Title
Choice-Based Assortment and Price Optimization

Author
Wang, Yanqiao

Publication Date
2018

Peer reviewed|Thesis/dissertation

Choice-Based Assortment and Price Optimization

by

Yanqiao Wang

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering - Civil and Environmental Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Zuo-Jun Max Shen, Chair
Professor Joan Walker
Professor Haiyan Huang

Spring 2018

Choice-Based Assortment and Price Optimization

Abstract

Choice-Based Assortment and Price Optimization

by

Yanqiao Wang

Doctor of Philosophy in Engineering - Civil and Environmental Engineering

University of California, Berkeley

Professor Zuo-Jun Max Shen, Chair

Online recommendation systems ask these questions everyday: How to describe customers' purchasing behavior? How to design a product assortment in order to maximize their expected profit/revenue with given customers' behavior? What is the optimal pricing strategy for an assortment? Even further, how to jointly design optimal assortment and pricing at the same time in an efficient way? The answers to these questions have a direct influence to the profitability and feasibility of the recommendation systems.

My thesis handles assortment and price optimization problems with various applications in online e-commerce and travel related recommendations, such as flight, rental car and hotels. For example, for car rentals in Expedia, they would like to offer customers a recommendation page of cars with different brands, prices, types, options, etc. What is a good recommendation for the customers so that they would have a good shopping and traveling experiences? First, it needs to be relevant to customers' choice behavior that can be learned from previous purchasing history or from marketing surveys; second, the recommendation cannot be too specified, which means that those rental cars in the recommendation page cannot be too similar in terms of their attributes; third, it cannot take too long to show the recommendations to the customers - an efficient algorithm is required. In this thesis, we will show our approach to assortment and price optimization problems.

The main contributions of my thesis is: 1) We formulate the assortment and price optimization problems in a choice-based way, which provides a good balance between relevance and variance of the products in an assortment; 2) We develop applicable recommendation algorithms that run in polynomial time and can be dynamically adapted; 3) Compared to the previous literature, our results are more advanced in terms of efficiency and applicability. Specifically, this thesis is consist of three essays in choice-based assortment and price optimization problems.

In the first essay, we study the joint constrained assortment and price optimization problem under the nested logit model with a no-purchase option in every choice stage. The cardinality or space constraints are imposed separately on the assortment of products that are offered in each nest. Specifically, cardinality constraint on a nest limits the total num-

ber of products that can be offered in that nest, and space constraint on a nest limits the total space consumption of products within that nest. The goal is to jointly determine the optimal assortment with optimal prices to maximize the expected profit per customer under cardinality or space constraints. By using our solution approach, this problem is simplified to find the fixed point of a single-variable unimodal expected profit function, where efficient searching algorithms can be applied. Furthermore, we provide a piecewise convex fixed point representation to facilitate computing. The optimal solution under cardinality constraints and a 2-approximate solution under space constraints can be obtained efficiently.

In the second essay, we study choice-based constrained assortment and price optimization problems under the multilevel nested logit model with a no-purchase option in every choice stage. For the constrained assortment optimization problem, each candidate product is associated with a fixed profit. The goal is to identify the optimal assortment satisfying cardinality or space constraints to maximize the expected profit per customer. Under cardinality constraints, there is a limitation imposed on nodes in the second lowest level. A polynomial-time algorithm with computational complexity $O(n \max\{m, k\})$ is provided to locate the optimal assortment for the m-level nested logit model with n products, where k is the maximum number of products within any node in level $m-1$. Under space constraints, every product consumes a certain amount of space and candidate assortments must satisfy the space limitation. However, the assortment optimization problem becomes NP-hard under space constraints, thus we develop an algorithm to find a 2-approximate solution in $O(mnk)$ operations. For the price optimization problem, we aim to find the profit-maximizing prices for all products. With product-differentiated price sensitivities, the expected profit function is no longer concave even under the two-level nested logit model, but we are able to reduce the multiproduct price optimization problem from a high dimensional optimization to the maximization of a unimodal function in single-dimensional searching space, in which the optimal prices can be found in a tractable manner.

In the third essay, we know that assortment and pricing decisions are of significant importance to firms and have huge influences on profit. How to jointly optimize over both assortment and prices draws increasing attention recently. However, in the most existing literature that considers joint optimization problem, they either impose strong restrictions on the choice structure or have strong assumptions on the price sensitivity parameters. Moreover, currently there is no flexible and comprehensive way to deal with the joint effect of assortment and pricing under multistage choice structure since the tangle between the assortment and prices makes the joint optimization problem less tractable. In this paper, we study the joint capacitated assortment and price optimization problem where the consumer choosing behavior is governed by the multistage tree logit model. Under the cardinality constraints, we develop an efficient algorithm that runs in polynomial time to find the optimal assortment with optimal prices. Under the space constraints, the assortment optimization problem is NP-hard even under tree logit model with only two levels. We can obtain a 2-approximate solution within the same time scale compared to the joint optimization problem under cardinality constraints. For a tree logit model with N candidate products, both algorithms run in $O(GN \log G)$ where G is the number of grid points for each node. The complexity can be further reduced to $O(GN \log K)$ under mild conditions, where K is the maximum number of children nodes that a nonleaf node could have in the tree structure.

To my parents,

Jiandong Wang and Yamei Wang

Contents

List of Figures

List of Tables

Acknowledgments

It is quite a journey for the past five years at UC Berkeley, I am extremely grateful for the guidance, belief and support from my advisors and friends.

I am deeply indebted to my advisor, Prof. Max Shen, who has always been a great mentor, a research advisor and a friend. His open mind and profound knowledge encourage me to explore interesting topics and provide a very open research environment. I would also like to thank Prof. Joan Walker for her generous support. She guided me through the first part of this journey and kept me on the right track. Her advice on research and career has been invaluable to me.

I would also like to thank my other committee member, Prof. Haiyan Huang, for her helpful comments and suggestions that have significantly improved the exposition of this dissertation. I spent one year at UC Berkeley PATH as a graduate student researcher, I would like to thank Weibin Zhang and Kun Zhou for their support and guidance.

My cool friends, including Andy Mao, Yulin Liu, Ying Cao, Sheng Liu, Mogeng Yin, Timothy Brathwaite, Ziheng Lin, and many others, made 116 McLaughlin a enjoyable and memorable home for me.

Lastly, I would like to thank my parents for their unconditional support and love.

Chapter 1

Introduction

In the first essay, the problem of choosing a set of products so as to maximize the expected profit is referred to as the assortment optimization problem. The goal of price optimization problem is to find the optimal pricing strategy to maximize the expected profit per customer. The multinomial logit model and the nested logit model are often used to describe the customer purchasing behavior. How to jointly optimize over both assortment and prices is of significant importance to study. However, assortment decisions greatly influence the pricing strategy and vice versa. We study the joint constrained assortment and price optimization problem under the nested logit model, which includes two stages of customers' choosing process. In the first stage, the customer chooses either to leave without purchasing (the no-purchase option) or to consider to buy within a nest of products that are grouped based on their attribute similarities. If she chooses to consider buying from a nest of products, then she can still either choose the no-purchase option within that nest or select an actual product in the second stage. Moreover, each nest has a scale parameter that is less than one to measure the similarity of products in that nest. For the joint optimization problem, we find a bridge that connects the pricing problem with the assortment optimization problem so as to resolve the tangle between assortment and pricing decisions. This bridge is a scalar that is defined as the node-specific adjusted markup. Due to practical operational limitation, we consider cardinality or space constraints on the assortment in each nest separately.

We first formulate the joint constrained optimization problem as a bilevel optimization program with assortment optimization and price optimization as its outer and inner problem, respectively. Then it can be simplified as an optimization over a single-variable unimodal function by observing the connection between the inner and outer optimization problems. Furthermore, the optimal solution has a piecewise convex fixed point representation. Our solution approach is one step further than 1) joint assortment and price optimization problem under the multinomial logit model that is studied in [48]; 2) price optimization problem under the nested logit model [19]; 3) assortment optimization problem under the nested logit model [18].

In the second essay, we consider the constrained assortment and price optimization problems under choice models, where we aim to offer products bundle with diversity to attract more customers and gain more profits. Preferences and choosing behaviors of customers can be modeled by discrete choice models that play an important role in revenue management and demand modeling. The multinomial logit model, developed by [34] according to random utility maximization (RUM) theory, is extensively used to capture the customer choosing behavior. More complicated choice processes have been modeled under the extensions of the multinomial logit model, such as the nested logit model, the mixed logit model, etc. In this essay, we consider the constrained assortment and price optimization problems under the multilevel nested logit model that is able to capture the multidimensional similarities of products. We assume that customers choose their desired attributes of the products sequentially based on an m-level tree structure, each level of which corresponds to a certain attribute of the product. Thus products are grouped by $(m-1)$-dimensional similarity in the m-level nested logit model if we consider the root node is in level 0.

Regarding the constrained assortment optimization problem, we assume that the prices of all products have already been exogenously given. The probability of choosing a product, which is a function of assortment, can be computed within the multilevel nested logit model framework. Our objective of the constrained assortment optimization problem is to find the expected-profit-maximizing assortment which also satisfies cardinality or space constraints. For cardinality constraints, there is a limitation of the number of products that can be offered within the nodes in level $m-1$ separately. Space constraints limit the available space for displaying products in the $m-1$ level separately, that can be the shelf space or volume space limitation in a physical retail store. We also study the multistage price optimization problem, where assortment and costs of products are fixed. However, instead of having a fixed profit, which is price minus cost, the price vector becomes the decision variable. The choice probability and profit of a product is determined by its own price, thus the price optimization problem becomes a multidimensional optimization problem with respect to the prices of all products.

In the third essay, a decision-making problem that firms always face is to choose a set of products that satisfy either cardinality or space constraints with proper prices to offer to the consumers in order to maximize their profit, which can be addressed as the joint capacitated assortment and price optimization problem. For the capacitated assortment optimization problem, the goal is to identify a set of products under certain constraints to offer to consumers so as to maximize the expected profit when the prices of candidate products are exogenously given. Price optimization refers to the problem of setting an expected-profit-maximizing price for each product within a fixed assortment selection, where the attractiveness of a product is inversely proportional to its price. The assortment optimization applies to the case where a firm cannot control the prices but is able to decide which products to offer to the consumers, and the price optimization is vice versa. For firms that have the ability to have control over both assortment and prices of products, joint optimization is necessary and worth studying. This fact is intuitive, for example, if some products with "good" quality or brand are added to a firm's consideration set, the optimal assortment and pricing strategy may change, since the demand for those products with even higher price may increase because consumers may have an overall "good" impression on the newly offered assortment and can tolerate high prices of some products.

However, assortment decisions are very sensitive to price changes, and similarly, optimal prices of products are also completely different given different offered assortment. On the other hand, the mutual dependence of capacitated assortment and pricing decisions increases the hardness of modeling and quantifying the *joint* effect of capacitated assortment and pricing. Most existing approaches do not have a tractable solution to jointly optimizing over capacitated assortment and prices, especially for the case where the consumers follow a complex choice structure, such as the tree logit model with an arbitrary number of levels and products.

To resolve the issues in modeling the joint effect and provide practical operations insights, we study the joint capacitated assortment and price optimization problem under the tree logit model [10]. Due to the practical display limitation, we consider both cardinality and space constraints on the assortment decisions, which impose cardinality and space limitation on nodes of the second lowest level in the tree structure, respectively. Under the tree logit model, consumers follow an m-level tree structure. The choosing process can be considered as a desired set reduction process, where the desired set is originally set to be the entire choice set and being reduced as choosing process goes on until only one product or the no-purchase option is left. For example, if a consumer wants to buy a history book on Amazon, after she specifies that the category is history under the "shop by category" list, all the other books that do not belong to this category will be eliminated from the desired set and not be purchased by her, in which case, history can be viewed as the first desired attribute that she wants from the book. Subsequently, the consumer further continues to choose the second attribute and so forth until a book with m desired attributes or the no-purchase option has been chosen in the end. Hence in this tree structure, the node in level l corresponds to a subset of products that share l attributes in common, and all the leaf nodes that share the same parent node stand for actual products having m common attributes. Specifically, the no-purchase option is in the first level of the tree. The choice model with multistage structure has practical motivation and usefulness; see recent studies in [30] and [22].

While retailers tend to set higher prices of products to gain more profit, consumers typically would consider less to buy a product with high price. So the products with higher price always have lower demand. Therefore, the joint effect of assortment and pricing can be translated as the tradeoff between promoting the willingness to buy of consumers and maximizing retailer's profit. The tangle in the joint effect can be unraveled by our efficient approach to the joint capacitated assortment and price optimization problem.

Chapter 2

Joint Nested Logit Model

2.1 Literature Review

[26] show an extensive review of the assortment optimization and price optimization problems under various choice models. For the work that is related to this essay, [18] study the constrained assortment optimization problem under the nested logit model and their approach can be adapted to solve the joint assortment and price optimization problem if feasible prices are restricted on finite grid points. Our model does not have this restriction by defining the price of a product on $\mathbb{R}_{\geqslant 0}$. [11] study the assortment optimization problems under the nested logit model with no-purchase options in all nests. [16] consider the constrained assortment optimization under the nested logit model with constraints across nests. [29] find structural conditions of the optimal assortment under the nested logit model. By assuming that the price-sensitivity parameters are identical within nests, [31] prove the expected profit function is concave with respect to market share vector. [19] show the expected profit function is unimodal under mild assumption on dissimilarity parameter and price-sensitivity parameters, whereas they relax the assumption in [31]. In this essay, we also have the same assumption that is in [19].

[39] study the constrained assortment optimization problem under the multinomial logit model. [47] considers the constrained assortment optimization problem under the general attraction model. Under the mixed multinomial logit model, customers are segmented into groups based on their social demographic information, [7], [41] and [23] study the assortment optimization problem. [18] consider the constrained assortment optimization problem under the nested logit model. [30] and [50] study the unconstrained and constrained assortment optimization under the multilevel level nested logit model, respectively. However, most of the research listed above only consider assortment optimization problems without the joint effect of assortment and pricing decisions.

[15] find the multinomial logit profit function is concave with respect to the market share vector. Under the nested logit model, [19] show the profit function is concave in terms of the aggregate market share and unimodal of the adjusted nest-level markup. Under the

multilevel nested logit model, [30] and [22] find an efficient approach to get the optimal pricing strategy. [50] consider no-purchase options in every choice stage under the multilevel nested logit model and study the price optimization problem.

For the literature considering the joint assortment and price optimization problem, [9] [32] obtain the structural properties of the optimal solution. [48] considers the multinomial logit model and proves that the joint optimization problem has a fixed point representation. Furthermore, [49] considers the search cost in the joint optimization problem. [6] study the joint problem in a game theory perspective. [27] consider the joint optimization problem under the nested logit model and obtain a competitive equilibrium. Under the nested logit model, [18] and [12] restrict prices on a grid of points. Our model does not have this restriction, while both approaches have real applications in practice. Under the multilevel nested logit model, [51] study the joint optimization problem, however, the authors only allows one no-purchase option. [37] proposes a linear program formulation with price bounds. [24] considers the nonparametric choice model for the joint assortment and price optimization problem.

2.2 Main Results and Contributions

We summarize our main results and contributions as follows:

1. We study the joint constrained assortment and price optimization problem under the nested logit model in this essay. We formulate the joint optimization problem as a bilevel optimization program with the price optimization problem and the constrained assortment optimization problem as its inner and outer problem, respectively. Focusing on the inner price optimization problem with a fixed nonempty assortment, we introduce a scalar that is referred to as the node-specific adjusted markup, which is proved to be a useful bridge connecting the inner and outer problems jointly.

2. In our problem setting, the consumer choosing process can be described under the nested logit model with m nests and N products, where $n_{\max}$ is the maximum number of products within any nest. We impose the cardinality or space constraints separately on each nest, which limits the number of products or space consumption of products within that nest. We first decompose the joint constrained assortment and price optimization problem into m bilevel joint subproblems, then we introduce an equivalent formulation of the joint subproblem that is referred to as the assortment subproblem. The assortment subproblem optimizes over a scalar instead of an assortment, and its objective function is convex, which makes it tractable to solve.

3. The main result in this essay is that we prove the joint constrained assortment and price optimization problem has a fixed point representation of a single-variable unimodal profit function. We also prove that the size of a collection that includes an optimal assortment under cardinality or space constraints is polynomially bounded by $O(N)$ or $O(n_{\max}N)$, respectively, thus a solution approach that is based on discretization can be applied to find the joint optimal solution. Furthermore, we propose a piecewise convex fixed point represen-

tation to further facilitate calculation. By applying our solution approach, the joint optimal solution and a 2-approximate solution can be obtained in an efficient way.

4. To the best of our knowledge, we are the first to study the joint constrained assortment and price optimization problem under the nested logit model where the utility of a product is a function of the price of this product. [48] considers the joint optimization problem under the cardinality constraints when the customer choosing behavior is governed by the multinomial logit model, which is tractable based on its linearity nature. [18] study the joint optimization problem the nested logit model where there exists nonlinearity in general. However, the authors only restrict the possible prices on some prespecified values and do not consider constrained assortment in their problem setting.

Organization

The organization of the essay is as follows. In Section 2.3, we present the modeling framework and problem formulation of the joint constrained assortment and price optimization problem under the nested logit model. Section 2.4 shows the joint optimization problem under cardinality constraints is tractable and an optimal solution can be efficiently obtained via a piecewise convex fixed point representation. In Section 2.5, we consider the joint optimization problem under space constraints and show how to find a 2-approximate solution. We illustrate our solution approach in Section 2.6 by showing a numerical example of the joint optimization problem under cardinality constraints.

2.3 Modeling Framework

Suppose that the customer purchasing behavior can be described by the nested logit model with m nests, the set of which is $M = \{1, 2, ..., m\}$. For a nest $i \in M$, there are n_i products, the set of which is denoted as $N_i = \{1, 2, ..., n_i\}$. The total number of products is denoted as $N = \sum_{i \in M} n_i$. For product $j \in N_i$, it is represented as a 2-tuple $\langle i, j \rangle$ and its price is p_{ij}, thus the price vector for nest i is $\mathbf{P}_i = (p_{i1}, p_{i2}, ..., p_{in_i})$. Price matrix $\mathbf{P} = (\mathbf{P}_1, \mathbf{P}_2, ..., \mathbf{P}_m)$ contains all prices for all the products in the nested logit model. A customer first chooses either to select a nest or to leave without purchasing anything; if she selects a nest of products in the first stage, then she can still choose either to exit without buying or to select a product in the second level. The no-purchase option in the first stage is denoted as a null tuple $\langle 0 \rangle$, and the no-purchase option within nest i is denoted as an order pair $\langle i, 0 \rangle$. For product $\langle i, j \rangle$ with price p_{ij}, the preference weight that is assigned by the customer is

$$v_{ij}(p_{ij}) = \exp(\alpha_{ij} - \beta_{ij} p_{ij}),$$

where α_{ij} is the price-independent deterministic utility and β_{ij} is the price-sensitivity parameter. Note that both α_{ij} and β_{ij} are different across products. By convention, we treat the preference weight of no-purchase options v_0 and v_{i0} as nonnegative constants.

The vector $S_i = (S_{i1}, S_{i2}, ..., S_{in_i}) \in \{0,1\}^{n_i}$ is used to denote the assortment of products that are offered in nest i. The binary decision variable S_{ij} equals to one if product $\langle i, j \rangle$ is offered and zero otherwise. If we offer S_i with price vector $\mathbf{P}_i$ for a nest $i \in M$, the probability of choosing product j out of N_i is

$$Q_{j|i}(S_i, \mathbf{P}_i) = \frac{v_{ij}(p_{ij})S_{ij}}{v_{i0}\mathbf{1}(S_i \neq \emptyset) + \sum_{j \in N_i} v_{ij}(p_{ij})S_{ij}},$$

where if $S_i \neq \emptyset$, the customer has a probability of $v_{i0}S_{ij}/(v_{i0} + \sum_{j \in N_i} v_{ij}(p_{ij})S_{ij})$ to leave without purchasing. Note that if $S_{ij} = 0$, then the price p_{ij} of product $\langle i, j \rangle$ becomes irrelevant to our goal.

For nest $i \in M$, the expected profit of assortment S_i with price vector $\mathbf{P}_i$ is

$$\begin{aligned} R_i(S_i, \mathbf{P}_i) &= \sum_{j \in N_i} Q_{j|i}(S_i, \mathbf{P}_i)(p_{ij} - c_{ij}) \\ &= \frac{\sum_{j \in N_i}(p_{ij} - c_{ij})v_{ij}(p_{ij})S_{ij}}{v_{i0}\mathbf{1}(S_i \neq \emptyset) + \sum_{j \in N_i} v_{ij}(p_{ij})S_{ij}}, \end{aligned}$$

where c_{ij} is the cost of product $\langle i, j \rangle$. We use $V_i(S_i, \mathbf{P}_i) = (v_{i0}\mathbf{1}(S_i \neq \emptyset) + \sum_{j \in N_i} v_{ij}(p_{ij})S_{ij})^{\gamma_i}$ to measure the attractiveness of S_i, where $\gamma_i \in (0, 1]$ measures the dissimilarity of products in nest i. The closer γ_i is to one, the less similarities between products within nest i are.

To guarantee the uniqueness of the optimal pricing strategy, we assume that $\max_{j \in N_i} \beta_{ij} / \min_{j \in N_i} \beta_{ij} < 1/(1 - \gamma_i)$ for $i \in M$ as in [19]. If we offer assortment matrix $S = (S_1, S_2, ..., S_m)$ over all nests, then the probability that a customer considers to buy a product in assortment S_i is

$$Q_i(S, \mathbf{P}) = \frac{V_i(S_i, \mathbf{P}_i)}{v_0 + \sum_{i \in M} V_i(S_i, \mathbf{P}_i)}.$$

We remark that $\sum_{i \in M} Q_i(S, \mathbf{P}) < 1$ and $\sum_{j \in N_i} Q_{j|i}(S_i, \mathbf{P}_i) < 1$ if $v_{i0} > 0$. The total expected profit for assortment S with prices $\mathbf{P}$ is written as

$$\Pi(S, \mathbf{P}) = \sum_{i \in M} Q_i(S, \mathbf{P}) R_i(S_i, \mathbf{P}_i) = \frac{\sum_{i \in M} V_i(S_i, \mathbf{P}_i) R_i(S_i, \mathbf{P}_i)}{v_0 + \sum_{i \in M} V_i(S_i, \mathbf{P}_i)}.$$

The cardinality or space constraints are imposed on each of the nest separately. For the cardinality constraints $\mathbb{C}_i$, it restricts the number of products that are offered in nest i to not exceeding $\mathbb{C}_i$, thus the set of feasible assortments at nest i is denoted as $\Im_i = \{S_i \in \{0,1\}^{n_i} : \sum_{j \in N_i} S_{ij} \leqslant \mathbb{C}_i\}$. Similarly, space constraints $\mathbb{S}_i$ limit the space consumption of assortment S_i, then the set of feasible assortments at nest i is $\Im_i = \{S_i \in \{0,1\}^{n_i} : \sum_{j \in N_i} w_{ij} S_{ij} \leqslant \mathbb{S}_i\}$ where $w_{ij} \leqslant \mathbb{S}_i$ is the space requirement of product $\langle i, j \rangle$. Without loss of generality, we assume that $\mathbb{C}_i \leqslant n_i$ and $\mathbb{S}_i \leqslant \sum_{j \in N_i} w_{ij}$ since it becomes uncapacitated assortment optimization otherwise. The feasible set of assortments over all nests is the cartesian product of $\Im_i$

over $i \in M$, which is denoted as $\Im = \Im_1 \times \Im_2 \times \cdots \times \Im_m$. We aim to jointly determine the expected-profit-maximizing assortment that satisfies the cardinality or space constraints and the optimal price matrix to the following problem

$$\max_{S \in \Im} \max_{\mathbf{P} \in \mathbb{R}^N_{\geqslant 0}} \Pi(S, \mathbf{P}). \tag{2.1}$$

Throughout the essay, we use $S^* = (S_i^* : i \in M)$ and $\mathbf{P}^* = (\mathbf{P}_i^* : i \in M)$ to denote the joint optimal solution and let $Z^* = \Pi(S^*, \mathbf{P}^*)$ be the maximum expected profit that we can obtain per consumer from the collection of feasible assortments $\Im$. We remark that if $S_{ij}^* = 0$, then the price of product $\langle i, j \rangle$ becomes irrelevant, thus we set $p_{ij}^* = 0$ as well. In the following sections, we show that problem (4.1) is tractable by building a bridge that connects the inner pricing problem and the outer assortment optimization problem.

2.4 Joint Optimization Under Cardinality Constraints

We present our solution approach to problem (4.1) under the cardinality constraints. In Section 2.4, we decompose problem (4.1) into joint subproblems. The union of the solutions to the joint subproblems at all nests is an optimal solution to problem (4.1). In Section 2.4, we show an equivalent formulation of the joint subproblem, which is referred to as the assortment subproblem. Then we show the expected profit function can be transformed to a single-variable unimodal function and there exists a piecewise convex fixed point representation of problem (4.1) in Section 2.4.

Joint Subproblem

In this section, we consider problem (4.1) by decomposing it into joint subproblems at all nests. The joint subproblem is a bilevel optimization problem with the price optimization and assortment optimization problem as its inner and outer problem, respectively. By solving the inner pricing problem, we present an equivalent formulation of the joint subproblem, which is only related to assortment decision variables.

Before introducing the joint subproblem, we first show that $R_i(S_i^*, \mathbf{P}_i^*)$ is at least as large as Z^* if S_i^* is not empty as in the following claim.

Claim 1. *If $S_i^* \neq \varnothing$, then we get $R_i(S_i^*, \mathbf{P}_i^*) \geqslant Z^*$.*

We defer the proof of this claim to Appendix A.2. The *joint subproblem* at nest $i \in M$ is defined as follows

$$\max_{S_i \in \Im_i} \max_{\mathbf{P}_i \in \mathbb{R}^{n_i}_{\geqslant 0}} V_i(S_i, \mathbf{P}_i)(R_i(S_i, \mathbf{P}_i) - Z^*). \tag{2.2}$$

The following claim shows the relationship between the union of the solutions to joint subproblems at all nests and problem (4.1).

Claim 2. *For each $i \in M$, let $(\hat{S}_i, \hat{\mathbf{P}}_i)$ be optimal to the joint subproblem (2.2) at nest i, then $(\hat{S}_1, \hat{S}_2, ..., \hat{S}_m; \hat{\mathbf{P}}_1, \hat{\mathbf{P}}_2, ..., \hat{\mathbf{P}}_m)$ is an optimal solution to problem (4.1).*

The proof of Claim 2 can be found in Appendix A.2. Joint subproblem (2.2) is a bilevel optimization problem: the inner optimization problem is a pricing problem for a fixed assortment; the outer optimization problem is a constrained assortment optimization problem where prices have already been set "optimally" for each feasible assortment.

Next we focus on the inner pricing problem at nest i with a fixed nonempty assortment S_i, which is shown as follows

$$\max_{\mathbf{P}_i(S_i) \in \mathbb{R}_{\geqslant 0}^{|S_i|}} V_i(S_i, \mathbf{P}_i(S_i))(R_i(S_i, \mathbf{P}_i(S_i)) - Z^*), \tag{2.3}$$

where the dimension of price vector $\mathbf{P}_i(S_i)$ is $|S_i|$ instead of n_i since the prices of the products that are not in assortment S_i are irrelevant. Note that $\mathbf{P}_i(S_i)$ is a function of assortment S_i, we use $\mathbf{P}_i$ to denote $\mathbf{P}_i(S_i)$ later in this essay for notational purpose. With a slight abuse of notation, S_i is also used to denote the set of products in assortment S_i. The next lemma shows the conditions that should be satisfied by the price vector at optimality of problem (2.3).

Lemma 1. *The optimality condition of problem (2.3) with a given nonempty assortment S_i and a constant Z^* is*

$$\theta_i = \gamma_i Z^* + (1 - \gamma_i) R_i(S_i, \theta_i) \tag{2.4}$$

where the node-specific adjusted markup $\theta_i = p_{ij} - c_{ij} - 1/\beta_{ij}$ is invariant for all $j \in S_i$.

Proof. Proof: For notational brevity, let $g_i = V_i(S_i, \mathbf{P}_i)(R_i(S_i, \mathbf{P}_i) - Z^*)$, $V_i = V_i(S_i, \mathbf{P}_i)$, $R_i = R_i(S_i, \mathbf{P}_i)$ and $Q(j|i) = Q_{j|i}(S_i, \mathbf{P}_i)$. The first derivative of the objective function g_i with respect to the price p_{ij} of product $j \in S_i$ is

$$\frac{\partial g_i}{\partial p_{ij}} = \frac{\partial V_i}{\partial p_{ij}}(R_i - Z^*) + V_i \frac{\partial R_i}{\partial p_{ij}},$$

where

$$\frac{\partial V_i}{\partial p_{ij}} = \gamma_i (v_{i0} + \sum_{j \in S_i} v_{ij})^{\gamma_i - 1}(-\beta_{ij} v_{ij}) = -\gamma_i \beta_{ij} V_i Q(j|i),$$
$$\frac{\partial R_i}{\partial p_{ij}} = -\beta_{ij} Q(j|i)(p_{ij} - c_{ij} - \frac{1}{\beta_{ij}} - R_i).$$

After plugging terms, we obtain

$$\frac{\partial g_i}{\partial p_{ij}} = -\beta_{ij} V_i Q(j|i)[\gamma_i (R_i - Z^*) + p_{ij} - c_{ij} - \frac{1}{\beta_{ij}} - R_i].$$

Let $\partial g_i/\partial p_{ij} = 0$, it follows that

$$p_{ij} - c_{ij} - \frac{1}{\beta_{ij}} = \gamma_i Z^* + (1-\gamma_i)R_i,$$

where the right hand side of the above equation does not depend on j, thus $p_{ij} - c_{ij} - 1/\beta_{ij}$ is independent of j, which can be denoted as θ_i. □

Lemma 1 shows that $\theta_i = p_{ij} - c_{ij} - 1/\beta_{ij}$ is invariant within assortment S_i at optimality of problem (2.3), which in turn indicates that the price p_{ij} of product $j \in S_i$ should be set as $\theta_i + c_{ij} + 1/\beta_{ij}$ at optimality. Therefore, the price vector $\mathbf{P}_i$ can be rewritten as $\mathbf{P}_i = (\theta_i + c_{ij} + 1/\beta_{ij} : j \in S_i)$, which has a one-to-one increasing correspondence with θ_i. Thus, it suffices to consider the node-specific adjusted markup θ_i that is a scalar instead of the price vector $\mathbf{P}_i$. Similar to the discussion before Lemma 1, note that θ_i is also a function of assortment S_i. Later in this essay, we focus on the scalar θ_i instead of the vector $\mathbf{P}_i$, thus we have $R_i(S_i, \mathbf{P}_i) = R_i(S_i, \theta_i)$, $V_i(S_i, \mathbf{P}_i) = V_i(S_i, \theta_i)$ and $Q_{j|i}(S_i, \mathbf{P}_i) = Q_{j|i}(S_i, \theta_i)$ at optimality accordingly. Throughout, θ_i^* is used to denote the scalar that corresponds to the optimal price vector $\mathbf{P}_i^*$.

Since Equation (2.4) is a necessary condition of the inner pricing problem for a given nonempty assortment S_i, if this S_i is the optimal assortment S_i^* to problem (4.1) and $S_i^* \neq \varnothing$, then S_i^* and the corresponding θ_i^* should also satisfy $\theta_i^* = \gamma_i Z^* + (1-\gamma_i)R_i(S_i^*, \theta_i^*)$. Specifically, we set $\theta_i^* = Z^*$ if S_i^* is empty without loss of generality. According to Claim 1, we have θ_i^* is a scalar that is always greater than or equal to Z^*. By observing the relationship between $R_i(S_i, \theta_i)$ and θ_i, the optimality condition (2.4) can be further simplified in next corollary.

Corollary 1. *The optimality condition (2.4) of problem (2.3) can be rewritten as*

$$Z^* = \delta_i(S_i, \theta_i)\theta_i - \omega_i(S_i, \theta_i), \tag{2.5}$$

where $\delta_i(S_i, \theta_i) = 1/\gamma_i - (1/\gamma_i - 1)\tau_i(S_i, \theta_i)$, $\tau_i(S_i, \theta_i) = \sum_{j \in S_i} Q_{j|i}(S_i, \theta_i)$ *and* $\omega_i(S_i, \theta_i) = (1/\gamma_i - 1)\sum_{j \in S_i} Q_{j|i}(S_i, \theta_i)/\beta_{ij}$.

Proof. Proof: For notational brevity, let $Q(j|i) = Q_{j|i}(S_i, \theta_i)$, $\tau_i = \tau_i(S_i, \theta_i)$ and $\omega_i = \omega_i(S_i, \theta_i)$. By the definition of θ_i and R_i, we get

$$R_i = \sum_{j \in S_i} Q(j|i)(\theta_i + \frac{1}{\beta_{ij}}) = \tau_i\theta_i + \frac{\gamma_i}{1-\gamma_i}\omega_i,$$

where $\omega_i = (1/\gamma_i - 1)\sum_{j \in S_i} Q(j|i)/\beta_{ij}$. Plug the above equation into Equation (2.4), we have

$$\theta_i = \gamma_i Z^* + (1-\gamma_i)(\tau_i\theta_i + \frac{\gamma_i}{1-\gamma_i}\omega_i).$$

Equation (2.5) is obtained after collecting terms. □

For these newly introduced quantities, we observe that $0 < \tau_i(S_i, \theta_i) \leqslant 1$ and $1 \leqslant \delta_i(S_i, \theta_i) < 1/\gamma_i$, where the inequalities are strict if $v_{i0} \neq 0$ for a nonempty assortment S_i. Next lemma shows there exists a one-to-one increasing correspondence between Z^* and θ_i in Equation (2.5).

Lemma 2. *Under the assumptions of price-sensitivity parameters, $g(\theta_i) = \delta_i(S_i, \theta_i)\theta_i - \omega_i(S_i, \theta_i)$ is a strictly increasing function of θ_i.*

Proof. Proof: Define $u_i(S_i, \theta_i) = \sum_{j \in S_i} \beta_{ij} Q_{j|i}(S_i, \theta_i)$. Let $g_i = g(\theta_i)$, $Q(j|i) = Q_{j|i}(S_i, \theta_i)$, $\delta_i = \delta_i(S_i, \theta_i)$, $\tau_i = \tau_i(S_i, \theta_i)$, $u_i = u_i(S, \theta_i)$ and $\omega_i = \omega_i(S_i, \theta_i)$ for notational purpose. The first derivative of g_i with respect to θ_i is

$$\frac{\partial g_i}{\partial \theta_i} = \delta_i + \theta_i \frac{\partial \delta_i}{\partial \theta_i} - \frac{\partial \omega_i}{\partial \theta_i},$$

where

$$\begin{aligned}
\frac{\partial \delta_i}{\partial \theta_i} &= -(\frac{1}{\gamma_i} - 1)\frac{\partial \tau_i}{\partial \theta_i} = -(\frac{1}{\gamma_i} - 1)\sum_{j \in S_i} \frac{\partial Q(j|i)}{\partial \theta_i} \\
&= -(\frac{1}{\gamma_i} - 1)\sum_{j \in S_i} Q(j|i)(u_i - \beta_{ij}) = (\frac{1}{\gamma_i} - 1)(1 - \tau_i)u_i = (\delta_i - 1)u_i, \\
\frac{\partial \omega_i}{\partial \theta_i} &= (\frac{1}{\gamma_i} - 1)\sum_{j \in S_i} \frac{1}{\beta_{ij}} \frac{\partial Q(j|i)}{\partial \theta_i} = (\frac{1}{\gamma_i} - 1)\sum_{j \in S_i} \frac{1}{\beta_{ij}} Q(j|i)(u_i - \beta_{ij}) = \omega_i u_i - (\frac{1}{\gamma_i} - 1)\tau_i,
\end{aligned}$$

thus we have

$$\frac{\partial g_i}{\partial \theta_i} = \frac{1}{\gamma_i} - \omega_i u_i + \theta_i(\delta_i - 1)u_i,$$

where $\delta_i \geqslant 1$, $\omega_i \geqslant 0$ and $u_i \geqslant 0$. At optimality, we have $\theta_i \geqslant 0$ since $Z^* = \delta_i\theta_i - \omega_i \geqslant 0$, otherwise $Z^* < 0$ if $\theta_i < 0$. To prove $\partial g_i / \partial \theta_i > 0$, we only need to show $1/\gamma_i - \omega_i u_i > 0$. We have

$$\begin{aligned}
\frac{\gamma_i}{1 - \gamma_i}\omega_i &= \sum_{j \in S_i} \frac{Q(j|i)}{\beta_{ij}} \leqslant \frac{\sum_{j \in S_i} Q(j|i)}{\min_{j \in S_i} \beta_{ij}} \leqslant \frac{\tau_i}{\min_{j \in N_i} \beta_{ij}} \leqslant \frac{1}{\min_{j \in N_i} \beta_{ij}} \\
u_i &= \sum_{j \in S_i} \beta_{ij} Q(j|i) \leqslant \max_{j \in S_i} \beta_{ij} \sum_{j \in S_i} Q(j|i) \leqslant \max_{j \in N_i} \beta_{ij}\tau_i \leqslant \max_{j \in N_i} \beta_{ij}.
\end{aligned}$$

Therefore, it follows that

$$\frac{1}{\gamma_i} - \omega_i u_i \geqslant \frac{1}{\gamma_i} - \frac{1 - \gamma_i}{\gamma_i} \frac{\max_{j \in N_i} \beta_{ij}}{\min_{j \in N_i} \beta_{ij}} > \frac{1}{\gamma_i} - \frac{1 - \gamma_i}{\gamma_i} \frac{1}{1 - \gamma_i} = 0,$$

where the last inequality is due to the parameter assumption that $\max_{j \in N_i} \beta_{ij} / \min_{j \in N_i} \beta_{ij} < 1/(1 - \gamma_i)$. □

Lemma 2 is insightful in a way that if the maximum profit Z^*, which is a constant, is known, then θ_i can be uniquely determined for a given assortment S_i. This implies that the inner pricing problem can be treated as a constraint satisfying Equation (2.5). We are ready to present an equivalent formulation of the joint subproblem at nest $i \in M$ in next proposition.

Proposition 1. *The bilevel joint subproblem (2.2) can be reformulated as an optimization program with respect to the assortment variable, which is written as follows*

$$\begin{aligned} \max_{S_i \in \Im_i} \quad & \frac{1}{1-\gamma_i} V_i(S_i, \theta_i)(\theta_i - Z^*) \\ s.t. \quad & Z^* = \delta_i(S_i, \theta_i)\theta_i - \omega_i(S_i, \theta_i). \end{aligned} \tag{2.6}$$

Proof. Proof: For notational brevity, we let $V_i = V_i(S_i, \theta_i)$, $\omega_i = \omega_i(S_i, \theta_i)$, $\tau_i = \tau_i(S_i, \theta_i)$ and $\delta_i = \delta_i(S_i, \theta_i)$. Since there exists a one-to-one strictly increasing correspondence between Z^* and θ_i for a give assortment S_i at the optimality condition of the inner price optimization problem, it implies that θ_i can be uniquely determined if both Z^* and S_i are given. Thus problem (2.2) can be rewritten as

$$\begin{aligned} \max_{S_i \in \Im_i} \quad & V_i(R_i - Z^*) \\ \text{s.t.} \quad & Z^* = \delta_i\theta_i - \omega_i. \end{aligned}$$

Next we show that the objective function can be reformulated as desired. At the optimality condition of the inner pricing problem of the joint subproblem (2.2), we obtain

$$\begin{aligned} R_i &= \tau_i\theta_i + \frac{\gamma_i}{1-\gamma_i}\omega_i, \\ Z^* &= \delta_i\theta_i - \omega_i, \end{aligned}$$

thus the objective function becomes

$$\begin{aligned} V_i(R_i - Z^*) &= V_i(-(\delta_i - \tau_i)\theta_i + \frac{1}{1-\gamma_i}\omega_i) \\ &= \frac{1}{1-\gamma_i} V_i[\omega_i - (\frac{1}{\gamma_i} - 1)(1-\tau_i)\theta_i] = \frac{1}{1-\gamma_i} V_i[\omega_i - (\delta_i - 1)\theta_i]. \end{aligned}$$

The desired result is established after collecting terms. □

Compared to problem (2.2) that is a bilevel optimization problem, formulation (2.6) is an optimization problem that is only related to the assortment decision variable S_i, where the inner pricing strategy is integrated in the constraints. Let $\hat{S}_i$ denote the optimal solution to problem (2.6) with corresponding node-specific adjusted markup $\hat{\theta}_i$ that can be determined through $Z^* = \delta_i(\hat{S}_i, \theta_i)\theta_i - \omega_i(\hat{S}_i, \theta_i)$. The objective function in formulation (2.6) is $V_i(S_i, \theta_i)(\theta_i - Z^*)/(1-\gamma_i)$, which implies that the optimal objective value satisfies

$V_i(\hat{S}_i, \hat{\theta}_i)(\hat{\theta}_i - Z^*)/(1-\gamma_i) \geqslant V_i(\varnothing, \theta_i)(\theta_i - Z^*)/(1-\gamma_i) = 0$ since $\hat{S}_i$ is always at least as good as the empty assortment $\varnothing$. Therefore, if the optimal solution $\hat{S}_i$ of nest $i \in M$ is nonempty, then $\hat{\theta}_i$ should be greater than or equal to Z^*.

We remark that if the preference weight of the no-purchase option in nest $i \in M$ is zero, i.e. $v_{i0} = 0$, then $\hat{S}_i$ is always nonempty. It is because $\delta_i(S_i, \theta_i)$ always equals to one if $v_{i0} = 0$, thus the objective function of problem (2.6) equals to $V_i(S_i, \theta_i)\omega_i(S_i, \theta_i)/(1 - \gamma_i)$. It is strictly positive for any feasible nonempty assortment S_i, which is strictly larger than $V_i(\varnothing, \theta_i)\omega_i(\varnothing, \theta_i)/(1 - \gamma_i) = 0$, which is the objective value of an empty assortment. Moreover, we have $|\hat{S}_i| = \mathbb{C}_i$ if $v_{i0} = 0$, which will be shown in Section 2.4.

The optimal assortment S_i^* and price θ_i^* satisfy $Z^* = \delta_i(S_i^*, \theta_i^*)\theta_i^* - \omega_i(S_i^*, \theta_i^*)$ since Equation (2.5) is a necessary condition. If both Z^* and S_i^* are known, then θ_i^* can be uniquely identified and problem (4.1) is solved. However, the concern is that it is not possible to get Z^* and S_i^* before solving problem (4.1). Even if Z^* is obtained in a magic way, problem (2.6) is still intractable since the size of $\Im_i$ is $\binom{n_i}{\mathbb{C}_i}$ under the cardinality constraint $\mathbb{C}_i$, which is too large even for a small $\mathbb{C}_i$. We propose a tractable approach in following sections to eliminate these concerns.

Assortment Subproblem

In this section, we first show how to get a polynomial-size collection $\mathcal{A}_i \subseteq \Im_i$ that includes an optimal solution to problem (4.1) by introducing the basic joint subproblem and then reformulate problem (2.6) as an optimization problem in terms of the scalar θ_i, which is referred to as the *assortment subproblem*.

According to the discussion before Corollary 1, we have $\theta_i^* = \gamma_i Z^* + (1 - \gamma_i)R_i(S_i^*, \theta_i^*)$ if $S_i^* \neq \varnothing$ and $\theta_i^* = Z^*$ if $S_i = \varnothing$. The *basic* joint subproblem at nest $i \in M$ is shown as follows

$$\max_{S_i \in \Im_i} \max_{\mathbf{P}_i \in \mathbb{R}^{n_i}_{\geqslant 0}} V_i(S_i, \mathbf{P}_i)^{1/\gamma_i}(R_i(S_i, \mathbf{P}_i) - \theta_i^*). \tag{2.7}$$

Similar to Claim 2, the following claim shows the the union of the optimal solution to problem (2.7) is also optimal to problem (4.1).

Claim 3. *For each $i \in M$, let $(\tilde{S}_i, \tilde{\mathbf{P}}_i)$ be optimal to the basic joint subproblem (2.7) at nest i, then $(\tilde{S}_1, \tilde{S}_2, ..., \tilde{S}_m; \tilde{\mathbf{P}}_1, \tilde{\mathbf{P}}_2, ..., \tilde{\mathbf{P}}_m)$ is an optimal solution to problem (4.1).*

We defer the proof of this claim to Appendix A.2. According to Claim 3, the union of the optimal solutions to the basic joint subproblem is also optimal to problem (4.1). Furthermore, problem (2.7) is easier to solve than problem (2.2) in a way that the objective function of problem (2.7) is linear in terms of S_i. Even though θ_i^* is unknown, it is possible to come up with a collection of assortments that includes an optimal solution to problem (4.1) by observing θ_i^* is essentially a nonnegative scalar. Next we focus on problem (2.7) and show its optimality condition in the following lemma.

Lemma 3. *The optimality condition of the inner pricing problem of basic joint subproblem (2.7) for a given nonempty assortment S_i and a nonnegative scalar θ_i^* is*

$$\theta_i = \theta_i^*,$$

where $\theta_i = p_{ij} - c_{ij} - 1/\beta_{ij}$ is invariant for all $j \in S_i$.

Proof. Let $g_i = V_i(S_i, \mathbf{P}_i)^{1/\gamma_i}(R_i(S_i, \mathbf{P}_i) - \theta_i^*)$, $V_i = V_i(S_i, \mathbf{P}_i)$, $R_i = R_i(S_i, \mathbf{P}_i)$ and $Q(j|i) = Q_{j|i}(S_i, \mathbf{P}_i)$ for notational purpose. We have

$$\begin{aligned}
\frac{\partial g_i}{\partial p_{ij}} &= \frac{\partial V_i^{1/\gamma_i}}{\partial p_{ij}}(R_i - \theta_i^*) + V_i^{1/\gamma_i}\frac{\partial R_i}{\partial p_{ij}} \\
&= -\beta_{ij}v_{ij}(R_i - \theta_i^*) + V_i^{1/\gamma_i}[-\beta_{ij}Q(j|i)(p_{ij} - c_{ij} - \frac{1}{\beta_{ij}} - R_i)] \\
&= -\beta_{ij}v_{ij}(R_i - \theta_i^* + p_{ij} - c_{ij} - \frac{1}{\beta_{ij}} - R_i) = -\beta_{ij}v_{ij}(p_{ij} - c_{ij} - \frac{1}{\beta_{ij}} - \theta_i^*).
\end{aligned}$$

Let $\partial g_i/\partial p_{ij} = 0$, we have $\theta_i = p_{ij} - c_{ij} - 1/\beta_{ij}$ for all $j \in S_i$ and $\theta_i = \theta_i^*$ at optimality. □

According to Lemma 3, the inner pricing problem of problem (2.7) is completely solved by setting $p_{ij} = \theta_i^* + c_{ij} + 1/\beta_{ij}$. Therefore, next proposition shows that the bilevel optimization problem (2.7) can be further simplified and solved efficiently.

Proposition 2. *The basic joint subproblem (2.7) at nest $i \in M$ with scalar θ_i^* can be reformulated as follows*

$$\max_{S_i \in \Im_i} \sum_{j \in S_i} \frac{v_{ij}(\theta_i^* + c_{ij} + 1/\beta_{ij})}{\beta_{ij}} - v_{i0}\mathbf{1}(S_i \neq \varnothing)\theta_i^*. \tag{2.8}$$

Furthermore, problem (2.8) can be solved within $O(n_i \log n_i)$ operations.

Proof. According to Lemma 3, problem (2.7) can be written as

$$\max_{S_i \in \Im_i} V_i(S_i, \theta_i^*)^{1\gamma_i}(R_i(S_i, \theta_i^*) - \theta_i^*),$$

which can be reformulated as problem (2.8) by the definition of $V_i(S_i, \theta_i^*)$ and $R_i(S_i, \theta_i^*)$. The goal of problem (2.8) is to locate $\mathbb{C}_i$ products with largest ratio of preference weight to price sensitivity parameter from N_j products within nest $i \in M$ and compare their summation with $v_{i0}\theta_i^*$, which only depends on the ordering of the ratio $v_{ij}(\theta_i^* + c_{ij} + 1/\beta_{ij})/\beta_{ij}$. If the largest summation is greater than $v_{i0}\theta_i^*$, then the optimal solution to problem (2.8) is corresponding $\mathbb{C}_i$ products, otherwise the optimal solution is an empty set. Hence the total number of operations needed is $O(N_j \log N_j)$ (sorting) + $O(\mathbb{C}_i)$ (printing the output) = $O(n_i \log n_i)$. □

The intuition of solving problem (2.8) is to include as many "good" products as possible to see whether the summation of the ratio $\sum_{j \in S_i} v_{ij}(\theta_i^* + c_{ij} + 1/\beta_{ij})/\beta_{ij}$ is larger than $v_{i0}\theta_i^*$ or not. If so, the optimal solution is nonempty with size $\mathbb{C}_i$; otherwise it is empty. However, if $v_{i0} = 0$, the optimal solution is always nonempty and the size of it is $\mathbb{C}_i$, which will be shown in next lemma. By observing the fact that θ_i^* is a unknown nonnegative scalar, we let $\tilde{S}_i(\theta_i)$ be an optimal solution to the following problem

$$\max_{S_i \in \Im_i} \sum_{j \in S_i} \frac{v_{ij}(\theta_i + c_{ij} + 1/\beta_{ij})}{\beta_{ij}} - v_{i0}\mathbf{1}(S_i \neq \varnothing)\theta_i, \tag{2.9}$$

Define $\mathcal{A}_i = \{\tilde{S}_i(\theta_i) : \theta_i \in \mathbb{R}_{\geqslant 0}\}$, then we can see that $\mathcal{A}_i$ includes $\tilde{S}_i(\theta_i^*)$ and $(\tilde{S}_1(\theta_1^*), \tilde{S}_1(\theta_1^*), ..., \tilde{S}_m(\theta_m^*); \theta_1^*, \theta_2^*, ..., \theta_m^*)$ is an optimal solution to problem (4.1) according to Claim 3. The next lemma shows the property of the size of $\tilde{S}_i(\theta_i)$.

Lemma 4. *For the optimal solution $\tilde{S}_i(\theta_i)$ to problem (4.9) with $\forall \theta_i \in \mathbb{R}_{\geqslant 0}$, we have either $|\tilde{S}_i(\theta_i)| = \mathbb{C}_i$ or $|\tilde{S}_i(\theta_i)| = 0$. Moreover, $|\tilde{S}_i(\theta_i)|$ always equals to $\mathbb{C}_i$ if $v_{i0} = 0$.*

Proof. Proof: We prove this lemma by contradiction, assume that $\tilde{S}_i(\theta_i)$ is nonempty and $|\tilde{S}_i(\theta_i)| < \mathbb{C}_i$. Then there exists a product j' satisfying that $j' \notin \tilde{S}_i(\theta_i)$ but $j' \in N_i$ since $\mathbb{C}_i \leqslant n_i$. Let $\tilde{S}_i'(\theta_i) = \tilde{S}_i(\theta_i) \cup j'$, then we have $\tilde{S}_i'(\theta_i)$ is feasible since $|\tilde{S}_i'(\theta_i)| \leqslant \mathbb{C}_i$ and $\tilde{S}_i'(\theta_i)$ strictly dominates $\tilde{S}_i(\theta_i)$ since $\sum_{j \in \tilde{S}_i(\theta_i)} v_{ij}(\theta_i + c_{ij} + 1/\beta_{ij})/\beta_{ij} - v_{i0}\theta_i > \sum_{j \in \tilde{S}_i'(\theta_i)} v_{ij}(\theta_i + c_{ij} + 1/\beta_{ij})/\beta_{ij} - v_{i0}\theta_i$, which contradicts with the hypothesis that $\tilde{S}_i(\theta_i)$ is the optimal solution to problem (4.9) at θ_i and $|\tilde{S}_i(\theta_i)| < \mathbb{C}_i$. Thus when $\tilde{S}_i(\theta_i)$ is nonempty, we have $|\tilde{S}_i(\theta_i)| = \mathbb{C}_i$; when $\tilde{S}_i(\theta_i)$ is empty, we have $|\tilde{S}_i(\theta_i)| = 0$. Furthermore, if $v_{i0} = 0$, then $\tilde{S}_i(\theta_i)$ is always nonempty since the objective value of an empty assortment is 0, which is strictly less than the objective value of any nonempty feasible assortment.

□

Lemma 4 is insightful when $\mathbb{C}_i = n_i$ and $v_{i0} = 0$, since the optimal assortment at nest i in this case is straightforward: $S_i^* = N_i$. If it applies to all nests, the joint optimization problem (4.1) is reduced to the pricing problem under the nested logit model with no-purchase options.

For problem (4.9) with $\theta_i \in \mathbb{R}_{\geqslant 0}$, the objective function of which is rewritten as $\sum_{j \in S_i} \exp(h_j(\theta_i))$, where linear function $h_j(\theta_i)$ is defined as $h_j(\theta_i) = \tilde{\alpha}_{ij} - \beta_{ij}\theta_i$ and $\tilde{\alpha}_{ij} = \alpha_{ij} - \beta_{ij}c_{ij} - \log(\beta_{ij}) - 1$ for all $j \in S_i$. We remark that only the ordering of these lines $h_j(\theta_i)$ matters for a given θ_i. [50] show problem (4.9) can be solved in $O(n_i^2)$ operations and the size of $\mathcal{A}_i$ is $O(n_i)$. Furthermore, next lemma shows the discontinuous property of $V_i(\tilde{S}_i(\theta_i), \theta_i)$.

Lemma 5. *At the changing point θ_i' of $\tilde{S}_i(\theta_i)$, $V_i(\tilde{S}_i(\theta_i), \theta_i)$ decreases discontinuously, both $\omega_i(\tilde{S}_i(\theta_i), \theta_i)$ and $\delta_i(\tilde{S}_i(\theta_i), \theta_i)$ increase discontinuously.*

Proof. Proof: Let $\underline{S}_i = \lim_{\epsilon \to 0} \tilde{S}_i(\theta_i' - \epsilon)$ and $\overline{S}_i = \lim_{\epsilon \to 0} \tilde{S}_i(\theta_i' + \epsilon)$ for a small $\epsilon > 0$. At any changing point, one product with larger price-sensitivity parameter of current assortment would be replaced by another product with smaller price-sensitivity parameter. Without loss of generality, assume that θ_i' is the intersection point of lines $h_{j_1}(\theta_i)$ and $h_{j_2}(\theta_i)$ and product j_1 is replaced by product j_2. We have $\overline{S}_i = (\underline{S}_i \backslash \{j_1\}) \cup \{j_2\}$ for $j_1, j_2 \in N_i$. Since product j_1 is replaced by product j_2 at θ_i', we have $h_{ij_1}(\theta_i') = h_{ij_2}(\theta_i')$ and $\beta_{ij_1} > \beta_{ij_2}$, which implies that $v_{ij_1}(\theta_i' + c_{ij_1} + 1/\beta_{ij_1})/\beta_{ij_1} = v_{ij_2}(\theta_i' + c_{ij_2} + 1/\beta_{ij_2})/\beta_{ij_2}$ and $v_{ij_1}(\theta_i' + c_{ij_1} + 1/\beta_{ij_1}) > v_{ij_2}(\theta_i' + c_{ij_2} + 1/\beta_{ij_2})$, thus $V_i(\underline{S}_i, \theta_i') > V_i(\overline{S}_i, \theta_i')$. It also implies that

$$\begin{aligned} \omega_i(\underline{S}_i, \theta_i') &= (\frac{1}{\gamma_i} - 1) \sum_{j \in \underline{S}_i} \frac{v_{ij}(\theta_i' + c_{ij} + 1/\beta_{ij})}{\beta_{ij}} \frac{1}{V_i(\underline{S}_i, \theta_i')^{1/\gamma_i}} \\ &< (\frac{1}{\gamma_i} - 1) \sum_{j \in \overline{S}_i} \frac{v_{ij}(\theta_i' + c_{ij} + 1/\beta_{ij})}{\beta_{ij}} \frac{1}{V_i(\overline{S}_i, \theta_i')^{1/\gamma_i}} = \omega_i(\overline{S}_i, \theta_i'), \end{aligned}$$

where the inequality holds because $\sum_{j \in \underline{S}_i} v_{ij}(\theta_i' + c_{ij} + 1/\beta_{ij})/\beta_{ij} = \sum_{j \in \overline{S}_i} v_{ij}(\theta_i' + c_{ij} + 1/\beta_{ij})/\beta_{ij}$. Therefore, $V_i(\tilde{S}_i(\theta_i), \theta_i)$ drops discontinuously at θ_i'. For similar reasons, we have $\delta_i(\underline{S}_i, \theta_i') < \delta_i(\overline{S}_i, \theta_i')$ since $\tau_i(\underline{S}_i, \theta_i') > \tau_i(\overline{S}_i, \theta_i')$.

□

Since $\mathcal{A}_i = \{\tilde{S}_i(\theta_i) : \theta_i \in \mathbb{R}_{\geqslant 0}\}$ includes an optimal solution at nest $i \in M$ to problem (4.1), using $\mathcal{A}_i$ to replace $\Im_i$ in problem (2.6) would not affect the optimality. Furthermore,we have one $\theta_i \in \mathbb{R}_{\geqslant 0}$ corresponds to one assortment $\tilde{S}_i(\theta_i)$ in collection $\mathcal{A}_i$, thus problem (2.6) can be further reformulated as *assortment subproblem* in terms of optimizing over $\theta_i \in \mathbb{R}_{\geqslant 0}$ as follows

$$\begin{aligned} \max_{\theta_i \in \mathbb{R}_{\geqslant 0}} \quad & \frac{1}{1-\gamma_i} V_i(\tilde{S}_i(\theta_i), \theta_i)(\theta_i - Z^*) \\ \text{s.t.} \quad & Z^* = \delta_i(\tilde{S}_i(\theta_i), \theta_i)\theta_i - \omega_i(\tilde{S}_i(\theta_i), \theta_i). \end{aligned} \tag{2.10}$$

For the constraints in problem (4.11), there may not exist a one-to-one correspondence between Z^* and θ_i according to Lemma 18, thus the feasible region of problem (4.11) is a set of points satisfying the constraints, the size of which may be greater than one. Therefore, we need to evaluate the objective function at those feasible points and select the point that has the maximum objective value. This is different from what we have in Lemma 2 since $\tilde{S}_i(\theta_i)$ in the constraints of problem (4.11) is a function of θ rather than a fixed assortment as in the constraints of problem (2.6).

Since Z^* is a unknown nonnegative scalar, we let $\mathcal{F}_i(z)$ be the optimal solution to the following problem

$$\begin{aligned} \max_{\theta_i \in \mathbb{R}_{\geqslant 0}} \quad & \frac{1}{1-\gamma_i} V_i(\tilde{S}_i(\theta_i), \theta_i)(\theta_i - z) \\ \text{s.t.} \quad & z = \delta_i(\tilde{S}_i(\theta_i), \theta_i)\theta_i - \omega_i(\tilde{S}_i(\theta_i), \theta_i), \end{aligned} \tag{2.11}$$

where $z \in \mathbb{R}_{\geqslant 0}$. If we define $\hat{S}_i(z) = \tilde{S}_i(\mathcal{F}_i(z))$ and $\hat{\mathscr{A}_i} = \{\hat{S}_i(z) : z \in \mathbb{R}_{\geqslant 0}\}$, it follows that $\hat{\mathscr{A}_i}$ includes $\hat{S}_i(Z^*)$ that is an optimal solution to problem (2.6). According to Claim 2, $\bigcup_{i \in M} \hat{S}_i(Z^*)$ is optimal to problem (4.1). If problem (2.11) can be easily solved for $\forall z \in \mathbb{R}_{\geqslant 0}$, then we claim that collection $\mathcal{A} = \{\bigcup_{i \in M} \hat{S}_i(z) : z \in \mathbb{R}_{\geqslant 0}\}$ includes an optimal solution to problem (4.1). We still have concerns about objective function in problem (2.11), if it does not have nice properties, the size of $\mathcal{A}$ would be so large that S^* and Z^* are hard to be obtained since z takes value in $\mathbb{R}_{\geqslant 0}$ in collection $\mathcal{A}$.

Let $l_i(\theta_i) = \delta_i(\tilde{S}_i(\theta_i), \theta_i)\theta_i - \omega_i(\tilde{S}_i(\theta_i), \theta_i)$, then $\{l_i^{-1}(z) : z \in \mathbb{R}_{\geqslant 0}\}$ is a set of points that satisfy the constraints in problem (2.11). Without loss of generality, we let $\theta_i^1, \theta_i^2, ..., \theta_i^K \in \{l_i^{-1}(z) : z \in \mathbb{R}_{\geqslant 0}\}$ and $S_i^k = \tilde{S}_i(\theta_i^k)$ for $k = 1, 2, ..., K$. The objective function in problem (2.11) for a fixed assortment S_i^k can be denoted as $T_i^k(z) = V_i(S_i^k, \theta_i^k)(\theta_i^k - z)/(1-\gamma_i)$ where θ_i^k is implicitly defined in $z = \delta_i(S_i^k, \theta_i^k)\theta_i^k - \omega_i(S_i^k, \theta_i^k)$. Next proposition shows that $T_i^k(z)$ is a decreasing convex function of z.

Proposition 3. *$T_i^k(z)$ is a decreasing convex function of z with $-V_i^k(S_i^k, \theta_i^k)$ as its first derivative. Moreover, if we have the assumption that $V_i(S_i^{k_1}, \theta_i^{k_1})$ does not intersect with $V_i(S_i^{k_2}, \theta_i^{k_2})$ in z domain, then $T_i^{k_1}(z)$ and $T_i^{k_2}(z)$ intersect at most once, where $k_1, k_2 \in \{1, 2, ..., K\}$.*

Proof. Proof: For ease of presentation, we denote $V_i^k = V_i^k(S_i^k, \theta_i^k)$, $T_i^k = T_i^k(z)$, $\omega_i^k = \omega_i(S_i^k, \theta_i^k)$, $\delta_i^k = \delta_i(S_i^k, \theta_i^k)$ and $u_i^k = u_i(S_i^k, \theta_i^k) = \sum_{j \in S_i} \beta_{ij} Q_{j|i}(S_i^k, \theta_i^k)$. We have

$$\begin{aligned}\frac{\partial T_i^k}{\partial z} &= \frac{1}{1-\gamma_i}[\frac{\partial V_i^k}{\partial \theta_i^k}\frac{\partial \theta_i^k}{\partial z}(\theta_i^k - z) + V_i^k(\frac{\partial \theta_i^k}{\partial z} - 1)] \\ &= \frac{1}{1-\gamma_i}[(-\gamma_i v_i^k u_i^k)\frac{\partial \theta_i^k}{\partial z}(\theta_i^k - \delta_i^k\theta_i^k + \omega_i^k) + V_i^k(\frac{\partial \theta_i^k}{\partial z} - 1)] \\ &= \frac{\gamma_i V_i^k}{1-\gamma_i}[\frac{1/\gamma_i - \omega_i^k u_i^k + (\delta_i^k - 1)u_i^k\theta_i^k}{\partial z/\partial \theta_i^k} - \frac{1}{\gamma_i}] = -V_i^k,\end{aligned}$$

where the last equality is due to the fact that $\partial z/\partial \theta_i^k = 1/\gamma_i - \omega_i^k u_i^k + (\delta_i^k - 1)u_i^k\theta_i^k$. The second derivative is

$$\frac{\partial^2 T_i^k}{\partial z^2} = -\frac{\partial V_i^k}{\partial \theta_i^k}\frac{\partial \theta_i^k}{\partial z} = \gamma_i V_i^k u_i^k \omega_i^k(\frac{1}{\gamma_i} - \omega_i^k u_i^k + (\delta_i^k - 1)u_i^k\theta_i^k)^{-1} > 0,$$

where the last inequality is due to Lemma 2. Without loss of generality, assume that $V_i^{k_1} > V_i^{k_2}$ in z domain, we get

$$\frac{\partial(T_i^{k_1} - T_i^{k_2})}{\partial z} = -(V_i^{k_1} - V_i^{k_2}) < 0,$$

which implies that $T_i^{k_1}$ and $T_i^{k_2}$ intersect at most once in z domain.

□

We have $|\mathcal{A}_i|$ decreasing convex curves, any two of which intersect at most once under the assumption in Proposition 3. Similar to the approach of solving problem (4.9), only the ordering of these convex curves matters to solving problem (2.11). In order to get collection $\hat{\mathcal{A}}_i$, it suffices to calculate the pairwise intersection points of these curves and select the highest curve, since $\tilde{S}_i(\mathcal{F}_i(z))$ does not change when z takes value between two consecutive intersection points. This calculation can be done very efficiently by binary search or golden ratio search due to its convexity nature. We immediately have the following corollary regarding the size of $\hat{\mathscr{A}}_i$.

Corollary 2. *For any $i \in M$, the size of $\hat{\mathscr{A}}_i$ is less than or equal to the size of $\mathscr{A}_i$ under the assumption in Proposition 3.*

If the assumption in Proposition 3 is met in our problem setting, it is possible that collection $\mathscr{A}$ is polynomially bounded. Fortunately, we mange to show $|\mathscr{A}|$ is $O(N)$ by proving the assumption in Proposition 3 is satisfied.

Theorem 1. *The collection $\mathcal{A} = \{\bigcup_{i \in M} \hat{S}_i(z) : z \in \mathbb{R}_{\geqslant 0}\}$ includes an optimal solution to problem (4.1). Furthermore, the size of $\mathcal{A}$ is bounded by $O(N)$.*

Proof. Proof: First, we define a subset of pairwise intersection points of $|\tilde{\mathscr{A}}_i|$ decreasing convex curves in problem (2.11) as $\mathcal{I}_i = \{z_i^0, z_i^1, \cdots, z_i^{U_i - 1}, z_i^{U_i}\}$ where $z_i^0 = -\infty$ and $z_i^{U_i} = \infty$, such that $\hat{S}_i(z) = \tilde{S}_i(\mathcal{F}_i(z))$ does not change when $z \in [z_i^{u-1}, z_i^u]$ for $u = 1, \cdots, U_i$. We get $|\mathcal{A}| = \sum_{i \in M} U_i$ according to the definition of collection $\mathcal{A}$.

Next, we prove the assumption in Proposition 3 is met. We define the set of changing points of $\tilde{S}_i(\theta_i)$ as $\mathcal{C}_i = \{\theta_i^0, \theta_i^1, \cdots, \theta_i^{D_i - 1}, \theta_i^{D_i}\}$ where $\theta_i^0 = -\infty$ and $\theta_i^{D_i} = \infty$, such that $\tilde{S}_i(\theta_i)$ does not change when $\theta_i \in [\theta_i^{d-1}, \theta_i^d]$ for $d = 1, \cdots, D_i$, where $D_i = |\mathcal{A}_i|$ according to the definition of set $\mathcal{A}_i$. We prove that for any two different assortments $S_i^{k_1}, S_i^{k_2} \in \mathcal{A}_i$, $V_i(S_i^{k_1}, \theta_i^{k_1})$ does not intersect with $V_i(S_i^{k_2}, \theta_i^{k_2})$ in z domain, where $\theta_i^{k_l}$ satisfies $z = \delta_i(S_i^{k_l}, \theta_i^{k_l})\theta_i^{k_l} - \omega_i(S_i^{k_l}, \theta_i^{k_l})$ for $l = 1, 2$. Without loss of generality, assume that $S_i^{k_1} = \tilde{S}_i(\theta_i^{k_1})$ where $\theta_i^{k_1} \in [\theta_i^{d_1 - 1}, \theta_i^{d_1}]$ and $S_i^{k_2} = \tilde{S}_i^{k_2}(\theta_i)$ where $\theta_i^{k_2} \in [\theta_i^{d_2 - 1}, \theta_i^{d_2}]$ with $d_1 < d_2$. In z domain, $V_i(S_i^{k_l}, z)$ is defined on $[\delta_i(S_i^{k_l}, \theta_i^{d_l - 1})\theta_i^{d_l - 1} - \omega_i(S_i^{k_l}, \theta_i^{d_l - 1}), \delta_i(S_i^{k_l}, \theta_i^{d_l})\theta_i^{d_l} - \omega_i(S_i^{k_l}, \theta_i^{d_l})]$ for $l = 1, 2$. By Lemma 18, we have $V_i(S_i^{k_1}, \theta_i^{k_1}) > V_i(S_i^{k_2}, \theta_i^{k_2})$, thus $V_i(S_i^{k_1}, \theta_i^{k_1})$ does not intersect with $V_i(S_i^{k_2}, \theta_i^{k_2})$ in z domain. By Proposition 3, $T_i^{k_1}(z)$ and $T_i^{k_2}(z)$ intersect at most once, where $k_1, k_2 \in \{1, 2, ..., K\}$, which implies that $U_i \leqslant D_i$ since the number changing points of $\hat{S}_i(z)$ from these D_i decreasing convex curves is at most D_i according to Corollary 2. As a result, we get $|\mathcal{A}| = \sum_{i \in M} U_i \leqslant \sum_{i \in M} D_i = \sum_{i \in M} |\mathcal{A}_i|$. Since $|\mathcal{A}_i|$ is bounded by $O(n_i)$, we have $|\mathcal{A}|$ is bounded by $O(N)$ where $N = \sum_{i \in M} n_i$ by definition.

□

Imaging Theorem 1 does not hold, then $\mathcal{A} = \mathcal{A}_1 \times \mathcal{A}_2 \times \cdots \times \mathcal{A}_m$ and we have $|\mathcal{A}| = O(n_{\max}^m)$ where $n_{\max} = \max_{i \in M} n_i$. Then we need at least $n_{\max}^m$ grid points of z to find the optimal solution to problem (4.1), which is intractable. In other words, $\hat{S}(z)$ is too sensitive to even a small change in z which leads to the intractability without the support of Theorem 1.

Fixed Point Representation

In this section, we first present a fixed point representation of problem (4.1). Second we show the profit function is unimodal of z. Third, we prove Z^* is the fixed point of a piecewise convex function.

Define $\tilde{S}(z) = \bigcup_{i\in M} \hat{S}_i(z)$, the optimal profit Z^* is the fixed point of the total expected profit function $\Pi(\tilde{S}(z), z)$, which is defined as follows

$$\Pi(\tilde{S}(z), z) = \frac{\sum_{i\in M} V_i(\hat{S}_i(z), \mathcal{F}_i(z)) R_i(\hat{S}_i(z), \mathcal{F}_i(z))}{v_0 + \sum_{i\in M} V_i(\hat{S}_i(z), \mathcal{F}_i(z))}.$$

Next lemma shows the properties of function $\Pi(\tilde{S}(z), z)$.

Lemma 6. *The expected profit function* $\Pi(\tilde{S}(z), z)$ *is unimodal of* z *and* Z^* *is the fixed point of* $\Pi(\tilde{S}(z), z)$.

Proof. Proof: For ease of reading, we let $\Pi = \Pi(\tilde{S}(z), z)$, $\mathcal{F}_i = \mathcal{F}_i(z)$, $V_i = V_i(\hat{S}_i(z), \mathcal{F}_i(z))$, $R_i = R_i(\hat{S}_i(z), \mathcal{F}_i(z))$, $Q_i = Q_i(\hat{S}_i(z), \mathcal{F}_i(z))$, $\omega_i = \omega_i(\hat{S}_i(z), \mathcal{F}_i(z))$, $\tau_i = \tau_i(\hat{S}_i(z), \mathcal{F}_i(z))$, $\delta_i = \delta_i(\hat{S}_i(z), \mathcal{F}_i(z))$ and $u_i = u_i(\hat{S}_i(z), \mathcal{F}_i(z))$. We have

$$\frac{\partial \Pi}{\partial \mathcal{F}_i} = \frac{\frac{\partial V_i}{\partial \mathcal{F}_i} R_i + V_i \frac{\partial R_i}{\partial \mathcal{F}_i} - \Pi \frac{\partial V_i}{\partial \mathcal{F}_i}}{v_0 + \sum_{i\in M} V_i} = \gamma_i Q_i u_i (\Pi - z),$$

since $\partial V_i/\partial \mathcal{F}_i = -\gamma_i V_i u_i$, $R_i = \tau_i \mathcal{F}_i + \gamma_i \omega_i/(1-\gamma_i)$ and $\partial R_i/\partial \mathcal{F}_i = \gamma_i[\omega_i u_i - (\delta_i - 1)u_i \mathcal{F}_i]/(1-\gamma_i)$. We also get $\partial z/\partial \mathcal{F}_i = 1/\gamma_i - \omega_i u_i + (\delta_i - 1)u_i \mathcal{F}_i > 0$ from Lemma 2, it follows that

$$\frac{\partial \Pi}{\partial z} = \sum_{i\in M} \frac{\partial \Pi/\partial \mathcal{F}_i}{\partial z/\partial \mathcal{F}_i} = (\Pi - z) \sum_{i\in M} \frac{\gamma_i^2 Q_i u_i}{1 - \gamma_i \omega_i u_i + (\delta_i - 1)\gamma_i u_i \mathcal{F}_i}.$$

To prove the unimodality of Π, it suffices to show that for any discontinuous point z' of Π, Π increases discontinuously at $z' \leqslant Z^*$ and decreases discontinuously at $z' > Z^*$, since it implies that Π increases when $z \leqslant Z^*$ and decreases when $z > Z^*$, which ensures unimodality of Π. We obtain that $V_i(\underline{S}_i, z')(R_i(\underline{S}_i, z') - z') = V_i(\overline{S}_i, z')(R_i(\overline{S}_i, z') - z')$ where $\underline{S}_i = \lim_{\epsilon\to 0} \hat{S}_i(z' - \epsilon)$ and $\overline{S}_i = \lim_{\epsilon\to 0} \hat{S}_i(z' + \epsilon)$, since z' is the intersection point of $V_i(\underline{S}_i, z')(R_i(\underline{S}_i, z') - z')$ and $V_i(\overline{S}_i, z')(R_i(\overline{S}_i, z') - z')$. For notational brevity, we denote $\underline{V}_i = V_i(\underline{S}_i, z')$, $\underline{R}_i = R_i(\underline{S}_i, z')$, $\overline{V}_i = V_i(\overline{S}_i, z')$ and $\overline{R}_i = R_i(\overline{S}_i, z')$. Let $\underline{S} = \lim_{\epsilon\to 0} \tilde{S}(z' - \epsilon)$ and $\overline{S} = \lim_{\epsilon\to 0} \tilde{S}(z' + \epsilon)$.

We obtain $\underline{V}_i \underline{R}_i - \overline{V}_i \overline{R}_i = (\underline{V}_i - \overline{V}_i)z$, it follows that

$$\begin{aligned}
\Pi(\underline{S}, z') &= \frac{\sum_{j\neq i} V_j R_j + \underline{V}_i \underline{R}_i}{v_0 + \sum_{j\neq i} V_j + \underline{V}_i} < \frac{\sum_{j\neq i} V_j R_j + \underline{V}_i \underline{R}_i - (\underline{V}_i \underline{R}_i - \overline{V}_i \overline{R}_i)}{v_0 + \sum_{j\neq i} V_j + \underline{V}_i - (\underline{V}_i - \overline{V}_i)} \\
&= \frac{\sum_{j\neq i} V_j R_j + \underline{V}_i \underline{R}_i - (\underline{V}_i - \overline{V}_i)z'}{v_0 + \sum_{j\neq i} V_j + \underline{V}_i - (\underline{V}_i - \overline{V}_i)} = \Pi(\overline{S}, z'),
\end{aligned}$$

when $z' \leqslant Z^*$ because of the fact that 1) $\underline{V}_i > \overline{V}_i$ according to Lemma 18; 2) $z' \leqslant \Pi(\underline{S}, z')$; 3) function $h(x) = \frac{A-xz'}{B-x}$ is increasing for $x > 0$ if $z' < A/B$. Similarly, we have $\Pi(\underline{S}, z') > \Pi(\overline{S}, z')$ when $z' > Z^*$. Therefore, the unimodality of Π holds. Furthermore, Π reaches its maximum when $\Pi = z$. □

We can use binary search for the unimodal profit function $\Pi(\tilde{S}(z), z)$ to find the optimal Z^*. If we define $f(z) = \sum_{i\in M} V_i(\hat{S}_i(z), \mathcal{F}_i(z))[R_i(\hat{S}_i(z), \mathcal{F}_i(z)) - z]$, then the fixed point representation $\Pi(\tilde{S}(z), z) = z$ can be rewritten as $f(z) = v_0 z$, implying that Z^* is the fixed point of function $f(z)/v_0$. Next proposition shows the piecewise convexity property of this representation.

Proposition 4. *$f(z)$ is a decreasing piecewise convex function of z. Moreover, the first derivative of $f(z)$ is increasing and there exists a unique solution Z^* that satisfies $f(z) = v_0 z$.*

Proof. Proof: Let set $\mathcal{I}_i = \{z_i^0, z_i^1, \cdots, z_i^{K_i-1}, z_i^{K_i}\}$ be the set of pairwise intersection points in problem (2.11) where $z_i^0 = -\infty$ and $z_i^{K_i} = \infty$, such that $\hat{S}_i(z) = \tilde{S}_i(\mathcal{F}_i(z))$ does not change when $z \in [z_i^{k-1}, z_i^k]$ for $k = 1, \cdots, K_i$. Let $T_i^k(z) = V_i(\hat{S}_i(z), \mathcal{F}_i(z))[R_i(\hat{S}_i(z), \mathcal{F}_i(z)) - z]$ for $z \in [z_i^{k-1}, z_i^k]$, then $T_i^k(z)$ is a decreasing convex function according to Proposition 3, thus $T_i(z) = V_i(\hat{S}_i(z), \mathcal{F}_i(z))[R_i(\hat{S}_i(z), \mathcal{F}_i(z)) - z]$ for $z \in \mathbb{R}_{\geqslant 0}$ is a piecewise decreasing convex function in $\mathbb{R}_{\geqslant 0}$, where the first derivative $\partial T_i(z)/\partial z$ changes discontinuously at z_i^k. To prove the first derivative of $f(z)$ is increasing, it suffices to show $\partial T_i(z)/\partial z$ increases discontinuously at z_i^k. Let $\underline{S}_i = \lim_{\epsilon\to 0} \hat{S}_i(z_i^k - \epsilon)$ and $\overline{S}_i = \lim_{\epsilon\to 0} \hat{S}_i(z_i^k + \epsilon)$, then we get $V_i(\underline{S}_i, z_i^k) > V_i(\overline{S}_i, z_i^k)$ since $V_i(\underline{S}_i, z_i^k)$ and $V_i(\overline{S}_i, z_i^k)$ do not intersect in z domain, which is shown in the proof of Theorem 1. It follows that

$$\lim_{z\to z_i^{k-}} \frac{\partial T_i(z)}{\partial z} - \lim_{z\to z_i^{k+}} \frac{\partial T_i(z)}{\partial z} = -(V_i(\underline{S}_i, z_i^k) - V_i(\overline{S}_i, z_i^k)) < 0,$$

which implies that $\partial T_i(z)/\partial z$ increases discontinuously at z_i^k for $k \in \{1, 2, ..., 3\}$. $f(z) = \sum_{i\in M} T_i(z)$ is also piecewise convex since piecewise convexity is preserved under addition. The fixed point of $f(z)/v_0$ is unique since $f(z)$ is a decreasing piecewise convex function. □

This piecewise convexity property can further facilitate computing Z^* [46]. To summarize, we propose the assortment subproblem by solving the basic assortment subproblem and replacing the feasible region of joint subproblem. Then we show there exists a piecewise convex fixed point representation of problem (4.1) which can be solved efficiently.

2.5 Joint Optimization Under Space Constraints

In this section, we consider the joint constrained assortment and price optimization problem (4.1) under space constraints. We show a 2-approximate solution can be found through a piecewise convex fixed point representation. Let $Z^\alpha = \Pi(S^\alpha, \mathbf{P}^\alpha)$ be an α-approximate solution to problem (4.1) where $\alpha Z^\alpha \geqslant Z^*$. First, we show how to construct a collection $\Im^\alpha$

that contains an α-approximate solution to problem (4.1). Then, we manage to obtain $\mathcal{A}^\alpha \subseteq \Im^\alpha$ that also includes an α-approximate solution to problem (4.1) with $|\mathcal{A}^\alpha| = O(n_{\max} N)$. Third, we show Z^α is the fixed point of a piecewise convex function.

As in Section 2.4, θ_i^* is still defined as $\theta_i^* = \gamma_i Z^* + (1-\gamma_i)R_i(S_i^*, \theta_i^*)$ if $S_i^* \neq \emptyset$ and $\theta_i^* = Z^*$ otherwise, then we have the following claim.

Claim 4. *For all nest $i \in M$, if $(\tilde{S}_i^\alpha, \tilde{\mathbf{P}}_i^\alpha)$ satisfies $V_i(\tilde{S}_i^\alpha, \tilde{\mathbf{P}}_i^\alpha)^{1\gamma_i}(\alpha R_i(\tilde{S}_i^\alpha, \tilde{\mathbf{P}}_i^\alpha) - \theta_i^*) \geqslant V_i(S_i^*, \mathbf{P}_i^*)^{1\gamma_i}(R_i(S_i^*, \mathbf{P}_i^*) - \theta_i^*)$, then $(\tilde{S}_1^\alpha, \tilde{S}_2^\alpha, ..., \tilde{S}_m^\alpha; \tilde{\mathbf{P}}_1^\alpha, \tilde{\mathbf{P}}_2^\alpha, ..., \tilde{\mathbf{P}}_m^\alpha)$ is an α-approximate solution.*

The proof of this claim is omitted because it is similar to the proof of Claim 3. One can check that Lemma 3 continues to holds under space constraints, thus we can focus on $\max_{S_i \in \Im_i} V_i(S_i, \theta_i^*)^{1\gamma_i}(R_i(S_i, \theta_i^*) - \theta_i^*)$. For $i \in M$, we let $\bar{S}_i'(\theta_i)$ be an α-approximate solution to the following problem

$$\max_{S_i \in \Im_i} \sum_{j \in S_i} \frac{v_{ij}(\theta_i + c_{ij} + 1/\beta_{ij})}{\beta_{ij}}. \tag{2.12}$$

Note the objective function in problem (2.12) is different from the one in problem (4.9) under cardinality constraints. Moreover, [18] show problem (2.12) is NP-hard. Let $\tilde{S}_i^\alpha(\theta_i) = \bar{S}_i'(\theta_i)$ if $\sum_{j \in \bar{S}_i'(\theta_i)} v_{ij}(\theta_i + c_{ij} + 1/\beta_{ij})/\beta_{ij} \geqslant v_{i0}\theta_i$; otherwise $\tilde{S}_i(\theta_i) = \emptyset$. Denote $\tilde{\mathbf{P}}_i^\alpha = \theta_i^*(\tilde{S}_i^\alpha(\theta_i^*))$. More precisely, $\tilde{\mathbf{P}}_i^\alpha = (\tilde{p}_{ij}^\alpha : j \in \tilde{S}_i^\alpha(\theta_i^*))$ where $\tilde{p}_{ij}^\alpha = \theta_i^* + c_{ij} + 1/\beta_{ij}$, then we have the following lemma.

Lemma 7. $(\tilde{S}_1^\alpha(\theta_i^*), \tilde{S}_2^\alpha(\theta_i^*), ..., \tilde{S}_m^\alpha(\theta_i^*); \tilde{\mathbf{P}}_1^\alpha, \tilde{\mathbf{P}}_2^\alpha, ..., \tilde{\mathbf{P}}_m^\alpha)$ *is an α-approximate solution to problem (4.1).*

Proof. Proof: For ease of reading, we denote $V_i^* = V_i(S_i^*, \mathbf{P}_i^*)$, $R_i^* = R_i(S_i^*, \mathbf{P}_i^*)$, $\tilde{S}_i^\alpha = \tilde{S}_i^\alpha(\theta_i^*)$, $\tilde{V}_i = V_i(\tilde{S}_i^\alpha(\theta_i^*), \tilde{\mathbf{P}}_i^\alpha)$, $\tilde{R}_i = R_i(\tilde{S}_i^\alpha(\theta_i^*), \tilde{\mathbf{P}}_i^\alpha)$ and $v_{ij} = v_{ij}(\theta_i^* + c_{ij} + 1/\beta_{ij})$. According to Claim 4, it suffices to prove $\tilde{V}_i^{1/\gamma_i}(\alpha\tilde{R}_i - \theta_i^*) \geqslant (V_i^*)^{1/\gamma_i}(R_i^* - \theta_i^*)$. If $S_i^* \neq \emptyset$, then we get

$$\begin{aligned}
\tilde{V}_i^{1/\gamma_i}(\alpha\tilde{R}_i - \theta_i^*) &= \alpha \sum_{j \in \tilde{S}_i^\alpha} v_{ij}(\theta_i^* + \frac{1}{\beta_{ij}}) - \sum_{j \in \tilde{S}_i^\alpha} v_{ij}\theta_i^* - v_{i0}\mathbf{1}(\tilde{S}_i^\alpha \neq \emptyset)\theta_i^* \\
&\geqslant \alpha \sum_{j \in \tilde{S}_i^\alpha} \frac{v_{ij}}{\beta_{ij}} - v_{i0}\mathbf{1}(\tilde{S}_i^\alpha \neq \emptyset)\theta_i^* \geqslant \alpha \sum_{j \in \tilde{S}_i^\alpha} \frac{v_{ij}}{\beta_{ij}} - v_{i0}\theta_i^* \\
&\geqslant \sum_{j \in S_i^*} \frac{v_{ij}}{\beta_{ij}} - v_{i0}\theta_i^* = \sum_{j \in S_i^*} \frac{v_{ij}}{\beta_{ij}} - v_{i0}\mathbf{1}(S_i^* \neq \emptyset)\theta_i^* = (V_i^*)^{1/\gamma_i}(R_i^* - \theta_i^*),
\end{aligned}$$

where the last inequality holds because S_i^* is a feasible solution to problem (2.12) at $\theta_i = \theta_i^*$ and $\tilde{S}_i^\alpha(\theta_i^*)$ is an α-approximate solution. If $S_i^* = \emptyset$, then we obtain $(V_i^*)^{1/\gamma_i}(R_i^* - \theta_i^*) = 0$. This inequality also holds since $\tilde{V}_i^{1/\gamma_i}(\tilde{R}_i - \theta_i^*) \geqslant 0$, implying $\tilde{V}_i^{1/\gamma_i}(\alpha\tilde{R}_i - \theta_i^*) \geqslant 0$. □

For nest $i \in M$, we define $\mathcal{A}_i^\alpha = \{\tilde{S}_i^\alpha(\theta_i) : \theta_i \in \mathbb{R}_{\geqslant 0}\}$. Set $\alpha = 2$, the size of $\mathcal{A}_i^\alpha$ is $O(n_i^2)$ if we apply the algorithm that is described in Section 5.1 in [18] by defining linear functions

$h_j(\theta_i) = \exp(\tilde{\alpha}_{ij} - \beta_{ij}\theta_i)$ where $\tilde{\alpha}_{ij} = \alpha_{ij} - \beta_{ij}c_{ij} - \log(\beta_{ij}) - 1$ for $j \in N_i$. [18] show that $\alpha = 2$ can be further refined to $\alpha = 1/(1-\epsilon)$ under certain assumptions of w_{ij}. By noting that θ_i^* is an unknown nonnegative scalar, $\Im^\alpha$ can be constructed as the cartesian product of all $\mathcal{A}_i^\alpha$ for $i \in M$. This finding is recorded in next proposition.

Proposition 5. *Collection* $\Im^\alpha = \mathcal{A}_1^\alpha \times \mathcal{A}_2^\alpha \times ... \times \mathcal{A}_m^\alpha$ *contains an* α*-approximate solution.*

Under space constraints, let $(S^\alpha, \mathbf{P}^\alpha)$ be optimal to the following problem

$$Z^\alpha = \max_{S \in \Im^\alpha} \max_{\mathbf{P} \in \mathbb{R}^N_{\geqslant 0}} \Pi(S, \mathbf{P}). \tag{2.13}$$

The joint subproblem under space constraints is formulated as follows

$$\max_{S_i \in \mathcal{A}_i^\alpha} \max_{\mathbf{P}_i \in \mathbb{R}^{n_i}_{\geqslant 0}} V_i(S_i, \mathbf{P}_i)(R_i(S_i, \mathbf{P}_i) - Z^\alpha). \tag{2.14}$$

Similar to Claim 2, if we let $(\hat{S}_i^\alpha, \hat{\mathbf{P}}_i^\alpha)$ be optimal to problem (2.14), then $(\hat{S}_1^\alpha, \hat{S}_2^\alpha, ..., \hat{S}_m^\alpha; \hat{\mathbf{P}}_1^\alpha, \hat{\mathbf{P}}_2^\alpha, ..., \hat{\mathbf{P}}_m^\alpha)$ is an optimal solution to problem (2.13), which is an α-approximate solution to problem (4.1).

By following the exact same logic in Section 2.4, problem (2.14) can be reformulated as follows

$$\begin{aligned} \max_{S_i \in \mathcal{A}_i^\alpha} \quad & \frac{1}{1-\gamma_i} V_i(S_i, \theta_i)(\theta_i - Z^\alpha) \\ \text{s.t.} \quad & Z^\alpha = \delta_i(S_i, \theta_i)\theta_i - \omega_i(S_i, \theta_i), \end{aligned} \tag{2.15}$$

Since $\mathcal{A}_i^\alpha$ is defined as $\mathcal{A}_i^\alpha = \{\tilde{S}_i^\alpha(\theta_i) : \theta_i \in \mathbb{R}_{\geqslant 0}\}$, implying that for every $\theta_i \in \mathbb{R}_{\geqslant 0}$, there is one corresponding assortment $\tilde{S}_i^\alpha(\theta_i) \in \mathcal{A}_i^\alpha$, then problem (2.15) can be rewritten as the following optimization problem in terms of decision variable $\theta_i \in \mathbb{R}_{\geqslant 0}$:

$$\begin{aligned} \max_{\theta_i \in \mathbb{R}_{\geqslant 0}} \quad & \frac{1}{1-\gamma_i} V_i(\tilde{S}_i^\alpha(\theta_i), \theta_i)(\theta_i - Z^\alpha) \\ \text{s.t.} \quad & Z^\alpha = \delta_i(\tilde{S}_i^\alpha(\theta_i), \theta_i)\theta_i - \omega_i(\tilde{S}_i^\alpha(\theta_i), \theta_i), \end{aligned} \tag{2.16}$$

which is referred to as the assortment subproblem under space constraints. Since Z^α is an unknown nonnegative scalar, we let $\mathcal{F}_i^\alpha(z)$ be the optimal solution to the following problem

$$\begin{aligned} \max_{\theta_i \in \mathbb{R}_{\geqslant 0}} \quad & \frac{1}{1-\gamma_i} V_i(\tilde{S}_i^\alpha(\theta_i), \theta_i)(\theta_i - z) \\ \text{s.t.} \quad & z = \delta_i(\tilde{S}_i^\alpha(\theta_i), \theta_i)\theta_i - \omega_i(\tilde{S}_i^\alpha(\theta_i), \theta_i), \end{aligned} \tag{2.17}$$

where $z \in \mathbb{R}_{\geqslant 0}$. If we define $\hat{S}_i^\alpha(z) = \tilde{S}_i^\alpha(\mathcal{F}_i^\alpha(z))$, then we have the following theorem.

Theorem 2. *The collection* $\mathcal{A}^\alpha = \{\bigcup_{i \in M} \hat{S}_i^\alpha(z) : z \in \mathbb{R}_{\geqslant 0}\}$ *includes* S^α*, the size of which is bounded by* $O(n_{\max} N)$*.*

The proof of Theorem 2 follows directly from the proof of Theorem 1, we have $|\mathscr{A}^\alpha| \leqslant \sum_{i\in M} |\mathscr{A}_i^\alpha| = \sum_{i\in M} n_i^2 \leqslant n_{\max} \sum_{i\in M} n_i = n_{\max} N$. Our approach is based on discretization of z to find the fixed point, Theorem 2 guarantees that the number of iterations is bounded.

Following the similar ideas in Section 2.4, we use the proposition below to end this section.

Proposition 6. *Define* $\tilde{S}^\alpha(z) = \bigcup_{i\in M} \hat{S}_i^\alpha(z)$, *then* Z^α *is the fixed point of function* $\Pi(\tilde{S}^\alpha(z), z)$ *that is defined as*

$$\Pi(\tilde{S}^\alpha(z), z) = \frac{\sum_{i\in M} V_i(\hat{S}_i^\alpha(z), \mathcal{F}_i^\alpha(z)) R_i(\hat{S}_i^\alpha(z), \mathcal{F}_i^\alpha(z))}{v_0 + \sum_{i\in M} V_i(\hat{S}_i^\alpha(z), \mathcal{F}_i^\alpha(z))}.$$

Furthermore, $\Pi(\tilde{S}^\alpha(z), z)$ *is a unimodal function of* z. *If we define* $f^\alpha(z) = \sum_{i\in M} V_i(\hat{S}_i^\alpha(z), \mathcal{F}_i^\alpha(z))[R_i(\hat{S}_i^\alpha(z), \mathcal{F}_i^\alpha(z)) - z]$, *then* $f^\alpha(z)$ *is a piecewise convex function of* z *and* Z^α *is the unique fixed point of* $f^\alpha(z)/v_0$.

2.6 Numerical Illustration

In this section, we illustrate our solution approach to problem (4.1) under cardinality constraints on an example of the nested logit model with 2 nests and 6 products. The nested structure is presented in Figure 4.1. Compared to the cardinality constraints, we can only get a 2-approximate solution to problem (4.9), other than this, the solution approach under space constraints is same as it is under cardinality constraints. Therefore, we can also use this illustration to demonstrate space constraints cases with minor adjustments.

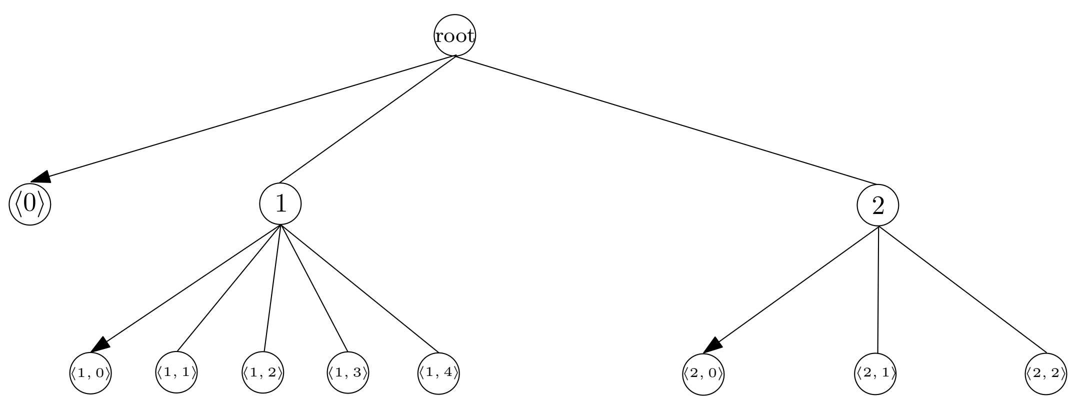

Figure 2.1: Nested structure

In the nested logit model, nest 1 has 4 products that are indexed by $\langle 1, j\rangle$ for $j \in \{1, 2, 3, 4\}$ and nest 2 has two products $\langle 2, 1\rangle$ and $\langle 2, 2\rangle$. There are three no-purchase options $\langle 0\rangle$, $\langle 1, 0\rangle$ and $\langle 2, 0\rangle$. Table 4.2 shows the input parameters of problem (4.1). Specifically, the preference weight of $\langle 2, 0\rangle$ is set to be zero.

product	$\langle 1,1\rangle$	$\langle 1,2\rangle$	$\langle 1,3\rangle$	$\langle 1,4\rangle$	$\langle 2,1\rangle$	$\langle 2,2\rangle$	nest	γ_i	v_{i0}	$\mathbb{C}_i$
α_j	30	25	17.6	9.5	25	10	1	0.92	5	2
β_j	6	3	1.2	0.5	3	2	2	0.85	0	2
c_j	1	3.4	8	10	4	2	V_0	5.5		

Table 2.1: Parameters setup for the joint optimization problem under cardinality constraints

One can check that it satisfies our assumption on price sensitivity parameter β_j. By Lemma 4, we have $\hat{S}_2(z) = \{\langle 2,1\rangle, \langle 2,2\rangle\}$ for $z \in \mathbb{R}$, thus we focus on nest 1. Figure 4.2 visualizes the optimization procedure of obtaining the optimal solution to problem (2.11) at nest 1.

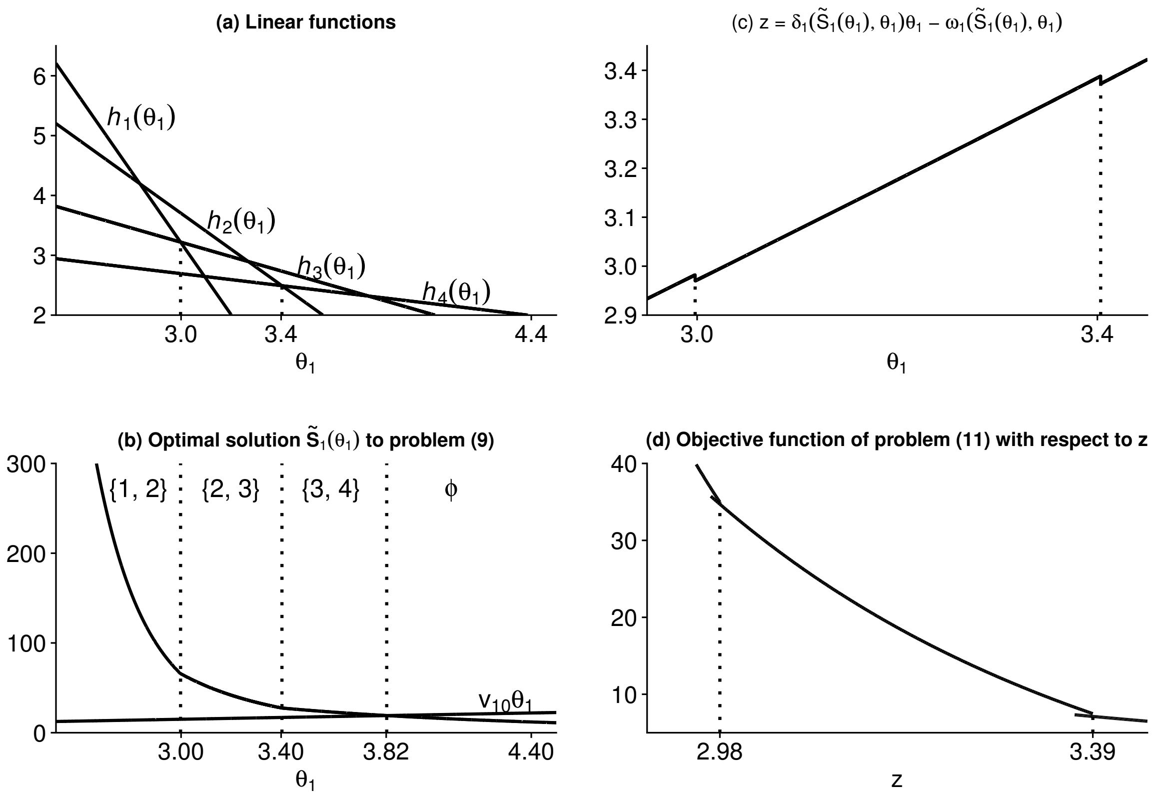

Figure 2.2: Solution approach visualization to problem (2.11) at nest 1

In Figure 4.2(a), there are 4 linear functions that are defined as $h_j(\theta_1) = \tilde{\alpha}_{1j} - \beta_{1j}\theta_i$ and $\tilde{\alpha}_{1j} = \alpha_{1j} - \beta_{1j}c_{1j} - \log(\beta_{1j}) - 1$ for $j \in \{1,2,3,4\}$. Since the cardinality $\mathbb{C}_1 = 2$, then we have $\tilde{S}'_1(\theta_1) = \{\langle 1,1\rangle, \langle 1,2\rangle\}$ when $\theta_1 \in [0,3]$; $\tilde{S}'_1(\theta_1) = \{\langle 1,2\rangle, \langle 1,3\rangle\}$ when $\theta_1 \in (3, 3.4]$; $\tilde{S}'_1(\theta_1) = \{\langle 1,3\rangle, \langle 1,4\rangle\}$ when $\theta_1 > 3.4$. In order to get the optimal solution $\tilde{S}_1(\theta_1)$ to problem

(4.9), we need to compare $\sum_{j \in \tilde{S}_1(\theta_1)} v_{1j}(\theta_1 + c_{1j} + 1/\beta_{1j})/\beta_{1j}$ with $v_{10}\theta_1$, which is shown in Figure 4.2(b): $\tilde{S}_1(\theta_1) = \{\langle 1,1\rangle, \langle 1,2\rangle\}$ when $\theta_1 \in [0,3]$; $\tilde{S}_1(\theta_1) = \{\langle 1,2\rangle, \langle 1,3\rangle\}$ when $\theta_1 \in (3,3.4]$; $\tilde{S}_1(\theta_1) = \{\langle 1,3\rangle, \langle 1,4\rangle\}$ when $\theta_1 \in (3.4, 3.82]$ and $\tilde{S}_1(\theta_1) = \varnothing$ when $\theta_1 > 3.82$. Figure 4.2(c) visualizes the constraint in problem (2.11), from which we can see that for certain ranges of z, there may not exist a one-to-one correspondence between z and θ_1. It is consistent with the discussion before problem (2.11). In Figure 4.2(d), the objective function of problem (2.11) in terms of z is consist of three convex curves, by selecting the highest curve, the optimal solution $\hat{S}_1(z) = \{\langle 1,1\rangle, \langle 1,2\rangle\}$ when $z \in [0, 2.98]$; $\hat{S}_1(z) = \{\langle 1,2\rangle, \langle 1,3\rangle\}$ when $z \in (2.98, 3.39]$; $\hat{S}_1(z) = \{\langle 1,3\rangle, \langle 1,4\rangle\}$ when $z \in (3.39, 3.80]$ and $\hat{S}_1(z) = \varnothing$ when $z > 3.80$. Therefore, we obtain $\tilde{S}(z) = \hat{S}_1(z) \bigcup \hat{S}_2(z) = \{\langle 1,1\rangle, \langle 1,2\rangle, \langle 2,1\rangle, \langle 2,2\rangle\}$ for $z \in [0, 2.98]$; $\tilde{S}(z) = \hat{S}_1(z) \bigcup \hat{S}_2(z) = \{\langle 1,2\rangle, \langle 1,3\rangle, \langle 2,1\rangle, \langle 2,2\rangle\}$ for $z \in (2.98, 3.39]$; $\tilde{S}(z) = \hat{S}_1(z) \bigcup \hat{S}_2(z) = \{\langle 1,3\rangle, \langle 1,4\rangle, \langle 2,1\rangle, \langle 2,2\rangle\}$ for $z \in (3.39, 3.80]$ and $\tilde{S}(z) = \hat{S}_1(z) \bigcup \hat{S}_2(z) = \{\langle 2,1\rangle, \langle 2,2\rangle\}$ for $z > 3.80$. Note that $|\mathscr{A}| \leqslant |\mathscr{A}_1| + |\mathscr{A}_2| \leqslant 6$ where $\mathscr{A} = \{\tilde{S}(z) : z \in \mathbb{R}\}$, $\mathscr{A}_1 = \{\tilde{S}_1(z) : z \in \mathbb{R}\}$ and $\mathscr{A}_2 = \{\tilde{S}_2(z) : z \in \mathbb{R}\}$, which supports Theorem 1.

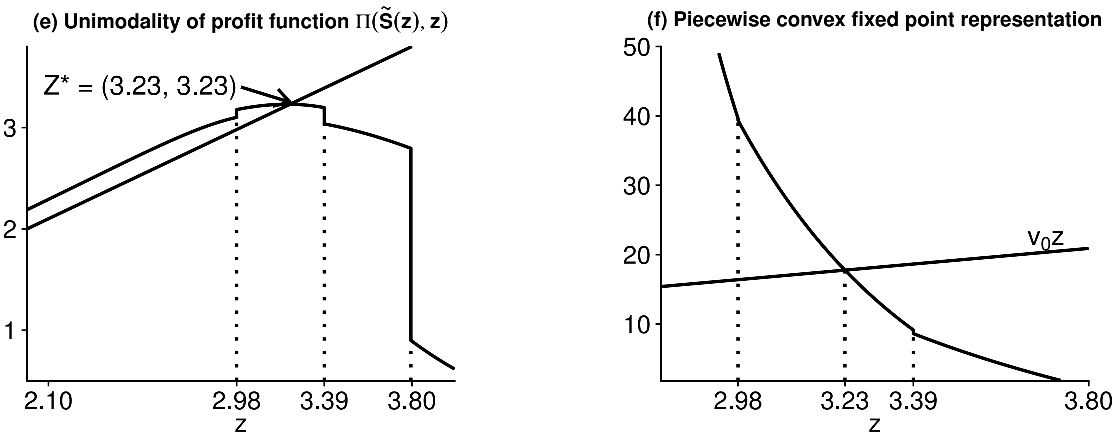

Figure 2.3: Unimodality and piecewise convexity

Figure 4.3 shows the unimodality of profit function $\Pi(\tilde{S}(z), z)$, which addresses Lemma 6; and the piecewise convexity that addresses Proposition 4. In Figure 4.3(e), the optimal Z^* satisfies the fixed point representation $\Pi(\tilde{S}(z), z) = z$, which is the intersection point of the solid 45°-line and the unimodal profit function $\Pi(\tilde{S}(z), z)$. We obtain $Z^* = 3.23$, implying that the optimal expected profit that we can get from this nested logit model with parameters in Table 4.2 is 3.23 and the optimal assortment is $\tilde{S}(3.23) = \{\langle 1,2\rangle, \langle 1,3\rangle, \langle 2,1\rangle, \langle 2,2\rangle\}$. We can see from Figure 4.3(e) that profit function $\Pi(\tilde{S}(z), z)$ jumps discontinuously at $2.98 < Z^*$ and drops discontinuously at $3.39 > Z^*$, which is an example of Lemma 6. Figure 4.3 shows the piecewise convex fixed point representation, line v_0z intersect with $f(z)$ at $Z^* = 3.23$.

By looking up previously stored table, the optimal prices are $p_{12}^* = 6.99, p_{13}^* = 12.09, p_{21}^* = 7.62, p_{22}^* = 5.79$.

Chapter 3

Constrained Assortment and Price Optimization

3.1 Literature Review

[26] provide us an extensive review of the assortment and price optimization problem under different models and scenarios. In this subsection, we review the literature on the different types of customer choice models.

If the customer choosing behavior is modeled under the multinomial logit model, then [42] formulate the unconstrained assortment optimization problem with a newsboy model considering its inventory cost. [44] point out that the optimal structure follows a simple form. [28] generalize their work to the network revenue management problem by proposing a linear program. [39] find the optimal solution in polynomial time under cardinality constraints with the existence of a no-purchase option. Similarly, [47] considers a generalized attraction model with the capacity constraints. [40] study the robust assortment optimization problem by assuming some of the true parameters to be unknown. As an extension of the multinomial logit model, the mixed multinomial logit model can be used to model more realistic choice scenarios. [7] develop a column generation algorithm to efficiently find an acceptable solution under the mixed multinomial logit model. [14] also present an approximation algorithm for the assortment optimization problem with capacity constraints under the mixed multinomial logit model.

[13] shows that the multinomial logit model suffers from independence of irrelevant alternatives (IIA), red-bus, blue-bus paradox is one of the most famous examples to show that the multinomial logit model is unrealistic in some cases ([34] and [5]). To resolve this limitation, [4] first introduced the nested logit model. [35] shows that the nested logit model is a member of generalized extreme value (GEV) models. For the unconstrained problems under the nested logit model, or the two-level nested logit model, [27] analyze both centralized and decentralized regimes. [11] propose a linear program to obtain the optimal assortment in polynomial time. [29] develop a greedy algorithm to find an approximate solution. The con-

strained assortment optimization problem becomes much harder to solve. [36] use an integer programming to analyze the assortment optimization problem with cardinality constraints under the latent class choice model. By assuming there is a fixed number of products, [38] develops an approximation algorithm for the assortment optimization problem with cardinality constraints under the two-level nested logit model. [18] study the assortment optimization problem with both cardinality and space constraints under the two-level nested logit model. The authors impose constraints on offered assortment in each of the nests separately, in this case, the assortment optimization problem with cardinality constraints can be solved by a linear program. However, this problem becomes NP-hard with space constraints. The authors propose an approximation algorithm with the performance guarantee of 2. [16] study the assortment optimization problem with cardinality and space constraints across nests under the two-level nested logit model. [52] consider the joint optimization of the constrained assortment and price optimization problem under the two-level nested logit model with a no-purchase option in every choice stage.

The two-level nested logit model only allows us to analyze one-dimensional dissimilarity between products. The multilevel nested logit model, including the two-level nested logit model as its special case, can describe the customer choosing behavior with multiple stages, which is closer to the real choice process. For applications of the multilevel nested logit model with more than two levels, a four-level nested logit model is applied to predict the recreational fishing demand [8]. [25] show applications of the multilevel nested logit model with an arbitrary number of levels in the recommendation system. The assortment optimization problem has also been studied under the multilevel nested logit model. [51] study the joint optimization of the assortment and pricing problem, but the authors only consider one no-purchase option in their model, where the customer can exit the system without buying only at the beginning of her choosing process. [30] consider the assortment optimization problem under the multilevel nested logit model with fixed number of products and develop a polynomial time algorithm to identify the optimal assortment. However, they do not consider cardinality or space constraints, and just study the case where there is only one no-purchase option in the first level of the tree structure; this means the authors assume that if a customer wants to leave without purchasing anything, she must make the exiting decision right after entering the system, otherwise she should buy a product in the end. This assumption is unrealistic for modeling the customer choosing behavior in real world scenarios. Our approach relaxes this assumption by allowing the no-purchase option in every stage of the customer choice process. This essay studies the constrained assortment optimization problem with both cardinality and space constraints under the multilevel nested logit model, where there is a no-purchase option in every nonleaf node. To the best of our knowledge, we are the first to study the constrained assortment optimization problem under the multilevel nested logit model. The algorithms that are used under the multinomial or two-level nested logit model, such as the linear program in [18], cannot be generalized to the multilevel nested logit model case. In Sections 16 and 13, we will develop an efficient algorithm for the multilevel nested logit model, which is of comparable complexity to the algorithms that solve the unconstrained assortment problem under the multilevel nested logit model.

For the price optimization problem under variants of the multinomial logit model, [20] firstly show that the multinomial logit profit function is not jointly concave in prices even when the price-sensitivity parameters are fixed to be identical. However, the profit function under the multinomial logit model is concave in market share variables that have a one-to-one mapping with the price variables ([43] and [15]). Under the multinomial logit model, [2] show that there is a unique optimal price vector satisfying first order condition, the method of which is also used by [21] and [1] to analyze the price optimization problem with respect to the markup variables that is defined as price minus cost. [48] also uses the multinomial logit model to study the pricing problem, which is generalized by [31] to the two-level nested logit model, where price-sensitivity parameters are assumed to be identical within each nest but different across nests. When the price-sensitivity parameters are different across all the products, the profit function is no longer concave in the market share vector even under the two-level nested logit model. [19] point out that the adjusted markup is constant within each nest by checking the first order condition. Furthermore, the multiproduct profit function can be reduced to a unimodal function via introducing the adjusted nest-level markup that has a one-to-one correspondence with the price vector. However, all the above literature only considers the price optimization problem under the one-level nested logit model (multinomial logit model), or the two-level nested logit model. [30] study the pricing problem under the multilevel nested logit model, but their iteration method can only find a local maximum because the authors still consider the pricing problem with respect to price vector, even though it has already been proved that the profit function is nonconcave under the two-level nested logit model. [22] study the centralized pricing problem under a tree structure, but the no-purchase option can only exist in the first level of their model and they do not consider the unconstrained or constrained assortment optimization either. To the best of our knowledge, we are the first to study the price optimization problem under the multilevel nested logit model with a no-purchase option existing in every nonleaf node of the tree structure.

The remainder of this essay is organized as follows. In Section 3.3, we address the constrained assortment optimization problem. Then Section 3.4 presents the price optimization problem with given assortment.

3.2 Main Results and Contributions

We summarize our main results and contributions as follows:

1. In this essay, we consider the constrained assortment and price optimization problems under the m-level nested logit model with n products and we allow a no-purchase option to appear in every period of customer choosing process. For the constrained assortment optimization problem, we discuss two subproblems and propose a way to stitch the optimal sub-assortments together to get the global optimal (an α-approximate) assortment under the cardinality (space) constraints in polynomial time. For the price optimization problem, we formulate it as a maximization of a unimodal profit function, thus it is tractable to find

the optimal pricing strategy. To the best of our knowledge, there is no work on constrained assortment optimization problem under the multilevel nested logit model in current literature. Moreover, there is no literature allowing the existence of a no-purchase option in every node of the multilevel tree structure for both assortment and price optimization problems.

2. For the constrained assortment optimization problem, we use an m-level tree with n products to describe the customer choice structure. The cardinality or space constraints are imposed on the nonleaf nodes in the second lowest level separately. Our main result is that the optimal assortment under the cardinality constraints can be obtained in $O(n \max\{m, k\})$ operations and a 2-approximate assortment under the space constraints can be obtained in $O(mnk)$ operations, where k is the maximum number of products within any node in level $m-1$. [30] study the unconstrained assortment optimization problem under an m-level nested logit model and their algorithm runs in $O(mn \log n)$ time, but it is not possible to implement their approach to deal with the constrained cases. It is interesting to find that our algorithm for the constrained problem is even more efficient than the unconstrained algorithm in [30] when k is relatively small. The reason why our constrained assortment algorithm is faster is that the core step of the unconstrained assortment algorithm of [30] is to compute the pairwise intersection points of lines, in which sorting algorithms are required. However, we manage to avoid constantly using sorting algorithm by revealing the hidden ordered properties of candidate sub-assortments.

3. In the multilevel nested logit model, every nonleaf node has a no-purchase option, which allows customer to exit at any period of their choosing process. For both constrained assortment and pricing optimization problems, it is a non-trivial extension of the case where there is only one no-purchase option that is associated with root node in the multilevel nested logit model. Particularly for the price optimization problem, the formulation of node-specific adjusted markup is more generalized and cannot be obtained by the approach in [22] that consider the pricing problem under multilevel choice structure with only one no-purchase option. We are able to show that the objective function can be reduced to a unimodal function by dimensional reduction of creating mapping between the node-specific adjusted markups.

4. Many existing literature regarding the assortment and price optimization problems under the multinomial logit model or the nested logit model turn out to be the special case of ours. For the constrained assortment optimization problem, we generalize the works of [39] (multinomial logit model) and [18] (two-level nested logit model) to the multilevel nested logit model with cardinality and space constraints. For the unconstrained assortment optimization problem in [30], it becomes a special case of the constrained problem when the constraints are set to be large enough, i.e. larger than k. [52] consider the joint optimization of assortment and price under the two-level nested logit model with no-purchase options. However, their approach cannot be directly applied to the multilevel nested logit model since the structure of the optimization model changes fundamentally when it comes to the nested logit model with the number of levels that is larger than three. Compared to [30], our approach is more general in three folds: first, consider both cardinality and space constraints; second, allow a no-purchase option in every stage of the customer choosing process; third,

the computational complexity $O(n \max\{m, k\})$ under the cardinality constraints and the complexity of $O(mnk)$ under the space constraints are comparable to $O(mn \log n)$, and even more efficient for a small k. Besides the constrained assortment optimization problem, we consider a price optimization problem under the multilevel nested logit model with product-differentiated price sensitivities. We generalize [48] (multinomial logit model) and [19] (two-level nested logit model) to the multilevel nested logit model. Furthermore, for the pricing problem under the multistage choice model, we are also able to generalize [22] in terms of letting a no-purchase option exist in every stage of the customer choosing process, which we believe is the first in the literature. [51] study the joint optimization of assortment and price problem under the multilevel nested logit model with only on no-purchase option that is connected to the root node. However, their approach fails to work when the no-purchase options are allowed to exist in every choosing stage.

In the following subsection, we address the literature based on the assortment and price optimization perspectives.

3.3 Constrained Assortment Optimization

In this section, we present the constrained assortment optimization problem and our solution approach. We first show the problem formulation, then discuss how to construct candidate assortments containing the optimal or α-approximate solutions. Before showing the algorithms for assortment optimization with cardinality or space constraints, we address the properties of the optimal or an α-approximate assortment for an arbitrary intermediate node.

Problem Formulation

We formulate the constrained assortment optimization problem under the multilevel nested logit model in this subsection. We use the multilevel nested logit model with m levels that is indexed by $M = \{1, 2, ..., m\}$ and n products to model a multistage decision-making process of the customer. In this tree structure, each node represents a subset of the entire choice space and each level stands for a choice criterion, or a specific attribute of products, such as price, quality, category, etc. Specifically, the root node includes all the candidate products. The node in level l ($1 \leqslant l \leqslant m-1$) represents the subset of products that satisfy all the first l choice rules. The customer choosing behavior can be described under this tree structure: start from the root node, then the customer has two options in general: either to choose the no-purchase option in the first level to leave without purchasing; or to choose one child node, which corresponds to a subset of products satisfying the first choice criterion, of the root node to narrow down her choice space. If she does not choose the no-purchase option in the first level, then she still has two possible choice alternatives: either to leave or to further narrow down her choice space. This choice procedure is being conducted repeatedly until she chooses a no-purchase option or an actual product in level m that is the lowest level of the tree structure.

Let V and E denote the set of nodes and edges in this tree structure, respectively. We use an l-dimensional ($1 \leqslant l \leqslant m$) vector $(i_1, i_2, ..., i_l)$ to denote the node i in level l. Moreover, the no-purchase option i_0 in level $l+1$ ($0 \leqslant l \leqslant m-1$) is denoted as $(i_1, i_2, ..., i_{l-1}, i_l, 0)$. A subset of products that satisfy the first l attributes, denoted as N_i, is affiliated with this node i. For example, buying clothes in a retail store can be formulated under a three-level nested logit model, the attribute for the first level is category, price is for the second level, actual clothes are in the third level. A node in the second level can correspond to a subset of clothes that have the following attributes: T-shirt (category) and \$100 - \$200 (price). The root node is in level 0, then the no-purchase option root_0 in level 1 is a one-dimensional vector (0). Specifically, the set of products of root, which is the entire choice space, is denoted as N_{root}. The total number of products is $n = |N_{\text{root}}|$. The set of products for a leaf node is the product itself. In this notation, we can see that an actual product (or a no-purchase option), which is a leaf node in the lowest level, is labeled as an m-dimensional vector $(i_1, i_2, ..., i_m)$, and specifically $i_m = 0$ for the no-purchase option. For the nonleaf node i, we use an $(l-1)$-dimensional vector $(i_1, i_2, ..., i_{l-1})$ and an $(l+1)$-dimensional vector $(i_1, i_2, ..., i_l, i_{l+1})$ to denote its parent and child node, respectively. We use i^P to denote i's parent node and i_C to denote the set of children nodes of node i. Then we have $N_i = \bigcup_{j \in i_C} N_j$. Particularly, $i_C = \varnothing$ if i is a leaf node or a no-purchase option; if i is root, then we define $i^P = \varnothing$. We can imagine that an edge in set E connecting one nonleaf node and its child node as a one-step choosing process: moving to its child node through this edge can be interpreted as starting to consider an additional attribute of products or choosing to leave without buying. Thus for a customer at a nonleaf node of this tree structure, she can either choose to leave the system without further considering any more attributes of the product (the no-purchase option), or to move to one of its children nodes.

A subset S_i of N_i is used to represent the assortment of node i. Particularly, if i is a leaf node, then S_i is the product i itself or $\varnothing$; if node i is a no-purchase option at an arbitrary level, then S_i is an empty set. For a nonleaf node i, we define $S_i = \bigcup_{j \in i_C} S_j$, then S_{root} is the assortment of the whole system.

Throughout the essay, the two types of constraints, cardinality and space constraints, are defined on the nodes in level $m-1$ separately. For the ease of presentation, a node in level $m-1$ is referred to as a *basic node*, the set of which is denoted as $\mathscr{B}$. For any node $i \in V$, we use $\Im_i$ to represent the collection of feasible assortments satisfying certain constraints. For all $i \in \mathscr{B}$, the cardinality constraints can be expressed as $\Im_i = \{S_i : S_i = \bigcup_{j \in i_C} S_j, |S_i| \leqslant \mathbb{C}_i\}$, where $\mathbb{C}_i$ is the maximum number of products for basic node i and $|S_i|$ represents the number of products in S_i; for the space constraints, $\Im_i = \{S_i : S_i = \bigcup_{j \in i_C} S_j, \sum_{j \in i_C} w_j \leqslant \mathbb{S}_i\}$, where $\mathbb{S}_i$ is the maximum available space for basic node i and w_j is the space consumption of product j. To make sure that all the products are eligible to be offered, we assume that $w_j \leqslant \mathbb{S}_i$ for any leaf node $j \in i_C$. If node i is neither a basic node nor a leaf node, then the feasible set $\Im_i$ is the cartesian product of its children nodes' feasible sets $\Im_i = \bigtimes_{j \in i_C} \Im_j$.

The upside-down tree in Figure 4.1 is used as an example to better explain the notation system and feasible sets for both constraints. We address the following nodes as representatives for illustrating our vector representation: $R_0 = (0), A = (1), A_0 = (1, 0), D = (1, 2), D_0 = (1, 2, 0)$ and $I = (1, 2, 1)$. For the node A, we have $A^P = R$, $A_C = \{C, D\}$. For cardinality constraints, if $\Im_C = \{S_C : \bigcup_{j \in i_C} S_j, |S_C| \leqslant 1\} = \{\varnothing, \{G\}, \{H\}\}$ and $\Im_D = \{S_D : \bigcup_{j \in i_D} S_j, |S_D| \leqslant 1\} = \{\varnothing, \{I\}, \{J\}\}$, then the feasible set of node A with cardinality constraints is $\Im_A = \Im_C \times \Im_D = \{\varnothing, \{G\}, \{I\}, \{H\}, \{J\}, \{G, I\}, \{G, J\}, \{H, I\}, \{H, J\}\}$. One feasible assortment of the node A is $S_A = \{H, I\} \subseteq \Im_A$. Similarly, for space constraints, if $w_G = w_I = 1$, $w_h = w_J = 2$, and $C_C = C_D = 2$, then the feasible set of node A with space constraints is $\Im'_A = \{\varnothing, \{G\}, \{I\}, \{H\}, \{J\}, \{G, I\}\}$. One feasible assortment of node A with these space constraints is $S'_A = \{H\} \subseteq \Im'_A$.

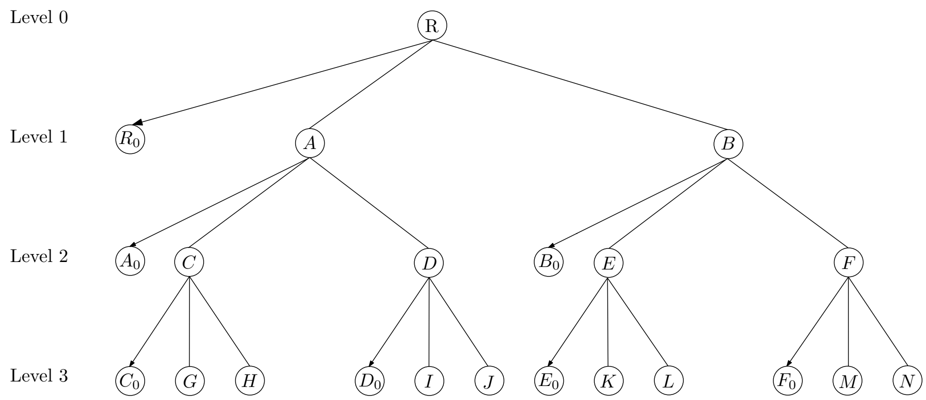

Figure 3.1: Example for the multilevel nested logit model

The preference weight of a leaf node describes the attractiveness of the product that is associated with this leaf node. The preference weight for a leaf node j is denoted as $V_j(\{j\}) = v_j$. Generally speaking, the preference weight V_{i0} for the no-purchase option i_0 is greater than or equal to zero. Then for each nonleaf node i, its preference weight is calculated recursively as

$$V_i(S_i) = \left(V_{i0} \mathbf{1}(S_i \neq \varnothing) + \sum_{j \in i_C} V_j(S_j) \right)^{\gamma_i},$$

where S_i is the assortment of node i. $\gamma_i \in (0, 1]$, the dissimilarity parameter for node i, is assumed to be a constant. For the root node, we set $\gamma_{\text{root}} = 0$ without loss of generality, thus $V_{\text{root}}(S_{\text{root}}) = 1$. The constraints on γ_i ensure that the multilevel nested logit model is consistent with utility maximization theory [34]. Then for node i, the correlation between the products' utilities of its children nodes is a decreasing function of γ_i, the closer γ_i is to one, the less positively correlated the utilities are. Moreover, if $\gamma_i = 1$, then the nest structure of node i degenerates: its children nodes directly connect to its parent node. If the dissimilarity parameter γ_i exceeds one, this model can still be a random utility model under some circumstances [33]. In this case, adding a product to assortment S_i can increase the probability of choosing other products in S_i. [11] relax the constraints on the dissimilarity parameter under the two-level nested logit model, then the synergistic effect within the assortments of children nodes can be modeled.

[30] assume that if a customer moves from the root node to a node that is not a no-purchase option, then she must make a purchase before leaving. Inspired by [18], our model formulation relaxes this strong assumption. The indicator function $\mathbf{1}(\cdot)$ in $V_i(S_i)$ allows us to model the scenario where even after reaching a specific node, the customer can still choose to leave the system without going deeper into the tree structure and making a purchase in the end. Our formulation is closer to reality because the customer will not notice node i when S_i is an empty set; otherwise when S_i is not empty, the customer still has the probability of $V_{i0}/V(S_i)^{1/\gamma_i}$ to leave without any purchasing. However, [18] consider this assumption relaxation under the two-level nested logit model for the constrained assortment optimization problem, which is generalized by our approach to the multilevel nested logit model.

In the multilevel nested logit model framework, if we assume that S_i is not empty and node j is one of its children nodes, then the conditional probability of choosing assortment S_j given S_i is computed as

$$Q(S_j|S_i) = \frac{V_j(S_j)}{V_{i0}\mathbf{1}(S_i \neq \varnothing) + \sum_{j \in i_C} V_j(S_j)}.$$

When S_i is empty, it means that we do not offer any products of node i so the customer simply will not consider purchasing anything in S_i. Hence we define $Q(S_j|S_i) = 0/0 = 0$ for $S_i = \varnothing$, indicating the customer makes purchases in an empty assortment with probability zero.

From here, we will present the formulation of the constrained assortment optimization problem. Using similar notations as in [30], $R_i(S_i)$ denotes the profit of the assortment S_i for any node $i \in V$. If i is a leaf node, then $R_i(S_i) = \mathbf{1}(S_i \neq \varnothing)r_i$, where r_i is the profit of the actual product i, and $R_i(\varnothing) = 0$ if i is a no-purchase option. It shows that if a customer chooses a non-empty leaf node, or an actual product, then a profit will be obtained with certain probability. Specifically, $R_i(S_i) = 0$ if node i is a no-purchase option or S_i is empty. If i is a nonleaf node, the expected profit is defined recursively as

$$\begin{aligned} R_i(S_i) &= \sum_{j \in i_C} Q(S_j|S_i) * R_j(S_j) \\ &= \frac{\sum_{j \in i_C} V_j(S_j) R_j(S_j)}{V_{i0}\mathbf{1}(S_i \neq \varnothing) + \sum_{j \in i_C} V_j(S_j)}. \end{aligned}$$

If $S_i = \varnothing$, then $Q(S_j|S_i) = 0$, so $R_i(S_i) = 0$. According to the above definition, the total expected profit from a customer is $R_{\text{root}}(S_{\text{root}})$. We use S^*_{root} and Z^* to denote the optimal solution and the corresponding maximum profit, respectively. Let S^α_{root} and Z^α denote an α-approximate solution and its profit, where $\alpha Z^\alpha = \alpha R_{\text{root}}(S^\alpha_{\text{root}}) \geqslant Z^* = R_{\text{root}}(S^*_{\text{root}})$.

Moreover, we formulate the α-approximate ($\alpha \geqslant 1$) assortment optimization problem as

$$Z^{\alpha} = \max_{S_{\text{root}} \subseteq \Im^{\alpha}_{\text{root}}} R_{\text{root}}(S_{\text{root}}), \tag{3.1}$$

where $\Im^{\alpha}_{\text{root}}$ is a subset of $\Im_{\text{root}}$ and contains an α-approximate solution as its best assortment. How to construct $\Im^{\alpha}_{\text{root}}$ will be shown in Section 3.3. Throughout the essay, we use S_i^{α} to denote the optimal solution to problem (3.1) at node $i \in V$. Specifically, we have $S_i^* = S_i^1$ and $\Im^*_{\text{root}} = \Im^1_{\text{root}}$. Problem (3.1) is highly nonlinear, the entire choice space is so large that it is impossible to find the optimal solution without an efficient algorithm.

Basic α-approximate Assortment Subproblem

In this subsection, we decompose problem (3.1) into α-approximate assortment subproblems that can be solved efficiently when the searching space has small size and propose an alternative formulation of the subproblem, which is referred to as *basic* α-approximate assortment subproblem.

Let h be the parent node of $i \in V$, we define

$$t_i^{\alpha} = \begin{cases} +\infty & \text{, if } R_i(S_i^{\alpha}) < t_h^{\alpha} \\ \gamma_i t_h^{\alpha} + (1-\gamma_i) R_i(S_i^{\alpha}) & \text{, if } R_i(S_i^{\alpha}) \geqslant t_h^{\alpha} \end{cases} \tag{3.2}$$

We set $t^{\alpha}_{\text{root}^P} = 0$ by convention, then $t^{\alpha}_{\text{root}} = R_{\text{root}}(S^{\alpha}_{\text{root}})$. The scalar t_i^{α} can be computed from top to bottom when all the S_i^{α} are known. Define the α-approximate assortment subproblem at an arbitrary nonleaf node $i \in h_C$ with parameter $\alpha \geqslant 1$ as

$$\max_{S_i \subseteq \Im_i^{\alpha}} \{V_i(S_i)\,(R_i(S_i) - t_h^{\alpha})\}. \tag{3.3}$$

Problem (3.3) at the root node is equivalent to (3.1) since $t^{\alpha}_{\text{root}^P} = 0$ and $V_{\text{root}}(S_{\text{root}}) = 1$. If $\Im_i^{\alpha}$, the collection of feasible α-approximate assortments, is a relatively small set, then the subproblem (3.3) is easy to solve by checking all its subsets. In the following sections, we aim to reduce the size of set $\Im_i^{\alpha}$.

We can see that S_i^{α} is not empty if $R_i(S_i^{\alpha}) \geqslant t_h^{\alpha}$. In other words, if we can find an assortment $\hat{S}_i \subseteq \Im_i^{\alpha}$ such that $R_i(\hat{S}_i) \geqslant t_h^{\alpha}$, then we know that it is worthwhile to offer a non-empty assortment at node i. Otherwise, if $R_i(S_i^{\alpha}) < t_h^{\alpha}$, then S_i^{α} is an empty assortment. Moreover, the scalar defined in (3.2) is $+\infty$ for all the descendants of node i. Then the α-approximate assortments for all the descendants of node i are empty, which is consistent with the fact that S_i is empty since S_i is defined recursively as $S_i = \bigcup_{j \in i_C} S_j$. Problem (3.3) can be solved if we know the value of all the scalars. However, knowing all these scalars beforehand is not possible since it requires the optimal solution to problem (3.3), but in an alternative way, we can get the candidate collection of assortments containing an α-approximate solution by letting those scalars vary from $-\infty$ to $+\infty$ because the true value lies in $\mathbb{R}$. Whereas it tremendously enlarges the searching space unless we can find the connection between solutions to the subproblems of the parent node and its children nodes. In light of [30], we claim that for an arbitrary nonleaf node i, S_i^α is optimal to problem (3.3) and the union of all the optimal assortments to problem (3.3) at its children nodes is also the optimal assortment to problem (3.3) at node i. For completeness, we provide the proof of this claim in B.2 since we still need to check if this claim holds when there is a no-purchase option in every choosing period.

The optimal solution to problem (3.3) at a nonleaf node i can be easily obtained if we can solve the subproblems at all its children nodes $j \in i_C$; otherwise the candidate collection $\Im_i^\alpha$ is the cartesian product of $\Im_j^\alpha$ for all $j \in i_C$, which makes problem (3.3) intractable. Thus we need to reduce $\Im_i^\alpha$ to a collection with smaller size, denoted as $\mathscr{A}_i^\alpha \subseteq \Im_i^\alpha$, and $\mathscr{A}_i^\alpha$ also includes an α-approximate assortment. Furthermore, it is still uncertain that how to get candidate collection for the basic nodes in the first place. We make an observation that problem (3.3) at node i is highly nonlinear, thus we propose an alternative formulation of problem (3.3), which is referred to as basic α-approximate assortment subproblem, as follows

$$\max_{S_i \subseteq \Im_i^\alpha} \{V_i(S_i)^{1/\gamma_i} \left(R_i(S_i) - t_i^\alpha\right)\}. \tag{3.4}$$

Lemma 8. *The optimal solution to problem (3.4) is also optimal to problem (3.3).*

Problem (3.4) is more tractable than problem (3.3), thus we turn our focus to problem (3.4) in following subsections.

Candidate Assortment Construction

In this subsection, we first come up with a way to construct candidate collection of α-approximate assortments for the basic nodes by utilizing the insights from Lemma 8. Then we present how to construct $\mathscr{A}_i^\alpha$ that has a reasonable size.

Lemma 9. *If we use $\tilde{S}_i^\alpha$ to denote the assortment satisfying $V_i(\tilde{S}_i^\alpha)^{1/\gamma_i}\left(\alpha R_i(\tilde{S}_i^\alpha) - t_i^*\right) \geqslant V_i(S_i^*)^{1/\gamma_i}\left(R_i(S_i^*) - t_i^*\right)$ at node i for all $i \in \mathscr{B}$ with parameter $\alpha \geqslant 1$, then the assortment $S_{\text{root}}^\alpha = \bigcup_{i\in\mathscr{B}} \tilde{S}_i^\alpha$ is an α-approximate solution.*

By applying Lemma 9, we can get an α-approximate solution if all the scalars are known for all basic nodes. Similarly, knowing all the scalars is impossible without already having the optimal solution to (3.1). As we discussed in subsection 3.3, problem (3.1) cannot be tractable unless the collection $\mathscr{A}_i^\alpha$ for a basic node i can be constructed to have a small size. How to construct the polynomial-size collection with cardinality constraints and space constraints for basic nodes will be shown in subsection 16 and subsection 13, respectively. We use the following proposition to summarize the above findings and answer the question that is asked at the end of Section 3.3 about how to build $\Im_{\text{root}}^\alpha$.

Proposition 7. *Assume the collection of assortments* $\{\mathscr{A}_i^\alpha : i \in \mathscr{B}\}$ *contain an* α*-approximate solution* S_i^α*, then* $\Im_{\text{root}}^\alpha = \bigtimes_{i\in\mathscr{B}} \mathscr{A}_i^\alpha$ *and there exists an assortment* $S_{\text{root}}^\alpha = \bigcup_{i\in\mathscr{B}} S_i^\alpha \in \Im_{\text{root}}^\alpha$ *such that* $\alpha R(S_{\text{root}}^\alpha) \geqslant Z^*$.

Even though the small-size candidate collection $\mathscr{A}_i^\alpha$ for a basic node i is known, we still need an algorithm to get a small-size candidate collection for the upper level nodes, which cannot be the cartesian product of $\mathscr{A}_i^\alpha$. A crucial observation is that t_i^α remains to be a constant once the entire searching space is fixed, thus an optimal solution to problem (3.4) would be found if we try all possible values of t_i^α. Let $\tilde{S}_i^\alpha(t_i)$ be optimal to the following problem

$$\max_{S_i \subseteq \Im_i^\alpha} \{V_i(S_i)^{1/\gamma_i}\,(R_i(S_i) - t_i)\},$$

and let $\hat{S}_j^\alpha(t_i)$ be optimal to the following problem

$$\max_{S_j \subseteq \mathscr{A}_j^\alpha} \{V_j(S_j)\,(R_j(S_j) - t_i)\},$$

where $\mathscr{A}_j^\alpha$ includes S_j^α and $j \in i_C$. The following claim shows some nice property and the relationship between $\tilde{S}_i^\alpha(t_i)$ and $\hat{S}_j^\alpha(t_i)$.

Claim 5. *If* $V_i(\bigcup_{j\in i_C} \hat{S}_j^\alpha(t_i))^{1/\gamma_i}(R_i(\bigcup_{j\in i_C} \hat{S}_j^\alpha(t_i)) - t_i)$ *is nonnegative, we have* $\tilde{S}_i^\alpha(t_i) = \bigcup_{j\in i_C} \hat{S}_j^\alpha(t_i)$*; otherwise* $\tilde{S}_i^\alpha(t_i) = \varnothing$*. Furthermore,* $\tilde{\mathscr{A}}_i^\alpha = \{\tilde{S}_i^\alpha(t_i) : t_i \in \mathbb{R}\}$ *includes* $\hat{S}_i^\alpha(t_h^\alpha)$.

Proof. Proof: $\max_{S_i \subseteq \Im_i^\alpha}\{V_i(S_i)^{1/\gamma_i}\,(R_i(S_i) - t_i)\}$ is equivalent to

$$\begin{aligned}
&\max_{S_i \subseteq \Im_i^\alpha} \sum_{j\in i_C} V_j(S_j)(R_j(S_j - t_i)) - t_i V_{i0}\mathbf{1}(S_i \neq \varnothing)\\
&= \sum_{j\in i_C} \max_{S_j \subseteq \mathscr{A}_j^\alpha} V_j(S_j)(R_j(S_j - t_i)) - t_i V_{i0}\mathbf{1}(\bigcup_{j\in i_C} S_j \neq \varnothing),
\end{aligned}$$

thus we have $\tilde{S}_i^\alpha(t_i) = \bigcup_{j\in i_C} \hat{S}_j^\alpha(t_i)$ if $V_i(\bigcup_{j\in i_C} \hat{S}_j^\alpha(t_i))^{1/\gamma_i}(R_i(\bigcup_{j\in i_C} \hat{S}_j^\alpha(t_i)) - t_i) \geqslant 0$, otherwise $\tilde{S}_i^\alpha(t_i) = \varnothing$.

When $t = t_i^\alpha$, $\hat{S}_j^\alpha(t_i^\alpha)$ is optimal to problem (3.3) at node j since $\mathscr{A}_j^\alpha$ also contains S_j^α. Then $\bigcup_{j\in i_C} \hat{S}_j^\alpha(t_i^\alpha)$ is the optimal solution to (3.3) at node i because of Claim 9. Because $\bigcup_{j\in i_C} \hat{S}_j^\alpha(t_i^\alpha) \in \tilde{\mathscr{A}}_i^\alpha$, then $\tilde{\mathscr{A}}_i^\alpha$ includes the optimal solution $\hat{S}_i^\alpha(t_h^\alpha)$ to problem (3.3) at node i. □

Next lemma shows some properties of $V_j(\hat{S}_j^\alpha(t_i))$ and $R_j(\hat{S}_j^\alpha(t_i))$ as a function of t_i.

Lemma 10.

1. *If* $V_j(\hat{S}_j^\alpha(t_i))(R_j(\hat{S}_j^\alpha(t_i)) - t_i)$ *is a continuous function of* t_i *in a certain range for all* $j \in i_C$*, then* $V_i(\tilde{S}_i^\alpha(t_i))^{1/\gamma_i}(R_i(\tilde{S}_i^\alpha(t_i)) - t_i)$ *is also a continuous function of* t_i.

2. $|\tilde{\mathscr{A}}_i^\alpha| = \sum_{j\in i_C} |\mathscr{A}_j^\alpha|$.

3. *If we have an assumption that $V_j(\hat{S}_j^\alpha(t_i))$ is a decreasing step function and $R_j(\hat{S}_j^\alpha(t_i))$ is an increasing step function, then $V_i(\tilde{S}_i^\alpha(t_i))$ is a decreasing step function and $R_i(\tilde{S}_i^\alpha(t_i))$ is an increasing step function, respectively.*

Proof. Proof: Since $V_i(\tilde{S}_i^\alpha(t_i))^{1/\gamma_i}(R_i(\tilde{S}_i^\alpha(t_i)) - t_i) = \sum_{j\in i_C} V_j(\hat{S}_j^\alpha(t_i))(R_j(\hat{S}_j^\alpha(t_i)) - t_i) - t_i V_{i0}\mathbf{1}(\tilde{S}_i^\alpha(t_i) \neq \emptyset)$, then $V_i(\tilde{S}_i^\alpha(t_i))^{1/\gamma_i}(R_i(\tilde{S}_i^\alpha(t_i)) - t_i)$ is also a continuous function with respect to t_i in a certain range because $V_j(\hat{S}_j^\alpha(t_i))(R_j(\hat{S}_j^\alpha(t_i)) - t_i)$ is continuous in terms of t_i within this range for all $j \in i_C$.

Assume that $\hat{S}_j^\alpha(t_i)$ only changes at some points, the set of which is denoted as $\mathscr{F}_j^\alpha = \{F_j^0, F_j^1, ..., F_j^{|\mathscr{A}_j^\alpha|}\}$ where $F_j^0 = 0$ and $F_j^{|\mathscr{A}_j^\alpha|} = +\infty$. Then $\tilde{S}_i^\alpha(t_i)$ also only changes at the point set $\mathscr{F}_i^\alpha = \bigcup_{j\in i_C} \mathscr{F}_j^\alpha = \{F_i^0, F_i^1, ..., F_i^{D_i}\}$ with $F_i^0 = 0$ and $F_i^{D_i} = +\infty$. We can see that $|\tilde{\mathscr{A}}_i^\alpha| = D_i = \sum_{j\in i_C} |\mathscr{A}_j^\alpha|$. For the rest of this lemma, we only need to prove $V_i(\tilde{S}_i^\alpha(t_i))$ is decreasing and $R_i(\tilde{S}_i^\alpha(t_i))$ is increasing discontinuously at any point $t_i' \in \mathscr{F}_i^\alpha$. Let $\underline{S}_i = \lim_{\epsilon\to 0} \tilde{S}_i^\alpha(t_i' - \epsilon)$ and $\overline{S}_i = \lim_{\epsilon\to 0} \tilde{S}_i^\alpha(t_i' + \epsilon)$. Since $V_i(\tilde{S}_i^\alpha(t_i')) = (V_{i0} + \sum_{j\in i_C} V_j(\hat{S}_j^\alpha(t_i')))^{\gamma_i}$ and $V_j(\hat{S}_j^\alpha(t_i'))$ is decreasing discontinuously at t_i', then $V_i(\tilde{S}_i^\alpha(t_i'))$ is also decreasing discontinuously at t_i', thus we have $V_i(\underline{S}_i) > V_i(\overline{S}_i)$. Since $V_i(\tilde{S}_i^\alpha(t_i))^{1/\gamma_i}(R_i(\tilde{S}_i^\alpha(t_i)) - t_i)$ is a continuous function t_i, then $V_i(\tilde{S}_i^\alpha(t_i))^{1/\gamma_i}(R_i(\tilde{S}_i^\alpha(t_i)) - t_i)$ is continuous at t_i', we have $V_i(\underline{S}_i)^{1/\gamma_i}(R_i(\underline{S}_i) - t_i') = V_i(\overline{S}_i)^{1/\gamma_i}(R_i(\overline{S}_i) - t_i')$. So that

$$R_i(\underline{S}_i) = \left(\frac{V_i(\overline{S}_i)}{V_i(\underline{S}_i)}\right)^{1/\gamma_i} (R_i(\overline{S}_i) - t_i') + t_i' < (R_i(\overline{S}_i) - t_i') + t_i' = R_i(\overline{S}_i),$$

where the inequality is due to the fact that $V_i(\underline{S}_i) > V_i(\overline{S}_i)$. This lemma holds because of the arbitrariness of t_i'. □

Even if we know $\tilde{\mathscr{A}}_i^\alpha = \{\tilde{S}_i^\alpha(t_i) : t_i \in \mathbb{R}\}$ and the set of changing points $\mathscr{F}_i^\alpha = \{F_i^0, F_i^1, ..., F_i^{|\tilde{\mathscr{A}}_i^\alpha|}\}$, we are still not able to stitch them together as $\mathscr{A}_h^\alpha = \{\bigcup_{i\in h_C} \tilde{S}_i^\alpha(t_i) : t_i \in \mathbb{R}\}$ to get $\mathscr{A}_h^\alpha$ that includes $\hat{S}_h^\alpha(t_{h^P}^\alpha)$, since $\tilde{S}_i^\alpha(t_i)$ depends on t_i that is different across $i \in h_C$. However, if we can obtain $\mathscr{A}_i^\alpha = \{\hat{S}_i^\alpha(t_h) : t_h \in \mathbb{R}\}$ where $\hat{S}_i^\alpha(t_h)$ is optimal to $\max_{S_i \subseteq \tilde{\mathscr{A}}_i^\alpha} \{V_i(S_i)\,(R_i(S_i) - t_h)\}$, then $\mathscr{A}_h^\alpha$ can be found as $\mathscr{A}_h^\alpha = \{\bigcup_{i\in h_C} \hat{S}_i^\alpha(t_h) : t_h \in \mathbb{R}\}$.

We aim to solve the following optimization problem

$$\max_{S_i \subseteq \tilde{\mathscr{A}}_i^\alpha} \{V_i(S_i)\,(R_i(S_i) - t_h)\}.$$

Let $\hat{S}_i^\alpha(t_h)$ be optimal to the above problem. We make an observation that $\hat{S}_i^\alpha(t_h)$ does not change in certain intervals where $V_i(\hat{S}_i^\alpha(t_h))\left(R_i(\hat{S}_i^\alpha(t_h)) - t_h\right)$ is the highest among these $|\tilde{\mathscr{A}}_i^\alpha|$ lines. Then our goal is reduced to find those intervals so that $\hat{S}_i^\alpha(t_h)$ does not change when t_h takes value in each interval. Assume that set $\tilde{\mathscr{A}}_i^\alpha = \{\tilde{S}_i^\alpha(t_i) : t_i \in \mathbb{R}\}$ is ordered such that $V_i(\tilde{S}_i^\alpha(t_i))$ is a decreasing step function and $R_i(\tilde{S}_i^\alpha(t_i))$ is an increasing step function of

Algorithm 1: statement for function AssortmentInitialization

Input: $\tilde{\mathscr{A}}_i^\alpha = \{\tilde{S}_i^\alpha(t_i) : t_i \in \mathbb{R}\}$ and $\mathscr{F}_i^\alpha = \{F_i^0, F_i^1, ..., F_i^{|\tilde{\mathscr{A}}_i^\alpha|}\}$;
for $g = 1, ..., |\tilde{\mathscr{A}}_i^\alpha|$ **do**
  $S_g = \lim_{\epsilon \to 0} \tilde{S}_i^\alpha(F_i^g - \epsilon)$;
end
Let $g \leftarrow |\tilde{\mathscr{A}}_i^\alpha| - 1$ and $E = \emptyset$;
while $g > 0$ **do**
  $G_g \leftarrow \text{Int}(g+1, g)$;
  $l \leftarrow 0$;
  while $G_{g-l} < 0$ **do**
    $\mathscr{F}_i^\alpha \leftarrow \mathscr{F}_i^\alpha \backslash \{F_i^{g-l}\}$;
    $l \leftarrow l + 1$;
    $G_{g-l} \leftarrow \text{Int}(g+1, g-l)$;
  end
  $E \leftarrow \{G_{g-l}\} \cup E$;
  $g \leftarrow g - l$;
end
Relabel $\mathscr{F}_i^\alpha$ as $\{O_i^0, O_i^1, ..., O_i^{n_i}\}$;
for $g = 1, ..., n_i$ **do**
  $S_g = \lim_{\epsilon \to 0} \tilde{S}_i^\alpha(O_i^g - \epsilon)$;
end
Output: $S = \{S_g : g = 1, 2, ..., n_i\}$ and $E = \{E_g : g = 1, 2, ..., n_i\}$.

t_i, respectively. The following two algorithms shows that $\mathscr{A}_h^\alpha = \{\bigcup_{i \in h_C} \hat{S}_i^\alpha(t_h) : t_h \in \mathbb{R}\}$ can be obtained in a tractable way.

In both Algorithms 1 and 2, function $\text{Int}(a, b)$ is to calculate the x-coordinate of the intersection points of line $V_i(S_a)(R_i(S_a) - t_h)$ and line $V_i(S_b)(R_i(S_b) - t_h)$ as $\text{Int}(a, b) = (V_i(S_a)R_i(S_a) - V_i(S_b)R_i(S_b)) / (V_i(S_a) - V_i(S_b))$.

Function AssortmentInitialization that is defined in Algorithm 1 calculates the positive consecutive intersection points of lines: $f(S_i, t_h) = V_i(S_i)(R_i(S_i) - t_h)$ for $S_i \in \tilde{\mathscr{A}}_i^\alpha = \{\tilde{S}_i^\alpha(t_i) : t_i \in \mathbb{R}\}$. By the ordering of $\tilde{S}_i^\alpha(t_i)$ where $t_i \in \mathbb{R}$ and the third item in Lemma 10, we make a remark that if the x-coordinate of the intersection point of two consecutive lines $f(S_n, t_h)$ and $f(S_{n+1}, t_h)$ is negative, then assortment S_n is dominated by S_{n+1}, which means that $f(S_{n+1}, t_h)$ is always larger than $f(S_n, t_h)$ as long as $t_h \geqslant 0$. So we can delete assortment S_n, and calculate the intersection points of lines $f(S_{n-1}, t_h)$ and $f(S_{n+1}, t_h)$, if it is still negative, we compute the intersection points of lines $f(S_{n-2}, t_h)$ and $f(S_{n+1}, t_h)$ and so forth until we get an positive intersection point and then record it in set E. We remark that the elements in set E is constructed in an increasing order. After deleting the dominated assortments, Algorithm 1 outputs the set of remaining candidate assortments

$S = \{S_g : g = 1, 2, ..., n_i\}$. The core step in Algorithm 1 is from line 5 to line 15, which includes deleting dominated assortments and calculating positive intersection points. We observe that the size of candidate assortments is reduced from $|\tilde{\mathscr{A}}_i^\alpha|$ to n_i. We then use the output of this function as an input to function AssortmentStitching that is stated in Algorithm 2 to get set $\mathscr{A}_i^\alpha = \{\hat{S}_i^\alpha(t_h) : t_h \in \mathbb{R}\}$.

Algorithm 2: statement for function AssortmentStitching

Input: $S = \{S_g : g = 1, 2, ..., n_i\}$ and $E = \{E_g : g = 1, 2, ..., n_i\}$.
Let $n \leftarrow 1$, $E_0 = 0$ and $g \leftarrow n_i$;
while $g > 0$ **do**
  $D_n \leftarrow E_g$;
  $m \leftarrow 1$;
  while $D_n < E_{g-m}$ **do**
    $m \leftarrow m + 1$
  end
  $D_n \leftarrow \text{Int}(g + 1, g - m + 1)$;
  $g \leftarrow g - m$;
  $n \leftarrow n + 1$;
end
Reverse the numerical array $\{D_1, D_2, ..., D_{n-1}\}$;
Let $D_0 \leftarrow -\infty$ and $D_n \leftarrow +\infty$;
for $l = 0, 2, ..., n - 1$ **do**
  $\hat{S}_i^\alpha(t_h) = \lim_{\epsilon \to 0} \tilde{S}_i^\alpha(D_l + \epsilon)$ for $t_h \in [D_l, D_{l+1}]$;
end
Output: $\mathscr{A}_i^\alpha = \{\hat{S}_i^\alpha(t_h) : t_h \in \mathbb{R}\}$ and $\mathscr{D}_i^\alpha = \{D_0, D_1, ..., D_n\}$.

In function AssortmentStitching, we have $E_g = \text{Int}(S_g, S_{g+1})$. The idea of Algorithm 2 is that if $E_g < E_{g-1}$, then line S_g can be deleted since line $f(S_g, t_h)$ is always lower than line $f(S_{g+1}, t_h)$ when t_h is nonnegative. It can be proved by using Lemma 10: since $V_i(S_{g-1}) > V_i(S_g) > V_i(S_{g+1})$ and $R_i(S_{g-1}) < R_i(S_g) < R_i(S_{g+1})$, we have $f(S_g, t_h) < f(S_{g+1}, t_h)$ for $t_h \geqslant 0$. Similar to Algorithm 1, it does not stop until $E_g \geqslant E_{g-m}$ where $m \geqslant 1$ by deleting assortments S_g, S_{g-1}, ..., S_{g-m+1}. Then we record the newly calculated intersection points in an array $\{D_1, D_2, ..., D_{n-1}\}$ and reverse it such that it has an increasing order. The output of this algorithm is $\mathscr{A}_i^\alpha = \{\hat{S}_i^\alpha(t_h) : t_h \in \mathbb{R}\}$ with corresponding changing points set $\mathscr{D}_i^\alpha = \{D_0, D_1, ..., D_n\}$ as desired.

The next proposition shows that $\mathscr{A}_i^\alpha = \{\hat{S}_i^\alpha(t_h) : t_h \in \mathbb{R}\}$ can be obtained in an efficient way by applying function AssortmentInitialization and function AssortmentStitching.

Proposition 8. *Let $\mathscr{A}_i^\alpha = \{\hat{S}_i^\alpha(t_h) : t_h \in \mathbb{R}\}$ where $\hat{S}_i^\alpha(t_h)$ is optimal to $\max_{S_i \subseteq \tilde{\mathscr{A}}_i^\alpha} \{V_i(S_i)\,(R_i(S_i) - t_h)\}$, then $\mathscr{A}_i^\alpha$ includes $\hat{S}_i^\alpha(t_h^\alpha)$. Furthermore, $|\mathscr{A}_i^\alpha| \leqslant |\tilde{\mathscr{A}}_i^\alpha|$ and*

$\mathscr{A}_i^\alpha$ can be computed within $O(|\tilde{\mathscr{A}_i^\alpha}|)$ operations by functions AssortmentInitialization and AssortmentStitching under the assumption in Lemma 10.

Proof. Proof: According to Claim 5, $\tilde{\mathscr{A}_i^\alpha}$ includes $\hat{S}_i^\alpha(t_h^\alpha)$, thus the feasible region of problem $\max_{S_i \subseteq \Im_i^\alpha}\{V_i(S_i)\,(R_i(S_i) - t_h)\}$ can be reduced from $\Im_i^\alpha$ to $\tilde{\mathscr{A}_i^\alpha}$ while preserving optimality. When $t_h = t_h^\alpha$, $\hat{S}_i^\alpha(t_h^\alpha) \in \mathscr{A}_i^\alpha$, thus $\mathscr{A}_i^\alpha$ contains $\hat{S}_i^\alpha(t_h^\alpha)$.

We use Algorithm 1 and 2 to compute $\mathscr{A}_i^\alpha = \{\hat{S}_i^\alpha(t_h) : t_h \in \mathbb{R}\}$ from $\tilde{\mathscr{A}_i^\alpha} = \{\tilde{S}_i^\alpha(t_i) : t_i \in \mathbb{R}\}$ and $\mathscr{F}_i^\alpha = \{F_i^0, F_i^1, ..., F_i^{|\tilde{\mathscr{A}_i^\alpha}|}\}$. In Algorithm 1, lines 1-3 take $O(|\tilde{\mathscr{A}_i^\alpha}|)$ operations. Lines 5-15 take $O(|\tilde{\mathscr{A}_i^\alpha}| - 1)$ since it deletes at most $|\tilde{\mathscr{A}_i^\alpha}| - 2$ assortments. Both line 16 and lines 17-19 take $O(n_i)$ where $n_i \leqslant |\tilde{\mathscr{A}_i^\alpha}|$. Thus the computational complexity of Algorithm 1 is $O(|\tilde{\mathscr{A}_i^\alpha}|) + O(|\tilde{\mathscr{A}_i^\alpha}| - 1) + O(n_i) + O(n_i) = O(|\tilde{\mathscr{A}_i^\alpha}|)$.

For Algorithm 2, lines 2-11 take $O(n_i)$ operations since it deletes at most n_i assortments. Line 12 takes $O(1)$ and lines 14-16 take $O(n)$ where $n \leqslant n_i$. So the computational complexity of Algorithm 2 is $O(n_i) + O(1) + O(n) = O(n_i)$.

For the size of $\mathscr{A}_i^\alpha$, we have $|\mathscr{A}_i^\alpha| = n \leqslant n_i \leqslant |\tilde{\mathscr{A}_i^\alpha}|$, and $\mathscr{A}_i^\alpha$ can be computed in $O(|\tilde{\mathscr{A}_i^\alpha}|) + O(n_i) = O(|\tilde{\mathscr{A}_i^\alpha}|)$ via Algorithms 1 and 2. □

By the Proposition 8, we immediately have the following two corollaries.

Corollary 3. *If $V_i(\tilde{S}_i^\alpha(t_i))$ is a decreasing step function and $R_i(\tilde{S}_i^\alpha(t_i))$ is an increasing step function, then $V_i(\hat{S}_i^\alpha(t_h))$ is a decreasing step function and $R_i(\hat{S}_i^\alpha(t_h))$ is an increasing step function.*

Corollary 4. *Let $\tilde{S}_h^\alpha(t_h) = \bigcup_{i \in h_C} \hat{S}_i^\alpha(t_h)$ and $\tilde{\mathscr{A}_h^\alpha} = \{\tilde{S}_h^\alpha(t_h) : t_h \in \mathbb{R}\}$ where h is the parent node of i, then $\tilde{\mathscr{A}_h^\alpha}$ can be computed in $O(\sum_{i \in h_c} |\mathscr{A}_i^\alpha|)$ operations.*

According to Corollary 3, Corollary 4 and the third item in Lemma 10, if we know $\tilde{\mathscr{A}_j^\alpha}$ for all the basic node$j \in \mathscr{B}$, a bottom-up method can be used repeatedly in a breath-first manner to get $\tilde{\mathscr{A}}_{\text{root}}^\alpha$, which can be further used to get $\mathscr{A}_{\text{root}}^\alpha$ via Algorithms 1 and 2, where $\mathscr{A}_{\text{root}}^\alpha$ is the candidate collection including an α-approximate solution S_{root}^α for the root node. After getting $\mathscr{A}_{\text{root}}^\alpha$, check the expected profit of every candidate assortment in $\mathscr{A}_{\text{root}}^\alpha$ to select the optimal assortment as S_{root}^α. The following theorem summarizes this finding.

Theorem 3. *Under the assumption in Lemma 10, for all the nonleaf nodes $i \in V$, we can construct $\mathscr{A}_i^\alpha$ with size $O(\sum_{j \in \mathscr{B}} |\tilde{\mathscr{A}_j^\alpha}|)$ in $O(\sum_{j \in \mathscr{B}} |\tilde{\mathscr{A}_j^\alpha}|)$ operations. The size of the candidate collection of assortments $\mathscr{A}_{\text{root}}^\alpha$ containing an α-approximate solution S_{root}^α satisfies $|\mathscr{A}_{\text{root}}^\alpha| \leqslant \sum_{j \in \mathscr{B}} |\tilde{\mathscr{A}_j^\alpha}|$. Constructing $\mathscr{A}_{\text{root}}^\alpha$ requires $O(m \cdot \sum_{j \in \mathscr{B}} |\tilde{\mathscr{A}_j^\alpha}|)$ operations.*

Proof. Proof: We claim that $\sum_{i:\text{level}(i)=l} |\mathscr{A}_i^\alpha| \leqslant \sum_{j \in \mathscr{B}} |\tilde{\mathscr{A}_j^\alpha}|$ for $l = 0, 1, ..., m-1$. This can be proved by induction on level l: 1) It is true for $l = m-1$; 2) Assume it is true for $l = L$ where $1 \leqslant m-1$: $\sum_{i:\text{level}(i)=L} |\mathscr{A}_i^\alpha| \leqslant \sum_{j \in \mathscr{B}} |\tilde{\mathscr{A}_j^\alpha}|$; 3) We have $\sum_{h:\text{level}(h)=L-1} |\mathscr{A}_h^\alpha| \leqslant \sum_{h:\text{level}(h)=L-1} |\tilde{\mathscr{A}_h^\alpha}| = \sum_{h:\text{level}(h)=L-1} \sum_{i \in h_C} |\mathscr{A}_i^\alpha| = \sum_{i:\text{level}(i)=L} |\tilde{\mathscr{A}_i^\alpha}| \leqslant \sum_{j \in \mathscr{B}} |\tilde{\mathscr{A}_j^\alpha}|$ because of Proposition 8 and Corollary 4. Thus constructing $\mathscr{A}_{\text{root}}^\alpha$ via Algorithms 1 and 2 requires $\sum_{L=0}^{m-1} \sum_{i:\text{level}(i)=L} |\tilde{\mathscr{A}_i^\alpha}| \leqslant \sum_{L=0}^{m-1} \sum_{j \in \mathscr{B}} |\tilde{\mathscr{A}_j^\alpha}| = O(m \cdot \sum_{j \in \mathscr{B}} |\tilde{\mathscr{A}_j^\alpha}|)$ operations. □

Theorem 3 tells us if $\tilde{\mathscr{A}}_j^\alpha$ has a polynomial size for all $j \in \mathscr{B}$, then constructing $\mathscr{A}_{\text{root}}^\alpha$ would also require polynomial size of operations. Furthermore, it also implies that problem (3.1) can be solved in polynomial time. In the following two subsections, we show how to construct the polynomial-size $\tilde{\mathscr{A}}_j^*$ under cardinality constraints and $\tilde{\mathscr{A}}_j^\alpha$ under space constraints for all basic nodes $j \in \mathscr{B}$, respectively. We also show that the assumption in Lemma 10 can be satisfied.

Cardinality Constraints

In this subsection, we introduce the Constrained Assortment Optimization Algorithm Under Cardinality Constraints (CAOA-C) that solves problem (3.1) with $\alpha = 1$ under cardinality constraints, which finds the optimal, or 1-approximate, assortment S_{root}^* in $O(n \max\{m, k\})$ time, where m is the number of levels in the multilevel nested logit model, n is the number of products and k is the maximum number of products of any basic nodes.

In order to obtain a candidate collection of assortments including the optimal solution S_j^* for a basic node $j \in \mathscr{B}$, we need to solve problem (3.4) with $\alpha = 1$ and $t_i^\alpha = t_j$ as follows

$$\max_{S_j \subseteq \Im_j^*} \{V_j(S_j)^{1/\gamma_j}(R_j(S_j) - t_j)\}, \tag{3.5}$$

where $\Im_j^*$ includes all the possible combination of feasible assortments that satisfy cardinality constraints and the size of $\Im_j^*$ is $\binom{N_j}{\mathbb{C}_j}$. We aim for reducing $\Im_j^*$ to a polynomial-size collection $\tilde{\mathscr{A}}_j^*$ such that $\tilde{\mathscr{A}}_j^* = \{\tilde{S}_j^*(t_j) : t_j \in \mathbb{R}\}$ where $\tilde{S}_j^*(t_j)$ represents the optimal solution to problem (3.5).

We observe that problem (3.5) at basic node j is more general than the constrained assortment optimization problem under the multinomial logit model [39]. The difference is that we need to consider the no-purchase option when constructing $\tilde{\mathscr{A}}_j^*$. [38] study the constrained assortment optimization problem under the multinomial logit model with space constraints and shows it is NP-hard. With cardinality constraints, assortment optimization problem under the multinomial logit model can be solved in polynomial time [39].

Problem (3.5) can be rewritten as

$$\max_{S_j \subseteq \Im_j^*} \{V_j(S_j)^{1/\gamma_j}(R_j(S_j) - t_j)\} = \max_{S_j \subseteq \Im_j^*} \{\sum_{k \in S_j} v_k(r_k - t_j) - V_{j0}\mathbf{1}(S_j \neq \varnothing)t_j\}.$$

It is a 0-1 knapsack problem with unit weight. If S_j is not empty, then the value of j for knapsack problem is $\sum_{k \in S_j} v_k(r_k - t_i) - V_{j0}t_j$ for a given t_j. After sorting the products by its valuein a decreasing order, the optimal solution $\tilde{S}_j^*(t_j)$ includes the first $\mathbb{C}_j$ products. To better illustrate the algorithm for this problem, inspired by [39], we define n linear functions $h_k(t_j) = v_k(r_k - t_j)$ for $k \in S_j$. When t_j takes values between two consecutive intersection points of the n lines, the ordering does not change so the optimal solution $\tilde{S}_j^*(t_j)$ would not change either. We define $h_0(t_j) = V_{j0}t_j$, then if $\sum_{k \in S_j} h_k(t_j) < h_0(t_j)$, we have $S_j = \varnothing$. We use Figure 4.2 as an example for three products with cardinality limitation of 2. In this case, there are 3 intersection points between the 3 linear functions and the real line is divided into 4 intervals by these 3 points. Under the cardinality constraints ($\mathbb{C} = 2$), the candidate collection contains 3 different feasible assortments: $\{\{1,2\}, \{2,3\}, \varnothing\}$, the corresponding range of t_j is shown in Table 3.1. Under mild conditions, [39] shows that the size of $\tilde{\mathscr{A}}_j^* = \{\tilde{S}_j^*(t_j) : t_j \in \mathbb{R}\}$ is $O(N_j)$ if the cardinality capacity $\mathbb{C}_j$ is fixed and constructing $\tilde{\mathscr{A}}_j^*$ requires $O(N_j^2)$ operations, where the set of changing points is denoted as $\mathscr{F}_j^* = \{F_j^0, F_j^1, ..., F_j^{|\tilde{\mathscr{A}}_j^*|}\}$.

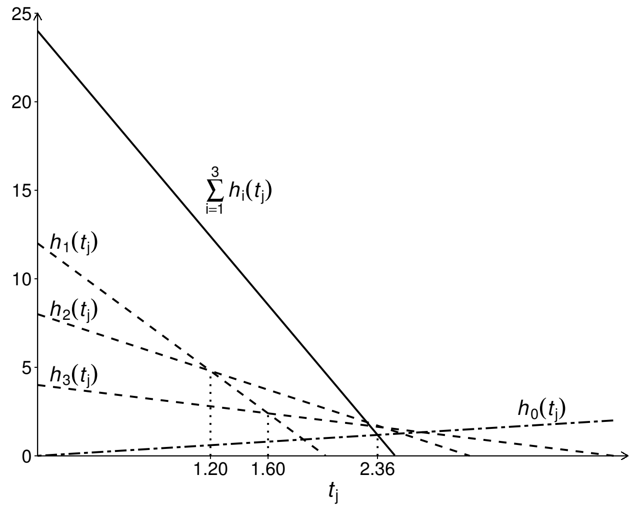

Figure 3.2: Three products with cardinality constraint $\mathbb{C}$= 2

t_j	$(-\infty, 1.6)$	$(1.6, 2.36)$	$(2.36, +\infty)$
$\tilde{S}_j^*(t_j)$	$\{1,2\}$	$\{2,3\}$	$\varnothing$

Table 3.1: The optimal solution $\tilde{S}_j^*(t_j)$ to problem (3.5)

By construction, $V_k(\hat{S}_k^\alpha(t_j))$ is a decreasing step function and $R_j(\hat{S}_j^\alpha(t_j))$ is an increasing step function because a product k with larger slope v_k and smaller profit r_k would always be replaced by a new product k' with smaller slope $v_{k'}$ and larger profit $r_{k'}$ as t_j increases. It satisfies the assumption of the third item in Lemma 10, thus we have $V_j(\tilde{S}_j^\alpha(t_j))$ is a decreasing step function and $R_j(\tilde{S}_j^\alpha(t_j))$ is an increasing step function. Therefore, we can use Algorithms 1 and 2 in a bottom-up manner and they run in polynomial-time by Proposition 8.

We are now ready to show the Constrained Assortment Optimization Algorithm Under Cardinality Constraints (CAOA-C) as follows. First, we get $\tilde{\mathscr{A}}_j^*$ and $\mathscr{F}_j^*$ for all basic nodes $j \in \mathscr{B}$; second, within each loop, we use $\tilde{\mathscr{A}}_j^*$ and $\mathscr{F}_j^*$ as the input of function AssortmentInitialization, then we feed the output of function AssortmentInitialization into function AssortmentStitching to obtain $\mathscr{A}_i^* = \{\hat{S}_i^*(t_h) : t_h \in \mathbb{R}\}$ and $\mathscr{D}_i^*$; third, we get $\tilde{\mathscr{A}}_h^* = \{\tilde{S}_h^*(t_h) : t_h \in \mathbb{R}\}$ where $\tilde{S}_h^*(t_h) = \bigcup_{i \in h_c} \hat{S}_i^*(t_h)$ and $\mathscr{F}_h^* = \bigcup_{i \in h_c} \mathscr{D}_i^*$ for node h that is the parent node of i. We call function AssortmentInitialization and AssortmentStitching repeatedly in a bottom-up manner to get $\mathscr{A}_{\text{root}}^*$, then the optimal assortment can be obtained as $S_{\text{root}}^* = \arg\max_{S_{\text{root}} \subseteq \mathscr{A}_{\text{root}}^*} R_{\text{root}}(S_{\text{root}})$ and the maximum profit is $Z^* = R_{\text{root}}(S_{\text{root}}^*)$.

Algorithm 3: Constrained Assortment Optimization Algorithm Under Cardinality Constraints (CAOA-C)

Input: v_k, r_k for product k, and γ_i for nonleaf node $i \in V$

1 **Initialization** Get $\tilde{\mathscr{A}}_j^* = \{\tilde{S}_j^*(t_j) : t_j \in \mathbb{R}\}$ and $\mathscr{F}_j^* = \{F_j^0, F_j^1, ..., F_j^{|\tilde{\mathscr{A}}_j^*|}\}$ for all $j \in \mathscr{B}$;

2 **for** $l = m-1, m-2, ..., 1$ **do**

3   **for** $i \in \text{Level}(l)$ **do**

4   **if** $l \neq m-1$ **then**

5             $\tilde{\mathscr{A}}_i^* = \{\tilde{S}_i^*(t_i) : t_i \in \mathbb{R}\}$ where $\tilde{S}_i^*(t_i) = \bigcup_{j \in i_C} \hat{S}_j^*(t_i)$;

6             $\mathscr{F}_i^* = \bigcup_{j \in i_c} \mathscr{D}_j^*$;

7   **end**

8         $S, E \leftarrow$ AssortmentInitialization$(\tilde{\mathscr{A}}_i^*, \mathscr{F}_i^*)$;

9         $\mathscr{A}_i^* = \{\hat{S}_i^*(t_h) : t_h \in \mathbb{R}\}$ and $\mathscr{D}_i^* \leftarrow$ AssortmentStitching(S, E);

10   **end**

11 **end**

12 $\mathscr{A}_{\text{root}}^* = \{\bigcup_{i \in \text{root}_C} \hat{S}_i^*(t_{\text{root}}) : t_{\text{root}} \in \mathbb{R}\}$;

13 $S_{\text{root}}^* = \arg\max_{S_{\text{root}} \subseteq \mathscr{A}_{\text{root}}^*} R_{\text{root}}(S_{\text{root}})$ and $Z^* = R_{\text{root}}(S_{\text{root}}^*)$;

Output: S_{root}^* and Z^*.

We use the following theorem to end this subsection, which summarizes above findings and shows that the assortment optimization problem under cardinality constraints can be solved in polynomial time.

Theorem 4. *The optimal assortment S_{root}^* to problem (3.1) with $\alpha = 1$ under cardinality constraints can be obtained within $O(n \max\{m, k\})$ operations by Algorithm CAOA-C.*

Proof. Proof: For the Algorithm CAOA-C, line 1 takes $O(\sum_{j \in \mathscr{B}} N_j^2)$ to get $\tilde{\mathscr{A}}_j^*$ with size $|\tilde{\mathscr{A}}_j^*| = O(N_j)$ for all $j \in \mathscr{B}$. According to Theorem 3, lines 2-11 take $O(m \cdot \sum_{j \in \mathscr{B}} |\tilde{\mathscr{A}}_j^*|) = O(m \cdot \sum_{j \in \mathscr{B}} N_j) = O(mn)$. Lines 12-13 also take $O(mn)$ operations. For the complexity of line 1, we have

$$\sum_{j\in\mathscr{B}} N_j^2 \leqslant \max_{j\in\mathscr{B}} N_j \cdot \sum_{j\in\mathscr{B}} N_j = k \cdot \sum_{j\in\mathscr{B}} N_j = kn,$$

thus the complexity of line 1 is $O(kn)$. Therefore, the total number of operations of Algorithm CAOA-C is $O(kn) + O(mn) + O(mn) = O(n \max\{m, k\})$. □

Space Constraints

Similar to Section 16, we show the Constrained Assortment Optimization Algorithm Under Space Constraints (CAOA-S) that solves problem (3.1) under space constraints, which finds an α-approximate assortment S^{α}_{root} in $O(mnk)$ time, where m is the number of levels in the multilevel nested logit model, n is the number of products and k is the maximum number of products of any basic nodes.

Similar to the problem with cardinality constraints, we show how to find an α-approximate collection $\tilde{\mathscr{A}}^{\alpha}_j$ for a basic node $j \in \mathscr{B}$ with $\alpha = 2$. We consider the following problem

$$\max_{S_j \subseteq \Im^{\alpha}_j} \{V_j(S_j)^{1/\gamma_j}(R_j(S_j) - t_j)\}, \tag{3.6}$$

where we use $\Im^{\alpha}_j$ to denote all the combinations of feasible assortments satisfying the space constraints. Our goal is to reduce $\Im^{\alpha}_j$ to a polynomial-size collection $\tilde{\mathscr{A}}^{\alpha}_j$ such that $\tilde{\mathscr{A}}^{\alpha}_j = \{\tilde{S}^{\alpha}_j(t_j) : t_j \in \mathbb{R}\}$ where $\tilde{S}^{\alpha}_j(t_j)$ satisfies that $V_j(\tilde{S}^{\alpha}_j(t_j))^{1/\gamma_j}\left(\alpha R_j(\tilde{S}^{\alpha}_j(t_j)) - t_j\right) \geqslant V_j(S^*_j)^{1/\gamma_j}\left(R_j(S^*_j) - t_j\right)$. If $\tilde{\mathscr{A}}^{\alpha}_i$ is known, then according to Lemma 9, $S^{\alpha}_{\text{root}} = \bigcup_{i\in\mathscr{B}} \tilde{S}^{\alpha}_i(t^*_j)$ is an α-approximate solution, thus $\Im^{\alpha}_{\text{root}} = \times_{j\in\mathscr{B}} \tilde{\mathscr{A}}^{\alpha}_j$ includes S^{α}_{root} when $t_j = t^*_j$ for all $j \in \mathscr{B}$. By applying Algorithm 1 and 2 repeatedly from the bottom to top, we manage to find the best assortment in $\Im^{\alpha}_{\text{root}}$, which is S^{α}_{root}. Next we show how to find corresponding $\tilde{\mathscr{A}}^{\alpha}_j$.

Problem (3.6) can be re-derived as

$$\max_{S_j \subseteq \Im^{\alpha}_j} \{V_j(S_j)^{1/\gamma_j}(R_j(S_j) - t_j)\} = \max_{S_j \subseteq \Im^{\alpha}_j} \{\sum_{k\in S_j} v_k(r_k - t_j) - V_{j0}\mathbf{1}(S_j \neq \varnothing)t_j\}. \tag{3.7}$$

Unfortunately, problem (3.7) at a basic node j is NP-hard with space constraints. Hence we use a linear relaxation method to solve problem $\max_{S_j \subseteq \Im^{\alpha}_j}\{\sum_{k\in S_j} v_k(r_k - t_j)\}$. Let $S'_j(t_j)$ denote an α-approximate solution such that

$$\alpha \sum_{k\in S'_j(t_j)} v_k(r_k - t_j) \geqslant \sum_{k\in S^*_j} v_k(r_k - t_j)$$

where $\alpha = 2$. By applying the algorithm in [39] and the approach in [18], we are able to construct the collection $\mathscr{A}'_j = \{S'_j(t_j) : t_j \in \mathbb{R}\}$ with size $O(N_j^2)$ for fixed $\mathbb{S}_j$ in $O(N_j^2)$ operations. Note that [18] also show that the performance of guarantee of 2 can be further refined to $\alpha = 1/(1-\epsilon)$ ($\epsilon \in [0, 1)$) under certain assumptions of the data, which also applies to our case.

We claim that $S'_j(t_j)$ satisfies $V_j(S'_j(t_j))^{1/\gamma_j}\left(\alpha R_j(S'_j(t_j)) - t_j\right) \geqslant V_j(S^*_j)^{1/\gamma_j}\Big(R_j(S^*_j) - t_j\Big)$ if $S^*_j \neq \varnothing$. Next we show how to prove this claim, we have

$$\begin{aligned} &V_j(S'_j(t_j))^{1/\gamma_j}\Big(\alpha R_j(S'_j(t_j)) - t_j\Big) = \alpha \sum_{k\in S'_j(t_j)} v_k r_k - V_{j0}\mathbf{1}(S'_j(t_j) \neq \varnothing)t_j - \sum_{k\in S'_j(t_j)} v_k t_j \\ &\geqslant \alpha \sum_{k\in S'_j(t_j)} v_k(r_k - t_j) - V_{j0}\mathbf{1}(S'_j(t_j) \neq \varnothing)t_j \geqslant \alpha \sum_{k\in S'_j(t_j)} v_k(r_k - t_j) - V_{j0}t_j \\ &\geqslant \sum_{k\in S^*_j} v_k(r_k - t_j) - V_{j0}t_j = \sum_{k\in S^*_j} v_k(r_k - t_j) - V_{j0}\mathbf{1}(S^*_j \neq \varnothing)t_j = V_j(S^*_j)^{1/\gamma_j}\Big(R_j(S^*_j) - t_j\Big), \end{aligned}$$

which establishes the claim.

The difference between problem (3.7) and the problem that is considered in [18] is that the objective function of problem (3.7) contains an additional term $-V_{j0}\mathbf{1}(S_j \neq \varnothing)t_j$ since we allow the no-purchase options to exist in every choice stage. Thus we need to check if $V_j(S'_j(t_j))^{1/\gamma_j}(\alpha R_j(S'_j(t_j)) - t_j) \geqslant 0$. If so, then $\tilde{S}^\alpha_j(t_j) = S'_j(t_j)$; otherwise $\tilde{S}^\alpha_j(t_j) = \varnothing$. Then $\tilde{S}^\alpha_j(t_j)$ satisfies that $V_j(S'_j(t_j))^{1/\gamma_j}(\alpha R_j(S'_j(t_j)) - t_j) \geqslant V_j(S^*_j)^{1/\gamma_j}(R_j(S^*_j) - t_j)$ because this inequality holds when $S^*_j \neq \varnothing$ due to the above claim and it also holds when $S^*_j = \varnothing$ since $V_j(S'_j(t_j))^{1/\gamma_j}(\alpha R_j(S'_j(t_j)) - t_j) \geqslant 0 = V_j(S^*_j)^{1/\gamma_j}(R_j(S^*_j) - t_j)$. In this way we can obtain the desired $\tilde{\mathscr{A}}^\alpha_j$ with corresponding set $\mathscr{F}^\alpha_j$ of changing points for all $j \in \mathscr{B}$.

Similar to the problem under cardinality constraints, we introduce the Constrained Assortment Optimization Algorithm Under Space Constraints (CAOA-S) as follows and show the complexity of CAOA-S in the next theorem.

Theorem 5. *A 2-approximate assortment S^α_{root} under space constraints can be obtained within $O(mnk)$ operations by Algorithm CAOA-S.*

Proof. Proof: For the Algorithm CAOA-S, line 1 takes $O(\sum_{j\in\mathscr{B}} N_j^2)$ to get $\tilde{\mathscr{A}}^\alpha_j$ with size $|\tilde{\mathscr{A}}^\alpha_j| = O(N_j^2)$ for all $j \in \mathscr{B}$. According to Theorem 3, lines 2-11 take $O(m \cdot \sum_{j\in\mathscr{B}} |\tilde{\mathscr{A}}^*_j|) = O(m \cdot \sum_{j\in\mathscr{B}} N_j^2)$. Lines 12-13 also take $O(m \cdot \sum_{j\in\mathscr{B}} N_j^2)$ operations. Because we have

$$\sum_{j\in\mathscr{B}} N_j^2 \leqslant \max_{j\in\mathscr{B}} N_j \cdot \sum_{j\in\mathscr{B}} N_j = k \cdot \sum_{j\in\mathscr{B}} N_j = kn,$$

thus the total number of operations of Algorithm CAOA-S is $O(kn) + O(mnk) + O(mnk) = O(mnk)$. □

3.4 Price Optimization

In this section, we study the price optimization problem, the goal of which is to maximize the expected profit per customer. The assortment S_{root} is assumed to be fixed as N_{root},

Algorithm 4: Constrained Assortment Optimization Algorithm Under Space Constraints (CAOA-S)

Input: v_k, r_k for product k, and γ_i for nonleaf node $i \in V$

Initialization Get $\tilde{\mathscr{A}_j^\alpha} = \{\tilde{S}_j^\alpha(t_j) : t_j \in \mathbb{R}\}$ and $\mathscr{F}_j^\alpha = \{F_j^0, F_j^1, ..., F_j^{|\tilde{\mathscr{A}_j^\alpha}|}\}$ for all $j \in \mathscr{B}$;

for $l = m-1, m-2, ..., 1$ **do**
- **for** $i \in \text{Level}(l)$ **do**
  - **if** $l \neq m-1$ **then**
    - $\tilde{\mathscr{A}_i^\alpha} = \{\tilde{S}_i^\alpha(t_i) : t_i \in \mathbb{R}\}$ where $\tilde{S}_i^\alpha(t_i) = \bigcup_{j \in i_C} \hat{S}_j^\alpha(t_i)$;
    - $\mathscr{F}_i^\alpha = \bigcup_{j \in i_c} \mathscr{D}_j^\alpha$;
  - **end**
  - $S, E \leftarrow \text{AssortmentInitialization}(\tilde{\mathscr{A}_i^\alpha}, \mathscr{F}_i^\alpha)$;
  - $\mathscr{A}_i^\alpha = \{\hat{S}_i^\alpha(t_h) : t_h \in \mathbb{R}\}$ and $\mathscr{D}_i^\alpha \leftarrow \text{AssortmentStitching}(S, E)$;
- **end**

end

$\mathscr{A}_{\text{root}}^\alpha = \{\bigcup_{i \in \text{root}_C} \hat{S}_i^\alpha(t_{\text{root}}) : t_{\text{root}} \in \mathbb{R}\}$;

$S_{\text{root}}^\alpha = \arg\max_{S_{\text{root}} \subseteq \mathscr{A}_{\text{root}}^\alpha} R_{\text{root}}(S_{\text{root}})$ and $Z^\alpha = R_{\text{root}}(S_{\text{root}}^\alpha)$;

Output: S_{root}^α and Z^α.

thus the decision variables are prices of all the products, denoted as $\mathbf{P}_{\text{root}} = (p_1, p_2, ..., p_n)$. $\mathbf{P}_{\text{root}} \in \mathbb{R}_+^n$ is a n-dimensional vector where n, as defined in Section 4.3, is the total number of products.

Problem Formulation

We use the same tree structure as in Section 4.3 to describe the customer choosing behavior. The n products are leaf nodes in the m-level tree. For every node in this tree, there can exist a no-purchase option associated with it. We assume that price of the no-purchase option is 0. For each node $i \in V$, it is assigned a preference weight $V_i(\mathbf{P}_i)$ by the customer. We define $V_i(\mathbf{P}_i)$ recursively as follows

$$V_i(\mathbf{P}_i) = \begin{cases} \exp(\alpha_i - \beta_i p_i) & \text{, if } i \text{ is a leaf node} \\ \left(V_{i0} + \sum_{j \in i_C} V_j(\mathbf{P}_j)\right)^{\gamma_i} & \text{, o.w.} \end{cases}$$

We can see that for a leaf node i, α_i can be interpreted as the price-independent part of the systematic utility of product i and $\beta_i > 0$ is the product-differentiated price sensitivity parameter. We define two scalars for nonleaf node $i \in V$:

$$\underline{B}_i = \begin{cases} \min_{j \in i_C}\{\beta_j\} & i \text{ is a basic node} \\ \min_{j \in i_C}\{\underline{B}_j \gamma_j\} & \text{o.w.} \end{cases}$$

and

$$\overline{B}_i = \begin{cases} \max_{j\in i_C}\{\beta_j\} & i \text{ is a basic node} \\ \max_{j\in i_C}\{\frac{\gamma_j^2\overline{B}_j}{1-(1-\gamma_j)\overline{B}_j/\underline{B}_j}\} & \text{o.w.} \end{cases}$$

As in [22], we also make the following assumption on the price-sensitivity parameters and dissimilarity parameters to guarantee that there is a unique optimal pricing solution.

Assumption 1. *For any nonleaf and nonroot node $i \in V$, we assume that*

$$\frac{\overline{B}_i}{\underline{B}_i} < \frac{1}{1-\gamma_i}.$$

We define the profit for node $i \in V$ as follows

$$R_i(\mathbf{P}_i) = \begin{cases} p_i - c_i & \text{, if } i \text{ is a leaf node} \\ \sum_{j\in i_C} V_j(\mathbf{P}_j)R_j(\mathbf{P}_j)/\left(V_{i0} + \sum_{j\in i_C} V_j(\mathbf{P}_j)\right) & \text{, o.w.} \end{cases}$$

Then the total expected profit can be expressed as $R_{\text{root}}(\mathbf{P}_{\text{root}})$. Therefore the price optimization problem can be formulated as

$$\max_{\mathbf{P}_{\text{root}}\in\mathbb{R}_+^n} R_{\text{root}}(\mathbf{P}_{\text{root}}). \tag{3.8}$$

We generalize the results of [19] in the following three folds: 1) the two-level nested logit model is generalized to the multilevel nested logit model; 2) the no-purchase option can exist in every stage of the customer choice process; 3) the adjusted nest-level markup is shown to be a special case of the *node-specific adjusted markup* which will be presented in the following subsection.

Constant Node-specific Adjusted Markup

In this subsection, we analyze properties of the price vector $\mathbf{P}_{\text{root}}$ at the optimality condition of problem (3.8). For the rest of this essay, we will use Q_i to represent the choice probability of product i. The markup for product i is defined as $m_i = p_i - c_i$. We use $\eta_{i,k}$ to denote the ancestor node of i in level k, where $(0 \leqslant k \leqslant m-1)$. With a slight abuse of notation, $\eta_{i,k}$ is also used to denote the collection of products that are associated with itself. The following lemma shows the expression of the first derivative of objective function that is defined in (3.8).

Lemma 11. *The first derivative of objective function $R_{\text{root}}(\boldsymbol{P}_{\text{root}})$ with respect to price p_i for any product i is*

$$
\begin{aligned}
&\partial R_{\text{root}}(\boldsymbol{P}_{\text{root}})/\partial p_i \\
&= \beta_i Q_i \Bigg(\frac{1}{\beta_i} - m_i + \lambda_{i,1}^{m-1} \sum_{i' \in \eta_{i,0}} m_{i'} Q_{i'} + \sum_{k=1}^{m-2} \sum_{\substack{i' \in \eta_{i,k} \\ i' \notin \eta_{i,k+1}}} m_{i'} \left(\sum_{t=1}^{k} (\lambda_{i,t+1}^{m-1} - \lambda_{i,t}^{m-1}) Q(i'|\eta_{i,t}) \right) \\
&+ \sum_{i' \in \eta_{i,m-1}} m_{i'} \left(\sum_{t=1}^{m-1} (\lambda_{i,t+1}^{m-1} - \lambda_{i,t}^{m-1}) Q(i'|\eta_{i,t}) \right) \Bigg),
\end{aligned}
$$

where $\lambda_{i,t}^{s} = \prod_{q=t}^{s} \gamma_{i,q}$.

Let the first derivative $\partial R_{\text{root}}(\mathbf{P}_{\text{root}})/\partial p_i = 0$, since $Q_i \neq 0$, after dividing $\beta_i Q_i$ and collecting terms, we have

$$
\begin{aligned}
m_i - 1/\beta_i = \lambda_{i,1}^{m-1} \sum_{i' \in \eta_{i,0}} m_{i'} Q_{i'} + \sum_{k=1}^{m-2} \sum_{\substack{i' \in \eta_{i,k} \\ i' \notin \eta_{i,k+1}}} m_{i'} \left(\sum_{t=1}^{k} (\lambda_{i,t+1}^{m-1} - \lambda_{i,t}^{m-1}) Q(i'|\eta_{i,t}) \right) \\
+ \sum_{i' \in \eta_{i,m-1}} m_{i'} \left(\sum_{t=1}^{m-1} (\lambda_{i,t+1}^{m-1} - \lambda_{i,t}^{m-1}) Q(i'|\eta_{i,t}) \right). \qquad (3.9)
\end{aligned}
$$

If we define the node-specific adjusted markup for product $i \in \mathscr{B}$ as $\theta_i = m_i - 1/\beta_i$, then Equation (3.9) implies that θ_i is identical for all $i \in \eta_{i,m-1}$. We make an important observation that the dimension of problem (3.8) has been reduced from n to $|\mathscr{B}|$ because there exists a one-to-one increasing mapping between the markup m_i for product $i \in \mathscr{B}$ and its node-specific adjusted markup θ_i. The idea behind this lemma is quite insightful, since the multidimensional price optimization problem can be reduced with regard to the number of decision variables. Originally, the number of decision variables equals the number of products in the assortment, by lemma 3, it can be reduced to the number of basic nodes.

The price for each product is a product-differentiated property, i.e. it is different across products. However, at optimality we can get the node-specific adjusted markup θ_i, which is invariant for the products in the basic nodes, via subtracting price by the product's own cost and the reciprocal of the price sensitivity parameter. We remark that at optimality, the node-specific adjusted markup is different across the basic nodes. Then the question is whether there still exists similar method so that the price optimization problem can be further simplified in terms of the problem dimensions. Before positively answering it, we first introduce the formal definition of node-specific adjusted markup, in which we use j to denote one of the children nodes of i.

Definition 1. *The node-specific adjusted markup for node i, the level of which is $0 \leqslant l \leqslant m-1$, is defined as*

$$\theta_i = \theta_j \delta_j(\theta_j) - \omega_j(\theta_j),$$

where $\delta_j(\theta_j)$ and $\omega_j(\theta_j)$, the level of node j is $1 \leqslant l \leqslant m-1$, are recursively defined as follows

$$\begin{aligned}
\delta_j(\theta_j) &= \frac{1}{\gamma_j} - (\frac{1}{\gamma_j} - 1)\tau_j(\theta_j), \\
\tau_j(\theta_j) &= \sum_{k \in j_C} \frac{Q(\theta_k|\theta_j)}{\delta_k(\theta_k)} \tau_k(\theta_k), \\
\omega_j(\theta_j) &= (\frac{1}{\gamma_j} - 1) \sum_{k \in j_C} \frac{Q(\theta_k|\theta_j)}{1-\gamma_k} \frac{\omega_k(\theta_k)}{\delta_k(\theta_k)},
\end{aligned}$$

where for leaf node k, we define $\delta_k(\theta_k) = \tau_k(\theta_k) = 1$, $\omega_k(\theta_k) = 1/\beta_k$ and $\gamma_k = 0$. For notational brevity, we denote $Q(\theta_k|\theta_j) = \frac{V_k(\theta_k)}{V_{i0} + \sum_{k \in j_C} V_k(\theta_k)}$.

For these quantities in the above definition, we have $0 \leqslant \tau_j(\theta_j) \leqslant 1$ and $1 \leqslant \delta_j(\theta_j) \leqslant 1/\gamma_j$. The *node-specific adjusted markup* is defined recursively, next we show an example of how to calculate it under the two-level nested logit model.

Example 1. *For the two-level nested logit model, in level 0, it is the root node; in level 1, there are $|\mathscr{B}|$ nests indexed by i; in level 2, there are products indexed by j. Then we have $\theta_i = p_j - c_j - 1/\beta_j$ and*

$$\begin{aligned}
\theta_{\text{root}} &= \theta_i \delta_i(\theta_i) - \omega_i(\theta_i) \\
&= \theta_i \left[\frac{1}{\gamma_i} - (\frac{1}{\gamma_i} - 1)\tau_i(\theta_i) \right] - (\frac{1}{\gamma_i} - 1) \sum_{j \in i_C} \frac{Q(\theta_j|\theta_i)}{1-\gamma_j} \frac{\omega_j(\theta_j)}{\delta_j(\theta_j)} \\
&= (p_j - c_j - 1/\beta_j) \left[\frac{1}{\gamma_i} - (\frac{1}{\gamma_i} - 1) \sum_{j \in i_C} Q(\theta_j|\theta_i) \right] - (\frac{1}{\gamma_i} - 1) \sum_{j \in i_C} \frac{Q(\theta_j|\theta_i)}{\beta_j}.
\end{aligned}$$

Define the following function recursively for each nonleaf node $i \in V$: $u_i(\theta_i) = \sum_{j \in i_C} \beta_j Q(\theta_j|\theta_i)$ if i is a basic node; otherwise
$u_i(\theta_i) = \sum_{j \in i_C} \frac{\gamma_j u_j}{\partial\theta_i/\partial\theta_j} Q(\theta_j|\theta_i)$. Next we assume that $\partial\theta_i/\partial\theta_j$ is bounded.

Assumption 2. *For any nonleaf node $i \in V$ and its child node j, we assume that*

$$\frac{\partial\theta_i}{\partial\theta_j} \leqslant \frac{1}{\gamma_j} - \frac{\omega_j(\theta_j) u_j(\theta_j)}{\delta_j(\theta_j)}.$$

Under Assumption 3 and 2, θ_i is an increasing function of θ_j. We record this finding in the following lemma.

Lemma 12. *Under Assumption 3 and 2, for any nonleaf node $i \in V$, we have*

$$\begin{aligned}\frac{\partial V_i(\theta_i)}{\partial \theta_i} &= -\gamma_i V_i(\theta_i) u_i(\theta_i),\\ \frac{\partial \omega_i(\theta_i)}{\partial \theta_i} &\leqslant \omega_i(\theta_i) u_i(\theta_i) - (\frac{1}{\gamma_i} - 1)\tau_i,\\ \frac{\partial \tau_i}{\partial \theta_i} &\leqslant -(1-\tau_i) u_i,\\ \frac{\partial \delta_i}{\partial \theta_i} &\geqslant (\delta_i - 1) u_i.\end{aligned}$$

Furthermore, we have

$$\omega_i u_i \leqslant \frac{1}{\gamma_i} \text{ and } \frac{\partial \theta_i}{\partial \theta_j} \geqslant 0.$$

The following proposition shows the properties of the *node-specific adjusted markup.*

Proposition 9. *For any nonleaf node i in level $0 \leqslant l \leqslant m-1$ and its arbitrary children nodes $j, j' \in i_C$, we have $\theta_i = \theta_j \delta_j(\theta_j) - \omega_j(\theta_j)$ and $\theta_i' = \theta_{j'} \delta_{j'}(\theta_{j'}) - \omega_{j'}(\theta_{j'})$ are equivalent under the optimality condition of problem (3.8). Furthermore, there exists a one-to-one increasing correspondence between the node-specific adjusted markup θ_i and θ_j under Assumption 3 and 2.*

The technique that we use in the above proposition is: for the *node-specific adjusted markup* θ_i, it can be obtained via subtracting θ_j by a value that only depends on N_j. By repeatedly using Proposition 9 from bottom to top, the price optimization problem (3.8), originally defined in the n-dimensional space, is reduced to a single-dimensional optimization problem by maximizing $R(\theta_{\text{root}})$ with respect to θ_{root}. Furthermore, the following theorem shows that $R(\theta_{\text{root}})$ is a unimodal function.

Theorem 6. *$R(\theta_{\text{root}})$ is strictly unimodal in θ_{root}. Moreover, we have $R(\theta^*_{\text{root}}) = \theta^*_{\text{root}}$ at optimality.*

We can use simple optimization algorithm, such as binary search, to find its optimal solution θ^*_{root} after having the correspondence of the *node-specific adjusted markup* between intermediate nodes. From Proposition 9, we know that there exists a one-to-one correspondence between θ_i, and θ_j for $j \in i_C$. Thus there also exists a one-to-one increasing mapping between θ^*_{root} and the optimal price p^*_k of product k. We denote this increasing mapping as $f_k(p^*_k) = \theta^*_{\text{root}}$. So if we know the optimal solution θ^*_{root}, then the optimal price of an actual product i can be calculated as $p^*_k = f_k^{-1}(\theta^*_{\text{root}})$. Therefore, the optimal price vector for problem (3.8) is $\mathbf{P}^*_{\text{root}} = F^{-1}(\theta^*_{\text{root}})$, where $F = (f_1^{-1}, f_2^{-1}, ..., f_n^{-1})$.

Numerical Example

Multiproduct price optimization problem (3.8) can be reduced to the maximization of a unimodal function

$$\max_{\theta_{\text{root}} \in \mathbb{R}_+} R_{\text{root}}(\theta_{\text{root}}). \tag{3.10}$$

At optimality, although θ^*_{root} has a one-to-one correspondence with p^*_k: $p^*_k = f_k^{-1}(\theta^*_{\text{root}})$, we cannot simply obtain a closed-form expression of f^{-1} due to its nonlinearity nature. However, starting from θ_{root}, we can recursively get θ_i according to Definition 1, then p_k can be found due to this linear relationship: $p_k = \theta_j + 1/\beta_k + c_k$ where the basic node j is parent of k . Therefore, $R(\theta_{\text{root}})$ can be identified, implying we can use some simple algorithms, such as golden section search and binary search, to locate the optimal solution $\theta^*_{\text{root}} = R(\theta^*_{\text{root}})$ to problem (3.10).

To better illustrate Theorem 6, we solve a small problem example with a three-level nested logit model that is shown in Figure 4.2. We can see for each nonleaf node, there is one no-purchase option associated with it, such as for node A, the no-purchase option is denoted as A_0. The parameters that are used in this numerical example are described in Table 4.2, which satisfies our assumptions.

Product	G	H	I	J	K1.8 *	L	M	N
α	15	12	13	11	10	8	14	9
β	1.8	1.6	1.7	2	2.2	2.1	2.4	1.8
cost	0.9	0.8	0.7	0.85	0.55	0.4	0.9	0.5
No-purchase	root_0	A_0	B_0	C_0	D_0	E_0	F_0	
V	0.15	0.12	0.1	0.05	0.07	0.06	0.04	
nonleaf Nodes	A	B	C	D	E	F		
γ	0.78	0.95	0.86	0.91	0.73	0.81		

Table 3.2: Parameters setup for the price optimization problem

The computation results are shown in Figure 4.3. For plot (a), we can clearly see that the objective function of problem (3.10) is unimodal with respect to θ_{root}. Moreover, when $R(\theta_{\text{root}})$ intersects with the 45°-line $f(\theta_{\text{root}}) = \theta_{\text{root}}$, $R(\theta_{\text{root}})$ reaches its optimal, i.e. $R(\theta^*_{\text{root}}) = \theta^*_{\text{root}} = 5.80$, as stated in Theorem 6. For plots $(b), (c)$ and (d), we recursively calculate the optimal node-specific adjusted markup $\theta^*_A = 5.917$, $\theta^*_B = 5.668$ and $\theta^*_C = 6.000$, $\theta^*_D = 5.953$, $\theta^*_E = 4.900$, $\theta^*_F = 5.500$ for basic nodes C, D, E and F, respectively. In the end, the optimal price of product k can be obtained by using this linear correspondence $p_k = \theta_j + c_k + 1/\beta_k$ where j is the parent node of k. For instance, $p^*_H = \theta^*_C + c_H + 1/\beta_H = 6.000 + 0.8 + 1/1.6 = 7.425$.

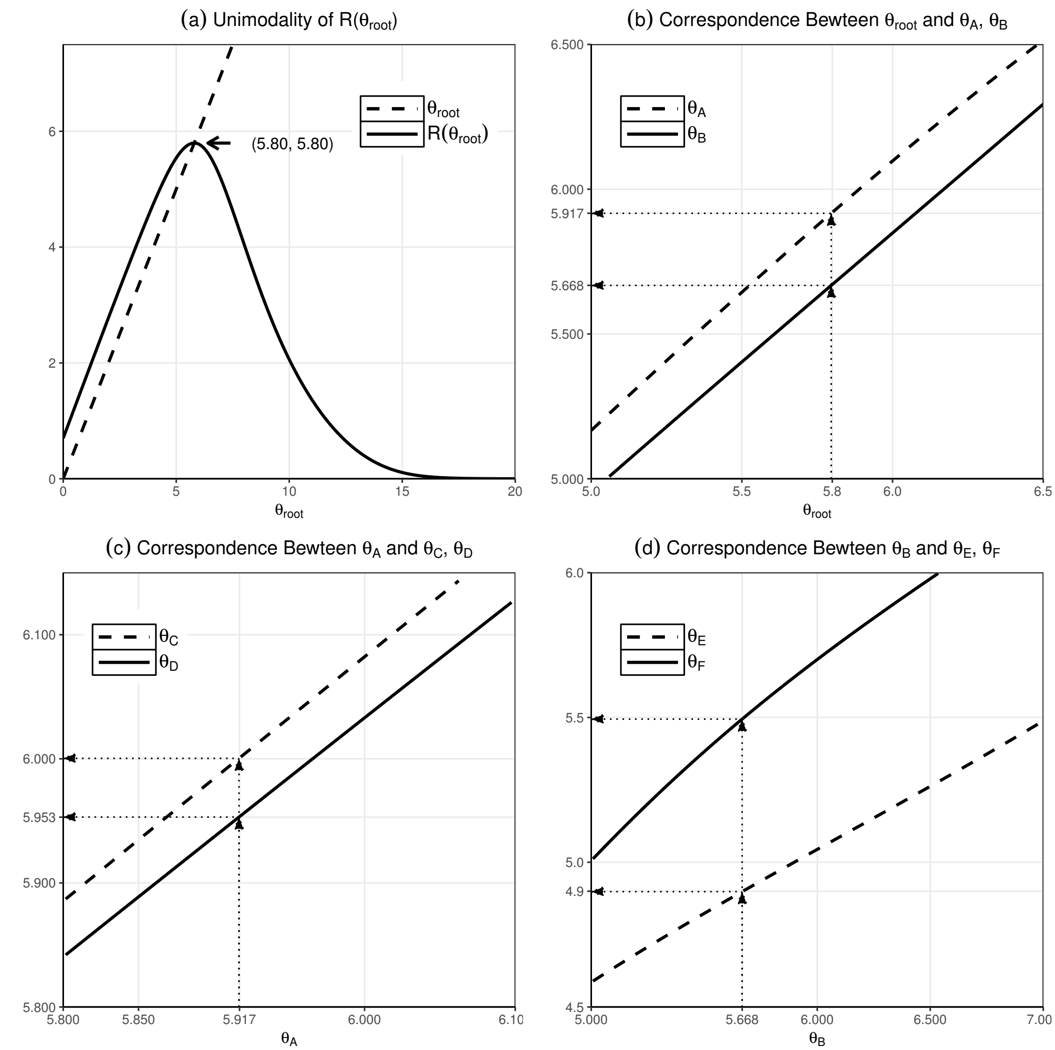

Figure 3.3: Pricing problem example under a three-level nested logit model

Chapter 4

Under Tree Logit Model: Joint Assortment and Price Optimization

4.1 Literature Review

We present the relevant literature based on assortment and/or pricing problems under variants of the multinomial and nested logit models. The assortment optimization problem is referred to as the problem where firms cannot control prices of products but are able to make assortment decisions to maximize their profit. The multinomial logit model is based on random utility maximization theory [34] and is successful in describing consumer choosing behavior among products with different attributes, such as price, brand and quality. Under the multinomial logit model, [42] show the structure of optimal assortment while considering inventory cost. [44] consider a no-purchase option in the multinomial logit model and show the optimal assortment policy is to include the set of products that are sorted by their revenues. [39] develop efficient algorithms to both static and dynamic assortment optimization problem. [47] consider capacity constraints on the offered assortment under the general attraction model including the multinomial logit model as its special case. [7], [41] and [23] study the assortment optimization problem assuming consumers choosing decisions follow the mixed multinomial logit model, which is able to segment consumers into groups and allows different choosing behavior in different groups. [17] consider the assortment problems under generalizations of the multinomial logit model and provide insight on the network revenue management problem.

The multinomial logit model suffers from the independence of irrelevant alternatives (IIA); see examples in [34] and [5]. The nested logit model, which is first proposed by [4], succeeds in avoiding IIA and draws attention in assortment optimization problems recently. When consumers choose according to the nested logit model, [38] study the assortment optimization problem with capacity constraints across different nests and develop polynomial-time approximation algorithm. [29] identify structural conditions for the optimal assortment and show a heuristic algorithm to compute the optimal solution. [11] study variants of assortment optimization problems, including considering the existence of no-purchase options in all the nests and the case where dissimilarity parameter exceeding one. The linear program formulated by them shows the ability to reduce the entire searching space to an acceptable size. [16] use a dynamic programming approach to solve the capacitated assortment optimization problems with cardinality and space constraints across nests. Although the nested logit model avoids the IIA property of the multinomial logit model, it still has IIA property within each nest. A tractable way is to generalize the two-level nested logit model to multistage tree logit model to alleviate the IIA property to the minimum extent. [30] and [50] study the unconstrained and constrained assortment optimization under the multilevel tree logit model, respectively. However, their approach do not generalize to the joint capacitated assortment and price optimization problem. Most of the research listed above focus only on assortment decisions without considering pricing effects that influence demand and profit.

For the case of price optimization problem involving choice models, a firm's objective is to come up with an optimal pricing strategy to maximize their profit with a fixed offered assortment by assuming that consumers choose according to a certain choice model. Under the multinomial logit model, [20] observe the fact that the profit function is not jointly concave with respect to price vectors. [3] and [21] discover the equal profit margin property of products at optimal prices. [43] and [15] both show that multinomial logit profit function is concave in terms of market share vectors instead of prices and there is a one-to-one relationship between the price and market share variable. This concavity property is further generalized to the nested logit model by [31] under the assumption that the price-sensitivity parameters are identical in each nest. [19] relax this assumption by adding mild restrictions of the dissimilarity parameter and price-sensitivity parameters and show the optimal prices can be obtained via maximizing a unimodal scalar function. Under the multilevel tree logit model, [30] work on an iterative algorithm that converges to a stationary point. In contrast, instead of working with price vectors, [22] assign an intermediate variable to each node and show there is an efficient way of getting the optimal solution. [50] generalize [19] by introducing multilevel adjusted markup and considering the no-purchase option to be associated with every node in the tree structure.

Most of the above study ignores the joint effect of assortment and pricing. For the joint assortment and price optimization problem, [9] study the mathematical properties of the optimal structure. [32] study the joint assortment and price optimization problem as well as considering inventory under a newsvendor model and derive the structural properties of optimal assortment and price. [48] works on the multinomial logit model with a general utility function and the offered assortments have cardinality constraints. He finds out that the adjusted markups are invariant for different products and further shows that the joint assortment and price optimization problem can be reduced to finding the fixed point of a single-dimensional function. [49] considers the search cost in the joint assortment and price optimization problem and shows that the optimal policy is to include the products with largest systematic utility. However, they assume the price-sensitivity parameters for products are all identical to one and this work is under the multinomial logit model. [6]

address the joint assortment and price competition in a competitive setting and derive the pure strategy equilibrium existence properties.

Under the nested logit model that alleviates the IIA property suffered by the multinomial logit model, [27] consider the joint assortment and price optimization problem with a type or brand primary choice model in both centralized and decentralized regime. They derive the properties and competitive equilibrium of joint optimal solution. [18] construct dummy products with different price levels and develop a linear program to find the joint optimal solution under the nested logit model. They do not assume any parametric relationship between the preference weight of products and its price, but their approach is not flexible in two folds. Firstly, for the joint assortment optimization problem, their method cannot consider constraints on the offered assortment; secondly, if the assortment is given and cannot be changed by the firm, their approach fails to work. However, our approach resolves these two limitations and generalize the nested logit model to the multilevel tree logit model. As a variant of our problem setting, our method turns out to work well when there is no parametric relationship between the preference weight and price of the product under the tree logit model. [12] study pricing problem under the nested logit model with a quality consistent constraint and it can be extended to joint assortment and pricing problem. Unlike most of the pricing literature that assume price is continuous, both of [18] and [12] consider the prices as discrete variables and do not have a parametric relationship with preference weight. In practice, considering the pricing problem as discrete or continuous optimization both have retail applications. [37] develop a linear program as approximation methods to the joint assortment and price optimization with price bounds under the nested logit model. [24] use a nonparametric choice modeling method to consider the joint assortment and price optimization problem and develop an expectation maximization (EM) algorithm to fit the model.

4.2 Main Results and Contributions

We summarize our main results and contributions as follows:

1. In this paper, we formulate the joint capacitated assortment and price optimization problem as a bi-level program where the inner problem is continuous price optimization and the outer problem is discrete optimization over assortment decisions. We show that the inner price optimization problem has a fixed point representation by introducing a scalar named as *node-specific adjusted markup*, which can be viewed as an important bridge connecting capacitated assortment and price optimization jointly.

2. For the joint capacitated assortment and price optimization problem, the consumer choice structure that we consider is a multilevel tree with N products. The cardinality or space constraints are imposed on the nonleaf nodes separately in the second last level. Our main result of this essay is that the joint optimal solution and a 2-approximate solution can be obtained in $O(GN \log G)$ time for the problem with cardinality and space constraints, respectively, where G is the number of grid points for each node in the tree structure.

Furthermore, under mild conditions on selecting grid points, our joint optimization algorithm can be further refined to run in $O(GN \log K)$ where K is the maximum number of children nodes that a nonleaf node can have in the tree structure. It is more efficient since K is always less than G in general. Surprisingly, the computational complexity is irrelevant of the number of levels in our tree structure. [30] studies the uncapacitated assortment optimization problem with fixed prices under d-level nested logit model, the algorithm of which runs in $O(dN \log N)$ time that is sensitive to the number of levels d. It is noticeable to find that our algorithm for joint optimization problem has the similar scale in terms of complexity compared to the algorithm for uncapacitated assortment-only problem in [30]. We show that the bi-level joint optimization problem can be decomposed to a single dimensional optimization problem over a scalar, which is tractable by our efficient algorithm.

3. To the best of our knowledge, we are the first to study joint capacitated assortment and price optimization problem under the multilevel tree logit model under both cardinality and space constraints. Many existing literature regarding the joint optimization problem under the multinomial logit model or the nested logit model turn out to be the special case of ours. For the earlier works on the joint capacitated assortment and price optimization, [48] shows that the joint optimization problem with cardinality constraints on assortments can be solved by finding a fixed of a one dimensional objective function when the consumers choose under the multinomial logit model. This problem is tractable based on the linearity nature of assortment optimization under the multinomial logit model that is considered in [39]. However, the nonlinearity arises when it comes to the nested logit model that has two choice stages. [18] consider the nested logit model and focus the possible prices of products only on a prespecified grid and do not assume any parametric relationship between the preference weight and price of products, while recent pricing literature under multistage choice structure does not work in this case; see [30] and [22]. Moreover, [18] do not consider constraints on the offered assortment in the joint assortment and price optimization problem.

4. Our approach generalizes earlier works not only to the tree logit model with arbitrary structure and an arbitrary number of products, but to the problem where there is a cardinality or space limitation on feasible assortments as well. The results in this essay is flexible and one step forward compared to the earlier and recent literature, we provide a systematic and complete solution to the following three problems: 1) Capacitated assortment optimization with fixed prices; 2) Price optimization with fixed assortment; 3) Joint assortment and price optimization problem under the multistage tree logit model, including the multinomial logit model and the nested logit model as special cases.

Organization

The organization of the essay is as follows. In the next section, we present the tree logit model and problem formulation of the joint capacitated assortment and price optimization. Section 4.4 considers joint optimization problem under cardinality constraints. We introduce the bi-level optimization program, solve the inner price optimization problem provided the assortment is fixed and introduce an intermediate variable that connects to the outer assortment optimization problem. Then we show this problem can be solved by an efficient algorithm. In Section 4.5, we study the joint optimization problem under space constraints and show a 2-approximate solution can be found in a tractable way by proposing an efficient algorithm. In Section 4.6, we illustrate our algorithm by testing a numerical example under a three-level tree structure.

4.3 Model and Problem Formulation

In this section, we first introduce the tree logit model and assumptions on the model parameters, such as the dissimilarity parameters and price-sensitivity parameters. Then we are ready to formulate the joint capacitated assortment and price optimization problem as a bi-level optimization program with price optimization as the inner problem and assortment optimization as the outer problem.

Tree Logit Model

We use an m-level tree structure, which is denoted as Tree $= (V, E)$ with vertices V and edges E, to describe the consumer choosing process under the tree logit model. In this tree structure, K is the maximum number of children nodes that a nonleaf node can have. Faced with various products, assume that a consumer has m specific requirements, such as product category, brands and rating, for the desired product that she wants to buy, which can be translated into the m-stage decision-making process described as follows. In the m-level decision tree, the consumer starts from the root node in level 0, and chooses whether to leave without purchasing any products, which corresponds to the no-purchase option in level 1, or to select a subset of all the products that satisfy her first requirement. The above choosing process can be viewed as moving from root to one of its children nodes. If we assume that she does not choose the no-purchase option, then she is now in level 1 and about to select another subset of products satisfying both of her first two requirements. As she is moving deeper down in the tree, she narrows down her set of desired products until she reaches a leaf node corresponding to an actual product that meets all her m requirements, which completes the choosing process. Parent(i) and Children(i) are used to denote the parent node and the set of children nodes of node i, respectively. Without loss of generality, we define Parent(root) $= \varnothing$ and Children$(\mathcal{L}) = \varnothing$ if node $\mathcal{L}$ is a leaf node. For any node $i \in V$ in level l, let N_i^l be the subset of products satisfying the consumer's first l requirements of the products. For notational brevity, we omit the superscript l of N_i^l throughout the paper. From the above choosing process, we can see that N_i is consist of products or leaf nodes sharing the same ancestor node i. Thus N_i can be defined recursively as follows, $N_i = \{i\}$ if node i is a leaf node; and $N_i = \bigcup_{j \in \text{Children}(i)} N_j$ for nonleaf node i that is in level $0 \leqslant l \leqslant m-1$. Using this notation, N_{root} represents the set of all the candidate products to be chosen from. Let N be the total number of candidate products, then the size of N_{root} is $N + 1$ since N_{root} is consist of N products and one no-purchase option.

We proceed to introduce the assortment that is offered to the consumer. In practice, it is always not optimal to offer all the product that a firm has to consumers because there is a display space limitation or it is not better off offering less attractive products with high prices. A decision that the firm always make is to choose a subset of products, or an assortment that is denoted as S_{root}, from N_{root} in order to achieve its objective, such as maximizing expected profit in our problem setting. Similarly, for any node $i \in V$, the assortment $S_i \subseteq N_i$ of node i is also defined recursively: if node i is a leaf node, $S_i = \{i\}$ or $\varnothing$, implying that we choose whether to include product i in the offered assortment or not; on the other hand, if node i is a nonleaf node, then $S_i = \bigcup_{j \in \text{Children(i)}} S_j$ that is an assortment of the subtree rooted at node i. We also assign a price vector $\mathbf{P}_i(S_i) \in \mathbb{R}^{|S_i|}$ to node $i \in V$, which includes the prices of products in assortment S_i. For instance, if node i is a leaf node, $\mathbf{P}_i(i) = (p_i)$ when $S_i = \{i\}$ and p_i is the price of product i or $\mathbf{P}_i(i) = \varnothing$ when $S_i = \varnothing$; for the root node, $\mathbf{P}_{\text{root}}(S_{\text{root}})$ contains the prices of all the products in assortment S_{root}. For notational purpose, we use $\mathbf{P}_i$ instead of $\mathbf{P}_i(S_i)$ throughout the paper.

As discussed in the above paragraph, the size of feasible assortment that a firm offers cannot be too large due to the display space limitation. In this paper, we work with cardinality/space constraints limiting the total number/space of products that are associated with the nodes in level $m-1$. For ease of presentation, we define those nodes in level $m-1$ as *basic* nodes, the set of which is denoted as $\mathscr{B}$. Then for the basic node $j \in \mathscr{B}$, the set of feasible assortments under cardinality constraints is defined as $\Im_j = \{S_j : S_j = \bigcup_{k \in \text{Children(j)}} S_k, |S_j| \leqslant \mathbb{C}_j\}$ where the prespecified $\mathbb{C}_j$ is a cardinality limitation on assortment S_j, implying the maximum number of products that node j can have should not exceed $\mathbb{C}_j$. The set of feasible assortments under space constraints is accordingly defined as $\Im'_j = \{S_j : S_j = \bigcup_{k \in \text{Children(j)}} S_k, \sum_{k \in \text{Children}(j)} w_k \leqslant \mathbb{S}_j\}$ where the prespecified $\mathbb{S}_j$ is a space limitation on assortment S_j and w_k is the space consumption of product k. For node i that is neither a leaf node nor a basic node, the feasible set of assortments is defined recursively as $\Im_i = \bigtimes_{j \in \text{Children}(i)} \Im_j$, where $\times$ stands for the Cartesian product. With this notation, the feasible set of assortments for node i is the Cartesian product of all the feasible sets of assortments for i's children nodes that are indexed by j.

We use the upside-down tree in Figure 4.1 as an example to illustrate our notational system. In the three-level tree structure, there are 11 leaf nodes including 10 products $\{g, h, ..., p\}$ in level 3 and a no-purchase option in level 1. For the node a, we have $\text{Parent(a)} = \text{root}$, $\text{Children(a)} = \{c, d\}$, $N_a = N_c \bigcup N_d = \{g, h, i\} \bigcup \{j, k, l\} = \{g, h, i, j, k, l\}$. We also have $N_g = \{g\}$, $\text{Children}(g) = \varnothing$ and $\text{Parent(root)} = \varnothing$. Let the cardinality constraints on basic nodes c and d be $\Im_c = \{S_c : S_c = \bigcup_{k \in \text{Children}(c)} S_k, |S_c| \leqslant 1\} = \{\varnothing, g, h, i\}$ and $\Im_d = \{S_d : S_d = \bigcup_{k \in \text{Children}(d)} S_k, |S_d| \leqslant 1\} = \{\varnothing, j, k, l\}$, respectively. Then the feasible set of assortments for node a is $\Im_a = \Im_c \times \Im_d = \{\varnothing, \{g\}, \{h\}, \{i\}, \{j\}, \{k\}, \{l\}, \{g, j\}, \{g, k\}, \{g, l\}, \{h, j\}, \{h, k\}, \{h, l\}, \{i, j\}, \{i, k\}, \{i, l\}\}$ and a feasible assortment for node a is $S_a = \{g, k\} \subseteq \Im_a$ with price vector $\mathbf{P}_a = (p_g, p_k)$. The size of $\Im_a$ is 15 that is very large for a system only has 10 products. If we impose the same cardinality constraints on nodes e and f, a feasible assortment for node root is $S_{\text{root}} = \{g, k, n\}$, in which $S_a = \{g, k\}$, $S_d = \{k\}$, $S_e = \{n\}$, $S_j = \varnothing$ and $S_n = \{n\}$.

The price vector of S_{root} is $\mathbf{P}_{\text{root}} = (p_g, p_k, p_n)$. From this toy example, we can still get a sense that the size of $\Im_{\text{root}} = \Im_a \times \Im_b$ is very large even under simple tree structure with small number of products, thus selecting an optimal assortment is not a trivial task and brute force clearly is not an option. Similarly, for space constraints, if we let $w_g = w_j = 1$, $w_h = w_k = 2$, $w_i = w_l = 3$ and $\mathbb{S}_c = \mathbb{S}_d = 2$, then the feasible set of assortments for node a is $\Im'_a = \{\emptyset, \{g\}, \{j\}, \{h\}, \{k\}, \{g, j\}\}$.

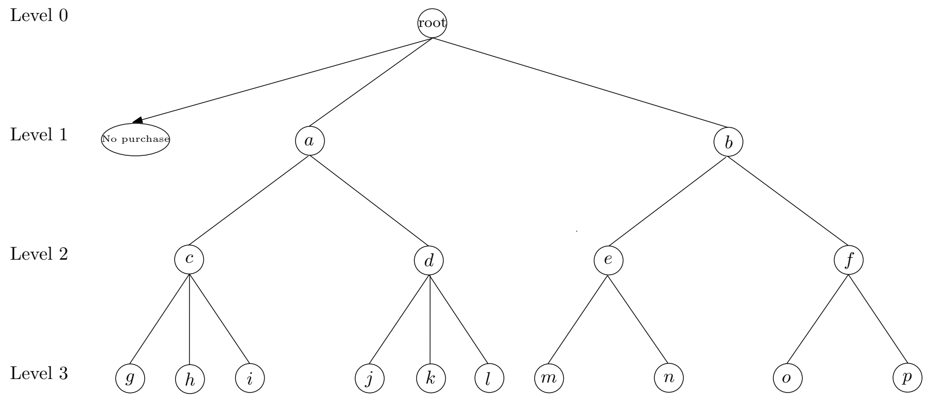

Figure 4.1: The tree logit model with a 3-level tree structure

To measure the attractiveness of assortment S_i with its corresponding price vector $\mathbf{P}_i$, we assign a preference weight $V_i(S_i, \mathbf{P}_i)$ for node $i \in V$ as a function of S_i and $\mathbf{P}_i$. Specifically, if node i is a leaf node and is not a no-purchase option, its preference weight is defined as

$$V_i(S_i, \mathbf{P}_i) = \exp(\alpha_i - \beta_i p_i) * \mathbf{1}(S_i \neq \emptyset),$$

where α_i is price-independent deterministic utility and β_i is the price-sensitivity parameter of product i. If the assortment S_i for leaf node i is empty, the indicator function $\mathbf{1}(\cdot)$ is zero, yielding the preference weight $V_i(\emptyset, \mathbf{P}_i)$ to be zero; if S_i is not empty, then $V_i(S_i.\mathbf{P}_i) = \exp(\alpha_i - \beta_i p_i)$ would be strictly positive. Moreover, for the no-purchase option in level 1 as a child node of the root node, let $V_{\text{root}_0} > 0$ be its preference weight. Then for a nonleaf node $i \in V$, its preference weight can be calculated recursively as

$$V_i(S_i, \mathbf{P}_i) = \left(\sum_{j \in \text{Children}(i)} V_j(S_j, \mathbf{P}_j) \right)^{\gamma_i},$$

where $\gamma_i \in (0, 1]$ is the dissimilarity parameter of nonleaf node i, S_i is the assortment associated with node i and $\mathbf{P}_i$ is its corresponding price vector. The preference weight of root node does not influence the objective profit function, which will be shown later of this section, so we set $\gamma_{\text{root}} = 0$ without loss of generality, which yields $V_{\text{root}}(S_{\text{root}}) = 1$ for arbitrary assortment S_{root}.

The restriction on γ_i guarantees that this tree logit model satisfies the random utility maximization theory [34]. The dissimilarity parameter γ_i for node i measures the dissimilarities between its children nodes. The larger γ_i is, the less similar the products of i's children nodes are, or the less positively related the random utilities of these products are. Specifically, when γ_i equals to one, the tree structure is degenerated by removing node i and connecting i's children nodes and parent node directly. If the dissimilarity parameter exceeds one, the two-level nested logit model is still consistent with the random utility maximization theory under some conditions; see [33] and [45]. Recently, [11] and [19] study the assortment optimization and price optimization problem under the two-level nested logit model with dissimilarity parameter exceeding one, respectively. In this paper, we impose the restriction that $\gamma_i \in (0, 1]$ for all $i \in V$, since the assortment optimization problem becomes NP-hard even for uncapaciated assortment with fixed prices under the two-level nested logit model. As in [22], we also make the following assumption on the price-sensitivity parameters and dissimilarity parameters to guarantee that there is a unique optimal pricing solution.

Assumption 3. *For any nonleaf node i, define*

$$\underline{B}_i = \begin{cases} \min_{j \in \text{Children}(i)} \{\beta_j\} & i \text{ is a basic node} \\ \min_{j \in \text{Children}(i)} \{\underline{B}_j \gamma_j\} & \text{o.w.} \end{cases}$$

and

$$\overline{B}_i = \begin{cases} \max_{j \in \text{Children}(i)} \{\beta_j\} & i \text{ is a basic node} \\ \max_{j \in \text{Children}(i)} \{\frac{\gamma_j^2 \overline{B}_j}{1-(1-\gamma_j)\overline{B}_j/\underline{B}_j}\} & \text{o.w.} \end{cases}$$

We assume that

$$\frac{\overline{B}_i}{\underline{B}_i} < \frac{1}{1-\gamma_i}.$$

Problem Formulation

The tree logit model is essentially a probability based choice model, with the notational system ready, we show how the choice probabilities of products are calculated and how they are related to the objective profit function as follows. If assortment S_i of node i is not empty, the conditional probability of choosing S_j where $j \in \text{Children}(i)$ can be formulated as

$$Q(S_j, \mathbf{P}_j | S_i, \mathbf{P}_i) = \frac{V_j(S_j, \mathbf{P}_j)}{\sum_{j \in \text{Children}(i)} V_j(S_j, \mathbf{P}_j)}.$$

If assortment S_i is empty, then we define $Q(S_j, \mathbf{P}_j | S_i, \mathbf{P}_i) = 0/0 = 0$ as the probability of choosing from an empty assortment is zero. We use $R_i(S_i, \mathbf{P}_i)$ to denote the profit for node $i \in V$ with assortment S_i and price vector $\mathbf{P}_i$. Particularly, if node i is a leaf node but not a no-purchase option, then $R_i(S_i, \mathbf{P}_i) = (p_i - c_i) * \mathbf{1}(S_i \neq \emptyset)$ where c_i is the cost of product

i and the indicator function $\mathbf{1}(\cdot)$ makes the profit to be zero if we do not offer product i in assortment S_i. For the no purchase, we define $R_{\text{root}_0}(S_{\text{root}_0}, P_{\text{root}_0}) = 0$ since we cannot get any profit if the consumer decides not to purchase. If node i is a nonleaf node, then the expected profit is defined recursively as follows

$$\begin{aligned} R_i(S_i, \mathbf{P}_i) &= \sum_{j \in \text{Children}(i)} Q(S_j, \mathbf{P}_j | S_i, \mathbf{P}_i) * R_j(S_j, \mathbf{P}_j) \\ &= \frac{\sum_{j \in \text{Children}(i)} V_j(S_j, \mathbf{P}_j) R_j(S_j, \mathbf{P}_j)}{\sum_{j \in \text{Children}(i)} V_j(S_j, \mathbf{P}_j)}. \end{aligned}$$

Thus for assortment S_{root} with price vector $\mathbf{P}_{\text{root}}$, the objective profit that is generated from this system is $R_{\text{root}}(S_{\text{root}}, \mathbf{P}_{\text{root}})$. Since the objective profit function $R_{\text{root}}(S_{\text{root}}, \mathbf{P}_{\text{root}})$ does not include the term V_{root}, we can set $\gamma_{\text{root}} = 0$ to make $V_{\text{root}} = 1$ without changing the objective.

The joint capaciated assortment and price optimization problem under the tree logit model can be formulated as the following bi-level optimization program

$$Z^* = \max_{S_{\text{root}} \subseteq \Im_{\text{root}}} \max_{\mathbf{P}_{\text{root}} \in \mathbb{R}_{\geqslant 0}^{|S_{\text{root}}|}} R_{\text{root}}(S_{\text{root}}, \mathbf{P}_{\text{root}}), \tag{4.1}$$

where Z^* denotes the maximum expected profit that we can obtain per consumer and $\Im_{\text{root}}$ is the collection of feasible assortments that satisfy the constraints, which can be either cardinality or space constraints. In this paper, we show that problem (4.1) is tractable by building a bridge that connects the inner price optimization problem and the outer assortment optimization problem.

4.4 Joint Optimization Under Cardinality Constraints

In this section, we consider the joint capacitated assortment and price optimization problem under the cardinality constraints. Thus, we have $\Im_j = \{S_j : S_j = \bigcup_{k \in \text{Children(j)}} S_k, |S_j| \leqslant \mathbb{C}_j\}$ for $\forall j \in \mathscr{B}$ throughout the section. First, we show the joint optimization problem can be decomposed to solving joint subproblem that are define on the nonleaf node by assuming we have already known both optimal assortment and optimal prices. Second, we solve the inner pricing problem of the joint subproblem. The remaining joint subproblem with solved inner pricing problem is referred to as assortment subproblem. Third, the assortment subproblem can be we reformulated by optimizing over a scalar. Last, we propose a polynomial-time approach to solve the joint optimization problem under cardinality constraints.

Problem Decomposition

We consider solving problem (4.1) by decomposing it into joint subproblem on every nonleaf node, which is inspired by the algorithm of uncapacitated assortment optimization problems;

see [30]. However, their approach can deal with the uncapacitated assortment optimization only and cannot be generalized to solve joint optimization over both capacitated assortment and prices. Let S_i^* and $\mathbf{P}_i^*$ be the optimal assortment and the optimal price vector for node $i \in V$, respectively. Then define the scalar $e_i^* = \gamma_i e_h^* + (1-\gamma_i)R_i(S_i^*, \mathbf{P}_i^*)$ where h is the parent node of i, and let $e_{\text{Parent(root)}}^* = 0$ for the boundary condition. By noting that $\gamma_{\text{root}} = 0$, we have $e_{\text{root}}^* = R_{\text{root}}(S_{\text{root}}^*, \mathbf{P}_{\text{root}}^*)$. If the optimal assortment S_{root}^* and optimal prices $\mathbf{P}_{\text{root}}^*$ are given, then $R_i(S_i^*, \mathbf{P}_i^*)$ for $i \in V$ can be calculated, thus all the scalars e_i^* can also be obtained in a top-down manner. For each nonleaf node i, the *joint subproblem* is defined as follows

$$\max_{S_i \subseteq \Im_i} \max_{\mathbf{P}_i \in \mathbb{R}_{\geqslant 0}^{|S_i|}} V_i(S_i, \mathbf{P}_i)(R_i(S_i, \mathbf{P}_i) - e_h^*). \tag{4.2}$$

Note that problem (4.2) at root node is identical to problem (4.1).

Compared to the assortment optimization problem, the joint optimization problem requires making decisions of both assortment and prices. In the capacitated assortment optimization problem under the two-level nested logit model [18] and uncapaciated assortment optimization problem under the d-level nested logit model [30], the optimal assortment S_i^* of a nonleaf node $i \in V$ can be an empty set if $R_i^*(S_i^*)$ is less than the scalar e_h^*. However, for the "joint" optimization problem, the optimal assortment S_i^* can never be empty since we have more control over both assortment and prices. The following lemma shows this difference between joint and assortment-only optimization problem.

Lemma 13. (Joint Subproblem and Optimal Assortment) *In the joint capacitated assortment and price optimization problem, for a nonleaf node $i \in V$ and its parent node h, the optimal assortment and price vector (S_i^*, $\mathbf{P}_i^*$) and optimal solution $(\hat{S}_i, \hat{P}_i)$ to joint subproblem at node i satisfy:*

1. $R_i(S_i^*, \mathbf{P}_i^*) > e_h^* \geqslant 0$;
2. S_i^* *is a nonempty set and* $\mathbf{P}_i^*$ *is a nonzero vector;*
3. $(S_i^*, \mathbf{P}_i^*) = (\hat{S}_i, \hat{\mathbf{P}}_i)$ *and* $\bigcup_{i \in \text{Children}(h)} (\hat{S}_i, \hat{\mathbf{P}}_i)$ *is optimal to joint subproblem at node h.*

With a slight abuse of notation, we denote $(\hat{S}_h, \hat{\mathbf{P}}_h) = \bigcup_{i \in \text{Children}(h)} (\hat{S}_i, \hat{\mathbf{P}}_i)$ throughout the essay for ease of presentation, where $\hat{S}_h = \bigcup_{i \in \text{Children}(h)} \hat{S}_i$ and $\hat{\mathbf{P}}_h = (\hat{\mathbf{P}}_{i_1}, \hat{\mathbf{P}}_{i_2}, ..., \hat{\mathbf{P}}_{i_n})$ where $i_1, i_2, ..., i_n \in \text{Children}(h)$. The proof of this lemma can be found in Online Appendix C.2.

From Lemma 13, it immediately follows that for node i in level l, both e_i^* and $R_i(S_i^*, \mathbf{P}_i^*)$ decrease as l becomes smaller, i.e. $e_i^* > e_{\text{Parent}(i)}^*$ and $R_i(S_i^*, \mathbf{P}_i^*) > R_{\text{Parent}(i)}(S_{\text{Parent}(i)}^*, \mathbf{P}_{\text{Parent}(i)}^*)$. We make an observation that the joint subproblem at node $i \in V$ is highly nonlinear in S_i even if the optimal price vector $\mathbf{P}_i^*$ is given. Therefore, we propose an alternative formulation of joint subproblem at node $i \in V$, which is referred to as *basic* joint subproblem, as follows

$$\max_{S_i \subseteq \Im_i} \max_{\mathbf{P}_i \in \mathbb{R}_{\geqslant 0}^{|S_i|}} V_i(S_i, \mathbf{P}_i)^{1/\gamma_i}(R_i(S_i, \mathbf{P}_i) - e_i^*). \tag{4.3}$$

For root node specifically, we define $V_{\text{root}}(R_{\text{root}}, \mathbf{P}_{\text{root}})^{1/\gamma_{\text{root}}} = \sum_{j \in \text{Children}(i)} V_j(R_j, \mathbf{P}_j)$, and the optimal objective value for problem (4.3) at root node is zero since $e^*_{\text{root}} = R_{\text{root}}(S^*_{\text{root}}, \mathbf{P}^*_{\text{root}})$. Problem (4.3) at a basic node turns out to be tractable, which we would show later in this section. The remaining question is what the relationship is between joint subproblem (4.2) and basic joint subproblem (4.3). We want to show the optimal solution to problem (4.3) is also optimal to problem (4.2), and the following example under the two-level nested logit model can be used to illustrate the intuition.

Example 1. *(Joint Optimization Under the two-level Nested Logit Model) Under the two-level nested logit model, where the disjoint nests are indexed by j. The feasible set of assortment for node j is denoted as $\Im_j$. The joint subproblem (4.2) at root node is the global joint optimization problem*

$$\theta^*_{\text{root}} = \max_{S_{\text{root}} \subseteq \Im_{\text{root}}} \max_{\mathbf{P}_{\text{root}} \in \mathbb{R}^{|S_{\text{root}}|}_{\geqslant 0}} R_{\text{root}}(S_{\text{root}}, \mathbf{P}_{\text{root}}),$$

which, according to the definition of $R_{\text{root}}(S_{\text{root}}, \mathbf{P}_{\text{root}})$, can be rewritten as

$$v_0 \theta^*_{\text{root}} = \max_{S_{\text{root}} \subseteq \Im_{\text{root}}} \max_{\mathbf{P}_{\text{root}} \in \mathbb{R}^{|S_{\text{root}}|}_{\geqslant 0}} \sum_{j \in \text{Childre(root)}} V_j(S_j, \mathbf{P}_j)(R_j(S_j, \mathbf{P}_j) - \theta^*_{\text{root}})$$

Move $v_0 \theta^*_{\text{root}}$ to the right hand side of the above equation, it becomes

$$0 = \max_{S_{\text{root}} \subseteq \Im_{\text{root}}} \max_{\mathbf{P}_{\text{root}} \in \mathbb{R}^{|S_{\text{root}}|}_{\geqslant 0}} V_{\text{root}}(R_{\text{root}}, \mathbf{P}_{\text{root}})^{1/\gamma_{\text{root}}}(R_{\text{root}}(S_{\text{root}}, \mathbf{P}_{\text{root}}) - \theta^*_{\text{root}}).$$

The right hand side of the above equation is the basic joint optimization problem (4.3) at root node. Therefore, the optimal solution to problem (4.3) is also optimal to problem (4.2) and they share the same optimal solution.

The following lemma follows the above discussion.

Lemma 14. *For any nonleaf node $i \in V$, the optimal solution to basic joint subproblem (4.3) is also optimal to joint subproblem (4.2).*

Even if e^*_h is known, we still cannot solve joint subproblem (4.2) at basic node i due to the nonlinearity of S_i in the preference weight $V_i(S_i, \mathbf{P}_i)$. However, by Lemma 14, we turn to deal with a more tractable basic joint subproblem (4.3). That is to say, if e^*_i is given $\forall i \in \mathscr{B}$, then the optimal solution to problem (4.1) can be obtained by taking the union of all the optimal solutions to problem (4.3) at node i by Lemma 14 and using the third item of Lemma 13 repeatedly.

Assortment Subproblem

In this subsection, we still assume the scalar e_i^* is given for any nonleaf node i and focus on building connections between the inner pricing and outer assortment optimization problem of the joint subproblem (4.2). For a given assortment $S_i \subseteq \Im_i$, the inner pricing problem of (4.2) at node i is

$$\max_{\mathbf{P}_i \in \mathbb{R}_{\geqslant 0}^{|S_i|}} V_i(S_i, \mathbf{P}_i)(R_i(S_i, \mathbf{P}_i) - e_h^*). \tag{4.4}$$

Problem (4.4) is a $|S_i|$-dimensional continuous optimization problem. The objective function is non-concave with respect to the price vector even under the 2-level tree logit model. We introduce an intermediate variable for node $i \in V$ to reduce the dimension of this problem, which is referred to as the *node-specific adjusted markup* and is defined as follows.

Definition 1. (Node-specific Adjusted Markup) *For nonleaf node $i \in V$, given assortment S_i, the node-specific adjusted markup for node i is defined as*

$$\theta_i = \theta_j - \omega_j(S_j, \theta_j),$$

where $j \in$ Children(i) and $\omega_j(S_j, \theta_j)$ is recursively defined as follows

$$\omega_j(S_j, \theta_j) = (\frac{1}{\gamma_j} - 1) \sum_{k \in \text{Children}(j)} \frac{Q(S_k, \theta_k | S_j, \theta_j)}{1 - \gamma_k} \omega_k(S_k, \theta_k).$$

For the boundary condition, if leaf node i is in level m, we set $\theta_i = p_i - c_i$, $\omega_i = 1/\beta_i$ and $\gamma_i = 0$; for the root *node, we define $\omega_{\text{root}}(S_{\text{root}}, \theta_{\text{root}}) = \theta_{\text{root}}$.*

By applying the first order condition repeatedly, the dimension of problem (4.4) can be reduced to one in the end. If the objective function has nice properties, then solving problem (4.4) with respect to the node-specific adjusted markup is not as a hard task as solving it with respect to the price vector, which in turn makes it possible to solve joint subproblem (4.2). Assumption 3 about input parameters, which are price-sensitivity parameters and dissimilarity parameters, is the sufficient condition that guarantees the uniqueness of the optimal solution to problem (4.4). At optimality conditions, the following lemma shows that the properties of node-specific adjusted markup.

Lemma 15. *Given assortment S_i for nonleaf node $i \in V$ in level d. Suppose node a, which is in level l $(d < l \leqslant m-1)$, is a descendant node of i, then for any nodes $b, b' \in$ Children(a), we have that $\theta_a = \theta_b - \omega_b(S_b, \theta_b)$ and $\theta'_a = \theta_{b'} - \omega_{b'}(S_{b'}, \theta_{b'})$ are equivalent under the optimality condition of problem (4.4). Under Assumption 3, there exists a one-to-one increasing correspondence between the node-specific adjusted markup θ_a and θ_b.*

According to the above lemma, we know that there is a one-to-one correspondence between the node-specific adjusted markup of one node's and its parent node's. Moreover, for product r in the lowest level m, its node-specific adjusted markup θ_r equals to $p_r - c_r$, so θ_r

has a one-to-one correspondence with its price p_r. Then we can conclude that for every node $i \in V$, there is a one-to-one correspondence between the node-specific adjusted markup θ_i and its price vector $\mathbf{P}_i$. Thus for a given θ_i, the prices of its descendant products' can also be uniquely determined. Therefore, the $|S_i|$-dimensional price vector $\mathbf{P}_i$ in problem (4.4) can be replaced by the scalar θ_i. As in the Definition 1, $Q(S_k, \theta_k|S_j, \theta_j)$ can be used to represent the conditional probability instead of $Q(S_k, \mathbf{P}_k|S_j, \mathbf{P}_j)$. We make a crucial observation that the multidimensional inner pricing subproblem (4.4) given assortment S_i can be rewritten as an optimization problem with respect to a scalar, which is shown as follows

$$\max_{\theta_i \in \mathbb{R}} V_i(S_i, \theta_i)(R_i(S_i, \theta_i) - e_h^*), \tag{4.5}$$

This observation is recorded in the following proposition.

Proposition 10. *The inner pricing subproblem (4.4) for a given assortment S_i at a nonleaf node $i \in V$ is equivalent to the optimization problem (4.5) with respect to the node-specific adjusted markup θ_i that is a scalar.*

We continue to study problem (4.5) and show the findings in the following lemma by applying the first order condition.

Lemma 16. *At the optimality condition of problem (4.5), θ_i should satisfy:*

$$\theta_i = \gamma_i e_h^* + (1 - \gamma_i) R_i(S_i, \theta_i).$$

where θ_i is the node-specific adjusted markup for node i.

Lemma 16 is insightful in a way that problem (4.5) is solved by finding θ_i that satisfy the optimality condition if S_i and e_h^* are known. However, solving problem (4.5) is still a non-trivial task since there might be multiple θ_i satisfying the optimality condition. Fortunately, problem (4.5) at root node can be solved uniquely and efficiently.

Corollary 5. *If i is root node, then $R_i(S_i, \theta_i)$ is strictly unimodal in θ_i and $R_i(S_i, \theta_i) = \theta_i$ at optimality for given S_i.*

In order to get the optimal solution to problem (4.5) at nodes besides the root in a tractable way, we aim for studying the optimality condition again by finding out what the scalar e_i^* essentially is. We define the *optimal node-specific adjusted markup* θ_i^* for node $i \in V$ as the optimal solution to inner pricing subproblem (4.5) given optimal assortment S_i^*: $\theta_i^* = \arg\max_{\theta_i \in \mathbb{R}} V_i(S_i^*, \theta_i)(R_i(S_i^*, \theta_i) - e_h^*)$.

Proposition 11. *For any node $i \in V$, the optimal node-specific adjusted markup θ_i^* is equivalent to the scalar e_i^*.*

Start from the optimality condition as shown in Lemma 4, the next proposition shows more of the structural relationship between θ_h^* and the expected profit $R_i(S_i, \theta_i)$ of node i given assortment S_i. This relationship can be viewed as the bridge that connects the inner pricing subproblem and the outer assortment optimization subproblem. How to exploit this property is crucial to solve the joint subproblem (4.2), which will be further discussed in this section.

Proposition 12. *$R_i(S_i, \theta_i)$ and θ_h^* should satisfy the following conditions at the optimality of problem (4.5):*

$$R_i(S_i, \theta_i) = \theta_h^* + \frac{1}{1-\gamma_i}\omega_i(S_i, \theta_i),$$

and

$$\theta_h^* = \theta_i - \omega_i(S_i, \theta_i).$$

Moreover, we have $R_i(S_i, \theta_i) > \theta_h^*$.

The fact that $R_i(S_i, \theta_i) > \theta_h^*$ is also consistent with the first item in Lemma 13. The next corollary eliminates the concerns of possibility that problem (4.5) has multiple solutions.

Corollary 6. *For a given θ_h^* and S_i, optimal solution to problem (4.5) can be uniquely determined by $\theta_h^* = \theta_i - \omega_i(S_i, \theta_i)$.*

We are ready to rewrite the bi-level joint subproblem (4.2) as an optimization program with respect to assortment variable only since inner pricing subproblem is completely solved by earlier discussion. We record this finding in the following theorem.

Theorem 7. (Assortment Subproblem) *The equivalent formulation of joint subproblem (4.2) is defined as follows:*

$$\begin{aligned} \max_{S_i \subseteq \Im_i} \quad & \frac{V_i(S_i, \theta_i)\omega_i(S_i, \theta_i)}{1-\gamma_i} \\ s.t. \quad & \theta_h^* = \theta_i - \omega_i(S_i, \theta_i), \end{aligned} \tag{4.6}$$

which is referred to as assortment subproblem at node i.

Note that for the root node, $\omega_{\text{root}}(S_{\text{root}}, \theta_{\text{root}}) = \theta_{\text{root}} = R_{\text{root}}(S_{\text{root}}, \theta_{\text{root}})$ at the optimality condition of problem (4.6). In the assortment subproblem (4.6), θ_i is essentially a function of S_i for a given θ_h^* since θ_i can be uniquely computed via the constraint in problem (4.6) by Corollary 6. However, the scalar θ_h^*, which is the optimal solution to the inner pricing subproblem (4.5) at node h given the optimal assortment S_h^*, cannot be identified before we obtain θ_{root}^* by solving $R(S_{\text{root}}^*, \theta_{\text{root}}) = \theta_{\text{root}}$; see [50] for numerical experiments.

Even if θ_h^* is provided, obtaining an analytical solution to problem (4.6) seems not possible due to the nonlinearity nature of the objective function. The optimal solution to problem (4.3) is also optimal to (4.2), since problem (4.3) is "linear" in S_i if a collection of candidate assortments for node $j \in \text{Children}(i)$ is given, we show the basic joint subproblem (4.3) can be reformulated as well in the next theorem.

Theorem 8. (Basic Assortment Subproblem) *The equivalent formulation of basic joint subproblem (4.3) is defined as follows:*

$$\begin{aligned} \max_{S_i \subseteq \Im_i} \quad & \sum_{j \in \text{Children}(i)} \frac{V_j(S_j, \theta_j)\omega_j(S_j, \theta_j)}{1 - \gamma_j} \\ s.t. \quad & \theta_i^* = \theta_j - \omega_j(S_j, \theta_j), \end{aligned} \tag{4.7}$$

which is referred to as basic assortment subproblem at nonleaf node i.

As the discussion at the end of section 4.4, problem (4.7) at basic nodes is of special interest to study, which can be solved in an efficient approach by the following corollary.

Corollary 7. *For basic node $j \in \mathscr{B}$, problem (4.7) at j can be simplified as*

$$\max_{S_j \subseteq \Im_j} \sum_{k \in \text{Children}(j)} \frac{V_k(S_k, \theta_j^* + c_k + 1/\beta_k)}{\beta_k}, \tag{4.8}$$

where $S_j = \bigcup_{k \in \text{Children}(j)} S_k$ and $S_k \in \mathscr{A}_k = \{\{k\}, \varnothing\}$ that is the candidate collection of assortments for leaf node $k \in \text{Children}(j)$. Problem (4.8) can be solved within $O(N_j \log N_j)$ operations.

By corollary 7, the next corollary shows that the union of the optimal assortments to problem (4.8) at all the basic nodes is the global optimal assortment to problem (4.1), which can be proved by applying Lemma 13 repeatedly.

Corollary 8. *For all the basic node $j \in \mathscr{B}$, let $\hat{S}_j$ be optimal to problem (4.8), then assortment $\hat{S}_{\text{root}} = \bigcup_{j \in \mathscr{B}} \hat{S}_j$ is the optimal assortment to problem (4.1).*

Equivalent Formulation of Assortment Subproblem

However, θ_j^* for $j \in \mathscr{B}$ in problem (4.8) remains unknown. Even if there is a limited number of optimal solution to problem (4.8) for $\theta_j^* \in \mathbb{R}$, we still need to deal with problem (4.6) at upper-level nodes, which is intractable since it is nonlinear with respect to the assortment variable and the feasible searching space grows exponentially as number of products increases. In this subsection, for basic nodes, we first prove that a candidate collection of assortments that include the global optimal assortment to problem (4.1) has reasonable size and can be obtained efficiently. Second, we can get a feasible collection $\mathscr{A}_i$ including the global optimal assortment S_i^* for upper-level node i in a bottom-up manner. Then we show that $\mathscr{A}_i$ has the size of $O(N)$ by reformulating problem (4.6) in terms of optimizing over a scalar θ_i rather than the assortment variable S_i.

For any basic node $j \in \mathscr{B}$, we need to know θ_j^* in order to solve the basic assortment subproblem, which in turn requires knowing S_j^*. However, knowing S_j^* is impossible unless we could solve problem (4.1). We remark that the scalar θ_j^* in problem (4.8) can also be negative; see [19]. For any $\theta_j \in \mathbb{R}$, let $\tilde{S}_j(\theta_j)$ be the optimal solution to the following problem

$$\max_{S_j \subseteq \Im_j} \sum_{k \in \text{Children}(j)} \frac{V_k(S_k, \theta_j + c_j + 1/\beta_k)}{\beta_k}, \tag{4.9}$$

Define $\mathscr{A}_j = \{\tilde{S}_j(\theta_j) : \theta_j \in \mathbb{R}\}$, then we claim that $\mathscr{A}_j$ includes the global optimal assortment S_j^*, since $S_j^* = \tilde{S}_j(\theta_j^*)$ according to Corollary 7 and Theorem 8. The next lemma shows that at optimality, the cardinality constraint in problem (4.9) is binding.

Lemma 17. *For the optimal solution $\tilde{S}_j(\theta_j)$ to problem (4.9) with $\forall \theta_j \in \mathbb{R}$, we have $|\tilde{S}_j(\theta_j)| = \mathbb{C}_j$.*

Lemma 17 is useful when the cardinality limitation is larger than the number of products that a basic node has, since the optimal assortment of this basic node in this case is straightforward: include them all. If it applies to all the basic nodes, the joint optimal assortment is $S^*_{\text{root}} = N_{\text{root}}$, and optimal prices can be found via $R_{\text{root}}(N_{\text{root}}, \theta^*_{\text{root}}) = \theta^*_{\text{root}}$.

Next we solve problem (4.9) for $\theta_j \in \mathbb{R}$, the objective function of which is rewritten as $\sum_{k \in \text{Children}(j)} \exp(\tilde{\alpha}_k - \beta_k \theta_j)$ where $\tilde{\alpha}_k = \alpha_k - \beta_k c_k - \log(\beta_k) - 1$ since $V_k(S_k, \theta_j + c_k + 1/\beta_k) = \exp(\alpha_k - \beta_k * (\theta_j + c_k + 1/\beta_k))$ according to Definition 1. If we define $f_k(\theta_j) = \exp(h_k(\theta_j))$ where $h_k(\theta_j) = \tilde{\alpha}_k - \beta_k \theta_j$, then there is at most one intersection point of any two functions in $\{f_k(\theta_j) : k \in \text{Children}(j)\}$. Furthermore, the intersection point, if any, of two functions $f_{k_1}(\theta_j)$ and $f_{k_2}(\theta_j)$ is same as the intersection point of the two linear functions $h_{k_1}(\theta_j)$ and $h_{k_2}(\theta_j)$, the x-coordinate of which can be calculated as

$$I(k_1, k_2) = \frac{\tilde{\alpha}_{k_1} - \tilde{\alpha}_{k_2}}{\beta_{k_1} - \beta_{k_2}}.$$

Problem (4.9) is essentially a 0-1 knapsack problem with unit weight, the optimal solution is to select top $\mathbb{C}_j$ products that are ordered by $f_k(\theta_j)$ for $k \in \text{Children}(j)$ with a given θ_j. The important observations are: 1) the ordering of $f_k(\theta_j)$ is identical to the ordering of $h_k(\theta_j)$; 2) the ordering does not change when θ_j takes values between two consecutive intersection points. For example, the basic node j has three products, indexed by $k = 1, 2, 3$, with cardinality limitation of two. Figure 4.2 shows how to solve problem (4.9) at node j, there are three lines: $h_k(\theta_j) = \tilde{\alpha}_k - \beta_k \theta_j$ for $k = 1, 2, 3$, which are intersected at 3 points and the real line is divided into 4 intervals. For the different ranges of θ_j, the optimal assortment $\tilde{S}_j(\theta_j)$ to problem (4.9) is shown in Table 4.1.

For $\theta_j \in \mathbb{R}$, the size of candidate collection $\mathscr{A}_j$ including optimal assortments to problem (4.8) at basic node $j \in \mathscr{B}$ is bounded by $O(N_j^2)$ since there are at most $\binom{N_j}{2} = O(N_j^2)$ intersection points of these lines. As in the above example, $\tilde{S}_j(\theta_j)$ does not change at intersection points a and c, and only changes at intersection point b with one product being replaced by a new product. Thus the set of intersection points can be further refined to a set of *changing* points where $\tilde{S}_j(\theta_j)$ actually changes. Moreover, the set of changing points has size $O(N_j)$ when the cardinality limitation $\mathbb{C}_j$ is prespecified [39].

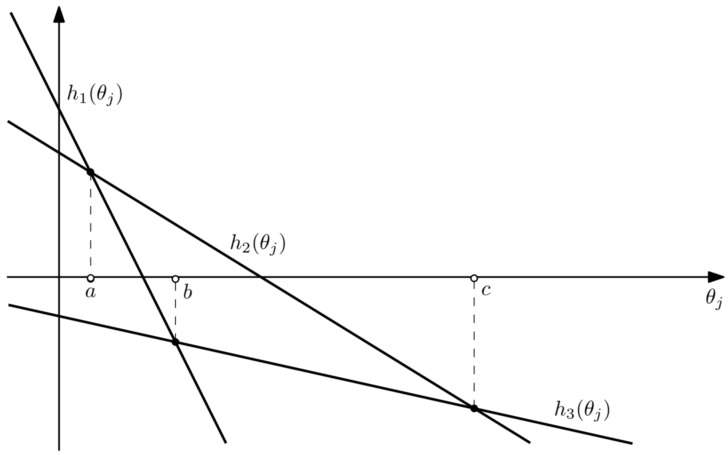

Figure 4.2: An illustration of solving problem (4.9) at basic node j

θ_j	$(-\infty, a)$	(a, b)	(b, c)	$(c, +\infty)$
$\hat{S}_j(\theta_j)$	$\{1, 2\}$	$\{1, 2\}$	$\{2, 3\}$	$\{2, 3\}$

Table 4.1: Optimal assortment $\hat{S}_j(\theta_j)$ to problem (4.9) for different ranges of θ_j

Lemma 18. *For basic node $j \in \mathscr{B}$, the preference weight $V_j(\tilde{S}_j(\theta_j), \theta_j)$ as a function of θ_j drops discontinuously at the changing point θ'_j.*

By the above discussion, the size of $\mathscr{A}_j = \{\tilde{S}_j(\theta_j) : \theta_j \in \mathbb{R}\}$ is $O(N_j)$. Denote i as the parent node of basic j, then $\mathscr{A}_i = \bigtimes_{j \in \text{Children}(i)} \mathscr{A}_j$ includes an optimal assortment to joint subproblem (4.2) at node i by Lemma 13. The size of $\mathscr{A}_i$ is $O(\prod_{j \in \text{Children}(i)} N_j)$, which is roughly $O(K^K)$ since K is the largest number of children nodes that a nonleaf node can have in the tree structure. Even if we have reduced the feasible searching space of the assortment subproblem (4.6) at node i from $\Im_i$ to $\mathscr{A}_i$, it is still intractable with a large K. To tackle this intractability, we next consider problem (4.6) at basic node j with smaller searching space. Then we stitch $\mathscr{A}_j$ for $\forall j \in \text{Children}(i)$ together in a systematic way to get $\mathscr{A}_i$, the size of which is at most $O(\sum_{j \in \text{Children}(i)} |\mathscr{A}_j|)$.

Since $\mathscr{A}_j$ includes the global optimal assortment S_j^*, we can use $\mathscr{A}_j$ to replace $\Im_j$ in problem (4.6) at node j without affecting the optimality. For $\theta_i \in \mathbb{R}$, we let $\hat{S}_j(\theta_i)$ be an optimal solution to the following problem

$$\begin{aligned} \max_{S_j \subseteq \mathscr{A}_j} \quad & \frac{V_j(S_j, \theta_j)\omega_j(S_j, \theta_j)}{1 - \gamma_j} \\ \text{s.t.} \quad & \theta_i = \theta_j - \omega_j(S_j, \theta_j). \end{aligned} \tag{4.10}$$

Then the set $\mathscr{A}_i = \{\bigcup_{j\in\text{Children}(i)} \hat{S}_j(\theta_i) : \theta_i \in \mathbb{R}\}$ includes a global optimal assortment S_i^*, since $S_i^* = \bigcup_{j\in\text{Children}(i)} \hat{S}_j(\theta_i^*)$ by lemma 13 and $\bigcup_{j\in\text{Children}(i)} \hat{S}_j(\theta_i^*) \in \mathscr{A}_i$. Recall that $\mathscr{A}_j = \{\tilde{S}_j(\theta_j) : \theta_j \in \mathbb{R}\}$ where $\tilde{S}_j(\theta_j)$ is the optimal solution to problem (4.9), thus we can reformulate problem (4.10) in terms of optimizing over the scalar $\theta_j \in \mathbb{R}$ as follows

$$\begin{aligned} \max_{\theta_j\in\mathbb{R}} \quad & \frac{V_j(\tilde{S}_j(\theta_j), \theta_j)\omega_j(\tilde{S}_j(\theta_j), \theta_j)}{1-\gamma_j} \\ \text{s.t.} \quad & \theta_i = \theta_j - \omega_j(\tilde{S}_j(\theta_j), \theta_j). \end{aligned} \tag{4.11}$$

The above finding is recorded in the next proposition.

Proposition 13. *Problem (4.11) and (4.10) are equivalent formulation at basic node j.*

For the constraint in problem (4.11), if there exists a one-to-one relationship between θ_i and θ_j, the feasible region of decision variable θ_j is a singleton for a given θ_i, then this problem is trivially solved by finding the matching θ_j. For $\theta_i = \theta_j - \omega_j(\tilde{S}_j, \theta_j)$, the one-to-one correspondence exists for a specific assortment $\tilde{S}_j$ by lemma 15. However, for the constraint $\theta_i = \theta_j - \omega_j(\tilde{S}_j(\theta_j), \theta_j)$, one θ_i may correspond to multiple θ_j's since $\tilde{S}(\theta_j)$ is also a function of θ_j. In the following lemma, we summarize our observations.

Lemma 19. *Let $\tilde{S}_j(\theta_j)$ be the optimal solution to problem (4.9) at basic node j for $\theta_j \in \mathbb{R}$, then $\theta_i = \theta_j - \omega_j(\tilde{S}_j(\theta_j), \theta_j)$ is discontinuous with respect to θ_j. Moreover, there exists $\underline{\theta}_i < \overline{\theta}_i$ such that one θ_i correspond to at least two θ_j's when $\theta_i \in [\underline{\theta}_i, \overline{\theta}_i]$.*

From the above lemma, there is no one-to-one correspondence for some range of θ_i, but for some other ranges of θ_i, the one-to-one correspondence do exist. We define $\mathcal{F}_j(\theta_i)$ to be the optimal solution to problem (4.11) at basic node $j \in \mathscr{B}$, then we have $\hat{S}_j(\theta_i) = \tilde{S}_j(\mathcal{F}_j(\theta_i))$ by proposition 13. Thus $\mathscr{A}_i = \{\bigcup_{j\in\text{Children}(i)} \tilde{S}_j(\mathcal{F}_j(\theta_i)) : \theta_i \in \mathbb{R}\}$ includes S_i^* by the argument discussed after introducing problem (4.10). For notational consistency, we define $\tilde{S}_i(\theta_i) = \bigcup_{j\in\text{Children}(i)} \tilde{S}_j(\mathcal{F}_j(\theta_i))$. The above results only hold for basic node j, the next proposition shows that these findings can be generalized to any nonleaf node i in the tree structure.

Proposition 14. *For any nonleaf node $i \in V$, let h and j represent its parent and children node, respectively. We define $\mathcal{F}_i(\theta_h)$ as the optimal solution to the following problem*

$$\begin{aligned} \max_{\theta_i\in\mathbb{R}} \quad & \frac{V_i(\tilde{S}_i(\theta_i), \theta_i)\omega_i(\tilde{S}_i(\theta_i), \theta_i)}{1-\gamma_i} \\ s.t. \quad & \theta_h = \theta_i - \omega_i(\tilde{S}_i(\theta_i), \theta_i), \end{aligned} \tag{4.12}$$

where $\tilde{S}_i(\theta_i)$ is recursively defined as follows

$$\tilde{S}_i(\theta_i) = \bigcup_{j\in\text{Children}(i)} \tilde{S}_j(\mathcal{F}_j(\theta_i)).$$

Then $\mathscr{A}_h = \{\bigcup_{i\in\text{Children}(h)} \tilde{S}_i(\mathcal{F}_i(\theta_h)) : \theta_h \in \mathbb{R}\}$ includes the optimal assortment S_h^* at node h.

$\mathscr{A}_h$ is created by stitching together $\tilde{S}_i(\mathcal{F}_i(\theta_h))$ through the scalar θ_h, but we still have concerns about the size of $\mathscr{A}_h$ since θ_h takes value in $\mathbb{R}$. The next lemma shows that problem (4.12) has nice properties, which helps us to prove that $\mathscr{A}_h$ has polynomial size of N and can be obtained efficiently.

Lemma 20. *Denote the objective function in problem (4.12) as $T_i(\tilde{S}_i(\theta_i), \theta_h)$ for a given θ_h, where θ_i is implicitly defined in $\theta_h = \theta_i - \omega_i(\tilde{S}_i(\theta_i), \theta_i)$, then*

1. *The first derivative of $T_i(\tilde{S}_i(\theta_i), \theta_h)$ with respect to θ_h is*
$$\frac{\partial T_i(\tilde{S}_i(\theta_i), \theta_h)}{\partial \theta_h} = -V_i(\tilde{S}_i(\theta_i), \theta_i);$$
2. *The objective function $T_i(\tilde{S}_i(\theta_i), \theta_h)$ of problem (4.12) is a convex function in terms of θ_h.*
3. *For two different assortments $S_1, S_2 \in \mathscr{A}_i$, if $V_i(S_1, \theta_i(S_1, \theta_h))$ and $V_i(S_2, \theta_i(S_2, \theta_h)))$ do not intersect, then $T_i(S_1, \theta_h)$ and $T_i(S_2, \theta_h)$ intersect at most once in θ_h domain.*

The objective function of problem (4.12) is a strictly decreasing convex function in terms of θ_h. Hence only the ordering of these convex curve matters, which is similar to the way that we used to study problem (4.9). For example, both θ_{i_1} and θ_{i_2} satisfy the constraint in problem (4.12) for a given θ_h. Let $S_{i_1} = \tilde{S}_i(\theta_{i_1})$ and $S_{i_2} = \tilde{S}_i(\theta_{i_2})$, the intersection points of $T_i(S_{i_1}, \theta_h)$ and $T_i(S_{i_2}, \theta_h)$ are easy to be identified by binary search due to the convexity nature of the objective function. When θ_h takes value between these intersection points, assortment $\tilde{S}_i(\mathcal{F}_i(\theta_h))$ does not change. The next proposition shows that the size of $\mathscr{A}_h$ is bounded and the size of $\{\tilde{S}_i(\theta_i) : \theta_i \in \mathbb{R}\}$ equals to the size of $\{\tilde{S}_i(\mathcal{F}_i(\theta_h)) : \theta_h \in \mathbb{R}\}$ since $T_i(S_{i_1}, \theta_h)$ and $T_i(S_{i_2}, \theta_h)$ intersect at most once.

Proposition 15. *For nonleaf node $i \in V$ and its parent node $h = \text{Parent}(i)$, let $\mathscr{A}_i = \{\tilde{S}_i(\theta_i) : \theta_i \in \mathbb{R}\}$, then $\mathscr{A}_h = \{\bigcup_{i \in \text{Children}(h)} \tilde{S}_i(\mathcal{F}_i(\theta_h)) : \theta_h \in \mathbb{R}\}$ includes S_h^* and we have $|\mathscr{A}_h| \leqslant \sum_{i \in \text{Children}(h)} |\mathscr{A}_i|$. Furthermore, the size of $|\mathscr{A}_i|$ is $O(N)$.*

The above proposition ensures that the size of $\mathscr{A}_i$ grows linearly as node i moves from bottom to top and it is bounded by $O(N)$. Otherwise, imagine that the above proposition does not hold, then $\mathscr{A}_i$ is the Cartesian product of $\mathscr{A}_j$ for all $j \in \text{Children}(i)$, and the size of $\mathscr{A}_i$ grows exponentially in a bottom-up manner. Our joint optimization algorithm to problem (4.1) with cardinality constraints that is about to be discussed in the next subsection is based on the discretization of the node-specific adjusted markup θ_i, thus we need at least $|\mathscr{A}_i|$ grid points to represent $\tilde{S}_i(\theta_i) \in \mathscr{A}_i$. If $\mathscr{A}_i$ includes too many different assortments, it makes $\tilde{S}_i(\theta_i)$ very sensitive to the change of θ_i, thus even more grid points are required, which in turn leads to the intractability of problem (4.1).

Joint Optimization Algorithm

We present the algorithm JCAPO-C to sovle the joint capacitated assortment and price optimization problem (4.1) under cardinality constraints. In earlier subsections, we show that problem (4.1) is reduced to the optimization problem (4.12) with respect to the node-specific adjusted markup θ_i at any nonleaf node $i \in V$. We discretize θ_i to a set of grid points $G_i = \{o_i^1, ..., o_i^g, ..., o_i^G\}$ with size G; see [37] for selection of grid points. For grid point $o_i^g \in G_i$, we get $\tilde{S}_i(o_i^g)$ by stitching together the assortments of its children nodes: $\tilde{S}_i(o_i^g) = \bigcup_{j \in \text{Children}(i)} \tilde{S}_j(\mathcal{F}_j(o_i^g))$, then we move up to upper-level nodes until we reach the root node by proposition 14, and obtain θ^*_{root} by a fix point representation. Therefore, the global optimal price vector $\mathbf{P}^*_{\text{root}}$ to problem (4.1) is uniquely defined due to the one-to-one correspondence between $\mathbf{P}^*_{\text{root}}$ and θ^*_{root} by Lemma 15 and the global optimal assortment S^*_{root} is $\tilde{S}_{\text{root}}(\theta^*_{\text{root}})$.

We define a mapping as $\lambda : o_i^g \to o_h^{g'}$ such that $o_h^{g'} \in G_h$ is the grid point that is closest to $o_i^g - \omega(\tilde{S}_i(o_i^g), o_i^g)$ where $h = \text{Parent}(i)$. Algorithm JCAPO-C, which is presented in Algorithm 5 as follows, solves problem (4.1) under cardinality constraints in a bottom-up manner, which starts from basic nodes and then moves to its parent node until the root node. It has three parts corresponding to i) problem (4.9) at basic nodes, ii) problem (4.12) at upper-level nodes and iii) solving the fixed point representation at root node, respectively.

In this algorithm, $\mathcal{F}_i(o_h^{g'})$ is initialized to $-M$ and then updated to the optimal solution to problem (4.12). For each update, it compares the objective function with θ_i taking value at current $\mathcal{F}_i(o_h^{g'})$ and new coming point o_i^g, which takes $O(1)$ since it only includes one comparison. Line 10 and line 22 essentially deals with problem (4.12) at basic node j and nonbasic node i, respectively, where the constraints in problem (4.12) are automatically satisfied due to the mapping λ. We make an observation that line 10 and line 22 run in $O(1)$ since it only takes a single numerical comparison. In line 17, $V_i(\tilde{S}_i(o_i^g), o_i^g) = (\sum_{j \in \text{Chilren}(i)} V_j(\tilde{S}_j(\mathcal{F}_j(o_i^g)), \mathcal{F}_j(o_i^g)))^{\gamma_i}$ and it can be calculated in $O(1)$ by looking up the previously stored preference weight of its children node j, and $\omega_i(\tilde{S}_i(o_i^g), o_i^g)$ can be computed in $O(1)$ as well. The next theorem states that the joint capacitated assortment and price optimization problem (4.1) under cardinality constraints can be solved in $O(GN \log G)$.

Theorem 9. *The computational complexity of Algorithm JCAPO-C is $O(GN \log G)$, where G is the number of grid points for each node in the tree structure and N is the total number of candidate products. Furthermore, if the spacings of grid points are identical, it can be further reduced to $O(GN \log K)$ where K is the maximum children nodes that a nonleaf node can have in the tree logit model.*

We remark that the computational complexity is irrelevant of the number of levels m and the total number of grid points is bounded by $O(GN)$ since the total number of nodes is less than twice the number of leaf nodes in a tree.

4.5 Joint Optimization Under Space Constraints

In this section, we consider the joint capacitated assortment and price optimization problem (4.1) under space constraints. We develop an algorithm that runs in $O(GN \log G)$ to obtain a 2-approximate solution. Similarly, this algorithm that is referred to as JCAOP-S is more efficient with same performance guarantee under mild adjustments. Under space constraints, the feasible assortment is $\Im_j = \{S_j : S_j = \bigcup_{k\in\text{Children(j)}} S_k, \sum_{k\in\text{Children}(j)} w_k \leqslant \mathbb{S}_j\}$ for $\forall j \in \mathscr{B}$. Let $S^\alpha_{\text{root}} = (S^\alpha_i : i \in V)$ denote the α-approximate assortment throughout this section. First, the joint optimization problem under space constraints can also be decomposed into assortment subproblem on a subset of feasible assortments $\Im^\alpha_i \subseteq \Im_i$ for all nonleaf node $i \in V$, where we assume $\Im^\alpha_i$ that includes S^α_i is known. Second, we show $\Im^\alpha_i$ can be constructed and we reduce the size of $\Im^\alpha_i$ to a new set $\mathscr{A}^\alpha_i$ that still contains S^α_i at node i. Third, we propose a polynomial-time algorithm to get a 2-approximate solution of the joint capacitated assortment and price optimization problem (4.1) under the space constraints.

Problem Decomposition

For problem (4.1) under space constraints, $\Im_{\text{root}} = \bigtimes_{j\in\mathscr{B}} \Im_j$ denotes set of all possible assortments that satisfy the space constraints. However, the assortment optimization problem under space constraints is NP-hard, even for the two-level tree logit model with fixed prices. Thus we aim to find an α-approximate solution to problem (4.1) under space constraints. Assume $\Im^\alpha_{\text{root}}$ that is a subset of $\Im_{\text{root}}$ is given and it includes an α-approximate assortment S^α_{root}. Let $(S^\alpha_{\text{root}}, \mathbf{P}^\alpha_{\text{root}})$ be the optimal solution to the following problem

$$Z^\alpha = \max_{S_{\text{root}}\subseteq \Im^\alpha_{\text{root}}} \max_{\mathbf{P}_{\text{root}}\in\mathbb{R}^{|S_{\text{root}}|}_{\geqslant 0}} R_{\text{root}}(S_{\text{root}}, \mathbf{P}_{\text{root}}), \tag{4.13}$$

where $Z^\alpha = R_{\text{root}}(S^\alpha_{\text{root}}, \mathbf{P}^\alpha_{\text{root}})$ is the maximum profit that can be obtained from $\Im^\alpha_{\text{root}}$. $(S^\alpha_{\text{root}}, \mathbf{P}^\alpha_{\text{root}})$ is an α-approximate solution to problem (4.1) under space constraints, thus $\alpha Z^\alpha \geqslant Z^* = R_{\text{root}}(S^*_{\text{root}}, \mathbf{P}^*_{\text{root}})$ where $(S^*_{\text{root}}, \mathbf{P}^*_{\text{root}})$ is the optimal solution to problem (4.1) under space constraints. For notational consistency, $(S^\alpha_i, \mathbf{P}^\alpha_i)$ denotes the optimal assortment and price vector to problem (4.13) at nonleaf node $i \in V$. When $\alpha = 1$, $(S^\alpha_i, \mathbf{P}^\alpha_i)$ is the optimal solution, thus denote $S^*_i = S^1_i$ and $\mathbf{P}^*_i = \mathbf{P}^1_i$. Similar to the problem with cardinality constraints in Section 4.4, we define the scalar $e^\alpha_i = \gamma_i e^\alpha_h + (1-\gamma_i) R_i(S^\alpha_i, \mathbf{P}^\alpha_i)$, which can also be calculated in a top-down manner. And for each nonleaf node $i \in V$, we introduce the α-joint subproblem as follows

$$\max_{S_i\subseteq \Im^\alpha_i} \max_{\mathbf{P}_i\in\mathbb{R}^{|S_i|}_{\geqslant 0}} V_i(S_i, \mathbf{P}_i)(R_i(S_i, \mathbf{P}_i) - e^\alpha_h), \tag{4.14}$$

where $\Im^\alpha_i$ is known since $\Im^\alpha_{\text{root}}$ is provided. If we define the α-optimal node-specific adjusted markup θ^α_i as the optimal solution to $\max_{\theta_i\in\mathbb{R}} V_i(S^\alpha_i, \theta_i)(R_i(S^\alpha_i, \theta_i) - e^\alpha_h)$, one can check that θ^α_i equals to e^α_i and satisfies Definition 1 with assortment S^α_i by using similar techniques as in Section 4.4. In this notation, we let $\theta^*_i = \theta^1_i$.

We claim that problem (4.14) has the following equivalent formulations, which is referred to as the α-assortment subproblem

$$\begin{aligned} \max_{S_i \subseteq \Im_i^\alpha} \quad & \frac{V_i(S_i, \theta_i)\omega_i(S_i, \theta_i)}{1-\gamma_i} \\ \text{s.t.} \quad & \theta_h^\alpha = \theta_i - \omega_i(S_i, \theta_i), \end{aligned} \tag{4.15}$$

and the optimal solution to the following problem, which is referred to as the basic α-assortment subproblem

$$\begin{aligned} \max_{S_i \subseteq \Im_i^\alpha} \quad & \sum_{j \in \text{Children}(i)} \frac{V_j(S_j, \theta_j)\omega_j(S_j, \theta_j)}{1-\gamma_j} \\ \text{s.t.} \quad & \theta_i^\alpha = \theta_j - \omega_j(S_j, \theta_j), \end{aligned} \tag{4.16}$$

is also optimal to problem (4.15). Proof of this claim follows directly from the results in Section 4.4 by changing our notation from $\{S_i^*, \theta_i^*, \Im_i\}$ to $\{S_i^\alpha, \theta_i^\alpha, \Im_i^\alpha\}$. Similar to the third item in Lemma 13, the following lemma also holds.

Lemma 21. *Let $\hat{S}_i$ be an optimal solution to α-assortment subproblem at nonleaf node $i \in V$, then $\bigcup_{i \in \text{Children}(h)} \hat{S}_i$ is also optimal to the α-assortment subproblem at i's parent node h.*

By applying the above lemma repeatedly, we have that $S_{\text{root}}^\alpha = \bigcup_{j \in \mathscr{B}} S_j^\alpha$ is the optimal assortment to problem (4.13) where S_j^α is the optimal solution to problem (4.15) at basic node $j \in \mathscr{B}$. Moreover, $\theta_{\text{root}}^\alpha$ is the solution to this fixed point representation: $R(S_{\text{root}}^\alpha, \theta_{\text{root}}) = \theta_{\text{root}}$, which satisfies $\alpha R(S_{\text{root}}^\alpha, \theta_{\text{root}}^\alpha) \geqslant Z^*$.

Candidate Assortment Construction

Although the joint optimization problem under space constraints can be decomposed to problem (4.15), we still have concerns since it requires to know $\Im_i^\alpha$ and θ_h^α. θ_h^α can be computed in a top-down manner if we know S_{root}^α since θ_h^α equals to e_h^α. In this subsection, we first show that S_{root}^α can be obtained if θ_i^* is known for $\forall i \in V$. Then we construct $\Im_i^\alpha$ that includes S_i^α and show how $\Im_i^\alpha$ can be simplified to $\mathscr{A}_i^\alpha$, the size of which is bounded by $O(N)$. We use these results in the next subsection to develop an algorithm running in polynomial time to solve problem (4.1) under space constraints with a performance guarantee. The following lemma shows that an α-approximate assortment S_{root}^α can be obtained provided that θ_i^* is given.

Lemma 22. *Let $\hat{S}_j^\alpha$ be an α-approximate solution to problem (4.8) at basic node $j \in \mathscr{B}$ under space constraints with parameter $\alpha \geqslant 1$. Then assortment $S_{\text{root}}^\alpha = \bigcup_{j \in \mathscr{B}} \hat{S}_j^\alpha$ is an α-approximate assortment.*

Once $\hat{S}_{\text{root}}^\alpha$ is known, we compute θ_h^α in a top-down manner. The next lemma clears the concern that we have at the beginning of this subsection by showing that $\Im_{\text{root}}^\alpha$ can also be built.

Lemma 23. *Define* $\mathscr{A}_j^\alpha = \{\tilde{S}_j^\alpha(\theta_j) : \theta_j \in \mathbb{R}\}$*, where* $\tilde{S}_j^\alpha(\theta_j)$ *is an* α*-approximate solution to problem (4.9) at basic node* $j \in \mathscr{B}$ *under space constraints, then* $\Im_{\text{root}}^\alpha = \times_{j\in\mathscr{B}} \mathscr{A}_j^\alpha$ *includes* S_{root}^α.

Proof of the above lemma is straight forward: since $S_j^\alpha = \tilde{S}_j^\alpha(\theta_j^*)$, then $\mathscr{A}_j^\alpha$ contains S_j^α. Thus $\Im_{\text{root}}^\alpha = \times_{j\in\mathscr{B}} \mathscr{A}_j^\alpha$ contains $S_{\text{root}}^\alpha = \bigcup_{j\in\mathscr{B}} \hat{S}_j^\alpha$. By applying the approach in [18], we are able to locate a 2-approximate solution to problem problem (4.9) under space constraints by finding the intersection points of these lines: $\{h_k(\theta_j) : k \in \text{Children}(j)\}$ where $h_k(\theta_j) = \tilde{\alpha}_k' - \beta_k\theta_j = \alpha_k - \beta_k c_k - \log(\beta_k) - 1$. Similar to the joint optimization problem under cardinality constraints, we define the mapping $\mathcal{F}_i^\alpha(\theta_h)$ as

$$\mathcal{F}_i^\alpha(\theta_h) = \arg\max_{\theta_i\in\mathbb{R}}\{\frac{V_i(\tilde{S}_i^\alpha(\theta_i),\theta_i)\omega_i(\tilde{S}_i^\alpha(\theta_i),\theta_i)}{1-\gamma_i} : \theta_h = \theta_i - \omega_i(\tilde{S}_i^\alpha(\theta_i),\theta_i)\},$$

where $\tilde{S}_i^\alpha(\theta_i)$ is recursively defined as follows

$$\tilde{S}_i^\alpha(\theta_i) = \bigcup_{j\in\text{Children}(i)} \tilde{S}_j^\alpha(\mathcal{F}_j^\alpha(\theta_i)).$$

$\mathcal{F}_i^\alpha(\theta_h)$ is essentially the optimal solution to problem (4.16) for a given $\theta_h^\alpha = \theta_h$. Note that $\mathcal{F}_i^\alpha(\theta_h^\alpha) = \theta_i^\alpha$. The next proposition shows that $\Im_{\text{root}}^\alpha$ can be reduced to $\mathscr{A}_{\text{root}}^\alpha$ with size $O(N)$.

Proposition 16. *Let* i *be any nonleaf node, assume* $\mathscr{A}_i^\alpha = \{\tilde{S}_i^\alpha(\theta_i) : \theta_i \in \mathbb{R}\}$ *includes* S_i^α*. Then there exists a collection of candidate assortments* $\mathscr{A}_h^\alpha = \{\bigcup_{i\in\text{Children}(h)} \tilde{S}_i^\alpha(\mathcal{F}_i^\alpha(\theta_h)) : \theta_h \in \mathbb{R}\}$ *that includes* S_h^α *where* h *is* i*'s parent node. We also have* $|\mathscr{A}_h^\alpha| \leqslant \sum_{i\in\text{Children}(h)} |\mathscr{A}_i^\alpha|$ *and* $|\mathscr{A}_i^\alpha|$ *has size* $O(N)$*.*

Joint Optimization Algorithm

We present the algorithm JCAPO-S of the joint capacitated assortment and price optimization problem (4.1) under space constraints in this subsection. Similar to Section 4.4, we build a set of grid points $G_i = \{o_i^1, ..., o_i^g, ..., o_i^G\}$ for each nonleaf node $i \in V$, the size of which is G. For each $o_i^g \in G_i$, $\tilde{S}_i^\alpha(o_i^g)$ can be obtained by $\tilde{S}_i^\alpha(o_i^g) = \bigcup_{j\in\text{Children}(i)} \tilde{S}_j^\alpha(\mathcal{F}_j^\alpha(o_i^g))$. The collection $\mathscr{A}_i^\alpha = \{\tilde{S}_i^\alpha(o_i^g) : o_i^g \in G_i\}$ includes an α-approximate assortment S_i^α at node i. After getting o_{root}^α by solving a fixed point representation, we are able to obtain $(S_{\text{root}}^\alpha, \mathbf{P}_{\text{root}}^\alpha)$ that is an α-approximate solution to problem (4.1) under space constraints.

The algorithm in Appendix C.3 is referred to as Algorithm JCAPO-C, which solves problem (4.1) under space constraints with a performance of guarantee of two. In this algorithm, the mapping λ is defined as $\lambda : o_i^g \to o_h^{g'}$ such that $o_h^{g'} \in G_h$ is the grid point that is closest to $o_i^g - \omega(\tilde{S}_i^\alpha(o_i^g), o_i^g)$ by assuming that $\tilde{S}_i^\alpha(\theta_i)$ is known for $\theta_i \in G_i$ and let node h be the parent node of i.

For the problem in line 4, we can get a 2-approximate solution in polynomial time [18]. Furthermore, 2-approximate solution to problem (4.1) with space constraints is obtained in

$O(GN \log G)$ time by applying Algorithm JCAPO-S, the findings of which is summarized in the following theorem.

Theorem 10. *A 2-approximate solution to problem (4.1) under space constraints can be found through Algorithm JCAPO-S, the complexity of which is $O(GN \log G)$. Furthermore, its complexity can be reduced to $O(GN \log K)$.*

4.6 Algorithm Illustration

In this section, we illustrate the joint optimization algorithm JCAOP-C on an instance of a three-level tree logit model, the structure of which is shown in Figure 4.1. The major difference between algorithm JCAOP-C and JCAOP-S is the optimization problem at basic nodes, but the rest part of two algorithms share the same idea. Hence this illustration can also be adapted to demonstrating JCAOP-S with minor adjustments.

A set of ten products $\{g, h, ..., p\}$ and a no-purchase option are considered in our setting. The remaining nodes in the tree structure are all nonleaf nodes, the set of which is denoted as $\{a, b, ..., f, \text{root}\}$. The following Table 4.2 shows the model parameters. The preference weight of no-purchase option is set to be 10, price-independent deterministic utility α_k, price-sensitivity parameter β_k and cost c_k are provided for each leaf node k, and dissimilarity parameter γ_i is given for each nonleaf node i. The bottom part of this table shows the cardinality constraint $\mathbb{C}_j$ on basic node j.

By Lemma 17, we have $S_c^* = \{m, n\}$ and $S_d^* = \{o, p\}$, then $S_b^* = \{m, n, o, p\}$. Thus we focus on the left portion of the tree. For basic node c and by line 4 in algorithm JCAOP-C, one can verify that $\tilde{S}_c(o_c) = \{g, h\}$ for grid point $o_c \in [0, 3.02]$ and $\tilde{S}_c(o_c) = \{g, i\}$ for $o_c \in [3.02, 10]$. Note that set $\mathscr{A}_c = \{\{g, h\}, \{h, i\}\}$ includes S_c^*. Then we plot $\theta_a = o_c - \omega_c(\tilde{S}_c(o_c), o_c)$ as a function of o_c as shown in Figure 4.3(a), where $\omega_c(\tilde{S}_c(o_c), o_c)$ jumps discontinuously at $o_c = 3.02$. One can see that a single θ_a corresponds to two grid points o_c's when $\theta_a \in [2.05, 2.21]$, which is an example of Lemma 19. We then consider line 22 in algorithm JCAOP-C, and the visualization of solving this optimization problem is demonstrated in Figure 4.3(b), where the two convex decreasing curves corresponding to assortments $\{g, h\}$ and $\{h, i\}$ intersect only once at $o_a = 2.14$. Thus $\tilde{S}_c(\mathcal{F}_c(o_a)) = \{g, h\}$ for $o_a \in [-0.73, 2.14]$ and $\tilde{S}_c(\mathcal{F}_c(o_a)) = \{h, i\}$ for $o_a \in [2.14, 9.02]$.

Similarly, we go through the above process for another basic node d. Then we obtain $\tilde{S}_a(o_a)$ by stitching together $\tilde{S}_c(\mathcal{F}_c(o_a))$ and $\tilde{S}_d(\mathcal{F}_d(o_a))$ via the grid point o_a. For instance, if $o_a \in [-0.73, 1.77]$, then $\tilde{S}_a(o_a) = \tilde{S}_c(\mathcal{F}_c(o_a)) \bigcup \tilde{S}_d(\mathcal{F}_d(o_a)) = \{g, h\} \bigcup \{j, k\} = \{g, h, j, k\}$; if $o_a \in [1.77, 2.14]$, $\tilde{S}_c(\mathcal{F}_c(o_a))$ is still $\{g, h\}$ and $\tilde{S}_d(\mathcal{F}_d(o_a))$ becomes $\{k, l\}$, thus $\tilde{S}_a(o_a) = \tilde{S}_c(\mathcal{F}_c(o_a)) \bigcup \tilde{S}_d(\mathcal{F}_d(o_a)) = \{h, i\} \bigcup \{j, k\} = \{g, h, k, l\}$; if $o_a \in [2.14, 8.91]$, then $\tilde{S}_a(o_a) = \{h, i, k, l\}$. Note that the collection $\mathscr{A}_a = \{\tilde{S}_a(o_a) : o_a \in G_a\} = \{\{g, h, j, k\}, \{g, h, k, l\}, \{h, i, k, l\}\}$ includes the global optimal assortment S_a^* to problem (4.1) with cardinality constraints. We also have that $|\mathscr{A}_a|$ is less than the sum of $|\mathscr{A}_c|$ and $|\mathscr{A}_d|$, which addresses Proposition 15.

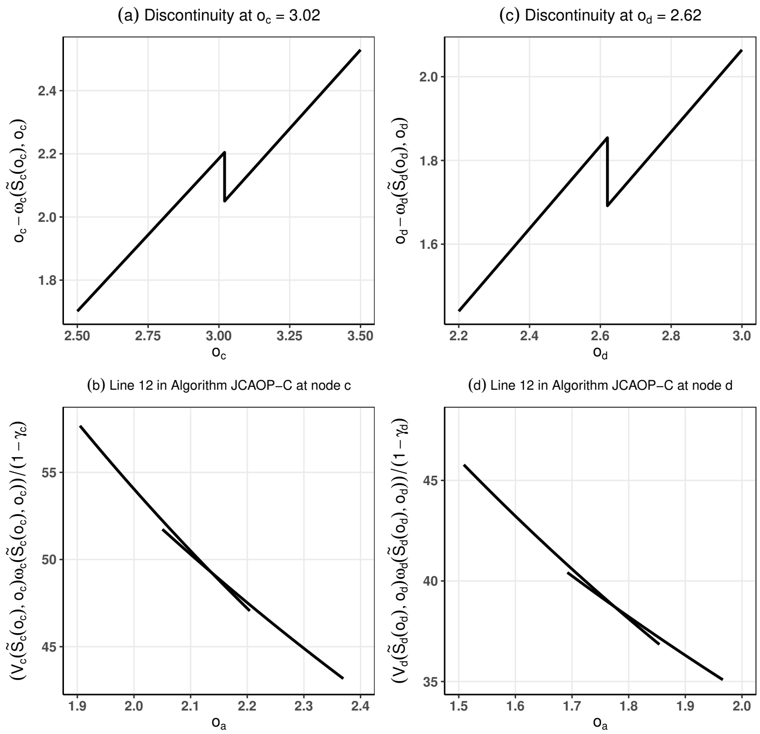

Figure 4.3: Computation in Table 4.3

The construction of $\tilde{S}_a(o_a)$ is shown in Table 4.3 with corresponding optimization process that is shown in the above Figure 4.3.

Next we build $\tilde{S}_{\text{root}}(o_{\text{root}})$ from $\mathscr{A}_a = \{\{g,h,j,k\},\{g,h,k,l\},\{h,i,k,l\}\}$ and $\mathscr{A}_b = S_b^* = \{m,n,o,p\}$. The construction of $\tilde{S}_{\text{root}}(o_{\text{root}})$ is given in Table 4.4, which shares a similar process as in Table 4.3 and Figure 4.3. For example, the bottom part of Table 4.4 shows that $\tilde{S}_{\text{root}}(o_{\text{root}}) = \{g,h,j,k,m,n,o,p\}$ for all the grid points in interval $[1.43, 1.75]$, and if o_{root} takes value in $[1.75, 8.50]$, $\tilde{S}_{\text{root}}(o_{\text{root}})$ changes to set $\{h,i,k,l,m,n,o,p\}$. The set $\mathscr{A}_{\text{root}} = \{\{g,h,j,k,m,n,o,p\},\{g,h,k,l,m,n,o,p\},\{h,i,k,l,m,n,o,p\}\}$ includes S_{root}^*, the size of which is only three and it is less than the total number of products.

Then we solve for optimal o_{root}^* via the fixed point representation: $R_{\text{root}}(\tilde{S}_{\text{root}}(o_{\text{root}}), o_{\text{root}}) = o_{\text{root}}$, which can be visualized in Figure 4.4. The solid curve in Figure 4.4 is the plot of the profit function with respect to o_{root} and the solid 45° -line intersects with it at $o_{\text{root}} = 3.22$. The objective function contains three segments with three corresponding different assortments and three different intervals of o_{root}. For $o_{\text{root}} \in [-1.01, 1.43]$, we have $\tilde{S}_{\text{root}}(o_{\text{root}}) = \{g,h,j,k,m,n,o,p\}$ and the maximum of the objective function is 2.16. Since there is no solution of the fixed point representation, we move to the second interval. If $o_{\text{root}} \in [1.43, 1.75]$, then $\tilde{S}_{\text{root}}(o_{\text{root}}) = \{g,h,j,k,m,n,o,p\}$ with 2.47 as its maximum of the objective function. There is still no solution to the fixed point representation, thus we consider $o_{\text{root}} \in [1.75, 8.50]$. In this interval, $\tilde{S}_{\text{root}}(o_{\text{root}}) = \{h,i,k,l,m,n,o,p\}$ and the maximum objective value is 3.22 that also satisfies $R_{\text{root}}(\tilde{S}_{\text{root}}(o^*_{\text{root}}), o^*_{\text{root}}) = o^*_{\text{root}} = 3.22$. Thus the optimal assortment S^*_{root} is $\{h, i, k, l, m, n, o, p\}$, optimal node-specific adjusted markup at the root node is $\theta^*_{\text{root}} = 3.22$ and the maximum profit is also 3.22. By looking up previous stored table, the optimal price vector for these ten products is $\mathbf{P}^*_{\text{root}} = (p^*_g, p^*_h, ..., p^*_p) = (0, 6.14, 6.11, 0, 6.03, 6.22, 4.73, 4.47, 4.55, 4.50)$. Therefore, the joint optimal assortment and price vector to problem (4.1) under cardinality constraints is $(S^*_{\text{root}}, \mathbf{P}^*_{\text{root}})$.

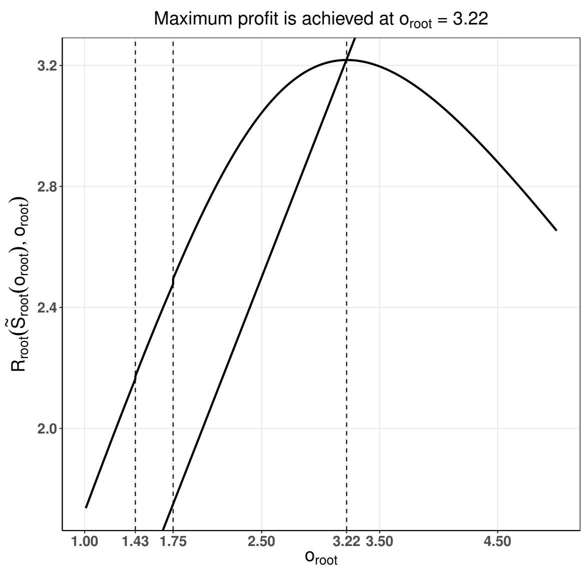

Figure 4.4: Objective function

Algorithm 5: Joint Capacitated Assortment and Price Optimization Under Cardinality Constraints (JCAPO-C)

Input: $\alpha_i, \beta_i, \gamma_i, G_i$ for $i \in V$, $\Im_j$ for $j \in \mathscr{B}$.
Initialization: Set $\mathcal{F}_j(o_i^g) = -M$ for $g = 1, 2, ..., G$ and $j \in \text{Children}(i)$;
for $j \in \mathscr{B}$ **do**
  for $g \leftarrow 1, 2, ..., G$ **do**
    get $\tilde{S}_j(o_j^g) = \arg\max_{S_j \subseteq \Im_j} \sum_{k \in \text{Children}(j)} V_k(S_k, o_j^g + c_k + 1/\beta_k)/\beta_k$;
    calculate $V_j(\tilde{S}_j(o_j^g), o_j^g)$ and $\omega_j(\tilde{S}_j(o_j^g), o_j^g)$;
    find g' such that $o_i^{g'} = \lambda(o_j^g)$;
    if $\mathcal{F}_j(o_i^{g'}) = -M$ **then**
      $\mathcal{F}_j(o_i^{g'}) \leftarrow o_j^g$;
    else
      $\mathcal{F}_j(o_i^{g'}) \leftarrow \arg\max_{\theta_j \in \{\mathcal{F}_j(o_i^{g'}), o_j^g\}} V_j(\tilde{S}_j(\theta_j), \theta_j)\omega_j(\tilde{S}_j(\theta_j), \theta_j)/(1 - \gamma_j)$;
    end
  end
end
for i in level $m-2, m-3, ..., 1$ **do**
  for $g \leftarrow 1, 2, ..., G$ **do**
    get $\tilde{S}_i(o_i^g) = \bigcup_{j \in \text{Children}(i)} \tilde{S}_j(\mathcal{F}_j(o_i^g))$;
    calculate $V_i(\tilde{S}_i(o_i^g), o_i^g)$ and $\omega_i(\tilde{S}_i(o_i^g), o_i^g)$;
    find g' such that $o_h^{g'} = \lambda(o_i^g)$;
    if $\mathcal{F}_i(o_h^{g'}) = -M$ **then**
      $\mathcal{F}_i(o_h^{g'}) \leftarrow o_i^g$;
    else
      $\mathcal{F}_i(o_h^{g'}) \leftarrow \arg\max_{\theta_i \in \{\mathcal{F}_i(o_h^{g'}), o_i^g\}} V_i(\tilde{S}_i(\theta_i), \theta_i)\omega_i(\tilde{S}_i(\theta_i), \theta_i)/(1 - \gamma_i)$;
    end
  end
end
for $g \leftarrow 1, 2, ..., G$ **do**
  get $\tilde{S}_{\text{root}}(o_{\text{root}}^g) = \bigcup_{i \in \text{Children(root)}} \tilde{S}_i(\mathcal{F}_i(o_{\text{root}}^g))$;
  calculate $R_{\text{root}}(\tilde{S}_{\text{root}}(o_{\text{root}}^g), o_{\text{root}}^g)$;
end
Solve for o_{root}^* in $o_{\text{root}} = R_{\text{root}}(\tilde{S}_{\text{root}}(o_{\text{root}}), o_{\text{root}})$, then get $S_{\text{root}}^* = \tilde{S}_{\text{root}}(o_{\text{root}}^*)$ and $P_{\text{root}}^* = P_{\text{root}}(o_{\text{root}}^*)$;
Output: $S_{\text{root}}^*, P_{\text{root}}^*$.

product	g	h	i	j	k	l	m	n	o	p
α_k	15	13	12	11	10	8	14	9	7	6
β_k	1.8	1.3	1.2	1.4	1.1	0.8	2.4	1.8	1.9	1.3
c_k	0.9	0.8	0.7	0.85	0.55	0.4	0.9	0.5	0.6	0.3
nonleaf nodes	a	b	c	d	e	f	root			
γ_i	0.83	0.95	0.45	0.52	0.73	0.81	0			
$V_{\text{No-purchase}}$	10									
basic nodes	c	d	e	f						
$\mathbb{C}_j$	2	2	2	2						

Table 4.2: Parameters setup for the joint optimization problem under cardinality constraints

$\tilde{S}_c(o_c)$	$\{g,h\}$	$\{h,i\}$	$\tilde{S}_d(o_d)$	$\{j,k\}$	$\{k,l\}$
o_c	[0, 3.02]	[3.02,10]	o_d	[0, 2.62]	[2.62, 10]
$o_c - \omega_c(\tilde{S}_c(o_c), o_c)$	[-0.73, 2.21]	[2.05, 9.02]	$o_d - \omega_d(\tilde{S}_d(o_d), o_d)$	[-0.73, 1.85]	[1.69, 8.91]
o_a	[-0.73, 2.14]	[2.14, 9.02]	o_a	[-0.73, 1.77]	[1.77, 8.91]

$\tilde{S}_a(o_a)$	$\{g,h,j,k\}$	$\{g,h,k,l\}$	$\{h,i,k,l\}$
o_a	[-0.73, 1.77]	[1.77, 2.13]	[2.13, 8.91]

Table 4.3: Construction of $\tilde{S}_a(o_a)$

$\tilde{S}_a(o_a)$	$\{g,h,j,k\}$	$\{g,h,k,l\}$	$\{h,i,k,l\}$	$\tilde{S}_b(o_b)$	$\{m,n,o,p\}$
o_a	[-0.73, 1.77]	[1.77, 2.13]	[2.13, 8.91]	o_b	$\mathbb{R}$
$o_a - \omega_a(\tilde{S}_a(o_a), o_a)$	[-1.01, 1.46]	[1.43, 1.79]	[1.75, 8.50]	$o_b - \omega_b(\tilde{S}_b(o_b), o_b)$	$\mathbb{R}$
o_{root}	[-1.01, 1.43]	[1.43, 1.75]	[1.75, 8.50]	o_{root}	$\mathbb{R}$

$\tilde{S}_{\text{root}}(o_{\text{root}})$	$\{g,h,j,k,m,n,o,p\}$	$\{g,h,k,l,m,n,o,p\}$	$\{h,i,k,l,m,n,o,p\}$
o_{root}	[-1.01, 1.43]	[1.43, 1.75]	[1.75, 8.50]

Table 4.4: Construction of $\tilde{S}_{\text{root}}(o_{\text{root}})$

Chapter 5

Conclusion

The first essay considers the joint constrained assortment and price optimization problem under the nested logit model. Under the cardinality (or space) constraints, the optimal (or a 2-approximate) solution can be identified by finding the fixed point of a unimodal function. Moreover, it can be further formulated as a piecewise convex fixed point representation. For the future research directions, one can consider the joint constrained optimization problem under the multilevel nested logit model with a no-purahse option in every choice stage by generalizing the results in [50] and [22]. The joint problem with the dissimilarity parameter exceeding one is also of interest to study.

In the second essay, we study the choice-based constrained assortment and price optimization problems under the multilevel nested logit model. Furthermore, we allow the no-purchase option in every nonleaf node within the tree structure. For the constrained assortment optimization problem, the optimal and a 2-approximate solutions can be located in polynomial time under cardinality and space constraints, respectively. Specifically, the computational time is $O(n \max\{m, k\})$ under the cardinality constraints and $O(mnk)$ under the space constraints, where m is number of levels in the multilevel nested logit model, n is the number of products and k is the maximum number of products of any basic nodes. For the price optimization problem, we reduce the nonconcave multiproduct price optimization problem to the maximization of a unimodal function, where the optimal price vector can be identified in a tractable manner. Regarding the extensions of our research, both the constrained assortment and price optimization problems with dissimilarity parameter exceeding one, are of interest for further study. One can also consider generalizing our price optimization results to multistage nested attraction models. [51] consider the joint optimization of assortment and price problem under the multilevel nested logit model with only one no-purchase option. It is interesting to study the joint optimization problem with multiple no-purchase options in the system by applying the results in this essay.

In the third essay, we consider joint capacitated assortment and price optimization problems under the tree logit model. With our efficient algorithm, we obtain the optimal solution under cardinality constraints and an approximate solution with performance guarantee under space constraints in polynomial time $O(GN \log G)$, where G is the number of grid points for each node in the tree structure and N is the total number of candidate products. With mild conditions, it can be further reduced to $O(GN \log K)$ where K is the maximum children nodes that a nonleaf node can have in the tree logit model. We formulate the joint optimization problem as a bi-level optimization program with pricing and assortment optimization problem as its inner and outer problem, respectively. Then by solving the inner pricing problem with fixed assortment, we succeed in building a bridge connecting to the outer assortment optimization problem. Finally, the bi-level optimization program is reduced to an optimization problem with respect to a scalar that is the node-specific adjusted markup and the feasible collection of assortments that include optimal solution can be constructed in a systematic way with a bounded size.

Our joint capaciated optimization algorithm and the uncapaciated assortment algorithm in [30] have the similar scale of complexity. The reason why the joint algorithm shares the similar complexity with the algorithm of the reduced assortment problem is that the core step in the uncapacitated assortment algorithm of [30] involves constantly computing the pairwise intersection points of lines, which requires sorting, however, the core part our algorithm is looking up a list of grid points that have already been sorted. One can also apply our algorithm with fixed prices to solve the uncapaciated assortment optimization as in [30] with better performance in terms of complexity.

Our study on the joint optimization problem includes all the results in earlier literatures that are based on the multinomial logit model, nest logit model and d-level nested logit model as its special cases. It also puts an end to the study on assortment/pricing/joint problems under the above three models. With minor adjustments, our approach can be adapted to solve the following three problem variants under multilevel tree logit model: 1) Assortment optimization with fixed prices; 2) Price optimization with fixed assortments; 3) Nonparametric joint assortment and price optimization with no functional assumption between preference weight and price variable. As for the extensions of our research, the joint optimization problem with dissimilarity parameter exceeding one is interesting to study. One can also consider generalizing joint optimization algorithm to the multistage nested attraction model where every node has a no-purchase option.

Bibliography

[1] Yalcin Akcay, Harihara Prasad Natarajan, and Susan H Xu. "Joint dynamic pricing of multiple perishable products under consumer choice". In: *Management Science* 56.8 (2010), pp. 1345–1361.

[2] Goker Aydin and Evan L Porteus. "Joint inventory and pricing decisions for an assortment". In: *Operations Research* 56.5 (2008), pp. 1247–1255.

[3] Goker Aydin and Jennifer K Ryan. "Product line selection and pricing under the multinomial logit choice model". In: *Proceedings of the 2000 MSOM conference*. Citeseer. 2000.

[4] Moshe E Ben-Akiva. *Structure of passenger travel demand models*. 526. 1974.

[5] Moshe E Ben-Akiva and Steven R Lerman. *Discrete choice analysis: theory and application to travel demand*. Vol. 9. MIT press, 1985.

[6] Omar Besbes and Denis Saure. "Product assortment and price competition under multinomial logit demand". In: *Production and Operations Management* 25.1 (2016), pp. 114–127.

[7] Juan José Miranda Bront, Isabel Méndez-Díaz, and Gustavo Vulcano. "A column generation algorithm for choice-based network revenue management". In: *Operations Research* 57.3 (2009), pp. 769–784.

[8] Richard T Carson, W Michael Hanemann, and Thomas C Wegge. "A nested logit model of recreational fishing demand in Alaska". In: *Marine Resource Economics* (2009), pp. 101–129.

[9] Kyle D Chen and Warren H Hausman. "Technical note: Mathematical properties of the optimal product line selection problem using choice-based conjoint analysis". In: *Management Science* 46.2 (2000), pp. 327–332.

[10] Andrew Daly. "Estimating tree logit models". In: *Transportation Research Part B: Methodological* 21.4 (1987), pp. 251–267.

[11] James M Davis, Guillermo Gallego, and Huseyin Topaloglu. "Assortment optimization under variants of the nested logit model". In: *Operations Research* 62.2 (2014), pp. 250–273.

[12] James M Davis, Huseyin Topaloglu, and David P Williamson. "Pricing Problems under the Nested Logit Model with a Quality Consistency Constraint". In: (2015).

[13] Gerard Debreu. "A social equilibrium existence theorem". In: *Proceedings of the National Academy of Sciences of the United States of America* 38.10 (1952), p. 886.

[14] Antoine Désir and Vineet Goyal. "Near-optimal algorithms for capacity constrained assortment optimization". In: *Available at SSRN 2543309* (2014).

[15] Lingxiu Dong, Panos Kouvelis, and Zhongjun Tian. "Dynamic pricing and inventory control of substitute products". In: *Manufacturing & Service Operations Management* 11.2 (2009), pp. 317–339.

[16] Jacob B. Feldman and Huseyin Topaloglu. "Capacity Constraints Across Nests in Assortment Optimization Under the Nested Logit Model". In: *Operations Research* 63.4 (2015), pp. 812–822.

[17] Guillermo Gallego, Richard Ratliff, and Sergey Shebalov. "A general attraction model and sales-based linear program for network revenue management under customer choice". In: *Operations Research* 63.1 (2014), pp. 212–232.

[18] Guillermo Gallego and Huseyin Topaloglu. "Constrained assortment optimization for the nested logit model". In: *Management Science* 60.10 (2014), pp. 2583–2601.

[19] Guillermo Gallego and Ruxian Wang. "Multiproduct Price Optimization and Competition Under the Nested Logit Model with Product-Differentiated Price Sensitivities". In: *Operations Research* 62.2 (2014), pp. 450–461.

[20] Ward Hanson and Kipp Martin. "Optimizing multinomial logit profit functions". In: *Management Science* 42.7 (1996), pp. 992–1003.

[21] Wallace J Hopp and Xiaowei Xu. "Product line selection and pricing with modularity in design". In: *Manufacturing & Service Operations Management* 7.3 (2005), pp. 172–187.

[22] Woonghee Tim Huh and Hongmin Li. "Technical Note–Pricing Under the Nested Attraction Model with a Multistage Choice Structure". In: *Operations Research* 63.4 (2015), pp. 840–850.

[23] Srikanth Jagabathula. "Assortment optimization under general choice". In: *Available at SSRN 2512831* (2014).

[24] Srikanth Jagabathula and Paat Rusmevichientong. "A nonparametric joint assortment and price choice model". In: *Available at SSRN 2286923* (2015).

[25] Hai Jiang, Xin Qi, and He Sun. "Choice-Based Recommender Systems: A Unified Approach to Achieving Relevancy and Diversity". In: *Operations Research* 62.5 (2014), pp. 973–993.

[26] A Gürhan Kök, Marshall L Fisher, and Ramnath Vaidyanathan. "Assortment planning: Review of literature and industry practice". In: *Retail Supply Chain Management*. Springer, 2015, pp. 175–236.

[27] A Gürhan Kök and Yi Xu. "Optimal and competitive assortments with endogenous pricing under hierarchical consumer choice models". In: *Management Science* 57.9 (2011), pp. 1546–1563.

[28] Sumit Kunnumkal and Huseyin Topaloglu. "A refined deterministic linear program for the network revenue management problem with customer choice behavior". In: *Naval Research Logistics (NRL)* 55.6 (2008), pp. 563–580.

[29] Guang Li and Paat Rusmevichientong. "A greedy algorithm for the two-level nested logit model". In: *Operations Research Letters* 42.5 (2014), pp. 319–324.

[30] Guang Li, Paat Rusmevichientong, and Huseyin Topaloglu. "The d-level nested logit model: Assortment and price optimization problems". In: *Operations Research* 62.2 (2015), pp. 325–342.

[31] Hongmin Li and Woonghee Tim Huh. "Pricing multiple products with the multinomial logit and nested logit models: Concavity and implications". In: *Manufacturing & Service Operations Management* 13.4 (2011), pp. 549–563.

[32] Bacel Maddah and Ebru K Bish. "Joint pricing, assortment, and inventory decisions for a retailer's product line". In: *Naval Research Logistics (NRL)* 54.3 (2007), pp. 315–330.

[33] Charles F Manski, Daniel McFadden, et al. *Structural analysis of discrete data with econometric applications*. Mit Press Cambridge, MA, 1981.

[34] Daniel McFadden. "Conditional logit analysis of qualitative choice behavior". In: (1973).

[35] Daniel McFadden. "Econometric models for probabilistic choice among products". In: *Journal of Business* (1980), S13–S29.

[36] Isabel Méndez-Díaz et al. "A Branch-and-Cut Algorithm for the Latent Class Logit Assortment Problem". In: *Electronic Notes in Discrete Mathematics* 36 (2010), pp. 383–390.

[37] W Zachary Rayfield, Paat Rusmevichientong, and Huseyin Topaloglu. "Approximation methods for pricing problems under the nested logit model with price bounds". In: *INFORMS Journal on Computing* 27.2 (2015), pp. 335–357.

[38] Paat Rusmevichientong, Zuo-Jun Max Shen, and David B Shmoys. "A PTAS for capacitated sum-of-ratios optimization". In: *Operations Research Letters* 37.4 (2009), pp. 230–238.

[39] Paat Rusmevichientong, Zuo-Jun Max Shen, and David B Shmoys. "Dynamic assortment optimization with a multinomial logit choice model and capacity constraint". In: *Operations research* 58.6 (2010), pp. 1666–1680.

[40] Paat Rusmevichientong and Huseyin Topaloglu. "Robust assortment optimization in revenue management under the multinomial logit choice model". In: *Operations Research* 60.4 (2012), pp. 865–882.

[41] Paat Rusmevichientong et al. “Assortment optimization under the multinomial logit model with random choice parameters”. In: *Production and Operations Management* 23.11 (2014), pp. 2023–2039.

[42] Garrett van Ryzin and Siddharth Mahajan. “On the relationship between inventory costs and variety benefits in retail assortments”. In: *Management Science* 45.11 (1999), pp. 1496–1509.

[43] Jing-Sheng Song and Zhengliang Xue. “Demand management and inventory control for substitutable products”. In: *Working Paper* (2007).

[44] Kalyan Talluri and Garrett van Ryzin. “Revenue management under a general discrete choice model of consumer behavior”. In: *Management Science* 50.1 (2004), pp. 15–33.

[45] Kenneth E Train. *Discrete choice methods with simulation*. Cambridge university press, 2009.

[46] Ider Tsevendorj. “Piecewise-convex maximization problems”. In: *Journal of Global Optimization* 21.1 (2001), pp. 1–14.

[47] Ruxian Wang. “Assortment management under the generalized attraction model with a capacity constraint”. In: *Journal of Revenue & Pricing Management* 12.3 (2013), pp. 254–270.

[48] Ruxian Wang. “Capacitated assortment and price optimization under the multinomial logit model”. In: *Operations Research Letters* 40.6 (2012), pp. 492–497.

[49] Ruxian Wang and Ozge Sahin. “The impact of consumer search cost on assortment planning and pricing”. In: *Available at SSRN 2503490* (2015).

[50] Yanqiao Wang and Zuo-Jun Max Shen. “Constrained Assortment and Price Optimization Problems Under the Tree Logit Model”. In: *Working Paper* (2015).

[51] Yanqiao Wang and Zuo-Jun Max Shen. “Joint Optimization of Capacitated Assortment and Pricing Problem Under the Tree Logit Model”. In: *Working Paper* (2017).

[52] Yanqiao Wang and Zuo-Jun Max Shen. “With No-purchase Options: Joint Constrained Assortment and Price Optimization Under the Nested Logit Model”. In: *Working Paper* (2017).

Appendices

Appendix A

Appendix to Chapter 1

A.1 Notation

For ease of reading, we summarize our notation as follows:

m The number of nests.

M $:= \{1, 2, ..., m\}$.

n_i The number of products in nest i.

N_i $:= \{1, 2, ..., n_i\}$.

N The total number of products.

$\mathbf{P}_i$ $:= (p_{i1}, p_{i2}, ..., p_{in_i})$.

$v_{ij}(p_{ij})$ The preference weight of product $\langle i, j \rangle$.

α_{ij} The price-independent deterministic utility.

β_{ij} The price-sensitivity parameter.

$\mathbb{S}_i$ The space limitation on node i.

$\mathscr{B}$ The set of basic nodes.

α_i The price-independent deterministic utility of product i.

β_i The price-sensitivity parameter of product i.

S_i $:= (S_{i1}, S_{i2}, ..., S_{in_i}) \in \{0, 1\}^{n_i}$.

$\mathbb{C}_i$ The cardinality constraint on nest i.

$\mathbb{S}_i$ The space constraint on nest i.

Z^* The optimal expected profit under cardinality/space constraints.

Z^α $:= \Pi(S^\alpha, \mathbf{P}^\alpha)$ and $\alpha Z^\alpha \geqslant Z^*$.

θ_i The node-specific adjusted markup for node i.

$\tilde{S}_i(\theta_i)$:= The optimal solution to problem (4.9).

$\mathscr{A}_i$ $:= \{\tilde{S}_i(\theta_i) : \theta_i \in \mathbb{R}_{\geqslant 0}\}$.

$\mathscr{F}_i(z)$ The optimal solution to problem (2.11).

$\hat{S}_i(z)$ $:= \tilde{S}_i(\mathcal{F}_i(z))$.

A.2 Technical Proof of Claims

Proof of Claim 1

Proof. We prove this claim by contradiction. Assume that $S_i^* \neq \emptyset$ and $R_i(S_i^*, \mathbf{P}_i^*) < Z^*$, then we construct a new assortment $\hat{S} = S^* \backslash \{S_i^*\}$ and price matrix $\hat{\mathbf{P}} = (\hat{\mathbf{P}}_1, ..., \hat{\mathbf{P}}_{i-1}, \vec{0}, \hat{\mathbf{P}}_{i+1}, ..., \hat{\mathbf{P}}_m)$. If $R_i(S_i^*, \mathbf{P}_i^*) < Z^*$, then the following inequality holds

$$\Pi(\hat{S}, \hat{\mathbf{P}}) = \frac{\sum_{i \in M} V_i(S_i^*, \mathbf{P}_i^*) R_i(S_i^*, \mathbf{P}_i^*) - V_i(S_i^*, \mathbf{P}_i^*) R_i(S_i^*, \mathbf{P}_i^*)}{v_0 + \sum_{i \in M} V_i(S_i^*, \mathbf{P}_i^*) - V_i(S_i^*, \mathbf{P}_i^*)}$$
$$> \frac{\sum_{i \in M} V_i(S_i^*, \mathbf{P}_i^*) R_i(S_i^*, \mathbf{P}_i^*)}{v_0 + \sum_{i \in M} V_i(S_i^*, \mathbf{P}_i^*)} = \Pi(S^*, \mathbf{P}^*),$$

which contradicts with the fact that $(S^*, \mathbf{P}^*)$ is the optimal solution to problem (4.1). □

Proof of Claim 2

Proof. By using the notation that is defined in Section 2.3, we have

$$Z^* = \frac{\sum_{i \in M} V_i(S_i^*, \mathbf{P}_i^*) R_i(S_i^*, \mathbf{P}_i^*)}{v_0 + \sum_{i \in M} V_i(S_i^*, \mathbf{P}_i^*)},$$

which implies that $v_0 Z^* = \sum_{i \in M} V_i(S_i^*, \mathbf{P}_i^*)[R_i(S_i^*, \mathbf{P}_i^*) - Z^*]$. Since $(\hat{S}_i, \hat{\mathbf{P}}_i)$ is optimal to the joint subproblem at nest i, we have $V_i(\hat{S}_i, \hat{\mathbf{P}}_i)(R_i(\hat{S}_i, \hat{\mathbf{P}}_i) - Z^*) \geqslant V_i(S_i^*, \mathbf{P}_i^*)(R_i(S_i^*, \mathbf{P}_i^*) - Z^*)$. Therefore, we have $v_0 Z^* \leqslant \sum_{i \in M} V_i(\hat{S}_i, \hat{\mathbf{P}}_i)(R_i(\hat{S}_i, \hat{\mathbf{P}}_i) - Z^*)$. It follows that $Z^* \leqslant \Pi(\hat{S}_1, \hat{S}_2, ..., \hat{S}_m; \hat{\mathbf{P}}_1, \hat{\mathbf{P}}_2, ..., \hat{\mathbf{P}}_m)$. On the other hand, $(\hat{S}_1, \hat{S}_2, ..., \hat{S}_m; \hat{\mathbf{P}}_1, \hat{\mathbf{P}}_2, ..., \hat{\mathbf{P}}_m)$ is a feasible solution to problem (4.1) since $S_i \in \Im_i$ and $\mathbf{P}_i \geqslant 0$, thus we get $Z^* \geqslant \Pi(\hat{S}_1, \hat{S}_2, ..., \hat{S}_m; \hat{\mathbf{P}}_1, \hat{\mathbf{P}}_2, ..., \hat{\mathbf{P}}_m)$. Therefore, we have $Z^* = \Pi(\hat{S}_1, \hat{S}_2, ..., \hat{S}_m; \hat{\mathbf{P}}_1, \hat{\mathbf{P}}_2, ..., \hat{\mathbf{P}}_m)$, it follows that $(\hat{S}_1, \hat{S}_2, ..., \hat{S}_m; \hat{\mathbf{P}}_1, \hat{\mathbf{P}}_2, ..., \hat{\mathbf{P}}_m)$ is an optimal solution to problem (4.1). □

Proof of Claim 3

Proof. The proof technique follows partially from Lemma 3 in [18], the difference from which is that we consider the joint optimization problem. For completeness purpose, we show the proof as follows.

We first prove $\Pi(\tilde{S}_1, \tilde{S}_2, ..., \tilde{S}_m; \tilde{\mathbf{P}}_1, \tilde{\mathbf{P}}_2, ..., \tilde{\mathbf{P}}_m) \geqslant Z^*$, then show $\Pi(\tilde{S}_1, \tilde{S}_2, ..., \tilde{S}_m; \tilde{\mathbf{P}}_1, \tilde{\mathbf{P}}_2, ..., \tilde{\mathbf{P}}_m) \leqslant Z^*$, which implies that $\Pi(\tilde{S}_1, \tilde{S}_2, ..., \tilde{S}_m; \tilde{\mathbf{P}}_1, \tilde{\mathbf{P}}_2, ..., \tilde{\mathbf{P}}_m) = Z^*$. It indicates that $(\tilde{S}_1, \tilde{S}_2, ..., \tilde{S}_m; \tilde{\mathbf{P}}_1, \tilde{\mathbf{P}}_2, ..., \tilde{\mathbf{P}}_m)$ is an optimal solution to problem (4.1).

In order to show $\Pi(\tilde{S}_1, \tilde{S}_2, ..., \tilde{S}_m; \tilde{\mathbf{P}}_1, \tilde{\mathbf{P}}_2, ..., \tilde{\mathbf{P}}_m) \geqslant Z^*$, we only need to show that $V_i(\tilde{S}_i, \tilde{\mathbf{P}}_i)(R_i(\tilde{S}_i, \tilde{\mathbf{P}}_i) - Z^*) \geqslant V_i(S_i^*, \mathbf{P}_i^*)(R_i(S_i^*, \mathbf{P}_i^*) - Z^*)$ due to the proof of Claim 2. Next we show this inequality is true. First, if S_i^* is empty, then $\theta_i^* = Z^*$ and we have $V_i(\tilde{S}_i, \tilde{\mathbf{P}}_i)^{1/\gamma_i}(R_i(\tilde{S}_i, \tilde{\mathbf{P}}_i) - Z^*) \geqslant V_i(\varnothing, \tilde{\mathbf{P}}_i)^{1/\gamma_i}(R_i(\varnothing, \tilde{\mathbf{P}}_i) - Z^*) = 0$. It implies that either $V_i(\tilde{S}_i, \tilde{\mathbf{P}}_i) = 0$ or $V_i(\tilde{S}_i, \tilde{\mathbf{P}}_i) > 0$ and $R_i(\tilde{S}_i, \tilde{\mathbf{P}}_i) - Z^* \geqslant 0$, which indicates that the target inequality holds. Second, if S_i^* is nonempty, then $\theta_i^* = \gamma_i Z^* + (1-\gamma_i)R_i(S_i^*, \theta_i^*)$. We have $V_i(\tilde{S}_i, \tilde{\mathbf{P}}_i)^{1/\gamma_i}(R_i(\tilde{S}_i, \tilde{\mathbf{P}}_i) - \gamma_i Z^* - (1-\gamma_i)R_i(S_i^*, \theta_i^*)) \geqslant V_i(S_i^*, \mathbf{P}_i^*)^{1/\gamma_i}(R_i(S_i^*, \mathbf{P}_i^*) - \theta_i^*)$ since $(S_i^*, \mathbf{P}_i^*)$ is a feasible solution to problem (2.7). We obtain

$$
\begin{aligned}
R_i(\tilde{S}_i, \tilde{\mathbf{P}}_i) - Z^* &\geqslant \left(\frac{V_i(S_i^*, \mathbf{P}_i^*)}{V_i(\tilde{S}_i, \tilde{\mathbf{P}}_i)}\right)^{1/\gamma_i} (R_i(S_i^*, \mathbf{P}_i^*) - \theta_i^*) + (1-\gamma_i)(R_i(S_i^*, \mathbf{P}_i^*) - Z^*) \\
&\geqslant [\gamma_i \left(\frac{V_i(S_i^*, \mathbf{P}_i^*)}{V_i(\tilde{S}_i, \tilde{\mathbf{P}}_i)}\right)^{1/\gamma_i} + (1-\gamma_i)](R_i(S_i^*, \mathbf{P}_i^*) - Z^*) \\
&\geqslant \frac{V_i(S_i^*, \mathbf{P}_i^*)}{V_i(\tilde{S}_i, \tilde{\mathbf{P}}_i)}(R_i(S_i^*, \mathbf{P}_i^*) - Z^*),
\end{aligned}
$$

where the last inequality holds because of the subgradient inequality and $R_i(S_i^*, \mathbf{P}_i^*) \geqslant Z^*$ due to Claim 1. □

Appendix B

Appendix to Chapter 2

B.1 Notation

For ease of reading, we summarize our notation as follows:

m	The number of levels in the tree structure.
n	The number of products.
k	The maximum number of products of within a basic node.
N_i	The set of products that are associated with node i.
S_i	The assortment of node i.
$\mathbf{P}_i$	The price vector of assortment S_i.
$\Im_i$	The set of feasible assortments that satisfies cardinality/space constraints.
$\mathbb{C}_i$	The cardinality limitation on node i.
$\mathbb{S}_i$	The space limitation on node i.
$\mathscr{B}$	The set of basic nodes.
α_i	The price-independent deterministic utility of product i.
β_i	The price-sensitivity parameter of product i.
γ_i	The dissimilarity parameter for nonleaf node i.
$\Im^{\alpha}_{\text{root}}$	The collection of feasible candidate assortments that includes S^{α}_{root}.
Z^*	The optimal expected profit under cardinality/space constraints.
Z^{α}	$:= \max_{S_{\text{root}} \subseteq \Im^{\alpha}_{\text{root}}} R_{\text{root}}(S_{\text{root}})$, and $\alpha Z^{\alpha} \geqslant Z^*$.
θ_i	The node-specific adjusted markup for node i.
$\tilde{S}^{\alpha}_i(t_i)$	$:= \arg\max_{S_i \subseteq \mathscr{A}^{\alpha}_i} \{V_i(S_i)^{1/\gamma_i} (R_i(S_i) - t_i)\}$.
$\tilde{\mathscr{A}}^{\alpha}_i$	$:= \{\tilde{S}^{\alpha}_i(t_i) : t_i \in \mathbb{R}\}$.
$\hat{S}^{\alpha}_i(t_h)$	$:= \arg\max_{S_i \subseteq \widetilde{\mathscr{A}}^{\alpha}_i} \{V_i(S_i) (R_i(S_i) - t_h)\}$.
$\mathscr{A}^{\alpha}_h$	$:= \{\bigcup_{i \in h_C} \hat{S}^{\alpha}_i(t_h) : t_h \in \mathbb{R}\}$.

B.2 Technical Proofs

PROOF OF LEMMA 8

Proof. Let $\tilde{S}_i$ be optimal to problem (3.4) at node i, then we have

$$V_i(\tilde{S}_i)^{1/\gamma_i}\left(R_i(\tilde{S}_i) - t_i^\alpha\right) \geqslant V_i(S_i^\alpha)^{1/\gamma_i}\left(R_i(S_i^\alpha) - t_i^\alpha\right).$$

Because of Claim 6, we have

$$V_i(\tilde{S}_i)\Big(R_i(\tilde{S}_i) - t_h^\alpha\Big) \geqslant V_i(S_i^\alpha)\Big(R_i(S_i^\alpha) - t_h^\alpha\Big) = \max_{S_i \subseteq \Im_i^\alpha}\{V_i(S_i)\left(R_i(S_i) - t_h^\alpha\right)\},$$

where the last equality is due to Claim 8. Therefore, $\tilde{S}_i$ is also optimal to problem (3.3), which completes the proof. □

PROOF OF LEMMA 9

Proof. The proof of this lemma is fairly straightforward. Let $h = i^P$ in level $m-2$ denote the parent node of i, then according to Claim 11, since assortment $\tilde{S}_i^\alpha$ for the basic node i satisfies $V_i(\tilde{S}_i^\alpha)^{1/\gamma_i}\left(\alpha R_i(\tilde{S}_i^\alpha) - t_i^*\right) \geqslant V_i(S_i^*)^{1/\gamma_i}\Big(R_i(S_i^*) - t_i^*\Big)$, then according to Claim 10, we have $V_i(\tilde{S}_i^\alpha)\left(\alpha R_i(\tilde{S}_i^\alpha) - t_h^*\right) \geqslant V_i(S_i^*)\Big(R_i(S_i^*) - t_h^*\Big) = \max_{S_i \subseteq \Im_i^*}\{V_i(S_i)(R_i(S_i) - t_h^*)\}$, where the last equality is due to Claim 8 by letting $\alpha = 1$. If we define $\tilde{S}_h^\alpha = \bigcup_{i \in h} \tilde{S}_i^\alpha$, then for any node h in level $m-2$, we have $V_h(\tilde{S}_h^\alpha)\Big(\alpha R_h(\tilde{S}_h^\alpha) - t_{h^P}^*\Big) \geqslant \max_{S_h \subseteq \Im_h^*}\{V_h(S_h)(R_h(S_h) - t_{h^P}^*)\}$ according to Claim 11. Take the union of α-approximate assortments of lower level nodes repeatedly, until we have $\tilde{S}_{\text{root}}^\alpha = \bigcup_{i \in \mathscr{B}} \tilde{S}_i^\alpha$, then due to Claim 11, $\tilde{S}_{\text{root}}^\alpha$ satisfies $\alpha R_{\text{root}}(\tilde{S}_{\text{root}}^\alpha) \geqslant \max_{S_{\text{root}} \subseteq \Im_{\text{root}}^*} R_{\text{root}}(S_{\text{root}}) = Z^*$. Thus $S_{\text{root}}^\alpha = \tilde{S}_{\text{root}}^\alpha = \bigcup_{i \in \mathscr{B}} \tilde{S}_i^\alpha$ is an α-approximate solution, establishing this lemma. □

PROOF OF PROPOSITION 9

Proof. For notational brevity, we omit the assortment S, let $\delta_j = \delta_j(\theta_j)$, $\omega_j = \omega_j(\theta_j)$ for all $j \in V$. We first prove θ_i and θ_i' are equivalent. We denote the actual product in the lowest level m as g. Without loss of generality, when node i is in level l $(0 \leqslant l \leqslant m-2)$, we assume that node i and $j \in \text{Children}(i)$ are ancestors of g, and when node i is in level $m-1$, node

j is the sibling node of g. From Lemma 11, we have

$$\begin{aligned}
&\partial R_{\text{root}}(\mathbf{P}_{\text{root}})/\partial p_g \\
&= \beta_g Q_g \Bigg(\frac{1}{\beta_g} - m_g + \lambda_{g,1}^{m-1} \sum_{g' \in \eta_{g,0}} m_{g'} Q_{g'} + \sum_{k=1}^{m-2} \sum_{\substack{g' \in \eta_{g,k} \\ g' \notin \eta_{g,k+1}}} m_{g'} \Bigg(\sum_{t=1}^{k} (\lambda_{g,t+1}^{m-1} - \lambda_{g,t}^{m-1}) Q(g'|\eta_{g,t}) \Bigg) \\
&+ \sum_{g' \in \eta_{g,m-1}} m_{g'} \Bigg(\sum_{t=1}^{m-1} (\lambda_{g,t+1}^{m-1} - \lambda_{g,t}^{m-1}) Q(g'|\eta_{g,t}) \Bigg) \Bigg),
\end{aligned}$$

We let the first derivative $\partial R_{\text{root}}(\mathbf{P}_{\text{root}})/\partial p_g = 0$, since $Q_g \neq 0$, after dividing $\beta_g Q_g$ and collecting terms, we have

$$\begin{aligned}
&1/\beta_g - m_g + \lambda_{g,1}^{m-1} \sum_{g' \in \eta_{g,0}} m_{g'} Q_{g'} + \sum_{k=1}^{m-2} \sum_{\substack{g' \in \eta_{g,k} \\ g' \notin \eta_{g,k+1}}} m_{g'} \left(\sum_{t=1}^{k} (\lambda_{g,t+1}^{m-1} - \lambda_{g,t}^{m-1}) Q(g'|\eta_{g,t}) \right) \\
&+ \sum_{g' \in \eta_{g,m-1}} m_{g'} \left(\sum_{t=1}^{m-1} (\lambda_{g,t+1}^{m-1} - \lambda_{g,t}^{m-1}) Q(g'|\eta_{g,t}) \right) = 0. \qquad \text{(B.1)}
\end{aligned}$$

Assume node i is in level l ($0 \leqslant l \leqslant m-1$), then according to (C.17) and the definition of node-specific adjusted markup, we have

$$\begin{aligned}
\theta_i &= \theta_j \delta_j - \omega_j \\
&= \lambda_{g,1}^{l} \sum_{g' \in \eta_{g,0}} m_{g'} Q_{g'} + \sum_{k=1}^{m-2} \sum_{\substack{g' \in \eta_{g,k} \\ g' \notin \eta_{g,k+1}}} m_{g'} \left(\sum_{t=1}^{l} (\lambda_{g,t+1}^{l} - \lambda_{g,t}^{l}) Q(g'|\eta_{g,t}) \right) \\
&+ \sum_{g' \in \eta_{g,m-1}} m_{g'} \left(\sum_{t=1}^{l} (\lambda_{g,t+1}^{l} - \lambda_{g,t}^{l}) Q(g'|\eta_{g,t}) \right). \qquad \text{(B.2)}
\end{aligned}$$

From the RHS of (C.19), we can see that θ_i is independent of $\eta_{g,s}$ where $0 \leqslant s \leqslant l$. Thus for another child node j' of node i, $\theta_i' = m_{j'} - \omega_{j'}$ also equals to the RHS of (C.19). Therefore, $\theta_i = \theta_i'$ for any node i in level $0, 1, ..., m-1$.

From Lemma 12, we obtain

$$\frac{\partial \theta_i}{\partial \theta_j} = \begin{cases} 1 & j \text{ is a leaf node} \\ \frac{1}{\gamma_j} - \omega_j u_j & \text{o.w.} \end{cases} > 0$$

Therefore, there exists a one-to-one increasing correspondence between the node-specific adjusted markup θ_i and θ_j, where $j \in$ Children(i). □

Proof. According to Lemma 11, the first derivative of the objective function $R_{\text{root}}(S_{\text{root}}, P_{\text{root}})$ with respect to price p_g for any product g is

$$\begin{aligned}
&\partial R_{\text{root}}(\mathbf{P}_{\text{root}})/\partial p_g \\
&= \beta_g Q_g \Bigg(\frac{1}{\beta_g} - m_g + \lambda_{g,1}^{m-1} \sum_{g' \in \eta_{g,0}} m_{g'} Q_{g'} + \sum_{k=1}^{m-2} \sum_{\substack{g' \in \eta_{g,k} \\ g' \notin \eta_{g,k+1}}} m_{g'} \Bigg(\sum_{t=1}^{k} (\lambda_{g,t+1}^{m-1} - \lambda_{g,t}^{m-1}) Q(g'|\eta_{g,t}) \Bigg) \\
&+ \sum_{g' \in \eta_{g,m-1}} m_{g'} \Bigg(\sum_{t=1}^{m-1} (\lambda_{g,t+1}^{m-1} - \lambda_{g,t}^{m-1}) Q(g'|\eta_{g,t}) \Bigg) \Bigg).
\end{aligned}$$

Use the notation for the *node-specific adjusted markup* in Definition 1 and according to Proposition 9, we have

$$\begin{aligned}
\partial R_{\text{root}}(\mathbf{P}_{\text{root}})/\partial p_g &= \lambda_{g,1}^{m-1} \beta_g Q_g \sum_{g' \in \eta_{g,0}} m_{g'} Q_{g'} + \beta_g Q_g (-\lambda_{g,1}^{m-1} \theta_{g,0}) \\
&= \lambda_{g,1}^{m-1} \beta_g Q_g R(\theta_{\text{root}}) + \beta_g Q_g (-\lambda_{g,1}^{m-1} \theta_{\text{root}}) = \lambda_{g,1}^{m-1} \beta_g Q_g \left(R(\theta_{\text{root}}) - \theta_{\text{root}} \right).
\end{aligned}$$

For product g in the lowest level m, we use $\text{An}(g, l)$ to denote the g's ancestor node in level l. Then we obtain

$$\begin{aligned}
&\frac{\partial R_{\text{root}}(\theta_{\text{root}})}{\partial \theta_{\text{root}}} \\
&= \sum_{g \in \eta_{g,0}} \frac{\partial R_{\text{root}}(S_{\text{root}}, P_{\text{root}})}{\partial p_g} \left(\frac{\partial \theta_{\text{An}(g,m-1)}}{\partial p_g} \right)^{-1} \left(\frac{\partial \theta_{\text{An}(g,m-2)}}{\partial \theta_{\text{An}(g,m-1)}} \right)^{-1} \cdots \left(\frac{\partial \theta_{\text{An}(g,0)}}{\partial \theta_{\text{An}(g,1)}} \right)^{-1} \left(\frac{\partial \theta_{\text{root}}}{\partial \theta_{\text{An}(g,0)}} \right)^{-1} \\
&= (R(\theta_{\text{root}}) - \theta_{\text{root}}) \sum_{g \in \eta_{g,0}} \lambda_{g,1}^{m-1} \beta_g Q_g \left(\frac{\partial \theta_{\text{An}(g,m-1)}}{\partial p_g} \right)^{-1} \left(\frac{\partial \theta_{\text{An}(g,m-2)}}{\partial \theta_{\text{An}(g,m-1)}} \right)^{-1} \cdots \left(\frac{\partial \theta_{\text{An}(g,0)}}{\partial \theta_{\text{An}(g,1)}} \right)^{-1}.
\end{aligned}$$

According to Proposition 9, we know

$$\left(\frac{\partial \theta_{\text{An}(g,m-1)}}{\partial p_g} \right)^{-1} \left(\frac{\partial \theta_{\text{An}(g,m-2)}}{\partial \theta_{\text{An}(g,m-1)}} \right)^{-1} \cdots \left(\frac{\partial \theta_{\text{An}(g,0)}}{\partial \theta_{\text{An}(g,1)}} \right)^{-1} > 0.$$

Therefore we can see that $R_{\text{root}}(\theta_{\text{root}})$ is strictly unimodal with respect to θ_{root} according to Lemma 2 in [19]. Moreover, let θ^*_{root} denote the solution to $R(\theta_{\text{root}}) = \theta_{\text{root}}$, then $\partial R_{\text{root}}(\theta^*_{\text{root}})/\partial \theta^*_{\text{root}} = 0$. Thus according to the unimodality of $R(\theta_{\text{root}})$, we have $R(\theta^*_{\text{root}}) = \theta^*_{\text{root}}$ at optimality. □

PROOF OF CLAIM 6

Claim 6. *For an arbitrary nonleaf node $i \in V$, assume $R_i(S_i^\alpha) \geqslant t_h^\alpha$ with parameter $\alpha \geqslant 1$. If there exists an assortment $\hat{S}_i \subseteq \Im_i^\alpha$ such that*

$$V_i(\hat{S}_i)^{1/\gamma_i} \Big(R_i(\hat{S}_i) - t_i^\alpha \Big) \geqslant V_i(S_i^\alpha)^{1/\gamma_i} \Big(R_i(S_i^\alpha) - t_i^\alpha \Big), \tag{B.3}$$

then we have

$$V_i(\hat{S}_i)\Big(R_i(\hat{S}_i) - t_h^\alpha\Big) \geqslant V_i(S_i^\alpha)\Big(R_i(S_i^\alpha) - t_h^\alpha\Big). \tag{B.4}$$

If the inequality in (B.3) is strict for some $j \in i_C$, so is the inequality in (B.4).

Proof. When $\hat{S}_i = \emptyset$, according to the above inequality, we have $R_i(S_i^\alpha) - t_i^\alpha \leqslant 0$, implying $R_i(S_i^\alpha) = t_i^\alpha = t_h^\alpha$. So (B.4) holds for $\hat{S}_i = \emptyset$ (Both of the LHS and RHS are zero). When $\hat{S}_i \neq \emptyset$, then $V_i(\hat{S}_i) \neq 0$ so we can divide the above inequality by $V_i(\hat{S}_i)^{1/\gamma_i}$. Because $R_i(S_i^\alpha) \geqslant t_h^\alpha$, then according to (3.2), $t_i^\alpha = \gamma_i t_h^\alpha + (1-\gamma_i)R_i(S_i^\alpha)$. Thus we have

$$\begin{aligned} R_i(\hat{S}_i) - t_h^\alpha &\geqslant \left[\frac{V_i(S_i^\alpha)}{V_i(\hat{S}_i)}\right]^{1/\gamma_i} (R_i(S_i^\alpha) - t_i^\alpha) + t_i^\alpha - t_h^\alpha \\ &= \left(\gamma_i \left[\frac{V_i(S_i^\alpha)}{V_i(\hat{S}_i)}\right]^{1/\gamma_i} + 1 - \gamma_i\right)(R_i(S_i^\alpha) - t_h^\alpha) \geqslant \frac{V_i(S_i^\alpha)}{V_i(\hat{S}_i)}(R_i(S_i^\alpha) - t_h^\alpha) \end{aligned}$$

For the last inequality, let's consider a convex function $f(x) = x^{1/\gamma_i}$, according to its Taylor expansion at $x = 1$, we have $f(x) = x^{1/\gamma_i} \geqslant 1 + (x-1)/\gamma_i \Rightarrow \gamma_i x^{1/\gamma_i} + 1 - \gamma_i \geqslant x$. Let $x = V_i(S_i^\alpha)/V_i(\hat{S}_i)$, establishing the inequality, thus $V_i(\hat{S}_i)(R_i(\hat{S}_i) - t_h^\alpha) \geqslant V_i(S_i^\alpha)(R_i(S_i^\alpha) - t_h^\alpha)$. It is also easy to check if the inequality in (B.3) is strict for some $j \in i_C$, then the inequality in (B.4) is also strict. □

PROOF OF CLAIM 7

Claim 7. *For an arbitrary nonleaf node $i \in V$, assume $R_i(S_i^\alpha) \geqslant t_h^\alpha$ with parameter $\alpha \geqslant 1$. If for all $j \in i_C$, there exists an assortment $\hat{S}_j \subseteq \Im_j^\alpha$ such that*

$$V_j(\hat{S}_j)\Big(R_j(\hat{S}_j) - t_i^\alpha\Big) \geqslant V_j(S_j^\alpha)\Big(R_j(S_j^\alpha) - t_i^\alpha\Big). \tag{B.5}$$

Define $h = i^P$ and let $\hat{S}_i = \bigcup_{j \in i_C} \hat{S}_j$, then we have

$$V_i(\hat{S}_i)\Big(R_i(\hat{S}_i) - t_h^\alpha\Big) \geqslant V_i(S_i^\alpha)\Big(R_i(S_i^\alpha) - t_h^\alpha\Big). \tag{B.6}$$

If the inequality in (B.5) is strict for some $j \in i_C$, so is the inequality in (B.6).

Proof. The logic of proving this lemma, the following proposition and theorem is similar to [30], for completeness, we provide the entire proof as follows. We first claim $S_i^\alpha \neq \emptyset$. It is true when i = root, then for all $i \in V\backslash$root, the scalar defined in (3.2) is either a positive value or $+\infty$, thus $R_i(S_i^\alpha) \geqslant t_i^\alpha \geqslant t_h^\alpha > 0$, which implies $S_i^\alpha \neq \emptyset$, establishing the claim. So we have $V_{i0}\mathbf{1}(\hat{S}_i \neq \emptyset) \leqslant V_{i0} = V_{i0}\mathbf{1}(S_i^\alpha \neq \emptyset)$. Since $V_j(\hat{S}_j)\Big(R_j(\hat{S}_j) - t_i^\alpha\Big) \geqslant V_j(S_j^\alpha)\Big(R_j(S_j^\alpha) - t_i^\alpha\Big)$, then

$$\begin{aligned}V_i(\hat{S}_i)^{1/\gamma_i}(R_i(\hat{S}_i) - t_i^\alpha) &= \sum_{j \in i_C} V_j(\hat{S}_j)(R_j(\hat{S}_j) - t_i^\alpha) - V_{i0}\mathbf{1}(\hat{S}_i \neq \varnothing)t_i^\alpha \\ &\geqslant \sum_{j \in i_C} V_j(S_j^\alpha)(R_j(S_j^\alpha) - t_i^\alpha) - V_{i0}\mathbf{1}(S_i^\alpha \neq \varnothing)t_i^\alpha = V_i(S_i^\alpha)^{1/\gamma_i}(R_i(S_i^\alpha) - t_i^\alpha)\end{aligned}$$

This claim holds because of Claim 6. □

PROOF OF CLAIM 8

Claim 8. *Let $S^\alpha = \{S_i^\alpha,\ \forall i \in V\}$ be optimal to problem (3.1). Then for all $i \in V$, S_i^α is an optimal solution to problem (3.3) at node i.*

Proof. Prove by induction on levels. It is true for root, then we assume that it is true for any node i in level l: $V_i(S_i^\alpha)(R_i(S_i^\alpha) - t_h^\alpha) = \max_{S_i \subseteq \Im_i^\alpha}\{V_i(S_i)(R_i(S_i) - t_h^\alpha)\}$.

Suppose on the contrary that there exists node $k \in i_C$ in level $l+1$ such that $V_k(S_k^\alpha)(R_k(S_k^\alpha) - t_i^\alpha) < \max_{S_k \subseteq \Im_k^\alpha}\{V_k(S_k)(R_k(S_k) - t_i^\alpha)\}$. For each $j \in i_C$, define $\hat{S}_j = \arg\max_{S_j \subseteq \Im_j^\alpha}\{V_j(S_j)(R_j(S_j) - t_i^\alpha)\}$. First, consider the case when $R_i(S_i^\alpha) \geqslant t_h^\alpha$. We have $S_j^\alpha \subseteq \Im_j^\alpha$ for all $j \in i_C$, then by construction: $V_j(\hat{S}_j)\Big(R_j(\hat{S}_j) - t_i^\alpha\Big) \geqslant V_j(S_j^\alpha)\Big(R_j(S_j^\alpha) - t_i^\alpha\Big)$, which is strict if $j = k$. Let $\hat{S}_i = \bigcup_{j \in i_C} \hat{S}_j$, then according to Claim 7, we have $V_i(\hat{S}_i)\Big(R_i(\hat{S}_i) - t_h^\alpha\Big) > V_i(S_i^\alpha)\Big(R_i(S_i^\alpha) - t_h^\alpha\Big)$. It contradicts the induction hypothesis: $V_i(S_i^\alpha)(R_i(S_i^\alpha) - t_h^\alpha) = \max_{S_i \subseteq \Im_i^\alpha}\{V_i(S_i)(R_i(S_i) - t_h^\alpha)\}$.

Then let's consider $R_i(S_i^\alpha) < t_h^\alpha$. Since $S_i = \varnothing$ is a feasible solution to the local problem and its objective value is 0. Thus $\max_{S_i \subseteq \Im_i^\alpha}\{V_i(S_i)(R_i(S_i) - t_h^\alpha)\} \geqslant 0$. Because node i is in level l, according to the induction hypothesis, we have $V_i(S_i^\alpha)(R_i(S_i^\alpha) - t_h^\alpha) = \max_{S_i \subseteq \Im_i^\alpha}\{V_i(S_i)(R_i(S_i) - t_h^\alpha)\} \geqslant 0$. However, we know $R_i(S_i^\alpha) < t_h^\alpha$, thus we have $V_i(S_i^\alpha)(R_i(S_i^\alpha) - t_h^\alpha) \leqslant 0$, so $V_i(S_i^\alpha) = 0$, which implies $S_i^\alpha = \varnothing$. Since $R_i(S_i^\alpha) < t_h^\alpha$, then according to Equation (3.2), $t_i^\alpha = +\infty$. Thus we have $0 = \max_{S_k \subseteq \Im_k^\alpha}\{V_k(S_k)(R_k(S_k) - t_i^\alpha)\}$. According to our assumption, $V_k(S_k^\alpha)(R_k(S_k^\alpha) - t_i^*) < \max_{S_k \subseteq \Im_k^\alpha}\{V_k(S_k)(R_k(S_k) - t_i^\alpha)\} = 0$, which implies $S_k^\alpha \neq \varnothing$. However, since $k \in i_C$, so $S_i^\alpha \supseteq S_k^\alpha$, which contradicts with $S_i^\alpha = \varnothing$. □

PROOF OF CLAIM 9

Claim 9. *For any nonleaf node i, if $\hat{S}_j$ is the optimal solution to problem (3.3) at node j for all $j \in i_C$, then $\hat{S}_i = \bigcup_{j \in i_C} \hat{S}_j$ is an optimal solution to problem (3.3) at node i.*

Proof. For any $j \in i_C$, since $\hat{S}_j$ is optimal to local problem at node j, then we have $V_j(\hat{S}_j)(R_j(\hat{S}_j) - t_i^\alpha) \geqslant V_j(S_j^\alpha)(R_j(S_j^\alpha) - t_i^\alpha)$.

When $R_i(S_i^\alpha) \geqslant t_h^\alpha$, let $\hat{S}_i = \bigcup_{j \in i_C} \hat{S}_j$, then according to Claim 7: $V_i(\hat{S}_i)(R_i(\hat{S}_i) - t_h^\alpha) \geqslant V_i(S_i^\alpha)(R_i(S_i^\alpha) - t_h^\alpha) = \max_{S_i \subseteq \Im_i^\alpha}\{V_i(S_i)(R_i(S_i) - t_h^\alpha)\}$. The last equality holds due to Claim 8. Therefore, $\hat{S}_i$ is the local optimal solution at node i when $R_i(S_i^\alpha) \geqslant t_h^\alpha$.

Then for $R_i(S_i^\alpha) < t_h^\alpha$, we have $t_i^\alpha = +\infty$ according to Equation (3.2). Thus for all $j \in i_C$, $\varnothing = \hat{S}_j = \arg\max_{S_j \subseteq \Im_j^\alpha} \{V_j(S_j^\alpha)(R_j(S_j^\alpha) - t_i^\alpha)\}$. Then $\hat{S}_i = \bigcup_{j \in i_C} \hat{S}_j = \varnothing$. As discussed in the proof of Claim 8, we can see that $S_i^\alpha = \varnothing$, which is the optimal solution to the local problem at node i. Therefore this lemma still holds when $R_i(S_i^\alpha) < t_h^\alpha$. □

PROOF OF CLAIM 10

Claim 10. *For all $i \in V$, let S_i^* be the optimal assortment at node i. For parameter $\alpha \geqslant 1$, if assortment $\widetilde{S}_i$ satisfies*

$$V_i(\widetilde{S}_i)^{1/\gamma_i}\left(\alpha R_i(\widetilde{S}_i) - t_i^*\right) \geqslant V_i(S_i^*)^{1/\gamma_i}\left(R_i(S_i^*) - t_i^*\right). \tag{B.7}$$

Define $h = i^P$, then we have

$$V_i(\widetilde{S}_i)\left(\alpha R_i(\widetilde{S}_i) - t_h^*\right) \geqslant V_i(S_i^*)\left(R_i(S_i^*) - t_h^*\right). \tag{B.8}$$

Proof. First consider $R_i(S_i^*) \geqslant t_h^*$. Due to the inequality (B.7), we can get $V_i(\widetilde{S}_i)^{1/\gamma_i}(\alpha R_i(\widetilde{S}_i) - t_i^*) \geqslant V_i(S_i^*)^{1/\gamma_i}(R_i(S_i^*) - t_i^*)$. Since $R_i(S_i^*) \geqslant t_h^*$, according to Equation (3.2), we have $t_i^* = \gamma_i t_h^* + (1-\gamma_i)R_i(S_i^*)$. After plugging t_i^* in the above inequality, we have

$$\begin{aligned} & V_i(\widetilde{S}_i)^{1/\gamma_i}(\alpha R_i(\widetilde{S}_i) - \gamma_i t_h^* - (1-\gamma_i)R_i(S_i^*)) \geqslant V_i(S_i^*)^{1/\gamma_i}(R_i(S_i^*) - \gamma_i t_h^* - (1-\gamma_i)R_i(S_i^*)) \\ \Rightarrow\quad & V_i(\widetilde{S}_i)^{1/\gamma_i}(\alpha R_i(\widetilde{S}_i) - t_h^*) \geqslant (\gamma_i V_i(S_i^*)^{1/\gamma_i} + (1-\gamma_i)V_i(\widetilde{S}_i)^{1/\gamma_i})(R_i(S_i^*) - t_h^*) \qquad \text{(B.9)} \end{aligned}$$

Multiply inequality (B.9) by $V_i(\widetilde{S}_i)^{1-1/\gamma_i}$, we get

$$V_i(\widetilde{S}_i)(\alpha R_i(\widetilde{S}_i) - t_h^*) \tag{B.10}$$

$$\geqslant V_i(\widetilde{S}_i)^{1-1/\gamma_i}(\gamma_i V_i(S_i^*)^{1/\gamma_i} + (1-\gamma_i)V_i(\widetilde{S}_i)^{1/\gamma_i})(R_i(S_i^*) - t_h^*) \tag{B.11}$$

Since $\gamma_i \in (0,1]$, due to the concavity of x^{γ_i}, we have $x^{\gamma_i} \leqslant \widetilde{x}^{\gamma_i} + \gamma_i \widetilde{x}^{\gamma_i - 1}(x - \widetilde{x}) = \widetilde{x}^{\gamma_i - 1}(\gamma_i x + (1-\gamma_i)\widetilde{x})$. Let $x = V_i(S_i^*)^{1/\gamma_i}$ and $\widetilde{x} = V_i(\widetilde{S}_i)^{1/\gamma_i}$, then we get

$$V_i(\widetilde{S}_i)^{1-1/\gamma_i}(\gamma_i V_i(S_i^*)^{1/\gamma_i} + (1-\gamma_i)V_i(\widetilde{S}_i)^{1/\gamma_i}) \geqslant V_i(S_i^*). \tag{B.12}$$

Since $R_i(S_i^*) > t_h^*$, multiply inequality (B.12) by $R_i(S_i^*) - t_h^*$, we have

$$V_i(\widetilde{S}_i)^{1-1/\gamma_i}(\gamma_i V_i(S_i^*)^{1/\gamma_i} + (1-\gamma_i)V(\widetilde{S}_i)^{1/\gamma_i})(R_i(S_i^*) - t_h^*) \tag{B.13}$$

$$\geqslant V_i(S_i^*)(R_i(S_i^*) - t_h^*) \tag{B.14}$$

Thus due to inequality (B.10) and inequality (B.13), we obtain $V_i(\widetilde{S}_i)(\alpha R_i(\widetilde{S}_i) - t_h^*) \geqslant V_i(S_i)(R_i(S_i^*) - t_h^*)$. Therefore, this theorem holds when $R_i(S_i^*) \geqslant t_h^*$. Now, let's see $R_i(S_i^*) < t_h^*$. According to previous discussion, we know $S_i^* = \varnothing$ and $t_i^* = +\infty$. Thus we have $V_i(\widetilde{S}_i)^{1/\gamma_i}(\alpha R_i(\widetilde{S}_i) - t_i^*) \geqslant V_i(S_i^*)^{1/\gamma_i}(R_i(S_i^*) - t_i^*) = 0$. Since $t_i^* = +\infty$, then $\widetilde{S}_i = \varnothing$. Therefore, inequality (B.8) still holds. □

PROOF OF CLAIM 11

Claim 11. *For an arbitrary nonleaf node i and parameter $\alpha \geqslant 1$, we assume that for all $j \in i_C$, there exists an assortment $\hat{S}_j \subseteq \Im_j$ such that*

$$V_j(\hat{S}_j)\Big(\alpha R_j(\hat{S}_j) - t_i^*\Big) \geqslant \max_{S_j \subseteq \Im_j^*} \{V_j(S_j)\,(R_j(S_j) - t_i^*)\}. \tag{B.15}$$

Define $h = i^P$ and let $\hat{S}_i = \bigcup_{j \in i_C} \hat{S}_j$, then we have

$$V_i(\hat{S}_i)\Big(\alpha R_i(\hat{S}_i) - t_h^*\Big) \geqslant \max_{S_i \subseteq \Im_i^*} \{V_i(S_i)\,(R_i(S_i) - t_h^*)\}. \tag{B.16}$$

Proof. According to Claim 8, $S_j^* = \arg\max_{S_j \subseteq \Im_j^*}\{V_j(S_j)(R_j(S_j) - t_i^*)\}$ and $S_i^* = \arg\max_{S_i \subseteq \Im_i^*}\{V_i(S_i)(R_i(S_i) - t_h^*)\}$. First when $R_i(S_i^*) \geqslant t_h^*$, then we have $S_i^* \neq \varnothing$ according to proof in Claim 7. Thus $\mathbf{1}(\hat{S}_i \neq \varnothing) \leqslant V_{i0} = V_{i0}\mathbf{1}(S_i^* \neq \varnothing)$. Since $V_j(\hat{S}_j)\Big(\alpha R_j(\hat{S}_j) - t_i^*\Big) \geqslant \max_{S_j \subseteq \Im_j^*}\{V_j(S_j)(R_j(S_j) - t_i^*)\} = V_j(S_j^*)\Big(R_j(S_j^*) - t_i^*\Big)$, then we have

$$\begin{aligned} V_i(\hat{S}_i)^{1/\gamma_i}(\alpha R_i(\hat{S}_i) - t_i^*) &= \sum_{j \in i_C} V_j(\hat{S}_j)(\alpha R_j(\hat{S}_j) - t_i^*) - V_{i0}\mathbf{1}(\hat{S}_i \neq \varnothing)t_i^* \\ &\geqslant \sum_{j \in i_C} V_j(S_j^*)(R_j(S_j^*) - t_i^*) - V_{i0}\mathbf{1}(S_i^* \neq \varnothing)t_i^* \\ &= V_i(S_i^*)^{1/\gamma_i}(R_i(S_i^*) - t_i^*) \end{aligned}$$

(B.16) holds for $\hat{S}_i = \varnothing$. When $\hat{S}_i \neq \varnothing$, because $R_i(S_i^*) \geqslant t_h^*$, then we know $t_i^* = \gamma_i t_h^* + (1 - \gamma_i)R_i(S_i^*)$ due to (3.2). Thus we have

$$\begin{aligned} \alpha R_i(\hat{S}_i) - t_h^* &\geqslant \left[\frac{V_i(S_i^*)}{V_i(\hat{S}_i)}\right]^{1/\gamma_i} (R_i(S_i^*) - t_i^*) + t_i^* - t_h^* \\ &= \left(\gamma_i \left[\frac{V_i(S_i^*)}{V_i(\hat{S}_i)}\right]^{1/\gamma_i} + 1 - \gamma_i\right)(R_i(S_i^*) - t_h^*) \geqslant \frac{V_i(S_i^*)}{V_i(\hat{S}_i)}(R_i(S_i^*) - t_h^*) \end{aligned}$$

The above inequality is equivalent to $V_i(\hat{S}_i)(\alpha R_i(\hat{S}_i) - t_h^*) \geqslant V_i(S_i^*)(R_i(S_i^*) - t_h^*) = \max_{S_i \subseteq \Im_i^*}\{V_i(S_i)(R_i(S_i) - t_h^*)\}$. This claim is true when $R_i(S_i^*) \geqslant t_h^*$.

Second, when $R_i(S_i^*) < t_h^*$, $S_i^* = \varnothing$ according to the proof in Claim 8. Because $R_i(S_i^*) < t_h^*$, then according to Equation (3.2), $t_i^* = +\infty$. Thus $S_j^* = \varnothing$ and $\hat{S}_j$ can only be empty to satisfy (B.15), which implies $\hat{S}_i = \bigcup_{j \in i_C} \hat{S}_j = \varnothing$ that satisfies (B.16). So it still holds when $R_i(S_i^*) < t_h^*$. □

PROOF OF LEMMA 11

Proof. We define the markup for product i as $m_i = p_i - c_i$. The first-order condition of the objective function $R_{\text{root}}(\mathbf{P}_{\text{root}})$ with respect to price p_i of product i is

$$\frac{\partial R_{\text{root}}(\mathbf{P}_{\text{root}})}{\partial p_i} = Q_i + m_i \frac{\partial Q_i}{\partial p_i} + \sum_{k=0}^{m-1} \sum_{\substack{i' \in \eta_{i,k} \\ i' \notin \eta_{i,k+1}}} m_{i'} \frac{\partial Q_{i'}}{\partial p_i}, \tag{B.17}$$

where we use Q_i to denote the choice probability of product i. We will derive $\partial Q_i / \partial p_i$ and $\partial Q_{i'} / \partial p_i$ in the remaining part of this proof, respectively. For all $i' \in \eta_{i,k}$ and $i' \notin \eta_{i,k+1}$ $(0 \leqslant k \leqslant m-1)$, $\partial Q_{i'} / \partial p_i$ can be calculated as

$$\begin{aligned}
\frac{\partial Q_{i'}}{\partial p_i} = \frac{\partial (\prod_{t=0}^{m-1} Q(\eta_{i',t+1}|\eta_{i',t}))}{\partial p_i} &= \sum_{t=0}^{m-1} \frac{\partial Q(\eta_{i',t+1}|\eta_{i',t})}{\partial p_i} * \frac{Q_{i'}}{Q(\eta_{i',t+1}|\eta_{i',t})} \\
&= \sum_{t=0}^{k} \frac{\partial Q(\eta_{i',t+1}|\eta_{i',t})}{\partial p_i} * \frac{Q_{i'}}{Q(\eta_{i',t+1}|\eta_{i',t})} \qquad \text{(B.18)} \\
&= \frac{\partial Q(\eta_{i',k+1}|\eta_{i,k})}{\partial p_i} * \frac{Q_{i'}}{Q(\eta_{i',k+1}|\eta_{i,k})} + \sum_{t=0}^{k-1} \frac{\partial Q(\eta_{i,t+1}|\eta_{i,t})}{\partial p_i} * \frac{Q_{i'}}{Q(\eta_{i,t+1}|\eta_{i,t})}. \qquad \text{(B.19)}
\end{aligned}$$

Equation (C.5) is due to the fact that $\partial Q(\eta_{i',t+1}|\eta_{i',t}) / \partial p_i = 0$, when $k+1 \leqslant t \leqslant m-1$. Equation (C.6) is established because $Q(\eta_{i',t+1}|\eta_{i',t}) = Q(\eta_{i,t+1}|\eta_{i,t})$ $(0 \leqslant t \leqslant k-1)$ when $i' \in \eta_{i,k}$ and $i' \notin \eta_{i,k+1}$. Especially, for $t = k$, we have $Q(\eta_{i',k+1}|\eta_{i',k}) = Q(\eta_{i',k+1}|\eta_{i,k})$. Thus, in order to compute $\partial Q_{i'} / \partial p_i$, we should first derive $\partial Q(\eta_{i',k+1}|\eta_{i,k}) / \partial p_i$ where $i' \in \eta_{i,k}$ and $i' \notin \eta_{i,k+1}$ $(0 \leqslant k \leqslant m-1)$. Then calculate $\partial Q(\eta_{i,t+1}|\eta_{i,t}) / \partial p_i$ for $0 \leqslant t \leqslant k-1$. For $\partial Q(\eta_{i',k+1}|\eta_{i,k}) / \partial p_i$, we have

$$Q(\eta_{i',k+1}|\eta_{i,k}) = \frac{V_{\eta_{i',k+1}}}{V_{\eta_{i,k}0} + \sum_{j \in Children(\eta_{i,k})} V_j}, \tag{B.20}$$

where $V_{\eta_{i,k}0}$ represents the no-purchase option of node $\eta_{i,k}$. For notational simplicity, we omit the price vector in the expression of preference weight. We can also see that $V_{\eta_{i',k+1}}$ does not contain the term p_i, but $V_{\eta_{i,k+1}} \in Children(\eta_{i,k})$ does. In order to obtain the partial derivative of (C.7) with respect to p_i, we need to derive $\partial V_{\eta_{i,k+1}} / \partial p_i$

$$\frac{\partial V_{\eta_{i,k+1}}}{\partial p_i} = \gamma_{i,k+1} \left(V_{\eta_{i,k+1}0} + \sum_{j \in Children(\eta_{i,k+1})} V_j \right)^{\gamma_{i,k+1}-1} \frac{\partial V_{\eta_{i,k+2}}}{\partial p_i}$$

It is an iterative form, we can finally get

$$\frac{\partial V_{\eta_{i,k+1}}}{\partial p_i} = -\beta_i V_i * \prod_{q=k+1}^{m-1} \gamma_{i,q} \left(V_{\eta_{i,q}0} + \sum_{j \in Children(\eta_{i,q})} V_j \right)^{\gamma_{i,q}-1}. \tag{B.21}$$

For notational simplicity, we denote $\lambda_{i,t}^{s} = \prod_{q=t}^{s} \gamma_{i,q}$. Based on Equations (C.7) and (C.8), we have

$$\begin{aligned}\frac{\partial Q(\eta_{i',k+1}|\eta_{i,k})}{\partial p_i} &= \beta_i \lambda_{i,k+1}^{m-1} * \frac{V_{\eta_{i',k+1}}}{V_{\eta_{i,k}0} + \sum_{j\in Children(\eta_{i,k})} V_j} * \prod_{q=k+1}^{m-1} Q(\eta_{i,q}|\eta_{i,q-1}) * Q(i|\eta_{i,m-1}) \\ &= \beta_i \lambda_{i,k+1}^{m-1} * Q(\eta_{i',k+1}|\eta_{i,k}) * Q(i|\eta_{i,k})\end{aligned}. \tag{B.22}$$

For $\partial Q(\eta_{i,t+1}|\eta_{i,t})/\partial p_i$, since $Q(\eta_{i,t+1}|\eta_{i,t}) = V_{\eta_{i,t+1}} / \left(V_{\eta_{i,t}0} + \sum_{j\in Children(\eta_{i,t})} V_j\right)$, then

$$\begin{aligned}\frac{\partial Q(\eta_{i,t+1}|\eta_{i,t})}{\partial p_i} &= (1 - \frac{V_{\eta_{i,t+1}}}{V_{\eta_{i,t}0} + \sum_{j\in Children(\eta_{i,t})} V_j}) * \frac{1}{V_{\eta_{i,t}0} + \sum_{j\in Children(\eta_{i,t})} V_j} * \frac{\partial V_{\eta_{i,t+1}}}{\partial p_i} \\ &= -\beta_i \lambda_{i,t+1}^{m-1} * (1 - Q(\eta_{i,t+1}|\eta_{i,t})) * Q(i|\eta_{i,t}).\end{aligned} \tag{B.23}$$

After plugging Equations (C.9) and (C.10) into Equation (C.6), we have

$$\begin{aligned}\frac{\partial Q_{i'}}{\partial p_i} &= \frac{\partial Q(\eta_{i',k+1}|\eta_{i,k})}{\partial p_i} * \frac{Q_{i'}}{Q(\eta_{i',k+1}|\eta_{i,k})} + \sum_{t=0}^{k-1} \frac{\partial Q(\eta_{i,t+1}|\eta_{i,t})}{\partial p_i} * \frac{Q_{i'}}{Q(\eta_{i,t+1}|\eta_{i,t})} \\ &= \beta_i \left(\sum_{t=1}^{k} (\lambda_{i,t+1}^{m-1} - \lambda_{i,t}^{m-1}) Q(i|\eta_{i,t}) Q_{i'} + \lambda_{i,1}^{m-1} Q_i Q_{i'} \right) \\ &= \beta_i Q_i \left(\sum_{t=1}^{k} (\lambda_{i,t+1}^{m-1} - \lambda_{i,t}^{m-1}) Q(i'|\eta_{i,t}) + \lambda_{i,1}^{m-1} Q_{i'} \right)\end{aligned}. \tag{B.24}$$

The last equality is due to the fact that $Q_i = Q(i|\eta_{i,t})Q(\eta_{i,t})$ and $Q_{i'} = Q(i'|\eta_{i,t})Q(\eta_{i,t})$, then $Q(i|\eta_t)Q_{i'} = Q(i'|\eta_t)Q_i$. Also note when $k = 0$, $\partial Q_{i'}/\partial p_i = \beta_i \lambda_{i,1}^{m-1} Q_i Q_{i'}$. In a similar way, for $\partial Q_i/\partial p_i$, we have

$$\frac{\partial Q_i}{\partial p_i} = \sum_{t=0}^{m-1} \frac{\partial Q(\eta_{i,t+1}|\eta_{i,t})}{\partial p_i} * \frac{Q_i}{Q(\eta_{i,t+1}|\eta_{i,t})} = \beta_i Q_i \left(\sum_{t=1}^{m-1} (\lambda_{i,t+1}^{m-1} - \lambda_{i,t}^{m-1}) Q(i|\eta_{i,t}) + \lambda_{i,1}^{m-1} Q_i - 1 \right). \tag{B.25}$$

Then plug Equations (C.11) and (C.12) into Equation (C.4), we obtain

$$\begin{aligned}\partial R_{\text{root}}(\mathbf{P}_{\text{root}})/\partial p_i &= Q_i + m_i \beta_i Q_i \left(\sum_{t=1}^{m-1} (\lambda_{i,t+1}^{m-1} - \lambda_{i,t}^{m-1}) Q(i|\eta_{i,t}) + \lambda_{i,1}^{m-1} Q_i - 1 \right) \\ &+ \sum_{\substack{i'\in\eta_{i,0} \\ i'\notin\eta_{i,1}}} m_{i'} \beta_i \lambda_{i,1}^{m-1} Q_i Q_{i'} + \sum_{k=1}^{m-2} \sum_{\substack{i'\in\eta_{i,k} \\ i'\notin\eta_{i,k+1}}} m_{i'} \beta_i Q_i \left(\sum_{t=1}^{k} (\lambda_{i,t+1}^{m-1} - \lambda_{i,t}^{m-1}) Q(i'|\eta_t) + \lambda_{i,1}^{m-1} Q_{i'} \right) \\ &+ \sum_{\substack{i'\in\eta_{i,m-1} \\ i'\neq i}} m_{i'} \beta_i Q_i \left(\sum_{t=1}^{m-1} (\lambda_{i,t+1}^{m-1} - \lambda_{i,t}^{m-1}) Q(i'|\eta_t) + \lambda_{i,1}^{m-1} Q_{i'} \right)\end{aligned}$$

This lemma is established after collecting terms. □

PROOF OF LEMMA 12

Proof. We prove this lemma by mathematical induction on the level in which node i lies. For notational brevity, we denote $V_i = V_i(\theta_i)$, $\omega_i = \omega_i(\theta_i)$, $\tau_i = \tau_i(\theta_i)$, $\delta_i = \delta_i(\theta_i)$, $u_i = u_i(\theta_i)$ and $Q(j|i) = Q(\theta_j|\theta_i)$. First, consider the case when i is a basic node, we have

$$\begin{aligned}
\frac{\partial V_i}{\partial \theta_i} &= \frac{\partial[(V_{i0} + \sum_{j\in i_C} V_j)^{\gamma_i}]}{\partial \theta_i} = \gamma_i(V_{i0} + \sum_{j\in i_C} V_j)^{\gamma_i - 1} \sum_{j\in i_C} \frac{\partial V_j}{\partial \theta_i} = \gamma_i V_i \frac{1}{V_{i0} + \sum_{j\in i_C} V_j} \sum_{j\in i_C} \frac{\partial V_j}{\partial \theta_j}\frac{\partial \theta_j}{\partial \theta_i} \\
&= \gamma_i V_i \frac{1}{V_{i0} + \sum_{j\in i_C} V_j} \sum_{j\in i_C} (-\beta_j V_j) = -\gamma_i V_i \sum_{j\in i_C} \beta_j Q(j|i) = -\gamma_i V_i u_i, \\
\frac{\partial \omega_i}{\partial \theta_i} &= (\frac{1}{\gamma_i} - 1) \sum_{j\in i_C} \frac{\partial Q(j|i)}{\partial \theta_i} \frac{1}{\beta_j} \\
&= (\frac{1}{\gamma_i} - 1) \sum_{j\in i_C} [\frac{\partial V_j}{\partial \theta_i}(V_{i0} + \sum_{j\in i_C} V_j) - V_j \sum_{j\in i_C} \frac{\partial V_j}{\partial \theta_i}] \frac{1}{(V_{i0} + \sum_{j\in i_C} V_j)^2} \frac{1}{\beta_j} \\
&= (\frac{1}{\gamma_i} - 1) \sum_{j\in i_C} [-\beta_j Q(j|i) - Q(j|i) \sum_{j\in i_C} (-\beta_j Q(j|i))] \frac{\partial \theta_i}{\partial \theta_j} \frac{1}{\beta_j} \\
&= (\frac{1}{\gamma_i} - 1) \sum_{j\in i_C} [-Q(j|i) + \frac{Q(j|i)}{\beta_j} \sum_{j\in i_C} (\beta_j Q(j|i))] = \omega_i u_i - (\frac{1}{\gamma_i} - 1)\tau_i, \\
\frac{\partial \tau_i}{\partial \theta_i} &= \sum_{j\in i_C} \frac{\partial Q(j|i)}{\partial \theta_i} = \sum_{j\in i_C} [-\beta_j Q(j|i) + Q(j|i) \sum_{j\in i_C} \beta_j Q(j|i)] = -u_i + \tau_i u_i = -(1-\tau_i)u_i, \\
\frac{\partial \delta_i}{\partial \theta_i} &= -(\frac{1}{\gamma_i} - 1)\frac{\partial \tau_i}{\partial \theta_i} = (\frac{1}{\gamma_i} - 1)(1-\tau_i)u_i = (\delta_i - 1)u_i.
\end{aligned}$$

And

$$\begin{aligned}
0 \leqslant \omega_i &= (\frac{1}{\gamma_i} - 1) \sum_{j\in i_C} \frac{Q(j|i)}{\beta_j} \leqslant (\frac{1}{\gamma_i} - 1) \frac{1}{\min_{j\in i_C}\{\beta_j\}} \sum_{j\in i_C} Q(j|i) = (\frac{1}{\gamma_i} - 1)\frac{1}{\underline{B}_i} \sum_{j\in i_C} Q(j|i) \\
&\leqslant (\frac{1}{\gamma_i} - 1)\frac{1}{\underline{B}_i}, \\
0 \leqslant u_i &= \sum_{j\in i_C} \beta_j Q(j|i) \leqslant \max_{j\in i_C}\{\beta_j\} \sum_{j\in i_C} Q(j|i) \leqslant \overline{B}_i \sum_{j\in i_C} Q(j|i) \leqslant \overline{B}_i.
\end{aligned}$$

Under Assumption 3, we have

$$\begin{aligned}
\omega_i u_i &\leqslant (\frac{1}{\gamma_i} - 1)\frac{\overline{B}_i}{\underline{B}_i} \leqslant \frac{1}{\gamma_i} \\
\frac{\partial \theta_i}{\partial \theta_j} &= 1 \geqslant 0,
\end{aligned}$$

where second equation is because $\theta_i = \theta_j - 1/\beta_j$. In general, $\frac{\partial \theta_i}{\partial \theta_j} = \delta_j + \theta_j \frac{\partial \delta_j}{\partial \theta_j} - \frac{\partial \omega_j}{\partial \theta_j} \geqslant \frac{1}{\gamma_j} - (\frac{1}{\gamma_j} - 1)\tau_j + \theta_j \frac{\partial \delta_j}{\partial \theta_j} - (\omega_i u_i - (\frac{1}{\gamma_i} - 1)\tau_i) \geqslant \frac{1}{\gamma_j} - \omega_j u_j$. Thus this lemma holds for basic nodes. Suppose it is also true for nonleaf node j in level l ($1 \leqslant l \leqslant m-1$), then for it's parent node i, we have

$$\begin{aligned}
\frac{\partial V_i}{\partial \theta_i} &= \frac{\partial[(V_{i0} + \sum_{j\in i_C} V_j)^{\gamma_i}]}{\partial \theta_i} = \gamma_i V_i \frac{1}{V_{i0} + \sum_{j\in i_C} V_j} \sum_{j\in i_C} \frac{\partial V_j}{\partial \theta_j}\frac{\partial \theta_j}{\partial \theta_i} \\
&= \gamma_i V_i \frac{1}{V_{i0} + \sum_{j\in i_C} V_j} \sum_{j\in i_C} (-\gamma_j V_j u_j)\frac{\partial \theta_j}{\partial \theta_i} = -\gamma_i V_i \sum_{j\in i_C} \frac{\gamma_j u_j}{\partial \theta_i/\partial \theta_j} Q(\theta_j|\theta_i) = -\gamma_i V_i u_i, \\
\frac{\partial \omega_i}{\partial \theta_i} &= (\frac{1}{\gamma_i} - 1)\sum_{j\in i_C} \frac{1}{1-\gamma_j}\left[A - B\right],
\end{aligned}$$

where

$$\begin{aligned}
A &= \frac{\frac{\partial V_j}{\partial \theta_j}\omega_j + \frac{\partial \omega_j}{\partial \theta_j} V_j}{(V_{i0} + \sum_{j\in i_C} V_j)\delta_j}\frac{\partial \theta_j}{\partial \theta_i} \leqslant \frac{-\gamma_j V_j u_j \omega_j + (\omega_j u_j - (\frac{1}{\gamma_j} - 1)\tau_j) V_j}{(V_{i0} + \sum_{j\in i_C} V_j)\delta_j}\frac{\partial \theta_j}{\partial \theta_i} \\
&= \frac{(1-\gamma_j) V_j u_j \omega_j - \frac{1}{\gamma_j}\tau_j V_j (1-\gamma_j)}{(V_{i0} + \sum_{j\in i_C} V_j)\delta_j}\frac{\partial \theta_j}{\partial \theta_i} = \frac{(1-\gamma_j) Q(j|i)(\omega_j u_j - \frac{\tau_j}{\gamma_j})}{\delta_j}\frac{\partial \theta_j}{\partial \theta_i},
\end{aligned}$$

and

$$\begin{aligned}
B &= \frac{(V_j\omega_j)[\delta_j \sum_{j\in i_C} \frac{\partial V_j}{\partial \theta_j}\frac{\partial \theta_j}{\partial \theta_i} + (V_{i0} + \sum_{j\in i_C} V_j)\frac{\partial \delta_j}{\partial \theta_j}\frac{\partial \theta_j}{\partial \theta_i}]}{(V_{i0} + \sum_{j\in i_C} V_j)^2 \delta_j^2} \\
&= \frac{(V_j\omega_j)[\delta_j \sum_{j\in i_C} (-\gamma_j V_j u_j)\frac{\partial \theta_j}{\partial \theta_i} + (V_{i0} + \sum_{j\in i_C} V_j)\frac{\partial \delta_j}{\partial \theta_j}\frac{\partial \theta_j}{\partial \theta_i}]}{(V_{i0} + \sum_{j\in i_C} V_j)^2 \delta_j^2} \\
&= -\frac{V_j \omega_j}{(V_{i0} + \sum_{j\in i_C} V_j)\delta_j} \sum_{j\in i_C} \frac{\gamma_j Q(j|i) u_j}{\partial \theta_i/\partial \theta_j} + \frac{V_j \omega_j \frac{\partial \delta_j}{\partial \theta_j}}{\partial \theta_i/\partial \theta_j (V_{i0} + \sum_{j\in i_C} V_j)\delta_j^2} \\
&= -\frac{Q(j|i)\omega_j}{\delta_j} u_i + \frac{Q(j|i)\omega_j \frac{\partial \delta_j}{\partial \theta_j}}{\partial \theta_i/\partial \theta_j \delta_j^2} = -\frac{Q(j|i)}{\delta_j}\left(\omega_j u_i - \frac{\omega_j \frac{\partial \delta_j}{\partial \theta_j}}{\partial \theta_i/\partial \theta_j \delta_j}\right),
\end{aligned}$$

Thus we have

$$\begin{aligned}
\frac{\partial\omega_i}{\partial\theta_i} &= (\frac{1}{\gamma_i}-1)\sum_{j\in i_C}\frac{1}{1-\gamma_j}\left[A-B\right] \\
&\leqslant (\frac{1}{\gamma_i}-1)\sum_{j\in i_C}\frac{1}{1-\gamma_j}\left[\frac{(1-\gamma_j)Q(j|i)(\omega_j u_j-\frac{\tau_j}{\gamma_j})}{\delta_j}\frac{\partial\theta_j}{\partial\theta_i}+\frac{Q(j|i)}{\delta_j}\left(\omega_j u_i-\frac{\omega_j\frac{\partial\delta_j}{\partial\theta_j}}{\partial\theta_i/\partial\theta_j\delta_j}\right)\right] \\
&= \left[(\frac{1}{\gamma_i}-1)\sum_{j\in i_C}\frac{1}{1-\gamma_j}\frac{Q(j|i)}{\delta_j}\omega_j\right]u_i \\
&+ (\frac{1}{\gamma_i}-1)\sum_{j\in i_C}\frac{Q(j|i)}{\delta_j}\left[\frac{(\omega_j u_j-\frac{\tau_j}{\gamma_j})}{\partial\theta_i/\partial\theta_j}-\frac{\omega_j\frac{\partial\delta_j}{\partial\theta_j}}{\partial\theta_i/\partial\theta_j\delta_j(1-\gamma_j)}\right] \\
&= \omega_i u_i-(\frac{1}{\gamma_i}-1)\sum_{j\in i_C}\frac{Q(j|i)}{\delta_j}\left[\frac{\omega_j\frac{\partial\delta_j}{\partial\theta_j}}{\partial\theta_i/\partial\theta_j\delta_j(1-\gamma_j)}-\frac{(\omega_j u_j-\frac{\tau_j}{\gamma_j})\delta_j(1-\gamma_j)}{\partial\theta_i/\partial\theta_j\delta_j(1-\gamma_j)}\right] \\
&\leqslant \omega_i u_i-(\frac{1}{\gamma_i}-1)\sum_{j\in i_C}\frac{Q(j|i)}{\delta_j}\tau_j\frac{\frac{1}{\gamma_j}-\frac{\omega_j u_j}{\delta_j}}{\partial\theta_i/\partial\theta_j}\leqslant \omega_i u_i-(\frac{1}{\gamma_i}-1)\sum_{j\in i_C}\frac{Q(j|i)}{\delta_j}\tau_j \\
&= \omega_i u_i-(\frac{1}{\gamma_i}-1)\tau_i,
\end{aligned}$$

where the last inequality is due to Assumption 2 and the fact that

$$\frac{1}{\gamma_j}-\frac{\omega_j u_j}{\delta_j}\geqslant\frac{1}{\gamma_j}-\frac{\frac{1}{\gamma_j}}{1}=0.$$

We also have

$$\begin{aligned}
\frac{\partial \tau_i}{\partial \theta_i} &= \sum_{j \in i_C} \frac{\frac{\partial (V_j \tau_j)}{\partial \theta_i}(V_{i0} + \sum_{j \in i_C} V_j)\delta_j - (V_j \tau_j)\frac{\partial (V_{i0} + \sum_{j \in i_C} V_j)\delta_j}{\partial \theta_i}}{(V_{i0} + \sum_{j \in i_C} V_j)^2 \delta_j^2} \\
&= \sum_{j \in i_C} \frac{\frac{\partial V_j}{\partial \theta_j}\tau_j + V_j \frac{\partial \tau_j}{\partial \theta_j}}{(V_{i0} + \sum_{j \in i_C} V_j)\delta_j} \frac{\partial \theta_j}{\partial \theta_i} - \sum_{j \in i_C} \frac{V_j \tau_j \left[\delta_j \sum_{j \in i_C} \frac{\partial V_j}{\partial \theta_j}\frac{\partial \theta_j}{\partial \theta_i} + (V_{i0} + \sum_{j \in i_C} V_j)\frac{\partial \delta_j}{\partial \theta_j}\frac{\partial \theta_j}{\partial \theta_i}\right]}{(V_{i0} + \sum_{j \in i_C} V_j)^2 \delta_j^2} \\
&\leqslant \sum_{j \in i_C} \frac{\frac{\partial V_j}{\partial \theta_j}\tau_j + V_j \frac{\partial \tau_j}{\partial \theta_j}}{(V_{i0} + \sum_{j \in i_C} V_j)\delta_j} \frac{\partial \theta_j}{\partial \theta_i} - \sum_{j \in i_C} \frac{V_j \tau_j \left[\delta_j \sum_{j \in i_C} \frac{\partial V_j}{\partial \theta_j}\frac{\partial \theta_j}{\partial \theta_i}\right]}{(V_{i0} + \sum_{j \in i_C} V_j)^2 \delta_j^2} \\
&\leqslant \sum_{j \in i_C} \frac{-\gamma_j V_j u_j \tau_j - V_j(1-\tau_j)u_j}{(V_{i0} + \sum_{j \in i_C} V_j)\delta_j} \frac{\partial \theta_j}{\partial \theta_i} - \sum_{j \in i_C} \frac{V_j \tau_j \left[\delta_j \sum_{j \in i_C}(-\gamma_j V_j u_j)\right]}{(V_{i0} + \sum_{j \in i_C} V_j)^2 \delta_j^2 \partial \theta_i / \partial \theta_j} \\
&= \sum_{j \in i_C} \frac{-\gamma_j V_j u_j \tau_j - V_j(1-\tau_j)u_j}{(V_{i0} + \sum_{j \in i_C} V_j)\delta_j} \frac{\partial \theta_j}{\partial \theta_i} - \sum_{j \in i_C} \frac{V_j \tau_j \left[\delta_j \sum_{j \in i_C}(-\gamma_j V_j u_j)\right]}{(V_{i0} + \sum_{j \in i_C} V_j)^2 \delta_j^2 \partial \theta_i / \partial \theta_j} \\
&= -\sum_{j \in i_C} \frac{\gamma_j Q(j|i) u_j (\frac{1}{\gamma_j} - (\frac{1}{\gamma_j} - 1)\tau_j)}{\partial \theta_i / \partial \theta_j \delta_j} + \sum_{j \in i_C} \frac{Q(j|i)}{\delta_j}\tau_j \sum_{j \in i_C} \frac{\gamma_j Q(j|i) u_j}{\partial \theta_i / \partial \theta_j} = -u_i + \tau_i u_i \\
&= -(1-\tau_i)u_i.
\end{aligned}$$

Thus we have

$$\frac{\partial \delta_i}{\partial \theta_i} = -(\frac{1}{\gamma_i} - 1)\frac{\partial \tau_i}{\partial \theta_i} \geqslant (\frac{1}{\gamma_i} - 1)(1-\tau_i)u_i = (\delta_i - 1)u_i.$$

And

$$\begin{aligned}
0 \leqslant \omega_i &= (\frac{1}{\gamma_i} - 1)\sum_{j \in i_C} \frac{Q(j|i)}{1-\gamma_j}\frac{\omega_j}{\delta_j} \leqslant (\frac{1}{\gamma_i} - 1)\sum_{j \in i_C} \frac{Q(j|i)}{1-\gamma_j}\frac{(\frac{1}{\gamma_j} - 1)\frac{1}{\underline{B}_j}}{1} \\
&\leqslant (\frac{1}{\gamma_i} - 1)\frac{1}{\min_{j \in i_C}\{\gamma_j \underline{B}_j\}}\sum_{j \in i_C} Q(j|i) \\
&\leqslant (\frac{1}{\gamma_i} - 1)\frac{1}{\min_{j \in i_C}\{\gamma_j \underline{B}_j\}} = \leqslant (\frac{1}{\gamma_i} - 1)\underline{B}_i, \\
0 \leqslant u_i &= \sum_{j \in i_C} \frac{\gamma_j u_j}{\partial \theta_i / \partial \theta_j} Q(j|i) \leqslant \sum_{j \in i_C} \frac{\gamma_j u_j}{\frac{1}{\gamma_j} - \omega_j u_j} Q(j|i) \leqslant \sum_{j \in i_C} \frac{\gamma_j \overline{B}_j}{\frac{1}{\gamma_j} - (\frac{1}{\gamma_j} - 1)\frac{\overline{B}_j}{\underline{B}_j}} Q(j|i) \\
&\leqslant \max_{j \in i_C}\{\frac{\gamma_j^2 \overline{B}_j}{1 - (1-\gamma_j)\overline{B}_j / \underline{B}_j}\}\sum_{j \in i_C} Q(j|i) \leqslant \overline{B}_i \sum_{j \in i_C} Q(j|i) \leqslant \overline{B}_i.
\end{aligned}$$

Under Assumption 3 and 2, we have

$$\omega_i u_i \leqslant (\frac{1}{\gamma_j} - 1)\frac{\overline{B}_i}{\underline{B}_i} \leqslant \frac{1}{\gamma_i}$$

$$\frac{\partial \theta_i}{\partial \theta_j} = \delta_j + \theta_j \frac{\partial \delta_j}{\partial \theta_j} - \frac{\partial \omega_j}{\partial \theta_j} \geqslant \frac{1}{\gamma_j} - (\frac{1}{\gamma_j} - 1)\tau_j + \theta_j \frac{\partial \delta_j}{\partial \theta_j} - (\omega_j u_j - (\frac{1}{\gamma_j} - 1)\tau_j) \geqslant \frac{1}{\gamma_j} - \omega_j u_j \geqslant 0,$$

By the principle of mathematical induction, this lemma holds for all nonleaf node $i \in V$. □

Appendix C

Appendix to Chapter 3

C.1 Notation

For ease of reading, we summarize our notation as follows:

m	The number of levels in the tree structure.
N	The number of products.
G	The number of grid points for each node in the tree structure.
K	The maximum number of children nodes that a node can have in the tree structure.
N_i	The set of products that are associated with node i.
S_i	The assortment of node i.
$\mathbf{P}_i$	The price vector of assortment S_i.
$\Im_i$	The set of feasible assortments that satisfies cardinality/space constraints.
$\mathbb{C}_i$	The cardinality limitation on node i.
$\mathbb{S}_i$	The space limitation on node i.
$\mathscr{B}$	The set of basic nodes.
α_i	The price-independent deterministic utility of product i.
β_i	The price-sensitivity parameter of product i.
γ_i	The dissimilarity parameter for nonleaf node i.
$V_i(S_i, \mathbf{P}_i)$	The preference weight for assortment S_i with price vector $\mathbf{P}_i$.
$R_i(S_i, \mathbf{P}_i)$	The expected profit for assortment S_i with price vector $\mathbf{P}_i$.
Z^*	The maximum expected profit under cardinality/space constraints.
θ_i	The node-specific adjusted markup for node i.
$\mathscr{A}_i$	The collection of assortments that include S_i^*.
$\mathcal{F}_i(\theta_h)$	The optimal solution to problem (4.12).
$\tilde{S}_j(\theta_j)$	The optimal solution to problem (4.9).
$\tilde{S}_i(\theta_i)$	$:= \bigcup_{j \in \text{Children}(i)} \tilde{S}_j(\mathcal{F}_j(\theta_i))$.
$(S^\alpha_{\text{root}}, \mathbf{P}^\alpha_{\text{root}})$	α-approximate solution to problem (4.1) under space constraints.
$\mathscr{A}_i^\alpha$	The collection of assortments that include S_i^α.
Z^α	$:= R_{\text{root}}(S^\alpha_{\text{root}}, \mathbf{P}^\alpha_{\text{root}})$.

C.2 Technical Proofs

Proof of Lemma 13

We use the following two claims to prove lemma 13.

Claim 12. *For an arbitrary nonleaf node $i \in V$ and its parent node h, assume $R_i(S_i^*, \mathbf{P}_i^*) > e_h^*$. If there exists an assortment $\hat{S}_i \subseteq \Im_i$ and a price vector $\hat{\mathbf{P}}_i$ such that*

$$V_i(\hat{S}_i, \hat{\mathbf{P}}_i)^{1/\gamma_i} \left(R_i(\hat{S}_i, \hat{\mathbf{P}}_i) - e_i^* \right) \geqslant V_i(S_i^*, \mathbf{P}_i^*)^{1/\gamma_i} \left(R_i(S_i^*, \mathbf{P}_i^*) - e_i^* \right). \tag{C.1}$$

Then we have

$$V_i(\hat{S}_i, \hat{\mathbf{P}}_i) \left(R_i(\hat{S}_i, \hat{\mathbf{P}}_i) - e_h^* \right) \geqslant V_i(S_i^*, \mathbf{P}_i^*) \left(R_i(S_i^*, \mathbf{P}_i^*) - e_h^* \right). \tag{C.2}$$

If the inequality in (C.1) is strict, then the inequality in (C.2) is strict as well.

Proof. This claim is a special case of the lemma 1 in [50], which uses the subgradient inequality in [18], we provide the proof for completeness. For notational brevity, we let $\hat{V}_i = V_i(\hat{S}_i, \hat{\mathbf{P}}_i)$, $V_i^* = V_i(S_i^*, \mathbf{P}_i^*)$, $\hat{R}_i = R_i(\hat{S}_i, \hat{\mathbf{P}}_i)$ and $R_i^* = R_i(S_i^*, \mathbf{P}_i^*)$ throughout the proof. After dividing both sides of inequality (C.1) by $\hat{V}_i^{1/\gamma_i}$ and subtracting both sides by e_h^*, we obtain

$$\begin{aligned} \hat{R}_i - e_h^* &\geqslant \left(\frac{V_i^*}{\hat{V}_i} \right)^{1/\gamma_i} (R_i^* - e_i^*) + e_i^* - e_h^* \qquad \text{(C.3)} \\ &= \gamma_i \left(\frac{V_i^*}{\hat{V}_i} \right)^{1/\gamma_i} (R_i^* - e_h^*) + (1-\gamma_i)(R_i^* - e_h^*) = \left(\gamma_i \left(\frac{V_i^*}{\hat{V}_i} \right)^{1/\gamma_i} + (1-\gamma_i) \right) (R_i^* - e_h^*), \end{aligned}$$

where we get the first equality from the definition of e_i^* as $e_i^* = \gamma_i e_h^* + (1-\gamma_i) R_i^*$ and the second equality is obtained after collecting terms. Consider function $f(x) = x^{1/\gamma_i}$, which is convex since $0 < \gamma_i \leqslant 1$, we have $x^{1/\gamma_i} \geqslant 1 + (x-1)/\gamma_i$ due to its convexity. After rearranging the terms, we obtain $\gamma_i x^{1/\gamma_i} + 1 - \gamma_i \geqslant x$. Set $x = V_i^*/\hat{V}_i$, we have

$$\gamma_i \left(\frac{V_i^*}{\hat{V}_i} \right)^{1/\gamma_i} + (1-\gamma_i) \geqslant \frac{V_i^*}{\hat{V}_i}.$$

Since $R_i^* \geqslant e_h^*$ according to our assumption in this claim, multiply $R_i^* - e_h^*$ on both sides of the inequality above, we get

$$\left(\gamma_i \left(\frac{V_i^*}{\hat{V}_i} \right)^{1/\gamma_i} + (1-\gamma_i) \right) (R_i^* - e_h^*) \geqslant \frac{V_i^*}{\hat{V}_i} (R_i^* - e_h^*).$$

From the inequality above and (C.3), we get that

$$\hat{R}_i - e_h^* \geqslant \frac{V_i^*}{\hat{V}_i}(R_i^* - e_h^*),$$

implying that $\hat{V}_i(\hat{R}_i - e_h^*) \geqslant V_i^*(R_i^* - e_h^*)$. If the inequality in (C.1) is strict, then the inequality in (C.3) is also strict, thus the inequality in (C.2) becomes strict as well. □

Claim 13. *For an arbitrary nonleaf node $i \in V$ and its parent node h, assume $R_i(S_i^*, \mathbf{P}_i^*) > e_h^*$. If for all $j \in \text{Children}(i)$, there exists an assortment $\hat{S}_j \subseteq \Im_j$ and a price vector $\hat{\mathbf{P}}_j$ such that*

$$V_j(\hat{S}_j, \hat{\mathbf{P}}_j)\left(R_j(\hat{S}_j, \hat{\mathbf{P}}_j) - e_i^*\right) \geqslant V_j(S_j^*, \mathbf{P}_j^*)\left(R_j(S_j^*, \mathbf{P}_j^*) - e_i^*\right).$$

Let $(\hat{S}_i, \hat{\mathbf{P}}_i) = \bigcup_{j \in \text{Children}(i)} (\hat{S}_j, \hat{\mathbf{P}}_j)$, then we have

$$V_i(\hat{S}_i, \hat{\mathbf{P}}_i)\left(R_i(\hat{S}_i, \hat{\mathbf{P}}_i) - e_h^*\right) \geqslant V_i(S_i^*, \mathbf{P}_i^*)\left(R_i(S_i^*, \mathbf{P}_i^*) - e_h^*\right).$$

Proof. This claim follows directly from Claim 12. For notational brevity, we let $\hat{V}_i = V_i(\hat{S}_i, \hat{\mathbf{P}}_i)$, $V_i^* = V_i(S_i^*, \mathbf{P}_i^*)$, $\hat{R}_i = R_i(\hat{S}_i, \hat{\mathbf{P}}_i)$ and $R_i^* = R_i(S_i^*, \mathbf{P}_i^*)$. Because the first inequality of this claim holds for all $j \in \text{Children}(i)$, we get

$$\hat{V}_i^{1/\gamma_i}(\hat{R}_i - e_i^*) = \sum_{j \in \text{Children}(i)} \hat{V}_j(\hat{R}_j - e_i^*) \geqslant \sum_{j \in \text{Children}(i)} V_j^*(R_j^* - e_i^*) = (V_i^*)^{1/\gamma_i}(R_i^* - e_i^*).$$

The second inequality holds because of Claim 12. □

Proof. Proof of Lemma 13: We prove this lemma by induction on levels. It is true for root node: 1) $R_{\text{root}}(S_{\text{root}}^*, \mathbf{P}_{\text{root}}^*) > e_{\text{Parent(root)}}^* = 0$; 2) S_{root}^* is a nonempty set and $\mathbf{P}_{\text{root}}^*$ is a nonzero vector, otherwise $R_{\text{root}}(S_{\text{root}}^*, \mathbf{P}_{\text{root}}^*)$ would be zero; 3) Because joint subproblem (4.2) at root node is equivalent to the global optimization problem (4.1), $(S_i^*, \mathbf{P}_i^*)$ is identical to $(\hat{S}_i, \hat{\mathbf{P}}_i)$ as well. Then we assume that the three items in this lemma are true for any nonleaf node i in level l $(1 \leqslant l \leqslant m-1)$.

First, we prove the third item in this lemma. Suppose on the contrary that there exists node $k \in \text{Children}(i)$ in level $l+1$ such that $V_k(S_k^*, \mathbf{P}_k^*)(R_k(S_k^*, \mathbf{P}_k^*) - e_i^*) < \max_{S_k \subseteq \Im_k} \max_{\mathbf{P}_k} \{V_k(S_k, \mathbf{P}_k)(R_k(S_k, \mathbf{P}_k) - e_i^*)\}$. For each $j \in \text{Children}(i)$, define $(\hat{S}_j, \hat{\mathbf{P}}_j) = \arg\max_{S_j \subseteq \Im_j} \max_{\mathbf{P}_j} V_j(S_j, \mathbf{P}_j)\left(R_j(S_j, \mathbf{P}_j) - e_i^*\right)$. Then by construction: $V_j(\hat{S}_j, \hat{\mathbf{P}}_j)\left(R_j(\hat{S}_j, \hat{\mathbf{P}}_j) - e_i^*\right) \geqslant V_j(S_j^*, \mathbf{P}_j^*)\left(R_j(S_j^*, \mathbf{P}_j^*) - e_i^*\right)$, the inequality in which is strict if $j = k$. Moreover, we have $R_i(S_i^*, \mathbf{P}_i^*) \geqslant e_h^*$ according to the induction hypothesis. Let $(\hat{S}_i, \hat{\mathbf{P}}_i) = \bigcup_{j \in \text{Children}(i)} (\hat{S}_j, \hat{\mathbf{P}}_j)$, then according to Claim 13, we have $V_i(\hat{S}_i, \hat{\mathbf{P}}_i)\left(R_i(\hat{S}_i, \hat{\mathbf{P}}_i) - e_h^*\right) > V_i(S_i^*, \mathbf{P}_i^*)\left(R_i(S_i^*, \mathbf{P}_i^*) - e_h^*\right)$. It contradicts the induction hypothesis that $V_i(S_i^*, \mathbf{P}_i^*)(R_i(S_i^*, \mathbf{P}_i^*) - e_h^*) = \max_{S_i \subseteq \Im_i} \max_{\mathbf{P}_i} \{V_i(S_i, \mathbf{P}_i)(R_i(S_i, \mathbf{P}_i) - e_h^*)\}$. Thus the first part of item 3 is true. Then the second part is also true by Claim 13.

Second, we prove the first two items in this lemma. For node $j \in \text{Children}(i)$, from the third item in this lemma, we know that $(S_j^*, \mathbf{P}_j^*)$ is the optimal solution to the joint subproblem at node j. Suppose on the contrary that $R_j(S_j^*, \mathbf{P}_j^*) \leqslant e_i^*$, then we obtain $\max_{S_j \subseteq \Im_j} \max_{\mathbf{P}_j} \{V_j(S_j, \mathbf{P}_j)(R_j(S_j, \mathbf{P}_j) - e_i^*)\} = V_j(S_j^*, \mathbf{P}_j^*)(R_j(S_j^*, \mathbf{P}_j^*) - e_i^*) \leqslant 0$. We construct another nonempty assortment $\tilde{S}_j$ with price vector $\tilde{P}_j$ for node j such that for any product $k \in \tilde{S}_j$, we set the price $p_k = e_i^* + \epsilon$ where ϵ is strictly larger than zero. By construction, we have $V_j(\tilde{S}_j, \tilde{P}_j)(R_j(\tilde{S}_j, \tilde{P}_j) - e_i^*) = \epsilon * V_j(\tilde{S}_j, \tilde{P}_j) > 0 \geqslant V_j(S_j^*, \mathbf{P}_j^*)(R_j(S_j^*, \mathbf{P}_j^*) - e_i^*)$, which contradicts that $(S_j^*, \mathbf{P}_j^*)$ is the optimal solution to the joint subproblem at node j. By the definition of e_i^* and the induction hypothesis, we get that $e_j^* \geqslant 0$. Thus the first item in this lemma is true. The second item also holds, otherwise $R_j(S_j^*, \mathbf{P}_j^*)$ would be zero, which contradicts with the first item in this lemma. □

Proof of Lemma 14

Proof. Denote $(\tilde{S}_i, \tilde{\mathbf{P}}_i)$ as the optimal solution to problem (4.3), then we have

$$V_i(\tilde{S}_i, \tilde{\mathbf{P}}_i)^{1/\gamma_i} \left(R_i(\tilde{S}_i, \tilde{\mathbf{P}}_i) - e_i^* \right) \geqslant V_i(S_i^*, \mathbf{P}_i^*)^{1/\gamma_i} \left(R_i(S_i^*, \mathbf{P}_i^*) - e_i^* \right).$$

Because of Claim 12, we have

$$\begin{aligned} V_i(\tilde{S}_i, \tilde{\mathbf{P}}_i) \left(R_i(\tilde{S}_i, \tilde{\mathbf{P}}_i) - e_h^* \right) &\geqslant V_i(S_i^*, \mathbf{P}_i^*) \left(R_i(S_i^*, \mathbf{P}_i^*) - e_h^* \right) \\ &= \max_{S_i \subseteq \Im_i} \max_{\mathbf{P}_i} V_i(S_i, \mathbf{P}_i)(R_i(S_i, \mathbf{P}_i) - e_h^*), \end{aligned}$$

where the last equality is due to Lemma 13. Therefore, $(\tilde{S}_i, \tilde{\mathbf{P}}_i)$ is also optimal to problem (4.2), which completes the proof. □

Proof of Lemma 15

Claim 14 and claim 15 are used to prove lemma 15 and lemma 16.

Claim 14. *For nonleaf node i in level d $(0 \leqslant d \leqslant m-1)$ and a given assortment S_i. The first derivative of $R_i(S_i, \mathbf{P}_i)$ and $V_i(S_i, \mathbf{P}_i)$ with respect to price p_r of any product r in the lowest level m is*

$$\begin{aligned} \frac{\partial R_i(S_i, \mathbf{P}_i)}{\partial p_r} &= \beta_r Q(r|i) \Bigg(\frac{1}{\beta_r} - m_r + \lambda_{r,d+1}^{m-1} \sum_{r' \in \eta_{r,d}} m_{r'} Q(r'|i) \\ &+ \sum_{k=d+1}^{m-2} \sum_{\substack{r' \in \eta_{r,k} \\ r' \notin \eta_{r,k+1}}} m_{r'} \left(\sum_{t=d+1}^{k} (\lambda_{r,t+1}^{m-1} - \lambda_{r,t}^{m-1}) Q(r'|\eta_{r,t}) \right) \\ &+ \sum_{r' \in \eta_{r,m-1}} m_{r'} \left(\sum_{t=d+1}^{m-1} (\lambda_{r,t+1}^{m-1} - \lambda_{r,t}^{m-1}) Q(r'|\eta_{r,t}) \right) \Bigg), \end{aligned}$$

and

$$\frac{\partial V_i(S_i, \mathbf{P}_i)}{\partial p_r} = -\beta_r Q(r|i)\lambda_d^{m-1} V_i(S_i, \mathbf{P}_i),$$

where the notations are defined as follows

$$\lambda_{r,t}^s = \begin{cases} \prod_{q=t}^s \gamma_{r,q} & t \leqslant s \\ 1 & t > s \end{cases}$$

$$Q(r|\eta_{r,l}) = Q(r|\eta_{r,l+1}) \frac{V_{\eta_{r,l+1}}}{\sum_{j \in \text{Children}(\eta_{r,l})} V_j},$$

$$m_r = p_r - c_r.$$

We use the following notations throughout the Appendix C.2. $\eta_{r,t}$ is defined as the product r's ancestor node in level t. For simplicity, $\eta_{r,t}$ is also used to represent the set of products that are associated with it. $Q(r|i)$ is the choosing probability of product r given node i and the markup for product r is denoted as $m_r = p_r - c_i$. $\gamma_{r,q}$ is the dissimilarity parameter of the r's ancestor in level q.

Proof. The markup for product r in level m is defined as $m_r = p_r - c_i$, then we have $R_i(S_i, \mathbf{P}_i) = \sum_{r \in \eta_{r,d}} m_r * Q(r|i)$ where $\eta_{r,d} = S_i$ that is the assortments of node i according to the definition. Thus the first derivative of $R_i(S_i, \mathbf{P}_i)$ with respect to price p_r of product r is

$$\frac{\partial R_i(S_i, \mathbf{P}_i)}{\partial p_r} = Q(r|i) + m_r \frac{\partial Q(r|i)}{\partial p_r} + \sum_{k=d}^{m-1} \sum_{\substack{r' \in \eta_{r,k} \\ r' \notin \eta_{r,k+1}}} m_{r'} \frac{\partial Q(r'|i)}{\partial p_r}. \tag{C.4}$$

We will derive $\partial Q(r|i)/\partial p_r$ and $\partial Q(r'|i)/\partial p_r$ in the remaining part of this proof, respectively. For all $r' \in \eta_{r,k}$ and $r' \notin \eta_{r,k+1}$ $(d \leqslant k \leqslant m-1)$, $\partial Q(r'|i)/\partial p_r$ can be calculated as

$$\frac{\partial Q(r'|i)}{\partial p_r} = \frac{\partial(\prod_{t=d}^{m-1} Q(\eta_{r',t+1}|\eta_{r',t}))}{\partial p_r} = \sum_{t=d}^{m-1} \frac{\partial Q(\eta_{r',t+1}|\eta_{r',t})}{\partial p_r} * \frac{Q(r'|i)}{Q(\eta_{r',t+1}|\eta_{r',t})}$$

$$= \sum_{t=d}^{k} \frac{\partial Q(\eta_{r',t+1}|\eta_{r',t})}{\partial p_r} * \frac{Q(r'|i)}{Q(\eta_{r',t+1}|\eta_{r',t})} \tag{C.5}$$

$$= \frac{\partial Q(\eta_{r',k+1}|\eta_{r,k})}{\partial p_r} * \frac{Q(r'|i)}{Q(\eta_{r',k+1}|\eta_{r,k})} + \sum_{t=d}^{k-1} \frac{\partial Q(\eta_{r,t+1}|\eta_{r,t})}{\partial p_r} * \frac{Q(r'|i)}{Q(\eta_{r,t+1}|\eta_{r,t})}. \tag{C.6}$$

Equation (C.5) is due to the fact that $\partial Q(\eta_{r',t+1}|\eta_{r',t})/\partial p_r = 0$ when $k+1 \leqslant t \leqslant m-1$. Equation (C.6) is established because $Q(\eta_{r',t+1}|\eta_{r',t}) = Q(\eta_{r,t+1}|\eta_{r,t})$ $(d \leqslant t \leqslant k-1)$ when $r' \in \eta_{r,k}$ and $r' \notin \eta_{r,k+1}$. Especially, for $t = k$, we have: $Q(\eta_{r',k+1}|\eta_{r',k}) = Q(\eta_{r',k+1}|\eta_{r,k})$. Thus, in order to compute $\partial Q(r'|i)/\partial p_r$, we should firstly derive $\partial Q(\eta_{r',k+1}|\eta_{r,k})/\partial p_r$ where

$r' \in \eta_{r,k}$ and $r' \notin \eta_{r,k+1}$ ($d \leqslant k \leqslant m-1$). Then secondly calculate $\partial Q(\eta_{r,t+1}|\eta_{r,t})/\partial p_r$ for $d \leqslant t \leqslant k-1$. For $\partial Q(\eta_{r',k+1}|\eta_{r,k})/\partial p_r$, we have

$$Q(\eta_{r',k+1}|\eta_{r,k}) = \frac{V_{\eta_{r',k+1}}}{\sum_{j\in\text{Children}(\eta_{r,k})} V_j}. \tag{C.7}$$

For notational simplicity, we omit the assortment and price vector in the expression of preference weight. We can also see that $V_{\eta_{r',k+1}}$ does not contain the term p_r, but $V_{\eta_{r,k+1}} \in \text{Children}(\eta_{r,k})$ does. In order to obtain the partial derivative of $Q(\eta_{r',k+1}|\eta_{r,k})$ that is defined in (C.7) with respect to p_r, we need

$$\frac{\partial V_{\eta_{r,k+1}}}{\partial p_r} = \gamma_{r,k+1}\left(\sum_{j\in\text{Children}(\eta_{r,k+1})} V_j\right)^{\gamma_{r,k+1}-1}\frac{\partial V_{\eta_{r,k+2}}}{\partial p_r}.$$

It is an iterative form, we can finally get

$$\frac{\partial V_{\eta_{r,k+1}}}{\partial p_r} = -\beta_r V_i * \prod_{q=k+1}^{m-1}\gamma_{r,q}\left(\sum_{j\in\text{Children}(\eta_{r,q})} V_j\right)^{\gamma_{r,q}-1}. \tag{C.8}$$

For notational simplicity, we denote $\lambda_{r,t}^{s} = \prod_{q=t}^{s}\gamma_{r,q}$. Therefore, we obtain

$$\frac{\partial V_i}{\partial p_r} = -\beta_r Q(r|i)\lambda_d^{m-1}V_i,$$

thus the second part of this claim holds. Based on Equations (C.7) and (C.8), we have

$$\begin{aligned}\frac{\partial Q(\eta_{r',k+1}|\eta_{r,k})}{\partial p_r} &= \beta_r\lambda_{r,k+1}^{m-1} * \frac{V_{\eta_{r',k+1}}}{\sum_{j\in\text{Children}(\eta_{r,k})} V_j} * \prod_{q=k+1}^{m-1} Q(\eta_{r,q}|\eta_{r,q-1}) * Q(r|\eta_{r,m-1}) \\ &= \beta_r\lambda_{r,k+1}^{m-1} * Q(\eta_{r',k+1}|\eta_{r,k}) * Q(r|\eta_{r,k}),\end{aligned} \tag{C.9}$$

where $Q(r|\eta_{r,k}) = Q(r|\eta_{r,k+1}) * V_{\eta_{r,k+1}}/\sum_{j\in\text{Children}(\eta_{r,k})} V_j$. For $\partial Q(\eta_{r,t+1}|\eta_{r,t})/\partial p_r$, since $Q(\eta_{r,t+1}|\eta_{r,t}) = V_{\eta_{r,t+1}}/\sum_{j\in\text{Children}(\eta_{r,t})} V_j$, then

$$\begin{aligned}\frac{\partial Q(\eta_{r,t+1}|\eta_{r,t})}{\partial p_r} &= \left(1 - \frac{V_{\eta_{r,t+1}}}{\sum_{j\in\text{Children}(\eta_{r,t})} V_j}\right) * \frac{1}{\sum_{j\in\text{Children}(\eta_{r,t})} V_j} * \frac{\partial V_{\eta_{r,t+1}}}{\partial p_r} \\ &= -\beta_r\lambda_{r,t+1}^{m-1} * (1 - Q(\eta_{r,t+1}|\eta_{r,t})) * Q(r|\eta_{r,t}).\end{aligned} \tag{C.10}$$

After plugging Equations (C.9) and (C.10) into Equation (C.6), we have

$$\begin{aligned}\frac{\partial Q(r'|i)}{\partial p_r} &= \frac{\partial Q(\eta_{r',k+1}|\eta_{r,k})}{\partial p_r} * \frac{Q(r'|i)}{Q(\eta_{r',k+1}|\eta_{r,k})} + \sum_{t=d}^{k-1} \frac{\partial Q(\eta_{r,t+1}|\eta_{r,t})}{\partial p_r} * \frac{Q(r'|i)}{Q(\eta_{r,t+1}|\eta_{r,t})} \\ &= \beta_r \left(\sum_{t=d+1}^{k} (\lambda_{r,t+1}^{m-1} - \lambda_{r,t}^{m-1}) Q(r|\eta_{r,t}) Q(r'|i) + \lambda_{r,d+1}^{m-1} Q(r|i) Q(r'|i) \right) \\ &= \beta_r Q(r|i) \left(\sum_{t=d+1}^{k} (\lambda_{r,t+1}^{m-1} - \lambda_{r,t}^{m-1}) Q(r'|\eta_{r,t}) + \lambda_{r,d+1}^{m-1} Q(r'|i) \right).\end{aligned} \tag{C.11}$$

The last equality is due to the fact that $Q(r|i) = Q(r|\eta_{r,t})Q(\eta_{r,t})$ and $Q(r'|i) = Q(r'|\eta_{r,t})Q(\eta_{r,t})$, then $Q(r|\eta_t)Q(r'|i) = Q(r'|\eta_t)Q_r$. Also note when $k = d$, $\partial Q(r'|i)/\partial p_r = \beta_r \lambda_{r,d+1}^{m-1} Q(r|i)Q(r'|i)$. In a similar way, for $\partial Q(r|i)/\partial p_r$, we have

$$\begin{aligned}\frac{\partial Q(r|i)}{\partial p_r} &= \sum_{t=d}^{m-1} \frac{\partial Q(\eta_{r,t+1}|\eta_{r,t})}{\partial p_r} * \frac{Q(r|i)}{Q(\eta_{r,t+1}|\eta_{r,t})} \\ &= \beta_r Q(r|i) \left(\sum_{t=d+1}^{m-1} (\lambda_{r,t+1}^{m-1} - \lambda_{r,t}^{m-1}) Q(r|\eta_{r,t}) + \lambda_{r,d+1}^{m-1} Q(r|i) - 1 \right).\end{aligned} \tag{C.12}$$

Then plug Equations (C.11) and (C.12) into Equation (C.4), we obtain

$$\begin{aligned}\partial R_i(S_i, \mathbf{P}_i)/\partial p_r &= Q(r|i) + m_r \beta_r Q(r|i) \left(\sum_{t=d+1}^{m-1} (\lambda_{r,t+1}^{m-1} - \lambda_{r,t}^{m-1}) Q(r|\eta_{r,t}) + \lambda_{r,d+1}^{m-1} Q(r|i) - 1 \right) \\ &+ \sum_{\substack{r' \in \eta_{r,d} \\ r' \notin \eta_{r,d+1}}} m_{r'} \beta_r \lambda_{r,d+1}^{m-1} Q(r|i) Q(r'|i) \\ &+ \sum_{k=d+1}^{m-2} \sum_{\substack{r' \in \eta_{r,k} \\ r' \notin \eta_{r,k+1}}} m_{r'} \beta_r Q(r|i) \left(\sum_{t=d+1}^{k} (\lambda_{r,t+1}^{m-1} - \lambda_{r,t}^{m-1}) Q(r'|\eta_t) + \lambda_{r,d+1}^{m-1} Q(r'|i) \right) \\ &+ \sum_{\substack{r' \in \eta_{r,m-1} \\ r' \neq r}} m_{r'} \beta_r Q(r|i) \left(\sum_{t=d+1}^{m-1} (\lambda_{r,t+1}^{m-1} - \lambda_{r,t}^{m-1}) Q(r'|\eta_t) + \lambda_{r,d+1}^{m-1} Q(r'|i) \right).\end{aligned}$$

This first part of this claim is established after collecting terms. □

Claim 15. *For any nonleaf node $i \in V$ and an assortment S_i, we have*

$$\frac{\partial V_i(S_i, \theta_i)}{\partial \theta_i} = -\gamma_i V_i(S_i, \theta_i) u_i(S_i, \theta_i), \tag{C.13}$$

and

$$\frac{\partial \omega_i(S_i,\theta_i)}{\partial \theta_i} = \omega_i(S_i,\theta_i)u_i(S_i,\theta_i) - (\frac{1}{\gamma_i} - 1), \tag{C.14}$$

where

$$u_i(S_i,\theta_i) = \begin{cases} \sum_{j\in \text{Children}(i)} \beta_j Q(S_j,\theta_j|S_i,\theta_i) & i \text{ is a basic node} \\ \sum_{j\in \text{Children}(i)} \frac{\gamma_j^2 u_j(S_j)}{1-\gamma_j\omega_j(S_j,\theta_j)u_j(S_j)} Q(S_j,\theta_j|S_i,\theta_i) & \text{o.w.} \end{cases}$$

Proof. For notational brevity, let $\omega_i = \omega_i(S_i,\theta_i)$, $u_i = u_i(S_i,\theta_i)$, $V_i = V_i(S_i,\theta_i)$ and $Q(j|i) = Q(S_j,\theta_j|S_i,\theta_i)$. We prove this claim by induction on depth of the node i. Firstly, if node i is a basic node, then $V_i = (\sum_{j\in \text{Children}(i)} V_j)^{\gamma_i}$ where $V_j = \exp(\alpha_j - \beta_j p_j)$. Thus we have

$$\begin{aligned} \frac{\partial V_i}{\partial \theta_i} &= \gamma_i(\sum_{j\in \text{Children}(i)} V_j)^{\gamma_i - 1} \sum_{j\in \text{Children}(i)} \frac{\partial V_j}{\partial p_j}\frac{\partial p_j}{\partial \theta_i} \\ &= -\gamma_i V_i \sum_{j\in \text{Children}(i)} \beta_j Q(j|i) \\ &= -\gamma_i V_i u_i, \end{aligned}$$

where the second equality is due to $Q(j|i) = V_j / \sum_{j\in \text{Children}(i)} V_j$ and $\theta_i = p_j - c_j - 1/\beta_j$, and the last equality is because of the definition of u_i for the basic node i. So Equation (C.14) holds if i is a basic node.

Recall that $\omega_i = (\frac{1}{\gamma_i} - 1)\sum_{j\in \text{Children}(i)} \frac{Q(j|i)}{\beta_j} = (\frac{1}{\gamma_i} - 1)\sum_{j\in \text{Children}(i)} \frac{V_j}{\beta_j \sum_{j\in \text{Children}(i)} V_j}$ from Definition 1, then we also have

$$\begin{aligned} \frac{\partial \omega_i}{\partial \theta_i} &= (\frac{1}{\gamma_i} - 1)\sum_{j\in \text{Children}(i)} \frac{\frac{\partial V_j}{\partial \theta_i}\sum_{j\in \text{Children}(i)} V_j - V_j \sum_{j\in \text{Children}(i)} \frac{\partial V_j}{\partial \theta_i}}{(\sum_{j\in \text{Children}(i)} V_j)^2 \beta_j} \\ &= (\frac{1}{\gamma_i} - 1)\sum_{j\in \text{Children}(i)} \frac{1}{\beta_j}\left[\frac{\frac{\partial V_j}{\partial p_j}\frac{\partial p_j}{\partial \theta_i}}{\sum_{j\in \text{Children}(i)} V_j} - Q(j|i)\sum_{j\in \text{Children}(i)} \frac{\partial V_j}{\partial p_j}\frac{\partial p_j}{\partial \theta_i}\right] \\ &= (\frac{1}{\gamma_i} - 1)\sum_{j\in \text{Children}(i)} \frac{1}{\beta_j}\left[-\beta_j Q(j|i) - Q(j|i)\sum_{j\in \text{Children}(i)} (-\beta_j Q(j|i))\right] \\ &= \omega_i u_i - (\frac{1}{\gamma_i} - 1). \end{aligned}$$

So Equation (C.14) holds for the basic node i.

Then assume Equations (C.13) and (C.14) are true for node $j \in \text{Children}(i)$, where node i is neither a basic node nor a leaf node

$$\frac{\partial V_j}{\partial \theta_j} = -\gamma_j V_j u_j, \tag{C.15}$$

$$\frac{\partial \omega_j}{\partial \theta_j} = \omega_j u_j - (\frac{1}{\gamma_j} - 1). \tag{C.16}$$

Since $\theta_i = \theta_j - \omega_j$, then

$$\frac{\partial \theta_j}{\partial \theta_i} = (1 - \frac{\partial \omega_j}{\partial \theta_j})^{-1} = (\frac{1}{\gamma_j} - \omega_j u_j)^{-1}.$$

We get

$$\begin{aligned}
\frac{\partial V_i}{\partial \theta_i} &= \gamma_i (\sum_{j \in \text{Children}(i)} V_j)^{\gamma_i - 1} \sum_{j \in \text{Children}(i)} \frac{\partial V_j}{\partial \theta_i} \\
&= \gamma_i (\sum_{j \in \text{Children}(i)} V_j)^{\gamma_i - 1} \sum_{j \in \text{Children}(i)} \frac{\partial V_j}{\partial \theta_j} \frac{\partial \theta_j}{\partial \theta_i} \\
&= \gamma_i (\sum_{j \in \text{Children}(i)} V_j)^{\gamma_i} (\sum_{j \in \text{Children}(i)} V_j)^{-1} \sum_{j \in \text{Children}(i)} (-\gamma_j V_j u_j)(\frac{1}{\gamma_j} - \omega_j u_j)^{-1} \\
&= -\gamma_i V_i \sum_{j \in \text{Children}(i)} \frac{\gamma_j^2 u_j}{1 - \gamma_j \omega_j u_j} Q(j|i) \\
&= -\gamma_i V_i u_i,
\end{aligned}$$

where the third equality is due to Equations (C.15) and (C.16) and the last equality is because of the definition of u_i. So Equation (C.13) is true for node i.

According to Definition 1, we have

$$\omega_i = (\frac{1}{\gamma_i} - 1) \sum_{j \in \text{Children}(i)} \frac{Q(j|i)}{1 - \gamma_j} \omega_j = (\frac{1}{\gamma_i} - 1) \sum_{j \in \text{Children}(i)} \frac{V_j \omega_j}{(1 - \gamma_j) \sum_{j \in \text{Children}(i)} V_j},$$

thus we can also obtain

$$
\begin{aligned}
\frac{\partial \omega_i}{\partial \theta_i} &= (\frac{1}{\gamma_i} - 1) \sum_{j \in \text{Children}(i)} \frac{1}{1-\gamma_j} \frac{\frac{\partial (V_j \omega_j)}{\partial \theta_i} \sum_{j \in \text{Children}(i)} V_j - (V_j \omega_j) \sum_{j \in \text{Children}(i)} \frac{\partial V_j}{\partial \theta_i}}{(\sum_{j \in \text{Children}(i)} V_j)^2} \\
&= (\frac{1}{\gamma_i} - 1) \sum_{j \in \text{Children}(i)} \frac{1}{1-\gamma_j} \left[\frac{\frac{\partial V_j}{\partial \theta_j} \frac{\partial \theta_j}{\partial \theta_i} \omega_j + V_j \frac{\partial \omega_j}{\partial \theta_j} \frac{\partial \theta_j}{\partial \theta_i}}{\sum_{j \in \text{Children}(i)} V_j} - \frac{(V_j \omega_j) \sum_{j \in \text{Children}(i)} \frac{\partial V_j}{\partial \theta_j} \frac{\partial \theta_j}{\partial \theta_i}}{(\sum_{j \in \text{Children}(i)} V_j)^2} \right] \\
&= (\frac{1}{\gamma_i} - 1) \sum_{j \in \text{Children}(i)} \frac{1}{1-\gamma_j} \left[\frac{(-\gamma_j V_j u_j)(\frac{1}{\gamma_j} - \omega_j u_j)^{-1} \omega_j + V_j (\omega_j u_j - (\frac{1}{\gamma_j} - 1))(\frac{1}{\gamma_j} - \omega_j u_j)^{-1}}{\sum_{j \in \text{Children}(i)} V_j} \right. \\
&\left. - \frac{(V_j \omega_j) \sum_{j \in \text{Children}(i)} (-\gamma_j V_j u_j)(\frac{1}{\gamma_j} - \omega_j u_j)^{-1}}{(\sum_{j \in \text{Children}(i)} V_j)^2} \right] \\
&= (\frac{1}{\gamma_i} - 1) \sum_{j \in \text{Children}(i)} \frac{1}{1-\gamma_j} \left[\frac{\gamma_j \omega_j u_j V_j (1-\gamma_j) - V_j (1-\gamma_j)}{(1-\gamma_j \omega_j u_j) \sum_{j \in \text{Children}(i)} V_j} \right] \\
&+ (\frac{1}{\gamma_i} - 1) \sum_{j \in \text{Children}(i)} \frac{\omega_j}{1-\gamma_j} Q(j|i) * \sum_{j \in \text{Children}(i)} \frac{\gamma_j u_j}{1-\gamma_j \omega_j u_j} Q(j|i) \\
&= (\frac{1}{\gamma_i} - 1) \sum_{j \in \text{Children}(i)} (-Q(j|i)) + \omega_i u_i \\
&= \omega_i u_i - (\frac{1}{\gamma_i} - 1).
\end{aligned}
$$

So Equation (C.14) also holds for node i. Therefore Equations (C.13) and (C.14) are true for all nonleaf node i according to the principle of mathematical induction. □

Proof. Proof of Lemma 15: For notational brevity, let $\omega_a = \omega_a(S_a, \theta_a)$, $\omega_b = \omega_b(S_b, \theta_b)$, and $g_i = g_i(\mathbf{P}_i) = V_i(S_i, \mathbf{P}_i)(R_i(S_i, \mathbf{P}_i) - e_h^*)$, $V_i = V_i(S_i, \mathbf{P}_i)$, $R_i = R_i(S_i, \mathbf{P}_i)$ for all $i \in V$ in this proof. We first prove θ_i and θ_i' are equivalent. We denote the actual product in the lowest level m as r. Without loss of generality, $0 \leqslant d \leqslant m-2$, we assume that node i and $j \in$ Children(i) are ancestor nodes of r, and when node $d = m-1$, node j is the sibling node of r. The first derivative of the objective function $g_i = V_i(R_i - e_h^*)$ in the inner pricing subproblem with respect to price p_r of product r is

$$\frac{\partial g_i}{\partial p_r} = \frac{\partial V_i}{\partial p_r}(R_i - e_h^*) + V_i \frac{\partial R_i}{\partial p_r}.$$

Due to Claim 14, the above equation can be rewritten as

$$\frac{\partial g_i}{\partial p_r} = -\beta_r Q(r|i)\lambda_d^{m-1} V_i(R_i - e_h^*) + V_i\beta_r Q(r|i)\Bigg(\frac{1}{\beta_r} - m_r + \lambda_{r,d+1}^{m-1}\sum_{r'\in\eta_{r,d}} m_{r'}Q(r'|i)$$
$$+ \sum_{k=d+1}^{m-2}\sum_{\substack{r'\in\eta_{r,k}\\ r'\notin\eta_{r,k+1}}} m_{r'}\Bigg(\sum_{t=d+1}^{k}(\lambda_{r,t+1}^{m-1} - \lambda_{r,t}^{m-1})Q(r'|\eta_{r,t})\Bigg)$$
$$+ \sum_{r'\in\eta_{r,m-1}} m_{r'}\Bigg(\sum_{t=d+1}^{m-1}(\lambda_{r,t+1}^{m-1} - \lambda_{r,t}^{m-1})Q(r'|\eta_{r,t})\Bigg)\Bigg).$$

We let the first derivative $\partial g_i/\partial p_r = 0$, since both $Q(r|i)$ and V_i are not zero, after dividing $\beta_r Q(r|i)V_i$ on both sides of the above equation and collecting terms, we have

$$-\lambda_d^{m-1}(R_i - e_h^*) + \Bigg(\frac{1}{\beta_r} - m_r + \lambda_{r,d+1}^{m-1}\sum_{r'\in\eta_{r,d}} m_{r'}Q(r'|i) \tag{C.17}$$
$$+ \sum_{k=d+1}^{m-2}\sum_{\substack{r'\in\eta_{r,k}\\ r'\notin\eta_{r,k+1}}} m_{r'}\Bigg(\sum_{t=d+1}^{k}(\lambda_{r,t+1}^{m-1} - \lambda_{r,t}^{m-1})Q(r'|\eta_{r,t})\Bigg)$$
$$+ \sum_{r'\in\eta_{r,m-1}} m_{r'}\Bigg(\sum_{t=d+1}^{m-1}(\lambda_{r,t+1}^{m-1} - \lambda_{r,t}^{m-1})Q(r'|\eta_{r,t})\Bigg)\Bigg) = 0. \tag{C.18}$$

We know nonleaf node a is in level l ($d \leqslant l \leqslant m-1$) and its children nodes are indexed by b. Then according to (C.17) and the definition of node-specific adjusted markup, we have

$$\theta_a = \theta_b - \omega_b$$
$$= \lambda_{r,d+1}^{l}\sum_{r'\in\eta_{r,d}} m_{r'}Q(r'|i) + \sum_{k=d+1}^{m-2}\sum_{\substack{r'\in\eta_{r,k}\\ r'\notin\eta_{r,k+1}}} m_{r'}\Bigg(\sum_{t=d+1}^{l}(\lambda_{r,t+1}^{l} - \lambda_{r,t}^{l})Q(r'|\eta_{r,t})\Bigg)$$
$$+ \sum_{r'\in\eta_{r,m-1}} m_{r'}\Bigg(\sum_{t=d+1}^{l}(\lambda_{r,t+1}^{l} - \lambda_{r,t}^{l})Q(r'|\eta_{r,t})\Bigg) + \lambda_d^{m-1}(R_i - e_h^*) \tag{C.19}$$

From the RHS of (C.19), we can see that θ_a is independent of $\eta_{r,s}$ where $d \leqslant s \leqslant l$. Thus for another child node b' of node a, $\theta_a' = m_{b'} - \omega_{b'}$ also equals to the RHS of (C.19). Therefore, $\theta_a = \theta_a'$ for any node a in level $d, d+1, ..., m-1$.

Since $\theta_a = \theta_b - \omega_b$, In order to prove that there exists a one-to-one increasing correspondence between the node-specific adjusted markup θ_a and θ_b where $b \in \text{Children}(a)$, we need to prove

$$\frac{\partial\theta_a}{\partial\theta_b} = 1 - \frac{\partial\omega_b}{\partial\theta_b} = 1 - (\omega_b u_b - (\frac{1}{\gamma_b} - 1)) = \frac{1}{\gamma_b} - \omega_b u_b > 0,$$

where the second equality is due to Claim 15.

Next, we prove by induction on level in which node a is that $0 < u_a \leqslant \overline{B}_i$ and $0 < \omega_a \leqslant \frac{1-\gamma_a}{\gamma_a}\frac{1}{\underline{B}_a}$. If node a is in level $m-1$, then $0 < u_a = \sum_{b\in\text{Children}(a)} \beta_b Q(b|a) \leqslant \max_{b\in\text{Children}(a)}\{\beta_b\} = \overline{B}_a$. And $0 < \omega_a = (\frac{1}{\gamma_a}-1)\sum_{b\in\text{Children}(a)}\frac{Q(b|a)}{\beta_b} \leqslant \frac{1-\gamma_a}{\gamma_a}\frac{1}{\underline{B}_a}$. Thus it is true for the basic node a, then assume this holds for nonleaf and non-basic node b

$$0 < u_b \leqslant \overline{B}_b,$$
$$0 < \omega_b \leqslant \frac{1-\gamma_b}{\gamma_b}\frac{1}{\underline{B}_b}.$$

So we have

$$0 < \omega_b u_b \leqslant \frac{1-\gamma_b}{\gamma_b}\frac{\overline{B}_b}{\underline{B}_b}.$$

Thus we get

$$u_a = \sum_{b\in\text{Children}(a)} \frac{\gamma_b^2 u_b}{1-\gamma_b\omega_b u_b} Q(b|a) > \sum_{b\in\text{Children}(a)} \gamma_b^2 u_b Q(b|a) > 0,$$

and

$$\begin{aligned} u_a = \sum_{b\in\text{Children}(a)} \frac{\gamma_b^2 u_b}{1-\gamma_b\omega_b u_b} Q(b|a) &\leqslant \sum_{b\in\text{Children}(a)} \frac{\gamma_b^2 u_b}{1-\gamma_b \frac{1-\gamma_b}{\gamma_b}\frac{\overline{B}_b}{\underline{B}_b}} Q(b|a) \\ &\leqslant \max_{b\in\text{Children}(a)}\{\frac{\gamma_b^2 \overline{B}_b}{1-\gamma_b\frac{1-\gamma_b}{\gamma_b}\frac{\overline{B}_b}{\underline{B}_b}}\} \leqslant \overline{B}_a. \end{aligned}$$

So for u_a, we have $0 < u_a \leqslant \overline{B}_a$. And for ω_a, we have

$$\begin{aligned} 0 < \omega_a = (\frac{1}{\gamma_a}-1)\sum_{b\in\text{Children}(a)} \frac{Q(b|a)}{1-\gamma_b}\omega_b &\leqslant (\frac{1}{\gamma_a}-1)\sum_{b\in\text{Children}(a)} \frac{Q(b|a)}{1-\gamma_b}\frac{1-\gamma_b}{\gamma_b}\frac{1}{\underline{B}_b} \\ &\leqslant (\frac{1}{\gamma_a}-1)\frac{1}{\min_{b\in\text{Children}(a)}\{\underline{B}_b\gamma_b\}} = \frac{1-\gamma_a}{\gamma_a}\frac{1}{\underline{B}_a}. \end{aligned}$$

By the principle of mathematical induction, for any nonleaf node a, we have

$$0 < \omega_a u_a \leqslant \frac{1-\gamma_a}{\gamma_a}\frac{\overline{B}_a}{\underline{B}_a} < \frac{1}{\gamma_a},$$

thus we obtain

$$\frac{\partial\theta_a}{\partial\theta_b} = \begin{cases} 1 & b \text{ is a leaf node} \\ \frac{1}{\gamma_b} - \omega_b u_b > 0 & \text{o.w.} \end{cases} \tag{C.20}$$

Therefore, there exists a one-to-one increasing correspondence between the node-specific adjusted markup θ_a and θ_b, where $b \in$ Children(a). □

Proof of Lemma 16

Proof. For notational brevity, let and $g_i = g_i(\theta_i)$, $V_i = V_i(S_i, \theta_i)$, $R_i = R_i(S_i, \theta_i)$ in this proof. We denote the actual product in the lowest level m as r, thus $\theta_r = p_r - c_r - 1/\beta_r$.

Due to Claim 14, we have

$$\frac{\partial V_i}{\partial \theta_r} = -\beta_r Q(r|i) \lambda_d^{m-1} V_i,$$

then by applying Lemma 15 repeatedly and use the notation for the node-specific adjusted markup in Definition 1, we have

$$\frac{\partial R_i}{\partial \theta_r} = \lambda_{r,d+1}^{m-1} \beta_r Q(r|i) \left(R_i - \theta_i\right).$$

Therefore, we get

$$\begin{aligned} \frac{\partial g_i}{\partial \theta_r} &= \frac{\partial V_i}{\partial \theta_r}(R_i - e_h^*) + V_i \frac{\partial R_i}{\partial \theta_r} \\ &= -\beta_r Q(r|i) \lambda_d^{m-1} V_i (R_i - e_h^*) + V_i \lambda_{r,d+1}^{m-1} \beta_r Q(r|i) \left(R_i - \theta_i\right) \\ &= V_i \lambda_{r,d+1}^{m-1} \beta_r Q(r|i) \Big((1 - \gamma_i) R_i + \gamma_i e_h^* - \theta_i \Big), \end{aligned}$$

where we use the fact that $\lambda_{r,d}^{m-1} = \lambda_{r,d+1}^{m-1} * \gamma_i$ to get the last equality.

For the first order condition, let $\partial g_i / \partial \theta_r = 0$, we can obtain

$$\theta_i = \gamma_i e_h^* + (1 - \gamma_i) R_i(S_i, \theta_i),$$

which completes the proof. □

Proof of Corollary 5

Proof. For notational brevity, let $R_i = R_i(S_i, \theta_i)$. If i is root node, then we have

$$\frac{\partial R_i}{\partial \theta_r} = \lambda_{r,d+1}^{m-1} \beta_r Q(r|i) \left(R_i - \theta_i\right),$$

Where product r in the lowest level m. We use $\mathrm{An}(r, l)$ to denote the r's ancestor node in level l ($d \leqslant l \leqslant m_1$), then we obtain

$$\begin{aligned} \frac{\partial R_i}{\partial \theta_i} &= \sum_{r \in \eta_{r,d}} \frac{\partial R_i}{\partial \theta_r} \left(\frac{\partial \theta_{\mathrm{An}(r,m-1)}}{\partial \theta_r} \right)^{-1} \left(\frac{\partial \theta_{\mathrm{An}(r,m-2)}}{\partial \theta_{\mathrm{An}(r,m-1)}} \right)^{-1} \cdots \left(\frac{\partial \theta_{\mathrm{An}(r,d+1)}}{\partial \theta_{\mathrm{An}(r,d+2)}} \right)^{-1} \left(\frac{\partial \theta_i}{\partial \theta_{\mathrm{An}(r,d+1)}} \right)^{-1} \\ &= (R_i - \theta_i) \sum_{r \in \eta_{r,d}} \lambda_{r,d+1}^{m-1} \beta_r Q(r|i) \left(\frac{\partial \theta_{\mathrm{An}(r,m-1)}}{\partial \theta_r} \right)^{-1} \left(\frac{\partial \theta_{\mathrm{An}(r,m-2)}}{\partial \theta_{\mathrm{An}(r,m-1)}} \right)^{-1} \cdots \left(\frac{\partial \theta_i}{\partial \theta_{\mathrm{An}(r,d+1)}} \right)^{-1}. \end{aligned}$$

According to Lemma 15, we know

$$\left(\frac{\partial\theta_{\mathrm{An}(r,m-1)}}{\partial\theta_r}\right)^{-1}\left(\frac{\partial\theta_{\mathrm{An}(r,m-2)}}{\partial\theta_{\mathrm{An}(r,m-1)}}\right)^{-1}\cdots\left(\frac{\partial\theta_i}{\partial\theta_{\mathrm{An}(r,d+1)}}\right)^{-1} > 0.$$

Therefore R_i is strictly unimodal with respect to θ_i according to Lemma 2 in [19]. According to the unimodality of R_i, we have $R_i = \theta_i$ at optimality. □

Proof of Proposition 11

Proof. For root node, given optimal assortment S^*_{root}, the optimal price vector P^*_{root} can be uniquely computed after getting θ^*_{root} by $R_{\mathrm{root}}(S^*_{\mathrm{root}}, \theta_{\mathrm{root}}) = \theta_{\mathrm{root}}$. Since there is a one-to-one correspondence between the node-specific adjusted markups, θ^*_i can also be uniquely identified once θ^*_{root} is given. Thus we have $R_i(S^*_i, \mathbf{P}^*_i) = R_i(S^*_i, \theta^*_i)$. From Lemma 16, we know $\theta^*_i = \gamma_i e^*_h + (1-\gamma_i)R_i(S^*_i, \theta^*_i) = \gamma_i e^*_h + (1-\gamma_i)R_i(S^*_i, \mathbf{P}^*_i) = e^*_i$, where is last equality is according to the definition of e^*_i. □

Proof of Proposition 12

We use lemma 24 to prove proposition 12.

Lemma 24. *For an assortment S_i of non-root node i in level d ($d \neq 0$), we have*

$$R_i(S_i, \theta_i) = \theta_i + \frac{\gamma_i}{1-\gamma_i}\omega_i(S_i, \theta_i)$$

Proof. We let $R_i = R_i(S_i, \theta_i)$, $\omega_i = \omega_i(S_i, \theta_i)$ and $Q(j|i) = Q(S_j, \theta_i|S_i, \theta_i)$ for notational brevity. This lemma can be proved by induction on level d. If node i is in level $d = m-1$, we can obtain

$$\begin{aligned} R_i &= \sum_{j\in\mathrm{Children}(i)} Q(j|i)R_j = \sum_{j\in\mathrm{Children}(i)} Q(j|i)(p_j - c_j) = \sum_{j\in\mathrm{Children}(i)} Q(j|i)(\theta_i + \frac{1}{\beta_j}) \\ &= \theta_i + \sum_{j\in\mathrm{Children}(i)} \frac{Q(j|i)}{\beta_j} && \text{(C.21)} \\ &= \theta_i + \frac{\gamma_i}{1-\gamma_i}\omega_i, && \text{(C.22)} \end{aligned}$$

where Equation (C.21) is due to the fact that there is no no-purchase option in the lowest level m, then $\sum_{j\in\mathrm{Children}(i)} Q(j|i) = 1$, and Equation (C.22) is because of Definition 1. Thus this claim holds for level $m-1$. Assume that this claim also holds for level l ($2 \leqslant l \leqslant m-1$) that node $j \in \mathrm{Children}(i)$ is in, then we have that $R_j = \theta_j + [\gamma_j/(1-\gamma_j)]\omega_j$. When it comes to level $l-1$, we get

$$R_i = \sum_{j\in \text{Children}(i)} Q(j|i)R_j = \sum_{j\in \text{Children}(i)} Q(j|i)(\theta_j + \frac{\gamma_j}{1-\gamma_j}\omega_j) \tag{C.23}$$

$$= \sum_{j\in \text{Children}(i)} Q(j|i)(\theta_i + \omega_j + \frac{\gamma_j}{1-\gamma_j}\omega_j) \tag{C.24}$$

$$= \theta_i + \sum_{j\in \text{Children}(i)} \frac{Q(j|i)}{1-\gamma_j}\omega_j = \theta_i + \frac{\gamma_i}{1-\gamma_i}\omega_i, \tag{C.25}$$

where Equation (C.23) is due to induction hypothesis, Equation (C.24) is because of the definition of θ_i and similarly Equation (C.25) is due to $\sum_{j\in \text{Children}(i)} Q(j|i) = 1$. Note that $\sum_{j\in \text{Children}(i)} Q(j|i) < 1$ only if i is root node. So it holds for all the non-root nodes, establishing the lemma. □

Proof. Proof of Proposition 12: Throughout the proof, we let $R_i = R_i(S_i, \theta_i)$, $\omega_i = \omega_i(S_i, \theta_i)$ and $Q(j|i) = Q(S_j, \theta_i|S_i, \theta_i)$ for notational brevity. First, if i is the root node, we have $\gamma_i = 0$ and $\theta_h^* = 0$, thus both equations hold according to Corollary 5 and Definition 1. Second, we consider the case when i is not the root node. By Lemma 16 and Lemma 24, we have

$$\theta_h^* = \frac{\theta_i - (1-\gamma_i)R_i}{\gamma_i} = \frac{\theta_i - (1-\gamma_i)(\theta_i + \frac{\gamma_i}{1-\gamma_i}\omega_i)}{\gamma_i} = \theta_i - \omega_i,$$

thus the second equation is true.

Thus after plugging the second equation into the equation in Lemma 24, we obtain

$$R_i = \theta_i + \frac{\gamma_i}{1-\gamma_i}\omega_i = \theta_h^* + \omega_i + \frac{\gamma_i}{1-\gamma_i}\omega_i = \theta_h^* + \frac{1}{1-\gamma_i}\omega_i,$$

which establishes the first equation. We also have $R_i > \theta_h^*$ because $\omega_i > 0$ for nonempty assortment S_i. □

Proof of Corollary 6

Proof. From the first equation in Proposition 12, we get

$$\frac{\partial \theta_h^*}{\partial \theta_i} = 1 - \frac{\partial \omega_i(S_i,\theta_i)}{\partial \theta_i} = 1 - \left[\omega_i(S_i,\theta_i)u_i(S_i,\theta_i) - (\frac{1}{\gamma_i} - 1)\right] = \frac{1}{\gamma_i} - \omega_i(S_i,\theta_i)u_i(S_i,\theta_i) > 0,$$

where the second equality is due to Claim 15 and the last inequality holds according to similar induction proof of Lemma 15. Thus there exists a one-to-one increasing correspondence between θ_h^* and θ_i, then given a θ_h^* and a fixed assortment S_i, corresponding θ_i that is optimal to problem (4.5) can be uniquely identified. □

Proof of Theorem 8

Proof. Similar to assortment subproblem, we can rewrite basic joint subproblem as

$$\max_{S_i \subseteq \Im_i} \max_{\theta_i \in \mathbb{R}} V_i(S_i, \theta_i)^{1/\gamma_i} (R_i(S_i, \theta_i) - \theta_i^*). \tag{C.26}$$

Given an assortment S_i, let $g_i'(\theta_i) = V_i(S_i, \theta_i)^{1/\gamma_i}(R_i(S_i, \theta_i) - \theta_i^*)$. For ease of presentation, we let $g_i' = g_i'(\theta_i)$, $V_i = V_i(S_i, \theta_i)$ and $R_i = R_i(S_i, \theta_i)$. r represents the descendant of node i in the lowest level m and $\text{An}(r, l)$ denotes r's ancestor node in level l ($d \leqslant l \leqslant m_1$). As proof of Lemma 4, we have

$$\begin{aligned}
\frac{\partial V_i}{\partial \theta_r} &= -\beta_r Q(r|i) \lambda_d^{m-1} V_i, \\
\frac{\partial R_i}{\partial \theta_r} &= \lambda_{r,d+1}^{m-1} \beta_r Q(r|i) \left(R_i - \theta_i\right).
\end{aligned}$$

Thus we obtain

$$\begin{aligned}
\frac{\partial g_i'}{\partial \theta_r} &= \frac{1}{\gamma_i} V_i^{1/\gamma_i - 1} \frac{\partial V_i}{\partial \theta_r}(R_i - \theta_i^*) + V_i^{1/\gamma_i} \frac{\partial R_i}{\partial \theta_r} \\
&= -V_i^{1/\gamma_i} \left(\theta_i - \theta_i^*\right) * \beta_r Q(r|i) \lambda_{r,d+1}^{m-1}.
\end{aligned}$$

So the first derivative with respect to θ_i is

$$\frac{\partial g_i'}{\partial \theta_i} = \sum_{r \in \eta_{r,d}} \frac{\partial g_i'}{\partial \theta_r} \left(\frac{\partial \theta_{\text{An}(r,m-1)}}{\partial \theta_r}\right)^{-1} \left(\frac{\partial \theta_{\text{An}(r,m-2)}}{\partial \theta_{\text{An}(r,m-1)}}\right)^{-1} \cdots \left(\frac{\partial \theta_{\text{An}(r,d+1)}}{\partial \theta_{\text{An}(r,d+2)}}\right)^{-1} \left(\frac{\partial \theta_i}{\partial \theta_{\text{An}(r,d+1)}}\right)^{-1}$$

$$= -V_i^{1/\gamma_i} \left(\theta_i - \theta_i^*\right) \sum_{r \in \eta_{r,d}} \beta_r Q(r|i) \lambda_{r,d+1}^{m-1} \left(\frac{\partial \theta_{\text{An}(r,m-1)}}{\partial \theta_r}\right)^{-1} \left(\frac{\partial \theta_{\text{An}(r,m-2)}}{\partial \theta_{\text{An}(r,m-1)}}\right)^{-1} \cdots \left(\frac{\partial \theta_i}{\partial \theta_{\text{An}(r,d+1)}}\right)^{-1}.$$

According to Lemma 15, we know

$$\left(\frac{\partial \theta_{\text{An}(r,m-1)}}{\partial \theta_r}\right)^{-1} \left(\frac{\partial \theta_{\text{An}(r,m-2)}}{\partial \theta_{\text{An}(r,m-1)}}\right)^{-1} \cdots \left(\frac{\partial \theta_i}{\partial \theta_{\text{An}(r,d+1)}}\right)^{-1} > 0.$$

We get $\theta_i = \theta_i^*$ by the first order condition. Thus problem (C.26) can be rewritten as

$$\max_{S_i \subseteq \Im_i} V_i(S_i, \theta_i^*)^{1/\gamma_i} (R_i(S_i, \theta_i^*) - \theta_i^*). \tag{C.27}$$

By Lemma 24, we have

$$R_i(S_i, \theta_i^*) = \theta_i^* + \frac{\gamma_i}{1 - \gamma_i} \omega_i(S_i, \theta_i^*).$$

Plug the above equation into problem (C.27), we obtain

$$\max_{S_i \subseteq \Im_i} V_i(S_i, \theta_i^*)^{1/\gamma_i} \frac{\gamma_i}{1 - \gamma_i} \omega_i(S_i, \theta_i^*),$$

which can be further simplified as problem (4.7) according to the definition of $\omega_i(S_i, \theta_i^*)$ in Definition 1. □

Proof of Corollary 7

Proof. By Theorem 8 and Definition 1, problem (4.7) can be written as

$$\max_{S_j \subseteq \Im_j} \sum_{k \in \text{Children}(j)} \frac{V_k(S_k, \theta_k)}{\beta_k}$$
$$\text{s.t.} \quad \theta_j^* = \theta_k - c_k - \frac{1}{\beta_k},$$

which can be further simplified as

$$\max_{S_j \subseteq \Im_j} \sum_{k \in \text{Children}(j)} \frac{V_k(S_k, \theta_j^* + c_k + 1/\beta_k)}{\beta_k}.$$

The above problem is essentially finding $\mathbb{C}_j$ leaf nodes with largest ratio of preference weight to price sensitivity parameter from N_j leaf nodes, which only depends on the ordering of the ratio $V_k(S_k, \theta_j^* + c_k + 1/\beta_k)/\beta_k$. The total number of operations in need is $O(N_j \log N_j)$ (sorting) + $O(\mathbb{C}_j)$ (printing the output of first $\mathbb{C}_j$ leaf nodes) = $O(N_j \log N_j)$. □

Proof of Corollary 8

Proof. Throughout the proof, we let $\hat{V}_j = V_j(\hat{S}_j, \theta_j^*)$, $V_j^* = V_j(S_j^*, \theta_j^*)$, $\hat{R}_j = R_j(\hat{S}_j, \theta_j^*)$ and $R_j^* = R_j(S_j^*, \theta_j^*)$ for ease of presentation. We index the parent node of the basic node j by i. Since $\hat{S}_j$ is the optimal solution to problem (4.8) at the basic node $j \in \mathscr{B}$, according to Lemma 14, it satisfies

$$\hat{V}_j(\hat{R}_j - \theta_i^*) = V_j^*(R_j^* - \theta_i^*).$$

Let $\hat{S}_i = \bigcup_{j \in \text{Children}(i)} \hat{S}_j$, due to Lemma 13, we obtain

$$\hat{V}_i(\hat{R}_i - \theta_{\text{Parent}(i)}^*) = V_i^*(R_i^* - \theta_{\text{Parent}(i)}^*).$$

Repeat this process based on Lemma 13 until we get $\hat{S}_{\text{root}} = \bigcup_{k \in \text{Children(root)}} \hat{S}_k = \bigcup_{j \in \mathscr{B}} \hat{S}_j$, then we have

$$\hat{V}_{\text{root}}(\hat{R}_{\text{root}} - \theta_{\text{Parent(root)}}^*) = V_{\text{root}}^*(R_{\text{root}}^* - \theta_{\text{Parent(root)}}^*).$$

According to the definition, we have $\hat{V}_{\text{root}} = V_{\text{root}}^* = 1$ and $\theta_{\text{Parent(root)}}^* = \theta_{\text{Parent(root)}}^* = 0$, implying $\hat{R}_{\text{root}} = R_{\text{root}}^*$. □

Proof of Lemma 17

Proof. Prove this lemma by contradiction, assume that $\tilde{S}_j(\theta_j)$ is the optimal solution to problem (4.9) at θ_j and $|\tilde{S}_j(\theta_j)| < \mathbb{C}_j$. Since $|\tilde{S}_j(\theta_j)| < \mathbb{C}_j$, there must exist a product

k' satisfying that $k' \notin \tilde{S}_j(\theta_j)$ but $k' \in \text{Children}(j)$. Let $\tilde{S}'_j(\theta_j) = \tilde{S}_j(\theta_j) \cup k'$, then we have that $\tilde{S}'_j(\theta_j)$ is feasible since $|\tilde{S}'_j(\theta_j)| \leqslant \mathbb{C}_j$ and $\tilde{S}'_j(\theta_j)$ strictly dominates $\tilde{S}_j(\theta_j)$ since $\sum_{k\in\tilde{S}'_j(\theta)} V_k(S_k, \theta_j + 1/\beta_k)/\beta_k > \sum_{k\in\tilde{S}_j(\theta)} V_k(S_k, \theta_j + 1/\beta_k)/\beta_k$, which contradicts with the hypothesis that $\tilde{S}_j(\theta_j)$ is the optimal solution to problem (4.9) at θ_j and $|\tilde{S}_j(\theta_j)| < \mathbb{C}_j$. □

Proof of Lemma 18

Proof. Let $\theta'_j \in \mathscr{C}_j$, $\underline{S}_j = \lim_{\epsilon\to 0} \tilde{S}_j(\theta'_j - \epsilon)$ and $\overline{S}_j = \lim_{\epsilon\to 0} \tilde{S}_j(\theta'_j + \epsilon)$ for a small $\epsilon > 0$. [39] show that at changing point, one product would be replaced by another product. Without loss of generality, assume that $\theta'_j = I(k_1, k_2)$ and $\overline{S}_j = (\underline{S}_j \backslash \{k_1\}) \cup \{k_2\}$ for $k_1, k_2 \in \text{Children}(j)$. Since product k_1 is replaced by product k_2 at θ'_j, we have $h_{k_1}(\theta'_j) = h_{k_2}(\theta'_j)$ and $\beta_{k_1} > \beta_{k_2}$, which implies that $V_{k_1}(S_{k_1}, \theta'_j + c_{k_1} + \beta_{k_1})/\beta_{k1} = V_{k_2}(S_{k_2}, \theta'_j + c_{k_2} + \beta_{k_2})/\beta_{k_2}$. Thus we have $V_{k_1}(S_{k_1}, \theta'_j + c_{k_1} + \beta_{k_1}) > V_{k_2}(S_{k_2}, \theta'_j + c_{k_2} + \beta_{k_2})$, so $V_j(\underline{S}_j, \theta'_j) > V_j(\overline{S}_j, \theta'_j)$. Thus the preference weight $V_j(\tilde{S}_j(\theta_j), \theta_j)$ of basic node j drops discontinuously at the changing point θ'_j. □

Proof of Lemma 19

Proof. According to Lemma 15, we know that there is an increasing one-to-one correspondence between θ_j and θ_i if $\tilde{S}_j(\theta_j)$ does not change, thus we only need to show θ_i is discontinuous in θ_j at the changing point, where $\tilde{S}_j(\theta_j)$ changes to a different assortment. Without loss of generality, assume $\theta'_j \in \mathscr{C}_j$ is the changing point, and let $\underline{S}_j = \lim_{\epsilon\to 0} \tilde{S}_j(\theta'_j - \epsilon)$ and $\overline{S}_j = \lim_{\epsilon\to 0} \tilde{S}_j(\theta'_j + \epsilon)$. By Lemma 18, we have that $V_j(\underline{S}_j, \theta'_j) > V_j(\overline{S}_j, \theta'_j)$, implying that

$$\begin{aligned}
\omega_j(\underline{S}_j, \theta'_j) &= (\frac{1}{\gamma_j} - 1) \sum_{k\in\text{Children}(j)} \frac{V_k(\underline{S}_k, \theta'_j + \beta_k)}{\beta_k} \frac{1}{V_j(\underline{S}_j, \theta'_j)^{1/\gamma_j}} \\
&< (\frac{1}{\gamma_j} - 1) \sum_{k\in\text{Children}(j)} \frac{V_k(\overline{S}_k, \theta'_j + \beta_k)}{\beta_k} \frac{1}{V_j(\overline{S}_j, \theta'_j)^{1/\gamma_j}} = \omega_j(\overline{S}_j, \theta'_j),
\end{aligned}$$

where the inequality holds due to the fact that $\sum_{k\in\text{Children}(j)} V_k(\underline{S}_k, \theta'_j + \beta_k)/\beta_k = \sum_{k\in\text{Children}(j)} V_k(\overline{S}_k, \theta'_j + \beta_k)/\beta_k$. Thus the function $\theta_i = \theta_j - \omega_j(\tilde{S}_j(\theta_j), \theta_j)$ drops discontinuously at θ'_j since $\theta'_j - \omega_j(\underline{S}_j, \theta'_j) > \theta'_j - \omega_j(\overline{S}_j, \theta'_j)$. Moreover, let $\underline{\theta}_i = \theta'_j - \omega_j(\overline{S}_j, \theta'_j)$ and $\overline{\theta}_i = \theta'_j - \omega_j(\underline{S}_j, \theta'_j)$, then when $\theta_i \in [\underline{\theta}_i, \overline{\theta}_i]$, it corresponds to two different θ_j's, which completes the lemma. □

Proof of Proposition 14

Proof. We prove this proposition by induction. It's true for basic nodes. Assume that $\mathscr{A}_i = \{\bigcup_{j\in\text{Children}(i)} \tilde{S}_j(\mathcal{F}_j(\theta_i)) : \theta_i \in \mathbb{R}\}$ includes an optimal solution to assortment subproblem

(4.6) at node i. To construct a candidate collection of assortments $\mathscr{A}_h$ including an optimal assortment at node h. Consider the assortment subproblem (4.6) at node i, let $\hat{S}_i(\theta_h)$ be an optimal solution to the following problem

$$\begin{aligned} \max_{S_i \subseteq \mathscr{A}_i} \quad & \frac{V_i(S_i, \theta_i)\omega_i(S_i, \theta_i)}{1-\gamma_i} \\ \text{s.t.} \quad & \theta_h = \theta_i - \omega_i(S_i, \theta_i), \end{aligned} \tag{C.28}$$

where we optimize over $\mathscr{A}_i$ instead of $\Im_i$ in problem (4.6) since $\mathscr{A}_i$ includes an optimal assortment by the induction hypothesis. Then the above problem is equivalent to the following optimization problem since $\mathscr{A}_i = \{\tilde{S}_i(\theta_i) : \theta_i \in \mathbb{R}\}$.

$$\begin{aligned} \max_{\theta_i \in \mathbb{R}} \quad & \frac{V_i(\tilde{S}_i(\theta_i), \theta_i)\omega_i(\tilde{S}_i(\theta_i), \theta_i)}{1-\gamma_i} \\ \text{s.t.} \quad & \theta_h = \theta_i - \omega_i(\tilde{S}_i(\theta_i), \theta_i). \end{aligned}$$

Since $\mathcal{F}_i(\theta_h)$ is optimal to the above problem, then $\tilde{S}_i(\mathcal{F}_i(\theta_h))$ is the optimal solution to problem (C.28), thus $\bigcup_{i \in \text{Children}(h)} \tilde{S}_i(\mathcal{F}_i(\theta_h^*))$ is the optimal solution to assortment problem (4.6) at node h by Lemma 13. From the definition of $\mathscr{A}_h$, we have that $\bigcup_{i \in \text{Children}(h)} \tilde{S}_i(\mathcal{F}_i(\theta_h^*)) \in \mathscr{A}_h$. □

Proof of Lemma 20

Proof. For ease of presentation, we denote $V_i = V_i(\tilde{S}_i(\theta_i), \theta_i)$, $\omega_i = \omega_i(\tilde{S}_i(\theta_i), \theta_i)$ and $T_i = T_i(\tilde{S}_i(\theta_i), \theta_h)$. We have

$$\frac{\partial T_i}{\partial \theta_h} = \frac{1}{1-\gamma_i}\left(\frac{\partial V_i}{\partial \theta_h}\omega_i + V_i \frac{\partial \omega_i}{\partial \theta_h}\right) = \frac{1}{1-\gamma_i}\left(\frac{\partial V_i}{\partial \theta_i}\omega_i + V_i \frac{\partial \omega_i}{\partial \theta_i}\right)\frac{\partial \theta_i}{\partial \theta_h}.$$

If i is a leaf node, then

$$\frac{\partial T_i}{\partial \theta_h} = \frac{1}{1-\gamma_i}\left(\frac{\partial V_i}{\partial p_i}\omega_i + V_i \frac{\partial \omega_i}{\partial p_i}\right)\frac{\partial p_i}{\partial \theta_h} = \left(-\beta_j V_j \frac{1}{\beta_j} + 0\right) * 1 = -V_j,$$

where the second equality is due to the fact that $\gamma_i = 0$ for all leaf node i from Definition 1. If i is nonleaf node, then according to Claim 15 and Equation (C.20), we can get

$$\begin{aligned} \frac{\partial T_i}{\partial \theta_h} &= \frac{1}{1-\gamma_i}\left(\frac{\partial V_i}{\partial \theta_i}\omega_i + V_i \frac{\partial \omega_i}{\partial \theta_i}\right)\frac{\partial \theta_i}{\partial \theta_h} \\ &= \frac{1}{1-\gamma_i}\left(-\gamma_i V_i u_i \omega_i + V_i\left(\omega_i u_i - (\frac{1}{\gamma_i} - 1)\right)\right)(\frac{1}{\gamma_i} - \omega_i u_i)^{-1} \\ &= -V_i. \end{aligned}$$

Thus item 1 of this lemma holds. Furthermore, the second derivative of T_i with respect to θ_h is

$$\frac{\partial^2 T_i}{\partial \theta_h^2} = -\frac{\partial V_i}{\partial \theta_h} = -\frac{\partial V_i}{\partial \theta_i}\frac{\partial \theta_i}{\partial \theta_h} = \gamma_i V_i u_i \omega_i (\frac{1}{\gamma_i} - \omega_i u_i)^{-1} > 0,$$

according to Lemma 15. So item 2 of this lemma is true. To prove the third item of this lemma, without loss of generality, we assume that $V(S_1, \theta_i(S_1, \theta_h)) > V(S_2, \theta_i(S_2, \theta_h)))$, then we have

$$\frac{\partial(T(S_1, \theta_h) - T(S_2, \theta_h))}{\partial \theta_h} = -\big(V_i(S_1, \theta_i(S_1, \theta_h)) > V_i(S_2, \theta_i(S_2, \theta_h)))\big) < 0,$$

which implies that $T(S_1, \theta_h)$ and $T(S_2, \theta_h)$ intersect at most once in θ_h domain. □

Proof of Proposition 15

Proof. First, we define set of *changing* points for nonleaf node $i \in V$ as $\mathscr{C}_i = \{\theta_i^0, \theta_i^1, \cdots, \theta_i^{D_i - 1}, \theta_i^{D_i}\}$ where $\theta_i^0 = -\infty$ and $\theta_i^{D_i} = \infty$, such that $\tilde{S}_i(\theta_i)$ does not change when $\theta_i \in [\theta_i^{d-1}, \theta_i^d]$ for $d = 1, \cdots, D_i$, and define set of *intersection* points for nonleaf node $i \in V$ as $\mathscr{T}_h^i = \{\theta_h^{i,0}, \theta_h^{i,1}, \cdots, \theta_h^{i,U_i - 1}, \theta_h^{i,U_i}\}$ where $\theta_h^{i,0} = -\infty$ and $\theta_h^{i,U_i} = \infty$, such that $\tilde{S}_i(\mathcal{F}_i(\theta_h))$ does not change when $\theta_h \in [\theta_h^{i,u-1}, \theta_h^{i,u}]$ for $u = 1, \cdots, U_i$. Then the set of changing points for node h is $\mathscr{C}_h = \bigcup_{i \in \text{Children}(h)} \mathscr{T}_h^i$ according to the above definitions. After relabeling, denote the points in $\mathscr{C}_h$ as $\{\theta_h^0, \theta_h^1, \cdots, \theta_h^{D_h - 1}, \theta_h^{D_h}\}$ where $\theta_h^0 = -\infty$ and $\theta_h^{D_h} = \infty$, we have $D_h \leqslant \sum_{i \in \text{Children}(h)} U_i$ according to the definition of $\mathscr{C}_h$ and $\mathscr{T}_h^i$.

Next, we prove that for any two different $S_1, S_2 \in \mathscr{A}_i$, $V(S_1, \theta_i(S_1, \theta_h))$ and $V(S_2, \theta_i(S_2, \theta_h)))$ do not intersect in θ_h domain. Without loss of generality, assume that $S_1(\theta_i) = \tilde{S}_i(\theta_i)$ where $\theta_i \in [\theta_i^{d_1 - 1}, \theta_i^{d_1}]$ and $S_2(\theta_i) = \tilde{S}_i(\theta_i)$ where $\theta_i \in [\theta_i^{d_2 - 1}, \theta_i^{d_2}]$ for $d_1 < d_2$. We denote $S_1 = S_1(\theta_i)$ and $S_2 = S_1(\theta_i)$ for notational purpose. We claim that $V_i(\tilde{S}_i(\theta_i), \theta_i)$ drops discontinuously at any changing point $\theta_i' \in \mathscr{C}_i$ which can be proved by induction on level of node i's since it is true for basic nodes by Lemma 18. Thus $V_i(S_1, \theta_i) > V_i(S_2, \theta_i)$ in θ_i domain.

We also have that $\sum_{j \in \text{Children}(i)} V_j(\tilde{S}_j(\mathcal{F}_j(\theta_i)), \mathcal{F}_j(\theta_i))\omega_j(\tilde{S}_j(\mathcal{F}_j(\theta_i)), \mathcal{F}_j(\theta_i))$ is a continuous function of θ_i because the optimal objective value $V_j(\tilde{S}_j(\mathcal{F}_j(\theta_i)), \mathcal{F}_j(\theta_i))\omega_j(\tilde{S}_j(\mathcal{F}_j(\theta_i)), \mathcal{F}_j(\theta_i))$ is continuous at intersection point $\theta_i^{j,u} \in \mathscr{T}_i^j$ for $u = 1, ..., U_j$. Thus $\omega_i(\tilde{S}_i(\theta_i), \theta_i) = (1/\gamma_i - 1)/V_i(\tilde{S}_i(\theta_i), \theta_i)^{1/\gamma_i} * \sum_{j \in \text{Children}(i)} V_j(\tilde{S}_j(\mathcal{F}_j(\theta_i)), \mathcal{F}_j(\theta_i))\omega_j(\tilde{S}_j(\mathcal{F}_j(\theta_i)), \mathcal{F}_j(\theta_i))/(1 - \gamma_j)$ increases discontinuously at any changing point $\theta_i' \in \mathscr{C}_i$, so we have $\theta_i(S_1, \theta_h) = \theta_h + \omega_i(S_1, \theta_i(S_1, \theta_h)) < \theta_h + \omega_i(S_2, \theta_i(S_2, \theta_h)) = \theta_i(S_2, \theta_h)$.

Therefore, we have that $V_i(S_1, \theta_i(S_1, \theta_h)) > V_i(S_2, \theta_i(S_1, \theta_h)) > V_i(S_2, \theta_i(S_2, \theta_h))$, where the first inequality is due to the argument at the end of second paragraph of this proof by setting $\theta_i = \theta_i(S_1, \theta_h)$ and the last inequality holds since $V_i(S_2, \theta_i)$ is a decreasing function of θ_i. Then by the third item in Lemma 20, the objective function of problem (4.12) for any two candidate assortments in $\mathscr{A}_i$ intersects at most once by the arbitrariness of S_1 and S_2. Thus the size of $\mathscr{T}_h^i$ is at most D_i, implying that $U_i \leqslant D_i$.

As a result, we get $|\mathscr{A}_h| = D_h \leqslant \sum_{i \in \text{Children}(h)} U_i \leqslant \sum_{i \in \text{Children}(h)} D_i = \sum_{i \in \text{Children}(h)} |\mathscr{A}_i|$.

Moreover, $|\mathscr{A}_{\text{root}}| \leqslant \sum_{h \in \text{Children(root)}} |\mathscr{A}_h| \leqslant \sum_{h \in \text{Children(root)}} \sum_{i \in \text{Children}(h)} |\mathscr{A}_i| \leqslant \cdots \leqslant \sum_{j \in \mathscr{B}} |\mathscr{A}_j|$, where $\mathscr{B}$ is the set of basic nodes. Since $|\mathscr{A}_j|$ has size $O(N_j)$, let $M > 0$ such that $|\mathscr{A}_j| \leqslant MN_j$ for any $j \in \mathscr{B}$. Thus we have $|\mathscr{A}_{\text{root}}| \leqslant \sum_{h \in \text{Children(root)}} |\mathscr{A}_h| \leqslant \sum_{j \in \mathscr{B}} |\mathscr{A}_j| \leqslant M \sum_{j \in \mathscr{B}} N_j = MN$, implying that $|\mathscr{A}_{\text{root}}|$ has size $O(N)$. Since $|\mathscr{A}_{\text{root}}| \geqslant |\mathscr{A}_i|$ for any $i \in V$, thus $|\mathscr{A}_i|$ is of size $O(N)$. □

Proof of Theorem 9

Proof. Due to Corollary 7, line 4 takes $O(N_j \log N_j)$ for basic node j. Then for all basic nodes and G grid points, the running time is $O(GN \log K)$ where K denotes the maximum products that a basic node can have in the tree logit model, since $G \sum_{j \in \mathscr{B}} N_j \log N_j \leqslant G \sum_{j \in \mathscr{B}} N_j \log K = GN \log K$. Line 5 takes $O(1)$, line 6 takes $O(\log G)$ by applying binary search and lines 7-11 run in $O(1)$ since line 22 involves one numerical comparison, thus lines 2-13 take $O(GN \log K + G|\mathscr{B}| + G|\mathscr{B}| \log G + G|\mathscr{B}|) = O(GN \log G)$ since G is larger than K in general.

Similarly, both line 16 and line 17 take $O(1)$. Line 18 also runs in $O(\log G)$ by using binary search and lines 19-23 take $O(1)$ as well. Since the number of nodes in level $m-2, m-1, ..., 0$ is less than twice the number of leaf nodes N, lines 14-25 takes $O(GN + GN + GN \log G + GN) = O(GN \log G)$.

For the final part of this algorithm, lines 26-28 take $O(G)$ and line 30 runs in $O(G)$ as well by looking up previous stored table. Hence, the overall complexity of Algorithm JCAPO-C is $O(GN \log G + GN \log G + G) = O(GN \log G)$. Moreover, if the spacings of grid points are identical, line 6 and line 18 take $O(1)$ by rounding up to the nearest grid points. Therefore, this algorithm would take $O(GN \log K + GN + G) = O(GN \log K)$, which is more efficient. □

Proof of Lemma 22

We first propose two claims to facilitate proving this lemma.

Claim 16. *For an arbitrary nonleaf node $i \in V$ and its parent node h, assume $R_i(S_i^*, \mathbf{P}_i^*) > e_h^*$. If there exists an assortment $\hat{S}_i \subseteq \Im_i$, a price vector $\hat{\mathbf{P}}_i$ and parameter $\alpha \geqslant 1$ such that*

$$\alpha V_i(\hat{S}_i, \hat{\mathbf{P}}_i)^{1/\gamma_i} \left(R_i(\hat{S}_i, \hat{\mathbf{P}}_i) - e_i^* \right) \geqslant V_i(S_i^*, \mathbf{P}_i^*)^{1/\gamma_i} \left(R_i(S_i^*, \mathbf{P}_i^*) - e_i^* \right). \tag{C.29}$$

Then we have

$$\alpha V_i(\hat{S}_i, \hat{\mathbf{P}}_i) \left(R_i(\hat{S}_i, \hat{\mathbf{P}}_i) - e_h^* \right) \geqslant V_i(S_i^*, \mathbf{P}_i^*) \left(R_i(S_i^*, \mathbf{P}_i^*) - e_h^* \right). \tag{C.30}$$

If the inequality in (C.29) is strict for some $j \in$ Children(i), then the inequality in (C.30) is strict as well.

Claim 17. *For an arbitrary nonleaf node $i \in V$ and its parent node h, assume $R_i(S_i^*, \mathbf{P}_i^*) > e_h^*$. If for all $j \in \text{Children}(i)$ and a parameter $\alpha \geqslant 1$, there exists an assortment $\hat{S}_j \subseteq \Im_j$ and a price vector $\hat{\mathbf{P}}_j$ such that*

$$\alpha V_j(\hat{S}_j, \hat{\mathbf{P}}_j)\left(R_j(\hat{S}_j, \hat{\mathbf{P}}_j) - e_i^*\right) \geqslant V_j(S_j^*, \mathbf{P}_j^*)\left(R_j(S_j^*, \mathbf{P}_j^*) - e_i^*\right).$$

Let $(\hat{S}_i, \hat{\mathbf{P}}_i) = \bigcup_{j \in \text{Children}(i)} (\hat{S}_j, \hat{\mathbf{P}}_j)$, then we have

$$\alpha V_i(\hat{S}_i, \hat{\mathbf{P}}_i)\left(R_i(\hat{S}_i, \hat{\mathbf{P}}_i) - e_h^*\right) \geqslant V_i(S_i^*, \mathbf{P}_i^*)\left(R_i(S_i^*, \mathbf{P}_i^*) - e_h^*\right).$$

Proof of Claim 16 and 17 follows directly from the proof of Claim 12 and 13, respectively.

Proof. Proof of Lemma 22: Since $\hat{S}_j^\alpha$ is an α-approximate solution to problem (4.8) at basic node j. Then the price vector $\hat{\mathbf{P}}_j^\alpha$ can be uniquely determined by θ_j^* and $\hat{S}_j^\alpha$ via $\theta_j^* = p_k - c_k - 1/\beta_k$ where $k \in \hat{S}_j^\alpha$. Problem (4.8) and problem (4.3) at node j under space constraints are equivalent formulations, thus $(\hat{S}_j^\alpha, \hat{\mathbf{P}}_j^\alpha)$ satisfies (C.29). According to Claim 16, $(\hat{S}_j^\alpha, \hat{\mathbf{P}}_j^\alpha)$ is also an α-approximate solution to problem (4.2) at node j under space constraints since it satisfies (C.30). By repeatedly applying Claim 17, $(S_{\text{root}}^\alpha, \mathbf{P}_{\text{root}}^\alpha) = \bigcup_{j \in \mathscr{B}} (\hat{S}_j^\alpha, \mathbf{P}_j^\alpha)$ is an α-approximate solution to problem (4.1) with space constraints. Thus assortment $S_{\text{root}}^\alpha = \bigcup_{j \in \mathscr{B}} \hat{S}_j^\alpha$ is an α-approximate assortment. □

C.3 Algorithm JCAOP-S

Algorithm 6: Joint Capacitated Assortment and Price Optimization Under Space Constraints (JCAPO-S)

Input: $\alpha_i, \beta_i, \gamma_i, G_i$ for $i \in V$, $\Im_j$ for $j \in \mathscr{B}$.

Initialization: Set $\mathcal{F}_j^\alpha(o_i^g) = -M$ for $g = 1, 2, ..., G$ and $j \in \text{Children}(i)$;

for $j \in \mathscr{B}$ **do**

  for $g \leftarrow 1, 2, ..., G$ **do**

    get $\tilde{S}_j^\alpha(o_j^g)$ as an α-approximate solution to $\max_{S_j \subseteq \Im_j} \sum_{k \in \text{Children}(j)} V_k(S_k, o_j^g + c_k + 1/\beta_k)/\beta_k$;

    calculate $V_j(\tilde{S}_j^\alpha(o_j^g), o_j^g)$ and $\omega_j(\tilde{S}_j^\alpha(o_j^g), o_j^g)$;

    find g' such that $o_i^{g'} = \lambda(o_j^g)$;

    if $\mathcal{F}_j^\alpha(o_i^{g'}) = -M$ **then**

      $\mathcal{F}_j^\alpha(o_i^{g'}) \leftarrow o_j^g$;

    else

      $\mathcal{F}_j^\alpha(o_i^{g'}) \leftarrow \arg\max_{\theta_j \in \{\mathcal{F}_j^\alpha(o_i^{g'}), o_j^g\}} V_j(\tilde{S}_j^\alpha(\theta_j), \theta_j)\omega_j(\tilde{S}_j^\alpha(\theta_j), \theta_j)/(1-\gamma_j)$;

    end

  end

end

for i in level $m-2, m-3, ..., 1$ **do**

  for $g \leftarrow 1, 2, ..., G$ **do**

    get $\tilde{S}_i^\alpha(o_i^g) = \bigcup_{j \in \text{Children}(i)} \tilde{S}_j^\alpha(\mathcal{F}_j^\alpha(o_i^g))$;

    calculate $V_i(\tilde{S}_i^\alpha(o_i^g), o_i^g)$ and $\omega_i(\tilde{S}_i^\alpha(o_i^g), o_i^g)$;

    find g' such that $o_h^{g'} = \lambda(o_i^g)$;

    if $\mathcal{F}_i^\alpha(o_h^{g'}) = -M$ **then**

      $\mathcal{F}_i^\alpha(o_h^{g'}) \leftarrow o_i^g$;

    else

      $\mathcal{F}_i^\alpha(o_h^{g'}) \leftarrow \arg\max_{\theta_i \in \{\mathcal{F}_i^\alpha(o_h^{g'}), o_i^g\}} V_i(\tilde{S}_i^\alpha(\theta_i), \theta_i)\omega_i(\tilde{S}_i^\alpha(\theta_i), \theta_i)/(1-\gamma_i)$;

    end

  end

end

for $g \leftarrow 1, 2, ..., G$ **do**

  get $\tilde{S}_{\text{root}}^\alpha(o_{\text{root}}^g) = \bigcup_{i \in \text{Children(root)}} \tilde{S}_i^\alpha(\mathcal{F}_i^\alpha(o_{\text{root}}^g))$;

  calculate $R_{\text{root}}(\tilde{S}_{\text{root}}^\alpha(o_{\text{root}}^g), o_{\text{root}}^g)$;

end

Solve for o_{root}^α in $o_{\text{root}} = R_{\text{root}}(\tilde{S}_{\text{root}}^\alpha(o_{\text{root}}), o_{\text{root}})$, then get $S_{\text{root}}^\alpha = \tilde{S}_{\text{root}}^\alpha(o_{\text{root}}^\alpha)$ and $P_{\text{root}}^\alpha = P_{\text{root}}(o_{\text{root}}^\alpha)$;

Output: $S_{\text{root}}^\alpha, P_{\text{root}}^\alpha$.