kappa_{il} - \eta_{il} \log w_{il}^* - z\right)}{W_i(\hat{S}_i, \boldsymbol{w}_i^*)^{1-\gamma_i}}$$
$$> \frac{\left(R_i(S_i^*, \boldsymbol{w}_i^*) - z\right) \left(W_i(S_i^*, \boldsymbol{w}_i^*) - (1-\gamma_i)\, w_{il}^*\right)}{W_i(S_i^*, \boldsymbol{w}_i^*)^{1-\gamma_i} + (1-\gamma_i)\, W_i(S_i^*, \boldsymbol{w}_i^*)^{-\gamma_i} \left(W_i(\hat{S}_i, \boldsymbol{w}_i^*) - W_i(S_i^*, \boldsymbol{w}_i^*)\right)}. \quad \text{(A.2)}$$

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, (A.1) and (A.2) 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 (3.17). 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 10

Theorem 10. *Letting Z^* and $\hat{z}$ respectively be the optimal objective values of problems* (3.2) *and* (3.24)*, we have $\hat{z} \geq Z^*$. Furthermore, for some $\rho \geq 0$, if the grid points in problem* (3.24) *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 (3.24), we use $\hat{z}$ and $(\hat{x}_1, \ldots, \hat{x}_m)$ to denote an optimal solution to this problem. By Proposition 2, Z^* corresponds to the value of z that satisfies (3.6), in which case, if we let x_i^* be the optimal objective value of the maximization problem on the right side of (3.6) 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 (3.6) 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 (3.24). Since $g_i(\cdot)$ is increasing and $y_i^* \le \bar{y}_i^{t_i+1}$, we have $g_i(y_i^*) \le g_i(\bar{y}_i^{t_i+1})$. Also, since $\gamma_i \le 1$ and $y_i^* \ge \bar{y}_i^{t_i}$, we have $(y_i^*)^{1-\gamma_i} \ge (\bar{y}_i^{t_i})^{1-\gamma_i}$. In this case, using the last two observations, we obtain

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

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 (3.24), 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 (3.24), we have $\hat{z} \ge \sum_{i \in M} \hat{x}_i$, 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 $\hat{z} \ge \sum_{i \in M} \hat{x}_i \ge \sum_{i \in M} x_i^* = Z^*$, which contradicts the fact that $\hat{z} < Z^*$. Therefore, we must have $\hat{z} \ge Z^*$.

Second, we show that $(1+\rho)\, Z^* \ge \hat{z}$. 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 (3.6) when this problem is solved with $z = Z^*$. Therefore, it follows that $x_i^* \ge (\bar{y}_i^t)^{\gamma_i}\, g_i(\bar{y}_i^t)/\bar{y}_i^t - (\bar{y}_i^t)^{\gamma_i}\, Z^*$ for all $t = 1, \ldots, \tau_i - 1$. If we multiply the last inequality by $1 + \rho$, then we obtain

$$\begin{aligned}(1+\rho)\, x_i^* &\ge (\bar{y}_i^t)^{\gamma_i} \frac{(1+\rho)\, g_i(\bar{y}_i^t)}{\bar{y}_i^t} - (\bar{y}_i^t)^{\gamma_i}\, (1+\rho)\, Z^* \\ &\ge (\bar{y}_i^t)^{\gamma_i} \frac{g_i(\bar{y}_i^{t+1})}{\bar{y}_i^t} - (\bar{y}_i^t)^{\gamma_i}\, (1+\rho)\, Z^*,\end{aligned}$$

where the second inequality follows by noting the fact that $g_i(\bar{y}_i^{t+1}) \le (1+\rho)\, g_i(\bar{y}_i^t)$

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 (3.24). Thus, the objective value provided by this solution for problem (3.24) 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 (3.24), we get $(1+\rho)\, Z^* \geq \hat{z}$. □

A.4 Proof of Proposition 3

We first note that $\lim_{\gamma_i \to 0} e^{(\mu_{ik} - \mu_{ij})/\gamma_i}$ is equal to 1 if $\mu_{ik} = \mu_{ij}$, 0 if $\mu_{ik} < \mu_{ij}$ and ∞ if $\mu_{ik} > \mu_{ij}$. To show the first part, note that $V_i(S_i, \gamma_i) = e^{\gamma_i \log(\sum_{j \in S_i} e^{\mu_{ij}/\gamma_i})}$. Using L'Hospital's rule, we note that

$$\lim_{\gamma_i \to 0} \gamma_i \log(\sum_{j \in S_i} e^{\mu_{ij}/\gamma_i}) = \lim_{\gamma_i \to 0} \frac{\log(\sum_{j \in S_i} e^{\mu_{ij}/\gamma_i})}{1/\gamma_i}$$

$$= \lim_{\gamma_i \to 0} \sum_{j \in S_i} \left\{ \frac{\mu_{ij}}{\sum_{k \in S_i} e^{(\mu_{ik} - \mu_{ij})/\gamma_i}} \right\} = \operatorname*{argmax}_{k \in S_i} \mu_{ik}$$

and the result follows. To show the second part, we have

$$\lim_{\gamma_i \to 0} \frac{e^{\mu_{ij}/\gamma_i}}{V_i(S_i, \gamma_i)^{1/\gamma_i}} = \lim_{\gamma_i \to 0} \frac{1}{\sum_{k \in N} e^{(\mu_{ik} - \mu_{ij})/\gamma_i}},$$

which is equal to 1 if $\mu_{ij} = \max_{k \in S_i} \mu_{ik}$ and 0 otherwise.

A.5 Proof of Lemma 6

We have

$$\sum_{j \in N} x_{ij}^*(\mathcal{S}, \boldsymbol{\gamma}) = \lambda Q_i(\mathcal{S}, \boldsymbol{\gamma}) + \alpha \lambda^r \Phi^{-1}(1 - c_i/p_i) Q_i(\mathcal{S}, \boldsymbol{\gamma})^r \sum_{j \in S_i} q_{j|i}(S_i, \gamma_i)^r$$

and

$$
\begin{aligned}
&\frac{\partial}{\partial \gamma_i}\Big[\sum_{j\in S_i} q_{j|i}(S_i,\gamma_i)^r\Big] \\
&\quad = \sum_{j\in S_i} r\bigg(\frac{e^{\mu_{ij}/\gamma_i}}{V_i(S_i,\gamma_i)^{1/\gamma_i}}\bigg)^{r-1}\frac{1}{\gamma_i^2}\bigg(-\mu_{ij}+\frac{\sum_{k\in S_i}\mu_{ik}e^{\mu_{ik}/\gamma_i}}{\sum_{k\in S_i}e^{\mu_{ik}/\gamma_i}}\bigg)\frac{e^{\mu_{ij}/\gamma_i}}{\sum_{k\in S_i}e^{\mu_{ik}/\gamma_i}} \\
&\quad = \frac{r}{\gamma_i^2}\bigg\{-\frac{\sum_{j\in S_i}\mu_{ij}e^{r\mu_{ij}/\gamma_i}}{(\sum_{k\in S_i}e^{\mu_{ik}/\gamma_i})^r}+\frac{\sum_{k\in S_i}\mu_{ik}e^{\mu_{ik}/\gamma_i}}{\sum_{k\in S_i}e^{\mu_{ik}/\gamma_i}}\frac{\sum_{j\in S_i}e^{r\mu_{ij}/\gamma_i}}{(\sum_{k\in S_i}e^{\mu_{ik}/\gamma_i})^r}\bigg\} \\
&\quad = \frac{r}{\gamma_i^2}\frac{\sum_{j\in S_i}e^{r\mu_{ij}/\gamma_i}}{(\sum_{j\in S_i}e^{\mu_{ij}/\gamma_i})^r}\bigg\{-\frac{\sum_{j\in S_i}\mu_{ij}e^{r\mu_{ij}/\gamma_i}}{\sum_{j\in S_i}e^{r\mu_{ij}/\gamma_i}}+\frac{\sum_{j\in S_i}\mu_{ij}e^{\mu_{ij}/\gamma_i}}{\sum_{j\in S_i}e^{\mu_{ij}/\gamma_i}}\bigg\}.
\end{aligned}
$$

The derivative of $\frac{\sum_{j\in S_i}\mu_{ij}e^{r\mu_{ij}/\gamma_i}}{\sum_{j\in S_i}e^{r\mu_{ij}/\gamma_i}}$ with respect to r is

$$
\frac{1}{\gamma_i}\Bigg[\frac{\sum_{j\in S_i}\mu_{ij}^2 e^{r\mu_{ij}/\gamma_i}}{\sum_{j\in S_i}e^{r\mu_{ij}/\gamma_i}}-\bigg(\frac{\sum_{j\in S_i}\mu_{ij}e^{r\mu_{ij}/\gamma_i}}{\sum_{j\in S_i}e^{r\mu_{ij}/\gamma_i}}\bigg)^2\Bigg],
$$

and the bracketed term is nonnegative by Jensen's Inequality. Thus, $\frac{\sum_{j\in S_i}\mu_{ij}e^{r\mu_{ij}/\gamma_i}}{\sum_{j\in S_i}\mu_{ij}e^{r\mu_{ij}/\gamma_i}}$ is increasing in r and is equal to $\frac{\sum_{j\in S_i}\mu_{ij}e^{\mu_{ij}/\gamma_i}}{\sum_{j\in S_i}e^{\mu_{ij}/\gamma_i}}$ when $r=1$, so

$$
\frac{\sum_{j\in S_i}\mu_{ij}e^{r\mu_{ij}/\gamma_i}}{\sum_{j\in S_i}e^{r\mu_{ij}/\gamma_i}} \le \frac{\sum_{j\in S_i}\mu_{ij}e^{\mu_{ij}/\gamma_i}}{\sum_{j\in S_i}e^{\mu_{ij}/\gamma_i}}\ \forall r\in[0,1).
$$

Thus, $\frac{\partial}{\partial\gamma_i}\big[\sum_{j\in S_i} q_{j|i}(S_i,\gamma_i)^r\big] \ge 0$. Since $\frac{\partial}{\partial\gamma_i}Q_i(\mathcal{S},\boldsymbol{\gamma}) \ge 0$ also, we have $\frac{\partial}{\partial\gamma_i}\sum_{j\in N}x^*_{ij}(\mathcal{S},\boldsymbol{\gamma}) \ge 0$. To show that $\frac{\partial}{\partial\gamma_i}\sum_{j\in N}x^*_{lj}(\mathcal{S},\boldsymbol{\gamma}) \le 0$ for all $l\ne i$, note that

$$
\sum_{j\in N}x^*_{lj}(\mathcal{S},\boldsymbol{\gamma}) = \lambda Q_l(\mathcal{S},\boldsymbol{\gamma}) + \alpha\lambda^r\Phi^{-1}(1-c_l/p_l)Q_l(\mathcal{S},\boldsymbol{\gamma})^r\sum_{j\in S_i}q_{j|l}(S_l,\gamma_l)^r
$$

and $q_{j|l}(S_l,\gamma_l)^r$ does not depend on γ_i, and we have $\frac{\partial}{\partial\gamma_i}Q_l(\mathcal{S},\boldsymbol{\gamma}) \le 0$ since $\frac{\partial}{\partial\gamma_i}V_i(S_i,\gamma_i) \ge 0$, so the result follows.

A.6 Proof of Lemma 7

By Proposition 3, the limit of the expected profit as $\gamma_i \to 0$ for a given assortment $\mathcal{S} = (S_1, \ldots, S_m)$ and parameters $\boldsymbol{\gamma}_{-i} \in (0, 1]^{m-1}$ is

$$\lambda \frac{(p_i - c_i)e^{\max_{k \in S_i} \mu_{ik}} + \sum_{l \neq i}(p_l - c_l)V_l(S_l, \gamma_l)}{v_0 + e^{\max_{k \in S_i} \mu_{ik}} + \sum_{t \neq i} V_l(S_l, \gamma_l)}$$
$$- \theta_i \left(\frac{e^{\max_{k \in S_i} \mu_{ik}}}{v_0 + e^{\max_{k \in S_i} \mu_{ik}} + \sum_{l \neq i} V_l(S_l, \gamma_l)} \right)^r \sum_{j \in N} \mathbf{1}\Big(j = \operatorname*{argmax}_{k \in S_i} \mu_{ik}\Big)^r$$
$$- \sum_{l \neq i} \theta_l \left(\frac{V_l(S_l, \gamma_l)}{v_0 + e^{\max_{k \in S_i} \mu_{ik}} + \sum_{t \neq i} V_t(S_t, \gamma_t)} \right)^r \sum_{j \in S_l} \left(\frac{e^{\mu_{lj}/\gamma_l}}{V_l(S_l, \gamma_l)^{1/\gamma_l}} \right)^r.$$

Suppose that $S_i \in \mathcal{P}$ and S_i is nonempty, so that $1 \in S_i$. Then $\max_{k \in S_i^*} \mu_{ik} = \mu_{i1}$. In addition, since we are assuming that the deterministic utilities of products within each type are distinct, the sum $\sum_{j \in N} \mathbf{1}(j = \operatorname{argmax}_{k \in S_i} \mu_{ik})^r$ is simply equal to 1 and does not depend on the choice of S_i. Thus, as γ_i approaches zero, either $\emptyset$ or $\{1\}$ must be an optimal assortment for product type i.

BIBLIOGRAPHY

[1] G. Allon, A. Federgruen, and M. Pierson. Price competition under multinomial logit demand functions with random coefficients. *Management Science*, 2011. To appear.

[2] A. Alptekinoglu and C. J. Corbett. Mass customization vs. mass production: Variety and price competition. *Manufacturing & Service Operations Management*, 10(2):204–207, 2008.

[3] G. Aydin and E. L. Porteus. Joint inventory and pricing decisions for an assortment. *Operations Research*, 56(5):1247–1255, 2008.

[4] G. Aydin and J. K. Ryan. Product line selection and pricing under the multinomial logit choice model. Technical report, Stanford University, Stanford, CA, 2000.

[5] D. R. Bell and J. M. Lattin. Shopping behavior and consumer preference for store price format: Why large basket shoppers prefer EDLP. *Marketing Science*, 17(1):66–68, 1998.

[6] M. Ben-Akiva and S. Lerman. *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press in Transportation Studies, Cambridge, MA, 1994.

[7] A. Borsch-Supan. On the compatibility of nested logit models with utility maximization. *Journal of Econometrics*, 43:373–388, 1990.

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

[9] R. E. Bucklin and S. Gupta. Brand choice, purchase incidence, and segmentation: An integrated modeling approach. *Journal of Marketing Research*, 29(2):201–215, 1992.

[10] J.M. Davis, G. Gallego, and H. Topaloglu. Assortment optimization under variants of the nested logit model. *Operations Research*, 2013. To appear.

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

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

[13] G. Gallego and H. Topaloglu. Constrained assortment optimization for the nested logit model. *Management Science*, 2012. To appear.

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

[15] V. Gaur and D. Honhon. Assortment planning and inventory decisions under a locational choice model. *Management Science*, 52(10):1528–1543, 2006.

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

[17] C. P. Haugtvedt, P. M. Herr, and F. R. Kardes, editors. *Handbook of Consumer Psychology*. Psychology Press, New York, NY, 2008.

[18] D. Honhon, V. Gaur, and S. Seshadri. Assortment planning and inventory decisions under stock-out based substitution. *Operations Research*, 58(5):1364–1379, 2010.

[19] W. J. Hopp and Y. Xu. Product line selection and pricing with modularity in design. *Manufacturing & Service Operations Management*, 7(3):172–187, 2005.

[20] S. Jagabathula, V. Farias, and D. Shah. Assortment optimization under general choice. In *INFORMS Conference, Charlotte, NC*, 2011.

[21] L. A. Johnson and D. C. Montgomery. *Operations Research in Production Planning, Scheduling and Inventory Control*. Wiley, New York, NY, 1974.

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

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

[24] H. Li and W. T. Huh. Pricing multiple products with the multinomial logit and nested models: Concavity and implications. *Manufacturing & Service Operations Management*, 2011. To appear.

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

[26] R. D. Luce. *Individual Choice Behavior: A Theoretical Analysis*. Wiley, New York, NY, 1959.

[27] B. Maddah and E. K. Bish. Joint pricing, assortment, and inventory decisions for a retailers product line. *Naval Research Logistics*, 54(3):315–330, 2007.

[28] D. McFadden. Conditional logit analysis of qualitative choice behavior. In P. Zarembka, editor, Frontiers in Economics, pages 105–142. Academic Press, 1974.

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

[30] J. Meissner, A. Strauss, and K. Talluri. An enhanced concave program relaxation for choice network revenue management. *Production and Operations Management*, 2012. To appear.

[31] I. Mendez-Diaz, J. J. M. Bront, G. Vulcano, and P. Zabala. A branch-and-cut algorithm for the latent-class logit assortment problem. *Discrete Applied Mathematics*, 2010. To appear.

[32] A. Morales, B. E. Kahn, L. McAllister, and S. M. Broniarczyk. Perceptions of assortment variety: The effects of congruency between consumers' internal and retailers' external organization. *Journal of Retailing*, 81(2):159–169, 2005.

[33] P. Rusmevichientong, Z.-J. M. Shen, and D. B. Shmoys. A PTAS for capacitated sum-of-ratios optimization. *Operations Research Letters*, 37(4):230–238, 2009.

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

[35] P. Rusmevichientong, D. B. Shmoys, C. Tong, and H. Topaloglu. Assortment

optimization under the multinomial logit model with random choice parameters. Technical report, Cornell University, School of Operations Research and Information Engineering, 2013.

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

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

[38] H. Topaloglu. Joint stocking and product offer decisions under the multinomial logit model. *Production and Operations Management*, 2012. To appear.

[39] G. van Ryzin and S. Mahajan. On the relationship between inventory costs and variety benefits in retail assortments. *Management Science*, 45(11):1496–1509, 1999.

[40] R. Wang. Assortment management under the generalized attraction model with a capacity constraint. Technical report, HP Labs, Palo Alto, CA, 2012.

[41] R. Wang. Capacitated assortment and price optimization under the multinomial logit model. *Operations Research Letters*, 2012. To appear.

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

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

[44] D. Zhang and Z. Lu. Assessing the value of dynamic pricing in network revenue management. *INFORMS Journal on Computing*, 2011. To appear.