Product Line Design, Pricing and Framing under General Choice Models

Anran Li

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2018

===

ABSTRACT

Product Line Design, Pricing and Framing under

General Choice Models

Anran Li

This thesis handles fundamental problems faced by retailers everyday: how do consumers make choices from an enormous variety of products? How to design a product portfolio to maximize the expected profit given consumers' choice behavior? How to frame products if consumers' choices are influenced by the display location? We solve those problems by first, constructing mathematical models to describe consumers' choice behavior from a given offer set, i.e., consumer choice models; second, by designing efficient algorithms to optimally select the product portfolio to maximize the expected profit, i.e., assortment optimization. This thesis consists of three main parts: the first part solves assortment optimization problem under a consideration set based choice model proposed by Manzini and Mariotti (2014) [Manzini, Paola, Marco Mariotti. 2014. Stochastic choice and consideration sets. Econometrica 82(3) 1153-1176.]; the second part proposes an approximation algorithm to jointly optimize products' selection and display; the third part works on optimally designing a product line under the Logit family choice models when a product's utility depends on attribute-level configurations.

Contents

List of Figures

List of Tables

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my advisor Prof. Guillermo Gallego who brought me to the OR/OM field and guided me throughout the course of my doctoral study. He has broadened my horizon with his knowledge, insight, sharpness, dedication and most importantly passion for research. The example he sets, as a researcher and mentor, will be invaluable to me throughout my career and my life. I feel a deep sense of gratitude toward his generous inputs to me. I have been incredibly fortunate to work with him.

My sincere gratitude also goes to Professor Van-anh Truong for her exceptional support and guidance. She widened my research from new perspectives and let me see the beauty of algorithms. My research trajectory would have been completely different (probably in a bad way) if not with her guidance and encouragement.

I am also fortunate to encounter mentors and collaborators outside Columbia. Chapter 4 is originated from an industrial project when I did my internship at Hewlett Packard under the supervision of Jose Beltran. Professor Kalyan Talluri also offered me tremendous help while doing another project on demand estimation. I would like to thank them for their help with my personal and professional development.

I would like to thank Professor Jay Sethuraman for being an invaluable source of advice and support during my graduate study. My interests in being a researcher in OR was ignited when I studied my first optimization course with Professor Sethuraman. I am indebted to Professor Awi Federgruen as being an exceptional role model. Professor Federgruen's insight, rigor, and passion for research will enlighten me throughout my career. I am very grateful to Professor Srikanth Jagabathula for serving on my thesis committee. I have benefited a lot from Professor Jagabathula's Choice Modeling course.

It is my genuine fortune to learn from and inspired by many great professors at Columbia. I am extremely grateful to Professor Donald Goldfarb, Professor Ward Whitt, Professor Garud Iyengar, Professor Cliff Stein, Professor David Yao and Professor Vineet Goyal for all the insightful discussions, help and encouragement they provided me at various stages of my academic journey.

I sincerely thank my fellow officemate Xinshang Wang for the stimulating discussions, for the days we worked hard together to prove theorems and write papers. Xinshang makes me see the magic of peer collaboration. I hope our career paths will cross again in the future.

Life during the past few years would not have been as much enjoyable without my fellow colleagues and friends at Columbia. I would like to thank Jing Dong, Yanan Pei, Xinyun Chen, Juan Li, Linan Yang, Chun Wang, Cun Mu, Jing Guo, Chun Ye, Wenqi Hu, Taier Gao and all the other fellow colleagues for sharing with me the happiness and supporting me through the hardship. I have learnt so much from them. I would also like to thank friends I met at HKUST, especially Xuan Wang. Thanks to Xuan for her cheerful spirit, and for the great food and deals we have explored together. My sincere gratitude also goes to the IEOR departmental staff, who worked relentlessly to make our Ph.D experience seamless.

Last but not the least, I would like to thank my parents for their unconditional support and constant love. Thanks them for indulging my pursuit even without truly understanding my job. To them I dedicated the thesis.

Anran Li

July 26, 2017

To my parents

Chapter 1

Introduction

1.1 Motivation

Walmart removed 15% of its SKUs in 2008 because a survey showed that customers would like less cluttered stores. However, it was forced to roll back most of the changes due to significant sales decline. A gelato shop in Santa Monica, California, overwhelmed its customers with over 100 flavors. As evidenced by Walmart and the gelato shop, getting the right subset of products offered to customers is not easy, yet it is critical to retail success. This dissertation focuses on some of the most important questions asked by retailers: which set of products to offer, at what prices, and how to present items in the subset to customers?

The problem of finding an optimal subset of products that generates the highest expected revenue is referred to as the assortment optimization problem, or equivalently, the product line selection problem. Picking the best assortment is a balancing act, as adding an appealing product to an assortment might cannibalize the sales from more profitable items in the current offer set.

Assortment optimization problem cannot stand apart from consumer choice models. With the ability to capture how demands are substituted among products, choice models are key to assortment optimization. An effective choice model should strike a delicate balance between the accuracy in capturing the purchase behavior and the efficiency in estimating model parameters and optimizing the assortment. Most existing proposals are based on the assumption that customers can maximize their utilities by rationally analyzing all available items and picking the best, for example, the Multinomial Logit choice model (MNL), and the Nested Logit choice model (NL). However, empirical works (Ghose and Yang 2009, Goeree 2008) provide the evidence of bounded rationality, indicating that consumers have limited ability to analyze all items. Therefore, the notion of consideration set — a subset of the offered assortment over which a consumer will make utility comparisons before arriving at a final purchase decision — has gained wide acceptance since its introduction by Howard and Sheth (1969). This is because that consideration sets explain, behaviorally, consumers' limited ability to process or acquire information; methodologically, it has been shown that ignoring consideration sets may lead to biased parameter estimates (Chiang, Chib and Narasimhan 1999). Chapters 2 and 3 focus on consideration set based choice models, while they are differentiated by the process by which consideration sets are formed.

Chapter 2 looks at the formation of consideration sets based on consumers' attentions associated with products. A common experience for consumers when they go shopping is that they tend to pay more attention to certain goods. For example, clothes with bright colors might grab a customer's eyes at first sight, so the customer might be more likely to bring them to the fitting room. Manzini and Mariotti (2014) propose a random consideration set based choice model (RCS) where consideration sets are formed stochastically with each product entering the consideration set with a certain probability, and the probability measures how much attention is attached to that product. The other salient feature of the model is that all consumers have a common preference ordering for the products, so the differences in choices are entirely due to consideration sets heterogeneity. This behavioral assumption is consistent with empirical findings (Jagabathula and Rusmevichientong 2016) and is further supported by our airline partner's transactional data showing that consumers always tend to buy the cheapest ticket available. While the economics literature Manzini and Mariotti (2014) focuses on rationalizing the choice model, we are concerned with its operational applicability. We show how to recover the full ordering and attention probabilities of the RCS model from transaction data. Empirical testing of the RCS model on our airline partner's data shows that it outperforms the Multinomial Logit model in the Chi-square measure over all the instances; compared with the Mixture of Multinomial Logit model the Chi-square is reduced by 15.6% on average, revealing that a choice model developed from a behavioral perspective can model choices better than a traditional parametric model. We show that an assortment that maximizes expected revenues can be found efficiently by dynamic programming. Moreover, the dynamic program can be adapted to solve several extensions, like the cardinality constrained assortment problem and the problem of finding all efficient sets, i.e., the collection of optimal assortments under various marginal costs, which demonstrates the potential practical usage of the RCS model.

Chapter 3 is motivated by another common experience that consumer choices are influenced by how the products are framed, or displayed, referred as the framing effect, meaning that people react to a particular choice in different ways depending on how it is presented. This phenomenon is more common in the ecommerce online retail setting: it has been observed that consumers' attention to a display decreases exponentially with the display's distance to the top (Feng, Bhargava, and Pennock 2007). During online shopping, it is cognitively harder for a typical consumer to visit sellers who are listed at the bottom of a web page, before or in addition to visiting those who are listed at the top. Thus, positioning a brand or product at a top position on a listing can improve both consumer attention to the brand, and consequently, consumer selection of the brand. Despite substantial evidence suggesting the impact of framing on consumers' choice outcome, there are very few models that have attempted to capture these effects. In this chapter, we introduce one of the first models for product framing and the first one for pricing that accounts explicitly for these effects. We present a model where a set of products are displayed, or framed, into a set of virtual web pages and build up the model based on consideration sets. We assume that consumers consider only products in the top pages, with different consumers willing to see different numbers of pages. Consumers select a product, if any, from these pages following a general choice model. Products are organized into virtual pages. Each page can hold a finite number, say p, of products. A consumer will examine only the first X pages, where

X is a random variable that may be personalized to the consumer's profile. The consumer forms a consideration set consisting of only products in the examined pages. From this consideration set, consumer makes a choice according to a general choice model. Thus, products that are placed in earlier pages are more likely to be considered, and therefore purchased, than those that are placed in later pages. We show this problem is NP-hard to solve and propose fast, easy-to-implement algorithms with worst-case performance guarantees.

Assortment optimization and pricing attract much attention in business for its potential to increase profits. It is also attracting the attention of people in academia interested in looking at this problem with new ideas and new tools. Most of the current research solves the problem at the stock-keeping units (SKUs) level, assuming the set of potential products is given, i.e., they are working on product line selection problem. However, the product manager cares not only about improving existing products; but also discovering potentially lucrative new products. Yet this problem cannot be analyzed using the current SKUs based assortment models due to the ambiguity of the potential product set. Chapter 4 addresses the product line design and pricing problem. To disentangle this problem, we need to: first, construct a choice model that is able to forecast demand of existing as well as new products; second, to jointly search for assortments and prices to maximize the profit. Yet the difficulties are two fold: to forecast demand of products that has never existed and to design products from a large number of attributes that can be configured at different quality levels. Based on the available choices, the number of potential products are in the billions and this makes the product line optimization problem very challenging. For example, Hewlett Packard (HP) used to apply hedonic regression to extract economic value of the attribute-level and estimate a new configuration's market price, but this procedure neglects the impact of a product configuration on consumers' demand. For example, a customer who buys a 17 inches HP laptop with core i7 processor at $1500 may choose to buy a 15 inch Mac book, if HP's price is increased by 5%. To deal with this problem, we apply an attribute-level dependent random utility maximization choice model positing that a consumer's utility for a product is not a direct function of the product itself, but instead depends on the product's attributes and the consumer's own tastes. Given a consumer's attribute-level's taste, we are able to forecast a new product's demand, as a new product merely is a new combination of attributes. We construct a graphical representation of the configurations and discover that optimal designs can be efficiently found by network flow algorithms under Multinomial Logit and Nested Logit models.

1.2 Background

In this section, we briefly introduce the dynamics of the choice models we study and the formulations of assortment problems.

1.2.1 Consumer Choice Models

Logit Family Models

The MNL model proposed by McFadden (1974) has attracted tremendous theoretical and empirical interests. It is derived from a random utility model where the random components are independent Gumbel random variables. The MNL model can characterize many choice situations conveniently, since it can provide a closed form for the underlying choice probability. Although criticized because of the independence of irrelevant alternatives (IIA) property, i.e., preferences for item A or B should not be changed by the inclusion of item X, it has remained a sustained assumption in many applications.

The standard MNL model has the following structure:

$$U_i = v_i + \varepsilon_i,$$

where U_i is the random utility of product i, which is equal to the sum of the expected utility, denoted by v_i, and the ε_i are independent standard Gumbel random variables.

Under this utility structure, the probability of a customer choosing product i from assortment S is

$$\pi_i(S) = \frac{V_i}{V_0 + \sum_{j \in S} V_j},$$

where $V_j = exp(v_j)$ is the attractiveness of product j; $V_0 = exp(v_0)$ is the attractiveness of the outside alternative.

The Multinomial Logit model has many applications due to its ease of estimation and optimization. Standard optimization packages can efficiently estimate parameters of the MNL model by maximizing the log-likelihood function, as the objective function is concave (McFadden 1974). Davis, Gallego and Topaloglu (2014) show that the assortment optimization can be easily solved by a linear program and the linear program can be extended to incorporate all the totally unimodular side constraints. Yet, the independence of irrelevant alternative (IIA) property (Luce 1959) is a serious limitation, meaning the ratio of the probabilities of choosing any two alternatives is independent of the attributes of any other alternative in the choice set. Debreu(1952) was among the first economists to discuss the implausibility of the independence from irrelevant alternatives assumption. The Nested Logit (NL) model, introduced by Williams (1977), has been developed to relax the assumption of independence between all the alternatives. Under the nested logit model, customers first select a nest, i.e., a subset of offered products that shares some common features, and then, an alternative within the selected nest, thus NL allows different substitution patterns within and between nests.

Given the nest structure, the conditional probability of an incoming consumer selecting product i from nest n under assortment S (let us call it product in for short) follows the MNL form

$$\pi(in|n,S) = \frac{exp(v_{in}/\gamma_n)}{\sum_{jn\in S_n} exp(v_{jn}/\gamma_n) + V_n},$$

where S_n is the subset of offered products belonging to nest n, and the probability of selecting from nest n is

$$\pi(n|S) = \frac{(\sum_{jn \in S_n} exp(v_{jn}/\gamma_n) + V_n)^{\gamma_n}}{\sum_{n'}(\sum_{jn' \in S_{n'}} exp(v_{jn'}/\gamma_{n'}) + V_{n'})^{\gamma_{n'}} + V_0}.$$

The probability of purchasing product i from nest n will be the product of these two probabilities

$$\pi(in|S) = \pi(in|n,S)\pi(n).$$

NL model assumes correlations of valuations among alternatives within a same nest while independence of valuations among products from different nests. γ measures the relative independence of the alternatives within the same submarket. The choice model parameter estimation becomes difficult because the log likelihood function is no longer jointly concave on expected utility v and the dissimilarity parameter γ. As a result, McFadden and Train (2000) recommend an iterative approach, which alternates between optimizing v and γ. Convergence to a stationary point is guaranteed because each sub-problem has a concave objective. Davis, Topaloglu and Gallego (2014b) show that the assortment problem under NL model can be solved by a linear program when the nest dissimilarity parameters of the choice model are less than one and the customers always make a purchase within the selected nest. Relaxing either of these assumptions renders the problem NP-hard.

Random Consideration Set Model

Let $N \equiv \{1,2,...,n\}$ be the set of products. The random consideration set (RCS) model developed by Manzini and Mariotti (2014) is characterized by a strict preference order $\prec$ on N, and by a vector λ of positive attention probabilities. We assume without loss of generality that the products are labeled so that $1 \prec 2 \prec 3 \prec \ldots \prec n$. For any subset $S \subset N$, the probability that product $i \in S$ is selected is given by

$$\pi_i(S) = \lambda_i \Pi_{j \succ i, j \in S}(1 - \lambda_j) \quad \forall \quad i \in S,$$

with $\pi_i(S) = 0$ if $i \notin S$. In words, $i \in S$ is chosen if and only if i is the highest ranked product in the random consideration set $C(S)$ that results from paying attention to product $j \in S$ with probability λ_j.

Clearly $\pi_i(\{i\}) = \lambda_i > 0$ and $1 - \lambda_i = \pi_0(\{i\})$ where 0 denotes the no-purchase alternative. This suggests that consumers include product i in a consideration set with probability λ_i equal to the probability that product i is preferred to the no-purchase alternative. But the λ_is may also arise from a filtering process that may entail several steps with λ_i measuring the probability that all steps in the process have been passed. Given a set S, consumers will select the no-purchase alternative only if no product in S is included in the consideration set. As a result,

$$\pi_0(S) = \Pi_{j \in S}(1 - \lambda_j),$$

with $\pi_0(S) > 0$ if and only if all the products $j \in S$ have positive inattention probabilities $1 - \lambda_j > 0$.

1.2.2 Assortment Problems and Its Variants

The assortment problem chooses a single set of products offered to customers to maximize the expected revenue or profit. Formally speaking, the retailer needs to select a subset S from a global product set $N = \{1, ..., n\}$, where each product i can generate revenue r_i. Therefore the assortment problem can be formulated as follows:

$$\max_{S \subseteq N} r_i \pi_i(S).$$

The difficulty of this optimization problem comes from its combinatorial nature. Since we cannot find the optimal assortment by enumerating all possible offer sets, we must resort to efficient algorithms.

Chapter 2 considers the pure assortment problem and its variants: assortment problem with cardinality constraints; as well as assortment pricing problem. Chapter 3 incorporates another dynamic into the assortment problem — products' display decision, as consumers' consideration sets will be influenced by the way that product are displayed. Chapter 4 studies product line design, so it optimizes configurations by assigning attribute-level directly to products.

Chapter 2

Attention, Consideration then Selection

2.1 Introduction

An effective choice model should strike a delicate balance between demand prediction accuracy and efficiency in parameter estimation and assortment optimization. In this chapter we study a consideration set based choice model that possesses these important properties. A consideration set is a subset of the offered assortment over which a consumer will make utility comparisons before arriving at a final purchase decision. There are two key ingredients in consideration set based choice models — the formation of consideration sets and the ordering of products' preference given a consideration set.

We are inspired by the random consideration set (RCS) model of Manzini and Mariotti (2014) emanating from the economics literature. Consideration sets are formed stochastically in this model with each product entering the consideration set with a certain probability. The other salient feature of the model is that all consumers have a common preference ordering for the products, so the randomness of choice is entirely due to consideration sets heterogeneity. While the economics literature focuses on rationalizing choice models (Eliaz and Spiegler 2011a, 2011b, Manzini and Mariotti 2014), we are concerned with the operations applicability of this model including demand estimation, assortment optimization with and without cardinality constraints, and optimal pricing for the Manzini and Mariotti (2014) RCS choice model.

Consideration sets are often justified as a form of bounded rationality and have ample support from the empirical literature, for example, Feng, Bhargava, and Pennock (2007), Agarwal, Hosanagar and Smith (2009) and Ghose and Yang (2009) in the online advertising setting; Chandon, Hutchinson, Bradlow and Young (2009), Corstjens and Corstjens (2012) in the brick-and-mortar retail setting. A vast literature has observed that the consideration set structure is in itself a significant explanatory factor of choice heterogeneity (Hauser 1978, Jagabathula and Rusmevichientong 2016).

2.1.1 Overview

The RCS model assumes that the composition of consideration sets is stochastic, where each alternative i is considered with a probability λ_i. A physical interpretation is that λ_i may indirectly measure the degree of brand awareness for a product, or the degree of salience of the product's display position, or the willingness of a consumer to seriously evaluate an alternative. An alternative explanation is that λ_is reflect the mood of consumers for product i. As is pointed out in the Manzini and Mariotti (2014), the assumption of λ_i being menu independent is a substantive one. It does have, however, empirical support in some contexts (Cheremukhin, Popova and Tutino 2011). At a theoretical level, the hypothesis of independent attention parameters is a natural starting point for analysis.

Essentially, the heterogeneity tends to appear mostly during the first stage of the purchasing decision when forming the consideration sets, whereas the preference ordering at the second stage are observed to be consistent and concentrated around one or a few central permutations over products. For example, our airline partner observed that consumers always tend to buy the cheapest ticket available. Exceptions to this behavior may be explained by consumer being unaware of the low fare when they buy a higher priced product even a lower fare is available. We now make some remarks about the plausible full order among products. When products are perceived to be a commodity or of the same quality it makes sense to rank them in decreasing order of prices. Consumers may also rank the products by their net value, or may rank products in decreasing order of quality but pay attention only to those products that fall within their budget constraint. Even the unique-ranking order model is quite rich and in general computationally intractable. Specifically, it was recently shown by Aouad, Farias, Levi and Segev (2015) that assortment optimization under the unique-ranking model is NP-hard to approximate within any constant factor except for some special cases (Aouad, Farias and Levi 2016, Honhon, Jonnalagedda and Pan 2012).

Our contributions in this chapter are as follows:

- We show how to recover the full ordering and attention probabilities of the RCS model given accurate estimates of choice probabilities or from empirical data. First we give the necessary condition and sufficient condition for assortments' variability in order to uniquely identify the model parameters, and then we introduce an alternating minimization algorithm which can discover model parameters efficiently in the empirical setting. A detailed analysis of data from our airline partner reveals that RCS model provides a significantly better fit than the Logit family model. Indeed, in all of the instances studied the RCS model outperformed the Multinomial Logit model in the Chi-square measure; compared with the Mixture of Multinomial Logit model, the Chi-square is reduced by 15.6% on average, revealing that a choice model developed from a behavioral perspective can model choices better than a traditional parametric model.

- We show that an assortment that maximizes expected revenues can be found in polynomial time by a dynamic program and the algorithm can be easily extended to solve a cardinality constrained assortment problem. In addition, we show that all of the efficient sets can be found in polynomial time. To our knowledge, our RCS model is the only choice model besides MNL and the nested MNL that admits efficient estimation, assortment optimization, capacitated assortment optimization and efficient sets discovery algorithms.

- We extend the Manzini and Mariotti RCS model to allow ties in preferences and show that a revenue-ordered assortment has a $1/2$ performance guarantee relative to the optimal assortment.

- We study the pricing problem where both the preference ordering and attention probability are price aware and show under mild assumptions that optimal profits are such that both the profit contributions and the net value to consumers are aligned with the value gap, defined as the difference between the value of the product to consumers and its unit wholesale cost.

Model Overview

Under the RCS choice model, customers include product i in their consideration set with probability λ_i. Once the consideration set is formed, consumers select the highest ranked product in the consideration set. As an example, suppose that we offer assortment $S = \{1,2,3\}$ with preference ordering $1 \prec 2 \prec 3$. The no-purchase alternative 0 is always available and will be chosen when the consideration set is null. Table 2.1 illustrates the consideration probabilities and the choice outcome for all possible consideration sets:

Table 2.1: Formation Probability and Final Choice under Different Consideration Sets

Consideration Set	Choice	Formation Probability
$\emptyset$	0	$(1-\lambda_1)(1-\lambda_2)(1-\lambda_3)$
$\{1\}$	1	$\lambda_1(1-\lambda_2)(1-\lambda_3)$
$\{2\}$	2	$(1-\lambda_1)\lambda_2(1-\lambda_3)$
$\{3\}$	3	$(1-\lambda_1)(1-\lambda_2)\lambda_3$
$\{1,2\}$	2	$\lambda_1\lambda_2(1-\lambda_3)$
$\{1,3\}$	3	$\lambda_1(1-\lambda_2)\lambda_3$
$\{2,3\}$	3	$(1-\lambda_1)\lambda_2\lambda_3$
$\{1,2,3\}$	3	$\lambda_1\lambda_2\lambda_3$

We can form the probability that product $i \in S$ is selected by adding the formation probabilities of consideration sets that lead to $i \in S$ as the final choice as illustrated in Table 2.2.

Table 2.2: Selection Probabilities for $S = \{1,2,3\}$

Choice	Probability
0	$(1-\lambda_1)(1-\lambda_2)(1-\lambda_3)$
1	$\lambda_1(1-\lambda_2)(1-\lambda_3)$
2	$\lambda_2(1-\lambda_3)$
3	λ_3

The RCS model is able to account for some plausible choice behaviors that cannot be captured by Luce's rule (Luce 1959) stating that the probability of selecting one item over another from an assortment of many items is not affected by the presence or absence of other items in the pool, as indicated by our numerical studies. We next present two motivating examples for illustration.

Choice Frequency Reversal In practice, we see choice frequency reversals, where product i is preferred to j in assortment S but j is preferred to i in a larger assortment. Reversals can be explained in the context of the RCS model when an alternative may be chosen more frequently than another by virtue of the attention paid to it as well as of its ranking as explained in the following table:

Table 2.3: Example of Choice Probability Reversion

Fare Classes	**Offer Set** $\{P,A,B\}$	**Offer Set.** $\{P,A\}$
Economic Plus (P)	**20%**	**20%**
Economic (A)	**5.6%**	**56%**
Economic (B)	**72%**	**N.A.**

In Table 2.3, an economic plus (P) ticket, which offers a free meal and free luggage handling may be chosen less frequently than an economic (A) ticket priced $5 lower without these ancillary services. This behavior may be due to the relatively low attention given to the more expensive economy plus ticket. The attention probabilities $\lambda_P = 0.2$ and $\lambda_A = 0.7$ result in the third column of Table 2.3 when the offer set is $\{A, P\}$). Suppose that a new economic class (B) ticket with attention probability $\lambda_B = 0.9$ is introduced at the same price as the original economic (A) ticket with free luggage handling resulting in the full order ranking $A \prec B \prec P$. Then the old economic (A) ticket will not be chosen whenever the new product is considered, so the introduction of the new economy (B) ticket can reverse the frequency with which the original economy (A) ticket and the economy plus ticket (P) are chosen. The basis for the choice frequency reversal in the RCS model is that while a better alternative may be chosen with lower probability than an inferior alternative in pairwise contests due to low attention for the superior product, the presence of an alternative that is mediocre among the three will reduce the probability that the inferior alternative is selected without affecting the probability that the superior alternative is noticed.

Red Bus, Blue Bus Paradox We now look into the classical red and blue buses example (Debreu 1960) where consumers select the train with probability 1/2 when only one of the buses is available. Under the MNL the inclusion of the second identical bus will change the probabilities to 1/3 for each of the three alternatives instead of keeping the probability of the train unchanged and splitting the 1/2 probability among the two buses.

Table 2.4: Example: Red Bus, Blue Bus Paradox

Transportation	**Choice Prob. wt. All Options Available**	**Choice Prob. wt. Blue Bus Removed**
Train	50%	50%
Red Bus	25%	50%
Blue Bus	25%	**N.A.**

The RCS model can resolve the red bus, blue bus paradox. To be more specific, let the transportations be ordered such that $Train \succ RedBus \sim BlueBus$. Note that $RedBus \sim BlueBus$ indicates that the choice probability will be equally split between when both can get chosen. And let the attention probability be such that $\lambda_T = 1/2$, $\lambda_{RB} = 1$, $\lambda_{BB} = 1$; therefore, the removal of blue bus will not affect the choice probability of train.

2.1.2 Related Works

Research on assortment optimization begins with the pioneering study by van Ryzin and Mahajan (1999). They show that the optimal assortment under a multinomial logit consumer choice model consists of a certain number of the highest utility products when they are equally profitable. Most of the focus has been on variants of this problem under the class of random utility maximization (RUM) models such as the Multinomial Logit (MNL) model, the Nested Logit (NL) model, and the Mixture of MNLs (MMNL) model. We refer readers to Talluri and van Ryzin (2004), Davis, Gallego and Topaloglu (2014), Desir and Goyal (2013) for a comprehensive review. Recent work in assortment optimization turns attention to new probabilisitic models leading to tractable assortment optimization (Blanchet, Gallego and Goyal 2016).

There are also several papers concerned about assortment optimization under consideration sets. Some work looks at endogenous consideration sets, where consumers pay a search cost to learn the value of a product and sequentially solve a optimal stopping time problem. Cachon (2005) shows that in equilibrium, the seller needs a larger assortment to attract more consumers, so ignoring consumer search will lead to less assortment variety. Sahin and Wang (2015) also study the assortment-optimization problem under endogenous consideration sets. They assume consumers are homogeneous in search cost and their search sequence is predetermined by all the products' expected utilities, which are common knowledge. There are a few papers studying assortment optimization under exogenous consideration sets. Feldman and Topaloglu (2015) study a model in which consumers of different types have different consideration sets, where the sets are fixed and nested, and each type of consumers chooses products according to a MNL model. They devise a fully polynomial-time approximation scheme for the assortment optimization problem. There are also papers that look at framing-dependent consideration sets. Davis, Topaloglu and Williamson (2015) study a problem in which a firm must sequentially add products to its assortment over time, so as to monotonically increase consumers' consideration sets. Gallego, Li, Truong and Wang (2016) model a phenomenon where in the online shopping setting, different consumers tend to view different number of pages and provide an algorithm with constant performance ratio relative to the optimal independent of inputs. Jagabathula and Rusmevichientong (2016) look into a two-stage choice process, where consumers consider the set of products with prices less than the threshold in the first stage and then choose the most preferred product from the set considered. They develop a nonparametric expectation maximization (EM) algorithm to fit the model and design a polynomial time approximation scheme to the joint assortment pricing problem.

Consideration set based choice models are closely related with preference list models as each preference list uniquely defines a consideration set. In order to be parsimonious, Rusmevichiengtong, Van Roy and Glynn (2006), Farias, Jagabathula and Shah (2013), van Ryzin and Vulcano (2015) assume the choice probability arises from a sparse distribution over preference lists. In this line of research, Farias, Jagabathula and Shah (2013) develop a robust estimation methodology by maximizing the worst case revenue. van Ryzin and Vulcano (2015) propose a "market discovery" algorithm: starting from an initial collection of preference lists, the support of the distribution is enlarged iteratively by generating a preference list that increases the log-likelihood value, using the column generation philosophy. In fact, sparsity of the preference list distribution is generally insufficient to alleviate the computational hardness of assortment optimization, and the problem was shown to be NP-hard to approximate by Aouad, Farias, Levi and Segev (2015). Honhon, Jonnalagedda and Pan (2012) develop tailor-made dynamic programming ideas for several special cases, which are subsumed by Aouad, Farias and Levi (2016), where customer preference lists are associated with structured set systems defined over a single preference ordering of the products. Essentially they require the variety of the consideration sets is in the order of $O(\log(N))$ where N is the number of products in order to keep the dynamic program tractable. However, our consideration sets' variety can grow exponentially with respect to the number of products. Our model differs from Aouad, Farias and Levi (2016) along another dimension. While Aouad, Farias and Levi (2016) focus on classical consideration set based models from the marketing literature, where consideration sets are generated from screening rules, for example, conjunctive or disjunctive rules (Pras and Summers 1975, Gilbride and Allenby 2004), we observe the bounded rationality among consumers as some available alternatives might be ignored even after the screening phase due to limited attention. Paul, Topaloglu and Feldman (2016) model choosy consumers whose rankings contain at most two products and show that even under this case the assortment optimization problem is NP-hard. Paul, Davis and Feldman (2016) assume the set of consumer classes is derived from paths in a tree where the order of nodes visited along each path gives the corresponding preference list, so the substitution will only be among neighboring products.

Moving to assortment pricing, some parametric structures need to be enforced. The most commonly used is the MNL model assuming linear utility structure (Hanson and Martin 1996, Song and Xue 2007). Under the MNL model with uniform price-sensitivity, the markup, defined as price minus cost, has been proven to be constant across all products at optimality (Anderson, de Palma and Thisse 1992, Hopp and Xu 2005, and Gallego and Stefanescu 2011). Li and Huh (2011) extend the concavity result to the NL model. Gallego and Wang (2014) extends the constant markup property to a general NL model with product-differentiated price-sensitivity parameters and arbitrary nest coefficients. More recently, Gallego, Li, Truong and Wang (2016) jointly price and design the product's display configuration when a framing effect influences consumer's choices. They show that products should be framed by sorting them in an increasing order of their value gap, which is essentially a product's surplus when it is priced at the wholesale cost. In addition, they show that at the optimum the products on a given page have a common markup and the markups increase with the page number.

2.2 Review of Random Consideration Set Model

Let $N \equiv \{1, 2, ..., n\}$ be the set of products. The random consideration set (RCS) model developed by Manzini and Mariotti (2014) is characterized by a strict preference order $\prec$ on N, and by a vector λ of positive attention probabilities. We assume without loss of generality that the products are labeled so that $1 \prec 2 \prec 3 \prec ... \prec n$. For any subset $S \subseteq N$, the probability that product $i \in S$ is selected is given by

$$\pi_i(S) = \lambda_i \Pi_{j \succ i, j \in S}(1 - \lambda_j) \quad \forall \quad i \in S, \tag{2.2.1}$$

with $\pi_i(S) = 0$ if $i \notin S$. In words, $i \in S$ is chosen if and only if i is the highest ranked product in the random consideration set $C(S)$ that results from paying attention to product $j \in S$ with probability λ_j.

Clearly $\pi_i(\{i\}) = \lambda_i > 0$ and $1 - \lambda_i = \pi_0(\{i\})$ where 0 denotes the no-purchase alternative. This suggests that consumers include product i in an assortment with probability λ_i equal to the probability that product i is preferred to the no-purchase alternative. But the λ_is may also arise from a filtering process that may entail several steps with λ_i measuring the probability that all steps in the process have been passed. Given a set S, consumers will select the no-purchase alternative only if no product in S is included in the consideration set. As a result,

$$\pi_0(S) = \Pi_{j\in S}(1-\lambda_j),$$

with $\pi_0(S) > 0$ if and only if all the products $j \in S$ have positive inattention probabilities $1-\lambda_j > 0$. Let us assume all the attention probabilities λs are strictly within the unit interval.

Manzini and Mariotti (2014) propose two choice axioms and demonstrate that a choice model satisfies the two choice axioms if and only if it satisfies (2.2.1). The axioms apply to all subsets $S, T \subseteq N$ such that $i, j \in S \cap T$, $i \neq j$.

- I-ASYMMETRY: $\frac{\pi_i(S-\{j\})}{\pi_i(S)} \neq 1 \Rightarrow \frac{\pi_j(S-\{i\})}{\pi_j(S)} = 1$,
- I-INDEPENDENCE: $\frac{\pi_i(S-\{j\})}{\pi_i(S)} = \frac{\pi_i(T-\{j\})}{\pi_i(T)}$ and $\frac{\pi_0(S-\{j\})}{\pi_0(S)} = \frac{\pi_0(T-\{j\})}{\pi_0(T)}$.

I-Asymmetry states that if the presence of product j influences the probability of i being chosen, then $i \prec j$, and therefore the presence of i does not influence the choice of j. I-Independence says that the impact of one product on another is menu-independent. Similar to Luce's IIA (Luce 1959) property, except that instead of relating to the odd ratios $\frac{\pi_i(S)}{\pi_j(S)}$ being menu-independent, it means that the impact $\frac{\pi_i(S-\{j\})}{\pi_i(S)}$ is menu-independent. The main result from Manzini and Mariotti (2014) is the following theorem:

Theorem 2.2.1 (Manzini & Mariotti)**.** *A random choice rule satisfies I-Asymmetry and I-Independence if and only if it is a random consideration set rule* $\pi_{\prec,\lambda}$*. Moreover, both* $\prec$

and λ *are unique, that is, for any random choice rule* $\pi_{\prec',\lambda'}$ *such that* $\pi_{\prec',\lambda'} = \pi$*, we have* $(\prec',\lambda') = (\prec,\lambda)$.

Therefore, a set of choice probabilities will uniquely define a random consideration set choice rule $(\prec,\lambda)$.

As is mentioned by Manzini and Mariotti (2014), a preference relation $j \succ i$ can be defined by $\pi_i(S-\{j\}) > \pi_i(S)$ for some $S \subseteq N$ and will be revealed uniquely. Moreover, given a preference order $\prec$, the attention probability can be calculated by the following formula:

$$\lambda_i = 1 - \sqrt{\frac{\pi_0(\{i,j\})\pi_0(\{i,k\})}{\pi_0(\{j,k\})}}.$$

According to Manzini and Mariotti (2014), to infer the preference relationship between i and j we need to know choice probabilities $\pi_i(S-\{j\})$ and $\pi_i(S)$ for at least one set S containing both i and j. And to identify the attention probabilities, it is sufficient to know the no-purchase probabilities for all binary menus. In the next section, we take a closer look into the identification problem. More precisely we are interested in finding minimal information structures that allow the unique identification of the preference order and the attention probabilities assuming that choices come from a random consideration rule $\pi_{\prec,\lambda}$. Once we identify such information structure we will show how to infer the preference order and the attention probabilities.

2.3 Estimation

In this section we look into the parameter identification problem of the RCS model under both noiseless and noisy settings. First, we focus on statistical identifiability of the model parameters under the noiseless setting where for any $i \in S_+$ we know the exact choice probability $\pi_i(S)$. We then turn our attention to empirical estimation where we have transaction data of the choices made by consumers over time under assortments $S_t, t = 1, \ldots, T$.

2.3.1 Identifiability

As is shown in Manzini and Mariotti (2014), the random choice rule $(\prec, \lambda)$ can be uniquely identified under certain information structures. In this section, we look further into the information structure required to uniquely identify $(\prec, \lambda)$.

Example 2.3.1. *To develop intuition we start by looking at a couple of information structures. Suppose we are given accurate choice probabilities under a collection of assortments:*

1. *Consider the collection of assortments* $\mathcal{S} = \{N, N-\{1\}, N-\{2\}, ..., N-\{n\}\}$*;*

   The preference order is easy to discover since $\pi_i(N-\{j\}) > \pi_i(N)$ *for all* $i \prec j$ *and* $\pi_i(N-\{j\}) = \pi_i(N)$ *for all* $i \succ j$*. And the attention probability can be calculated after the preference order is discovered since* $1-\lambda_j = \frac{\pi_i(N)}{\pi_i(N-\{j\})}$ *if* $i \prec j$*.*

2. *Consider the collection of assortments* $\mathcal{S} = \{N, \{1\}, \{2\}, ..., \{n\}\}$*;*

   Without loss of generality, let us assume products are labeled as $1 \prec 2 \prec \ldots \prec n$*.*

Product n, the "best" product, can be identified immediately as $\pi_n(N) = \pi_n(\{n\})$ *and* $\pi_j(N) < \pi_j(\{j\})$ *for all* $j \neq n$*. After we identify n, the product at the top of the preference list, we remove it from set N and update the information, in other words, we can infer choice probabilities of all products under assortment* $N - \{n\}$*, as* $\pi_j(N - \{n\}) = \frac{\pi_j(N)}{1-\lambda_n}$*. Now, as product* $n-1$ *is at the top of preference list, we can immediately identify it since it is the only product such that* $\pi_j(\{j\}) = \pi_j(N - \{n\})$*. We repeat the procedures and sequentially identify all products' preference order and attention probabilities.*

3. *There are 3 products with preference ordering* $1 \prec 2 \prec 3$ *and we are only given choice probabilities under a set of menus* $\mathcal{S} = \{\{1,2\},\{3\}\}$*. The preference list cannot be identified because there is not enough information about the change in probability caused by adding or removing a product.*

The following result uncovers the minimal information requirements to uniquely identify the model parameters. Minimal information requirements mean that any subset of the information cannot fully identify model parameters.

Theorem 2.3.2 (Necessary Condition)**.** *The parameters of the RCS model can be identified only if we are given the choice probabilities under a collection of assortments where every two products who rank consecutively in the preference list are offered together at least once and separately at least once.*

Proof. **Necessity:** Suppose that products i and j rank consecutively with $i \prec j$ and $\nexists k \in N, k \neq i, j$ and $i \prec k \prec j$. If products i and j are always offered together, then we cannot check if $i \prec j$ or $j \prec i$. This is because the presence or absence of other products have the same impact on both i and j. Moreover, we cannot see the influence to i's purchase probability due to removal of j, nor we can see the influence to j due to removal of i, so the relative ordering of i and j cannot be identified. This shows that it is necessary to offer the two products separately at least once. On the other hand, if i and j are never offered together, we cannot see the impact to the purchase probability of one product due to the presence of the other product, so we cannot infer the ranking between them. This is illustrated in the third instance in Example 2.3.1. □

Theorem 2.3.3 (Sufficient Condition)**.** *The parameters of the RCS model can be identified if first, we are given the choice probabilities under a collection of assortments where every two products who rank consecutively in the preference list are offered together at least once and separately at least once. Second, for each product $i \in N$, there exist at least two offered subsets $S, S' \subseteq N$ such that $i \in S \cap S'$ and $S \neq S'$; let $L_i(S) = \{j \in S : j \succ i\}$, then the choice probability of the no-purchase $\pi_0(L_i(S)) \neq \pi_0(L_i(S'))$ as long as $L_i(S) \neq L_i(S')$.*

Proof. **Sufficiency:** Suppose that two conditions hold. Without loss of generality, let us assume the underlying true preference order is $1 \prec 2 \prec \ldots \prec n$. We can identify product n because $\pi_n(S) = \lambda_n$ for all $S \subseteq N$ such that $n \in S$. And we argue it is the only product whose choice probability is menu independent. This is due to the other assumption $\pi_0(L_i(S)) \neq \pi_0(L_i(S'))$, because for any product i, $\pi_i(S) = \lambda_i \pi_0(L_i(S))$. This implies that if the choice probability of a product is menu independent then it must be at the top of the preference ordering. After identifying product n and its attention probability λ_n, we remove product

n from all S such that $S \ni n$ and update the probability information to $\pi_j(S-\{n\}) = \frac{\pi_j(S)}{1-\lambda_n}$ for all $j \in S-\{n\}$. Since product $n-1$ is at the top of the preference list at this moment, it will be the only product whose choice probability is menu-independent. This allow us to identify product $n-1$ and its attention probability λ_{n-1}. By repeating the identification-elimination procedures we can identify the preference order and attention probabilities of all of the products. □

2.3.2 Empirical Estimation

The setting considered in the previous section assumes that we know all of the choice probabilities for every assortment in the collection $\mathcal{S}$ that satisfies the conditions of Theorem 2.3.3. Now we transfer our attention to parameter estimation in the real setting, where we are only given the historical transaction data.

Data Model

We assume that available data is in the form of a sequence of consumer choices over offered assortments. More specifically, consumers sequentially arrive over T arrival epochs. At epoch $t = 1, ..., T$, we observe (c_t, S_t), where S_t is the assortment offered to the consumer t and $c_t \in S_t \cup \{0\}$ is the chosen product. Assortments can change over time due to stockout or managerial decisions. The type of data is standard in the operations literature (see Jagabathula and Rusmevichientong 2016, van Ryzin and Vulcano 2015). Particularly, we assume we can observe the selection of the no-purchase option. This corresponds to the case where on-line retailers can record consumers visiting their web-sites, or brick- and-mortar stores can track consumer flows by monitors. Without the no-purchase information will require the joint estimation on market size and attention probability, which has been well studied in the literature (see Haensel, Koole and Erdman 2011).

Estimation Methodology

We use the maximum likelihood estimation (MLE) method to estimate our model parameters. Our choice model is defined by the preference ordering $\prec$ and attention probabilities $\lambda_i, i \in N$. Let $x_{ij} \in \{0,1\}$ denote the preference order of products i and j, i.e., $x_{ij} = 1$ if $i \succ j$ and $x_{ij} = 0$ otherwise. With this notation we can write the purchase probability of product j under assortment S as $\pi_j(S) = \prod_{i \in S-\{j\}} (1-\lambda_i)^{x_{ij}} \lambda_j$.

Assuming that each data point is generated from an independent draw of the choice model, then the log likelihood is given by

$$\mathcal{L}(x, \lambda | Data) = \sum_{t=1}^{T} \sum_{i \in S_t} x_{ic_t} \times \log(1-\lambda_i) + \log \lambda_{c_t}.$$

From the likelihood function we see that a higher purchase probability can be due to ranking at the upper level of the preference list or having larger attention probability, or both. We can model the problem of maximizing the log likelihood function as:

$$\max_{x,\lambda} \quad \sum_{t=1}^{T}\sum_{i\in S_t} x_{ic_t} \times \log(1-\lambda_i) + \log\lambda_{c_t} \tag{2.3.1}$$

$$\text{s.t.} \quad x_{ij} + x_{ji} = 1 \quad \forall i,j \in N, i \neq j \tag{2.3.2}$$

$$x_{ij} + x_{jk} + x_{ki} \leq 2 \quad \forall i,j,k \in N, i \neq j \neq k \tag{2.3.3}$$

$$x_{ij} \in \{0,1\} \quad \forall i,j \in N$$

$$\lambda_i \in (0,1) \quad \forall i \in N,$$

while $x_{i0} = 1$ are fixed for all $i \in N$ and $\lambda_0 = 1$, meaning consumers are always aware of the no-purchase alternative.

The first equality constraint (2.3.2) ensures that for any two different products i and j, either $i \prec j$ or $j \prec i$, one and only one of them can happen. The second inequality constraint (2.3.3) ensures the linear ordering among the products' preference, indicating the transitivity among the products' preferences. The third constraint requires the preference indicator x_{ij} to be binary, and the last constraint requires the attention probabilities to be well defined.

The Mixed Integer Nonlinear program is in general hard to solve. As a result we settle for finding a local maximum or a stationary point.

Observe that if we fix the value of x, the objective (2.3.1) is strictly concave with respect to λ, and it is possible to solve for the optimal λ in closed form. If we fix the value of λ, the objective (2.3.1) is linear on x, then the problem is reduced to a pure integer program. So we iteratively optimize on x and λ respectively. We start with arbitrary initial estimates of $\hat{\lambda}$ and then optimize on the binary variables x by solving the pure integer program. Let us refer to this as M1 step. Then, conditional on the estimation $\hat{x}$ from the M1 step, we maximize with respect to λ. Let us refer to it as M2 step. We repeat the procedure again and again until we reach at a coordinatewise maximum point. This is the Alternating Minimization Algorithm, or Two-Block Gradient Descent Algorithm, which is commonly used to simplify the maximization of log likelihood functions when parameters can be partitioned into two sets, but the log likelihood function becomes globally concave when one of the block is observed. The following lemmas show some structural properties in the M1 and M2 steps.

Lemma 2.3.4 (Conditional optimization of x for the M1 step.)**.** *Given model parameter estimates $\hat{\lambda}$, the maximization of the conditional log likelihood function with respect to x boils down to the classical Rank Aggregation Problem.*

Rank Aggregation aims to output a preference list by aggregating pairwise comparison information and is proved to be NP-hard (Dwork, Kumar, Naor and Sivakumar 2001). We refer readers to a comprehensive review of Ali and Melia (2012).

Lemma 2.3.5 (Conditional optimization of λ for the M2 step.)**.** *Given model parameter estimates $\hat{x}$, the maximization of the conditional log likelihood function with respect to λ can be solved to closed form:*

$$\lambda_i^* = \frac{\sum_{t=1}^{T} \mathbb{1}(c_t = i)}{\sum_{t=1}^{T} \mathbb{1}(c_t = i) + \hat{x}_{ic_t} \mathbb{1}(i \in S_t)}, \tag{2.3.4}$$

where $\mathbb{1}(\cdot)$ is an indicator function.

This directly follows the first order condition for maximizing objective (2.3.1) conditional on x, which is quite intuitive since it is the proportion of time that product i is purchased when it ranks at the top of the offer set.

Based on above results, the estimation procedure can be summarized as follows:

Step 0 [Initialization.] Set $\hat{\lambda}$ to arbitrary feasible initial values.

Step 1 [M1 step.] Compute x^*, the optimal estimates of x conditional on $\lambda's$ value being $\hat{\lambda}$

Step 2 [M2 step.] Compute λ^*, the optimal estimates of λ conditional on $x's$ value being $\hat{x}$ by formula (2.3.4)

Step 3 [Check stopping condition.] If $Dist((\hat{\lambda},\hat{x}),(\lambda^*,x^*)) > \varepsilon$, then $\hat{\lambda} \leftarrow \lambda^*$ and $\hat{x} \leftarrow x^*$ and go to Step 1. Otherwise, terminate and output λ^* and x^*. We can use any metric as the distance measure, for example the Euclidean distance.

An intuitive question is whether the coordinatewise maximum solution is a stationary point. Unfortunately, convergence to a stationary point is not guaranteed due to the integer constraint in the M1 step (Xu and Yin 2017) even if we can solve the NP-hard problem to optimality. Empirical studies by Reinelt (1985) and Reineilt, Grotschel and Junger (1983), show that the LP relaxation of M1 performs exceptionally well and often finds optimal and binary solution in practice. Therefore we relax the binary constraint in the implementation. Interestingly, all the convergent points in our experiments are integer, which are guaranteed to be stationary points.

Now we turn our attention to operational decisions. The first problem we look into is the product assortment when products configurations and prices are fixed.

2.4 Assortment Optimization

Let $r_i > 0$ denote the revenue associated with product i and let

$$R(S) = \sum_{i \in S} r_i \pi_i(S),$$

denote the expected revenue from assortment S. Our goal is to find an optimal assortment S^*, such that

$$R(S^*) = \max_{S \subseteq N} R(S).$$

The following lemma will be helpful in finding an optimal assortment.

Lemma 2.4.1. *Let $S \subseteq N$, and let $k \in N$ be such that $i \prec k$ for all $i \in S$. Then,*

$$R(S \cup \{k\}) = (1 - \lambda_k) R(S) + \lambda_k r_k, \tag{2.4.1}$$

and consequently $R(S \cup \{k\}) \geq R(S)$ if and only if $r_k \geq R(S)$.

For $i \in S$, $i \prec k$, so $\pi_i(S \cup \{k\}) = \pi_i(S)(1 - \lambda_k)$, while $\pi_k(S \cup \{k\}) = \lambda_k$, from which (2.4.1) follows directly. Moreover $R(S \cup \{k\}) = R(S) + \lambda_k(r_k - R(S)) \geq R(S)$ if and only if $r_k \geq R(S)$. So we have the following greedy-like assortment optimization algorithm.

Algorithm 2.1 Assortment Optimization Algorithm

Set $S_0 = \emptyset$, $\tilde{S}_0 = \emptyset$ and $H_0 = 0$.
for $i \in N$ **do**
  $H_i = H_{i-1} + \lambda_i (r_i - H_{i-1})^+$;
  $S_i = S_{i-1} \cup \{i\}$;
  $\tilde{S}_i = \{j \in S_i : r_j \geq H_{j-1}\}$
end for

where $(x)^+ \equiv \max(x, 0)$. The sequence H_i, $i \in N$ considers at stage i, whether or not to add product i to $\tilde{S}_{i-1}$ to form $\tilde{S}_i$ and does this if and only if $r_i \geq H_{i-1}$. For each $i \in N$, the algorithm requires one comparison, and three arithmetic operations, so the algorithm runs in $O(n)$ time.

Theorem 2.4.2. *The sequence H_i is weakly monotonically increasing in $i \in N$, with*

$$H_i = R(\tilde{S}_i) \geq R(S) \quad \forall \quad S \subseteq S_i, \;\; i \in N.$$

Therefore $\tilde{S}_n$ maximizes $R(S)$.

Proof. The order $H_i \geq H_{i-1}$ follows directly from its definition. Clearly $H_1 = \lambda_1 r_1 > 0$, so $\tilde{S}_1 = S_1 = \{1\}$ and $R(\tilde{S}_1) = H_1$. Assume that the result holds for $i-1$, so $H_{i-1} = R(\tilde{S}_{i-1}) \geq R(S)$ for all $S \subset S_{i-1}$. We will show that $H_i = R(\tilde{S}_i) \geq R(S)$ for all $S \subset S_i$.

From the algorithm,

$$\begin{aligned} H_i &= H_{i-1} + \lambda_i (r_i - H_{i-1})^+. \\ &= R(\tilde{S}_{i-1}) + \lambda_i (r_i - R(\tilde{S}_{i-1}))^+. \end{aligned}$$

If $r_i < H_{i-1}$ then $H_i = H_{i-1} = R(\tilde{S}_{i-1}) = R(\tilde{S}_i)$ on account of $\tilde{S}_i = \tilde{S}_{i-1}$. On the other hand, if $r_i \geq H_{i-1}$ then $\tilde{S}_i = \tilde{S}_{i-1} \cup \{i\}$, and

$$H_i = (1-\lambda_i)R(\tilde{S}_{i-1}) + \lambda_i r_i = R(\tilde{S}_i),$$

where the last equality follows from equation (2.4.1). To complete the proof we need to show that $R(\tilde{S}_i) \geq R(S)$ for all $S \subset S_i$. If $i \notin S$, then $R(S) \leq R(\tilde{S}_{i-1}) \leq R(\tilde{S}_i)$ by the inductive hypothesis. On the other hand, if $i \in S$, then we can write $S = T \cup \{i\}$ for some $T \subset S_{i-1}$. In this case

$$\begin{aligned}
R(S) = R(T \cup \{i\}) &= (1-\lambda_i)R(T) + \lambda_i r_i \\
&\leq (1-\lambda_i)R(\tilde{S}_{i-1}) + \lambda_i r_i = (1-\lambda_i)H_{i-1} + \lambda_i r_i \\
&\leq H_{i-1} + \lambda_i (r_i - H_{i-1})^+ = H_i \\
&= R(\tilde{S}_i).
\end{aligned}$$

□

2.4.1 Cardinality Constraints

Finding an optimal assortment subject to a cardinality constraint is also of key interest for industry and academia. We show that this can be done by introducing a state variable that represents number of products in the assortment. Let

$$H_i(k) = \max_{S:\, S \subseteq S_i, |S| \leq k} R(S).$$

Using the dynamic program,

$$H_i(k) = \max\{(1-\lambda_i)H_{i-1}(k-1) + \lambda_i r_i, H_{i-1}(k)\}.$$

By sequentially computing $H_i(k), i \in N, k = 1,\ldots n$, we can find optimal assortments of sizes $k = 1,\ldots,n$. For each k, we can compute $H_i(k), i \in N$ in $O(n)$ times, so finding $H_i(k), i \in N$ and $\forall k \leq n$ takes $O(n^2)$ time.

2.4.2 Efficient Sets

In single-leg revenue management problem, we are often interested in solving the problem

$$\mathcal{R}(z) \equiv \max_{S \subseteq N} R(S,z) = \max_{S \subseteq N} \sum_{i \in S} (r_i - z)\pi_i(S), \tag{2.4.2}$$

where z is the marginal value of capacity and will change over time, so at anytime, the optimal assortment should maximize $R(S,z)$ for some z. Let $\Pi(S) = \sum_{j \in S} \pi_j(S)$ be the sales probability for $S \subseteq N$. The problem of finding solutions to (2.4.2) is related to the following linear program.

$$\begin{aligned} Q(\rho) = \max \quad & \sum \alpha(S) R(S) \\ \text{s.t.} \quad & \sum \alpha(S) \Pi(S) \leq \rho \\ & \sum \alpha(S) = 1 \\ & \alpha(S) \geq 0 \quad \forall S \subseteq N. \end{aligned} \tag{2.4.3}$$

The linear program (2.4.3) finds the fraction of time, $\alpha(S)$, that assortment S should be offered to maximize expected revenues without exceeding a bound ρ on the average sales rate. The dual of this problem can be written as $Q(\rho) = \min_{z\geq 0}[\mathcal{R}(z) + \rho z]$, so for each ρ, there is a corresponding dual variable $z(\rho) \geq 0$ and an assortment that maximizes $R(S, z(\rho))$. Conversely, given z there is an optimal assortment S that maximizes $R(S,z)$ with $\Pi(S) = \rho$, such that $Q(\rho) = R(S)$.

The notion of efficient sets was first introduced by Talluri and van Ryzin (2004). Gallego and Topaloglu (2016) define as efficient any assortment E such that $Q(\Pi(E)) = R(E)$. Suppose that $\emptyset = E_0, E_1, \ldots, E_m$ is the collection of efficient sets. Without loss of generality we can label the sets so that $0 = \Pi(E_0) < \Pi(E_1) < \ldots < \Pi(E_m) \leq 1$. It can be shown, see Talluri and van Ryzin (2004) and Gallego and Topaloglu (2016), that the problem of solving $\mathcal{R}(z)$ can be reduced to a search over efficient sets only.

Theorem 2.4.3 (Talluri and van Ryzin 2004)**.** *For any $z \geq 0$,*

$$\mathcal{R}(z) = \max_i R(E_i, z).$$

Moreover, E is efficient if and only if there exists a $z \geq 0$ such that $R(E,z) = \mathcal{R}(z)$.

As a result, there will be cost thresholds $0 < z_m < z_{m-1} < \ldots < z_1$ such that the efficient set E_m is optimal for all z such that $0 \leq z \leq z_m$; E_j is optimal for all z such that $z_{j+1} < z \leq z_j$ for $j = 1, \ldots, m-1$; and the empty set E_0 is optimal for all z such that $z_1 < z$.

For $z = 0$, the dynamic program algorithm 2.1 finds $E_m = S^* = \tilde{S}_n$ that maximizes $R(S, 0)$. A brute force approach would be to use the same algorithm for each $z \geq 0$. More precisely, let $H_0(z) = 0$, and for $i \in N$, let $H_i(z) = H_{i-1}(z) + \lambda_i(r_i - z - H_{i-1}(z))^+$, and $\tilde{S}_i(z) = \{j \in S_i : r_j - z \geq H_{j-1}(z)\}$. From our previous results, we know that $H_i(z) = R(\tilde{S}_i(z), z) \geq R(S, z)$ for all $S \subseteq S_i$, and in particular $S^*(z) = \tilde{S}_n(z)$ maximizes $R(S, z)$ over all $S \subseteq N$. By running the algorithm for all $z \geq 0$, we can find the cost thresholds $0 < z_m < z_{m-1} < \ldots < z_1$ and the efficient sets $E_0 = \emptyset$, and $E_j = S(z_j), j = 1, \ldots, m$. In this section we provide a more expedient way of finding efficient sets. For general choice models, the efficient sets need not be nested. We next show that the efficient sets are nested for the RCS model.

Lemma 2.4.4. *The efficient sets are nested, i.e.,*

$$E_0 \subsetneq E_1 \subsetneq \ldots \subsetneq E_m.$$

To identify the efficient sets, let $J_0 = 0$ and $J_i = \Pi(\tilde{S}_i), i = 1, \ldots, n$ be the sales rate associated with the set $\tilde{S}_i$. Consider the graph $(J_i, H_i), i = 0, 1, \ldots, n$, and let $u_i = (H_i - H_{i-1})/(J_i - J_{i-1})$ for all $i \in N$. Notice that for $i \notin S^*$, the ratio is of the form $u_i = 0/0$, which for convenience we set as $0/0 = \infty$. Notice that for $i \in S^*$, the slopes are positive and finite. Let $m = |E_m|$ and let i be such that $u_i < u_j$ for all $j \leq n$. We will show that $z_m = u_i$ and that $E_{m-1} = S^*/\{i\} = E_m/\{i\}$. To see this, consider the smallest value of z such that it is optimal to drop a product from S^* at z^+. Thus $S^*(z) = S^* = E_m$ but a product, say $i \in S^*$ is not in $S^*(z+)$. Then,

$$r_i - z = \sum_{j \in \tilde{S}_{i-1}} (r_j - z)\pi_j(\tilde{S}_{i-1}) = H_{i-1} - zJ_{i-1},$$

i.e.,

$$z = \frac{r_i - H_{i-1}}{1 - J_{i-1}} = u_i,$$

where the last equality follows from simple algebra. Moreover, i must be a product in S^* with the smallest u_i ratio as otherwise it would be optimal to drop another product first. This confirms that $z_m = u_i$ is the first breakpoint and that product i is the one to exit E_m to form set E_{m-1}. In case of ties, it may be necessary to remove more than one product at a time. Once product i is removed, and $z_m = u_i$ is recorded, we form the set $E_{m-1} = E_m/\{i\}$, and update the slopes and repeat the process until the last product is eliminated. More precisely, given z_{k+1} and E_k, we need to compute the points $(J_j(z_{k+1}), H_j(z_{k+1}))$ and the corresponding slopes $u_j(z_{k+1})$, and then find the smallest slope $z_k = \min_{i \in E_k} u_i(z_{k+1})$ and form the set $E_{k-1} = E_k/\{i(k)\}$, where $i(k)$ is the index of the product with the smallest slope. Because, each time we need to compute at most $m \leq n$ slopes, and remove at least one product, the algorithm to identify all of the efficient sets is $O(m^2)$. Since the efficient sets are nested, there can be at most $m \leq n$ efficient sets implying that it takes at most $O(n)$ effort to find them all.

Table 2.5: Example of the Efficient Sets Discovery Process

Product	λ	**Price**	E_5	u	E_4	u	E_3	u	E_2	u	E_1
1	**0.017**	**50**	**1**	**50.00**	**1**	**(50.00)**	**0**	∞	**0**	∞	**0**
2	**0.055**	**60**	**1**	**60.17**	**1**	**60.17**	**1**	**(60.00)**	**0**	∞	**0**
3	**0.044**	**68**	**1**	**68.78**	**1**	**68.78**	**1**	**68.47**	**1**	**(68.00)**	**0**
4	**0.100**	**75**	**1**	**76.67**	**1**	**76.67**	**1**	**76.21**	**1**	**75.32**	**1**
5	**0.089**	**52**	**1**	**(47.89)**	**0**	∞	**0**	∞	**0**	∞	**0**

From the above example we see that the order of elimination does not need to be aligned with prices order.

2.4.3 Ordered Preference with Ties

Manzini and Mariotti (2014) assume the preference is strictly ordered, while in reality partial ranking happens often as consumers might be indifferent among choices. For example, hotels are rated from one star to five stars. Consumers might choose each of the identically rated ones with equal probability. Tied preference has been widely observed and over the past decades, for example, statisticians has spent efforts in identifying tied preferences via experiment (Glenn and David 1960, Ennis and Ennis 2012). So in this section we relax the strict preference ordering assumption and look into the assortment problem when consumers can have equal preferences with some products, namely, when the preference is partially ordered. In this situation, some of the products are non-comparable and have tied preference. For example, consumers evaluate the set $N = \{1,2,...,n\}$ as $1 \prec 2 \sim 3 \sim 4 \prec 5... \prec n$, where consumers are indifferent between products 2, 3, and 4. The randomness still results from inattention, where each product i is considered with probability λ_i.

Before presenting our assortment algorithm for the entire problem, let us first take a look at a subproblem – assortment optimization when consumers have equal preference for all products but the consideration sets are randomly generated. Equal preference means that conditional on the consideration set, the probability of purchase is uniformly distributed among the considered products. Again, let λ_i be the attention probability of product i and from now on let us assume all products in assortment S have equal preference. The following example shows that assortment problem is intractable and does not have the nested by revenue structure.

Example 2.4.5. *There are eight products among which consumers are indifferent on preferences and they are considered with attention probabilities*

$$\lambda = (0.477, 0.831, 0.467, 0.046, 0.015, 0.492, 0.150, 0.267),$$

and revenues

$$r = (55, 41, 42, 44, 67, 86, 8, 11).$$

The optimal expected revenue is 50.03 given by assortment $\{1,2,5,6\}$.

Notice that $r_4 > r_3 > r_2$ but product 3 and 4 are not in the optimal assortment. We will show that the best revenue ordered assortment can guarantee at least 50% of the optimal revenue.

From Example 2.4.5 we can see that unlike the assortment problem for the MNL, the optimal assortment under this random consideration set model does not necessarily have the nested by revenue structure; moreover, some products might be introduced even if their prices are lower than the expected total revenue.

Now let us introduce notations. For every $S \subseteq N$, the probability that consumers form consideration set $T \subseteq S$, is given by

$$\Pi(T,S) = \Pi_{i\in T}\lambda_i \times \Pi_{j\in S\setminus T}(1-\lambda_j).$$

For any assortment T, we will denote by $r(T) = \sum_{i \in T} r_i / |T|$ the average revenue obtained from consideration set T. This is because consumers purchase products from the consideration set with uniform probabilities. In order to distinguish from the revenue of strictly ordered preference, let us denote the revenue from assortment S with tied preference by $J(S)$. With this notation, we can write the expected revenue from offering assortment S as

$$J(S) = \sum_{T \subseteq S} r(T) \times \Pi(T, S). \tag{2.4.4}$$

In the next lemma, we show there indeed is a minimum requirement of products' revenue in order to be in the optimal assortment.

Lemma 2.4.6. *Let S^* be the optimal assortment, and $\underline{r}(S^*) = \min\{r_i, i \in S^*\}$, then $\underline{r}(S^*) \geq \frac{1}{2}J(S^*)$.*

Lemma 2.4.6 shows that the lowest price in the optimal assortment should be at least half of the optimal revenue. We will give a detailed proof in the appendix. Now we are ready to present our main result:

Theorem 2.4.7. *Let $\bar{S} \equiv \{i \in N : r_i \geq \frac{1}{2}J(S^*)\}$, then the expected revenue $J(\bar{S}) \geq \frac{1}{2}J(S^*)$.*

This theorem means that offering all products with prices above a threshold will have 0.5-optimality guarantee. But as the threshold $\frac{1}{2}J(S^*)$ comes at the cost of knowing optimal revenue, we have the following more practical corollary.

Corollary 2.4.8. *Consider the assortment optimization among products with tied preference, the optimal nested by revenue assortment can guarantee at least $\frac{1}{2}$ of the optimal total revenue.*

We remark that finding an optimal nested by revenue assortment still requires us to compute $J(S)$, which asks for enumerating exponentially many consideration sets. We now look at an easy to implement heuristic which has the same performance guarantee but does not require computing $J(S)$. To this end, consider an MNL model with attraction probabilities $v_i = \frac{\lambda_i}{1-\lambda_i}$ for all $i \in N$ and $v_0 = 2$, and revenues $r_i, i \in N$. Let $\tilde{S}$ be an optimal assortment for this MNL model. Let $J(\tilde{S})$ be the expected revenue under this assortment under the true model, and recall that $J(S^*)$ denotes the expected revenue under an optimal assortment.

Theorem 2.4.9.

$$J(\tilde{S}) \geq \frac{1}{2} J(S^*),$$

where

$$\tilde{S} \in \arg\max_{S \subseteq N} \frac{\sum_{i \in N} r_i v_i}{v_0 + \sum_{i \in N} v_i}.$$

Finding $\tilde{S}$ is as easy as finding an optimal assortment for the standard MNL model, which can be solved in polynomial time since the optimal assortment should be nested by prices, and the optimal assortment $\tilde{S}$ should be $\{j \in N : r_j \geq J_{\text{MNL}}(\tilde{S})\}$, where $J_{\text{MNL}}(\tilde{S})$ is the optimal expected revenue of the constructed MNL model. The intuition behind is that for a product to enter the optimal assortment set S^*, $J_{\text{MNL}}(\tilde{S})$ is the price lower bound.

However, we can show that $J_{\text{MNL}}(\tilde{S})$ is greater than $\frac{1}{2}J(S^*)$. If we look closely at the MNL expected revenue, $J_{\text{MNL}}(S^*)$ will drop to $\frac{1}{2}J(S^*)$ only when attention probabilities are converging to zero.

Given that we have a $1/2$ approximation guarantee for assortment optimization among products with tied preference, we are ready to check the assortment optimization problem for the entire product set. Let us partition the global product set $N = \{1,2,...,n\}$ in a way such that $\mathcal{Q} = \{Q_1,Q_2,...,Q_T\}$ is the collection of partitions, where $\forall i,j \in Q_t$, we have $i \sim j$ and $\forall i \in Q_t,\ j \in Q_{t+1}$, we have $i \prec j$. So all products in the same partition have equal preferences and products from different partitions are ordered. $|\mathcal{Q}| = T$ indicates there are T preferences levels.

Algorithm 2.2 Assortment Optimization Algorithm When Preferences Have Ties

Set $\mathcal{Q}_0 = \emptyset$ and $H_0(\mathcal{Q}_0) = 0$;
for $t = 1,\ldots,T$ **do**
  $\mathcal{Q}_t = \mathcal{Q}_{t-1} \cup Q_t$;
  $H_t(\mathcal{Q}_t) = H_{t-1}(\mathcal{Q}_{t-1}) + \max_{S\subseteq Q_t} J(r - H_{t-1}(\mathcal{Q}_{t-1}), S)$;
  $\tilde{\mathcal{Q}}_t = \tilde{\mathcal{Q}}_{t-1} \cup \arg\max_{S\subseteq Q_t}\ J(r - H_{t-1}(\mathcal{Q}_{t-1}), S)$
end for

The sequence $H_t(\mathcal{Q}_t), t = 1,...,T$ sequentially considers whether or not to add subset S of products in Q_t to $\tilde{\mathcal{Q}}_{t-1}$ to form $\tilde{\mathcal{Q}}_t$ and does this if and only if $S = \arg\max_{S\subseteq Q_t}\ J(r - H_{t-1}(\mathcal{Q}_{t-1}), S)$.

Theorem 2.4.10. *The sequence $H_t(\mathcal{Q}_t)$ is weakly monotonically in $t = 1,...,T$, with*

$$H_t(\tilde{\mathcal{Q}}_t) = R(r,\tilde{\mathcal{Q}}_t) \geq R(r,S)\ \ \forall\ \ S \subseteq \tilde{\mathcal{Q}}_t,\ \ t = 1,...,T.$$

so $\tilde{\mathcal{Q}}_T$ is an optimal assortment.

The logic is similar to the logic behind theorem (2.4.2), as we can express $R(T \cup S)$, $T \subseteq Q_{t-1}$ and $S \subseteq Q_t$ by

$$R(T \cup S) = R(T) + J(r - R(T), S).$$

So we will omit the proof.

Theorem 2.4.11. *If the subproblem* $\max_{S \subseteq Q_t} J(r - H_{t-1}(Q_{t-1}), S)$ *can be solved to* $\alpha-$*optimal for all* $t = 1, \ldots, T$*, then the entire problem can be at least* $\alpha-$*optimal.*

This theorem states that if we can solve each subproblem to $\alpha-$optimal then we can solve the entire problem to $\alpha-$optimal. This can be easily shown by induction on $t = 1 \ldots T$.

Corollary 2.4.12. *Solving the subproblem* $\max_{S \subseteq Q_t} J(r - H_{t-1}(Q_{t-1}), S)$ *heuristically by choosing optimal assortments under the constructed MNL model for all* $t = 1, \ldots, T$ *can guarantee at least* $1/2$ *of the optimal revenue.*

2.5 Pricing under the RCS model

Retailers may have the freedom to select assortments and to price product at the same time. In this section we parameterize the attention probability using the Logit model, see Assumption 2.5.1 below, to reflect the fact that people are more likely to include high utility products in their consideration sets, but less likely to include expensive products. In particular, we are interested in uncovering the form of the optimal prices relative to the products' quality measured by α_i and its unit wholesale cost measured by z_i. We will see that at optimality prices are chosen so that the ordering of the value gap $\alpha_i - z_i$ is preserved by $\alpha_i - r_i$ and by the contribution margins $w_i = r_i - z_i$.

Assumption 2.5.1. *Given the quality α_i and price r_i of each product $i \in N$, the inclusion probability of product i into the consideration set follows the Logit form:*

$$\lambda_i = \frac{exp(\alpha_i - r_i)}{1 + exp(\alpha_i - r_i)}. \tag{2.5.1}$$

Moreover, consumers have common valuations on products' quality and price, so they will choose the product with the highest utility $\alpha_i - r_i$ from the consideration set; therefore, consumers rank the products in the order of $\alpha_i - r_i$.

We consider the case when the seller wants to maximize his profit by taking into account

z_i, the unit cost of product i, then the pricing problem can be formulated as

$$\max_{r,x} \quad \sum_{i=1}^{n}\sum_{j=i+1}^{n}(1-\lambda_{(j)})\lambda_{(i)}(r_{(i)}-z_{(i)}) \tag{2.5.2}$$

$$\text{s.t.} \quad \lambda_{(i)} = \sum_{j=1}^{n}\lambda_j x_{j(i)} \quad \forall i \in N \tag{2.5.3}$$

$$r_{(i)} = \sum_{j=1}^{n} r_j x_{j(i)} \quad \forall i \in N \tag{2.5.4}$$

$$z_{(i)} = \sum_{j=1}^{n} z_j r_{j(i)} \quad \forall i \in N \tag{2.5.5}$$

$$\sum_{j=1}^{n}(\alpha_j - r_j)x_{j(i)} \leq \sum_{j=1}^{n}(\alpha_j - r_j)x_{j(i+1)} \quad \forall i = 1...n-1 \tag{2.5.6}$$

$$\sum_{j=1}^{n} x_{j(i)} \leq 1 \quad \forall i \in N \tag{2.5.7}$$

$$\sum_{i=1}^{n} x_{j(i)} \leq 1 \quad \forall j \in N \tag{2.5.8}$$

$$x_{j(i)} \in \{0,1\} \quad \forall i \& j \in N. \tag{2.5.9}$$

Driving r_i to infinity is equivalent to eliminating it from the market, so our formulation solves both the assortment and pricing problem. As the prices will affect the value of $\alpha_i - r_i$ and thus affect the products' ranking, we introduce the binary variable $x_{j(i)}$ indicating whether product j is at the i_{th} position of consumer's preference list. Therefore, it must satisfy the utility ordering, i.e., utility of the i_{th} product from the preference list should not be greater than the utility of the $i+1_{st}$ product: $\sum_{j=1}^{n}(\alpha_j - r_j)x_{j(i)} \leq \sum_{j=1}^{n}(\alpha_j - r_j)x_{j(i+1)} \quad \forall i = 1...n-1$.

Theorem 2.5.2. *Without loss of generality, suppose the products are ordered such that*

$0 < \alpha_1 - z_1 < \alpha_2 - z_2 < \ldots < \alpha_n - z_n$, *and let* $w_i = r_i - z_i$, $i \in N$ *be the contribution margins. Then at optimality, products should be priced such that the net utility* $\alpha_i - r_i$ *has the same ordering as the value gap* $\alpha_i - z_i$, *i.e.,* $0 < \alpha_1 - r_1 < \alpha_2 - r_2 < \ldots < \alpha_n - r_n$, *and* $1 \prec 2 \prec \ldots \prec n$. *Moreover, the optimal margins increases monotonically:* $w_1 < w_2 < \ldots < w_n$, *so the markups are increasing in i.*

This result tells us that if product i has a larger value gap than product j, i.e., $\alpha_i - z_i > \alpha_j - z_j$, then they should be priced so that both $\alpha_i - r_i > \alpha_j - r_j$ and $r_i - z_i > r_j - z_j$. This is equivalent to saying that a product with a large value gap will be priced so that it is both ranked higher by the consumer and will have a higher profit to the retailer conditioned on sales. Moreover, $\alpha_i - r_i > \alpha_j - r_j$ implies that $\lambda_i > \lambda_j$, so the product will garner more attention from consumers. Suppose now that the products are priced as suggested by Theorem 2.5. Would all the products be part of an optimal assortment? The answer is yes, as $H_1 = \lambda_1(r_1 - z_1) > 0$ and $r_i - z_i > H_{i-1}$ for all i as H_{i-1} is a sub-convex combination of $r_1 - z_1 < \ldots < r_{i-1} - z_{i-1}$ and consequently $H_{i-1} < r_{i-1} - z_{i-1} < r_i - z_i$. As a result, we have the following corollary.

Corollary 2.5.3. *If the products are priced so that* $\alpha_i - z_i > \alpha_j - z_j$ *implies* $r_i - z_i > r_j - z_j$ *and* $\alpha_i - r_i > \alpha_j - r_j$, *then* λ_j *is increasing in j and the optimal assortment is* $N = \{1, \ldots, n\}$.

Table 2.6 gives optimal prices under the given α and z parameters. We see that both the order of the utility $\alpha - r^*$ and the profit margin $r^* - z$ are aligned with the order of value gap $\alpha - z$. Notice that the price is not necessarily ordered, for example $r_3^* > r_4^*$ as product 3 has higher cost. But if we compare prices of product 3 and 5, we have $r_3^* < r_5^*$

Table 2.6: Example of RCS Pricing

Product	α	z	$\alpha - z$	r^*	$\alpha - r^*$	$r^* - z$	λ	$\lambda \times (r^* - z)$
1	**8**	**5**	**3**	**7.56**	**0.44**	**2.56**	**0.608**	**1.56**
2	**9**	**4**	**5**	**8.39**	**0.61**	**4.39**	**0.648**	**2.84**
3	**13**	**6**	**7**	**12.34**	**0.66**	**6.34**	**0.659**	**4.18**
4	**10**	**1**	**9**	**9.32**	**0.68**	**8.32**	**0.664**	**5.52**
5	**15**	**4**	**11**	**14.31**	**0.69**	**10.31**	**0.666**	**6.87**

even though product 3's cost is higher simply because product 5 has higher perceived value to consumers. Notice that the pricing behavior is much richer than that of the multinomial logit model which results in a constant markup.

One thing worth mentioning is that we can extend the utility model $\alpha_i - r_i$ to a more general form by introducing price sensitivity parameter β, therefore, the utility of product i should be $\alpha_i - \beta r_i$ and attention probability $\lambda_i = \frac{exp(\alpha_i - \beta r_i)}{1+exp(\alpha_i - \beta r_i)}$. We argue that the structural result still holds under the more general case. Indeed, the adjusted value gap: $\alpha - \beta z$ gives the ordering of the net quality $\alpha - \beta r$ and the margins $v = r - z$.

To summarize, optimal net utilities and markups will preserve the value gap ordering. Comparing with MNL under which all products share a common markup at optimality (Anderson, de Palma and Thisse 1992), our parametrized RCS model shows that products with higher utilities will have higher markups than products with lower utilities under optimal pricing.

2.6 Computational Study

The practical applicability of a choice model depends on its predictive power. In this section we numerically test our Random Consideration Set (RCS) model on both real world transactional data and synthetic data. We compare the predictive power of the RCS model with the the MMNL(3) model, which represents a mixture of MNLs model with three market segments. We find that the RCS model obtains an average of 15.6% improvement over the benchmark on a chi-square metric for data obtained by our airline partner. Compared with MNL model, our RCS always outperforms it over all markets.

2.6.1 Predictive Accuracy: Real World Data

Data Description and Assumption

We are given tickets sales data of 107 different itineraries departing over a collection of 120 days from one of the world's largest airline companies. We eliminate 16 of them due to data collection error. We fit the data and test the model predictive efficiency for each flight separately. The sales data are collected for each flight leg departing on each single date in a set of reading days before its departure. On each reading day, we are presented with the cumulative sales and nested protection levels of all fare classes (as the airline company uses the nested capacity allocation policy). So we can infer the sales of each fare class within two reading days by taking the difference of two cumulative sales and infer the offer set by checking the inventory availabilities on the two consecutive reading days. Furthermore, because a consumer's arrival with no-purchase are not observable, we focus on capturing the substitution behavior among different fare classes within a single flight by assuming this is the entire market. We deal with sparsity of the data by looking into a single flight's sales among all departure dates and assuming the same choice behavior for consumers taking the same flight even on different days. For example, the attention probability and the preference ordering remain the same for consumers who take flight 32139 regardless of whether it departs on May-08-2015 or on June-10-2016.

Model Fit and Results

For a single itinerary, we treat flights departing on the first 90 days as the training set and the rest as the testing set. We compare the predictive accuracy of our RCS model against the popular Mixture of Multinomial Logit (MMNL) benchmark on the same data set. Under the MMNL model, an incoming consumer belongs to a market segment m with probability w_m and chooses a product according to MNL. So essentially we are maximizing the following Log-likelihood function:

$$\max_{w,v} \quad \sum_{t=1}^{T} \log\Big(\sum_{m} w_m \frac{exp(v_{c_t}^m)}{\sum_{i\in S_t} exp(v_i^m)}\Big),$$

where the parameter v_i^m is the expected utility of product i evaluated by market segment m; c_t is the product chosen at time t. The problem is difficult to solve in general. We pick the number of mixtures at three and use the EM algorithm described in Train (2008) to find a stationary point.

$$\chi_E^2 = \frac{1}{\sum_{t=1}^{T} |S_t|} \sum_{t=1}^{T} \sum_{i\in S_t} \frac{(\hat{n}^{model}(i,S_t) - n^{actual}(i,S_t))^2}{0.5 + n^{actual}(i,S_t)}.$$

where $\hat{n}^{model}(i,S_t)$ is the predicted sales of product i at time t and $n^{actual}(i,S_t)$ is the actual sales of product i at t. And we compute the predicted number of sales by multiplying the predicted choice probability with the total sales at time t. This metric is analogous to the popular chi-square measure $(O-E)^2/E$ which measures the goodness of fit and it has been used in Jagabathula and Rusmevichientong (2016).

The percentage improvements in the χ_E^2 metrics under our RCS model relative to the MMNL model can be defined as

$$\frac{\chi_E^{2^{MMNL}} - \chi_E^{2^{RCS}}}{\chi_E^{2^{MMNL}}}.$$

We compare these two models over the 91 markets. The result is summarized in table 2.7 where we also provide comparisons to the MNL model.

Table 2.7: Improvements over Benchmark Models

Improvement over MNL			**Improvement over MMNL(3)**		
Max	**Average**	**Pr(RCS≻MNL)**	**Max**	**Average**	**Pr(RCS≻MMNL(3))**
68.2%	**22.3%**	**100%**	**44.4%**	**15.6%**	**67.0%**

The comparison result shows that RCS is a more relevant model to capture demand substitution. As stated in the Introduction, consumers tend to buy the cheapest ticket available. And our estimation result shows that the preference ordering and the ticket fare are heavily correlated. It is worth highlighting that the RCS model outperforms MNL over all markets, i.e., Pr(RCS≻MNL)=100% and even compared with the Mixture of MNLs model, RCS can give better fit on 67.0% of the markets.

2.6.2 Predictive Accuracy: Synthetic Data

In this section we test the predictive power of our RCS model against the benchmark MMNL model on synthetic data. We first explain how the simulations are set up and then present the results. We assume there are $n = 10$ products and $T = 600$ time periods. At each time point, a random subset is drawn from the global product set as the offered assortment. Then the random purchase is simulated according to a ground truth model. We first assume the ground truth model is MMNL(3), using it to simulate random purchases and then fit the choice data by an MMNL with the same number of segments and by our RCS model. Later we assume the ground truth model is an RCS choice model, and then fit the data by MMNL(3) model as well as RCS model and compare their predictive accuracy.

1. MMNL(3) model parameter generation: Assuming MMNL(3) being the ground truth, we randomly generate the parameter v_i^m, valuation of product i of segment m by uniform distribution from the interval $[-5,5]$. And the weight parameter $w_m = \frac{exp(u_m)}{\sum_j exp(u_j)}$, where u_m follows uniform distribution from the interval $[-2,2]$ for all market segments m.

2. Random Consideration Set model parameter generation: Assuming RCS being the ground truth, we randomly generate the attentions probabilities λ_i for all products $i = 1,..,n$, where each λ_i is an independent draw from a uniform$[0,1]$ distribution. And we generate product rankings by randomly permuting all the products.

We compare these two models in terms of the MAPE and Chi-square metrics defined as:

$$MAPE^{model} = \frac{1}{\sum_{t=1}^{T} |S_t|} \sum_{t=1}^{T} \sum_{i \in S_t} \frac{|\hat{\pi}^{model}(i,S_t) - \pi^{actual}(i,S_t)|}{\pi^{actual}(i,S_t)},$$

$$\chi^{2^{model}} = \frac{1}{\sum_{t=1}^{T} |S_t|} \sum_{t=1}^{T} \sum_{i \in S_t} \frac{(\hat{\pi}^{model}(i,S_t) - \pi^{actual}(i,S_t))^2}{\pi^{actual}(i,S_t)},$$

where $\hat{\pi}^{model}(i,S_t)$ is the predicted sales probability of product i at time t according to the assumed model, and $\pi^{actual}(i,S_t)$ is the actual sales probability of product i at t according to the ground truth.

Table 2.8 summarizes the comparison result. Basically neither MMNL nor RCS can dominate the other.

Table 2.8: Predictive Accuracy on Synthetic Data

Ground Truth	**MMNL(3)**	**RCS**
Improvement over MMNL(3) Measured by Chi-square	**-79.9%**	**21.0%**
Improvement over MMNL(3) Measured by MAPE	**-45.2%**	**53.3%**
Prob. RCS Outperforms MMNL(3)	**11.5%**	**88.6%**

2.6.3 Approximation Algorithm Performance: Tied Preference

Recall that in Section 2.4.3 we show that our approximation algorithm has performance guarantee $1/2$. In this section we numerically test its performance. We generate the model parameters attention probability λ, and preference ordering as in Section 2.6.2; while prices are simulated from uniform$[0,1000]$ distribution. Since the choice probability is already intractable under this model, we test the algorithm on a relative small data set, i.e., the number of products $n \in \{5, 10, 15\}$ and find the optimal assortment by exhaustive search.

Table 2.9: Heuristic Performance under Tied Preference

n	**Average Gap**	**Max Gap**	**Pr(Gap$<1\%$)**	**Pr(Gap$<5\%$)**
5	**0.27%**	**8.14%**	**89.65%**	**99.50%**
10	**0.80%**	**10.52%**	**73.95%**	**97.45%**
15	**1.37%**	**11.25%**	**62.10%**	**94.90%**

2.7 Conclusion

This chapter contributes to the consideration-set-based choice models literature by bringing the random consideration set model of Manzini and Mariotti (2014) to the operations management field. As empirical works (Jagabathula and Rusmevichientong 2016) support the unique preference ordering among all products, we assume the differences in the purchase is due to consideration sets variability as consumers pay random attention to the products. We study the parameter identification problem of the RCS model under both noiseless and noisy settings. Under the noiseless setting, we give both necessary condition and sufficient condition to uniquely identify parameters. Under the noisy setting we design an algorithm that can efficiently discover locally optimal parameters. When testing the RCS model on our airline partner's data, we find that RCS model outperforms MNL over all markets, i.e., Pr(RCS$\succ$MNL)=100% and even compared with the Mixture of MNLs model, RCS can give better fit on 67.0% of the markets. We also optimize the RCS model to determine the revenue-maximizing assortment. We show that an optimal assortment can be found efficiently in $O(n)$ time by dynamic programming. Adding a cardinality constraint increases the complexity to $O(n^2)$. We also extend the RCS model to allow ties in products' ranks, which guarantees $1/2$ revenue from the optimal approach. Finally, we look into the assortment pricing problem and prove that under mild assumptions optimal prices preserves the value gap ordering. In the future, it is worthwhile to extend the model to the case of multiple market segments where each market segment can not only have a different attention probability but also a different products ranking. But we suspect that this problem will be intractable, as Aouad, Farias, Levi and Segev (2015) show that even under a general unique-ranking model the assortment optimization problem is NP-hard to approximate within any constant factor.

Chapter 3

Approximation Algorithms for Product Framing and Pricing

3.1 Introduction

In this chapter, we propose one of the first models of *product framing* and pricing. Framing refers to the way in which the choice among available alternatives is influenced by how the alternatives are framed, or displayed (Tversky and Kahneman 1986). For example, empirical works by Agarwal, Hosanagar and Smith (2009) and Ghose and Yang (2009) in online advertising show that ads that are placed higher on a webpage attract more clicks from consumers. Johnson, Moe, Fader, Bellman and Lohse (2004) examine the average number of websites, sorted by product categories, that are actively visited by households each month. They observe that in a typical search session, consumers search from fewer than two stores. Their data show that 70% of CD shoppers, 70% of book shoppers, and 42%

of travel shoppers, are loyal to just one site. Brynjolfsson, Dick, and Smith (2010) find on a website that catalogs price and product information from multiple retailers, that only 9% of users select offers that are listed beyond the first page. In related search contexts, Baye, Gatti, Kattuman and Morgan (2009) have found that a consumer's likelihood of visiting a firm and purchasing from it is strongly related to the order in which the firm is listed on a webpage by a search engine. They find that a firm receives about 17% fewer clicks for every competitor listed above it on the screen, all other things being equal. chapter

This well-documented framing effect is a natural outcome of the cognitive burden of processing larger and larger assortments. During online shopping, it is cognitively harder for a typical consumer to visit sellers who are listed at the bottom of a web page, before or in addition to visiting those who are listed at the top (Animesh, Viswanathan and Agarwal 2011). In the context of online retailing, it has been observed that consumers' attention to a display decreases exponentially with the display's distance to the top (Feng, Bhargava, and Pennock 2007). Thus, positioning a brand or product at a top position on a listing can improve both consumer attention to the brand, and consequently, consumer selection of the brand (Chandon, Hutchinson, Bradlow, and Young 2009).

3.1.1 Model overview

Despite substantial evidence suggesting the impact of framing on consumers' choice outcome, there are very few models that have attempted to capture these effects. In this chapter, we introduce one of the first models for product framing and the first one for pricing that accounts explicitly for these effects.

We base our model on the notion of *consideration set*. A consideration set is a set of products over which a consumer will make utility comparisons before arriving at the final purchase decision. Consideration sets have gained considerable acceptance since their introduction in the seminal work of Howard and Sheth (1969). A widely used approach to modeling choice in psychology and marketing is to assume that a consumer will first form a consideration set. Then he will choose from among the alternatives in the set. Consideration sets explain, behaviorally, consumers' limited ability to process or acquire information (Manrai and Andrews 1998). Methodologically, it has been shown that ignoring consideration sets may lead to biased parameter estimates (Chiang, Chib and Narasimhan 1999), whereas including consideration sets improves the predictability of choice models (Hauser and Gaskin 1984, Silk and Urban 1978). As an example, Hauser (1978) finds that a disproportionate 78% of the explainable uncertainty in consumer choice can be accounted for by consideration sets, whereas the Multinomial Logit Model (MNL) can only capture the remaining 22%.

We model the effect of framing on the formation of consideration sets as follows. Products are organized into virtual pages. Each page can hold a finite number, say p, of products. A consumer will examine only the first X pages, where X is a random variable that may be *personalized* to the consumer's profile. The consumer forms a consideration set consisting of only products in the examined pages. From this consideration set, the consumer makes a choice according to a *general choice model*. Thus, products that are placed in earlier pages are more likely to be considered, and therefore purchased, than those that are placed in later pages.

Given the behavior described above, we study two problems that are faced by an e-retailer who is managing n different products in a particular product category. The retailer's *product framing problem* is how to determine an assortment and a distribution of the products in the assortment into the different pages in order to maximize the expected revenue. The retailer's *price framing problem* is to determine both the framing and pricing of the products in order to maximize the expected revenues.

For the product framing problem, we capture additional effects that come into play after the consideration sets have been formed. We call these effects *location preference*. Location preference works as follows. First, given that a collection of pages enters into a consumer's consideration set, products that are displayed higher on a page are more likely to be chosen than those displayed lower on the same page due to the evaluation effect, all other factors being equal. Second, products that are listed in earlier pages are more likely to be chosen than products that are listed in later pages due to the attraction effect, all other factors being equal. With location preference, we require that the choice model be the MNL model. We capture the effect of location on choice by using location-dependent preference weights, which we will describe in greater detail in Section 3.8.

3.1.2 Results and Implications for Retailers

Our contributions in this chapter are the following:

- We propose one of the first models of framing effects. Our model is more general than those previously proposed in two important ways: It allows for a general choice model and a more general framing structure.

- We prove that the product framing problem is NP-hard, even when there are just 2 pages and the choice model is the MNL model.
- We propose fast, easy-to-implement algorithms with worst-case performance guarantees. Our algorithms are significantly simpler than existing algorithms and offer strong performance bounds. The ease and simplicity of the algorithms mean that they can be personalized on-line for each arriving consumer. The algorithms also apply to an extended model when the product's display location affects both attention (i.e., the formation of consideration sets) and valuation (i.e., the comparison of utilities).
- We prove new structural results for pricing under framing effects. We show that given a fixed placement of products, at optimality, each page is filled with products as many as possible. All products on the same page have the same page-level markup, which increases monotonically with respect to page indices. Products with higher value gaps are given higher priority, where the value gap is defined as expected utility when priced at cost. This implies that the optimal markup is higher for less attractive products. This last finding is contrary to the findings of Arbatskaya (2007), who focuses on the oligopoly market and argues that sellers lower on a list will charge lower prices; thus, consumers with lower search costs will search longer and obtain better deals.

We remark that our model, although tailored to e-commerce and virtual pages, can also be used to model brick-and-mortar retailers, with a suitable interpretation of what consumers are willing to look at. Some consumers, for example, would only look at the most prominent displays (see Chandon, Hutchinson, Bradlow and Young 2009, Corstjens and Corstjens 2012), while others may enter a store and look at some aisles or the rest of the store. Our location-preference model also applies in brick-and-mortar settings to signify the value of having products at eye level versus waist level, versus shoe level; and the value of end-of-aisle locations. Our pricing results also have implications for brick-and-mortar retailers. The most prominent displays should have the highest utility products at the lowest markups.

3.1.3 Relation to assortment planning

This chapter falls within the literature on assortment planning, which is currently a very active area of research. Assortment planning began with a stylized model introduced by van Ryzin and Mahajan (1999). Van Ryzin and Mahajan (1999) show that under the MNL model, an optimal assortment consists of a certain number of highest-utility products when the products are equally profitable. When the products' prices are given exogenously and the choice model is the MNL, Talluri and van Ryzin (2004) prove that an optimal assortment includes a certain number of products with the highest revenues.

The assortment-planning problem is easy to solve for the MNL model over a given consideration set. Davis, Gallego and Topaloglu (2013) show that this problem can be formulated as a linear program with totally unimodular constraints. Davis, Gallego and

Topaloglu (2014) also propose that under the nested logit (NL) model, the assortment-planning problem can be solved by a linear program when the nest-dissimilarity parameters of the choice model are less than one, and a consumer always makes a purchase within the selected nest. Relaxing either of these assumptions renders the problem NP-hard.

The assortment-planning problem is NP-hard for general choice models. Indeed, Bront, Mendes-Diaz and Vulcano (2009) show that under the mixed multinomial logit (MMNL) choice model, the assortment-planning problem with a fixed number of mixtures is NP-hard. Desir and Goyal (2013) show that this problem is even NP-hard to approximate within a factor of $O(n^{1-\varepsilon})$, for any fixed $\varepsilon > 0$. They give approximation schemes that tradeoff running time with solution quality, but the running time for their approach grows exponentially with the number of mixtures.

Few papers have studied assortment planning with location effects, but ignoring consideration sets. Davis, Gallego and Topaloglu (2013) model location effects by introducing location-dependent item weights to the MNL model. The resulting assortment-optimization problem reduces to a linear program with totally unimodular constraints.

Assortment planning under consideration-set-based choice models have been studied by a number of authors. One stream works with endogenous consideration sets that arise as a result of search. Cachon (2005) shows that ignoring consumer search will lead to less assortment variety, since in equilibrium, the seller needs a larger assortment to attract more consumers. Sahin and Wang (2015) also study the assortment-optimization problem with search costs. They assume consumers are homogeneous and their search sequence is predetermined by all the products' expected utilities, which are common knowledge.

Feldman and Topaloglu (2015) study a model in which consumers choose products according to the MNL model, but consumers of different types have different consideration sets, and the sets are fixed and nested. They devise a fully polynomial-time approximation scheme for this problem. To our knowledge, there are only two papers that model a framing-dependent formation of consideration sets. Davis, Topaloglu and Williamson (2015) study a problem in which a firm must sequentially add products to its assortment over time, thereby monotonically increasing consumers' consideration sets. They provide an algorithm with constant relative performance. The decision space for this problem is much more constrained than ours and the application context is very specific. Aouad and Segev (2016) consider a variant of our model, where the number of products that can be displayed on each page is one, the choice model is the MNL, and all products must be displayed even if doing so is suboptimal.

3.1.4 Relation to assortment pricing

Our work also falls within the area of assortment pricing. Hanson and Martin (1996) are among the first to notice that the expected revenue function fails to be concave in pricing problems, even under the MNL model. Song and Xue (2007) show that the expected revenue is concave with respect to the market shares. Under the MNL model with uniform price-sensitivity parameter, the markup, defined as price minus cost, has been shown to be constant across all products at optimality (Anderson, de Palma and Thisse 1992, Hopp and Xu 2005, and Gallego and Stefanescu 2011). By assuming that the price sensitivities of the products are constant within each nest and the nest dissimilarity parameters are restricted to the unit interval, Li and Huh (2011) extend the concavity result to the NL model. Gallego and Wang (2014) consider the general NL model with product-differentiated price-sensitivity parameters and arbitrary nest coefficients. They find that the adjusted nest-level markup is also constant across all the nests.

We extend the assortment-pricing literature to model framing effects. Under the MNL revenue- (profit-) maximization model, we find that the constant price (markup) property still holds at the page level. We also show that the price is higher for less attractive products, which is contrary to the findings of Arbatskaya (2007).

3.2 Product-Framing Problem

Consider n products. Product i has revenue r_i, $i \in N = \{1,\ldots,n\}$. The revenue can be the profit contribution net of costs in some applications. Products are organized into virtual pages. Each page can hold up to p products. Potentially all of the products may be offered, but offering all of the products is not a hard requirement. Consumers who arrive at the system have consideration sets that are governed by a random variable X taking values in a set M that consists of all the positive integers, or is of the form $\{1,\ldots,m\}$ for some finite positive integer m. Let $\lambda(x) = \mathbf{P}[X = x]$, and $\Lambda(x) = \mathbf{P}[X \geq x]$ for all $x \in M$. A consumer who draws $X = x \in M$ has consideration set $\{1,\ldots,x\}$, i.e., he will consider all products in the first x pages. From this consideration set, the consumer purchases at most one product according to a general choice model.

The product-framing problem is to distribute the products among the pages to maximize the expected revenue that can be obtained from an arriving consumer. We assume that we *do not know* the number of pages that a consumer is willing to view when he arrives into the system. We assume, however, that the distribution of X is known and is independent of the framing of the products. Knowledge of X can be acquired from observing click data and by computing the frequency of consumers who examine $x \in M$ pages. By the law of large numbers, these frequencies converge to the probability distribution of X. We call this multi-page assortment-optimization problem, the *product-framing problem*. Although we will refer to a single random variable X, it is easy to see that X can be personalized to heterogeneous consumer types based on available information about the distribution of pages they are willing to see. Information that may change the distribution of X includes, but is not limited to, prior purchases, zip code, age, and gender.

We will assume that consumers choose according to a general choice model that is independent of the consumer type $x \in M$. Later, we will relax this assumption and show that under mild conditions we can still guarantee a constant ratio of the expected revenue relative to the upper bound.

The product-framing problem can be formulated in terms of decision variables $y_{ij} \in \{0,1\}$, $i \in N$, $j \in M$, where $y_{ij} = 1$ if item i is displayed on page j and is zero otherwise. Let $\pi(i,S)$ denote the purchase probability of item i when the consideration set is $S \subseteq N$, with $\pi(i,S) = 0$ if $i \notin S$. The formulation in terms of the variables y_{ij} is given by

$$\begin{aligned}
\text{OPT} = & \max_{y_{ij}} \sum_{x \in M} \lambda(x) \sum_{i=1}^{n} r_i \pi(i, \{k \in N : \sum_{l=1}^{x} y_{kl} = 1\}) \\
\text{s.t.} \; & \sum_{j \in M} y_{ij} \leq 1, \; \forall i \in N \\
& \sum_{i=1}^{n} y_{ij} \leq p, \; \forall j \in M \\
& y_{ij} \in \{0,1\}, \; \forall i \in N, \; j \in M.
\end{aligned} \tag{3.2.1}$$

We will show that problem (3.2.1) is NP-hard. Therefore, to derive performance bounds for our algorithms, we will first find an upper bound on (3.2.1), which can be easily computed.

3.3 Upper Bound on Optimal Revenue for Product Framing

Consider the following assortment-optimization problem, which constrains the number of products in an assortment to be at most c.

$$\begin{aligned}
G(c) \; & = \max_{S \subseteq N} \sum_{i \in S} r_i \pi(i,S) \\
\text{s.t.} \; & |S| \leq c.
\end{aligned} \tag{3.3.1}$$

Let $R(x) = G(x \cdot p)$ be the optimal expected revenue from consumers who see $x \in M$ pages. Let $S(x) \subset N$ be an optimal solution associated with the revenue $R(x)$. If we had the luxury of knowing the number of pages $x \in M$ upon the arrival of a consumer, we would offer him assortment $S(x)$, and would earn expected revenue

$$E[R(X)] = \sum_{x \in M} \lambda(x) R(x). \tag{3.3.2}$$

The following result shows that this $E[R(X)]$ is an upper bound on the expected value V^{OPT} of OPT.

Theorem 3.3.1. $\mathbf{E}[R(X)] \geq V^{OPT}$.

Proof. Suppose y^* is an optimal solution to (3.2.1). We must have, for any $x \in M$,

$$R(x) \geq \sum_{i=1}^{n} r_i \pi(i, \{k \in N : \sum_{l=1}^{x} y^*_{kl} = 1\})$$

$$\Longrightarrow \sum_{x \in M} \lambda(x) R(x) \geq V^{OPT}.$$

□

3.4 Hardness of Framing Problem

We show that problem (3.2.1) is NP-hard even in the special case that $m = 2$ and the choice model is the MNL model. We do this by reducing the well-known *2-PARTITION* problem to a special case of our model. The *2-PARTITION* problem is defined as follows

Definition 3.4.1 (2-PARTITION)**.** *Given a set of n non-negative numbers* $w_1, w_2, \ldots, w_n$*, determine whether there is a set* $S \subseteq \{1, 2, \ldots, n\}$ *such that* $\sum_{i \in S} w_i = \sum_{i \notin S} w_i$.

Our reduction works as follows. Starting with any instance of 2-PARTITION, we design an instance of problem (3.2.1). We show that the solution to the continuous relaxation of this problem takes a certain value if and only if there is a solution to the 2-PARTITION problem.

Theorem 3.4.2. *Problem* (3.2.1) *is NP-hard even when all consumers follow the same MNL model.*

3.5 Assumptions for Analysis of Algorithms

Given that the product-framing problem is NP-hard, one of our goals will be to propose algorithms with guaranteed performance ratios relative to OPT. Towards this goal, we will make three innocuous assumptions:

Assumption 3.5.1.

$$\pi(i,S) \geq \pi(i,T) \ \text{ for all } \ i \in S, \ \text{ and } S \subseteq T \subseteq N.$$

Assumption 3.5.2. *In polynomial time, we can obtain a solution with expected revenue* $\bar{G}(c)$ *to problem (3.3.1) such that* $\bar{G}(c) \geq (1-\varepsilon)G(c)$ *for some constant* $\varepsilon \in (0,1]$.

Assumption 3.5.3. *X has new better than used in expectation (NBUE) distribution.*

Assumption 3.5.1 is very general as it holds for all random utility models. Davis, Topaloglu and Williamson (2015) have shown that Assumption 3.5.1 leads to the following results which we will use in the analysis.

Lemma 3.5.4 (Davis, Topaloglu and Williamson 2015)**.** *For any set $S \subseteq N$ with $|S| \geq 2$, there exists $i \in S$ such that*

$$\frac{\sum_{k\in S, k\neq i} r_k \pi(k, S\setminus\{i\})}{|S|-1} \geq \frac{\sum_{k\in S} r_k \pi(k,S)}{|S|}.$$

Lemma 3.5.5 (Davis, Topaloglu and Williamson 2015)**.** *$R(x)/x$ is decreasing in $x \in M$.*

Assumption 3.5.2 states that we can approximately solve the capacity-constrained problem (3.3.1) within a constant approximation ratio. For the MNL and Nested Logit models, the capacity-constrained problem can be solved exactly in polynomial time ($\varepsilon = 0$) (Gallego and Topaloglu 2014). The Mixed Multinomial Logit model with a constant number of mixtures can also be solved within any given error $\varepsilon > 0$ in polynomial time . For ease of exposition, in the rest of the chapter we will just assume that $\varepsilon = 0$, but all of our algorithms and bounds can be easily extended to the case of $\varepsilon > 0$ by replacing the revenue function $R(x)$ with an approximate value, and scaling the corresponding bound by $(1-\varepsilon)$.

Let $q(x) = E(X-x|X>x)$ for all $x \in M$. Assumption 3.5.3 is equivalent to $q(x) \leq q(0)$ for all $x \in M$. Assumption 3.5.3 implies that the additional number of pages that a consumer will see is no more than the expected number of pages he or she would like to see before the search.

3.6 Algorithms with constant Performance Bounds for Product Framing

We now propose algorithms for problem (3.2.1), which have guaranteed constant performance ratios relative to *OPT* under Assumptions 3.5.1, 3.5.2, and 3.5.3. All of our performance guarantees in this section are tight relative to the upper bound (3.3.2) in the sense that our proofs exhibit ways to construct instances in which the bounds are achieved by our algorithms.

Our first algorithms, called NEST, start by truncating the number of pages to an arbitrary integer y, and by selecting an optimal assortment, say $S(y)$, for the first y pages. Then they select an optimal assortment for the first $y-1$ pages, say $S(y-1)$, by looking only at the products in $S(y)$. This procedure continues until the content of all pages have been determined. With a little bit abuse of notations, we let $R(S)$ denote the expected revenue when a consumer considers all of the products in the set S.

NEST(y) Algorithms:

- Let y be an integer and let $S(y)$ be the set of products with revenue $R(y)$. Without loss of generality, we assume that $|S(y)| > (y-1)p$. In other words, a minimum of y pages are required to hold all products in $S(y)$. If this fails, then we can select a smaller value of y until this assumption holds.

- For $x = y-1$ down to 1, let $S(x)$ be the set of products to be displayed on the first x pages, given that $S(x) \subset S(x+1)$ and $|S(x)| = x \cdot p$, and let $\tilde{R}(x) = R(S(x))$. Using

Lemma 3.5.4, we can set $S(x)$ to be a subset of $S(x+1)$ such that

$$\frac{\tilde{R}(x)}{x} \geq \frac{\tilde{R}(x+1)}{x+1}. \tag{3.6.1}$$

This implies that $\tilde{R}(x)/x \geq \tilde{R}(y)/y = R(y)/y$ for all $x < y$.

- Use any heuristic to fill products into pages $x > y$, such that the total expected revenue of products in the first $x > y$ pages is at least $R(y)$. As a default, we can leave pages $x > y$ blank.

3.6.1 Constant lower bound

Notice that the revenue of NEST(y) for pages $x < y$ is guaranteed to be at least $\tilde{R}(x) \geq xR(y)/y$, and the expected revenue for $x \geq y$ is at least $yR(y)/y$. Consequently,

$$\begin{aligned}
V^{NEST(y)} &\geq \sum_{x=1}^{y-1} \tilde{R}(x)\lambda(x) + R(y)\Lambda(y) \\
&\geq \frac{R(y)}{y}\sum_{x=1}^{y-1} x\lambda(x) + R(y)\Lambda(y) \quad \text{(by (3.6.1))} \\
&= \frac{R(y)}{y}\sum_{x=1}^{y-1} x\lambda(x) + \frac{R(y)}{y} y\Lambda(y) \\
&= \frac{R(y)}{y} E[\min(X, y)].
\end{aligned}$$

Let $NEST = NEST(y)$ be an algorithm in the above class where y is selected to maximize the lower bound. More precisely, let y be the largest integer in the set

$$\arg\max_{x\in M} E[\min X,x]R(x)/x.$$

Then $V^{NEST} \geq \frac{R(y)}{y}E[\min(X,y)]$ for this choice of y. We will show that $NEST$ achieves at least $6/\pi^2 = 0.607921\ldots$ of the optimal expected revenue.

The idea of the proof is to minimize $\frac{R(y)}{y}E[\min(X,y)]$ over all functions $R(\cdot)$ satisfying Lemma 3.5.5, and over all distributions X satisfying Assumption 3.5.3. For convenience we will scale $R(\cdot)$ without loss of generality so $E[R(X)] = 1$. This leads to a min max problem that can be formulated as follows:

$$\begin{aligned}
\gamma = \quad \min_{R,\,\Lambda} \quad & \max_{x\in M} \frac{R(x)}{x}E[\min(X,x)], && (3.6.2)\\
s.t. \quad & 1 = \Lambda(1) \geq \Lambda(2) \geq \cdots \geq 0 \\
& \Lambda(x+1) \times \sum_{y>0} \Lambda(y) \geq \sum_{y>x} \Lambda(y), \quad x \in M \\
& R(x) \leq R(x+1), \quad x \in M, \\
& \frac{R(x)}{x} \geq \frac{R(x+1)}{x+1}, \quad x \in M, \\
& \sum_{x\in M} \lambda(x)R(x) = 1 \\
& R(x) \geq 0 \;\; x \in M.
\end{aligned}$$

The first constraint ensures that Λ corresponds to a valid tail distribution. The second constraint ensures that X has the NBUE property. The third and fourth constraint ensure that R is increasing and satisfies Lemma 3.5.5. The fifth constraint normalizes $E[R(X)]$ to 1, and the last ensures that R is non-negative.

To uncover the structure of an optimal solution to (3.6.2) we will first characterize the functions $R(\cdot)$ and $\Lambda(\cdot)$ in the worst case. In the process of establishing the bounds, we will not use special notation, say $\gamma^*, R^*(\cdot)$ or $\Lambda^*(\cdot)$ to denote the optimal solution to program (3.6.2). This comes at a small cost of ambiguity, but makes the exposition a bit cleaner.

Let y be the largest integer in the set $\arg\max_{x\in M} \frac{R(x)}{x}E[\min(X,x)]$ and $\gamma = \frac{R(y)}{y}E[\min(X,y)]$. We next show the worst-case structure for $R(x)$ for all $x \in M$.

Lemma 3.6.1. $R(x) = \gamma\frac{x}{E[\min(X,x)]}$ *for all* $x \in M$.

The lemma implies that in the worst case, the maximum value of $E[\min(X,x)]R(x)/x$ is achieved by all points x in the set M. In other words, the worst-case function R and the worst-case distribution for X are such that the value $E[\min(X,x)]R(x)/x$ is constant for all $x \in M$.

Proof. We first show that the function $\frac{R(x)}{x}E[\min(X,x)]$ cannot decrease in x. Suppose for a contradiction that there is a smallest x, $x > 1$, such that the function decreases strictly from $x-1$ to x. That is, $\frac{R(x-1)}{x-1}E[\min(X,x-1)] > \frac{R(x)}{x}E[\min(X,x)]$. Then we can revise the function as follows. For points $y \in \{x, x+1, x+2, ..., m\}$ such that $R(y) = R(x)$, we increase their value $R(y)$, such that $R(y) \leftarrow (1+\varepsilon)R(x)$. Let us denote this set of $y's$ by $\mathcal{Y}$. Note it might not be a singleton, since there might be consecutive $y's$ such that $R(y) = R(x)$. In this meantime, we scale down the $R(\cdot)$ at all other points $y \in M - \mathcal{Y}$, i.e., $R(y) \leftarrow (1-\varepsilon')R(y)$. We can properly chose the values of ε and ε' such that $E[R(X)] = 1$ is maintained. Let z

be the largest index in the set $\mathcal{Y}$, then the only two constraints that might get violated by this revision are $\frac{R(x-1)}{x-1} \geq \frac{R(x)}{x}$ and $R(z) \leq R(z+1)$. However, $\frac{R(x-1)}{x-1}E[\min(X,x-1)] > \frac{R(x)}{x}E[\min(X,x)]$ implies $\frac{R(x-1)}{x-1} > \frac{R(x)}{x}$, which means the decreasing unit reward condition will not be violated as long as ε and ε' are chosen small enough; moreover, $R(z) \leq R(z+1)$ will not be violated either since by the definition of set $\mathcal{Y}$ we know $R(z) < R(z+1)$ before the revision. So again, the increasing reward condition will not be violated as long as ε and ε' are chosen small enough. Thus the optimality condition meets contradiction, as we can strictly decrease the value of γ.

We now show that the function $\frac{R(x)}{x}E[\min(X,x)]$ cannot increase in x either. Suppose there is a smallest x, $x > 1$ such that the function increases strictly from $x-1$ to x, or equivalently $\frac{R(x-1)}{x-1}E[\min(X,x-1)] < \frac{R(x)}{x}E[\min(X,x)]$. Then we can scale up the value of $R(\cdot)$ at all points $y < x$, and scale down $R(\cdot)$ at all points $y \geq x$, while maintaining the constraint $E[R(X)] = 1$. This adjustment is valid because the only violation might occur at pair $(x-1,x)$ for the constraint $R(x-1) \leq R(x)$. However, $\frac{R(x-1)}{x-1}E[\min(X,x-1)] < \frac{R(x)}{x}E[\min(X,x)]$ implies $R(x-1) < R(x)$, as $\frac{E[\min(X,x-1)]}{x-1} > \frac{E[\min(X,x)]}{x}$. Consequently, the condition that $R(\cdot)$ be increasing will not be violated as long as we properly choose the scaling parameter. This leads to a contradiction as we can strictly decrease the value of γ. □

Given the worst case structure for $R(\cdot)$, we next show the relation between γ and the distribution $\Lambda(\cdot)$ by using the normalization condition $E[R(X)] = 1$.

Theorem 3.6.2. *Let Y be a random variable that is independent of the worst case X and* *has the same distribution as X, i.e., $P(Y \geq x) = \Lambda(x)$, $\forall x \in M$. Then*

$$\frac{1}{\gamma} = E\left[\frac{X}{E[\min(X,Y)|X]}\right]. \tag{3.6.3}$$

Proof. By Lemma 3.6.1, $R(x) = \gamma \frac{x}{E[\min(Y,x)]}$, where Y is a random variable with the same distribution as X. Then

$$1 = E[R(X)] = \gamma \sum_{x \in M} \frac{x}{E[\min(Y,x)]} \lambda(x) = \gamma E\left[\frac{X}{E[\min(X,Y)|X]}\right].$$

The result follows after dividing the equation by γ. □

Our next goal is to show that $E[\frac{X}{E[\min(X,Y)|X]}] \leq \pi^2/6$ among all non-negative, NBUE distributions. In our analysis, we will allow for continuous distributions to obtain the bound. We will later show that there is a discrete distribution that in the limit achieves the bound.

Lemma 3.6.3. *If X is non-negative with mean μ and NBUE, and Z is exponential with the same mean, then $X \leq^{icx} Z$ in the increasing convex ordering.*

Proof. Let $\bar{F}_X$ denote the CCDF of X and $\bar{F}_Z$ denote the CCDF of Z. Proving $X \leq^{icx} Z$ is equivalent to proving $\int_a^\infty \bar{F}_X(v)dv \leq \int_a^\infty \bar{F}_Z(v)dv = \mu e^{-a/\mu}$ for any $a \geq 0$. Since X has NBUE property, we have $E(X - t | X > t) \equiv \frac{\int_t^\infty \bar{F}_X(v)dv}{\bar{F}_X(t)} \leq \mu$ for all $t \geq 0$, indicating

$$\frac{\bar{F}_X(t)}{\int_t^\infty \bar{F}_X(v)dv} \geq \frac{1}{\mu}$$

We integrate both sides over $t \in [0,a]$, and notice that the left hand side can be written as $-d\ln\int_t^\infty \bar{F}_X(v)dv$. We have

$$\ln\int_0^\infty \bar{F}_X(v)dv - \ln\int_a^\infty \bar{F}_X(v)dv \geq \frac{a}{\mu}$$

Since $\ln\int_0^\infty \bar{F}_X(v)dv = \ln E(X) = \ln\mu$, the above reduces to the desired expression:

$$\int_a^\infty \bar{F}_X(v)dv \leq \mu e^{-a/\mu} = \int_a^\infty \bar{F}_Z(v)dv.$$

□

Corollary 3.6.4. *If Y is a non-negative, NBUE random variable with mean μ, and Z is exponential with the same mean, then $E[\min(Y,x)] \geq E[\min(Z,x)]$ for all $x \geq 0$.*

Proof. Applying the Lemma 3.6.3 to Y and Z, we see that $E[(Y-x)^+] \leq E[(Z-x)^+]$, so $E[\min(Y,x)] = E[Y] - E[(Y-x)^+] \geq E[Z] - E[(Z-x)^+] = E[\min(Z,x)]$. □

Corollary 3.6.5. *For any independently and identically distributed X and Y that are NBUE and have mean μ, $E\left[X/E[\min(X,Y)|X]\right] \leq E\left[W/E[\min(Z,W)|W]\right]$, where Z and W are independent exponentials with mean μ*

Proof. Define $H(x) \equiv x/\mathbf{E}[\min(Z,x)]$. From Corollary 3.6.4, $x/\mathbf{E}[\min(Y,x)] \leq x/\mathbf{E}[\min(Z,x)] = H(x)$. Since $H(x)$ is increasing convex, and $X \leq^{icx} W$, it follows from Lemma 3.6.3 that $\mathbf{E}[H(X)] \leq \mathbf{E}[H(W)]$. □

We are now ready to present the bound of our algorithm.

Theorem 3.6.6. *The expected revenue V^{NEST} is at least $\frac{6}{\pi^2}V^{OPT}$.*

Proof. By Lemma 3.6.1, $R(x) = \gamma \frac{x}{E[\min(X,x)]}$, and by Theorem 3.6.2, and Corollary 3.6.5 , $1/\gamma = E\left[\frac{W}{E[\min(Z,W)|W]}\right]$, where W and Z are independent exponentially distributed random variables with the same mean, say μ. We will now show that $E\left[Z/E[\min(Z,W)|W]\right] = \pi^2/6$, which is independent of μ. This will imply that

$$\frac{V^{NEST}}{V^{OPT}} \geq \gamma = 6/\pi^2,$$

completing the proof.

Since W is exponential with mean μ, we have $E[\min(Z,W)|Z] = \mu(1-\exp(-Z/\mu))$. Substituting this expression into the denominator, we obtain

$$E\left[\frac{Z}{\mu(1-\exp(-Z/\mu))}\right] = \int_0^\infty \frac{z/\mu \exp(-z/\mu)}{1-\exp(-z/\mu)} dz/\mu = \int_0^\infty \frac{ue^{-u}}{1-e^{-u}} du = \sum_{x=1}^{\infty} \frac{1}{x^2} = \frac{\pi^2}{6},$$

where the first equality follows from the distribution of Z, the second from the transformation $u = z/\mu$, the third equality is a well know result from calculus, and the last equality is an important problem in number theory, posed by Mengoli in 1644. This problem remained open for 90 years until Euler solved it in 1734 at the age of 28.

□

Our results show that the bound for Problem (3.6.2) is tight for the exponential distribution. The reader may wonder whether there is a discrete distribution over the non-negative integers such that the bound for Problem (3.6.2) is also tight. The following corollary asserts that this is indeed the case.

Corollary 3.6.7. *The performance ratio* $6/\pi^2$ *with respect to the upper bound (3.3.2) is attained when* X *and* Y *have geometric distributions with mean* $1/(1-p)$ *as* $p \uparrow 1$.

Proof. Let X, Y be geometrically distributed with mean $\frac{1}{1-p}$. That is, $P[X = x] = P[Y = x] = p^{x-1}(1-p)$. Then we can write

$$E\left[\frac{X}{E[\min(X,Y)|X]}\right] = \sum_{y=1}^{\infty} \frac{p^{y-1}(1-p)^2 y}{1-p^y}.$$

For $p < 1$, we have

$$\begin{aligned}
\sum_{y=1}^{\infty} \frac{p^{y-1}(1-p)^2 y}{1-p^y} &= (1-p)^2 \sum_{y=1}^{\infty} \frac{y}{p} \frac{p^y}{1-p^y} \\
&= (1-p)^2 \sum_{y=1}^{\infty} \frac{y}{p} \sum_{n=1}^{\infty} p^{yn} \\
&= \frac{(1-p)^2}{p} \sum_{n=1}^{\infty} p^n \sum_{y=1}^{\infty} y(p^n)^{y-1} \\
&= \frac{(1-p)^2}{p} \sum_{n=1}^{\infty} p^n \frac{d}{dp^n} \frac{1}{1-p^n} \\
&= \frac{(1-p)^2}{p} \sum_{n=1}^{\infty} p^n \frac{1}{(1-p^n)^2} \\
&= \sum_{n=1}^{\infty} p^{n-1} \frac{(1-p)^2}{(1-p^n)^2}.
\end{aligned}$$

The above series is an increasing function of p. It is maximized both locally and globally at $p = 1$. To find the limit at $p = 1$, we use two applications of L'Hospital's rule to obtain $\lim_{p\to 1}\sum_{n=1}^{\infty} p^{n-1}\frac{(1-p)^2}{(1-p^n)^2} = \sum_{n=1}^{\infty}\frac{1}{n^2} = \frac{\pi^2}{6}$.

□

3.6.2 Improvement to the algorithms

We can refine the NEST Algorithms by extending the algorithms to pages beyond y. In particular, let $\tilde{R}(x)$ be the performance of the NEST Algorithms described above for all $x \leq y$. Set $x = y+1$ and solve the x-page problem subject to the constraint that the first y pages are composed of $S(y)$. Then select the products for page $x = y+1$ by solving the problem for page x among all products in the complement of $S(y)$. For higher values of x, continue the procedure, fixing the pages for all $z < x$ and selecting page x among the complement of the products used in pages $1,\ldots,x-1$. Let $NEST+(y)$ be an algorithm in this improved class. When y is chosen as the largest integer in the set

$$\arg\max_{x\in M} V^{NEST+(x)}$$

then we call $NEST+(y)$ as $NEST+$. Note that $NEST+$ continues to have a worst-case performance bound of $\frac{6}{\pi^2}$ as we have proved in Section 3.6.1.

Corollary 3.6.8. *If $R(x)$ has nested solutions, i.e, $S(1) \subseteq S(2) \subseteq ... \subseteq S(m)$, where $S(x)$ is the optimal assortment for $R(x)$, then NEST+(y) returns an optimal solution. A sufficient condition would be that the capacitated assortment problem $G(c)$ has nested solutions.*

3.7 Product Framing with Type-Dependent Consumer Choice Models

In this section we relax the requirement that the consumer choice model must be the same for all customers. This relaxation allows us to model *keen customers* who are more determined to buy, and *picky customers* who are more price- or quality-sensitive.

To this end, we allow the choice model to be type dependent. Accordingly, for a consumer of type x, who intends to view x pages, we let $\pi_x(i,S)$ be the purchase probability of item i in assortment S. We will show that under mild assumptions, we can still design an algorithm to guarantee an expected revenue of $1/3$ of the upper bound.

Define

$$r(x,S) = \sum_{i \in S} r_i \pi_x(i,S) \tag{3.7.1}$$

to be the expected revenue from presenting assortment S to a consumer of type x. We will still use $R(x)$ to denote the optimal revenue from the capacitated assortment problem for consumers with type x, namely,

$$R(x) = \max_{S \in N} \; r(x,S)$$

$$\text{s.t.} |S| \leq x \cdot p.$$

It is easy to verify that in this extended model Theorem 3.3.1 is still valid, i.e., $\mathbf{E}[R(X)]$

is still an upper bound on the optimal expected revenue, as we only need to replace $\pi(\cdot,\cdot)$ with $\pi_x(\cdot,\cdot)$ in the proof of Theorem 3.3.1.

We make the following assumptions:

Assumption B1. *$r(x,S) \leq r(y,S)$ for all $x \leq y$ and $S \subseteq N$ with $|S| \leq x \cdot p$*

Assumption B2. *$R(x)/x$ is decreasing in x.*

Assumption B3. *Same as Assumption 3.5.2.*

Assumption B4. *X has Increasing Failure Rate (IFR).*

Assumption B1 was implicitly true in previous sections as $r(x,S) = r(y,S)$ when all consumers follow the same choice model. Assumption B2 is weaker than Assumption 3.5.1 because, according to Lemma 3.5.5, the former is a result of the latter.

Assumption B4 is stronger than Assumption 3.5.3 as it is sufficient for Assumption 3.5.3 (Shaked and Shanthikuma 2007). It is equivalent to $h(x) \equiv \frac{\lambda(x)}{\Lambda(x)}$ increasing in $x \in M$. Assumption B4 implies that the probability that a consumer will view the next page is decreasing in the number of pages he or she has viewed.

To gauge the appropriateness of our algorithms in settings where our assumptions might not hold, we perform computational experiments in these settings in Section 3.10.2. The experiments indicate that our algorithms significantly outperform greedy heuristics even in settings where the choice model changes drastically with x and the monotonicity of $r(x,S)$ in x is violated.

3.7.1 TRUNCATE (TRUNC) Algorithms

This algorithm simply optimizes for the set of items to be included in the first x pages, for a well-chosen x. In other words, it truncates the number of pages to exactly x. It does not try to optimize for the placement of these items within the x pages. The idea is to cater only to consumers who will view at least x pages.

Algorithm:

1. Choose $y = \max \arg\max_{x \in M} R(x)\mathbf{P}[X \geq x]$. Let $S(y)$ be the set of items with expected revenue $R(y)$.

2. Use any heuristic to fill in the first y pages such that the set of all items in the first y pages is $S(y)$.

3. Use any heuristic to fill in pages $x > y$ such that the total expected revenue of items in the first $x > y$ pages is at least $R(y)$. As a default, we can leave pages $x > y$ blank. This directly follows Assumption B1.

Let $TRUNC$ be an algorithm in the above class. We next show that $TRUNC$ achieves at least $1/3$ of the optimal expected revenue. The main idea of the proof is as follows. Clearly $R(y)\Lambda(y)$ is a lower bound on the expected revenue V^{TRUNC} of $TRUNC$. We will minimize the quantity $R(y)\Lambda(y)$ over all IFR distributions of X and all increasing functions $R(\cdot)$ satisfying Assumption (B2). We will scale $R(\cdot)$ without loss of generality so the upper bound $E[R(X)]$ is normalized to 1. We will then show that the smallest value of $R(y)\Lambda(y)$ is at least $1/3$. Doing this is equivalent to proving the same lower bound on the following optimization problem:

$$\begin{aligned} r = \min_{R,\,\Lambda} \quad & \max_{x \in M} R(x)\Lambda(x) \\ s.t. \quad & 1 = \Lambda(1) \geq \Lambda(2) \geq \cdots \geq 0 \\ & \Lambda(x+1)\Lambda(x-1) \leq \Lambda(x)^2, \quad x \geq 2 \\ & R(1) \leq R(2) \leq \cdots \\ & R(x) \geq \frac{x}{x+1} R(x+1), \quad x \geq 1, \\ & \sum_{x \in M} (\Lambda(x) - \Lambda(x+1)) R(x) = 1 \\ & R(x) \geq 0 \;\; x \in M. \end{aligned} \tag{3.7.2}$$

All the constraints follow the same logic as those in optimization problem (3.6.2). However, note that the third constraint that R be increasing is just a necessary condition of Assumption B1.

We are now ready to present our main result for the TRUNC algorithm. Please refer to the Appendix for the detailed analysis.

Theorem 3.7.1. *Let R and Λ be an optimal solution of* (3.7.2)*. Then in the worst case X is geometric with mean $E[X] = y$, and*

$$r = \frac{1}{3 - 2/y} \geq \frac{1}{3}. \tag{3.7.3}$$

Notice the this bound is sharper than $1/3$ when y is small. We have $r = 1$ if $y = E[X] = 1$ as expected because then people only see page one and TRUNC is optimal. The result is

$r \geq 1/2$ when $y = E[X] = 2$, and decreases slowly to $1/3$ as y increases to infinity. Thus, the 1/3 bound is attained in the limit in the unrealistic case that people see in expectation an infinite number of pages, and there are an infinite number of products.

As a corollary, we have the main result for this section.

Corollary 3.7.2. *Under Assumptions B1, B2, B3, and B4,* $r \geq 1/3$*, and consequently* $V^{TRUNC} \geq 1/3 V^{OPT}$.

3.8 Product Framing with Location Preferences

In online retail, consumers may be more likely to choose products that are displayed at the top among search results, since consumers tend to associate high valuation with products that are displayed earlier (Chandon, Hutchinson, Bradlow, and Young 2009). In this section, we augment our model to capture the phenomenon that a consumer is more likely to buy a product that is displayed earlier, even if the consumer has determined his consideration set. We model this phenomenon by introducing location-dependent preference weights for all products. We use ν_{ixq} to denote the preference weight of product i when this product is displayed at location q on page x. Without loss of generality, we assume that there are as many locations as there are products so that we can offer all products at once. If the number of possible locations is smaller than the number of products, then we can define additional locations with $\nu_{ixq} = 0$ for all $i \in N$ for each additional location (x, q). In this case, using one of these additional locations for a product is equivalent to not displaying the product at all. To capture the product-framing decisions, we use

$y = \{y_{ixq} : i \in N, x \in M, q \in \mathcal{P}\} \in \{0,1\}^{n\times m\times p}$, where $y_{ixq} = 1$ if we offer product i in location q of page x; otherwise $y_{ixq} = 0$. If the product offer decisions are given by y, then we obtain an expected revenue of $R_x(y) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{x}\sum_{q=1}^{p} y_{ijq}\nu_{ijq}r_i}{1+\sum_{i=1}^{n}\sum_{j=1}^{x}\sum_{q=1}^{p} y_{ijq}\nu_{ijq}}$ from consumers who only view the first x pages.

Recall that we can obtain the upper bound of the problem without location preference by $E(R(X)) \equiv \sum_{x=1}^{m} \lambda(x)R(x)$, where $R(x) \equiv G(x\cdot p)$ is the highest expected revenue from consumer who is willing to view x pages. Under the MNL choice model, for the problem with location preference, the problem $G(x\cdot p)$ is still polynomially solvable. This is because the problem is formulated as

$$
\begin{aligned}
\max_{y} \quad & \frac{\sum_{i=1}^{n}\sum_{j=1}^{x}\sum_{q=1}^{p} y_{ijq}\nu_{ijq}r_i}{1+\sum_{i=1}^{n}\sum_{j=1}^{x}\sum_{q=1}^{p} y_{ijq}\nu_{ijq}} \\
\text{s.t.} \quad & \sum_{i\in N} y_{ijq} \le 1 \;\; \forall j = 1,\ldots,x, \;\; q \in \mathcal{P};, \\
& \sum_{j=1}^{x}\sum_{q\in\mathcal{P}} y_{ijq} \le 1 \;\; \forall i \in N; \\
& y_{ijq} \in \{0,1\} \;\; \forall i \in N, \;\; j = 1,\ldots,x, \;\; q \in \mathcal{P}.
\end{aligned}
$$

where the first set of constraints ensures that each product is offered in at most one location and the second set of constraints ensure that each location is used by at most one product. The constraint matrix is that of an assignment problem, which is totally unimodular; see Corollary 2.9 in Chapter III.1 of Nemhauser and Wolsey (1988). With the linear fractional objective function, we know the problem is easily solvable; see Davis, Gallego and Topaloglu (2013).

We prove the following generalization of Lemma 3.5.5:

Lemma 3.8.1. *Assume that* $\nu_{ixq} \leq \nu_{ix'q'}$ *if* $(x-1)\times p+q > (x'-1)\times p+q'$ *for all products* $i \in N$*, i.e, the preference weight will decrease if the item is displayed further in the rear. Then* $\frac{R(x)}{x} \leq \frac{R(x')}{x'}$ *for any* $x, x' \in M$ *and* $x > x'$*.*

Proof. It is easy to see that the value of $R(x)$ is a root of the following equation:

$$R(x) = \max_{y} \sum_{i=1}^{n}\sum_{j=1}^{x}\sum_{q=1}^{p}(r_i - R(x))\nu_{ijq}y_{ijq}$$

Any suboptimal assortment yields the left hand side bigger than the right hand side. Under the optimal solution yielding $R(x)$, we can pick the $x' \times p$ products that give the highest $(r_i - R(x))\nu_{ijq}$. Let us denote this set of products by S. Then it must be true that

$$\sum_{i\in S}\sum_{j=1}^{x}\sum_{q=1}^{p}(r_i - R(x))\nu_{ijq}y_{ijq} \geq x'\frac{R(x)}{x}$$

If any item $i \in S$ is displayed in some page later than x', for example at position jq, then there must exist one product i' at position $j'q'$ such that $j' \leq x'$ and $i' \notin S$. We replace product i' by product i, then $\nu_{ij'q'} \geq \nu_{ijq}$ by assumption. Under the updated configuration y' it must be

$$\sum_{i\in S}\sum_{j=1}^{x'}\sum_{q=1}^{p}(r_i - R(x))\nu_{ijq}y'_{ijq} \geq x'\frac{R(x)}{x}$$

Since we are dealing with more restrictive cardinality constraint, i.e., $x' < x$, we have

$R(x') \leq R(x)$, therefore

$$R(x') \geq \sum_{i \in S} \sum_{j=1}^{x'} \sum_{q=1}^{p} (r_i - R(x')) \nu_{ijq} y'_{ijq} \geq x' \frac{R(x)}{x},$$

which implies that

$$\frac{R(x')}{x'} \geq \frac{R(x)}{x}.$$

□

With Lemma 3.8.1 and the increasing-failure-rate property, we can show that the $\frac{6}{\pi^2}$ performance bound still holds.

Theorem 3.8.2. *Assume that* $\nu_{ixq} \leq \nu_{ix'q'}$ *if* $(x-1) \times p + q > (x'-1) \times p + q'$ *for all products* $i \in N$. *Then all bounds proved in Sections 3.6 continue to hold for the model with location preference.*

3.9 Price Framing Problem

In practice, the retailer may care not only about how to select and display the products, but also how to price them to maximize expected revenues. In this section we consider the problem faced by a retailer who is jointly framing and pricing all products. The framing policy will still determine consumers' consideration sets, i.e., each customer will consider products on the first x pages with probability $\lambda(x)$, $x = 1, \ldots, m$. However, prices will influence consumers' valuation of the products. More specifically, under a given consideration set, we assume that consumers make choices according to the MNL model, with utility following a linear form: $u_i = a_i - \beta r_i + \varepsilon_i$. This formulation is commonly used in Economics, Marketing and Psychology (Berry, Levinsohn and Pakes 1995; Fader and Hardie 1996; Shugan 1980). Here, a_i is the price-independent quality of item i, $\beta > 0$ is the price sensitivity parameter, and r_i is the price of product i. Without loss of generality, the i.i.d. random perturbation terms $\varepsilon' s$ follow a standard Gumbel distribution and the expected utility of the outside alternative is zero, i.e., $E(u_0) = 0$.

Before rushing into the joint optimization problem, let us first consider the case where the products are framed and the only issue is to find optimal prices. More precisely, we assume that an assignment of products to pages is given, where $y_{ij} \in \{0,1\}$, and $y_{ij} = 1$ if and only if item i is displayed on page j. Naturally $\sum_j y_{ij} = 1$ and $\sum_i y_{ij} \leq 1$. Without loss of generality, we first look into revenue maximization problem by assuming cost $c = 0$. The decision variables then become the price vector $r = (r_1, ..., r_n) \in R^n_+$, where r_i denotes the price of item i. Later we will see the result can be easily extended to the profit maximization case with cost $c \neq 0$.

Let $S_x = \{i : y_{ix} = 1\}$ be the set of products displayed on page x and $\mathcal{S}_x = \cup_{j=1}^{x} S_j$ be the consideration set of consumers who view x pages. Assuming that the sets S_x, $\forall x \in M$, are fixed, we can express the pricing problem as follows:

$$R^{\text{pricing}} = \max_r R(r|S), \tag{3.9.1}$$

where

$$R(r|S) \equiv \sum_{x \in M} \lambda(x) R_x(r|\mathcal{S}_x)$$

is the expected total revenue under price r. Here, $R_x(r|\mathcal{S}_x) = \sum_{i \in \mathcal{S}_x} r_i \pi(i, \mathcal{S}_x)$ is the expected revenue from consumers who view exactly x pages, and $\pi(i, \mathcal{S}_x)$ is the probability that such a consumer purchases item i:

$$\pi(i, \mathcal{S}_x) = \begin{cases} \frac{\exp(a_i - \beta r_i)}{1 + \sum_{j \in \mathcal{S}_x} \exp(a_j - \beta r_j)}, & \text{if } i \in \mathcal{S}_x; \\ 0, & \text{otherwise.} \end{cases}$$

The following theorem states the structural results of the optimal pricing solution. We omit the proof because of space constraints.

Theorem 3.9.1. *Assume that the assignment of products onto m pages is given. To maximize expected revenue, all products on the same page have a common price. That is, there are page-dependent parameters* $\theta = (\theta_1, \ldots, \theta_m)$ *such that* $r_i = r_k = \theta_x$ *if both products i and k are displayed on page x. Moreover, the page-dependent parameters monotonically increase with the page indices, i.e.,* $\theta_1 < \theta_2 < \ldots < \theta_m$.

This result extends the classical optimal assortment-pricing structure under the MNL model. It is known that the optimal prices under the MNL profit-optimization problem have a constant markup (see for example, Anderson, de Palma and Thisse (1992), Hopp and Xu (2005) and Gallego and Stefanescu (2011)). We are considering the revenue-maximization problem, and therefore, we set the costs to zero, treating them as sunk costs. We obtain an analogous result that the prices should be equal for products that are displayed on the same page. We can easily extend the result to the profit maximization problem by treating a_i as $a_i - \beta c_i$, then we will see the markups should be page-dependent.

Therefore, we can reduce the decision variables from n prices, one for each item, to m prices, one for each page. The resulting problem in terms of θ_x, $x \in M$, is given by

$$R^{\text{pricing}} = \max_{\theta} R(\theta|S)$$

where

$$R(\theta|S) \equiv \sum_{x=1}^{m} \lambda(x) R_x(\theta|\mathcal{S}_x) \equiv \sum_{x=1}^{m} \lambda(x) \frac{\sum_{j=1}^{x} \theta_j \sum_{i \in S_j} \exp(a_i - \beta\theta_j)}{1 + \sum_{j=1}^{x} \sum_{i \in S_j} \exp(a_i - \beta\theta_j)}.$$

The first order condition yields the system of equations

$$\frac{\partial R(\theta|S)}{\partial \theta_x} = \sum_{j=x}^{m} \lambda(j)\pi(S_x, \mathcal{S}_j)(1 - \beta\theta_x + \beta R_j(\theta|\mathcal{S}_j)) = 0 \quad \forall x \in M,$$

which are equivalent to

$$\theta_x = \frac{1}{\beta} + \frac{\sum_{j=x}^{m} \lambda(j)\pi(S_x, \mathcal{S}_j) R_j(\theta|\mathcal{S}_j)}{\sum_{j=x}^{m} \lambda(j)\pi(S_x, \mathcal{S}_j)}.$$

In other words, the (common) price for products listed on page t is a constant, plus the weighted average of the expected revenue from the consumers who consider these products.

In general, the expected revenue $R(\theta|S)$ may not be jointly concave in the θ vector.

Example 3.9.2. *Two products are to be displayed on two pages. Suppose that* $\pi(X = 1) = 56\%$ *and* $\pi(X = 2) = 44\%$*. Product one has quality* $a_1 = 4$ *and product two has* $a_2 = 2$*. The* *price sensitivity is equal to one. The graph below shows how the expected revenue changes with respect to the price. As we can see, the expected revenue is not jointly concave in the price vector.*

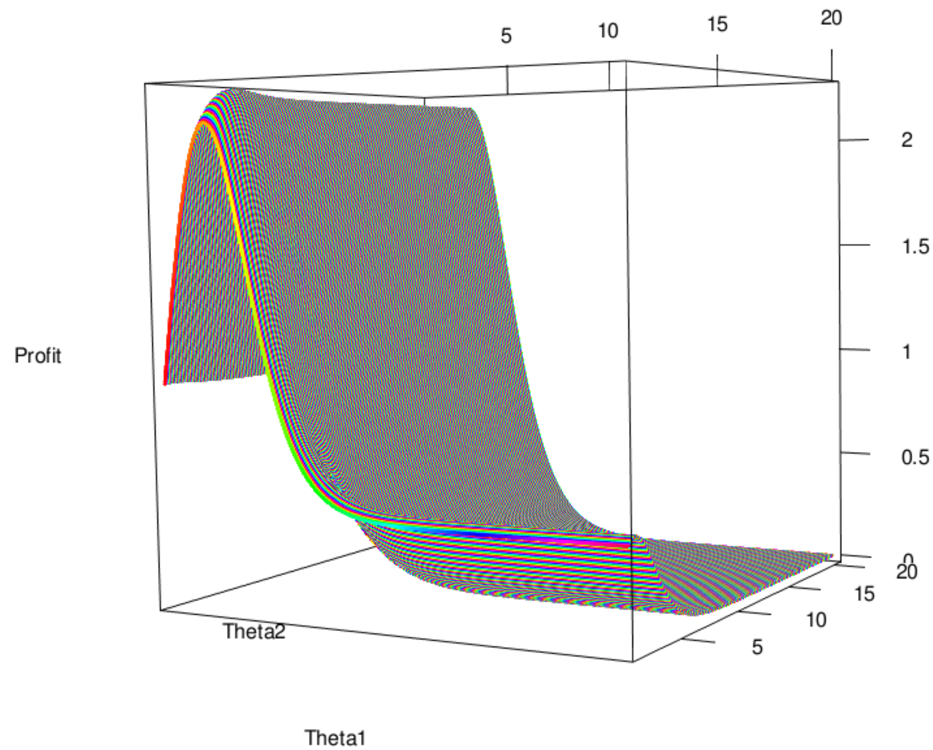

Figure 3.1: Pricing under Framing Effect

With the above optimal pricing structure, we now look into the problem of how to display the products. A related question is whether all products should be displayed. By the presumed utility structure, we see that products are differentiated by their quality parameters $a_i,\ i = 1,\ldots,n$. Without loss of generality, we will assume that the products are ordered in decreasing quality, so that $a_1 \geq a_2 \geq \ldots \geq a_n$.

Theorem 3.9.3. *Assume that products differ in their quality parameters a_i, $i = 1, \ldots, n$ such that $a_1 \geq a_2 \geq \ldots \geq a_n$. At optimality, each page will be filled with products until all products are displayed. The products are displayed in the order of their indices, so that higher quality products are displayed earlier.*

In essence, Theorems 3.9.1 and 3.9.3 tell us that higher-quality products should be displayed first and given lower prices. This result might appear unintuitive, and therefore requires some explanation. The prices are of the form $r_i = \theta_{x(i)}$, where $x(i)$ is the index of the page for product i, with $\theta_1 \ldots < \theta_m$. Thus products seen by fewer people have a higher price, with the lowest price enjoyed by products in page one, which are seen by everybody. Interestingly, the price θ_1, levied on the highest quality products that appear in the first page is higher than the optimal price, that would result if $\pi(X \leq 1) = 1$; the price θ_m, levied on the lowest quality products that appear in the last page is lower than the optimal price, that would result if $\pi(X \geq m) = 1$. In essence, consumers are penalized for not being willing to see all products and are compensated for being willing to see more products. Another way to understand the result is that lower-quality products are charged higher prices and serve to steer consumers to the higher-quality products. This result is in sharp contrast with the result for pricing in an oligopoly market, where several firms are deciding prices, see Arbatskaya (2007).

Corollary 3.9.4. *Under optimal pricing and frame, the expected revenue $R_x(r|\mathcal{S}_x)$ for a consumer who views x pages is increasing and concave with respect to x.*

This result is intuitive as the objective function is increasing in the number of products, and consumers who look at more pages also look at more products.

The optimal structure can also be extended to the profit-maximization problem. Interestingly, products now should be ordered by their value gaps, defined as the quality minus price sensitivity times price, i.e., the expected utility when products are priced at their costs, which is first introduced by Gallego, Li and Beltran (2016).

Corollary 3.9.5. *For the profit-maximization problem, at optimality, products are displayed in decreasing order of their value gaps, i.e.,* $\tilde{v}_i \equiv a_i - \beta c_i$*, and they are priced with page-level constant markups, i.e,* $r_i - c_i = \theta_{x(i)}$*, where the page-level markups should monotonically increase with the page indices. Therefore the products are displayed in decreasing order of their expected net utility as well.*

This result can be easily seen by replacing a_i with the value gap $a_i - \beta c_i$. For details, see Gallego, Li and Beltran (2016).

3.10 Computational Experiments

In this section, we numerically test the performance of the following framing heuristics:

- The NEST algorithm introduced in Section 3.6. Recall that NEST(y) leaves pages $x > y$ empty.
- The enhanced algorithm NEST+ introduced in Section 3.6.2, which is designed to improve practical performance.

- A heuristic, SORT_1, which sorts and displays products in increasing order of price.
- A heuristic, SORT_2, which sorts and displays products in decreasing order of attractiveness.
- A greedy heuristic, BOTTOM-UP (BU), which starts with the first page $(x = 1)$ and sequentially fills in products that would maximize the expected revenue for type x customers, $x = 2, \ldots, m$, such that this assortment includes the assortment for type-$x-1$ customers.
- A greedy heuristic, TOP-DOWN (TD) which starts by fitting all products into m pages. Then in the k-th step, $k = 1, \ldots, m-1$, the heuristic finds an assortment that maximizes revenue for type-$m-k$ customers, from the assortment for type $m-k+1$ customers.

3.10.1 Experimental Setup

We proceed by describing the instances being tested in our experiments. In all of our test problems, we assume that consumer choice is governed by the Multinomial Logit Model. The MNL associates the attractiveness $\{v_i : i \in N\}$ with the products. If the set of products $S(x)$ is displayed in the first x pages, then conditional on consumer type x's arrival, he will buy product $i \in S(x)$ with probability

$$\frac{v_i}{1 + \sum_{i' \in S(x)} v_{i'}}.$$

By convention, $v_i = e^{a_i - \beta r_i}$, where a_i is the product quality. In the tests, we independently draw every r_i from a uniform distribution over $[50, 100]$. We set $\beta = 1.02$ and $a_i = r_i + \varepsilon_i$, where $\varepsilon_i \in [-0.3, 0.3]$ is a noise added to the quality of product i.

For all test cases of the framing algorithms, we use $n = 300$ and $m = 20$. We test three different distributions of X: geometric, uniform and Poisson. We also differentiate the simulation scenarios by the expectation of X: Small ($E(X) = 2$), Median ($E(X) = 4$), and Large ($E(X) = 8$). We also vary the number p of products in each page.

For each test case, we simulate $1,000$ replicates and report the average gaps between the heuristics and the upper bound.

3.10.2 Experimental Result

Results of Framing Algorithms

Refer to Tables 3.1 to 3.9. In all scenarios and according to our metrics, $SORT_2$ outperforms $SORT_1$; TD, same as BU dominates SORT1 and SORT2; and unsurprisingly $NEST+$ outperforms $NEST$. $NEST+$ dominates all other heuristics $SORT_1$, $SORT_2$, TD and BU, with the average optimality gap of just 0.99% in the worst case, compared to 35.18% for $SORT_2$, 3.03% for TD and 3.83% for BU. The optimality gaps for $SORT_1$ and $SORT_2$ are relatively uniform across different scenarios, whereas the optimality gaps for $NEST$ and $NEST+$ tend to be larger for the geometric and exponential distributions, which we have shown to be worst-case distributions for these algorithms. When the page capacity and expected number of pages viewed by a consumer increase, NEST+, TD and BU tend to give same performance as they become optimal.

Table 3.1: Performance of NEST when $\mathbf{E}[X] = 2$ and X follows a geometric distribution.

	Avg gap						Avg number of displayed products			
p	NEST	NEST+	$SORT_1$	$SORT_2$	TD	BU	NEST	NEST+	TD	BU
1	12.22%	0.76%	25.12%	12.71%	1.06%	3.83%	2.1	20.0	20.0	20.0
3	7.43%	0.26%	31.49%	23.84%	0.43%	0.55%	6.0	60.0	60.0	60.0
9	12.23%	0.01%	37.52%	32.60%	0.01%	0.03%	9.0	83.0	83.0	83.0
15	8.02%	0.00%	38.73%	34.50%	0.00%	0.01%	15.0	83.2	83.2	83.2

Table 3.2: Performance of NEST when $\mathbf{E}[X] = 4$ and X follows a geometric distribution.

	Avg gap						Avg number of displayed products			
p	NEST	NEST+	$SORT_1$	$SORT_2$	TD	BU	NEST	NEST+	TD	BU
1	10.46%	0.99%	29.39%	20.12%	2.01%	1.68%	4.5	20.0	20.0	20.0
3	9.63%	0.16%	35.15%	29.20%	0.60%	0.24%	9.0	60.0	60.0	60.0
9	7.34%	0.01%	37.92%	33.77%	0.01%	0.01%	18.0	83.1	83.1	83.1
15	11.07%	0.00%	37.12%	33.60%	0.00%	0.00%	18.0	83.1	83.1	83.1

Table 3.3: Performance of NEST when $\mathbf{E}[X] = 8$ and X follows a geometric distribution.

	Avg gap						Avg number of displayed products			
p	NEST	NEST+	$SORT_1$	$SORT_2$	TD	BU	NEST	NEST+	TD	BU
1	10.49%	0.56%	33.28%	26.17%	2.44%	0.74%	7.0	20.0	20.0	20.0
3	9.81%	0.08%	37.32%	32.34%	0.56%	0.11%	13.2	60.0	60.0	60.0
9	5.42%	0.00%	36.26%	32.87%	0.01%	0.01%	27.0	83.2	83.2	83.2
15	5.38%	0.00%	33.20%	30.58%	0.00%	0.00%	30.0	83.0	83.0	83.0

Picky customers

We investigate in this section whether our algorithms continue to make good decisions in settings in which $v_0(x)$ increases linearly or exponentially in x. Specifically, in the linear

Table 3.4: Performance of NEST when $\mathbf{E}[X] = 2$ and X follows a uniform distribution.

	Avg gap						Avg number of displayed products			
p	NEST	NEST+	$SORT_1$	$SORT_2$	TD	BU	NEST	NEST+	TD	BU
1	0.79%	0.79%	24.43%	12.13%	0.88%	3.34%	3.0	3.0	3.0	20.0
3	6.24%	0.17%	32.03%	24.77%	0.53%	0.40%	6.0	9.0	9.0	60.0
9	2.66%	0.01%	38.20%	33.45%	0.02%	0.02%	18.0	27.0	27.0	83.2
15	10.01%	0.00%	39.32%	35.09%	0.00%	0.00%	15.0	45.0	45.0	83.3

Table 3.5: Performance of NEST when $\mathbf{E}[X] = 4$ and X follows a uniform distribution.

	Avg gap						Avg number of displayed products			
p	NEST	NEST+	$SORT_1$	$SORT_2$	TD	BU	NEST	NEST+	TD	BU
1	4.90%	0.64%	29.42%	20.48%	2.26%	1.17%	5.0	7.0	7.0	20.0
3	4.87%	0.10%	36.25%	30.69%	0.67%	0.15%	12.0	21.0	21.0	60.0
9	8.97%	0.01%	38.91%	34.79%	0.02%	0.01%	18.0	63.0	63.0	82.9
15	4.85%	0.00%	37.70%	34.24%	0.00%	0.00%	30.0	83.1	83.1	83.1

case, we set $v_0(x) = 1 + (x-1)\beta$, for page index $x \in \{1, 2, ..., m\}$. In the exponential case, we set $v_0(x) = e^{(x-1)\beta}$, for page index $x \in \{1, 2, ..., m\}$. When computing NEST(y), we set $v_0 = v_0(y)$. We explore the range of values $[0.2, 2]$ for β in the linear case and $[0.1, 1]$ in the exponential case.

Since $NEST+$ inherits the structure of both $NEST$ and $TRUNC$, and is optimized for performance, we have not added separate computational experiments for $TRUNC$.

Tables 3.10 and 3.11 display the results in the linear and exponential case, respectively. The results show that our leading algorithm, namely $NEST+$ continues to dominate all heuristics by substantial amounts in nearly all cases, except for a small difference in one case, where β is very small. Thus, these experiments indicate that our model serves as a good approximation of the more complex setting where the choice model may change dras-

Table 3.6: Performance of NEST when $\mathbf{E}[X] = 8$ and X follows a uniform distribution.

	Avg gap						Avg number of displayed products			
p	NEST	NEST+	SORT_1	SORT_2	TD	BU	NEST	NEST+	TD	BU
1	7.26%	0.32%	34.08%	27.63%	2.75%	0.42%	8.0	15.0	15.0	20.0
3	6.14%	0.05%	38.36%	33.72%	0.60%	0.06%	18.0	45.0	45.0	60.0
9	3.69%	0.00%	36.50%	33.31%	0.01%	0.00%	35.4	83.0	83.0	83.0
15	2.24%	0.00%	32.63%	30.25%	0.00%	0.00%	45.0	83.0	83.0	83.0

Table 3.7: Performance of NEST when $\mathbf{E}[X] = 2$ and X follows a Poisson distribution.

	Avg gap						Avg number of displayed products			
p	NEST	NEST+	SORT_1	SORT_2	TD	BU	NEST	NEST+	TD	BU
1	10.11%	0.49%	24.84%	12.24%	0.94%	3.53%	2.0	17.1	17.5	20.0
3	6.19%	0.18%	31.91%	24.48%	0.49%	0.45%	6.0	48.4	48.4	60.0
9	14.09%	0.02%	38.02%	33.21%	0.02%	0.02%	9.0	83.1	83.1	83.1
15	9.51%	0.01%	39.16%	34.96%	0.01%	0.00%	15.0	83.1	83.1	83.1

tically with the customer type. Thus, our model is a good starting point for an investigation of framing decisions involving heterogeneous customers.

3.11 Conclusion and Future Work

In this work, we propose one of the first models of "framing effects" for pricing and assortment optimization. We introduce a model in which a set of products must be organized sequentially into a set of virtual pages and priced appropriately. Each consumers will only consider a random number of pages, and will select an item, if any, from these pages following a general choice model. We show that this product-framing problem is NP-hard. We derive algorithms with guaranteed relative performance. Our algorithms are fast and easy to implement. We also show new structural results for pricing under framing effects. Direc-

Table 3.8: Performance of NEST when $\mathbf{E}[X] = 4$ and X follows a Poisson distribution.

	Avg gap						Avg number of displayed products			
p	NEST	NEST+	SORT$_1$	SORT$_2$	TD	BU	NEST	NEST+	TD	BU
1	7.69%	0.34%	28.91%	20.13%	2.21%	0.94%	4.0	20.0	20.0	20.0
3	8.91%	0.04%	36.44%	31.08%	0.74%	0.08%	9.0	60.0	60.0	60.0
9	3.46%	0.00%	39.27%	35.18%	0.02%	0.00%	27.0	82.9	83.0	82.9
15	5.09%	0.00%	38.13%	34.69%	0.00%	0.00%	30.0	82.9	82.9	82.9

Table 3.9: Performance of NEST when $\mathbf{E}[X] = 8$ and X follows a Poisson distribution.

	Avg gap						Avg number of displayed products			
p	NEST	NEST+	SORT$_1$	SORT$_2$	TD	BU	NEST	NEST+	TD	BU
1	4.32%	0.05%	34.15%	27.99%	3.03%	0.13%	7.9	20.0	20.0	20.0
3	5.08%	0.01%	39.25%	34.86%	0.65%	0.01%	18.0	60.0	60.0	60.0
9	2.03%	0.00%	37.41%	34.24%	0.01%	0.00%	45.0	83.1	83.1	83.1
15	0.76%	0.00%	33.01%	30.79%	0.00%	0.00%	60.0	82.9	82.9	82.9

tions for future research include to endogenize the number of pages consumers are willing to see. In the context of dynamic search, it would be convenient to allow for correlations between X and the choice model.

Table 3.10: Performance of NEST and other heuristics. $\mathbf{E}[X] = 4$; p = 3; X follows a Poisson distribution; $v_0(i) = 1 + (i-1)\beta$, where i is the page index.

	Avg gap					Max gap				
β	NEST+	SORT_1	SORT_2	BU	TD	NEST+	SORT_1	SORT_2	BU	TD
0.2	0.02%	32.40%	25.93%	0.74%	0.03%	0.10%	34.28%	28.85%	1.70%	0.14%
0.4	0.01%	29.65%	22.13%	0.60%	0.01%	0.04%	32.18%	24.50%	1.57%	0.07%
0.6	0.00%	27.44%	19.03%	0.57%	0.00%	0.02%	31.46%	21.40%	1.35%	0.02%
0.8	0.00%	26.05%	16.76%	0.51%	0.00%	0.00%	29.60%	19.32%	1.18%	0.00%
1	0.00%	24.81%	14.78%	0.45%	0.00%	0.00%	28.70%	17.01%	1.11%	0.00%
1.2	0.00%	24.05%	13.54%	0.39%	0.00%	0.00%	27.74%	15.84%	0.91%	0.00%
1.4	0.00%	22.84%	12.11%	0.35%	0.00%	0.02%	27.08%	15.24%	0.93%	0.02%
1.6	0.00%	22.61%	11.35%	0.29%	0.00%	0.06%	27.84%	13.48%	0.70%	0.06%
1.8	0.01%	21.63%	10.48%	0.26%	0.01%	0.11%	27.07%	13.35%	0.62%	0.11%
2	0.02%	21.52%	9.92%	0.26%	0.02%	0.08%	27.85%	14.01%	0.66%	0.08%

Table 3.11: Performance of NEST and other heuristics. $\mathbf{E}[X] = 4$; p = 3; X follows a Poisson distribution; $v_0(i) = e^{(i-1)\beta}$, where i is the page index.

	Avg gap					Max gap				
β	NEST+	SORT_1	SORT_2	BU	TD	NEST+	SORT_1	SORT_2	BU	TD
0.1	0.02%	33.77%	27.79%	0.61%	0.03%	0.09%	36.16%	29.81%	1.78%	0.09%
0.2	0.02%	30.95%	24.06%	0.73%	0.49%	0.10%	34.73%	26.15%	1.73%	1.27%
0.3	0.01%	28.17%	19.98%	0.76%	2.30%	0.08%	30.48%	21.91%	1.46%	4.55%
0.4	0.03%	25.99%	16.69%	0.72%	2.37%	0.10%	30.66%	19.13%	1.37%	4.71%
0.5	0.09%	24.22%	14.08%	0.69%	2.03%	0.23%	29.65%	15.82%	1.49%	4.66%
0.6	0.17%	23.37%	12.19%	0.59%	1.79%	0.38%	29.12%	13.60%	1.47%	4.48%
0.7	0.25%	22.53%	10.72%	0.58%	1.74%	0.53%	27.61%	12.74%	1.31%	4.29%
0.8	0.27%	21.82%	9.81%	0.60%	1.53%	0.61%	27.22%	13.05%	1.35%	3.83%
0.9	0.31%	21.46%	9.09%	0.65%	1.47%	0.76%	28.61%	11.12%	1.24%	4.29%
1	0.40%	21.76%	8.53%	0.60%	1.51%	1.07%	28.15%	10.81%	1.36%	3.94%

Chapter 4

Product Line Design and Pricing under Logit Models

Assortment optimization and pricing are getting more attention in business for its potential to increase profits. It is also attracting the attention of people in academia interested in looking at this problem with new ideas and new tools. Most of the current research solves the problem at the stock-keeping units (SKUs) level, assuming the set of potential products is given. However, the product manager cares not only about improving over the existing products; but also discovering the potentially lucrative new products. Yet this cannot be analyzed under current SKUs based assortment models due to the ambiguity of the potential product set. To disentangle the new product introduction and pricing problem, we need to: First, construct a choice model which is able to forecast demand of current and new products; Second, jointly search for assortments and prices to maximize the profit.

Each SKU is actually a unique combination of a set of discrete and tangible attributes, which may have several levels, and consumers typically view a product based on its attributes. To forecast a new product's demand, it is important to understand how consumers integrate the attributes' valuation into the product's. For example, customer A buys a 17 inches HP laptop with core i7 processor at $1500, but he may alter his choice to the 15 inches Mac book if HP's price is increased by 5%. We are looking for a powerful but parsimonious choice model, to calibrate consumer choice based on attributes and prices, such that the joint assortment and pricing problem is tractable.

There is a long history of research papers modeling a multi-attribute SKUs choice behavior based on consumers' preference over attribute levels. As is written in Fader and Hardie (1996): "The theoretical justification for such an approach can be found in psychology (Fishbein 1967) and economics (Lancaster 1971; Quandt 1956). Moreover, in both the theoretical (e.g., Anderson, de Palma, and Thisse 1992) and empirical (e.g., Berry 1994) literature on discrete choice models of product differentiation, it is common to posit that a consumer's utility for a product is not a direct function of the product itself, but instead depends on the product's attributes and the consumer's tastes."

The Multinomial Logit (MNL) model has attracted tremendous theoretical and empirical interests. It was first proposed by McFadden (1974), derived from a random utility model where the random components are independent Gumbel random variables. The MNL model can characterize many choice situation conveniently, since it can provide a convenient closed form for the underlying choice probability without any integration. Although criticized because of the independence of irrelevant alternatives (IIA) property (Luce 1959), it has remained a maintained assumption in many applications. The Nested Logit (NL) model, introduced by Williams (1977), has been developed to relax the assumption of independence between all the alternatives. Under the Nested Logit model, customers first select a nest, and then, a product within the selected nest, thus NL allows different substitution patterns within and between nests. In this work we model consumers' demand by a Logit model, characterizing utility on each SKU by the integrated attribute-level valuation. The valuation of a new product, even if it never appears on the market before, can be easily calculated. Therefore, given an assortment, we can forecast the market share of any product.

Assortment optimization intends to provide insight into structural properties of optimal assortments and is an active area of research. The stylized model research began with a pioneering paper by van Ryzin and Mahajan (1999). They show that the optimal assortment consists of a certain number of the highest utility products when they are equally profitable under the MNL consumer choice model. Talluri and van Ryzin (2004) show that under MNL choice model the optimal assortment includes a certain number of products with the largest revenues when the products' prices are given exogenously. Davis et al. (2014) show that the assortment problem under NL model can be solved by a linear program when the nest dissimilarity parameters of the choice model are less than one and the customers always make a purchase within the selected nest. Relaxing either of these assumptions renders the problem NP-hard. The problem becomes more complicated when more general choice models are considered.

For the assortment pricing problem, Hanson and Martin (1996) notice that the expected revenue function fails to be concave in prices for the multinomial logit model. Song and Xue (2007) and Dong et al. (2009) make significant progress by formulating the expected revenue as a concave function with respect to market shares; Li and Huh (2011) extend the concavity result to the Nested Logit model by assuming that the price sensitivities of the products are constant within each nest and the nest dissimilarity parameters are restricted to be in the unit interval. Derived from first order condition, the markup, defined as price minus cost, is constant across all the products of the firm at optimality under MNL model with uniform price-sensitivity parameter (Anderson and de Palma 1992, Hopp and Xu 2005, and Gallego and Stefanescu 2011). Gallego and Wang (2014) consider the general NL model with product-differentiated price-sensitivity parameters and arbitrary nest coefficients. They find that the adjusted nest-level markup is also constant across all the nests. Li and Huh (2011) and Gallego and Wang (2014) assume that the customers always make a purchase within the selected nest, which cannot incorporate the type-primary choice process in Kok and Xu (2011). Kok and Xu (2011) work with two nest structures: brand-primary choice model where customers first select a brand, and then a product type within the selected brand; Type-primary choice model where customers first select a product type, and then a brand for the selected product type. The authors characterize the structure of the optimal solution, but they do not provide an explicit result under the type-primary choice model or when the number of brands is large.

There are two mainstreams in product line design. One is referred as product line selection problem, which consists of two stages. In the first stage, a set of potentially optimal products is generated; In the second stage, an optimal product line is selected from the predetermined reference set. The other stream referred as one-step product line design constructs an optimal product line directly by assigning levels to product attributes. Most of the literatures on the product line selection problem focus on the second stage (e.g. Chen and Hausman 2000, Schon 2010), which essentially is an assortment problem and does not provide an instruction on how to obtain a reference set. We can generate the reference set by exhausting the combinations of attributes and levels, but the set's size will explode unmanageably. For example, for the laptop industry, there will be over a billion potential products if we fully enumerate the combination of attributes and levels. So we refer to the product line design problem as the one-step design case. Zufryden(1977, 1982) was the first to consider product-line design problems from idiosyncratic part-worth preference functions estimated using conjoint analysis. Assuming a deterministic first choice model, he formulates the product line design problem as an integer program to maximize the market share. His formulation is NP-hard. Kohli and Sukumar (1990) extend this line of research by proposing a heuristic algorithm from the dynamic programming perspective. Belloni et al. (2008) compare the performance of different heuristics for product line design and find that the greedy heuristic performs extremely well. Instead of using conjoint analysis, Chong et al. (2001) present consumer's preference by a choice model based on the Guadagni and

Little (1983) brand-share model and propose a local improvement heuristic algorithm to design the product line. Fisher and Vaidyanathan (2014) apply several heuristic algorithms for product line design under their attribute-level substitution based choice model. Luo (2011) models consumers choice by a random coefficient multinomial logit model, and instead of focusing on discovering structural properties of product configuration, she presents a model that simultaneously considers both marketing and engineering factors.

We optimize the product line on the attributes space under the MNL model. Each attribute may have different levels of quality and different unit wholesale costs. A product is defined by the level of its attributes. Thus, a two attribute product may have attribute one at level one and attribute two at level three. A key to designing products that are both profitable and have high value for consumers is to look for attribute-level combinations with high value gap. The value gap of an attribute-level combination is defined as the net utility to the customer when the attribute-level is priced at wholesale cost. Based on this notion, we can use a greedy approach to solving the problem of finding the K-best assortments. This beautiful structure makes a greedy algorithm, derived from the K-shortest paths algorithm (Yen 1971), able to find an optimal K products' configuration in polynomial time. This structural result also applies to the NL model with respect to each nest, meaning that given the product line length constraint for each nest, we can find an optimal configuration in polynomial time by running K-shortest paths algorithm nest-wise.

Conditional on the optimal assortment, we optimize over prices and find that the constant mark-up property still applies for MNL model, meaning every product should be priced at its cost plus a constant, which is invariant across all the SKUs. Davis et al. (2014) show the pure assortment problem is NP-hard under the type-primary NL model, while we prove the problem is polynomially solvable if we can jointly optimize the assortment and prices. Under our general NL model, which incorporates both the brand-primary choice process and type-primary choice process, the markup is constant for all the products within each nest at optimality. In addition, there is an optimal adjusted nest-level markup invariant across all the nests, reducing the price optimization problem to a single-dimensional maximization of a continuous function over a bounded interval. Moreover, under both choice models, the profit function is unimodal with respect to the single dimension parameter.

The remainder of this chapter is organized as follows. In section 4.1, we first show the constant markup property of the MNL pricing problem, and then the structural properties of optimal products configuration. We present a K-shortest paths algorithm which can solve the optimal product line design problem in polynomial time. We bound the loss due to the line length constraint in the extension. In section 4.2, we prove that the nest level constant markup property holds for NL model and the structural properties of optimal configuration also apply to NL model within each nest, which will guarantee that the polynomial algorithm still works for the NL product design problem.

4.1 Product Design and Pricing under Multinomial Logit Demand

The MNL model has attracted tremendous theoretical and empirical interests. It was first proposed by McFadden (1974), derived from a random utility model where the random components are independent Gumbel random variables. The MNL model can characterize many choice situation conveniently, since it can provide a convenient closed form for the underlying choice probability without any integration. Although criticized because of the independence of irrelevant alternatives (IIA) property, it has remained a maintained assumption in many applications.

4.1.1 Multinomial Logit Model

The standard MNL model has the following structure:

$$U_i = v_i + \varepsilon_i$$

where U_i is the random utility of product i, which is equal to the sum of the expected utility, denoted by v_i, and the ε_i are independent standard Gumbel random variables.

We model the expected utility of a product as a linear function of the attributes that define the product. By the attributes we mean the product features and the price, e.g., CPU, RAM etc. for a laptop. Each attribute can have several levels, so the attribute-level combination will uniquely define a product. In particular, if x_i is the vector that defines product i, then

$$v_i = \beta_0 + \beta' x_i - \beta_r r_i \tag{4.1.1}$$

where β_0 is a constant, a common valuation across all products, probably used for normalization; β is the utility vector corresponding to all the attribute-level, e.g., the k-th component of vector β, β_k represents attribute-level k's utility; so basically we represent a product by a configuration vector, where the k-th component of x_i, $x_{ik} = 1$ if product i has attribute-level k as one component and $x_{ik} = 0$ otherwise; β_r is the price sensitivity parameter; r_i represents i's price. The term $\beta_0 + \beta' x_i$ is the expected utility of the product at price zero. Consequently, the expected utility is the product expected quality minus the product of the price and the sensitivity to price.

Under this utility structure, the probability of an incoming customer choosing product i from assortment S is

$$\pi_i(S) = \frac{V_i}{V_0 + \sum_{j \in S} V_j} \tag{4.1.2}$$

While V_j is the attractiveness of product j, i.e., $V_j = exp(v_j)$; $V_0 = exp(v_0)$ is the attractiveness of the outside alternative.

Now we can see the advantage of this formulation: when a new product is introduced, even if it never exists in the market before, as long as its configuration vector x_{new} is given, we can calculate its valuation by the equation $v_{new} = \beta_0 + \beta' x_{new} - \beta_r r_{new}$. Therefore its market share can be computed under any given assortment.

4.1.2 Product Design and Optimal Pricing

Problem Formulation

Assume there is a manufacturer who designs, prices and sells products. His goal is to improve his profits by discontinuing some existing products, introducing some new products, and adjusting the prices.

Here we show how the manufacturer can design products to maximize his expected profit assuming that customers choose according to the MNL model described earlier and assuming that the cost of the products is equal to the sum of costs of the attributes included in the product. We will also assume, without loss of generality that the market size is one, and that the manufacturer wants to design a portfolio of products of size no more than K.

$$\max_S \quad \max_r \quad R(S,p) \equiv \sum_{i\in S}(r_i - c'x_i)\pi_i(S) \tag{4.1.3}$$

$$\text{s.t.} \quad |S| \leq K \tag{4.1.4}$$

$$x_i \cdot x_j \leq |F| - 1, \forall i \in S \tag{4.1.5}$$

$$Ax_i = e, \forall i \in S \tag{4.1.6}$$

$$x_i \in \{0,1\}^{|FL|}, \forall i \in S \tag{4.1.7}$$

The manufacturer has control over both the set of products, or assortment, $S := \{1,2,...,K\}$, where each product i is defined by its configuration vector x_i, and the price vector r for the products in the assortment. Each product in the assortment S is determined by its configuration vector x, where $x_{il} = 1$ if product i has attribute-level l as a component and $x_{il} = 0$ otherwise. The vector c is the attribute-level wholesale price vector, and the lth component c_l represents attribute-level l's cost. Assuming that the unit production cost is equal to the sum of the wholesale prices of its components, the term $r_i - c'x_i$ is the profit of selling one unit of product i, and $\pi_i(S)$ defined in (4.1.2) is the probability of selling product i under assortment S and price vector r.

We now describe the constraints. Constraint (4.1.4), is a cardinality constraint telling us that we allow at most K products. Constraint (4.1.5), $x_i \cdot x_j \leq |F| - 1$ guarantees that any two products in S cannot have the same configuration, where F is the set of attributes; as each attribute needs to assign one level, $x_i \cdot x_j = |F|$ indicates i and j have same configuration. Constraint (4.1.6) repeatedly appears in marketing and operations research literatures of product design (e.g. Zufryden 1982, Kohli and Sukumar 1990, Fisher and Vaidyanathan 2014). A is a configuration constraint matrix, and it imposes that some attribute levels are mutually exclusive but each product must contain exactly one level for each attribute. Each row of A represents an attribute, and the columns the levels. For instance, for the laptop, row m can correspond to attribute RAM, and $A_{m,l} = 1$ if l is one of the levels of RAM, and $A_{m,l} = 0$ otherwise. The constraint $\sum_l A_{m,l} x_{i,l} = 1$ implies that one and only one level of the RAM should be part of the configuration of product i. This constraint has flexibility in its nature, for example, it can also incorporate the case that some attribute can be omitted in the product by introducing one level representing the decision of not having the attribute. As an example, in configuring a laptop we may opt for not having a DVD reader. The last constraint (4.1.7) requires a binary configuration so for each attribute level the variable can be either zero or one. Here $|FL|$ is the total number of attribute levels, i.e., the length of configuration vector. This is a Mixed Integer Nonlinear Optimization problem, that we now analyze in depth.

Optimal Pricing

Assuming for the moment that the set of products is fixed, and that our only challenge is to price the products. The optimization problem reduces to

$$\max_{r} \sum_{i \in S} (r_i - \tilde{c}_i) \frac{exp(\alpha_i - \beta_r r_i)}{V_0 + \sum_{j \in S} exp(\alpha_j - \beta_r r_j)}, \tag{4.1.8}$$

where S is the set of products to be offered, and for each $i \in S$, $\tilde{c}_i = c' x_i$ is the product's cost which is equal to the components' costs sum and $\alpha_i = \beta_0 + \beta' x_i$ is the product's quality. Driving some of the products' prices to positive infinity is equivalent to pushing those products out of market, so optimizing over price also enables us to refine the assortment. Therefore the question is: should we discontinue some products and will there be some nice structure on the optimal prices?

Theorem 4.1.1 (Constant Markup)**.** *Assuming the set of products S is fixed, optimal prices are of the form $r_i^* = \tilde{c}_i + \theta_S^*$ for all $i \in S$. Moreover, θ_S^* is increasing in assortment S and is the maximizer of the profit function $R(S|\theta) = \theta \sum_{i \in S} \pi_i(S|\theta)$, where $\pi_i(S|\theta)$ is the probability* *that product $i \in S$ is selected given markup θ. In addition, $R(S|\theta)$ is unimodular on θ and θ_S^* is the root of the equation $\beta_r \theta \pi_0(S|\theta) = 1$, where $\pi_0(S|\theta)$ is the no purchase probability. This result is due to Gallego and Wang (2014).*

Optimal Product Design

Optimal Products Configuration

Armed with a method to find optimal prices for a given assortment, we now concentrate on the product line configuration problem. A brute force method would call for exhausting all of the possible assortments of size K and to compute the profit $R(S|\theta_S^*)$ for each such assortment and pick the best, where θ_S^* is the optimal markup under assortment S. However, the number of products and assortments to consider grows very quickly as we add attributes and levels. Our goal here is to find a systematic way of designing products that are profitable and highly valued by consumers. To do this we define the concept of value gaps.

Definition 4.1.2 (Value Gap Index). *The value gap of attribute-level k is defined as $g_k = \beta_k - \beta_r c_k$. This is the difference between the part worth of attribute-level k and the cost disutility $\beta_r c_k$. We define the value gap of product i as*

$$\tilde{v}_i = \beta_0 + g' x_i = \beta_0 + \beta' x_i - \beta_r \tilde{c}_i.$$

Notice that $v_i = \tilde{v}_i$ when $r_i = \tilde{c}_i = c' x_i$, i.e., when the price of product i is equal to its cost. Since $r_i \geq \tilde{c}_i$, it follows that $v_i \leq \tilde{v}_i$ for all i, so $\tilde{v}_i$ is an upper bound on v_i.

Theorem 4.1.3 (Highest Value Gap Index Property)**.** *Assume that the product configurations x_j, for all $j \in S$ are fixed except for attribute m of product $i \in S$. Then it is optimal to select level l available for attribute m with the highest value gap r_l. More precisely, l is selected as the attribute-level with highest g_l among all available attribute levels q such that $A_{m,q} = 1$.*

The idea behind is that we can always improve profits by switching to the level with the highest value gap.

Although this theorem gives us some sense of how to configure the last attribute if everything else is fixed, it is still unclear how to design the rest of the products given the interaction effects, since selecting attribute-level l for product i may preclude l from product j.

Theorem 4.1.4 (Principle of Priority)**.** *Products with higher value gaps are prioritized for picking attribute levels in the product line design.*

The intuition is that under the optimal pricing, adding an attribute-level with positive value gap to a product with higher value gap will always generate a higher profit than that from adding it to a product with lower value gap.

This is consistent with the majorization results in van Ryzin and Mahajan (1999). Knowing the optimal structure, given a set of attribute levels, there will be a polynomial algorithm to find K optimal products.

Algorithm

To describe the algorithm we will use the following notation:

- Let f, m represent attributes, and F is the set of attributes
- Let l, q be the indices for attribute-levels, and let FL be the set of all attribute levels
- Let $F(m)$ be the set of all attribute levels of attribute m
- Let F^{-1} map attribute-level to the attribute it belongs to, i.e., F^{-1}:$FL \to F$ maps each $l \in FL$ to $m \in F$ st. $F^{-1}(l) = m$ if and only if $l \in F(m)$

Then construct a graph with key characteristics as the following:

1. There is one node s representing the source;
2. There is one node t representing the sink;
3. There are $|FL|$ additional nodes, each representing an attribute-level $l \in FL$;
4. Order the attributes, i.e., assign them index $m = 1, \ldots, |F|$. (We can see later that the way how the attributes are ordered does not make a difference, since this is only to differentiate the attributes);
5. There are arcs connecting the source to every $l \in F(1)$ with cost $c_{sl} = \beta_0$, i.e., these are directed arcs from the source to every level of attribute one with cost β_0;

6. There are arcs connecting every $l \in F(|F|)$ to the sink t with cost $c_{lt} = g_l$, i.e., there are directed arcs from every level of the last attribute to the sink with cost c_{lt} equal to the value gap g_l;

7. $\forall l, q \in FL$ such that $F^{-1}(l) + 1 = F^{-1}(q)$, indicating attribute-level l and q belong to two consecutive attributes, then there is one directed arc originating from l connecting them, with cost of l's value gap $c_{lq} = g_l$.

An interpretation of the graph is that: each layer corresponds to an attribute; any node in a layer is one level of the corresponding attribute; so every path from s to t will pass by one and only one level of each attribute and then will uniquely define a product configuration. Since every arc's originating node defines a component and the cost of that arc is defined by the originating attribute level's value gap, cost of the path is the total value gap of the configured product returned by the path.

From our analysis in the last section, it follows that the first best product is the one with highest value gap, and shown on the graph it is the longest path; the second best product will differ from a single optimal product by one attribute-level, which gives the second highest total value gap, and represented on the graph it is the second longest path. Inductively, designing K-optimal products is equivalent to finding K-longest paths on the graph. However, it is well known that the longest path problem is NP-hard. A nice property of our problem is that the graph is "layered", i.e., every feasible flow from the source to the sink has to first pass by one node in the layer one (attribute one), then one node in the layer two (attribute two), and so on. In another word, the graph in our formulation is a directed acyclic graph. Moreover, the longest path is the shortest path in the negative graph, in which the costs are replaced by their negative value. Due to the layered structure, no negative cycles can be created in the negative graph. Yen(1971) proposed an efficient algorithm to find K shortest loopless paths when there are no negative loops in the network, by making Kl calls to Dijkstra's algorithm (Dijkstra 1959), where l is length of the spur paths. In our problem l is no more than the length of each path, $|F|+1$. And the time complexity of Dijkstra's algorithm with $|FL|+2$ nodes is $O(|FL|^2)$. So the overall time complexity becomes $O(K|F||FL|^2)$.

Theorem 4.1.5 (K shortest paths algorithm for optimal product configuration)**.** *The optimal K product design can be solved in $O(K|F||FL|^2)$ by running the Yen's K shortest paths algorithm on the negative graph.*

Example 4.1.6. *Two attributes need to be configured for a product line, and there are three levels for each attribute, as is show in the table below, value in each cell is the value gap index for every attribute level:*

Table 4.1: Product Line Configuration Example

Value Gap	**Levels**		
	1	2	3
Attribute a	30	20	10
Attribute b	20	15	0

Let us construct the graph as below:

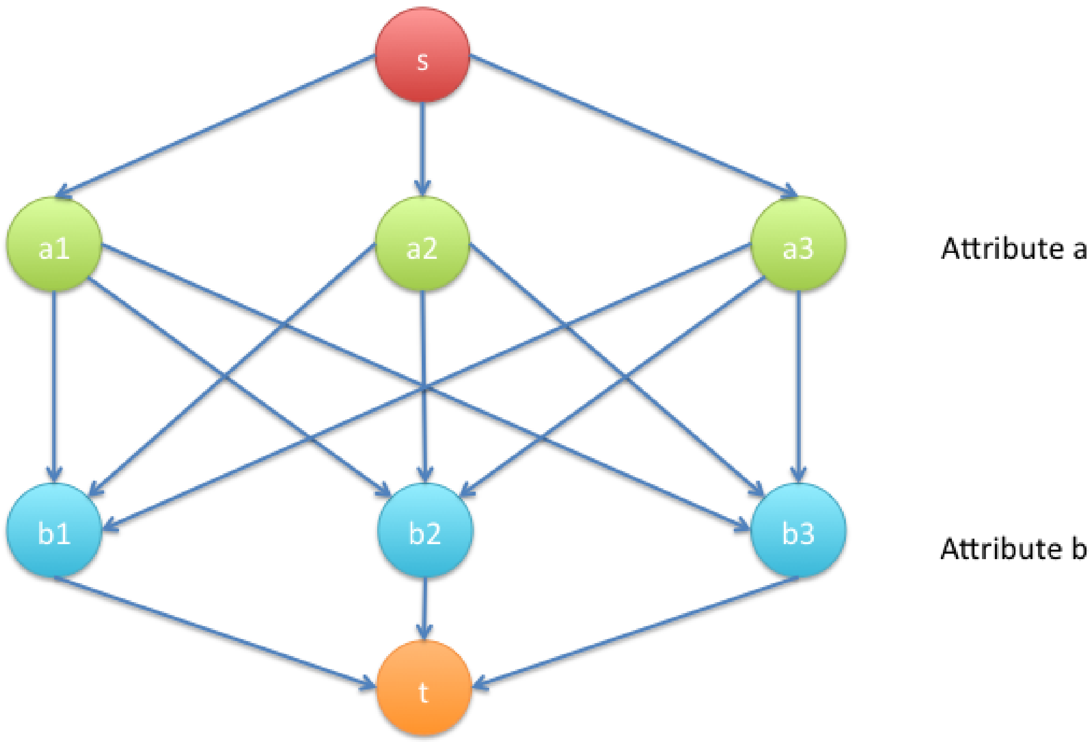

Figure 4.1: Network Flow for the Product Line Design Algorithm

We then negate the cost and run Yen's K shortest paths algorithm on the negative graph. The first shortest path $s \rightarrow a1 \rightarrow b1 \rightarrow t$ with total length -50 presents the best single product (with highest total value gap); the second shortest path $s \rightarrow a1 \rightarrow b2 \rightarrow t$ with total cost -45 indicates we get the second best product by replacing $b1$ in the best single product to $b2$. Finally, the optimal list of configuration (ordered by total value gap): $S^* = \{\{a1,b1\},\{a1,b2\},\{a2,b1\},\{a2,b2\},\{a3,b1\},\{a1,b3\},\{a3,b2\},\{a2,b3\},\{a3,b3\}\}$, and the optimal K products configuration is to select the first K set elements from S^*.

4.1.3 Extension: Loss Bound due to Line Length Constraint

We assume the product line length constraint is given exogenously. In the optimal configuration, the line length should reach its maximum since the optimal profit increase monotonically on the line length. In this section, we want to study the ratio bound of the loss caused by the cardinality constraint to the optimal profit under full product line length.

Proposition 4.1.7. *(Ratio Bound of Cardinality Constraint) Order products by their residual valuation $\tilde{v}$, the top K products in the production line can capture at least $\frac{\sum_{i=1}^{K} exp(\tilde{v}_i)}{\sum_{i=1}^{\bar{K}} exp(\tilde{v}_i)}$ of the optimal profit from a product line of bigger length $\bar{K}$.*

We use simulation to get some numerical bound. When the products' residual valuations are $[0,1]$ uniformly distributed, 15 out of 30 products can capture 95% of the total profit. When the residual valuations are $[0,3]$ uniformly distributed, 9 out of 30 products can capture 98% of the total profit. The higher the variation among the products' residual valuation is, the fewer products we need to capture a significant percentage of the total profit.

4.2 Product Design and Pricing under Nested Logit Demand

The Multinomial Logit model has many applications due to its ease of computation and the existence of a number of computer programs. Yet, the independence of irrelevant alternative (IIA) property (Luce 1959) is a serious limitation, meaning the ratio of the probabilities of choosing any two alternatives is independent of the attributes of any other alternative in the choice set. Debreu(1952) was among the first economists to discuss the implausibility of the independence from irrelevant alternatives assumption. The Nested Logit (NL) model, introduced by Williams (1977), has been developed to relax the assumption of independence between all the alternatives. Under the Nested Logit model, customers first select a nest, and then, an alternative within the selected nest, thus NL allows different substitution patterns within and between nests.

4.2.1 Nested Logit Model

In our study we consider two different hierarchical choice process illustrated in figure 4.2.1 (take the laptop industry as an example). Under the brand-primary model (Li and Huh 2011; Gallego and Wang 2014), a consumer first decides to leave without purchase or to buy from one brand. Once the brand is selected, the consumer choose one product from this brand. When consumers have strong brand loyalty and the functionalities among the attribute levels are not much differed, it is proper to use the brand-primary model. Type-primary model (Currim, Meyer, and Le 1988) assumes that consumers make hierarchical decision in the form of a decision tree (Tversky and Sattath 1979), where a sequence of decisions are made to screen the attribute levels. A nest is a set of correlated alternatives sharing similar attribute levels. In contrast, the type-primary model is suitable when consumers are more picky about the characteristics of the product than its brand and various attribute levels are aiming for different usage. Both brand-primary and type-primary Nested Logit fit data better than the MNL does; However, some empirical papers show that the type-primary structure gives more accurate estimate than does brand-primary model (Kannan and Wright 1991). We can also incorporate the attribute-incompatibility constraint into the type-primary model by representing the constraints by the branches.

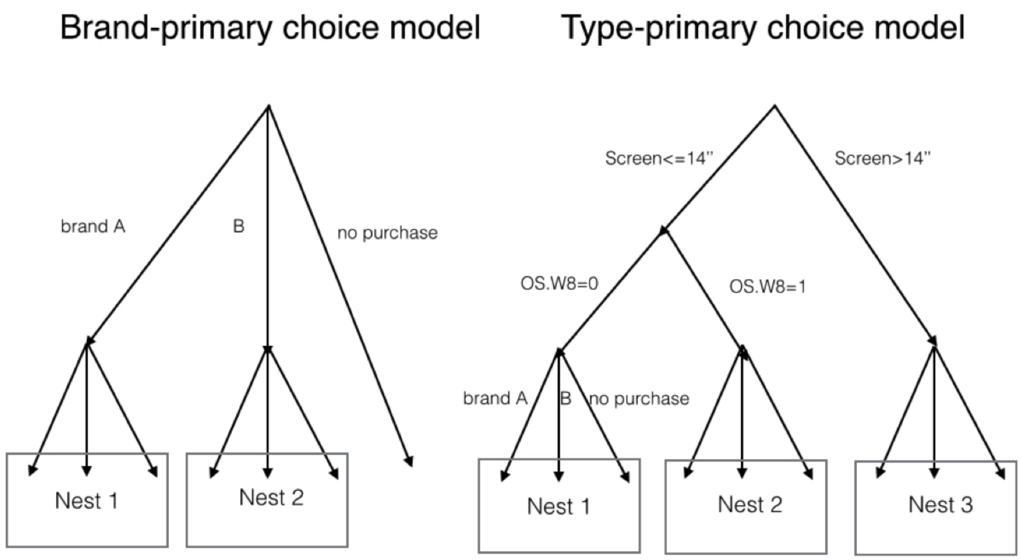

Figure 4.2: Brand-Primary and Type-Primary Hierarchical Consumer Choice Models

Given the nest structure, the conditional probability of an incoming consumer selecting product i from nest n under assortment S (let us call it product in for short) follows the Logit form

$$\pi(in|n,S) = \frac{exp(\frac{\alpha_{in}-\beta_r^n r_{in}}{\gamma_n})}{\sum_{jn\in S_n} exp(\frac{\alpha_{jn}-\beta_r^n r_{jn}}{\gamma_n}) + V_n} \tag{4.2.1}$$

where S_n is the set of products from nest n; and the probability of selecting from nest n is

$$\pi(n|S) = \frac{(\sum_{jn \in S_n} exp(\frac{\alpha_{jn} - \beta_r^n r_{jn}}{\gamma_n}) + V_n)^{\gamma_n}}{\sum_{n'} (\sum_{jn' \in S_{n'}} exp(\frac{\alpha_{jn'} - \beta_p^{n'} r_{jn'}}{\gamma_{n'}}) + V_{n'})^{\gamma_{n'}} + V_0} \tag{4.2.2}$$

The final probability of purchasing product i from nest n will be the multiplication of the two probabilities

$$\pi(in|S) = \pi(in|n,S)\pi(n|S) \tag{4.2.3}$$

Consumers' valuations of two products from different nests are independent; and γ is a measure of the relative independence of the alternatives within the same submarket. To be compatible with the random utility maximization theory, we restrict the γ coefficients lie within the unit interval (Williams 1977, McFadden 1980), indicating the valuations for two products within a same group are positively correlated. When γ equals 1, it reduces to the Multinomial Logit model. α_{in} is the expected quality valuation of product in in nest n. Applying the linear utility structure, we have $\alpha_{in} = \beta_0^n + \beta^n \cdot x_{in}$, where β_0^n is nest n's inherent valuation and β^n is the attribute-level valuation vector for nest n, i.e., the kth component $(\beta^n)_k$ is attribute-level k's valuation in nest n. Here we assume the attribute-level valuation is homogeneous within a nest and heterogeneous across nests. x_{in} is the product in's configuration vector, still $x_{in}^k = 1$ if product in has attribute-level k as a component and it is 0 otherwise. β_r^n is the price sensitivity parameter for nest n. r_{in} represents product in's price. Product in's determinant valuation v_{in} still follows $v_{in} = \alpha_{in} - \beta_r^n r_{in}$, the product quality valuation minus the price disutility.

V_0 is the attractiveness of the outside alternative outside nests and without loss of generality, we normalize it to 1.

Our formulation is general in the sense that it can incorporate both the brand-primary choice process and the type-primary choice process. When two different branded products belong to two different sets and the attractiveness of the inner nest outside alternative V_n equals 0, it follows the brand-primary choice process, and is a common demand formulation for assortment pricing (Li and Huh 2011; Gallego and Wang 2014). When the nests are grouped by SKU's other attribute levels but brands, then the demand follows type-primary choice process; since within each type, there might be products offered by multiple firms, from a manufacturer's point of view there is outside alternative within each nest, i.e., V_n may be greater than zero. This formulation is similar to that in Kok and Xu (2011), but we formulate consumers' preference endogenously as the aggregated attribute-level valuation, and allow heterogeneity on the valuation parameters across nests, and we also allow for more than two products within each nest. Davis et al. (2014) show the pure assortment problem is NP-hard when the customers can make no purchase within the selected nest, while we prove the problem is polynomially solvable if we can jointly optimize over the assortment and prices.

4.2.2 Product Design and Optimal Pricing

Problem Formulation

There is a manufacturer who sells a category of products directly to the market, where the choice behavior follows Nested Logit model. By discontinuing some current products, introducing some new products, and pricing them in an optimal way, he wants to maximize his profit. Assuming an SKU's cost is the linear sum of its components' cost, and without loss of generality, assuming the market size is normalized to one, the manufacturer is solving the following problem:

$$\max_{S} \quad \max_{r} \quad R(S,r) \equiv \sum_{n} \sum_{in \in S_n} (r_{in} - c' x_{in}) \pi(in|S) \tag{4.2.4}$$

$$\text{s.t.} \quad |S_n| \leq K_n, \ \forall \ n = 1, ..., N \tag{4.2.5}$$

$$x_{in} \cdot x_{jn} \leq |F| - 1, \ \forall \ in \neq jn \in S_n \tag{4.2.6}$$

$$A_n x_{in} = e, \ \forall \ in \in S_n \tag{4.2.7}$$

$$x_{in} \in \{0,1\}^{|FL|}, \ \forall \ in \in S_n$$

This is a Mixed Integer Nonlinear Optimization problem. We have control over both assortment $S := S_1 \cup S_2 \cup ... \cup S_N$ and price r. Let the assortment of nest n be $S_n := \{1n, 2n, ...\}$, then each product in is determined by its configuration vector x_{in}. As is illustrated in the choice model, the lth component $x_{in}^l = 1$ if product in has attribute-level l and it is 0 otherwise. c is the attribute-level cost vector, and the lth component c_l represents attribute-level l's cost. By assuming the SKU cost is the linear sum of the attribute-level cost, $c' x_{in}$ is product in's cost. So the term $r_{in} - c' x_{in}$ is the profit of selling one unit of product in, and the term $\pi(in|S)$ is the probability of selling product in under assortment S and price vector r, as is defined in the choice model.

We consider cardinality constraints (4.2.5) on the offered assortment, which respectively limit the number of the products offered in each nest (Gallego and Topaloglu 2014). Constraint (4.2.6) $x_{in} \neq x_{jn}$ guarantees that any two products cannot have the same configuration. A_n is a configuration constraint matrix for nest n, and it imposes several constraints. First, some levels of the attributes are mutually exclusive but one of them must contain in the product, i.e., each row represents an attribute, for instance, for the laptop, row m can correspond to attribute RAM, and $A_n(m,l) = 1$ if l is one of the levels of RAM, and $A_n(m,l) = 0$ otherwise. Then $\sum_l A_n(m,l)x_{in}^l = 1$ means one and only one level of the RAM should be a component of the laptop. This constraint has flexibility in its nature, for example, it can also incorporate the case that some attribute can be skipped in the product by introducing one level representing "NONE" of that attribute. Second, this constraint precludes some attribute levels for nest n. Again taking the laptop for instance, nest n is Apple's market, and let row f correspond to the brand attribute and column q correspond to attribute-level "brand Apple", all the elements in row f are 0 except that $A_n(f,q) = 1$, then constraint $\sum_l A_n(f,l)x_{in}^l = 1$ enforces all products in nest n should be of Apple brand. Last constraint is the binary configuration requirement, and $|FL|$ is the total number of attribute levels.

Optimal Pricing

Assuming the set of products is fixed (i.e., S_n is given for each nest n), now we are dealing with the inner problem: maximizing the profit over prices, i.e.,

$$\max_{r} \quad R(r) \equiv \sum_{n} \sum_{in \in S_n} (r_{in} - \tilde{c}_{in}) \pi(in|S) \tag{4.2.8}$$

Recall that the product in's cost $\tilde{c}_{in} = c' x_{in}$ is the component-wise sum of attribute-level costs. When the set of products' configuration is fixed, $\tilde{c}$ is constant.

Driving some of the products' prices to positive infinity is equivalent to pushing those products out of market, so optimizing over price also enables us to refine the assortment. Therefore the question is: should we discontinue some products and will there be some nice structure on the optimal prices?

Theorem 4.2.1 (Constant Mark-up Pricing)**.** *Assuming the set of products is fixed, the markup, defined as price minus cost, is nest level invariant at optimality for all the products, i.e., there is a* θ_n^* *for each nest n such that* $r_{in}^* = \tilde{c}_{in} + \theta_n^*$

This result can be discovered from the first order condition, and for each nest it reduces a multi-variable optimization problem to a single dimension. Let θ_n denote the constant markup for all the products in nest n, i.e., $\theta_n = r_{in} - \tilde{c}_{in}$ for all products in nest n. Plug this relation into the choice model, then we can get the conditional choice probability of product *in* in nest *n*:

$$\pi(in|n,S) = \frac{exp(\tilde{v}_{in} - b_n\theta_n)}{\sum_{jn\in S_n} exp(\tilde{v}_{jn} - b_n\theta_n) + V_n} \tag{4.2.9}$$

The probability of choosing nest *n*:

$$\pi(n|S) = \frac{(\sum_{jn\in S_n} exp(\tilde{v}_{jn} - b_n\theta_n) + V_n)^{\gamma_n}}{\sum_{n'}(\sum_{jn'\in S_{n'}} exp(\tilde{v}_{jn'} - b_{n'}\theta_{n'}) + V_{n'})^{\gamma_{n'}} + 1} \tag{4.2.10}$$

By letting $\tilde{v}_{in} := \frac{\alpha_{in} - \beta_r^n \tilde{c}_{in}}{\gamma_n}$ and $b_n := \frac{\beta_r^n}{\gamma_n}$, which are deterministic when the assortment is fixed.

Instead of optimizing over prices, which are heterogeneous across all product in every nest, we optimize over the nest-level constant mark-up θ:

$$\begin{aligned}
\max_{\theta} R(\theta|S) &= \sum_n \sum_{in\in S_n} \theta_n \pi(in|S) \\
&= \sum_n \theta_n \pi(n|S) \sum_{in\in S_n} \pi(in|n,S) \\
&= \sum_n \theta_n \pi(n) \pi(S_n|n,S) \\
&= \sum_n \theta_n \pi(S_n|S) \qquad (4.2.11)
\end{aligned}$$

Where $\pi(S_n|n,S) := \sum_{in\in S_n} \pi(in|n,S)$ and $\pi(S_n|S) := \sum_{in\in S_n} \pi(in|S)$. Also define that $\pi(S_0|n,S) := 1 - \pi(S_n|n,S)$

Theorem 4.2.1 tells us that products in a same nest are either opted out of market all together or priced with a same finite markup. When we optimize the expected profit over the markups, the first order condition gives us the following result:

Theorem 4.2.2 (Optimal Nest Level Markup)**.** *All products will be priced finitely and the adjusted nest level markup, defined as* $\phi_n = \theta_n \pi(S_n|n,S) + \frac{\theta_n}{\gamma_n}\pi(S_0|n,S) - \frac{1}{\beta_r^n}$ *is nest invariant at optimality, i.e.,* $\phi_n = \phi$ *for all n.*

Thus we further reduce price optimization problem to a single dimension optimization with respect to the adjusted nest level markup ϕ, i.e.,

$$\begin{aligned} \max_{\phi} \quad & R(\phi|S) = \sum_n \theta_n \pi(S_n|S) \\ \text{s.t.} \quad & \phi = \theta_n \pi(S_n|n,S) + \frac{\theta_n}{\gamma_n}\pi(S_0|n,S) - \frac{1}{\beta_r^n}, \forall n \end{aligned} \tag{4.2.12}$$

Corollary 4.2.3. *The price optimization problem can be reduced to a single dimension optimization with respect to the adjusted nest level markup* ϕ*, where* $R(\phi|S)$ *is unimodal on* ϕ*, and the optimal solution can be determined from the unique fixed point* $R(\phi^*|S) = \phi^* = \max_\phi R(\phi|S)$.

Gallego and Wang (2014) draw the similar conclusions. However their proofs do not apply to our type-primary choice model.

Optimal Inner Nest Structure

Having found the optimal pricing structure, now we should focus on the outer problem – how to optimally configure the assortment.

Define $g_k^n := \frac{\beta_k^n - \beta_r^n c_k}{\gamma_n}$ as attribute k's value gap index in nest n, and $\tilde{v}_{in} := \frac{\alpha_{in} - \beta_r^n \tilde{c}_{in}}{\gamma_n} = \frac{\beta_0^n}{\gamma_n} + \sum_k g_k^n x_{in}^k$ as product in's value gap index. Those are similar to the definition in the MNL model except that now they are normalized by the nest coefficients γ. We can see this index still plays a key role in the NL optimal product configuration.

Theorem 4.2.4 (Greedy Inner Nest Structure)**.** *A feasible attribute-level with higher value gap can bring more profit to the product line. Given the available set of attribute-level, in the optimal product line design, within each nest, products with higher value gap have priority in picking the higher value gaped attribute levels.*

There are three important outputs from Theorem 4.2.4: First, the value gap index plays a key role in designing products; second, higher value gaped attribute-level can always bring more profit than lower value gaped attribute levels can; third, adding an attribute-level to one product may preclude it from another product in the same nest, but the higher value gaped product always has priority in selecting attribute levels. Thus, the greedy algorithm in section 4.1 also works for the NL model.

Corollary 4.2.5. *We can run the K_n- shortest path algorithm for every nest and get the optimal product configuration in polynomial time.*

4.3 Conclusions and Future Research

Different from the traditional assortment optimization problem aiming at discontinuing some existing suboptimal products, our chapter proposes a method in designing an optimal product line by configuring all the products with attribute levels. Formulating consumers' utility as the aggregated part-worth valuation, we are able to forecast the market share of any product under a given assortment even the product has not existed in the market before.

We optimize the product line configuration on the attributes space and find that under the MNL model, the product with higher value gap has priority in selecting attributes, and an attribute level with higher residual valuation can bring more profit than other levels of the attribute can. This structure makes a greedy algorithm, derived from the K-shortest paths algorithm, able to find an optimal K products' configuration in polynomial time. This structural result also applies to the NL model with respect to each nest, meaning that we can find an optimal product line configuration for every nest in polynomial time by running K-shortest paths algorithm nest-wise.

Conditional on the optimal product line, we optimize over prices and find that the constant markup property still applies for MNL model, meaning every product should be priced at its cost plus a constant, which is invariant across all the SKUs. Our NL formulation incorporates both brand-primary choice process and type-primary choice process. We show that under the NL model, the markup is constant for all the products within a same nest at optimality. In addition, there is an optimal adjusted nest-level markup invariant across all the nests, which reduce the price optimization problem to a single-dimensional maximization of a continuous function over a bounded interval. Moreover, under both choice model, the profit function is unimodal with respect to the single dimension parameter.

In our NL choice model formulation, we cluster all the products along the same path in the decision tree to a single nest. A direction in the future may be modeling the consumer choice by a d-level Nested Logit model and optimize the product line on it. Another direction would be product line configuration on Mixed Multinomial Logit model. The problem is NP-hard, so it would interesting to design some efficient heuristics and study their perfomances.

Chapter 5

Conclusions

We studied product line design, pricing and framing problems under a variety of choice models in this thesis.

In Chapter 2, we looked at the operations applicability of the Random Consideration Set based choice model proposed by Manzini and Marrioti (2014). We showed how to recover the full ordering and attention probabilities given accurate estimates of choice probabilities or from noisy empirical data. Empirical testing of the Random Consideration Set model on our airline partner's data showed that it outperformed the Multinomial Logit choice model in test data over all markets. The Random Consideration Set model also performed better than the Mixture of MNLs model giving a better fit on 67.0% of the markets and with an average improvement of 15.6%. We showed that an assortment that maximizes expected revenues could be found in $O(n)$ time where n is the number of products. Adding a cardinality constraint increases the complexity to $O(n^2)$. We showed that the efficient sets discovery problem can be solved in $O(n^2)$ where the goal is to find an assortment to maximize revenues net of the marginal value of capacity. We extended the model to allow ties in preferences and showed that a revenue-ordered assortment has a $1/2$ performance guarantee relative to the optimal assortment. We studied the pricing problem where the preference ordering are price aware and showed under mild assumptions that optimal profits are such that both the profit contributions and the net utilities to consumers are aligned with the value gap defined as the difference between the value of the product to consumers and its unit wholesale cost.

In Chapter 3, we proposed one of the first models of "product framing" and pricing. We presented a model where a set of products are displayed, or framed, into a set of virtual web pages. We showed that the product framing problem is NP-hard. We derived algorithms with guaranteed performance relative to an optimal algorithm under reasonable assumptions. Our algorithms are fast and easy to implement. We also presented structural results for pricing under framing effects. For profit maximization problem, at optimality products are sorted in descending order of value gap, which is defined as expected utility when the product is priced at cost; and markups are page-dependent, with higher markups associated with products on later pages, so that products in the first page are of the highest utility and have the lowest markups.

In Chapter 4, we studied a manufacturer who wants to design and price a set of products which are defined by attribute-level combinations. We assumed that demand for products is based on a Logit model (MNL or NL) that measures quality as the aggregate value of the attribute-levels. In addition, the Logit model penalizes prices by assuming that customers are price sensitive. Assuming that the unit production cost is equal to the sum of the wholesale prices of its components we showed how to solve the corresponding product line design problem. We showed that optimal product configurations give priority to attribute levels with high value gaps. The K best configurations can be obtained through a greedy algorithm derived from the K-shortest path problem. The resulting pricing problem has the well known constant mark-up property, which combines well with the design problem and allows us to find an optimal configuration in polynomial time.

To summarize, this thesis studied assortment optimization related problems under general choice models. The analysis on consumer choice models showed that the models developed from a behavioral perspective can model choices better than a traditional parametric model, and might admit efficient estimation and assortment optimization algorithms in the meantime (Gallego and Li 2017). In the future, it is important to develop more consumer choice models that delicately balance the realism in terms of capturing consumers? purchase behavior versus the practicability in terms of ensuring understandable estimation and assortment optimization algorithms.

Bibliography

[1] Agarwal, Ashish, Kartik Hosanagar, Michael D. Smith. 2011. Location, location, location: An analysis of profitability of position in online advertising markets. *Journal of Marketing Research*. **48**(6) 1057-1073.

[2] Alnur Ali, Marina Meil?, Experiments with Kemeny ranking: What works when?, *Mathematical Social Sciences*, Volume 64, Issue 1, 2012, Pages 28-40, ISSN 0165-4896

[3] Amos Tversky, Daniel Kahneman. 1986. Rational choice and the framing of decisions. *The Journal of Business* **59**(4) S251-S278.

[4] Anderson, S. P., A. de Palma, J. Thisse. 1992. *Discrete Choice Theory of Product Differentiation.* 1st ed. The MIT Press.

[5] Animesh, Animesh, Siva Viswanathan, Ritu Agarwal. 2011. Competing ?creatively? in sponsored search markets: The effect of rank, differentiation strategy, and competition on performance. *Information Systems Research* **22**(1) 153-169.

[6] Aouad, Ali, Vivek Farias, Retsef Levi. 2016. Assortment Optimization Under Consider-Then-Choose Choice Model. Working papers, MIT.

[7] Aouad, Ali, Vivek Farias, Retsef Levi, Danny Segev. 2015. The approximability of assortment optimization under ranking preferences. Tech. rep., Working paper.

[8] Arbatskaya, Maria. 2007. Ordered search. *The RAND Journal of Economics* **38**(1) 119-126.

[9] Baye, Michael R., J. Rupert J. Gatti, Paul Kattuman, John Morgan. 2009. Clicks, discontinuities, and firm demand online. *Journal of Economics & Management Strategy* **18**(4) 935-975.

[10] Berry, S. T. 1994. Estimating Discrete-Choice Models of Product Differentiation. *The RAND Journal of Economics.* **71** 242-262

[11] Berry, S., J. Levinsohn, and A. Pakes Z.1995. Automobile Prices in Market Equilibrium. *Econometrica*. **63** 841-890.

[12] Blanchet, Jose, Guillermo Gallego, Vineet Goyal. 2016. A markov chain approximation to choice modeling. *Operations Research* **64**(4) 886-905.

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

[14] Brynjolfsson, Erik, Astrid A. Dick, Michael D. Smith. 2010. A nearly perfect market? : differentiation vs. price in consumer choice.

[15] Cachon, Gerard P., Christian Terwiesch, Yi Xu. 2005. Retail assortment planning in the presence of consumer search. *Manufacturing & Service Operations Management* **7**(4) 330-346.

[16] Chandon, Pierre, J. Wesley Hutchinson, Eric T. Bradlow, Scott H. Young. 2009. Does in-store marketing work? effects of the number and position of shelf facings on brand attention and evaluation at the point of purchase. *Journal of Marketing* **73**(6) 1-17.

[17] Chiang, Jeongwen, Siddhartha Chib, Chakravarthi Narasimhan. 1998. Markov chain monte carlo and models of consideration set and parameter heterogeneity. *Journal of Econometrics* **89**(1-2) 223-248.

[18] Cheremukhin, Anton A., Anna Popova, Antonella Tutino. 2011. Experimental evidence on rational inattention. Working Papers 1112, Federal Reserve Bank of Dallas.

[19] Corstjens, J., M. Corstjens. 2012. The battlefield for Mindspace and Shelfspace. John Wiley and Sons, Inc., 69-82.

[20] Chong, J. K., T. H. Ho, C. S. Tang. 2001. A modeling framework for category assortment planning. *Manufacturing & Service Operations Management.* **3**(3) 191-210.

[21] Currim, I. S., R. J. Meyer and N. T. Le. 1988. Disaggregate Tree-Structured Modeling of Consumer Choice Data. *Journal of Marketing Research.* **25** (August), 253-265.

[22] Davis, James, Guillermo Gallego, Huseyin Topaloglu. 2014a. Assortment planning under the multinomial logit model with totally unimodular constraint structures. Tech. rep., Working paper.

[23] Davis, J. M., G. Gallego, and H. Topaloglu. 2014b. Assortment optimization under variants of the nested logit model. *Operations Research* **62** (2) 250-273.

[24] Davis, James M., Huseyin Topaloglu, David P. Williamson. 2015. Assortment optimization over time. *Operations Research Letters* **43**(6) 608- 611.

[25] Debreu, G. 1952. A social equilibrium existence theorem. *Proceedings of the National Academy of Sciences.* **38**(10) 886-893.

[26] Desir, Antoine, Vineet Goyal. 2013. Near-optimal algorithms for capacity constrained assortment optimization. Tech. rep., Working paper.

[27] Dijkstra, E. W. A note on two problems in connexion with graphs. *Numer. Math. I.* **4** 269-271.

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

[29] Dwork, Cynthia and Kumar, Ravi and Naor, Moni and Sivakumar, D. 2001. Rank Aggregation Methods for the Web. Proceedings of the 10th International Conference on World Wide Web

[30] Eliaz, Kfir, Ran Spiegler. 2011a. Consideration sets and competitive marketing. *The Review of Economic Studies* **78**(1) 235-262.

[31] Eliaz, Kfir, Ran Spiegler. 2011b. On the strategic use of attention grabbers. *Theoretical Economics* **6**(1) 127-155.

[32] Daniel M. Ennis, John M. Ennis, Accounting for no difference/preference responses or ties in choice experiments, *Food Quality and Preference,* Volume 23, Issue 1, 2012, Pages 13-17, ISSN 0950-3293,

[33] Fader, P. S., B. G. S. Hardie. 1996. Modeling consumer choice among SKUs. *J. Marketing Res.* **33** 442-452.

[34] Farias, Vivek F., Srikanth Jagabathula, Devavrat Shah. 2013. A nonparametric approach to modeling choice with limited data. *Management Science* **59**(2) 305-322.

[35] Feldman, J., H. Topaloglu. 2015. Assortment optimization under the multinomial logit model with nested consideration sets. Tech. rep., Working Paper.

[36] Feng, Juan, Hemant K. Bhargava, David M. Pennock. 2007. Implementing sponsored search in web search engines: Computational evaluation of alternative mechanisms. *INFORMS Journal on Computing* **19**(1) 137-148.

[37] Fishbein, M. and Ajzen, I. 1981. Attitudes and voting behavior: An application of the theory of reasoned action. In G. M. Stephenson & J. M. Davis (Eds.), *Progress in Applied Social Psychology* (Vol. 1, pp. 253-313). London: Wiley.

[38] Fisher, M., and R. Vaidyanathan. 2014. A Demand Estimation Procedure for Retail Assortment Optimization with Results from Implementations. *Management Science* **60**(10):2401-2415.

[39] Gallego, Guillermo, Anran Li, 2017. Attention,Consideration then Selection.Tech. rep., working paper.

[40] Gallego, Guillermo, Anran Li, Jose Beltran. 2016. Product line design and pricing under logit model. Tech. rep., working paper.

[41] Gallego, Guillermo, Anran Li, Van anh Truong, Xinshang Wang. 2016. Approximation algorithms for product framing and pricing. Tech. rep., working paper.

[42] Gallego G, Stefanescu C. 2011. Upgrades, Upsells and Pricing in Revenue Management. Working paper, Columbia University, New York.

[43] Gallego, Guillermo, Huseyin Topaloglu. 2016. Handbook in revenue management and dynamic pricing. Tech. rep., Book to appear.

[44] Gallego G, Topaloglu H. 2014. Constrained assortment optimization for the nested logit model. *Management Science*. **60** (10) 2583-2601

[45] Gallego, G., and R. Wang. 2013. Multi-product price optimization and competition under the nested logit model with product-differentiated price sensitivities. *Operations Research* **62** (2) 450-461.

[46] Ghose, Anindya, Sha Yang. 2009. An empirical analysis of search engine advertising: Sponsored search in electronic markets. *Management Science* **55**(10) 1605-1622.

[47] Gilbride, Timothy J., Greg M. Allenby. 2004. A choice model with conjunctive, disjunctive, and compensatory screening rules. *Marketing Science* **23**(3) 391-406.

[48] Guadagni, P.M., J.D.C. Little. 1983. A logit model of brand choice calibrated on scanner data. *Marketing Science.* **2**(3) 203-238.

[49] Haensel, Alwin, Ger Koole, Jeroen Erdman. 2011. Estimating unconstrained customer choice set demand: A case study on airline reservation data. *Journal of Choice Modelling* **4**(3) 75-87.

[50] Hanson W, Martin K. 1996. Optimizing multinomial logit profit functions. *Management Sci.* **42**(7):992-1003.

[51] Hauser, John R. 1978. Testing the accuracy, usefulness, and significance of probabilistic choice models: An information-theoretic approach. *Operations Research* **26**(3) 406-421.

[52] Hauser, John R., Steven P. Gaskin. 1984. Application of the defender consumer model. *Marketing Science* **3**(4) 327-351.

[53] Honhon, Dorothee, Sreelata Jonnalagedda, Xiajun Amy Pan. 2012. Optimal algorithms for as- sortment selection under ranking-based consumer choice models. *Manufacturing & Service Operations Management* **14**(2) 279-289.

[54] Hopp WJ, Xu X. 2005. Product line selection and pricing with modularity in design. *Management Sci.* **7**(3):172-187.

[55] Howard, J.A., J.N. Sheth. 1969. The theory of buyer behavior. Wiley marketing series, Wiley.

[56] Jagabathula, Srikanth, Paat Rusmevichientong. 2016. A nonparametric joint assortment and price choice model. *Management Science* 0(0) null.

[57] Johnson, Eric J., Wendy W. Moe, Peter S. Fader, Steven Bellman, Gerald L. Lohse. 2004. On the depth and dynamics of online search behavior. *Manage. Sci. 50*(3) 299-308.

[58] Kannan, P. K. and Gordon P. Wright 1991. Modeling and Testing Structured Markets: A Nested Logit Approach. *Marketing Science.* **10** (Winter) 58-82.

[59] Kohli, R., R. Sukumar. 1990. Heuristics for product-line design using conjoint analysis. *Management Sci.* **36**(12) 1464-1478.

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

[61] Luo, L. 2011. Product Line Design for Consumer Durables: An Integrated Marketing and Engineering Approach. *Journal of Marketing Research*. **48** (1) 128-139.

[62] Lancaster, Kelvin. 1971. *Consumer Demand: A New Approach.* New York: Columbia University Press.

[63] Li G, Rusmevichientong P, Topaloglu H. 2014. The d-level nested logit model: Assortment and price optimization problems. Working paper, Marshall School of Business, University of Southern California, Los Angeles.

[64] Li, H., W.T. Huh. 2011. Pricing multiple products with the multinomial logit and nested logit models: Concavity and implications. *Manufacturing Service Oper. Management* . **13**(4) 549-564.

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

[66] Manrai, Ajay K, Rick L Andrews. 1998. Two-stage discrete choice models for scanner panel data: An assessment of process and assumptions. *European Journal of Operational Research* **111**(2) 193 -215.

[67] Manzini, Paola, Marco Mariotti. 2014. Stochastic choice and consideration sets. *Econometrica* **82**(3) 1153-1176.

[68] McFadden, D. 1974. Conditional logit analysis of qualitative choice behavior. P. Zarembka, ed., *Frontiers in Econometrics. Academic Press*, New York, 105-142.

[69] McFadden, D. 1978. Modelling the choice of residential location. A. Karlvist, L. Lundqvist, F. Snickars, J. Weibull, eds., *Spatial Interaction Theory and Planning Models.* Amsterdam: North-Holland, 75-96.

[70] McFadden, D. 1980. Econometric models for probabilistic choice among products. *The Journal of Business* **53**(3) 13-29.

[71] Mittal, Shashi, Andreas S. Schulz. 2013. A general framework for designing approximation schemes for combinatorial optimization problems with many objectives combined into one.*Operations Research* **61**(2) 386-397.

[72] Nemhauser, George L., Laurence A. Wolsey. 1988. Integer and Combinatorial Optimization. Wiley Interscience, New York, NY, USA.

[73] Paul, Alice, Huseyin Topaloglu, Jacod Feldman. 2016. Assortment optimization for choosy customers. Tech. rep., working paper.

[74] Pras, Bernard, John Summers. 1975. A comparison of linear and nonlinear evaluation process models. *Journal of Marketing Research* **12**(3) 276-281.

[75] Quandt, Richard E. 1956. A Probabilistic Theory of Consumer Behavior. *Quarterly Journal of Economics.* **70** (November), 507-36.

[76] Rusmevichientong, Paat, Benjamin Van Roy, Peter W. Glynn. 2006. A nonparametric approach to multiproduct pricing. *Operations Research* **54**(1) 82-98.

[77] Sahin, Ozge, Ruxian Wang. 2014. The impact of consumer search cost on assortment planning and pricing. Tech. rep., Working Paper.

[78] Shaked, Moshe, J. George Shanthikumar. 2007. Stochastic Orders. Springer New York.

[79] Shugan, Steven M. 1980. The cost of thinking. *Journal of Consumer Research* **7**(2) 99-111.

[80] Song, Jing-sheng, Zhengliang Xue. 2009. Demand management and inventory control for substitutable products. Tech. rep., Working Paper.

[81] Silk, Alvin J, Glen L Urban. 1978. Pre-test-market evaluation of new packaged goods: A model and measurement methodology. *Journal of marketing Research* 171191.

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

[83] Train, Kenneth E. 2008. Em algorithms for nonparametric estimation of mixing distributions. *Journal of Choice Modelling* **1**(1) 40-69.

[84] W. Z. Rayfield, P. Rusmevichientong, and H. Topaloglu. 2015. Approximation methods for pricing problems under the nested logit model with price bounds. *INFORMS Journal on Computing*. Forthcoming.

[85] Tversky, A. and S. Sattath. 1979. PreferenceTrees. *Psychological Review.* **86** 542-573.

[86] van Ryzin, G., S. Mahajan. 1999. On the Relationship between Inventory Costs and Variety Benefits in Retail Assortments. *Management Sci.* **45** 1496-1509.

[87] van Ryzin, Garrett, Gustavo Vulcano. 2015. A market discovery algorithm to estimate a general class of nonparametric choice models. *Management Science* **61**(2) 281-300.

[88] Schon, C. 2010. On the Optimal Product Line Selection Problem with Price Discrimination. *Management Science*, **56** 896-902.

[89] Song J, Xue Z. 2007. Demand management and inventory control for substitutable products. Working paper, Duke University, Durham, NC.

[90] J. Y. Yen. 1971. Finding the K shortest loopless paths in a network. *Management Science* **17** 712-716.

[91] Zufryden. 1977. A Conjoint-Measurement-Based Approach for Optimal New Product Design and Product Positioning. *Analytical Approaches to Product and Market Planning* (A. D. Shocker, Ed.), Marketing Science Institute, Cambridge, MA, 100-14.

— 1982. Product Line Optimization by Integer Programming. *Proc. Annual Meeting of ORSA/ TIMS.* San Diego, CA.

Appendix A

Appendix for Chapter 2: Attention, Consideration then Selection

A.1 Proofs in Section 2.4

A.1.1 Proof of Lemma 2.4.4

Proof. Let $S^*(z)$ be the efficient set that maximizes $R(S,z)$. We will show that $S^*(z) \subset S^*$ when $z > 0$. This is sufficient because we can always transform the problem of comparing the efficient sets for $z_1 > z_2 > 0$ to the problem of comparing the efficient sets for $z_1 - z_2$ and 0 by transforming the revenues $r_j \leftarrow r_j - z_2$ for all $j \in N$. Our conclusion $S^*(z) \subset S^*$ will follow if we can show that $i \notin S^*$ implies that $i \notin S^*(z)$. Let $i = \arg\min\{j \in N, j \notin S^*\}$, then product i is rejected by the algorithm on account of $r_i < H_{i-1}$, but then by adding and subtracting $z\Pi(\tilde{S}_{i-1})$ to H_{i-1} we see that $r_i < H_i(z) + z\Pi(\tilde{S}_{i-1})$, and this directly implies that $r_i - z < H_{i-1}(z)$, so $\tilde{S}_i(z) \subset \tilde{S}_i$. By induction, assume $\tilde{S}_{k-1}(z) \subset \tilde{S}_{k-1}$ and $k \notin S^*$, or equivalently $r_k < H_{k-1}$, then the same argument shows that $r_k - z < H_{k-1}(z)$ so $\tilde{S}_k(z) \subset \tilde{S}_k$. Consequently, $S^*(z) = \tilde{S}_n(z) \subset \tilde{S}_n = S^*$ completing the proof. □

A.1.2 Proofs of Subsection 2.4.3

To obtain structural results, we need the following Lemmas.

Lemma A.1.1. *Suppose $k \notin S$. If $k \in T$, then*

$$\Pi(T, S \cup \{k\}) = \lambda_k \Pi(T \setminus \{k\}, S),$$

and if $k \notin T$, then

$$\Pi(T, S \cup \{k\}) = (1 - \lambda_k)\Pi(T, S).$$

This Lemma shows an effective way to compute $\Pi(\cdot, S \cup \{k\})$ based on the knowledge of $\Pi(\cdot, S)$. Next let us see how the revenue $J(S)$ updates.

Lemma A.1.2. *If there is a product $k \in N$ such that $J(S) > J(S \cup \{k\})$, then $J(S) > J(S \cup \{l\})$ for all l such that $r_l < r_k$. Moreover, for any $T \subset \{l \in N : r_l \leq r_k\}$, $J(S) > J(S \cup T)$.*

Proof. Lemma A.1.1 allows us to write

$$J(S \cup \{k\}) = \lambda_k \sum_{T \subset S} r(T \cup \{k\}) \times \Pi(T, S) + (1 - \lambda_k) J(S),$$

so

$$J(S \cup \{k\}) - J(S) = \lambda_k \sum_{T \subset S} [r(T \cup \{k\}) - r(T)] \times \Pi(T,S). \tag{A.1}$$

Consequently $J(S) > J(S \cup \{k\})$ if and only if

$$\sum_{T \subset S} [r(T \cup \{k\}) - r(T)] \Pi(T,S) < 0.$$

Since $r(T \cup \{l\}) \leq r(T \cup \{k\})$ for all l such that $r_l \leq r_k$, it follows that

$$\sum_{T \subset S} [r(T \cup \{l\}) - r(T)] \Pi(T,S) < 0,$$

and consequently that $J(S) > J(S \cup \{l\})$.

The value of $J(S \cup T)$ can be expressed by

$$\begin{aligned} J(S \cup T) &= \Pi(\emptyset, T) J(S) + \sum_{\substack{S_a \subset T \\ S_a \neq \emptyset}} \Big(\Pi(S_a, T) \sum_{S_l \subset S} \Pi(S_l, S) r(S_l \cup S_a) \Big) \\ &= J(S) + \sum_{\substack{S_a \subset T \\ S_a \neq \emptyset}} \Big(\Pi(S_a, T) \sum_{S_l \subset S} \Pi(S_l, S) \big(r(S_l \cup S_a) - J(S) \big) \Big) \\ &\leq J(S) + \sum_{\substack{S_a \subset T \\ S_a \neq \emptyset}} \Big(\Pi(S_a, T) \sum_{S_l \subset S} \Pi(S_l, S) \big(r(S_l \cup \{k\}) - J(S) \big) \Big) \\ &< J(S) \end{aligned}$$

The first line expresses the value of $J(S \cup T)$ as the sum of expected revenues when consumer pays no attention to product in T and when they pay attention to some of them. The second equality is because $\Pi(\emptyset, T) = 1 - \sum_{\substack{S_l \subset T \\ S_l \neq \emptyset}} \Pi(S_l, S)$ and $\sum_{S_l \subset S} \Pi(S_l, S) = 1$. The first inequality is due to $r_k \geq r_i, \forall i \in T$, and the last inequality is because $\sum_{S_l \subset S} \Pi(S_l, S)\big(r(S_l \cup \{k\}) - J(S)\big) < 0$. □

Proof of Lemma 2.4.6

Proof. A necessary condition for S^* to be optimal is that $J(S^* \setminus \{j\}) \leq J(S^*)$ for any $j \in S^*$. We will show that this necessary condition guarantees that $r_j \geq \frac{1}{2} J(S^*)$, $\forall j \in S^*$. Let us prove it by induction on the size of assortment S^*.

Suppose first that $S^* = \{j\}$ has a single element. Then $J(S^*) = r_j \lambda_j$, and consequently $r_j \geq J(S^*)/2 = r_j \lambda_j / 2$ on account of $2 > 1 \geq \lambda_j$. Suppose it is true for $|S^*| \leq n-1$, consider now the case where $|S^*| = n$. Assume that $r_j \leq r_i$ for all $i \in S^*$, and let $S^*_{-j} = S^* / \{j\}$. From equation (A.1)

$$0 \leq J(S^*) - J(S^*_{-j}) = \lambda_j \sum_{T \subset S^*_{-j}} [r(T \cup \{j\}) - r(T)] \times \Pi(T, S^*_{-j})$$

Since the right hand side is increasing in r_j and is negative if $r_j = 0$, there is a value of r_j for which the right hand equal to zero, and this is the smallest value that r_j can take to preserve the optimality of S^*, with $J(S^*) - J(S^*_{-j}) = 0$.

Notice that the terms $r(T \cup \{j\}) - r(T) < 0$ except when $T = \emptyset$. Therefore, if we keep only the sets T of cardinality $|T| \leq 1$, we have

$$0 \leq J(S^*) - J(S^*_{-j}) \leq \lambda_j \left(r_j \times \Pi(\emptyset, S^*_{-j}) + \sum_{i \in S^*_{-j}} \frac{r_j - r_i}{2} \times \Pi(i, S^*_{-j}) \right).$$

Solving for r_j we obtain

$$r_j \geq \frac{\sum_{i\in S^*_{-j}} r_i v_i}{2+\sum_{i\in S^*_{-j}} v_i}, \tag{A.2}$$

where $v_i = \lambda_i/(1-\lambda_i)$.

We want to know how small the lower bound (A.2) can be with respect to $J(S^*_{-j})$, i.e.,

$$\min_{r,\lambda} \quad \frac{\frac{\sum_{i\in S^*_{-j}} r_i v_i}{2+\sum_{i\in S^*_{-j}} v_i}}{J(S^*_{-j}).}$$

We can formulate it as an optimization problem while normalizing $J(S^*_{-j}) = 1$.

$$\begin{aligned} \min_{r,\lambda} \quad & \frac{\sum_{i\in S^*_{-j}} r_i v_i}{2+\sum_{i\in S^*_{-j}} v_i} \\ \text{s.t.} \quad & J(S^*_{-j}) = \sum_{T\subset S^*_{-j}} r(T)\times \Pi(T, S^*_{-j}) = 1 \end{aligned}$$

We want to show the objective value $\frac{\sum_{i\in S^*_{-j}} r_i v_i}{2+\sum_{i\in S^*_{-j}} v_i} \geq \frac{1}{2}J(S^*_{-j}) = \frac{1}{2}$, equivalently,

$$\sum_{i\in S^*_{-j}} (r_i - \frac{1}{2})v_i \geq 1$$

by multiplying both sides by the denominator and combining common terms. Therefore, as long as we can show the following optimization problem has objective value greater than or equal to one, then we are done

$$\min_{r,\lambda} \quad \sum_{i\in S^*_{-j}} (r_i - \tfrac{1}{2}) v_i \tag{A.3}$$

$$\text{s.t} \quad . \quad J(S^*_{-j}) = \sum_{T\subset S^*_{-j}} r(T) \times \Pi(T, S^*_{-j}) = 1 \tag{A.4}$$

Let write down the constraint (A.4) explicitly

$$\sum_{\substack{T\subset S^*_{-j}\\ T\neq\emptyset}} \prod_{k\in T} \lambda_k \prod_{k\in S^*_{-j}\setminus T} (1-\lambda_k) \times \sum_{i\in T} \frac{r_i}{|T|} = 1$$

If we divide and multiply left hand side by $\Pi(\emptyset, S^*_{-j})$, we have

$$\Pi(\emptyset, S^*_{-j}) \Big(\sum_{\substack{T\subset S^*_{-j}\\ T\neq\emptyset}} \prod_{k\in T} v_k \times \sum_{i\in T} \frac{r_i}{|T|} \Big) = 1$$

which is equivalent to

$$\Pi(\emptyset, S^*_{-j}) \Big(\sum_{i\in S^*_{-j}} \big(\sum_{T\subset S^*_{-j}\setminus\{i\}} \frac{\prod_{k\in T} v_k}{|T|+1} \big) \times v_i r_i \Big) = 1 \tag{A.5}$$

Observe that both the objective (A.3) and the constraint (A.5) are linear on $r's$. Fixing the values of $\lambda's$, we notice that the problem is a continuous knapsack problem on $r's$. We would like to increase the value of r_i to the largest possible amount if product i has the smallest ratio for v_i to $\sum_{T\subset S^*_{-j}\setminus\{i\}} \frac{\prod_{k\in T} v_k}{|T|+1} \times v_i$, which are the coefficient before r_i in the objective function and the coefficient before r_i in the constraint respectively, i.e., increase the value of r_i to its upper bound for

$$i \in \arg\min_{l \in S^*_{-j}} \frac{v_l}{\sum_{T \subset S^*_{-j} \setminus \{l\}} \frac{\prod_{k \in T} v_k}{|T|+1} \times v_l} = \frac{1}{\sum_{T \subset S^*_{-j} \setminus \{l\}} \frac{\prod_{k \in T} v_k}{|T|+1}}$$

and to preserve the constraint we decrease the value of r_l to their lower bound for all other $l \neq i$. Moreover, observe the above ratio, for product i and l, if $\lambda_i < \lambda_l$, we can easily show that

$$\frac{1}{\sum_{T \subset S^*_{-j} \setminus \{i\}} \frac{\prod_{k \in T} v_k}{|T|+1}} < \frac{1}{\sum_{T \subset S^*_{-j} \setminus \{l\}} \frac{\prod_{k \in T} v_k}{|T|+1}}$$

Therefore, under any fixed λ, we increase the value of r_i to its upper bound such that $i \in \arg\min\{k \in S^*_{-j} : \lambda_k\}$ and decrease all the other $r_l's$, $l \in S^*_{-j} \setminus \{i\}$ to their lower bounds, so it will be indifferent from introducing product l or not

$$r_i \times \lambda_i = J(S^*_{-j}) = 1$$

If we look back to the objective function (A.3), we have $(r_k - \frac{1}{2}) \times v_k \geq 0$ for all $k \in S^*_{-j}$. This is because $r_k \geq \frac{1}{2} J(S^*_{-j}) = \frac{1}{2}$ for all $k \in S^*_{-j}$ by induction. Recall that $|S^*_{-j}| = n-1$, and by induction we have assumed when $|S^*_{-j}| \leq n-1$ that $r_k \geq \frac{1}{2} J(S^*_{-j})$ for all $k \in S^*_{-j}$, in order to guarantee $J(S^*_{-j} \setminus \{k\}) \leq J(S^*_{-j})$. Note that we need $J(S^*_{-j} \setminus \{k\}) \leq J(S^*_{-j})$, since if there is a $k \in S^*_{-j}$, such that $J(S^*_{-j} \setminus \{k\}) > J(S^*_{-j})$, then $J(S^*_{-j} \setminus \{k\}) > J(S^*)$ due to lemma (A.1.2) as $r_k > r_j$, which violates the optimality of S^*. Therefore, we show that the objective function is lower bounded by

$$(r_i - \frac{1}{2})v_i$$

Now the entire problem is reduced to proving if the following problem has objective value greater than or equal to one:

$$\begin{aligned} \min_{r_i, \lambda_i} \quad & (r_i - \frac{1}{2})v_i \\ \text{s.t.} \quad & r_i \lambda_i = 1 \end{aligned}$$

It can be easily proved that the above problem is optimized when $\lambda_i \to 0$ and the objective value equals to one. Now the lemma gets proved. $\square$

Proof of Theorem 2.4.7

Proof. From Lemma 2.4.6 we know the optimal assortment $S^* \subset \bar{S}$, so let $\hat{S} \equiv \bar{S} \setminus S^*$, then the value of $J(\bar{S}) = J(S^* \cup \hat{S})$ can be expressed by

$$
\begin{aligned}
J(S^* \cup \hat{S}) &= \Pi(\emptyset, \hat{S}) J(S^*) + \sum_{\substack{S_a \subset \hat{S} \\ S_a \neq \emptyset}} \Big(\Pi(S_a, \hat{S}) \sum_{S_l \subset S^*} \Pi(S_l, S^*) r(S_l \cup S_a) \Big) \\
&= J(S^*) + \sum_{\substack{S_a \subset \hat{S} \\ S_a \neq \emptyset}} \Big(\Pi(S_a, \hat{S}) \sum_{S_l \subset S^*} \Pi(S_l, S^*) \big(r(S_l \cup S_a) - J(S^*) \big) \Big) \\
&\leq J(S^*) - \big(\sum_{\substack{S_a \subset \hat{S} \\ S_a \neq \emptyset}} \Pi(S_a, \hat{S}) \sum_{S_l \subset S^*} \Pi(S_l, S^*) \big) \frac{1}{2} J(S^*) \\
&< \frac{1}{2} J(S^*)
\end{aligned}
$$

The first line expresses the value of $J(S^* \cup \hat{S})$ in terms of S^* and subsets of $\hat{S}$. The second equality is because $\Pi(\emptyset, \hat{S}) = 1 - \sum_{\substack{S_l \subset \hat{S} \\ S_l \neq \emptyset}} \Pi(S_l, \hat{S})$ and $\sum_{S_l \subset S^*} \Pi(S_l, S^*) = 1$. The first inequality is due to $r(S_l \cup S_a) \geq \frac{1}{2} J(S^*)$ by the set's definition, and the last inequality is because $\sum_{\substack{S_a \subset \hat{S} \\ S_a \neq \emptyset}} \Pi(S_a, \hat{S}) \sum_{S_l \subset S^*} \Pi(S_l, S^*) < 1$. □

Proof of Corollary 2.4.8

This result immediately follows Lemma A.1.2 and Theorem 2.4.7.

Proof of Theorem 2.4.9

Proof. The first part is to prove that $r_j \geq \frac{1}{2}J(S^*), \forall j \in \tilde{S}$. This is equivalent to proving $V(\tilde{S}) \geq \frac{1}{2}J(S^*)$, where $V(\tilde{S})$ is the MNL expected revenue. Still let us normalize $J(S^*) = 1$, so the problem boils down to proving the problem below has lower bound $\frac{1}{2}$

$$\begin{aligned} \min_{r,\lambda} \max_{S \subset N} \quad & \frac{\sum_{i \in S} r_i v_i}{2 + \sum_{i \in S} v_i} \\ \text{s.t.} \quad & J(S^*) = 1 \end{aligned} \tag{A.6}$$

Since we know

$$\max_{S \subset N} \frac{\sum_{i \in S} r_i v_i}{2 + \sum_{i \in S} v_i} \geq \frac{\sum_{i \in S^*} r_i v_i}{2 + \sum_{i \in S^*} v_i}$$

a lower bound of the objective function (A.6) will be $\min_{p,\lambda} \frac{\sum_{i \in S^*} r_i v_i}{2 + \sum_{i \in S^*} v_i}$, so if we can prove the problem below has lower bound $1/2$, then we are done

$$\begin{aligned} \min_{r,\lambda} \quad & \frac{\sum_{i \in S^*} r_i v_i}{2 + \sum_{i \in S^*} v_i} \\ \text{s.t.} \quad & J(S^*) = 1 \end{aligned} \tag{A.7}$$

The objective value is indeed greater than or equal to $1/2$ as is proved in the Lemma 2.4.6. Next we show that $\tilde{S}$ is inferior to the optimal nested by prices assortment but is superior to $\bar{S} \equiv \{i \in N : r_i \geq \frac{1}{2}J(S^*)\}$. Inferiority is easy to check since $\tilde{S}$ is one of the nested by prices assortments. Superiority can be proved by $J(\tilde{S}) \geq J(\bar{S})$. This is because first we know, $\tilde{S} \subset \bar{S}$; if $\tilde{S} \neq \bar{S}$, there should exist $j \in \arg\max\{r_i : i \in \bar{S} \setminus \tilde{S}\}$. As long as we can prove $J(\tilde{S}) > J(\tilde{S} \cup \{j\})$, we are able to conclude $J(\tilde{S}) > J(\bar{S})$ which follows Lemma A.1.2. Recall that from Lemma 2.4.6, especially formula (A.2), we know a necessary condition for $J(\tilde{S}) \leq J(\tilde{S} \cup \{j\})$ is such that $r_j \geq \frac{\sum_{i \in \tilde{S}} r_i v_i}{2 + \sum_{i \in \tilde{S}} v_i}$, which cannot be true due to the optimality of $\tilde{S}$ for the $V(S)$. Therefore, $J(\tilde{S}) > J(\tilde{S} \cup \{j\})$ and $J(\tilde{S}) > J(\bar{S})$. So the superiority of $\tilde{S}$ gets proved, and it has $1/2$ performance guarantee. □

A.2 Proofs in Section 2.5

A.2.1 Proof of Theorem 2.5.2

Proof. We show the ordering result by contradiction. Suppose there exist two position $i-1$ and i such that $\alpha_{(i-1)} - z_{(i-1)} - w_{(i-1)} = \alpha_{(i-1)} - r_{(i-1)} < \alpha_{(i)} - r_{(i)} = \alpha_{(i)} - z_{(i)} - w_{(i)}$ but $\alpha_{(i-1)} - z_{(i-1)} > \alpha_{(i)} - z_{(i)}$. Without loss of generality, let product k and l be ranked at the $i-1_{st}$ and i_{th} position respectively. By assumption, we know $\alpha_k - z_k - w_k < \alpha_l - z_l - w_l$ but $\alpha_k - z_k > \alpha_l - z_l$. To have this order, it must be that the markups are ordered such that $w_k > w_l$. Now let us update product k's and l's markups to w'_k and w'_l such that the net utilities are switched, i.e.,

$$\alpha_k - z_k - w'_k = \alpha_l - z_l - w_l$$

$$\alpha_l - z_l - w'_l = \alpha_k - z_k - w_k$$

equivalently, we update the markups

$$w'_k = (\alpha_k - z_k) - (\alpha_l - z_l) + w_l = \Delta + w_l$$

$$w'_l = -(\alpha_k - z_k) + (\alpha_l - z_l) + w_k = -\Delta + w_k$$

where $\Delta > 0$ as $\alpha_k - z_k > \alpha_l - z_l$. The above two equations indicate that

$$w'_k - w_l = \Delta \quad \text{and} \quad w'_l - w_k = -\Delta$$

As this change of markup will switch the preference order of k and l, therefore

$$w'_{(i-1)} - w_{(i-1)} = \Delta \quad \text{and} \quad w'_{(i)} - w_{(i)} = -\Delta$$

$$\lambda'_{(i-1)} = \lambda_{(i-1)} \leq \lambda'_{(i)} = \lambda_{(i)}$$

Taking difference on expected profits after and before, the expected profit will change by

$$\prod_{j=i+1}^{n} (1 - \lambda_{(j)})\lambda_{(i)}\Delta - \prod_{j=i+1}^{n} (1 - \lambda_{(j)})(1 - \lambda_{(i)})\lambda_{(i-1)}\Delta > 0$$

This is because $\lambda_{(i)} > (1 - \lambda_{(i)})\lambda_{(i-1)}$ as $\lambda_{(i)} > \lambda_{(i-1)}$. Using the same logic, we can show that it is impossible that $a_{i-1} - r_{i-1} = a_i - r_i$ either. Therefore desired contradiction gets shown. The ordering of $a_i - r_i$ should be consistent with the ordering of $a_i - z_i$.

Next we show the that the optimal markups should be ordered in the same way as value gap $a_i - z_i$, i.e, $w_1 < w_2 < ... < w_n$. The argument will be again by contradiction. Suppose there exist two consecutive products $i - 1$ and i such that $w_{i-1} > w_i$ at optimality. From the proof of the first part, we are convinced that the sequence $a_i - z_i - w_i$ should have the same ordering as the sequence $a_i - z_i$, i.e., $a_{i-2} - z_{i-2} < a_{i-1} - z_{i-1} < a_i - z_i < a_{i+1} - z_{i+1}$ indicates $a_{i-2} - z_{i-2} - w_{i-2} < a_{i-1} - z_{i-1} - w_{i-1} < a_i - z_i - w_i < a_{i+1} - z_{i+1} - w_{i+1}$ under optimal markups. Therefore, we can increase w_i or decrease w_{i-1} or do both by a small amount without violating the ordering condition.

Now let us look into a sub-problem: fixing markups of all the other products, and ignoring all the constraints, let us just focus on the profit maximization problem with respect to product $i's$ markup:

$$\max_{w_i} \lambda_i w_i + (1-\lambda_i)\lambda_{i-1}w_{i-1} + (1-\lambda_i)(1-\lambda_{i-1})\lambda_{i-2}w_{i-2} + ... + \prod_{j=2}^{i}(1-\lambda_j)\lambda_1 w_1$$

Fixing all the other products' markups, then profits from products preferred before i will be fixed, as i exerts no cannibalization to them. Moreover, we can rewrite the above formula as

$$\max_{w_i} \lambda_i w_i + (1-\lambda_i)R(\{1,...,i-1\}) \equiv \max_{w_i} \lambda_i \times \Big(w_i - R(\{1,...,i-1\})\Big) + R(\{1,...,i-1\}) \tag{A.1}$$

Where $R(\{1,...,i-1\}) \equiv \lambda_{i-1}w_{i-1} + (1-\lambda_{i-1})\lambda_{i-2}w_{i-2} + ... + \prod_{j=2}^{i-1}(1-\lambda_j)\lambda_1 w_1$ is the profit from assortment $\{1,...,i-1\}$ under the fixed markups, so it is independent of w_i. As a consequence, w_i will affect no other terms except λ_i.

Now we check the first order condition of problem (A.1). Let w_i^* denote the optimal solution, then it must satisfy:

$$w_i^* = \theta\Big(i, R(\{1,...,i-1\})\Big) + R(\{1,...,i-1\}) + 1 \tag{A.2}$$

Where $\theta(i, R(\{1,...,i-1\})) = \lambda_i^* \times \Big(w_i^* - R(\{1,...,i-1\})\Big)$, i.e., the maximum value of $\lambda_i \times \Big(w_i - R(\{1,...,i-1\})\Big)$. Moreover, the objective (A.1) is unimodal with respect to w_i, i.e, it monotonically increases when $w_i < w_i^*$ and monotonically decreases when $w_i > w_i^*$. As increasing the value of w_i by a small amount can still preserve the optimal ordering condition, it must be true that $w_i \geq w_i^*$, otherwise we can just raise w_i and therefore increase the expected profit. Similarly, if we look into the problem of markup optimization on product $i-1$, we know that $i-1$ is facing an identical problem:

$$\max_{w_{i-1}} \lambda_{i-1} \times \Big(w_{i-1} - R(\{1,...,i-2\})\Big) + R(\{1,...,i-2\})$$

and the first order condition shows

$$w_{i-1}^* = \theta\Big(i-1, R(\{1,...,i-2\})\Big) + R(\{1,...,i-2\}) + 1 \tag{A.3}$$

will be a maximizer to the above problem, where $\theta\Big(i-1, R(\{1,...,i-2\})\Big) = \lambda_{i-1}^* \times \Big(w_{i-1}^* - R(\{1,...,i-2\})\Big)$. Since again, decreasing w_{i-1} is feasible, so it must be true that $w_{i-1} < w_{i-1}^*$.

According to the presumed relationship $w_{i-1} > w_i$ and above derivation, we should have $w_{i-1}^* > w_i^*$. We now show that this ordering is impossible. Recall the equilibrium formulations (A.2) and (A.3) which give the values of w^*_{i-1} and w^*_i, we have $R(\{1,...,i-2\}) < R(\{1,...,i-1\})$, so the rest concern is on the comparison between $\theta(i-1, R(\{1,...,i-2\}))$ and $\theta(i, R(\{1,...,i-1\}))$. As long as we can prove $\theta(i, R(\{1,...,i-1\})) - \theta(i-1, R(\{1,...,i-2\})) > R(\{1,...,i-2\}) - R(\{1,...,i-1\})$, we are able to show the contradiction. Since $\theta\Big(i, R(\{1,...,i-1\})\Big)$ is the optimal value of $\lambda_i \times \Big(w_i - R(\{1,...,i-1\})\Big)$ and $\theta\Big(i-1, R(\{1,...,i-2\})\Big)$ is the optimal value of $\lambda_{i-1} \times \Big(w_{i-1} - R(\{1,...,i-2\})\Big)$, we can easily show that the gap between these two values can be at most $R(\{1,...,i-2\}) - R(\{1,...,i-1\})$ by setting w_{i-1} to optimal, i.e., $w_{i-1} = w^*_{i-1}$ and assigning w_i a suboptimal value w^*_{i-1}. Therefore, it is impossible to have $w^*_{i-1} > w^*_i$, and thus impossible to have $w_{i-1} > w_i$. This proves the ordering of the optimal markups. $\square$

Appendix B

Appendix for Chapter 3: Approximation Algorithms for Product Framing and Pricing

B.1 Proofs in Section 3.4

B.1.1 Proof of Theorem 3.4.2.

Proof. Fix any instance of a 2-PARTITION problem with d numbers $w_1, w_2, ..., w_d$. We reduce this instance to a special case of our model with $m = 2$, $p = 200d^2$ and $n = 400d^2 + 2d$. The attractiveness of the 'no-purchase' option is 1. The prices and attractiveness of the n products are as follows:

- Each of the first d products corresponds to a number in the two-partition problem.

For $i = 1, 2, ..., d$, we set $r_i = 34$ and $\nu_i = M + \varepsilon w_i$, where ε is some small value which we will define shortly, and M is determined by ε via

$$2d \cdot M + \varepsilon \sum_{i=1}^{d} w_i = 2.$$

- For $i = d+1, d+2, ..., 2d$, we set $r_i = 34$ and $\nu_i = M$.
- For $i = 2d+1, 2d+2, ..., 3d$, we set $r_i = 59$ and $\nu_i = 2\delta$ where

$$\delta \equiv \frac{1}{p} = \frac{1}{200d^2}.$$

- For $i = 3d+1, 3d+2, ..., n$, we set $r_i = 59$ and $\nu_i = \delta$.

Given the special case constructed above, we argue that for an optimal frame, it is critical to decide which of the first $2d$ products with price 34 should be offered on the first page. We prove that the total expected revenue is a quasi-concave function of the total attractiveness of the first $2d$ products that are offered on the first page. In particular, if the 2-PARTITION problem has a solution, then we are able to recover that solution from the maximizer of the quasi-concave function (i.e., the optimal solution of our model). Therefore, we can solve the 2-PARTITION problem by optimizing the expected total revenue of our model.

We first observe the following structural properties of the special case of our model:

1. Since the revenues of all products are at most 59, the expected total revenue should be strictly less than 59 (due to the no-purchase option). Thus, it is never optimal to

leave any space in the two pages unfilled, as there are plenty of products with revenue 59.

2. It is easy to check that it always improves revenue to greedily replace a product with price 59 and attractiveness δ by, if any, a *spare* product with price 59 and attractiveness 2δ. Furthermore, whenever there is a product with price 59 and attractiveness δ on the first page and a product with price 59 and attractiveness 2δ on the second page, it is better to greedily swap the two products. Thus, products with price 59 and attractiveness 2δ should all be put on the first page.

3. Starting with an optimal solution, if we remove all the (at most $2d$) products with price 34 from the first two pages, we end up with at least $2p - 2d \approx 2p$ products with price 59 remaining in the two pages (among which only d products have attractiveness 2δ). It is easy to check that the resulting expected revenue is

$$\approx \frac{59 \cdot d \cdot 2\delta + 59 \cdot 2p \cdot \delta}{1 + d \cdot 2\delta + 2p \cdot \delta} \approx 39.33$$

for customers who view two pages. Thus, when we put those products with price 34 back into the solution, the expected revenue for customers who view two pages should be strictly greater than 34. This implies that in the optimal solution, no product with revenue 34 should be put on the second page, because 34 is lower than the expected revenue of the assortment consisting of products on the two pages.

In summary, the optimal solution must put some of the products with price 34 on the first page if any, all products with price 59 and attractiveness 2δ on the first page, and fill in all other spots using products with price 59 and attractiveness δ.

Let $S \subseteq \{1,2,...,2d\}$ denote the set of products with price 34 that are put in the first page. We set $\lambda(1) = \lambda(2) = 0.5$. The expected revenue under decision S is

$$\begin{aligned} R(S) =&\lambda(1)\frac{34\sum_{i\in S}\nu_i + 59\cdot 2\delta\cdot d + 59\cdot\delta\cdot(p-|S|-d)}{1+\sum_{i\in S}\nu_i + 2\delta\cdot d + \delta\cdot(p-|S|-d)} \\ &+\lambda(2)\frac{34\sum_{i\in S}\nu_i + 59\cdot 2\delta\cdot d + 59\cdot\delta\cdot(2p-|S|-d)}{1+\sum_{i\in S}\nu_i + 2\delta\cdot d + \delta\cdot(2p-|S|-d)} \\ =&0.5\frac{59(1-(|S|-d)\delta)+34\sum_{i\in S}\nu_i}{2-(|S|-d)\delta+\sum_{i\in S}\nu_i} + 0.5\frac{59(2-(|S|-d)\delta)+34\sum_{i\in S}\nu_i}{3-(|S|-d)\delta+\sum_{i\in S}\nu_i}. \end{aligned} \tag{B.1}$$

Now consider the following function defined for $l = 1,2,...,2d$ and $\theta \in \mathbb{R}$.

$$\begin{aligned} f(l,\theta) \equiv& 0.5\frac{59(1-(l-d)\delta)+34\left(\frac{l}{d}+\theta\right)}{2-(l-d)\delta+\frac{l}{d}+\theta} + 0.5\frac{59(2-(l-d)\delta)+34\left(\frac{l}{d}+\theta\right)}{3-(l-d)\delta+\frac{l}{d}+\theta} \\ =& 0.5\frac{59(1-\frac{l-d}{200d^2})+34\left(\frac{l}{d}+\theta\right)}{2-\frac{l-d}{200d^2}+\frac{l}{d}+\theta} + 0.5\frac{59(2-\frac{l-d}{200d^2})+34\left(\frac{l}{d}+\theta\right)}{3-\frac{l-d}{200d^2}+\frac{l}{d}+\theta}. \end{aligned}$$

By this definition, we have $R(S) = f(|S|, \sum_{i\in S}\nu_i - |S|/d)$ and $\lim_{\varepsilon\to 0} R(S) = f(|S|,0)$. Furthermore, we can prove the following properties for $f(\cdot,\cdot)$:

- $f(l,0)$ is quasi-concave in l for $l \geq 0$. When $d \geq 1$ and l is relaxed to a non-negative continuous variable, the only solution for $\frac{\partial f(l,0)}{\partial l} = 0$ is

$$l = \frac{-21d - 25000d^3 + 25d^2\left(-427+4\sqrt{-441-7350d+360000d^2}\right)}{7(3+25d)(-1+200d)} \geq 0.$$

It is easy to check that this is a local maximizer for $f(l,0)$. Therefore, $f(l,0)$ is quasi-concave in l for $l \geq 0$.

- When l can only take non-negative integral values, the maximizer for $f(l,0)$ is $l = d$. We can deduce that

$$f(d,0) - f(d-1,0) = \frac{-49 + 9600d + 5000d^2}{2(1-200d+600d^2)(1-200d+800d^2)} > \frac{1}{200d^2}, \ \forall d \geq 1,$$

$$f(d,0) - f(d+1,0) = \frac{-49 + 9600d + 75000d^2}{2(-1+200d+600d^2)(-1+200d+800d^2)} > \frac{1}{200d^2}, \ \forall d \geq 1.$$

Therefore, since $f(l,0)$ has at most one local maximizer for $l \geq 0$, $l = d$ must be the unique maximizer for $f(l,0)$ when l is a non-negative integer.

- When $l = d$ and $\theta \in [-1,1]$, the unique maximizer for $f(d,\theta)$ is $\theta = 0$ as shown by the following calculation.

$$f(d,\theta) = \frac{8}{\theta+4} - \frac{4.5}{\theta+3} + 34$$

$$\Longrightarrow \frac{\partial f(d,\theta)}{\partial \theta} = \frac{4.5}{(\theta+3)^2} - \frac{8}{(\theta+4)^2}.$$

When $\theta \in [-1,1]$,

$$\frac{\partial f(d,\theta)}{\partial \theta} = 0 \Longrightarrow \theta = 0.$$

It is easy to check that $\theta = 0$ is a maximizer for $f(d,\theta)$.

Thus,

$$f(d,\theta) < f(d,0), \ \ \forall\theta \neq 0, \theta \in [-1,1]. \tag{B.2}$$

- When $l \neq d$, $l \in \{0,1,2,...,2d\}$ and $\theta \in [-0.5, 0.5]$, we can deduce that

$$\begin{aligned}
f(l,\theta) =& 0.5\frac{59(1-\frac{l-d}{200d^2})+34\left(\frac{l}{d}+\theta\right)}{2-\frac{l-d}{200d^2}+\frac{l}{d}+\theta}+0.5\frac{59(2-\frac{l-d}{200d^2})+34\left(\frac{l}{d}+\theta\right)}{3-\frac{l-d}{200d^2}+\frac{l}{d}+\theta} \\
\leq& 0.5\frac{59(1-\frac{l-d}{200d^2})+34\frac{l}{d}}{2-\frac{l-d}{200d^2}+\frac{l}{d}+\theta}+0.5\cdot 34|\theta|+0.5\frac{59(2-\frac{l-d}{200d^2})+34\frac{l}{d}}{3-\frac{l-d}{200d^2}+\frac{l}{d}+\theta}+0.5\cdot 34|\theta| \\
& \text{(because } 2-\frac{l-d}{200d^2}+\frac{l}{d}+\theta \geq 1 \text{ and } 3-\frac{l-d}{200d^2}+\frac{l}{d}+\theta \geq 1\text{)} \\
=& 0.5\frac{59(1-\frac{l-d}{200d^2})+34\frac{l}{d}}{2-\frac{l-d}{200d^2}+\frac{l}{d}+\theta}+0.5\frac{59(2-\frac{l-d}{200d^2})+34\frac{l}{d}}{3-\frac{l-d}{200d^2}+\frac{l}{d}+\theta}+34|\theta| \\
\leq& 0.5\frac{59(1-\frac{l-d}{200d^2})+34\frac{l}{d}}{2-\frac{l-d}{200d^2}+\frac{l}{d}}\left(1+\frac{2|\theta|}{2-\frac{l-d}{200d^2}+\frac{l}{d}}\right)+ \\
& 0.5\frac{59(2-\frac{l-d}{200d^2})+34\frac{l}{d}}{3-\frac{l-d}{200d^2}+\frac{l}{d}+\theta}\left(1+\frac{2|\theta|}{3-\frac{l-d}{200d^2}+\frac{l}{d}}\right)+34|\theta| \\
& \text{(since } \theta \in [-0.5,0.5]\text{)} \\
\leq& 0.5\frac{59(1-\frac{l-d}{200d^2})+34\frac{l}{d}}{2-\frac{l-d}{200d^2}+\frac{l}{d}}(1+2|\theta|)+0.5\frac{59(2-\frac{l-d}{200d^2})+34\frac{l}{d}}{3-\frac{l-d}{200d^2}+\frac{l}{d}+\theta}(1+2|\theta|)+34|\theta| \\
=& f(l,0)+(2f(l,0)+34)|\theta| \\
\leq& f(l,0)+(2f(d,0)+34)|\theta|.
\end{aligned}$$

- Combining these results, for $l \neq d$, $l \in \{0,1,...,2d\}$, if

$$|\theta| \leq \frac{1}{2f(d,0)+34}\cdot\frac{1}{200d^2}, \tag{B.3}$$

we must always have $|\theta| < 0.5$ and

$$f(l,\theta) \leq f(l,0) + (2f(d,0)+34)|\theta| \leq f(l,0) + \frac{1}{200d^2} < f(d,0),$$

where the last inequality follows from the bound in the second bullet point.

In our model, we set

$$\varepsilon = \frac{1}{2\sum_{i=1}^{d} w_i} \cdot \frac{1}{2f(d,0)+34} \cdot \frac{1}{200d^2}.$$

Then we can bound the total expected revenue $R(S)$ using $f(d,0)$.

$$\begin{aligned}
R(S) &= f(|S|, \sum_{i\in S} \nu_i - |S|/d) \\
&= f(|S|, |S|M - |S|/d + \sum_{i\in S\cap\{1,2,\ldots,d\}} \varepsilon w_i) \\
&= f(|S|, |S|\frac{2-\varepsilon\sum_{i=1}^{d} w_i}{2d} - |S|/d + \varepsilon \sum_{i\in S\cap\{1,2,\ldots,d\}} w_i) \\
&= f(|S|, \varepsilon\left[\frac{-|S|\sum_{i=1}^{d} w_i}{2d} + \sum_{i\in S\cap\{1,2,\ldots,d\}} w_i\right]).
\end{aligned}$$

Since

$$\begin{aligned}
&\left|\varepsilon\left[\frac{-|S|\sum_{i=1}^{d} w_i}{2d}+\sum_{i\in S\cap\{1,2,\ldots,d\}} w_i\right]\right| \\
=&\left|\frac{1}{2\sum_{i=1}^{d} w_i}\cdot\frac{1}{2f(d,0)+34}\cdot\frac{1}{200d^2}\left[\frac{-|S|\sum_{i=1}^{d} w_i}{2d}+\sum_{i\in S\cap\{1,2,\ldots,d\}} w_i\right]\right| \\
\leq&\frac{1}{2\sum_{i=1}^{d} w_i}\cdot\frac{1}{2f(d,0)+34}\cdot\frac{1}{200d^2}\left[\sum_{i=1}^{d} w_i+\sum_{i\in S\cap\{1,2,\ldots,d\}} w_i\right] \\
\leq&\frac{1}{2f(d,0)+34}\cdot\frac{1}{200d^2},
\end{aligned}$$

we must have, according to (B.2) and (B.3),

$$R(S)=f(|S|,\sum_{i\in S}\nu_i-|S|/d)\begin{cases}=f(d,0),\ \text{if } |S|=d,\sum_{i\in S}\nu_i=1\\ <f(d,0),\ \text{otherwise.}\end{cases} \tag{B.4}$$

This implies that $R(S)=f(d,0)$ if and only if $|S|=d,\sum_{i\in S}\nu_i=1$. Furthermore, if $R(S)=f(d,0)$, then S is an optimal solution to our model (not vice versa).

Let S^* be an optimal solution to our model. We prove the theorem by showing that the following two conditions are equivalent:

- $R(S^*)=f(d,0)$, i.e., the optimal expected revenue of our model is $f(d,0)$.
- The 2-PARTITION problem has a solution.

The above equivalence shows that the ranking problem reduces to the 2-partition problem as follows. If we can solve the ranking problem, then we can find the optimal value. If the optimal value is $f(d,0)$, then we can conclude that there is a solution to the 2-partition problem. If the optimal value is not $f(d,0)$, then we can conclude that there is no solution to the 2-partition problem.

First, suppose the 2-PARTITION problem has a solution $T \subset \{1,2,...,d\}$ such that

$$\sum_{i\in T} w_i = \frac{1}{2}\sum_{i=1}^{d} w_i.$$

We construct a solution $S \subset \{1,2,...,2d\}$ to our model as

$$S = T \cup \{d+1, d+2, ..., 2d-|T|\}.$$

We can check that $|S| = d$ and

$$\sum_{i\in S} \nu_i = dM + \sum_{i\in T} \varepsilon w_i = dM + \frac{1}{2}\varepsilon \sum_{i=1}^{d} w_i = 1.$$

Therefore, according to (B.4) , S is an optimal solution to our model which gives expected revenue $R(S) = f(d,0)$.

On the other hand, suppose S^* is an optimal solution to our model and $R(S^*) = f(d,0)$. According to (B.4), we must have $|S^*| = d$ and $\sum_{i\in S^*} \nu_i = 1$, which gives

$$\sum_{i\in S} \nu_i = 1$$

$$\Longrightarrow |S|M + \sum_{i\in S\cap\{1,2,...,d\}} \varepsilon w_i = 1$$

$$\Longrightarrow d\frac{2-\varepsilon\sum_{i=1}^{d} w_i}{2d} + \sum_{i\in S\cap\{1,2,\ldots,d\}} \varepsilon w_i = 1$$

$$\Longrightarrow \sum_{i\in S\cap\{1,2,\ldots,d\}} w_i = \frac{1}{2}\sum_{i=1}^{d} w_i.$$

This proves that the 2-PARTITION problem has a solution. □

B.2 Proofs in Section 3.7

To uncover the structure an optimal solution to (3.7.2) we will first establish some elementary results concerning the function $g(x) = x\Lambda(x)$ which turns out to play an important role in the analysis.

Lemma B.2.1. *If X has an IFR distribution, then $g(x)$ is unimodal, and the largest maximizer, say y, is the smallest integer x such that $(x+1)h(x) > 1$. Moreover, if X is geometric with $h(x) = \theta$ for all $x \in M$, then $1/(y+1) < \theta \leq 1/y$.*

Proof of Lemma B.2.1

A little algebra shows that $g(x) \leq g(x+1)$ if and only if $h(x) \leq 1/(x+1)$. Since $h(x)$ is increasing and $1/(x+1)$ is decreasing, there is a smallest x, say y, such that $h(y) > 1/(y+1)$. Then, $h(x) \leq 1/(x+1)$ for all $x < y$ implies that $g(1) \leq \ldots \leq g(y)$. On the other hand, $h(x) > 1/(x+1)$ for all $x \geq y$, implies that $g(y) > g(y+1) > \ldots$. In the geometric case, $h(x) = \theta$ for all x implies that at $\theta = h(y-1) \leq 1/y$ and $\theta = h(y) > 1/(y+1)$, together imply that $1/(y+1) < \theta \leq 1/y$ as claimed.

Lemma B.2.2. *Let R and Λ be an optimal solution of* (3.7.2)*, and let* $y = \max \arg \max_{x \in M} R(x)\Lambda(x)$*. Then*

1. $R(x) = x\frac{R(y)}{y}$ *for all* $x \geq y$*; and*

2. $h(x) = h(y)$ *for all* $x \in M,\ x \geq y$*.*

The intuition behind the above result is that if the above properties are not satisfied, then we can find another set of values for $R(\cdot)$ and $\Lambda(\cdot)$ satisfying the constraints but generating a lower objective value in program (3.7.2). The result also hints at a geometric or truncated geometric distribution for values of $x \geq y$.

Proof of Lemma B.2.2.

Proof. Consider $x > y$. By Assumption B2, $R(x)/x \leq R(y)/y$, so $R(x) \leq xR(y)/y$ for all $x > y$. If any of the $R(x), x > y$ values is not at its upper bound, then we can rescale the values of $R(x), x > y$, and reduce $R(x), x \leq y$, while maintaining the constraint $E[R(X)] = 1$. But this impossible because it would contradict the optimality of Problem (3.7.2).

We next verify that $h(x) = h(y)$ for all $x \in \{y, \ldots, m\}$. This implies that $\Lambda(x+1)\Lambda(x-1) = \Lambda(x)^2$ for all $x = y, \ldots, m-1$. If not, there is a largest z such that $\Lambda(z+1)\Lambda(z-1) < \Lambda(z)^2$, and we can Increase $\Lambda(z+1), \ldots, \Lambda(m)$ by a small amount while maintaining the IFR property. This adjustment has the effect of increasing $E[R(X)]$. To maintain $E[R(X)] = 1$ we would need to scale down all the $R's$ and in the process reduce $R(y)\Lambda(y)$, again contradicting the optimality of Problem (3.7.2).

□

We will now show that $y = \max \arg \max_{x \in M} R(x)\Lambda(x)$ is also the largest maximizer of $g(x)$ in M.

Lemma B.2.3. *Let R and Λ be an optimal solution of* (3.7.2)*, then $y = \max \arg \max_{x \in M} R(x)\Lambda(x)$ is the largest maximizer of $g(x)$ in M.*

Proof of Lemma B.2.3.

Proof. By Lemma B.2.2, $R(x)/x = R(y)/y$ for all $x > y$. Since y is the largest maximizer of $R(x)\Lambda(x)$ it follows that $R(x)\Lambda(x) < R(y)\Lambda(y)$ for all $x > y$. Multiplying both sides by $x/R(x) = y/R(y)$, we obtain $x\Lambda(x) < y\Lambda(y)$ for all $x > y$. For all $x < y$, $R(x)\Lambda(x) \leq R(y)\Lambda(y)$ and $x/R(x) \leq y/R(y)$, multiplying the inequalities, we obtain $x\Lambda(x) \leq y\Lambda(y)$. Therefore, y is the largest maximizer of $g(x)$ over $x \in M$.

□

Let $r = R(y)\Lambda(y)$. We next show the structure of $R(x)$ for all $x \in \{1, \ldots, y\}$.

Lemma B.2.4. *$R(x) = r/\Lambda(x)$ for all $x \in \{1, \ldots, y\}$.*

Proof of Lemma B.2.4.

Proof. From $r = R(y)\Lambda(y)$, it follows that $R(x) \leq r/\Lambda(x)$ for all $x \in \{1, \ldots, y\}$. We will show that $R(x) = r/\Lambda(x)$ for all $x \in \{1, \ldots, y\}$. Suppose for a contradiction that this is false, and let z be the smallest $x \in \{1, \ldots, y\}$ such that $R(x) < r/\Lambda(x)$. We first argue that $z > 1$, for if $z = 1$, then we could increase $R(1)$ a bit, without disrupting $R(x)/x$ decreasing in x.

This would then allow us to scale down $R(x), x > 1$ and reduce $r = R(y)\Lambda(y)$, contradicting the optimality of Problem (3). Consequently, $z > 1$. We now argue that $R(z)/z = R(z-1)/(z-1)$. Otherwise, we could increase $R(z)$ and decrease $R(x)$, $x > z$, and again get a contradiction. Then from $r = R(z-1)\Lambda(z-1) > R(z)\Lambda(z) = R(z-1)z/(z-1)\Lambda(z)$, we obtain $g(z-1) > g(z) < g(y)$, but this contradicts the uni-modality of $g(x)$. Consequently, $R(x) = r/\Lambda(x)$ for all $x \in \{1, \ldots, y\}$.

□

This implies that the $R(x)\Lambda(x) = r$ for all $x \in \{1, \ldots, y\}$, so the maximizers of $R(x)\Lambda(x)$ are consecutive. It also implies that $R(x)\lambda(x) = rh(x)$ for all $x \in \{1, \ldots, y\}$.

B.2.1 Proof of Theorem 3.7.2.

Proof. We will work with the condition $1 = E[R(X)] = \sum_{x \in M} R(x)\lambda(x)$ to show that in the worst case the distribution of X is geometric with mean $E[X] = y$, and from this conclude that $r \geq 1/(3-2/y) \geq 1/3$. Let $h(y) = \theta$ for some $\theta > 0$ and recall from Lemma B.2.2 that $h(x) = \theta$ for all $x \in M, x \geq y$. This implies that X has tail probabilities $\Lambda(x) = (1-\theta)^{x-1}$ for all $x \in M, x \geq y$. For any such X, we have $E[X|X \geq y] = E[X|X > y-1] \leq y - 1 + E[X] = y - 1 + 1/\theta$ with the upper bound attained by the distribution with tail probabilities $\Lambda(x) = (1-\theta)^{x-1}$ for all $x \geq y$. Finally notice that $R(x) = r/\Lambda(x), x < y$, implies that

$R(x)\lambda(x) = rh(x) \leq r\theta$. Consequently,

$$\begin{aligned} 1 = E[R(X)] &= \sum_{x=1}^{y-1} R(x)\lambda(x) + \sum_{x>y-1} R(x)\lambda(x) \\ &= r\sum_{x=1}^{y-1} h(x) + \frac{r}{y}\sum_{x>y-1} x\frac{\lambda(x)}{\Lambda(y)} \\ &= r\sum_{x=1}^{y-1} h(x) + \frac{r}{y}E[X|X > y-1] \\ &\leq r\left((y-1)\theta) + \frac{1}{y}(y-1+1/\theta)\right) \end{aligned}$$

where the bound is attained by the geometric distribution with hazard rate $h(x) = \theta$ and mean $E[X] = 1/\theta$. Because Lemma B.2.1, it follows that $1/(y+1) < \theta \leq 1/y$. Since the expression is brackets is convex in θ, the maximum is at an extreme point and it is easy to see that it is maximized when $\theta = 1/y$, or equivalently when $E[X] = y$. Evaluating the last equation at $\theta = 1/y$, we see that $1 \leq r(3-2/y)$, or equivalently $r \geq 1/(3-2/y) \geq 1/3$.

□

B.3 Proofs in Section 3.9

B.3.1 Proof of Theorem 3.9.1

Proof. Given the sets $S_x,\ \forall x = 1,\ldots,m$ fixed, we maximize solely over prices. That is, we want to compute

$$\max_r R(r|S) = \sum_{x=1}^{m} \lambda(x)R_x(r|\mathcal{S}_x) = \sum_{x=1}^{m} \lambda(x)\sum_{i\in\mathcal{S}_x} r_i\pi(i,\mathcal{S}_x),$$

where $R_x(r|\mathcal{S}_x)$ is the expected revenue from consumer who has consideration set $\mathcal{S}_x$ and $\pi(i,\mathcal{S}_x) = \frac{\exp(a_i - \beta r_i)}{1+\sum_{k\in\mathcal{S}_x}\exp(a_k-\beta r_k)}$ is the probability of choosing item i if i is in consumer x's consideration set. Taking partial derivative of $\pi(i,\mathcal{S}_x)$ with respect to product i's and k's prices r_i and r_k respectively, we have the following formulas:

$$\frac{\partial\pi(i,\mathcal{S}_x)}{\partial r_i} = \beta\pi(i,\mathcal{S}_x)(\pi(i,\mathcal{S}_x)-1), \quad \text{and}$$

$$\frac{\partial\pi(i,\mathcal{S}_x)}{\partial r_k} = \beta\pi(i,\mathcal{S}_x)\pi(k,\mathcal{S}_x).$$

Taking the first order derivative of the expected revenue $R_x(r)$ with respect to r_i, we obtain

$$\begin{aligned}\frac{\partial R_x(r|\mathcal{S}_x)}{\partial r_i} &= \pi(i,\mathcal{S}_x) + r_i\frac{\partial\pi(i,\mathcal{S}_x)}{\partial r_i} + \sum_{k\neq i} r_k \frac{\partial\pi(k,\mathcal{S}_x)}{\partial r_i} \\ &= \beta\pi(i,\mathcal{S}_x)\{\frac{1}{\beta} + \sum_{k\in\mathcal{S}_x} r_k\pi(k,\mathcal{S}_x) - r_i\} \\ &= \beta\pi(i,\mathcal{S}_x)\{\frac{1}{\beta} + R_x(r) - r_i\}.\end{aligned}$$

Let $x(i) \equiv \{x : i \in S_x\}$. That is, $x(i)$ denotes the page where item i is displayed. Taking partial derivative of the total expected revenue with respect to r_i, we obtain

$$\frac{\partial R(r|\mathcal{S})}{\partial r_i} = \beta\sum_{j=x(i)}^{m}\lambda(j)\pi(i,\mathcal{S}_j)\{\frac{1}{\beta} + R_j(r|\mathcal{S}_j) - r_i\}. \tag{B.1}$$

Equivalently,

$$\sum_{j=x(i)}^{m} \lambda(j)\pi(i,\mathcal{S}_j)\{\frac{1}{\beta}+R_j(r|\mathcal{S}_j)\} = \sum_{j=x(i)}^{m} \lambda(j)\pi(i,\mathcal{S}_j)r_i. \tag{B.2}$$

Notice that equation (B.2) is satisfied either when $\pi(i,\mathcal{S}_j)=0,\ \forall j=1,...,m$, i.e, when $r_i=+\infty$ meaning item i is priced out of the market; or when

$$r_i = \frac{\sum_{j=x(i)}^{m} \lambda(j)\pi(i,\mathcal{S}_j)\{\frac{1}{\beta}+R_j(r|\mathcal{S}_j)\}}{\sum_{j=x(i)}^{m} \lambda(j)\pi(i,\mathcal{S}_j)} \tag{B.3}$$

It is easy to check that in the above equation (B.3), the right hand side is invariant for all i and k such that $x(i)=x(k)$, i.e, invariant for all products that are displayed on the same page. Thus, for every finitely priced item i, there should be a page-level invariant price $\theta_{x(i)}$, such that $r_i=\theta_{x(i)}$ at optimality.

For the monotonicity of the page-level price, notice that equation (B.3) tells us that the price is a weighted average of $\frac{1}{\beta}+R_j(r|\mathcal{S}_j)$, where $R_j(r|\mathcal{S}_j)$, the expected revenue from consumer j, must be nondecreasing. To see, suppose that there is a drop at j. This could only be because we set prices too low for products in page j. Then we would just raise their prices and therefore increase the expected revenue consumer j. Moreover, we argue the expected revenue from the later market can also be increased due to the unimodularity of $R_j(\cdot)$ function with respect to all $r's$ (See Gallego, Li and Beltran (2016)). And the saddle point will shift right as more products are considered. According to Theorem (3.9.3), indeed all products should be introduced, so $R_j(r|\mathcal{S}_j)$ must be monotonically increasing in j. Thus, the page-level price must be monotonically increasing in j.

To prove the optimal prices for all products are finite, we want to show two facts: (1) the optimal price for all pages, i.e., $\theta_1 \ldots \theta_m$ are finite; and (2) supposing that at optimality, some products are priced at infinity, then the item offering cannot be optimal.

The first fact is equivalent to having $\theta_m < +\infty$, since we know that the page-level price θ_x is monotonically increasing in x. The first-order condition given by page m implies $\theta_m = \frac{1}{\beta} + R_m(\theta|\mathcal{S}_m)$, which proves the finiteness of θ_m.

Consider the second fact that we need to prove. Notice that having some prices equal to infinity is equivalent to pricing the corresponding products out of the market, or not displaying them at all. Equivalently, there is some page x displaying fewer than p products. Therefore we can introduce one more item without increasing the total attractiveness of page x. This can be achieved by increasing products' prices on page x. In this situation, the expected revenue from consumer segment $x' = 1 \ldots x-1$ will not be affected, and the expected revenue from segment x and larger will strictly increase. That is, $R_{x'}(r|\mathcal{S}_{x'})$, $x' = x \ldots m$, will become larger. Thus, the total expected revenue will strictly increase when we fill all pages with finitely priced products.

$\square$

B.3.2 Proof of Theorem 3.9.3

Proof. Suppose that the page configuration S_x is fixed for all $x = 1,...,m$. Given the quality vector $a = (a_1,...,a_n)$, we want to optimize with respect to prices. That is, we want to solve

$$R(a) = \max_r \quad R(r|\mathcal{S}) \equiv \textstyle\sum_{x=1}^m \lambda(x) R_x(r|\mathcal{S}_x). \tag{B.4}$$

Now, suppose we can change the vector a. We wish to determine how the optimal revenue $R(a)$ will change. Let us take the partial derivative of $R(a)$ with respect to a_i.

$$\begin{aligned}\frac{\partial R(a)}{\partial a_i} &= \sum_{j=x(i)}^{m} \lambda(j) \frac{\partial R_j(r(a)|\mathcal{S}_j)}{\partial a_i} \\ &= \sum_{j=x(i)}^{m} \lambda(j)\{r_i(a) - R_j(r(a)|\mathcal{S}_j)\}\pi(i,\mathcal{S}_j),\end{aligned} \tag{B.5}$$

where $r(a)$ is the optimal price vector under quality vector a, which must satisfy the first-order condition given by (B.3), i.e, $\sum_{j=x(i)}^m \lambda(j)\pi(i,\mathcal{S}_j) r_i(a) = \sum_{j=x(i)}^m \lambda(j)\pi(i,\mathcal{S}_j)\{\frac{1}{\beta} + R_j(r(a)|\mathcal{S}_j)\}$. Plug this relationship into (B.5), we obtain

$$\frac{\partial R(a)}{\partial a_i} = \frac{1}{\beta} \sum_{j=x(i)}^{m} \lambda(j)\pi(i,\mathcal{S}_j). \tag{B.6}$$

From the above relationship, we can draw several conclusions. First, the optimal revenue $R(a)$ is increasing in a_i for all $i = 1,...,n$. Second, suppose there are two item i and k with the same quality $a_i = a_k$, but $x(i) < x(k)$, in other words, i is displayed before k so it can be viewed by more consumer. Equation (B.6) tells us $\frac{\partial R(a)}{\partial a_i} > \frac{\partial R(a)}{\partial a_k}$, so if we can either increase a_i or a_k, it is always better to prioritize a_i. This indicates that products should be displayed in decreasing order of their quality.

To prove that all products should be displayed, notice that eliminating item i is equivalent to letting its quality a_i go to negative infinity. However, since $\frac{\partial R(a)}{\partial a_i} > 0$ and it is always better to put products with higher quality ahead of products with lower quality, we can conclude that at optimality, all products should be displayed; and they are displayed in the order of their indices. Each page is filled up with products until its capacity is saturated.

□

Appendix C

Appendix for Chapter 4: Product Line Design and Pricing under Logit Models

C.1 Proofs in Section 4.1

C.1.1 Proof of Theorem 4.1.1.

Proof. Given the potential set of products S, we purely maximize over prices:

$$\max_r R(r|S) = \sum_{j\in S}(r_j - \tilde{c}_j)\pi_i(S)$$

where $\pi_j(S) = \frac{exp(a_j-\beta_r r_j)}{V_0+\sum_{j\in S} exp(a_j-\beta_r r_j)}$, is the probability of choosing product j from set S under price vector r. Taking partial derivative of product i's and product j's choice probabilities with respect to i's price, we have the following formulas:

$$\frac{\partial \pi_i(S)}{\partial r_i} = \beta_r \pi_i(S)(\pi_i(S) - 1)$$

$$\frac{\partial \pi_j(S)}{\partial r_i} = \beta_r \pi_i(S)\pi_j(S)$$

Take first order derivative of expected profit $R(p)$ with respect to product i's price p_i:

$$\begin{aligned}\frac{\partial R(r|S)}{\partial r_i} &= \pi_i(S) + (r_i - \tilde{c}_i)\frac{\partial \pi_i(S)}{\partial r_i} + \sum_{j \neq i}(r_j - \tilde{c}_j)\frac{\partial \pi_j(S)}{\partial r_i} \\ &= \beta_r \pi_i(S)\{\frac{1}{\beta_r} + \sum_{j \in S}(r_j - \tilde{c}_j)\pi_j(S) - (r_i - \tilde{c}_i)\} \end{aligned} \tag{C.1}$$

From (C.1) we see, to satisfy the first order condition (FOC), either $\pi_i(S) = 0$, meaning $r_i = +\infty$, i.e., product i is priced out of market; Or the second term of (C.1) is equal to 0, i.e.,

$$r_i - \tilde{c}_i = \frac{1}{\beta_r} + \sum_{j \in S}(r_j - \tilde{c}_j)\pi_j(S) \tag{C.2}$$

In the above equation, the right hand side is invariant for all i, meaning for every finitely priced product i, there should be a constant mark-up, let us call it θ, such that $r_i = \tilde{c}_i + \theta$ at optimality.

Assuming the set S^* of products with finite price is a proper subset of S, then there should be at least one product $i \in S \setminus S^*$ not offered. The expected profit when set S^* is offered at prices with constant mark-up θ is:

$$R^{S^*}(\theta) = \sum_{j \in S^*} \theta \frac{exp(\alpha_j - \beta_r \tilde{c}_j - \beta_r \theta)}{V_0 + \sum_{j \in S^*} exp(\alpha_j - \beta_r \tilde{c}_j - \beta_r \theta)}$$

The expected profit by offering assortment $S^* \cup \{i\}$ at prices with the same mark-up is strictly higher. Therefore it is optimal to offer all the available products with a constant mark-up.

So the problem can be simplified to the one dimensional optimization problem

$$\max_{\theta} R(\theta|S) = \frac{\theta \sum_{i \in S} exp(\alpha_i - \beta_r(\theta + \tilde{c}_i))}{V_0 + \sum_{i \in S} exp(\alpha_i - \beta_r(\theta + \tilde{c}_i))}$$

and we take first order derivative of $R(\theta|S)$ with respect to θ, then

$$\frac{\partial R(\theta|S)}{\partial \theta} = (1 - \pi_S{}^{\theta}(0))(1 - \beta\theta\pi_S{}^{\theta}(0))) \tag{C.3}$$

In order to satisfy the first order condition, the right hand side of equation (C.3) should be 0, yet the first part can never be 0 since $\pi_0(S) = 1$ is impossible; Therefore it must be the case that at optimality:

$$\pi_0{}^{\theta}(S) = \frac{1}{\theta\beta_r}$$

Observe that second term of the right hand side of equation (C.3) $1 - \beta\theta\pi_S{}^{\theta}(0))$ monotonically increases with respect to θ, there must be a unique sign change of (C.3) from positive to negative. So it can be concluded that $R(\theta)$ is unimodal in θ. □

C.1.2 Proof of Theorem 4.1.3

Proof. First, let us assume that the configuration of $S \setminus \{i\}$ is fixed, and all the features of i except for feature m have picked a level. This indicates we only need to select one level from feature m to product i.

Let us compare two potential levels – k and q of feature m: with feature-level valuation $\beta_k > \beta_q$ and costs $c_k > c_q$ (otherwise the problem is trivial, just select the level who has higher valuation and lower cost). The the profit under price r_i and feature level k configuration is:

$$R(k, r_i) = \frac{R_{-i} + (r_i - c_0 - c_k)exp(\alpha_0 + \beta_k - \beta_r r_i)}{V_{-i} + exp(\alpha_0 + \beta_k - \beta_r r_i)}$$

And the profit under price p'_i and feature level q configuration is:

$$R(q, p'_i) = \frac{R_{-i} + (r'_i - c_0 - c_q)exp(\alpha_0 + \beta_q - \beta_r r'_i)}{V_{-i} + exp(\alpha_0 + \beta_q - \beta_r r'_i)}$$

Where $R_{-i} = \sum_{j \neq i}(r_j - \tilde{c}_j)exp(\beta_0 + \beta' x_j - \beta_r r_j)$ and $V_{-i} = \sum_{j \neq i} exp(\beta_0 + \beta' x_j - \beta_r r_j)$. Since have assumed that configurations and prices of all products except for i are fixed, R_{-i} and V_{-i} are just constant. Let c_0 be the cost of product i's current configuration x_i^0, i.e., $c_0 = c' x_i^0$ and a_0 be its current configuration's expected quality evaluation, i.e., $a_0 = \beta_0 + \beta' x_i^0$. Those are also constant.

Now let p_i be the price such that product i has same utility before and after, i.e.,

$$\alpha_0 + \beta_k - \beta_r r_i = \alpha_0 + \beta_q - \beta_r r'_i$$

$$\Leftrightarrow$$

$$r_i = \frac{\beta_k - \beta_q}{\beta_r} + p_i'$$

This will guarantee the two configurations have the same total market share. Replace r_i by that quantity, then

$$R(k, r_i) = \frac{R_{-i} + (r_i' - c_0 - c_k + \frac{\beta_k - \beta_q}{\beta_r}) exp(\alpha_0 + \beta_q - \beta_r r_i')}{V_{-i} + exp(\alpha_0 + \beta_q - \beta_r r_i')}$$

Compare it with $R(q, r_i)$, then

$$R(k, r_i) \geq R(q, r_i')$$

$$\Leftrightarrow$$

$$r_i' - c_0 - c_k + \frac{\beta_k - \beta_q}{\beta_r} \geq r_i' - c_0 - c_q$$

$$\Leftrightarrow$$

$$\frac{\beta_k}{\beta_r} - c_k \geq \frac{\beta_q}{\beta_r} - c_q$$

So as long as $\frac{\beta_k}{\beta_r} - c_k \geq \frac{\beta_q}{\beta_r} - c_q$, under whatever price of configuration q, we can always find another price, so the profit under k will strictly dominate.

□

C.1.3 Proof of Theorem 4.1.4

Proof. Assume the current assortment S with products $i, j \in S$ is fixed. Except that a feature-level k with positive adjusted value gap $r_k = \beta_k - \beta_r c_k$ is available to either i or j. Assuming that under current configuration, value gap of product i: $\tilde{v}_i = \beta_0 + \sum_l g_l x_{i,l}$ is greater than product j's $\tilde{v}_j = \beta_0 + \sum_l g_l x_{j,l}$, the question is: should we add the feature-level k to i or to j?

Recall theorem 4.1.1 – the constant adjusted mark-up property, so for each configuration, we are actually solving the following program:

$$\max_{\theta} R(\theta, i) = \theta \frac{V_{-ij}(\theta) + exp(\tilde{v}_i - \beta_r\theta + g_k) + exp(\tilde{v}_j - \beta_r\theta)}{V_0 + V_{-ij}(\theta) + exp(\tilde{v}_i - \beta_r\theta + g_k) + exp(\tilde{v}_j - \beta_r\theta)}$$

or

$$\max_{\theta} R(\theta, j) = \theta \frac{V_{-ij}(\theta) + exp(\tilde{v}_i - \beta_r\theta) + exp(\tilde{v}_j - \beta_r\theta + g_k)}{V_0 + V_{-ij(\theta)} + exp(\tilde{v}_i - \beta_r\theta) + exp(\tilde{v}_j - \beta_r\theta + g_k)}$$

Since we have assumed all the other products' configurations are fixed then the total valuation of them $V_{-ij}(\theta) = \sum_{i' \neq i,j} exp(\tilde{v}_{i'} - \beta_r\theta)$ is just a function on θ. And recall that V_0 is the valuation of outside alternative.

Let us price both assortment at the same θ and compare their profits, which is equivalent to comparing:

$$\frac{V_{-ij} + e^{u_i + z} + e^{u_j}}{V_0 + V_{-ij} + e^{u_i + z} + e^{u_j}}$$

with

$$\frac{V_{-ij}+e^{u_i}+e^{u_j+z}}{V_0+V_{-ij}+e^{u_i}+e^{u_j+z}}$$

They are simplified by defining constants$V_{-ij}=V_{-ij}(\theta)$, $u_i=\tilde{v}_i-\beta_r\theta$, $u_j=\tilde{v}_j-\beta_r\theta$ and $z=g_k$. Given that $u_i>u_j$ and $z>0$, we can easily prove that $R(\theta,i)>R(\theta,j)$. So we can see the higher adjusted value gaped product tends to attract more to its value gap to increase the profit, which proves our theorem. □

Proof of PROPOSITION 4.1.7.

Proof. Assume in the current assortment, we put K-optimal products; According to Theorem 4.1.1, the optimal markup θ_K should satisfy the equation:

$$\theta_K=R(\theta_K)+\frac{1}{\beta_r}$$

i.e., the optimal constant markup should be equal to the total profit under this pricing strategy and the reciprocal of price sensitivity. Also by Theorem 4.1.1, we also have at optimality:

$$\pi_0(K)=\frac{1}{\theta_K\beta_r}$$

Combining the above two results:

$$\begin{aligned}\theta_K&=R(\theta_K)+\frac{1}{\beta_r}\\&=\frac{\sum_{i=1}^{K}exp(\tilde{v}_i-\beta_r\theta_K)+1}{\beta_r}\end{aligned}$$

Therefore,

$$R(\theta_K) = \frac{\sum_{i=1}^{K} exp(\tilde{v}_i - \beta_r \theta_K)}{\beta_r}$$

So

$$\begin{aligned} R(\theta_{K+1}) - R(\theta_K) &= \frac{\sum_{i=1}^{K+1} exp(\tilde{v}_i - \beta_r \theta_{K+1}) - \sum_{i=1}^{K} exp(\tilde{v}_i - \beta_r \theta_K)}{\beta_r} \\ &< \frac{exp(\tilde{v}_K - \beta_r \theta_K)}{\beta_r} \end{aligned}$$

Since $\theta_K < \theta_{K+1}$. Hence,

$$\frac{R(\theta_{K+1}) - R(\theta_K)}{R(\theta_K)} < \frac{exp(\tilde{v}_{K+1} - \beta_r \theta_K)}{\beta_r \theta_K - 1}$$

And for any larger set $\bar{K}$, the ratio:

$$\begin{aligned} \frac{R(\theta_{\bar{K}}) - R(\theta_K)}{R(\theta_K)} &< \frac{\sum_{i=K+1}^{\bar{K}} exp(\tilde{v}_i - \beta_r \theta_K)}{\beta_r \theta_K - 1} \\ &= \frac{\sum_{i=K+1}^{\bar{K}} exp(\tilde{v}_i - \beta_r \theta_K)}{\sum_{i=1}^{K} exp(\tilde{v}_i - \beta_r \theta_K)} \\ &= \frac{\sum_{i=K+1}^{\bar{K}} exp(\tilde{v}_i)}{\sum_{i=1}^{K} exp(\tilde{v}_i)} \end{aligned}$$

□

C.2 Proofs in Section 4.2

C.2.1 Proof of Theorem 4.2.1

Proof. Before proving the theorems, let us look into the differential equations of choice probabilities (4.2.1), (4.2.2) and (4.2.3) with respect to prices, which will be used iteratively in the later proofs.

The partial derivative of product *in*'s conditional choice probability with respect to its own price is:

$$\frac{\partial \pi(in|n,S)}{\partial r_{in}} = \frac{\beta_r^n}{\gamma_n}\pi(in|n,S)(\pi(in|n,S)-1)$$

The partial derivative of product *in*'s conditional choice probability with respect to the price of another product in the same submarket is:

$$\frac{\partial \pi(in|n,S)}{\partial r_{jn}} = \frac{\beta_r^n}{\gamma_n}\pi(in|n,S)\pi(j|n)$$

The partial derivative of product *in*'s conditional choice probability with respect to another product from a different submarket is:

$$\frac{\partial \pi(in|n,S)}{\partial r_{kn'}} = 0$$

The partial derivative of nest n's market share with respect to the price of one of its products is:

$$\frac{\partial \pi(n|S)}{\partial r_{in}} = \beta_r^n \pi(in|S)(\pi(n|S)-1)$$

The partial derivative of nest *n*'s market share with respect to the price of a product in a different submarket is:

$$\frac{\partial \pi(n|S)}{\partial r_{kn'}} = \beta_r^n \pi(n|S)\pi(kn'|S)$$

The partial derivative of product *in*'s market share with respect its own price is:

$$\frac{\partial \pi(in|S)}{\partial\, r_{in}} = \frac{\beta_r^n}{\gamma_n}\pi(in|S)(\pi(in|n,S)-1)+\beta_r^n\pi(in|S)^2-\beta_r^n\pi(in|n,S)\pi(in|S)$$

The partial derivative of product *in*'s market share with respect the price of another product from the same submarket is:

$$\frac{\partial \pi(in|S)}{\partial r_{jn}} = \frac{\beta_r^n}{\gamma_n}\pi(in|n,S)\pi(jn|S)+\beta_r^n\pi(in|S)\pi(jn|S)-\beta_r^n\pi(jn|S)\pi(in|n,S)$$

The partial derivative of product *in*'s market share with respect the price of another product from the different submarket is:

$$\frac{\partial \pi(in|S)}{\partial r_{jn'}} = \beta_r^n \pi(in|S)\pi(jn'|S)$$

Plug the above results into the partial derivative of the expected profit with respect to product *jn*'s price:

$$\frac{\partial R(r|S)}{\partial r_{jn}} = \sum_{n' \neq n} \sum_{i} (r_{in'} - \tilde{c}_{in'}) \frac{\partial \pi(in'|S)}{\partial r_{jn}} + \sum_{i \neq j} (r_{in} - \tilde{c}_{in}) \frac{\partial \pi(in|S)}{\partial r_{jn}} + \pi(jn|S) + (r_{jn} - \tilde{c}_{jn}) \frac{\partial \pi(jn|S)}{\partial r_{jn}}$$

$$= \beta_r^n \pi(jn|S) \{ \sum_{n'} \sum_{i} (r_{in'} - \tilde{c}_{in'}) \pi(in'|S) + (\frac{1}{\gamma_n} - 1) \sum_{i} (r_{in} - \tilde{c}_{in}) \pi(in|n,S) + \frac{1}{\beta_r^n} - \frac{r_{jn} - \tilde{c}_{jn}}{\gamma_n} \} \tag{C.1}$$

In order to satisfy the first order condition, (C.1) must equal to 0, which can be achieved either by letting $\pi(jn|S) = 0 \Leftrightarrow r_{jn} = +\infty$ or the second term of (C.1) be 0, i.e.,

$$\sum_{n'} \sum_{i} (r_{in'} - \tilde{c}_{in'}) \pi(in'|S) + (\frac{1}{\gamma_n} - 1) \sum_{i} (r_{in} - \tilde{c}_{in}) \pi(in|n,S) + \frac{1}{\beta_r^n} - \frac{r_{jn} - \tilde{c}_{jn}}{\gamma_n} = 0 \tag{C.2}$$

Observe that the first term on the left hand side of equation (C.2) is the expected profit $R(r|S)$; and the second term, which can be denoted by $R(r|n)$, is invariant for all products belonging to the nest *n*, so (C.2) is equivalent to

$$\gamma_n R(r|S) + (1 - \gamma_n) R(r|n) + \frac{\gamma_n}{\beta_r^n} = r_{jn} - \tilde{c}_{jn} \tag{C.3}$$

The left hand side of the above equation is invariant within nest *n*, so there must be a

θ_n for each nest n such that,

$$r_{jn} = \theta_n + \tilde{c}_{jn} \tag{C.4}$$

$$\theta_n = \gamma_n R(p) + (1-\gamma_n) R(p|n) + \frac{\gamma_n}{\beta_r^n} \tag{C.5}$$

Assume that at optimality, S_n^*, set products in nest n priced with the constant markup is a proper subset of the available set of products in nest n, denoted by S_n, i.e., there exists at least one product in such that $in \in S_n \setminus S_n^*$.

Let us fix the optimal price for all nests except nest n, and denote it by vector r_{-n}^*; To be more specific, let the optimal price vector of nest n' be denoted as $r_{n'}^*$. Let the adjusted attractiveness of nest n' be $a(r_{n'}^*) = (\sum_j exp(\frac{\alpha_{jn'} - \beta_r^{n'} r_{jn'}^*}{\gamma_{n'}}) + V_{n'})^{\gamma_{n'}}$, which is a constant under the fixed optimal price vector $r_{n'}^*$. And denote the adjusted attractiveness over all nests except n be $\tilde{a}(r_{-n}^*) = \sum_{n' \neq n} a(r_{n'}^*)$. And at optimality, another parameter ρ^*, defined as

$$\rho^* = \frac{\tilde{a}(r_{-n}^*) + (\sum_{j \in S_n^*} exp(\frac{\alpha_{jn} - \beta_n^p r_{jn}^*}{\gamma_n}) + V_n)^{\gamma_n}}{\tilde{a}(r_{-n}^*) + (\sum_{j \in S_n^*} exp(\frac{\alpha_{jn} - \beta_n^p r_{jn}^*}{\gamma_n}) + V_n)^{\gamma_n} + 1} \tag{C.6}$$

should be fixed also. We can interpret ρ as the total nests' market share.

Equation (C.6) is equivalent to

$$(\sum_{j \in S_n^*} exp(\frac{\alpha_{jn} - \beta_n^p r_{jn}^*}{\gamma_n}) + V_n)^{\gamma_n} = \frac{\rho^*}{1-\rho^*} - \tilde{a}(r_{-n}^*) \tag{C.7}$$

According to the constant markup property, there must be a $\theta_n(S_n^*)$ such that $r_{jn} =$

$\tilde{c}_{jn} + \theta_n(S_n^*)$ at optimality, and by replacing $\theta_n(S_n^*)$ into equation (C.7), we see that $\theta_n(S_n^*)$ should satisfy

$$(\sum_{j \in S_n^*} exp(\frac{\alpha_{jn} - \beta_n^p(\theta_n(S_n^*) + \tilde{c}_{jn})}{\gamma_n}) + V_n)^{\gamma_n} = \frac{\rho^*}{1-\rho^*} - \tilde{a}(r_{-n}^*) \tag{C.8}$$

Let us look into optimal profit function in terms of ρ, θ_n and r_{-n}

$$R(r^*|S) = \theta_n(S_n^*)(\rho^* - (1-\rho^*)\tilde{a}(r_{-n}^*)) \frac{(\frac{\rho^*}{1-\rho^*} - \tilde{a}(r_{-n}^*))^{\frac{1}{\rho^*}} - V_n}{(\frac{\rho^*}{1-\rho^*} - \tilde{a}(r_{-n}^*))^{\frac{1}{\rho^*}}} \tag{C.9}$$

$$+ \sum_{n' \neq n} (1-\rho^*) a(r_{n'}^*) \sum_i (r_{in'}^* - \tilde{c}_{in'}) \frac{exp(\frac{\alpha_{in'} - \beta_r^{n'} r_{in'}^*}{\gamma_{n'}})}{a(r_{n'}^*)^{\frac{1}{\gamma_{n'}}}} \tag{C.10}$$

By fixing ρ^* and r_{-n}^*, the objective value will monotonically increase only with respect to $\theta_n(S_n^*)$. Observe from equation (C.8), when we expand the set to $S_n^* \cup i$, θ_n's value will strictly increase to keep the equation hold, i.e., $\theta_n(S_n^* \cup i^*) > \theta_n(S_n^*)$, which will strictly increase the expected profit and that contradicts to the optimality of S_n^*. Therefore, all the available products within a same nest should be priced with the same markup.

□

C.2.2 Proof of Theorem 4.2.2

Proof. Let us first look into some partial differential equations of the choice probability defined in (4.2.9) and (4.2.10) with respect to the nest level constant markup.

The partial derivative of product *in*'s conditional choice probability with respect to nest *n*'s markup θ_n:

$$\frac{\partial\pi(in|n,S)}{\partial\theta_n} = -b_n\pi(in|n,S)\pi(S_0|n,S)$$

The partial derivative of nest *n*'s market share with respect to its own markup θ_n:

$$\frac{\partial\pi(n|S)}{\partial\theta_n} = \beta_r^n\pi(S_n|S)(\pi(n|S)-1)$$

The partial derivative of nest *n*'s market share with respect to another submarket *n*'s markup $\theta_{n'}$:

$$\frac{\partial\pi(n|S)}{\partial\theta_{n'}} = \beta_p^{n'}\pi(n|S)\pi(S_{n'}|S)$$

Recall that $\pi(S_n|S) = \sum_i \pi(in|S)$, $\pi(S_n|n,S) = \sum_i \pi(in|n,S)$, $\pi(S_0|n,S) = 1-\sum_i \pi(in|n,S)$.

Take partial derivative of the expected profit $R(\theta|S)$ with respect to nest *n*'s markup θ_n:

$$\frac{\partial R(\theta|S)}{\partial\theta_n} = \beta_r^n\pi(S_n|S)\{\frac{1}{\beta_r^n} + \sum_n \theta_n\pi(S_n|S) - \theta_n\pi(S_n|n,S) - \frac{\theta_n}{\gamma_n}\pi(S_0|n,S)\} \tag{C.11}$$

If nest *n* is offered, meaning $\pi(S_n|S) > 0$, then to satisfy the first order condition, the second term of the right hand side of equation (C.11) must be 0, i.e.

$$R(\theta|S) = \theta_n\pi(S_n|n,S) + \frac{\theta_n}{\gamma_n}\pi(S_0|n,S) - \frac{1}{\beta_r^n} \tag{C.12}$$

Observe that the left hand side is nest invariant, therefore all we need to prove now is

$\pi(S_n|S) > 0$ for every nest n at optimality, which is equivalent to all optimal mark-ups θ_n being finite.

Preparatory Material

Lemma C.2.1. *Problem (4.2.11) can alternatively be solved by computing the fixed point of an appropriately defined scaler function:*

$$\eta = \max_{\theta} \sum_n \{\theta_n (\sum_j e^{\tilde{v}_{jn} - b_n\theta_n} + V_n)^{\gamma_n} \frac{\sum_j exp(\tilde{v}_{jn} - b_n\theta_n)}{\sum_j exp(\tilde{v}_{jn} - b_n\theta_n) + V_n} - (\sum_j e^{\tilde{v}_{jn} - b_n\theta_n} + V_n)^{\gamma_n}\eta\} \tag{C.13}$$

This is due to Rayfield et al. 2015.

Proof. Proof of Lemma C.2.1. First observe that the left hand side to equation (C.13) is strictly increasing with respect to η and the right hand side is strictly decreasing with respect to η. When $\eta = 0$, the left hand side is strictly less than the right. Those conditions guarantee the uniqueness and existence of the fixed point satisfying (C.13). Letting $\hat{\eta}$ denote the value of η satisfying (C.13), we can argue that $\hat{\eta}$ is the optimal objective value of problem (C.13). To see this, let us denote the optimal solution to problem (4.2.11) by $(\theta_1^*, \theta_2^*, .., \theta_n^*)$, then we must have

$$\hat{\eta} \geq \sum_n \{\theta_n^* (\sum_j e^{\tilde{v}_{jn} - b_n\theta_n^*} + V_n)^{\gamma_n} \frac{\sum_j exp(\tilde{v}_{jn} - b_n\theta_n^*)}{\sum_j exp(\tilde{v}_{jn} - b_n\theta_n^*) + V_n} - (\sum_j e^{\tilde{v}_{jn} - b_n\theta_n^*} + V_n)^{\gamma_n}\hat{\eta}\}$$

This is because $\hat{\eta}$ is the value satisfying ((C.13)) and θ^* is a feasible solution but not necessarily an optimal solution to the maximization problem at the right hand side of ((C.13)) under $\eta = \hat{\eta}$. If we collect the term $\hat{\eta}$ in the above inequality, we get $\hat{\eta} \geq R(\theta^*)$, i.e., $\hat{\eta}$'s value is no smaller than the optimal profit in problem (4.2.11).

Furthermore, since $\hat{\eta}$ satisfies the equation in ((C.13)), there must exist a vector $(\hat{\theta}_1, \hat{\theta}_2, .., \hat{\theta}_n)$ satisfying the equation:

$$\hat{\eta} = \sum_n \{\hat{\theta}_n (\sum_j e^{\tilde{v}_{jn} - b_n \hat{\theta}_n} + V_n)^{\gamma_n} \frac{\sum_j exp(\tilde{v}_{jn} - b_n \hat{\theta}_n)}{\sum_j exp(\tilde{v}_{jn} - b_n \hat{\theta}_n) + V_n} - (\sum_j e^{\tilde{v}_{jn} - b_n \hat{\theta}_n} + V_n)^{\gamma_n} \hat{\eta}\}$$

Collect the term $\hat{\eta}$ again in the above equation, then we will find $\hat{\eta} = R(\hat{\theta}) \leq R(\theta^*)$, since θ^* optimizes (4.2.11). Therefore, the lemma gets proved.

□

Now we are ready to prove that all θ's value is finite at optimality.

An obvious advantage of formulation in (C.13) is that the whole maximization problem in (4.2.11) can be separated into a single dimensional optimization problem for each nest, e.g., for nest n, given η's value, it solves the following problem:

$$\eta_n(\theta_n^*) = \max_{\theta_n} \theta_n (\sum_j e^{\tilde{v}_{jn} - b_n \theta_n} + V_n)^{\gamma_n} \frac{\sum_j exp(\tilde{v}_{jn} - b_n \theta_n)}{\sum_j exp(\tilde{v}_{jn} - b_n \theta_n) + V_n} - (\sum_j e^{\tilde{v}_{jn} - b_n \theta_n} + V_n)^{\gamma_n} \eta \tag{C.14}$$

If problem (C.14) has finite maximizer under all values of η, then obviously all optimal mark-ups in problem (4.2.11) will be finite.

Observe that when $\theta_n \to +\infty$, (C.14)'s objective value $\eta_n(+\infty) = -{V_n}^{\gamma_n}\eta$; Therefore, a finite maximizer of (C.14) exists if and only if we can find a finite θ_n satisfying $\eta(\theta_n) > \eta(+\infty)$, i.e., there exists a $\theta_n \in [0, +\infty)$ such that

$$\theta_n(\sum_j e^{\tilde{v}_{jn}-b_n\theta_n} + V_n)^{\gamma_n} \frac{\sum_j exp(\tilde{v}_{jn} - b_n\theta_n)}{\sum_j exp(\tilde{v}_{jn} - b_n\theta_n) + V_n} - (\sum_j e^{\tilde{v}_{jn}-b_n\theta_n} + V_n)^{\gamma_n}\eta$$
$$>$$
$$-{V_n}^{\gamma_n}\eta \quad \text{(C.15)}$$

Collecting the term of η, we see (C.15) is equivalent to

$$\theta_n \frac{\sum_j exp(\tilde{v}_{jn} - b_n\theta_n)}{\sum_j exp(\tilde{v}_{jn} - b_n\theta_n) + V_n} > (1 - (\frac{V_n}{\sum_j exp(\tilde{v}_{jn} - b_n\theta_n) + V_n})^{\gamma_n})\eta \quad \text{(C.16)}$$

Let $\theta_n = \eta$ and plug its value into (C.15) to check the inequality, which is equivalent to checking

$$\frac{\sum_j exp(\tilde{v}_{jn} - b_n\eta)}{\sum_j exp(\tilde{v}_{jn} - b_n\eta) + V_n} > 1 - (\frac{V_n}{\sum_j exp(\tilde{v}_{jn} - b_n\eta) + V_n})^{\gamma_n}$$
$$\Leftrightarrow$$
$$(\frac{V_n}{\sum_j exp(\tilde{v}_{jn} - b_n\eta) + V_n})^{\gamma_n} > \frac{V_n}{\sum_j exp(\tilde{v}_{jn} - b_n\eta) + V_n} \quad \text{(C.17)}$$

(C.17) is satisfied because $\frac{V_n}{\sum_j exp(\tilde{v}_{jn} - b_n\eta) + V_n} < 1$. Therefore, we can find a finite θ_n such that $\eta_n(\theta_n) > \eta(+\infty)$, which proves the existence of finite optimizer for problem (4.2.11).

□

Proof of Corollary 4.2.3

Proof. Let function $\phi_n(\theta) := \theta_n \pi(S_n|n,S) + \frac{\theta_n}{\gamma_n}\pi(S_0|n,S) - \frac{1}{\beta_r^n}$. Taking partial derivative of $\phi_n(\theta)$ with respect to θ_n:

$$\frac{\partial \phi_n(\theta)}{\partial \theta_n} = \pi(S_n|n,S) + \frac{1}{\gamma_n}\pi(S_0|n,S) + (\frac{b_n\theta_n}{\gamma_n} - b_n\theta_n)\pi(S_n|n,S)\pi(S_0|n,S) \tag{C.18}$$

Since $\gamma_n < 1$, then term $\frac{b_n\theta_n}{\gamma_n} - b_n\theta_n > 0$, so $\phi_n(\theta)$ monotonically increases with respect to θ_n. Therefore, we know that in equation (4.2.12), every ϕ will uniquely determine a vector of θ.

Consider the derivative of $R(\phi)$ with respect to ϕ,

$$\begin{aligned}
\frac{\partial R(\phi)}{\partial \phi} &= \sum_n \frac{\partial R(\theta)/\partial \theta_n}{\partial \phi/\partial \theta_n} \\
&= \sum_n \frac{\beta_r^n \pi(S_n|S)\{\frac{1}{\beta_r^n} + \sum_n \theta_n \pi(S_n|S) - \theta_n \pi(S_n|n,S) - \frac{\theta_n}{\gamma_n}\pi(S_0|n,S)}{\pi(S_n|n,S) + \frac{1}{\gamma_n}\pi(S_0|n,S) + (\frac{b_n\theta_n}{\gamma_n} - b_n\theta_n)\pi(S_n|n,S)\pi(S_0|n,S)} \\
&= (R(\phi) - \phi)\sum_n \frac{\beta_r^n \pi(S_n|S)}{\pi(S_n|n,S) + \frac{1}{\gamma_n}\pi(S_0|n,S) + (\frac{b_n\theta_n}{\gamma_n} - b_n\theta_n)\pi(S_n|n,S)\pi(S_0|n,S)}
\end{aligned} \tag{C.19}$$

Observe the second part of (C.19) is always greater than 0, so the sign of the first order derivative will purely depend on $R(\phi) - \phi$. As the value of ϕ increase, there will be a unique sign change from positive to negative for the derivative, and the optimal ϕ^* is determined from the unique fixed point $R(\phi) = \phi$.

□

C.2.3 Proof of Theorem 4.2.4

Proof. Assume that all nests except nest n are configured, and nest n is also "partially" configured as S_n, where only one feature-level k is available for any product in the nest n. Then to which product should the feature-level be added?

Recall that under the constant markup scheme, the conditional choice probability of product i in nest n:

$$\pi(in|n,S) = \frac{exp(\tilde{v}_{in} - b_n\theta_n)}{\sum_j exp(\tilde{v}_{jn} - b_n\theta_n) + V_n}$$

The probability of choosing nest n:

$$\pi(n|S) = \frac{(\sum_j exp(\tilde{v}_{jn} - b_n\theta_n) + V_n)^{\gamma_n}}{\sum_{n'}(\sum_j exp(\tilde{v}_{jn'} - b_{n'}\theta_{n'}) + V_{n'})^{\gamma_{n'}} + 1}$$

The expected profit:

$$R(\theta) = \sum_n \theta_n \pi(n|S) \sum_i \pi(in|n,S)$$

where $\tilde{v}_{in} = \beta_0^n/\gamma_n + \sum_k g_k^n x_{in}^k$ is product in's value gap index, entirely determined by the product's configuration. And every feature-level k in nest n also has its own identity g_k^n, the feature-level value gap.

From the above formulation we can see, under a given constant markup, adding feature-level k to product i in nest n is equivalent to increasing product in's value gap by g_k^n; Similarly, adding feature-level k to product j in nest n is equivalent to increasing product jn's value gap by g_k^n; So all we need to compare is which manipulation can increase the expected profit by a larger amount.

First, we need to look into the partial derivatives of the choice probabilities with respect to the value gaps.

The partial derivative of product in's conditional choice probability with respect to its own value gap:

$$\frac{\partial \pi(in|n,S)}{\partial \tilde{v}_{in}} = \pi(in|n,S)(1-\pi(in|n,S))$$

The partial derivative of another product jn's conditional choice probability with respect to in's value gap:

$$\frac{\partial \pi(j|n)}{\partial \tilde{v}_{in}} = \pi(in|n,S)\pi(j|n)$$

The partial derivative of nest n's choice probability with respect to in's value gap:

$$\frac{\partial \pi(n|S)}{\partial \tilde{v}_{in}} = \gamma_n \pi(in|S)(1-\pi(n|S))$$

The partial derivative of nest n''s choice probability with respect to in's value gap:

$$\frac{\partial \pi(n')}{\partial \tilde{v}_{in}} = \gamma_n \pi(n')\pi(in|S)$$

Plug them into the partial derivative of $R(\tilde{v})$ with respect to $\tilde{v}_{in}$. According to the Envelop Theorem,

$$\begin{aligned}\frac{\partial R(\tilde{v})}{\partial \tilde{v}_{in}} &= \sum_{n'} \theta_{n'} \{ \frac{\partial \pi(n')}{\partial \tilde{v}_{in}} \pi(S|n') + \pi(n') \sum_{j} \frac{\partial \pi(j|n')}{\partial \tilde{v}_{in}} \} \\ &= \gamma_n \pi(in|S) \{ \theta_n \pi(S_n|n,S) - R(\theta) + \frac{\theta_n}{\gamma_n} \pi(S_0|n,S) \} \end{aligned} \quad \text{(C.20)}$$

Where θ is the optimal nest-level mark-up vector, satisfying the first order condition given in (C.11). So the partial derivative can be reduced to:

$$\frac{\partial R(\tilde{v})}{\partial \tilde{v}_{in}} = \frac{\pi(in|S)}{b_n} \quad \text{(C.21)}$$

Compare the partial derivative with respect to different products' value gaps,

$$\frac{\partial R(\tilde{v})}{\partial \tilde{v}_{in}} = \frac{\pi(in|S)}{b_n} > \frac{\partial R(\tilde{v})}{\partial \tilde{v}_{jn}} = \frac{\pi(jn|S)}{b_n} > 0$$

because $\pi(in|S) > \pi(jn|S)$ when $\tilde{v}_{in} > \tilde{v}_{jn}$. So adding the positive valued feature-level to the product with higher value gap brings more profits than the lower value gap combination does.

□