January 2020

# Customer Choice Modeling For Retail Category Assortment Planning And Product-Line Extension

Elham Nosratmirshekarlou
*Wayne State University*

**CUSTOMER CHOICE MODELING FOR RETAIL CATEGORY ASSORTMENT PLANNING AND PRODUCT-LINE EXTENSION**

by

**ELHAM NOSRATMIRSHEKARLOU**

**DISSERTATION**

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

**DOCTOR OF PHILOSOPHY**

2020

MAJOR: INDUSTRIAL ENGINEERING

Approved by:

Addvisor Date

## ACKNOWLEDGEMENTS

I would like to use this opportunity to thank my advisors Dr. Ratna Babu Chinnam and Dr. Evrim Dalkiran for their guidance, and technical advise in this research. I am also grateful to my committee members, for supporting this thesis. A special thank you to Dr. Pooya Alavian for providing valuable suggestions throughout this study. I also thank my family and friends for their continuous support.

TABLE OF CONTENTS

LIST OF FIGURES

## LIST OF TABLES

# CHAPTER 1

## Introduction

In today's highly competitive global markets, many leading retailers are struggling to keep their profit, and market shares are shrinking for many leading companies. Retail bankruptcies specifically are expected to continue to increase [14]. There are several reasons for these difficulties, including, lack of product innovation, slow recognition of trends, poor product assortment planning and supply chain management, and other factors such as low quality, marketing, and service [11].

One of the biggest problems with retailing today is neglecting to manage and optimize the assortment of products and services provided by the company, and a firm's ability to uncover its customers' preferences towards different products, and to utilize this knowledge to decide which products to keep offering, which products to retire, and which new products to introduce is crucial to its profitability and competitiveness.

From extant literature, one of the most effective approaches for addressing these problems is proper Customer Choice Modeling (CCM). It aims to explain the preferences individual customers have that make them choose from a set of products. In a retail setting, often the choices are discrete, leading to discrete choice models (e.g., preference for product A over B). CCM theoretically or empirically models choices made by customers among a finite set of available alternatives to choose from. In the absence of their preferred product at the time of purchase (e.g., due to stock out), the customer can select the next most preferred product in the available product assortment or 'walk away' without making a purchase. When implemented properly, choice modeling can be used for prediction of demand for products as well as optimization of assortments, and for these reasons, in the past decade, it has been gaining increasing attention [6].

Customer choice modelling under revealed preferences (e.g., through purchase transactions data from stores) becomes 'empirical' choice modeling. The goal of it is to learn choice preferences from store transaction data which record what specific products were bought by individual customers in the presence of the available set of products at the time of purchase. At an abstract level, a choice model can be thought of as a conditional probability distribution that gives us the probability of an arriving customer purchasing a given product in an offer set [20].

Another key aspect of product assortment planning is proper identification of customer segments. Customer segmentation is the practice of dividing a customer base into groups of individuals that are similar in specific ways relevant to marketing, such as age, gender, interests, or spending habits. For example, a demographic segment might include all female customers between the ages of 25 and 35 who are single. Simple examples of psychographic segments could be customers who like to play golf or listen to classical music. Behavioral segments are based on customer shopping habits, and its examples include frequent shoppers [28]. CCM should account for purchasing preferences/differences across different segments of customers (either explicitly or implicitly).

Empirical discrete choice models can be in general categorized into 'parametric' and 'non-parametric' models. Since models of choice are essentially high dimensional objects, the usual approach to dealing with it is a parametric model that captures choice behavior [20]. In general, parametric models statistically relate the choice made to the 'attributes' of the customers and the attributes of the alternatives presented to them. The explicit modeling of choice in terms of attributes allows parametric models the ability to 'interpolate' choice preferences over the attribute space for potentially introducing 'new' products that might improve the assortment. There is vast literature spanning marketing, economics, and psychology that are devoted to the construction of parametric choice models and their estimation from data. Some of the popular models include the multinomial logit (MNL), nested logit (NL), and mixed multinomial logit models (MX). The issue with parametric models is that they can be substantially sub-optimal when making predictions [6]. Also, apart from the fact that one can never be sure that the chosen parametric structure is a 'good' representation of the underlying truth, parametric models are prone to having over-fitting and under-fitting issues [4, 20].

Non-parametric choice models introduced in recent years have been shown to significantly outperform parametric models [20], [5]. They are generic models of consumer choice that obtain distributions over list of product preferences, and the best choice model is automatically selected based on the data [20]. The identified preference lists can be seen to represent customer segments with similar preferences as well [7]. While non-parametric models offer better accuracy, the models do not have a provision to explicitly incorporate product attributes, which is a limitation for evaluating the attractiveness of new/non-existing products that might rely on some new combination of the attributes within current assortments. In summary, parametric models can explicitly handle product attributes but lack accuracy (i.e., can suffer from over-fitting and under-fitting) whereas non-parametric models yield better accuracy but cannot explicitly account for product attributes. To overcome the limitations of both techniques while leveraging their strengths, we propose a 'hybrid' approach for discrete choice modeling and assortment optimization. Our hybrid approach relies on non-parametric models for establishing 'customer segments' and parametric models within each segment for modeling choice. Using real data, we show that a few number of rankings ('segments') explain the behavior of most customers; similar results are also reported by [6].

So far all our discussion was around the common theme among the majority of CCM methods, meaning the single-category nature of the developed models, i.e., these approaches model the customers in an isolated environment in which only one category of products (the "target" category) is offered, and study the brand choice behavior within individual product categories [37]. A category consists of a group of products (SKUs) with the same functionality which are competing with each other [37]. Major manufacturers often own multiple brands in a vast number of product categories. The downstream customers of these manufacturers are large supermarket chains and department stores, who are multi-category firms as well. CCM is capable to analyze the market in multiple categories of products.

The main problem with the single-category models is that they fail to work under the following main two conditions:

I) if a customer does not buy any of the existing products and walks away,

II) if a new product with a new attribute is introduced to the market and there is not historical data available.

While in general, concluding any information from the above-mentioned conditions seem to be extremely difficult, there are cases where some level of inference is in fact possible. For instance, if a dairy manufacturer is considering the introduction of a new attribute (fat-free) for ice cream for the first time, it may be able to infer some insight based on the demand for fat-free milk or yogurt, for which historical data exists.

As mentioned before, studies have shown a frequent empirical finding is that customers have different choices and behaviors from one another in each single category and there are a few main groups of behaviors to be extracted that can cover a good number of customers [32]. The goal here is to find if customers exhibit similarities in their choice behavior across multiple categories and if these information can be used to make better predictions about existing and new products attractiveness (i.e., a quantitative measure of products' probability of being chosen). As an example, consider an apparel store that sells both shirts and pants. Shirts are classified by their color, size, and material, whereas pants are classified by color and size but not by material (e.g., all pants are currently produced using one material). The question is how the store can estimate customer inclination toward new pants with a new material from customers behavior toward material variety of shirts.

While this question has attracted interest for a long time, appropriate approaches to tackle the problem have not been developed until recent years [3]. The findings in these studies have established the bases in the development of multi-category customer choice models, i.e., models in which customers' preferences have a joint distribution that allows *correlatedness* across categories.

In other words, multi-category models study multiple categories of products together and pro-

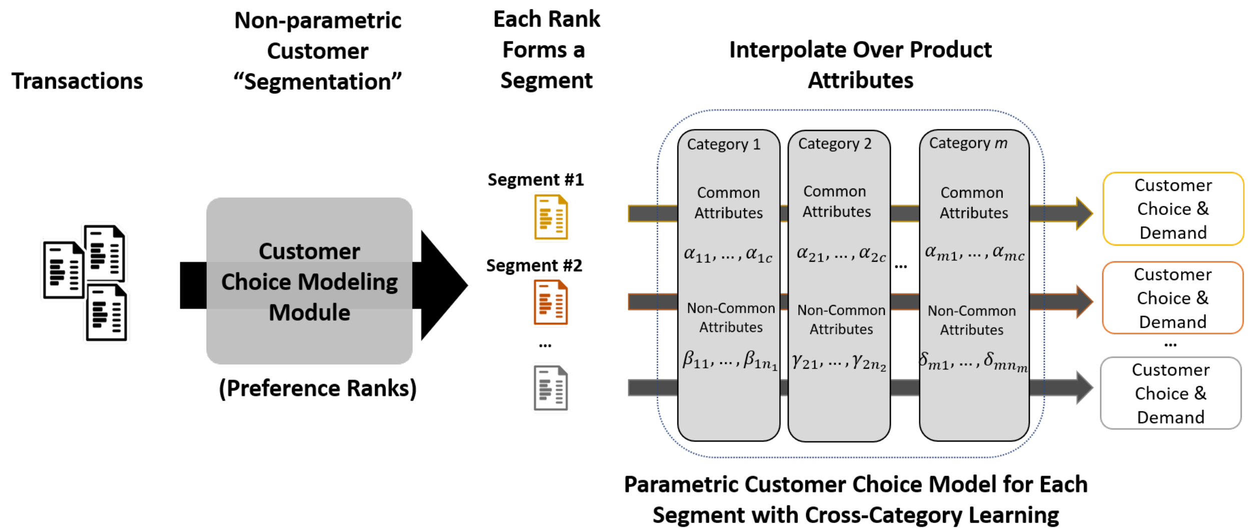



Figure 1.1: Illustration of major steps of the proposed method for characterizing customer choice

vide a joint probability model for customer behavior in each category. Even though these methods are not as mature and accurate as single-category models, they try to address the issues with conditions I and II above, at least for the cases that required data about the customer or products are available.

Our focus in this dissertation is on the use of observed household purchase data to analyze preferences in multiple categories. To benefit from the advantages of both methods, we present hybrid approaches of both single and multi category methods. Our proposed methodology uses a single-category CCM at its core prediction model. This is to utilize the simplicity and higher accuracy of single-category models. Next, to utilize the data from customers' behavior across other categories, we apply a multi-category model to our problem. This model studies a category along with another category that can provide us with additional information about it. The result of the multi-category analysis is leveraged to provide a better estimate of customer choices for the target category. The overall approach in this study is shown in Figure 1.1.

The rest of the dissertation is organized as follows: Chapter 2 describes the proposed single category customer choice modeling for new product development, Chapter 3 provides solution to the same problem using cross category customer choice modeling, and Chapter 4 offers some concluding remarks and directions for future research.

# CHAPTER 2

## Single Category Customer Choice Modeling for Improved Product Development

### 2.1 Introduction

In today's global and highly competitive markets, many leading retailers are struggling and retail bankruptcies are expected to continue unabated [14]. Market shares are also shrinking for leading companies. For example, the top 25 food and beverage companies in the U.S. lost $18 billion in market share since 2009 and 90% of the top 100 consumer packaged goods brands lost market share in 2017 [13]. In the apparel industry, J. Crew suffered $1 billion in write-downs, Aeropostale's market value is halved, and Gap, the quintessential American brand, announced it would close about a quarter of its North American stores, hobbled by years of market share losses [39]. L Brands announced difficulties with same-store sales at its Victoria's Secret stores and plans to close 53 stores in 2019 [25]. There are a multitude of reasons for these difficulties, including, lack of product innovation, slow recognition of market trends, poor store product assortment planning, poor supply chain management, and other factors such as poor quality, marketing, and service [11].

One of the key problems with retailing today is neglecting to effectively manage and optimize the assortment of goods and services provided by the firm. The following examples are illustrative [11]: 1) A small Walmart store in Mexico City that operates under the name Bodega Aurrera Express stocks 50 different SKUs (stock keeping units) of toilet paper, which is an inefficient use of resources; 2) A highly sophisticated supermarket in Hamburg, which sells some 30 varieties of yogurt, neglects to stock the 10% cream variety that many of its most discerning customers prefer. A firm's ability to uncover its customers' preferences towards different products, and to utilize this knowledge to decide which products to keep offering, which products to retire, and which new products to introduce is crucial to its profitability and competitiveness. In a recent study, researchers from Nielsen, a British information, data and measurement firm, helped several grocery stores make changes across condiment categories by focusing on their expandable features [36]. Their analysis concluded that the sheer number of products on the shelf is cluttering the consumer experience and many of the new products are chasing the same benefits. They found that stores that preformed optimization of assortments strongly outperformed competitors that changed their selection minimally or did not participate in the study.

From extant literature, one of the most effective approaches for addressing these problems is proper Customer Choice Modeling (CCM). It aims to explain the choices individual customers make in choosing from a set of products based on their stated (e.g., through surveys) or revealed preferences. In a retail setting, often the choices are discrete, leading to discrete choice modeling (e.g., preference for product A over B). It theoretically or empirically models choices made by customers among a finite set of available alternatives. In the absence of their preferred product at the time of purchase (e.g., due to stock out), the customer can select the next most preferred product in the available product assortment or 'walk away' without purchase. When carried out properly, choice modeling can facilitate the prediction of demand for products as well as optimization of assortments, and is gaining increasing attention in the past decade [6].

Customer discrete choice modelling under revealed preferences (e.g., through purchase transaction data and inventory records from stores) becomes 'empirical' discrete choice modeling. The goal is to learn choice preferences from store transaction data which record what specific products were bought by individual customers in the presence of the available set of products at the time of purchase. At an abstract level, a choice model can be thought of as a conditional probability distribution that for any offer set yields the probability that an arriving customer purchases a given product in that set [20]. Another key aspect of product assortment planning is proper identification of customer segments. Customer segmentation is the practice of dividing a customer base into groups of individuals that are similar in specific ways relevant to marketing, such as age, gender, interests and spending habits. For example, a demographic segment might include all female customers between the ages of 20 and 30 who are married. Simple examples of psycho-graphic segments could be customers who like golf or classical music. Behavioral segments are based on customer shopping habits, one example being frequent purchasers [28]. Overall, CCM should account for purchasing preferences/differences across segments (either explicitly or implicitly).

Empirical discrete choice models can be broadly categorized into ‘parametric’ and ‘non-parametric’ models. Since models of choice are inherently high dimensional objects, the typical approach to dealing with this problem is positing, a-priori, a parametric model that one believes adequately captures choice behavior [20]. In general, parametric models statistically relate the choice made to the ‘attributes’ of the person and the attributes of the alternatives available to the person. The explicit modeling of choice in terms of attributes allows parametric models the ability to ‘interpolate’ choice preferences in the attribute space for potential introduction of ‘new’ products to improve the assortment. There is vast literature spanning marketing, economics, and psychology devoted to the construction of parametric choice models and their estimation from data. Some of the popular models include the multinomial logit (MNL), nested logit (NL), and mixed multinomial logit models (MX). For a good review of the topic, see [38]. However, parametric approaches can be substantially sub-optimal in scenarios where one cares about using the choice model learned to make fine-grained predictions [6]. Apart from the fact that one can never be sure that the chosen parametric structure is a ‘good’ representation of the underlying ground truth, parametric models are prone to over-fitting and under-fitting issues [4, 20].

Non-parametric choice models introduced in recent years have been shown to significantly outperform classic parametric models (i.e., MNL, NL, and MX) in a variety of settings [20, 5]. They are ‘generic’ models of consumer choice, namely, distributions over product ‘preference lists’, and the data automatically selects the ‘right’ choice model [20]. The identified preference lists can be seen to identify customer segments as well [7]. In addition, these models also subsume essentially all extant choice models. While non-parametric models offer better accuracy, the models proposed to date do not have a provision to explicitly incorporate product attributes of the assortment, which is a limitation for evaluating the attractiveness of potentially new (currently non-existing) products that might rely on some new combination of the attributes within current assortments.

In summary, parametric models can explicitly handle product attributes but lack accuracy (i.e., can suffer from over-fitting and under-fitting) whereas non-parametric models yield better accuracy but cannot explicitly account for product attributes. To overcome the limitations of both techniques while leveraging their strengths, we propose a 'hybrid' approach for discrete choice modeling and category product assortment optimization. Our hybrid approach relies on non-parametric models for establishing 'customer segments' and parametric models within each segment for modeling choice at the attribute level.

We validate the proposed methods using data from a leading apparel retailer. In particular, we rely on primary sales transaction data available through point-of-sale (POS) systems along with product inventory/stock-out data, and product specifications (in terms of attributes) for our modeling. First, we construct a non-parametric customer choice model to learn the dominant preference lists (i.e., lists of product rankings) to form the dominant 'customer segments'. The resulting model is then employed to categorize/assign the observed sales transactions to the most likely preference list/segment. The transactions associated with each segment are in turn leveraged for learning an 'attribute-based' parametric customer choice model to be able to quantify the relative importance of each attribute to customers of each segment and also enable analytical evaluation of potentially new products and their performance comparison with existing products. Using data from the apparel retailer, we show that a handful of product rankings (segments) can adequately explain the behavior of most customers for individual product categories (e.g., yoga pants category); similar results are also reported by [6].

Thus, two main problems must be solved to develop the proposed hybrid approach for CCM for improved assortment planning and product-line extension. The first problem (**P1**) is to develop a robust approach to customer segmentation based on sales transaction and inventory data, such that customers have same order (rank) of preferences towards the offered products within each segment. The order of preferences becomes important, when seller's inventory changes. In particular, the set of products observed by the potential customer changes under stock-outs of select products. In fact, two customers belonging to two different segments may buy the same product when the seller does not consistently offer complete assortments at the store. In consumer theory, substitute products are products that a consumer perceives as similar or comparable, so that having more of one product makes them desire less of the other product. This product substitution can take the form of 'static' assortment-based substitution (e.g., product is not available in the preferred color by the retailer) or 'dynamic' stock-out based substitution (product is temporarily unavailable due to stock-outs). The algorithm to be developed for this problem must be able to identify the customers who are not buying their 'first preference' but their second, third, etc. choice due to incomplete store assortments and not to mix them with those who are purchasing their first preference. This adds to complexity of customer segmentation when relying on POS data. The algorithm must also provide a measure of the size of each segment (e.g., the fraction of customers belonging to each segment). From a theoretical perspective, **P1** can be modeled as a clustering/segmentation problem.

The second problem (**P2**) is to develop a robust approach to translate customers' preferences toward 'products' to preferences towards 'attributes'. We try to find preferences in customer choice in each segment in terms of a set of defined attributes (e.g., color, size, material, fashion, etc.). The objective, is to develop a procedure to assign importance to each attribute. By solving **P2**, a firm can answer the question of which products to keep offering and which products to discontinue. In addition to its help in understanding the dynamics of the market towards current products, solving **P2** enables decision-makers to identify the important attributes and their best values for each customer segment so that the firm can introduce new products with maximum likelihood of sales (preferably attracting additional customers rather than cannibalizing sales for other products within the assortment). From a theoretical perspective, **P2** can be modeled as a regression problem.

Once the customer segments are established from **P1**, the firm can now develop policies for introducing new products, dismissal of the current ones, and marketing and advertising (e.g., to show appropriate ads to each customer segment). For multi-regional firms with distributed branches, this research can be used for supply chain optimization by finding the optimal distribution of products based on regional customer segments. In summary, this research will focus on the following: 1) Identifying customer segments based on customers' choices using POS and store inventory data. 2) Estimating relative attractiveness of existing products for each segment based on products' attributes. 3) Predicting attractiveness of potential new products based on observed attributes in competing products. 4) Demonstrating the effectiveness of the proposed approach using data from an apparel retailer.

The rest of this chapter is organized as follows: Section 2.2 reviews the literature related to each of the problems introduced. Section 2.3 explains the proposed methodology. Section 2.4 presents results based on real-world data from an apparel retailer. Finally, Section 2.5 concludes and suggests directions for future research.

## 2.2 Literature Review

Customer choice models and their applications have been studied extensively in the literature across multiple fields such as allocation decisions [27], price optimization [24], online purchase behaviour [40], and market segmentation [21]. The logit and the multinomial logit (MNL) models are the most extensively studied customer choice models in the literature. Originally, logit models were introduced for binary choice models by [29] and generalization of this model to more than two alternatives led to the MNL model, which is popularized by [31]. The primary reason for the popularity of logit and MNL models is their ease of implementation. Researchers are often interested in understanding how price, promotions, and other marketing mix variables impact a firm's market share. Being stochastic and allowing admission of decision variables was the initial appeal for the MNL model ([15]; [34]). [23] used the MNL model to understand the effects of various marketing variables on consumer choice among product alternatives. They demonstrated the statistical significance of brand loyalty, size loyalty, store promotions, shelf price, and price cuts on market share and other explanatory variables. [22] compared MNL model to other regression models, and showed that MNL is superior for understanding household preferences as well as cross-sectional multi-attribute choice modeling. Much research has been done addressing the basic and applied assumptions of the MNL models. [26] proposed other logit models such as nested and mixed multinomial logit models. For a good review of MNL models and their variants, see [38].

Although the MNL models allow us to parameterize utility as a function of product attributes and are widely employed in practice, as noted earlier, they are limited in their accuracy and also lack the ability to establish customer segments. As an alternative to parametric models, several generic choice models have been proposed that make minimal structural assumptions and are capable of representing a wide variety of choice models. Earliest studies on 'non-parametric' choice models appeared in the economics and psychology literature [10]. They were introduced into operations literature by [30] who also showed that non-parametric models capture a variety of parametric models as special cases. Several studies focus on the use of non-parametric models in CCM. In particular, [20] proposed a general model of choice, where one represents choice behavior by a probability distribution over all of the possible rankings of the products. They showed, using both synthetic and real data, that their revenue predictions in the automotive sector are more accurate than those produced by parametric models such as MNL and Mixed MNL. Another general model that has recently been proposed is the Markov chain model of customer choice [8]. In this approach, products are modeled as states in a Markov chain and substitution behavior is modeled by transitions. They showed that such a model provides a good approximation to any choice model based on random utility maximization. The shortcoming of this method is being computationally intractable since it considers all possible rankings as states.

A recent computationally efficient method for non-parametric CCM is proposed by [6]. The computational advantage of this method stems from the fact that it does not consider all possible rankings of products, instead, it efficiently identifies a small subset of them over the products, and assigns a probability distribution over this set using POS data.

Overall, without loss of generality, we shall rely on MNL models for modeling choice within individual customer segments and rely upon the non-parametric method by [6] for identifying the dominant product rank lists (segments) in developing our hybrid approach for product assortment planning and product-line extension.

## 2.3 Methodology

In this study, we aim to characterize customer preferences utilizing attributes of products within a category and identify new product(s) in the existing attribute space that might increase demand for the updated product assortment. Toward this goal, we use a combination of parametric and non-parametric CCM. Our proposed approach benefits from both methods while avoiding the shortcomings of either method. Specifically, we use non-parametric classification approach to distribute heterogeneous POS transactions and inventory data among various customer segments, each with a specific product ranking. Hence, a higher homogeneity is achieved within each customer segment. Then, we develop a distinct parametric choice model for each customer segment to quantify the attractiveness of product attributes for the segment. For illustrative purposes, without loss of generality, we chose the non-parametric CCM approach proposed by [6] and the MNL regression approach for characterizing customer choice in the attribute space for each customer segment.

The proposed framework relies on a set of attributes representing different aspects of the products within the category under study for characterizing customer choice. These attributes are selected such that their combination uniquely identifies products, and each attribute is important in customers' choices. A product attribute is a feature of products that is recognizable by observing the product, collectively exhaustive (i.e., each product is uniquely identifiable by its attributes), and its level is precisely determined for each product [19]. The questions of what attributes to focus upon and how to choose them are highly subject-dependent. In the case of personal computers, plausible attributes are screen size, weight, processor, memory, ports, storage, etc., while in athletic apparel product attributes can be design/style, size, color, fabric, etc. Evidently, the choice of attributes depends on the products in the assortment as well. If say, Apple wants to analyze

Table 2.1: Notation

| | |
|---|---|
| $\mathcal{N} = \{0, 1, \ldots, N\}$ | set of products |
| $\mathcal{D} = \{1, \ldots, D\}$ | set of attributes |
| $\mathcal{K} = \{1, \ldots, K\}$ | set of customer segments |
| $\mathcal{M} = \{1, \ldots, M\}$ | set of assortments available to customers during the data collection period |
| $\sigma^k$ | ranking (permutation) of products in $\mathcal{N}$ for customers in segment $k, k \in \mathcal{K}$ |
| $\lambda^k$ | fraction of customers belonging to segment $k, k \in \mathcal{K}$ |
| $\mathcal{S}_m$ | set of options available in assortment $m, m \in \mathcal{M}$ |
| $\mathcal{T}_m$ | set of transactions during the period when assortment $m$ is present, $m \in \mathcal{M}$ |
| $\mathcal{T}^k$ | set of transactions assigned to customer segment $k, k \in \mathcal{K}$ |
| $v_{mn}$ | fraction of customers purchasing $n^{th}$ option when assortment $\mathcal{S}_m$ is presented |
| $X_n$ | vector of size $D + 2$ representing values of attributes for option $n, n \in \mathcal{N}$ |

its MacBook laptop line, it does not need to take the brand and number of connection ports into account since they are not differentiating attributes; Best Buy, on the other hand, might have to consider these two attributes.

The notation employed is borrowed from [6]. $N$ denotes the number of products, indexed from 1 to $N$. Index 0 is used to denote the 'outside' or 'no-purchase' alternative (the possibility that the customer does not purchase any of the offered products by the firm). We refer to the set $\mathcal{N} = \{0, 1, \ldots, N\}$, the set of products together including the no-purchase alternative, as available options. A customer always chooses exactly one of the available options. The no-purchase option 0 is always available, while the other options may be out-of-stock. $\sigma^1, \cdots, \sigma^K$ denote the rankings (or permutations) of options in $\mathcal{N}$, where each ranking $\sigma^k$ orders the options from most preferred to least preferred. A customer having ranking $\sigma^k$ prefers option $i$ to $j$ if and only if $\sigma^k(i) < \sigma^k(j)$, where $\sigma^k(i)$ is the position of option $i$ in ranking $\sigma^k$. There are $(n+1)!$ different possible rankings of options. The fraction of customers who have ranking $\sigma^k$ is denoted as $\lambda^k$ (i.e., market share), where $\sum_{k=1}^{K} \lambda^k = 1$. Let $M$ denote the number of different available assortments to customers throughout the data collection period, where $\mathcal{S}_m$ is the set of options offered in assortment $m$. Let $\mathcal{T}_m$ be the set of transactions during the periods when assortment $m$ is offered. Similarly, let $\mathcal{T}^k$ be the set of transactions assigned to customer segment $k$. There exists a vector $v_m$ of the same dimension, corresponding to each $\mathcal{S}_m$, whose $n^{th}$ element denotes the fraction of customers selecting $n^{th}$ option in $\mathcal{S}_m$, where $\sum_{n\in\mathcal{S}_m} v_{mn} = 1$. The attributes form a $D$-dimensional attribute space, $\mathcal{R}^D$, where $D$ is the number of attributes. Option $n$ is represented by the vector $X_n$ in $\mathcal{R}^{D+2}$. The first element is 1, representing the bias term for the MNL model and the second element is 1 for the $N$ products and 0 for no-purchase option. The rest of the elements in $X_n$ are the values of attributes for option $n$. Table 2.1 summarizes the notation.

The steps of the proposed approach are as follows:

1. Perform a non-parametric CCM utilizing transaction and inventory data to obtain the dominant product rankings and associated customer segments and fraction of customers in each segment, i.e., $\sigma^k$ and $\lambda^k, k \in \mathcal{K}$.
2. Based on likelihood, allocate all historical transactions to most likely customer segments obtained in Step 1, i.e., $\mathcal{T}^k, k \in \mathcal{K}$.
3. Fit an MNL regression model to each customer segment from Step 1 using the assigned transaction data from Step 2 to find model coefficients, i.e., $\beta^k = [\beta^k_b, \beta^k_p, \beta^k_1, \cdots, \beta^k_D], k \in \mathcal{K}$.
4. Compute the attractiveness of potentially new line-extension products using their attribute vectors and MNL model coefficients from Step 3, i.e., $X_{new}\beta^k$, and find their ranks for each customer segment.

We use non-parametric choice algorithm for segmenting customers in the first step of the proposed method. Although there are several proposed methods in this area as discussed in Section 2.2, without loss of generality, we utilize the method proposed by [6] for implementation. To set the stage, let matrix $A$ be vertically concatenated matrices of $A_m$, where $n^{th}$ row and $k^{th}$ column of matrix $A_m$, $(A_m)_{nk}$, is 1 if a customer belonging to ranking $\sigma^k$ would choose $n^{th}$ option when offered assortment $m$ and is 0 otherwise. Matrix $A$ and vector $v$ are related to each other; they satisfy $A\lambda = v$, where each column of $A$ and each element of $\lambda$ correspond to a ranking and each row of $A$ and $v$ correspond to an option in an assortment. Since there are $(n+1)!$ columns

**Algorithm 1** Algorithm for finding the customer segments [6]

**Require:** Assortments $\mathcal{S}_1, \ldots, \mathcal{S}_M$, and $v$
**Ensure:** $\sigma^1, \ldots, \sigma^K$, and $\lambda^1, \ldots, \lambda^K$

```
Initialize K to 0
Set A to an empty matrix
α, μ, λ ← π_1(A, v)
z, a ← π_2(S_1, ..., S_M, α, μ)
while (−α^T a − μ < 0) do
    K ← K + 1
    σ^K(i) = Σ_{j=0, j≠i}^{n} z_ij
    A ← [A, a]
    α, μ, λ ← π_1(A, v)
    z, a ← π_2(S_1, ..., S_M, α, μ)
end while
```

in $A$, the problem becomes computationally intractable. A practical approach is proposed by [6] based on column generation heuristic, as presented in Algorithm 1. Given vector $v$ and assortments $\mathcal{S}_1, \ldots, \mathcal{S}_M$, Algorithm 1 provides the customer segments (rankings) and fraction of customers in each segment, $\sigma^k$ and $\lambda^k$, respectively.

In the above algorithm, mathematical programs $\pi_1(A, v)$ and $\pi_2(\mathcal{S}_1, \ldots, \mathcal{S}_M, \alpha, \mu)$ are defined as follows:

$$\pi_1(A, v): \qquad \min_{\lambda, \epsilon^+, \epsilon^-} \quad \mathbf{1}^T \epsilon^+ + \mathbf{1}^T \epsilon^- \tag{2.1a}$$

$$\text{s.t.} \quad A\lambda + \epsilon^+ - \epsilon^- = v \tag{2.1b}$$

$$\mathbf{1}^T \lambda = 1, \tag{2.1c}$$

$$\lambda, \epsilon^+, \epsilon^- \geq 0. \tag{2.1d}$$

$\pi_1(A, v)$ minimizes the $l_1$ error between $A\lambda$ and $v$, and returns optimal $\lambda$, and optimal dual variables $\alpha$ and $\mu$ corresponding to constraints in (2.1b) and (2.1c), respectively. In our numerical experiments, Algorithm 1 terminated with a nonzero objective function value for Problem $\pi_1$. The variables representing the error, $\epsilon^+$ and $\epsilon^-$, are treated equally in the objective function in (2.1a) for every product in every assortment presented to customers. Instead, different weights may be given to different assortments depending on their importance. One such weight can be defined as the number of transactions that occurred while an assortment is present. Accordingly, the vector $\mathbf{1}$ in the objective function can be replaced by $[|\mathcal{T}_1|, \ldots, |\mathcal{T}_1|, |\mathcal{T}_2|, \ldots, |\mathcal{T}_2|, \ldots, |\mathcal{T}_M|, \ldots, |\mathcal{T}_M|]$, where $|\mathcal{T}_m|$ is replicated $|\mathcal{S}_m|$ times for each assortment, $m \in \mathcal{M}$.

The integer programming problem $\pi_2(S_1, \ldots, S_M, \alpha, \mu)$ determines the new ranking, $\sigma$, and the associated column, $a$, to be added to matrix $A$ in Problem $\pi_1$ at the next iteration. The binary variable $a_{im}$ is 1 if $i^{th}$ option is preferred over the options in $\mathcal{S}_m$ by customers with ranking $\sigma$. Similarly, the binary variable $z_{ij}$ is 1, if $i^{th}$ option is preferred over $j^{th}$ option by customers in the new segment. The objective function in (2.2a) identifies a binary column $a$ with the smallest reduced cost coefficient for Problem $\pi_1$. The constraints in (2.2b) define the relation between $a$- and $z$-variables, whereas constraints in (2.2c) and (2.2d) enforce strong preference between any two options and transitivity, respectively.

$$
\begin{aligned}
\pi_2(S_1, \ldots, S_M, \alpha, \mu) : \qquad & \min_{z,a} \quad -\alpha^T a - \mu && \text{(2.2a)}\\
& \text{s.t.} \quad a_{im} \leq z_{ij}, \;\; \forall m \in \mathcal{M}, i, j \in \mathcal{S}_m, i \neq j, && \text{(2.2b)}\\
& \qquad\quad z_{ij} + z_{ji} = 1, \;\; \forall i, j \in \mathcal{N}, i \neq j, && \text{(2.2c)}\\
& \qquad\quad z_{ij} + z_{jk} - z_{ik} \leq 1, \;\; \forall i, j, k \in \mathcal{N}, i \neq j \neq k, && \text{(2.2d)}\\
& \qquad\quad z_{ij} \in \{0, 1\}, \;\; \forall i, j \in \mathcal{N}, i \neq j, && \text{(2.2e)}\\
& \qquad\quad a_{im} \in \{0, 1\}, \;\; \forall m \in \mathcal{M}, i \in \mathcal{S}_m. && \text{(2.2f)}
\end{aligned}
$$

After segmenting customers to insure homogeneity within each segment, transactions are assigned to the customer segments based on their likelihood. Each transaction is represented using a tuple $t_i = (\mathcal{S}_{m(i)}, P_{(i)})$, where $P_{(i)}$ is the product bought (regardless of the quantity purchased) and $\mathcal{S}_{m(i)}$ is the assortment presented to the customer at the time of the transaction $t_i$. For each transaction $t_i$, we identify the set of customer segments, $\mathcal{K}_i \in \{1, \ldots, K\}$, in which customers would purchase $P_{(i)}$ when assortment $\mathcal{S}_{m(i)}$ is offered, i.e, $\mathcal{K}_i = \{k \in \{1, \ldots, K\} : P_{(i)} = \text{argmin}_{j \in \mathcal{S}_{m(i)}} \sigma^k(j)\}$. Then, the transaction $t_i$ is randomly assigned to customer segment $k$ with probability:

$$\frac{\lambda^k}{\sum_{l \in \mathcal{K}_i} \lambda^l}, \forall k \in \mathcal{K}_i. \tag{2.3}$$

Next, we estimate the MNL model parameters $\beta^k = [\beta_b^k, \beta_p^k, \beta_1^k, \cdots, \beta_D^k]$ using the transaction data assigned to segment $k, \mathcal{T}^k$. The parameters $\beta_b^k$ and $\beta_p^k$ correspond to the bias term and purchasing a product, respectively. The other parameters, $\beta_d^k, d \in \mathcal{D}$, denote the 'importance' of each attribute for customer segment $k$. Next, the transaction data in terms of options and assortments is restructured to represent the transactions in attribute space. For each transaction $t_i = (\mathcal{S}_{m(i)}, P_{(i)}), t_i \in \mathcal{T}^k, |\mathcal{S}_{m(i)}|$ number of input vectors are created, one for each option in assortment $m(i)$. The input vectors are in the form of $[Y_j^i, X_j]$, where $Y_j^i$ is 1 for the purchased option $P_{(i)}$ and is 0 for the other options in the assortment $m(i)$. Recall that $X_j$ represent the option $j$ in attribute space.

The predicted probability of choosing option $n$ by customers of segment $k$ when the complete product set is available is formulated as $Pr^k(P_n|\mathcal{N}) = \frac{e^{\beta^k X_n}}{\sum_{i \in \mathcal{N}} e^{\beta^k X_i}}$. Similarly, if the offered set of products is $S_m$, the the predicted purchase probability of option $n$ by customers of segment $k$ is defined as $Pr_m^k(P_n|\mathcal{S}_m) = \frac{e^{\beta^k X_n}}{\sum_{i \in \mathcal{S}_m} e^{\beta^k X_i}}$. New options (products) with potentially large estimated utilities for customer segment $k$ can be identified based on the signs of $\beta_d^k, d \in \mathcal{D}$. For example, if $\beta_d^k$ is negative, the level for attribute $d$ is selected as zero, and if $\beta_d^k$ is positive, the level for attribute $d$ is selected as one. This vector $X_{new}$ represents the levels of attributes for the option estimated to be most attractive to this segment. If this product option is not part of the current assortment, it should be considered for addition. The market share of the new option when the complete product set is offered to customer segment $k$ is $Pr^k(P_{new}|\mathcal{N}) = \frac{e^{\beta^k X_{new}}}{e^{\beta^k X_{new}} + \sum_{i \in \mathcal{N}} e^{\beta^k X_i}}$. It is of course also possible to explore and estimate utilities for other combinations of product attributes to identify other options for product-line extension.

### 2.4 Case Study

In this section, we illustrate the proposed method for CCM and product-line extension through a case study and demonstrate its effectiveness. The analysis is based on the data provided by a leading global apparel brand. Data includes a list of currently sold products and their attributes as well as sales and product inventory data for individual stores for a 6-month period. For this case study, we focus on four stores from a specific geographic region, known for very similar customer demographics and purchase habits. Among hundreds of product categories, we selected 'yoga pants' category and size 'small'. After pooling data from the four stores, we arrive at seven distinct SKUs (stock keeping units) with sufficient number of purchases for each product and nine different assortments experienced by customers due to temporary product stock-outs at the products during the study period.

At the pre-processing step, POS and inventory data is analyzed to identify the assortments offered during the study period, $\mathcal{S}_1, \ldots, \mathcal{S}_M$, and the proportion of sales for each product option, $v_1, \ldots, v_M$. We utilize the inventory data to determine whether a product was present or not for each store and for each day during the 6-month study period. We assume that if a product is listed in the inventory table for a particular day, it is available to the customers throughout the day, and if the inventory of a product is zero, it is not available for customers to purchase. While both assumptions may be occasionally violated in the real-world due to inevitable data entry errors and transactions throughout the day, our proposed approach manages to provide robust results as demonstrated via sensitivity analysis in Section 2.4.2. The different assortments available to the customers at distinct points during the study period are identified in Table 2.2.

Table 2.2: Sample output of yoga pant product assortments offered at a particular store on two particular days.

| Date | Yoga pant products present |
|---|---|
| 4/4/2016 | $P_1, P_2, P_3, P_5$ |
| 4/5/2016 | $P_1, P_2, P_5, P_6, P_7$ |
| ... | ... |

Next, we aggregate the assortment data in Table 2.2 to find the number of days each assortment is presented and the list of assortments offered during the study period. The end result of this step

Table 2.3: Number of retail days during which each distinct assortment is offered during the study period.

| Assortment no. | Assortment | Days present |
|---|---|---|
| 1 | $P_1, P_2, P_3, P_5$ | 10 |
| 2 | $P_1, P_2, P_3, P_4, P_7$ | 49 |
| 3 | $P_1, P_2, P_3, P_5, P_6, P_7$ | 28 |
| 4 | $P_1, P_2, P_3$ | 14 |
| 5 | $P_1, P_2, P_3, P_4, P_5, P_6, P_7$ | 38 |
| 6 | $P_1, P_2, P_4$ | 9 |
| 7 | $P_1, P_2$ | 11 |
| 8 | $P_1, P_2, P_5, P_6, P_7$ | 20 |
| 9 | $P_1, P_2, P_5, P_7$ | 17 |

Table 2.4: Total sales for each product grouped by assortment for the study period

| Sales data | | **Assortment** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* |
| SKU | $P_1$ | 2 | 9 | 5 | 1 | 4 | 4 | 2 | 2 | 1 |
| | $P_2$ | 3 | 18 | 16 | 1 | 27 | 1 | 1 | 1 | 7 |
| | $P_3$ | 7 | 58 | 3 | 6 | 1 | 0 | 0 | 0 | 0 |
| | $P_4$ | 0 | 39 | 0 | 0 | 1 | 1 | 0 | 0 | 0 |
| | $P_5$ | 4 | 0 | 16 | 0 | 19 | 0 | 0 | 9 | 1 |
| | $P_6$ | 0 | 0 | 20 | 0 | 29 | 0 | 0 | 12 | 0 |
| | $P_7$ | 0 | 0 | 7 | 0 | 24 | 0 | 0 | 6 | 2 |
| No purchase, $P_0$ | | 1 | 12 | 6 | 0 | 10 | 0 | 0 | 3 | 1 |

is shown in Table 2.3. $P_1$ and $P_2$ were never stocked out during the study period, whereas the complete assortment, $P_1, \ldots, P_7$ was only offered for 38 days. During the study period, customers encountered in total nine different assortments, $K = 9$.

For each observed assortment, we total the sales for each of the seven products using POS data and the dates that the assortment is offered to customers, as shown in Table 2.4. The stores did not record lost sales from 'serious' customers (i.e., non 'window shoppers' that had the intent to buy a product if presented a good choice) due to product stock-outs or due to dissatisfaction with offered product assortment. While it is not trivial to collect such information accurately, it is possible to make reasonable inferences about these estimates as a function of average customer traffic into the

Table 2.5: Defined attributes for yoga pant category and their meanings

| Attribute name | Value 0 | Value 1 |
|---|---|---|
| Bias term: Always ‘1’ | – | – |
| Identifier for no-purchase | No-purchase | Physical product |
| Material | No mesh | With mesh |
| Length | Full length | Capri |
| Style | Campus | Flat |
| Fashion | Fashionable | Basic |
| Color | Black | Colored |

store and average transactions under full assortment [20]. For each assortment, the number of no-purchase events is estimated as 10% of total sales. The total sales data in Table 2.4 is normalized by dividing each total sales value by the sum of total sales in the corresponding assortment to find $v_m, \forall m \in \mathcal{M}$, which is used in the non-parametric choice algorithm for customer segmentation in the first step of our proposed approach.

The required inputs for the second step are the attribute-based representation of options (products and no-purchase option) and transactions assigned to each customer segment. We utilize the product database to find distinguishing features among the products. Among the seven attributes we identified, five attributes correspond to product features (i.e., material, length, style, fashion, and color), one attribute is used to distinguish no-purchase option from other products, and one attribute represents the bias term. Table 2.5 displays the selected attributes and the characteristics for the two attribute levels, whereas Table 2.6 shows the attributes levels for each option, $\mathcal{N} = \{0, 1, \ldots, 7\}$.

### 2.4.1 Results

In this section, we demonstrate the effectiveness of the proposed approach using data from a leading global apparel brand, as discussed in the previous section. We commence with non-parametric choice model to segment customers and allocate transactions to customer segments. Next, we continue with parametric choice model, namely MNL, to estimate the importance of

Table 2.6: Product attributes for yoga pants

| SKU | Bias term | Identifier | Material | Length | Style | Fashion | Color |
|---|---|---|---|---|---|---|---|
| $P_0$ | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| $P_1$ | 1 | 1 | 0 | 1 | 0 | 1 | 1 |
| $P_2$ | 1 | 1 | 0 | 1 | 1 | 1 | 1 |
| $P_3$ | 1 | 1 | 0 | 0 | 1 | 1 | 1 |
| $P_4$ | 1 | 1 | 0 | 1 | 0 | 0 | 0 |
| $P_5$ | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| $P_6$ | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| $P_7$ | 1 | 1 | 0 | 1 | 0 | 1 | 0 |

Table 2.7: Identified top customer segments (product order rankings) and estimated market shares

| # | Segment | $\lambda^i$ | $\sum_{j=1}^{i} \lambda^j$ |
|---|---|---|---|
| 1 | $\sigma^1 : P_5 \succ P_3 \succ P_1 \succ P_0$ | 0.23 | 0.23 |
| 2 | $\sigma^2 : P_6 \succ P_3 \succ P_1 \succ P_0$ | 0.21 | 0.44 |
| 3 | $\sigma^3 : P_7 \succ P_3 \succ P_1 \succ P_0$ | 0.18 | 0.62 |
| 4 | $\sigma^4 : P_2 \succ P_0$ | 0.15 | 0.77 |
| 5 | $\sigma^5 : P_1 \succ P_0$ | 0.09 | 0.85 |
| 6 | $\sigma^6 : P_6 \succ P_0$ | 0.07 | 0.93 |
| 7 | $\sigma^7 : P_0$ | 0.04 | 0.97 |
| 8 | $\sigma^8 : P_4 \succ P_3 \succ P_0$ | 0.01 | 0.98 |

product attributes for each customer segment. We compare the observed sales data with the predicted sales using the results of MNL models for validation of the proposed approach.

As a first step, we implement the column generation approach proposed by [6] for solving the non-parametric choice problem utilizing the transaction data in Table 2.4. The run is executed on a PC with 64-bit processor and 2.5 Ghz Intel(R) Core(TM) system utilizing IBM Ilog-Cplex version 12.3. The column generation algorithm terminates when the objective function of the sub-problem, as formulated in (2.2), is zero. The algorithm generated 25 dominant product order rankings, i.e., customer segments. Like other column generation approaches, we observed the tailing off effect where the fast improvement in the objective function is followed by a little progress per iteration as we approach to the optimum solution. Table 2.7 displays the eight customer segments which have more than 1% market share, cumulatively accounting for 98% of the customer transactions. The customer segments are sorted in the decreasing order of market shares, where the ranking $\sigma^1$ is assigned to the permutation $P_5 \succ P_3 \succ P_1 \succ P_0$ representing 23% of the customers.

From the top four rankings/customer segments, we observe that each segment has a separate favorite product; products $P_5, P_6, P_7$, and $P_2$ for customer segments $\sigma^1, \sigma^2, \sigma^3$, and $\sigma^4$, respectively. Furthermore, product $P_3$ is the second favorite product for the top three rankings. Due to stock-out based substitution, $P_3$ is also expected to sell significantly. The sales data in Table 2.4 generated via POS data also confirms that $P_2, P_3, P_5, P_6$, and $P_7$ are high selling products. The outlier in our analysis is product $P_4$ and we closely examine its sales behavior. 41 units of $P_4$ were sold during the study period, accounting for almost 10% of the sales, whereas the non-parametric choice model revealed only one ranking that includes $P_4$, with a market share of 1%. When we investigate the data in Tables 2.3 and 2.4, we observe that almost all $P_4$ units were sold when $\mathcal{S}_2 = \{1, 2, 3, 4, 7\}$ was present. Surprisingly, only one $P_4$ unit is sold when presented in $\mathcal{S}_5 = \{1, 2, 3, 4, 5, 6, 7\}$ and $\mathcal{S}_6 = \{1, 2, 4\}$. Because pricing and promotion data was not available, the unusual sales behavior for $P_4$ is not thoroughly explainable.

Next, we allocate the transaction data to customer segments. For brevity and conciseness of the analysis, we focus on the top four customer segments in the remainder of this section, accounting for 77% of the customer transactions. The POS sale data allocated to the top four rankings is displayed in Table 2.8. The *Other* row in Table 2.8 represents the transactions that are not assigned to these four segments. This analysis can be readily extended to all 25 customer segments identified by the non-parametric choice model.

The MNL model coefficients for attributes for the top four customer segments are displayed in Table 2.9. For the first customer segment, $\sigma^1 : P_5 > P_3 > P_1 > P_0$, the attributes material, length, and fashion are relatively important. Referring to attribute levels in Table 2.5, customers in this segment prefer mesh (material), full length (length), and basic (fashion). An ideal product based on the MNL model for this segment, purely from a sales perspective (not account-

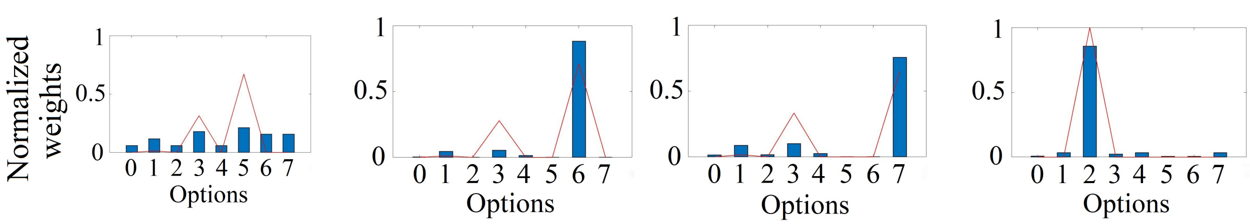
Table 2.8: POS transactions allocated to identified top four customer segments.

| Segment | $P_1$ | $P_2$ | $P_3$ | $P_4$ | $P_5$ | $P_6$ | $P_7$ | $P_0$ (walk) |
|---|---|---|---|---|---|---|---|---|
| $\sigma^1 : P_5 > P_3 > P_1 > P_0$ | 1 | 0 | 23 | 0 | 49 | 0 | 0 | 0 |
| $\sigma^2 : P_6 > P_3 > P_1 > P_0$ | 2 | 0 | 24 | 0 | 0 | 61 | 0 | 0 |
| $\sigma^3 : P_7 > P_3 > P_1 > P_0$ | 1 | 0 | 21 | 0 | 0 | 0 | 39 | 0 |
| $\sigma^4 : P_2 > P_0$ | 0 | 75 | 0 | 0 | 0 | 0 | 0 | 0 |
| Other | 26 | 0 | 7 | 41 | 0 | 0 | 0 | 33 |

ing for profitability and other factors), is full length, black, campus style, basic yoga pants with mesh. In order to assess the validity of the MNL model for this segment, we compute the purchase probabilities of the products using the attribute coefficients using the following formulation: $\widehat{Pr}^1(P_n|\mathcal{N}) = \dfrac{e^{\widehat{\beta}^1 X_n}}{\sum_{i\in\mathcal{N}} e^{\widehat{\beta}^1 X_i}}, n \in \mathcal{N}$. Note that we assume the presence of a complete assortment while calculating the estimated purchase probabilities. The blue bars in Figure 2.1a displays the purchase probabilities estimated using MNL model and the red line shows the purchase percentages of products within the assigned transaction data to this segment. Qualitatively, the MNL model performs satisfactorily and identifies the top two products with the highest sales volume for

Table 2.9: MNL coefficients for attributes for the top four segments.

| Segment | $\beta_{Material}$ | $\beta_{Length}$ | $\beta_{Style}$ | $\beta_{Fashion}$ | $\beta_{Color}$ |
|---|---|---|---|---|---|
| $\sigma^1 : P_5 > P_3 > P_1 > P_0$ | 0.97 | -1.10 | -0.66 | 0.97 | -0.31 |
| $\sigma^2 : P_6 > P_3 > P_1 > P_0$ | 10.49 | -7.68 | -7.53 | -6.14 | 7.35 |
| $\sigma^3 : P_7 > P_3 > P_1 > P_0$ | -37.81 | -1.77 | -1.62 | 3.42 | -2.16 |
| $\sigma^4 : P_2 > P_0$ | -5.21 | 3.65 | 3.20 | 0.00 | 0.00 |




(a) $\sigma^1 : 5 > 3 > 1 > 0$ (b) $\sigma^2 : 6 > 3 > 1 > 0$ (c) $\sigma^3 : 7 > 3 > 1 > 0$ (d) $\sigma^4 : 2 > 0$

Figure 2.1: Purchase probabilities of products using MNL model results (blue bars) and purchase percentages of products based on allocated transaction data to each segment (red line).

this segment, $P_5$ and $P_3$.

The MNL model coefficients for customer segment 2, $\sigma^2 : P_6 > P_3 > P_1 > P_0$, reveal that customers have strong opinions towards all attributes. Customers value the products with mesh the most. They prefer full length, campus style, fashionable, colored yoga pants. Attribute-wise, $P_6$ is the closest to the ideal, followed by $P_3, P_1$, and $P_4$ for customers in segment 2. The customers in segment 3 seem to display very strong dislike towards products with mesh, i.e., $P_5$ and $P_6$. The other attributes differentiate the preferences for the remaining products without mesh. Being very similar to the ideal product for this segment, $P_7$ captures a significant portion of the demand, followed by $P_3, P_1$, and $P_4$. The results of the MNL model for segment 4 demonstrates that material, length, and style are the only important attributes that seem to matter for the customers. The ideal product for this segment is $P_2$ in terms of sales volume and it captures nearly all the demand. In summary, the MNL models for the top four segments seem to directionally capture customer behavior and satisfactorily identify customer preferences towards products.

### 2.4.2 Robustness Analysis

In this section, we conduct an analysis to test the predictive power of the proposed approach for CCM. Toward this goal, we artificially 'remove' some of the prominent products (one product at a time) from the transaction data and conduct the customer-choice modeling analysis with the resulting partial product set and transaction data. The results from the complete product set transaction data analysis in Section 2.4.1 are then utilized to verify the product attractiveness estimates for the removed products from the partial product set transaction datasets. Furthermore, we study the robustness of the proposed approch in identifying the customer segments with associated share fractions and product purchase probabilities under incomplete or partially missing transaction dataset.

#### 2.4.2.1 Predictive Power in Estimating Attractiveness for Missing Products

In this section, we assess the effectiveness of the proposed approach for estimating the demand for products that are yet to be offered to customers. The experiments for this objective can be designed in various ways depending on the data and resource availability. A straightforward design would be creating and offering a new product and comparing the observed demand with the one generated by the proposed algorithm. Though this design is not far from ideal, it is impractical since studied period does not involve a new product launch, and design, production, and distribution of a new product requires a lengthy process. Hence, we chose to create an artificial transaction history by removing transactions for a select product. Let $\bar{P}$ represent the selected product for artificial removal from the historical dataset. In the real-world, a customer whose preference is $\bar{P}$ would either select another product from the assortment or walk away. Hence, a transaction that is removed in this experimental design should be recounted as a purchase of another product or walk away, according to customers' preferences. However, preferences of individual customers are not known, and recounting the removed transactions as walk-away or purchase of another product would bring bias to the analysis. Hence, we chose to simply delete the transactions for $\bar{P}$ and implement Algorithm 1 in Section 2.3 using only the remaining POS data. We identify new set of customer segments and fraction of customers in each segment. Then, the remaining POS data is randomly distributed among the customer segments with probabilities following (2.3), and an MNL model is fit based on the data assigned to each customer segment. We estimate the market shares of each product, including $\bar{P}$, using the estimated MNL model coefficients. Hence, the attractiveness of $\bar{P}$ in comparison to other products is estimated for each customer segment without using any transaction data related to $\bar{P}$.

We performed this analysis for three different products, $P_5, P_6$, and $P_7$. While the results using $\bar{P} = P_6$ are presented and discussed in this section, we report the results using $\bar{P} = P_5$ and $P_7$ in the Appendix. The POS data upon removing the transactions for $P_6$ is displayed in Table 2.10. We implement the non-parametric choice approach and obtain the customer segments and fraction of customers as in Table 2.11. For clarity in presentation, we present the top four segments with the largest fractions. While three out of these four customer segments are also identified in the original analysis, we observe a new customer segment with ranking $5 \succ 4 \succ 0$ in the absence of transactions for $P_6$. The POS data is distributed among the customer segments and an MNL model is fit to the top four segments. The MNL coefficients are displayed in Table 2.11, where the purchase probabilities of products including $\bar{P} = P_6$ are given in Table 2.12.

The top part of Table 2.12 presents the demand distribution of products as well as walkaway probability for the top four customer segments using the MNL model coefficients in Table 2.11, i.e., $\widehat{Pr}^k(P_n|\mathcal{N}') = \frac{e^{\widehat{\beta}^k X_n}}{\sum_{i\in\mathcal{N}'} e^{\widehat{\beta}^k X_i}}, n \in \mathcal{N}', k \in \mathcal{K}, \mathcal{N}' = \mathcal{N} \setminus \{6\}$. The bottom part of Table 2.12 displays the estimated demand distribution when $P_6$ is offered to customers, $\widehat{Pr}^k(P_n|\mathcal{N}) = \frac{e^{\widehat{\beta}^k X_n}}{\sum_{i\in\mathcal{N}} e^{\widehat{\beta}^k X_i}}, n \in \mathcal{N}, k \in \mathcal{K}$. We observe that $P_6$ becomes the first choice or the most attractive product for customers with rankings $P_5 > P_3 > P_1 > P_0$ and $P_5 > P_4 > P_0$ before $P_6$ is introduced. Customers with ranking $\sigma_1$ and $\sigma_3$ are not particularly interested in $P_6$ and the introduction of $P_6$ does not alter the demand distribution in these customer segments. Furthermore, we observe a decrease in walkaway probability for customers in $\sigma^2$ and $\sigma^4$. Using the customer segment sizes given in Table 2.11, we estimate that $P_6$ is expected to be as popular as $P_2$ and $P_5$, which are the highest selling products among customers from the top four segments.

#### 2.4.2.2 Robustness of the Customer Segments with Partial Data

In Section 2.4.1, customer segments and fraction of customers in each segment were identified. The goal of this section is to investigate the robustness of the proposed method in consistently

Table 2.10: Sales data upon removing transactions for $P_6$

| Sales data | | **Assortment** | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *9* |
| SKU | $P_1$ | 2 | 9 | 5 | 1 | 4 | 4 | 2 | 3 |
| | $P_2$ | 3 | 18 | 16 | 1 | 27 | 1 | 1 | 8 |
| | $P_3$ | 7 | 58 | 3 | 6 | 1 | 0 | 0 | 0 |
| | $P_4$ | 0 | 39 | 0 | 0 | 1 | 1 | 0 | 0 |
| | $P_5$ | 4 | 0 | 16 | 0 | 19 | 0 | 0 | 10 |
| | $P_7$ | 0 | 0 | 7 | 0 | 24 | 0 | 0 | 8 |
| No purchase, $P_0$ | | 1 | 12 | 6 | 0 | 10 | 0 | 0 | 4 |

Table 2.11: Customer segments, shares, and MNL model coefficients for attributes using data from Table 2.10 ($\bar{P} = P_6$).

| | | | MNL Coefficients | | | | |
|---|---|---|---|---|---|---|---|
| Segment | $\lambda^i$ | $\sum_{j=1}^{i} \lambda^j$ | $\beta_{Material}$ | $\beta_{Length}$ | $\beta_{Style}$ | $\beta_{Fashion}$ | $\beta_{Color}$ |
| $\sigma^1 : P_7 > P_3 > P_1 > P_0$ | 0.24 | 0.24 | -1.46 | -1.13 | -0.99 | 2.24 | -1.25 |
| $\sigma^2 : P_5 > P_3 > P_1 > P_0$ | 0.22 | 0.46 | 3.44 | -1.13 | -0.96 | 0 | 0.96 |
| $\sigma^3 : P_2 > P_0$ | 0.19 | 0.65 | -3.64 | 3.41 | 3.05 | 0 | 0 |
| $\sigma^4 : P_5 > P_4 > P_0$ | 0.12 | 0.78 | 2.91 | 0 | 0 | -1.56 | 0 |

identifying the segments and associated share fractions under data sampling or if the dataset is quite small. For brevity, we limit our analysis to five rankings with the highest fraction of customers. We study robustness under low and high data loss levels by randomly selecting 80% and 40% of the transaction data, respectively. For each level of data loss, we replicate the modeling 70 times to test the robustness of the proposed method. In each replication $r$, the non-parametric choice modeling step is implemented using the selected data to find the top five rankings having the highest fraction of customers $\sigma_r^i$ and the corresponding fractions $\lambda_r^i$, $i = 1, ...5$ and $r = 1, ..., 70$. Our goal is to observe the level of agreement/consistency between $\sigma_r^i$ and the original $\sigma^i$. Table 2.13 displays the

Table 2.12: Demand distributions without and with $\bar{P} = P_6$ for the top four segments.

| | Demand Distribution without $P_6$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | $P_0$ | $P_1$ | $P_2$ | $P_3$ | $P_4$ | $P_5$ | $P_7$ | |
| $\sigma^1 : P_7 > P_3 > P_1 > P_0$ | 5.3% | 14.2% | 5.3% | 16.3% | 5.3% | 4.3% | 49.5% | |
| $\sigma^2 : P_5 > P_3 > P_1 > P_0$ | 4.6% | 12.0% | 4.6% | 14.3% | 4.6% | 55.2% | 4.6% | |
| $\sigma^3 : P_2 > P_0$ | 1.1% | 3.9% | 82.3% | 2.7% | 3.9% | 2.2% | 3.9% | |
| $\sigma^4 : P_5 > P_4 > P_0$ | 5.1% | 3.5% | 3.5% | 3.5% | 16.7% | 64.2% | 3.5% | |
| | Demand Distribution with $P_6$ | | | | | | | |
| | $P_0$ | $P_1$ | $P_2$ | $P_3$ | $P_4$ | $P_5$ | $P_7$ | $P_6$ |
| $\sigma^1 : P_7 > P_3 > P_1 > P_0$ | 5.2% | 14.0% | 5.2% | 16.1% | 5.2% | 4.2% | 48.9% | 1.2% |
| $\sigma^2 : P_5 > P_3 > P_1 > P_0$ | 1.9% | 4.9% | 1.9% | 5.9% | 1.9% | 22.6% | 1.9% | 59.0% |
| $\sigma^3 : P_2 > P_0$ | 1.1% | 3.8% | 80.6% | 2.7% | 3.8% | 2.1% | 3.8% | 2.1% |
| $\sigma^4 : P_5 > P_4 > P_0$ | 3.1% | 2.1% | 2.1% | 2.1% | 10.2% | 39.1% | 2.1% | 39.1% |

Table 2.13: The number of times (%) each of the top five rankings from Section 2.4.1 are indeed observed as $i^{th}$ largest segment using randomly selected 80% of transaction data under 70 replications.

| Ranking | Largest | 2nd largest | Segment<br>3rd largest | 4th largest | 5th largest |
|---|---|---|---|---|---|
| $P_5 > P_3 > P_1 > P_0$ | 32 (46%) | 23 (33%) | 8 (11%) | 3 (4%) | 0 (0%) |
| $P_6 > P_3 > P_1 > P_0$ | 28 (40%) | 18 (26%) | 11 (16%) | 5 (7%) | 3 (4%) |
| $P_7 > P_3 > P_1 > P_0$ | 6 (9%) | 16 (23%) | 24 (34%) | 10 (14%) | 7 (10%) |
| $P_2 > P_0$ | 0 (0%) | 9 (13%) | 18 (26%) | 24 (34%) | 13 (19%) |
| $P_1 > P_0$ | 0 (0%) | 0 (0%) | 4 (6%) | 23 (33%) | 36 (51%) |
| Other | 4 (6%) | 4 (6%) | 5 (7%) | 5 (7%) | 11 (16%) |

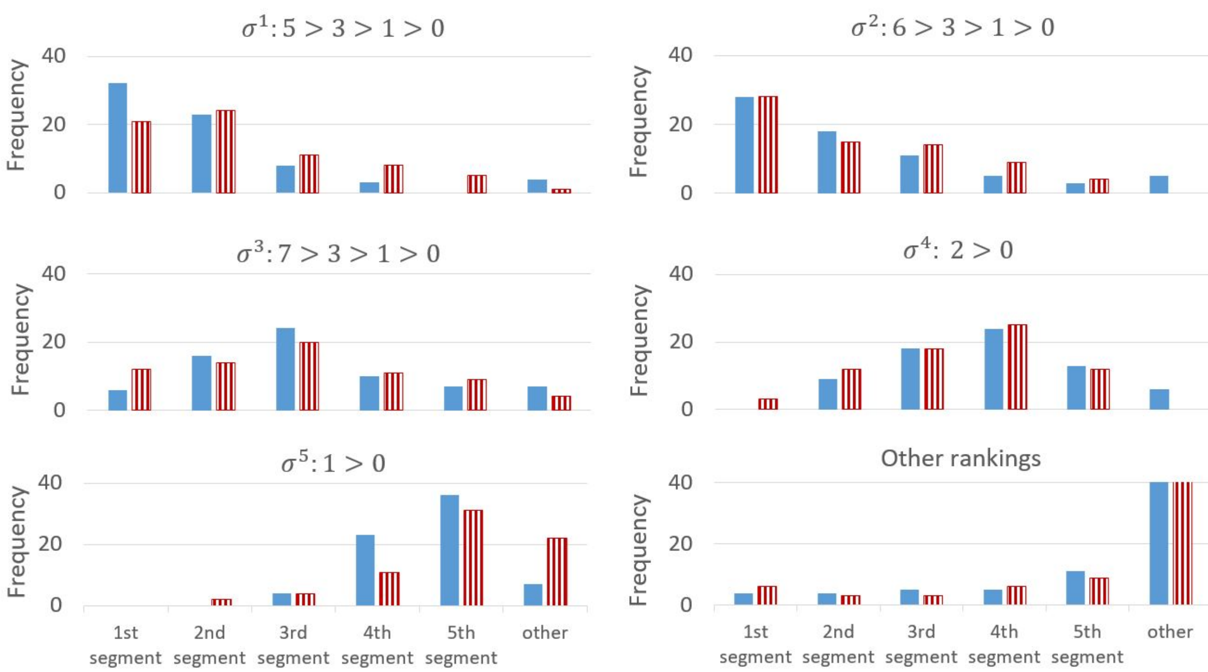



Figure 2.2: Distribution of the number of times each of the top five rankings from Section 2.4.1 observed as $i^{th}$ largest segment using randomly selected 80% (blue bars) and 40% (red-striped bars) of transaction data with 70 replications.

number of times and percentages each of the top five rankings from Section 2.4.1 are observed as $i^{th}$ largest segment, $i = 1, \ldots, 5$ using randomly selected 80% of the transaction data. Figure 2.2 compares robustness under low and high level data losses.

The robustness analysis reveals that the original ranking 1, $\sigma^1 : P_5 > P_3 > P_1 > P_0$, has a

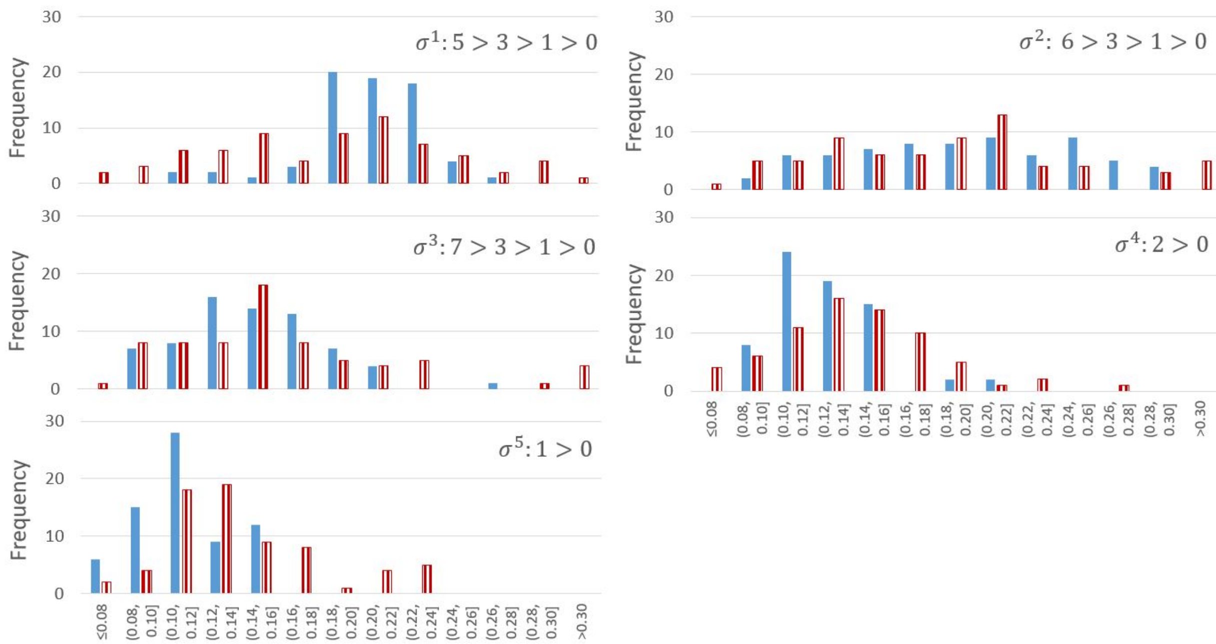



Figure 2.3: Distribution of market shares for customer segments using randomly selected 80% (blue) and 40% (red-striped) of the dataset, replicated 70 times.

higher tendency to be the top ranking even under sample datasets (first row in Table 2.13). Similarly, original rankings $\sigma^3$, $\sigma^4$, and $\sigma^5$ have higher propensity to be $3^{rd}, 4^{th}$ and $5^{th}$ segments, respectively. However, $\sigma^2 : P_6 > P_3 > P_1 > P_0$ is observed to take the top position more often that the second position (28 vs. 18 in second row of Table 2.13). More importantly, this illustrative experiment reveals that in 91% of all the 70 experiments, the same rankings remained in the top 5, which provides strong evidence for the robustness of the proposed approach.

Furthermore, the results obtained using only 40% of the transaction data (high data loss) are in agreement with the results obtained using 80% of the transaction data (low data loss), though exhibiting higher variances and smoother distributions, as displayed in Figure 2.2.

Next, we investigate the distribution of fraction of customers in each segment, or market share, as displayed in Figure 2.3, where the blue and red-striped bars correspond to experiments utilizing randomly selected 80% (low data loss) and 40% (high data loss) of transaction data. In Figure 2.3, each bin covers a range with width 0.02, except the first and last bins which covers values

<0.08 and >0.30, respectively. The $y-$axes in Figure 2.3 show the number of times observed market share values fall into these bins. The market shares of $\sigma^1, \sigma^3, \sigma^4$, and $\sigma^5$ follow bell-shape distributions with means 0.21, 0.16, 0.14, and 0.12, and standard deviations 0.030, 0.036, 0.026, and 0.023, respectively. On the order hand, the market share for $\sigma^2$ follows a more dispersed distribution, close to uniform, with mean and standard deviation of 0.20 and 0.55, respectively. When only 40% of the transaction data is utilized, the distributions of market shares becomes more dispersed, with sample means 0.20, 0.19, 0.17, 0.15, and 0.15 and standard deviations 0.058, 0.062, 0.056, 0.039, and 0.039, respectively for $\sigma^1, \sigma^2, \sigma^3, \sigma^4$, and $\sigma^5$. Though the variations are increased with less data, the results stayed reasonably stable, as observed in Figure 2.3.

#### 2.4.2.3 Robustness of Product Purchase Probabilities with Partial Data

The robustness of customer segmentation and market share when partial transaction data is available is examined in Section 2.4.2.2. In this section, we investigate the robustness of product purchase probabilities obtained by the MNL models using partial transaction data. As in Section 2.4.2.2, the low and high data loss correspond to analysis using 80% and 40% of the randomly

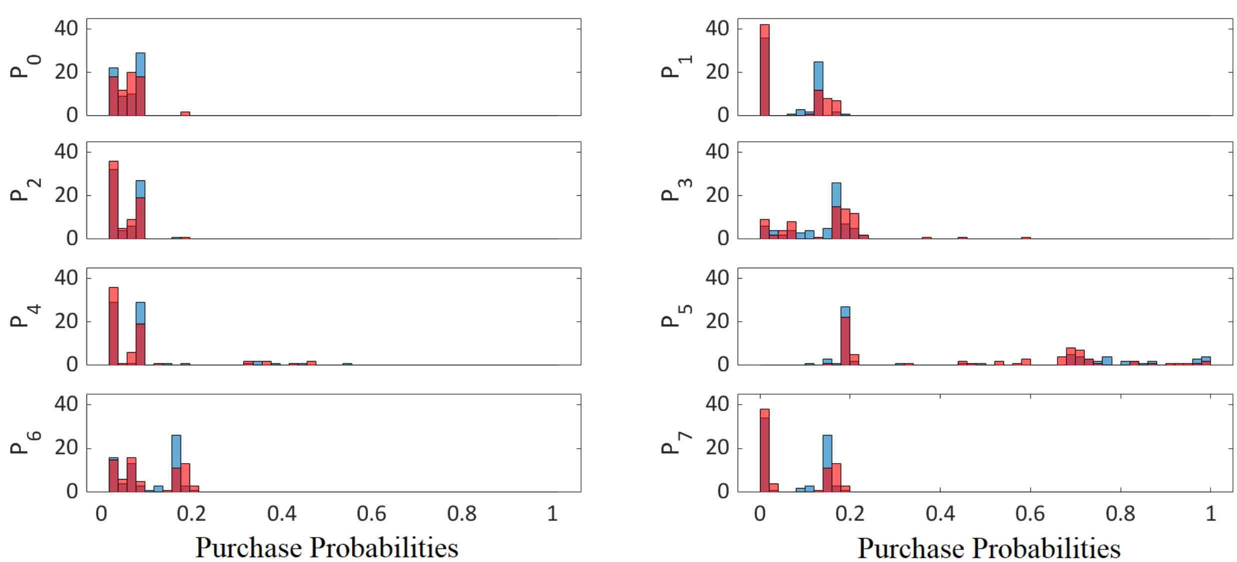



Figure 2.4: Histogram of product purchase probabilities by customers in $\sigma^1 : P_5 > P_3 > P_1 > P_0$ using 80% (blue) and 40% (red) of randomly selected transaction data.

selected transaction dataset. We replicate the experiments 70 times for each level of data loss. For each replication, we obtain the customer segments and associated market shares as discussed in Section 2.4.2.2. Next, the randomly selected transaction data is assigned to identified customer segments and an MNL model is fit for each segment. Using the MNL model parameter estimates for product attributes, we calculate the purchase probability of product $i$ for segment $k$ in replication $r$ by $\widehat{Pr}_r^k(P_i|\mathcal{N}) = \frac{e^{\widehat{\beta}_r^k X_i}}{\sum_{j\in\mathcal{N}} e^{\widehat{\beta}_r^k X_j}}, i \in \mathcal{N}, k \in \mathcal{K}$.

For brevity of analysis, histogram of product purchase probabilities only for $\sigma^1 : P_5 > P_3 > P_1 > P_0$ and $\sigma^3 : P_7 > P_3 > P_1 > P_0$ are presented in Figures 2.4 and 2.5, respectively. The width of each bin on $x$-axes is 0.02 and the $y$-axes display the number of times estimated product purchase probability falls in the corresponding bin among 70 replications. The results in Figures 2.4 and 2.5 demonstrate that our proposed approach is robust in identifying the most favorite product for segments, $P_5$ and $P_7$, respectively for $\sigma^1 : P_5 > P_3 > P_1 > P_0$ and $\sigma^3 : P_7 > P_3 > P_1 > P_0$, even when only 40% of the transaction data is available for the analysis. While the second favorite product is successfully identified for $\sigma^1$, for segment $\sigma^3$ $P_4$ was incorrectly identified as the second

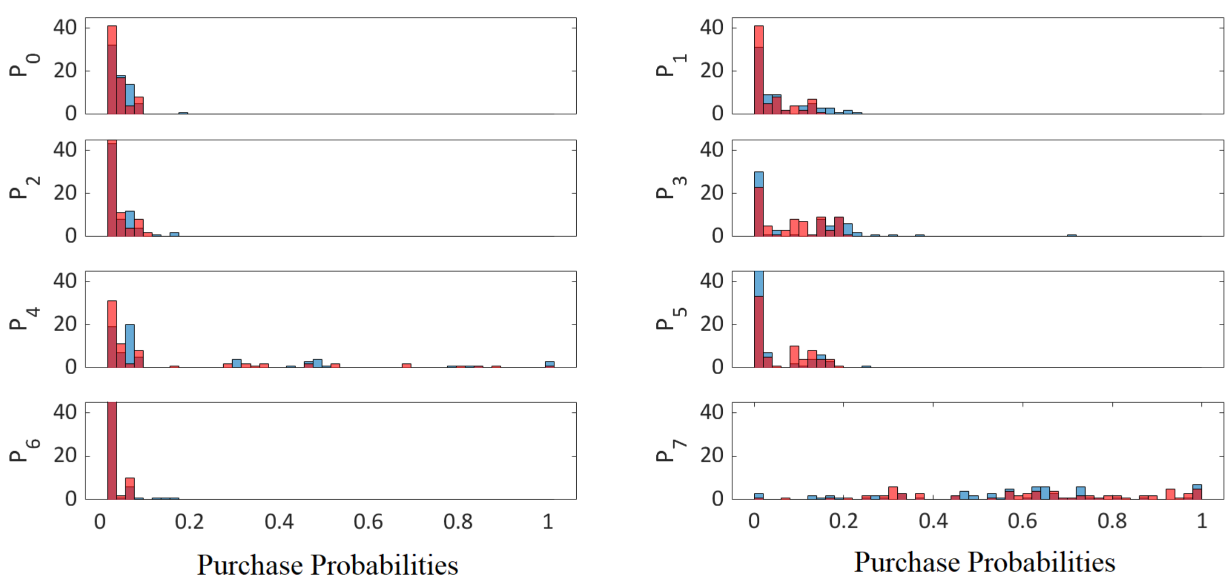



Figure 2.5: Histogram of product purchase probabilities by customers in $\sigma^3 : P_7 > P_3 > P_1 > P_0$ using 80% (blue) and 40% (red) of randomly selected transaction data.

favorite product while it should be $P_3$, which is identified as the third favorite product.

## 2.5 Conclusion

Customer choice modeling is challenging in practice due to limitations around the quality of the data available for modeling and potentially complex choice behaviors. We propose a hybrid modeling approach that relies on both parametric and non-parametric methods to derive effective recommendations for product assortment planning and product-line extension.

We recommend the utilization of non-parametric choice models to first extract an accurate ranking-based product choice model from sales transactions and inventory records. The resulting model is utilized to establish customer segments and derive more actionable product attribute-based parametric models for each segment that can be employed for product assortment optimization as well as product-line extension. The proposed modeling approach is validated using data from a leading global apparel retailer as well as synthetic experiments to evaluate the robustness of the proposed approach.

This study can be seen as a base for more comprehensive research on hybrid customer choice models. There are several avenues for future research. First and foremost, the methods should be tested across more application settings (both retail and other) for effectiveness. The methods should also be extended to account for learning across 'sister' product categories to not only improve choice modeling accuracy but potentially expand the space of product attributes for uncovering product-line extension opportunities. For example, choices learned from 'yoga pants' category (e.g., fabric types and colors) might hold promise to make inferences about choices for new 'yoga tops'. The employed MNL model only accounted for first order effects of attributes and did not address any potential interactions. The ability to comprehend and leverage these interaction effects and evaluation of other parametric choice models for assortment planning is also worthy of study. Finally, most of the non-parametric methods for learning distributions over product 'preference lists' currently cannot scale to large categories (with hundreds of SKUs). Effective heuristic methods need to be developed.

# CHAPTER 3

# Cross Category Analysis in Market Prediction and Product Development

## 3.1 Introduction

While a large number of new products are introduced every year, most of these products fail in the market place and this is particularly true for the highly competitive retail industry. These failed products often carry significant costs in many areas such as development, production, marketing, and brand damage. Overall, 94% of retail managers expect the rate of recent retail bankruptcies to continue unabated, with a substantial 40% share predicting that retail bankruptcy filings will grow [14]. It is estimated that $18 billion is lost in market share by top 25 food and beverage companies since 2009 and that 90% of top 100 consumer packaged goods brands lost market share in recent years [13]. Among the big store chains losing market share are TJ Maxx, Kohl's, Home Depot, Target, Victoria's Secret, Best Buy, Williams-Sonoma and Gap [12], quintessential U.S. brands.

Consumer packaged goods (CPG) market is also highly concentrated, with only a few companies accounting for majority of the overall global market [1]. Each of the major manufacturers, such as Unilever, Procter and Gamble, Kraft, and General Mills owns brands in a vast number of product categories. The most important customers of these manufacturers are large supermarket chains and department stores, who are multi-category firms as well. For these businesses, having a good understanding of customers' preferences and behavior that surpass product categories can be a source of gaining strategic advantage [2]. Thus, it is of vital importance for the manufacturers and retailers to develop a deep understanding for customer preferences and purchasing behaviors in order to offer better sets of products with greater chances of survival.

The science of studying customer shopping behavior and predicting their purchase pattern is known as Customer Choice Modeling (CCM). A customer choice model helps one understand and analyze the choices that customers make in a specific market. Customer choice models can assist companies in obtaining knowledge about factors such as product availability or pricing to influence customers' choices [19]. Firms can also use CCM to develop marketing campaigns that are tailored to specific market segments (or even individual customers) to gain advantage over their competitors.

Empirical CCM involves analyzing the purchasing behaviors of customers the market using sales and inventory data from categories of products. A category consists of a group of products (SKUs) with the same general functionality that are competing with each other [37]. Products in a category can be characterized by their set of attributes [19]. For instance, in the category of men's T-shirts, the attributes can be color, pattern, collar, sleeve length, brand, size, etc. Different products have different values for each attribute. No two products can have exactly the same attributes; if two items have the same attributes, we are missing one or more distinguishing attributes. Customer choice modeling methods may also consider other indicators and intrinsic customer properties such as demographics, income level, etc. in their prediction models.

Since the inception of CCM, scientists have introduced a variety of models to improve choice models. A common theme among the majority of these solutions has been using a single-category of products in the developed models, i.e., these approaches model the customers in an isolated environment in which only the "target" category is considered, and study the customer choice behavior within individual product categories [37]. Some of the most popular parametric approaches for single category CCM are multinomial logit (MNL) and mixed-logit (ML). While these methods provide satisfactory results in predicting customers choices within existing products, they often fail to predict the demand for new products that come with attributes, not initially observed in the existing products but might be present in other adjacent "sister" categories. The main limitation with single-category models is that they fail to work under the following two conditions: I) when a customer does not buy any of the existing products and walks away, II) when a new product with a new attribute is introduced to the market and there is not historical data available.

While in general, concluding any information from the above-mentioned conditions seem to be extremely difficult or even unfeasible, there are cases where some level of inference is in fact possible. For instance, if a dairy manufacturer is considering the introduction of a new attribute (fat-free) for ice cream for the first time, it may be able to infer some insight into potential customer demand based on the demand for fat-free milk or fat-free yogurt categories, for which historical data might exist.

Our studies have shown, as a frequent empirical finding, that although customers show different behaviors from one another in different categories, there are a few main sets of behaviors that can cover a vast number of customers' choices [32]. The goal here is to find out if customers exhibit similarities in their choice behavior across multiple categories and if these information can be used to make better predictions about existing and new products attractiveness (i.e., a quantitative measure of products' probability of being chosen). As an example, consider an apparel store that sells both shirts and pants. Shirts are classified by their color, size, and fabric, whereas pants are classified by color and size but not by fabric (e.g., all pants are currently produced using one material). The question is how the store can estimate customer inclination toward new pants with a new fabric from customers' behavior toward fabric variety of shirts.

While the notion of using multiple categories of products has been attracting interest for a long time (for instance, see a pioneering work by [9], which dates back to 1976), appropriate approaches to tackle the problem have not been developed until more recently (e.g., [3]). In their paper, [3] used data from five product categories, and found high correlations in price, display and feature sensitivity of households between these categories. Similarly, [17] in his research concluded that customers' preferences for a brand name is correlated across multiple categories of products. The findings in these studies have established the bases in the development of cross-category customer choice models, i.e., models in which customers' preferences have a joint distribution that allows *correlated-ness* across categories.

In other words, cross-category models study multiple categories of products together and provide a joint probability model for customer behavior in each category. Even though these methods are not as mature and accurate as single-category models, they try to address the issues with conditions I and II above, at least for the cases, where the required data on the customers or products are available.

Our focus in this research is on the use of observed household purchase data to analyze preferences in multiple categories. To benefit from the advantages of both methods, we present a hybrid approach of both single and multi category methods in this research. Our proposed approach uses a single-category CCM at the core of its prediction model. This is to utilize the simplicity and accuracy of single-category models. Next, to employ the data from customers' behavior across other categories, we append a cross-category model to our existing CCM. The cross-category model studies the target category and one other category at the same time. The result of the cross-category analysis is leveraged to provide a better estimate of customer choices for the target category.

The remainder of this chapter is organized as follows: Section 3.2 provides a literature review of customer choice modeling methods in the literature, with an added emphasis on [37] and [32]. Section 3.3 provides the methodology used in this study including the model and the solution. An application of the developed method in a case study is presented in Section 3.4. Section 3.5 closes the chapter with the conclusion and future steps.

### 3.2 Literature Review

Single-category CCMs only work with the data from the product category that customers will buy from; they may or may not take intrinsic properties of the customers (such as income, demographics, etc.) into account. This problem has been studied since 1980s and sizeable promising results have been developed since then. Multinomial logit (MNL), nested logit and mixed-logit are three of the most well-known approaches in solving these models. A vast literature on marketing studies the choice behavior within individual product categories ([32]). A consistent empirical finding across a number of studies is that consumer heterogeneity in preferences and sensitivity to marketing mix variables such as price and promotions, explains a significant portion of the variation in choices of customers. In other words, customers are very different from one another within each category of products. Considering this, an important question is whether a household exhibits similarities in it's choice behavior across different categories.

Reference [3] has been the first study to appropriately address this issue and develop methods for cross-category CCM. Soon after, [17] and [16] in their research discovered that consumers' preferences towards similar choices are correlated across different categories. Since then, cross-category choice models have gained attention in the literature. Cross-category models are referred to the ones that consider not only the target category from which customers choose, but also one or more other categories that are somewhat related to the target category in terms of attributes and nature of the products. The study of cross-categorical models is not as mature as single-category ones, and the results accuracy are not comparable yet. However, the idea of extracting useful information to predict customers' decision in one category by studying their behavior in other categories seems promising. These models are based on the finding that consumer preferences for brands and their responsiveness to marketing activities in each category have a joint distribution that allows correlated-ness across multiple categories. Potential application areas include branding and advertising ([17], [18]), brand equity and its extendability ([33]), cross-category promotions ([35]), etc.

The idea here, is to view products as their attributes, which was originally introduced by [19] in the context of modeling brand choice in an individual category. Customer's preference for a product in a category can then be viewed as a function of their preferences for the set of attributes of the product, and their responsiveness to marketing-mix variables. The vector of customers' preferences for attributes and responsiveness to various marketing variables is modeled as a function of observable household characteristics (i.e. demographic variables) and a small number of unobserved "factors" that originate from product attributes [37]. These components together capture the correlated-ness both within a single category and across multiple ones. This model can be viewed as an extension of the study conducted by [3], which proposed a variance component approach to model the cross-category choices.

### 3.2.1 Single-category analysis methods

In summary, the goal in single-category analysis is to find a function $U$, which explains the utility of each item or product $p_i$ based on their attribute vectors. Consider a category $C$ with $n$ items, $p_1, ..., p_n$ and $m$ attributes $a_1, ..., a_m$. Each item in this category can be uniquely identified in terms of the attributes by a vector, $p_i = [x_{i1}, ..., x_{im}]$, where $x_{ij}$ is the value of attribute $a_j$ for item $p_i$. The utility function can be written as $U(p_i) = \beta x_i$. The goal is to estimate these coefficients $\beta$. As mentioned in the literature, estimating $\beta$ has been studied for many years and several methods have been developed thus far, the two most important ones being the Multinomial-logit (MNL) and mixed-logit (ML). Both of these methods (as well as many more) are parametric models, with strong assumptions about customer heterogeneity, which as shown in [32] are not generally true, and, as a result, decrease the overall accuracy of the method.

This issue is taken care of in the model of [32]. This is a hybrid approach, combining parametric and non-parametric CCM models. The method starts with a non-parametric method to classify customers into several segments with the same behavior towards products using transactional sales data, guaranteeing customers in each segment have similar preferences towards the products in that category. Therefore, their behavior can be assumed heterogeneous, and the parametric method can be applied with much higher accuracy. Customer segmentation may be carried out using any non-parametric model. In [32], we used a model developed by [6], although other non-parametric models would work as well. The output of the non-parametric model is a set of customer segments $\Sigma = \{\sigma_1, ..., \sigma_k\}$, with each segment $\sigma_k$ representing $\lambda_k$ fraction of the population, $0 < \lambda_k < 1$. Then, performing a parametric model such as MNL on each segment $\sigma_k$, we obtain a vector of $\beta_k$, which shows the importance of each attribute in each segment and can be used to evaluate the attractiveness of existing products and any new proposed products for customers of this segment based on the attributes.

### 3.2.2 Cross-category analysis methods

As explained in the literature review, there are several cross-category CCM models, but in this study we utilize a method similar to the one developed in [37]. While other cross-category models in the literature work for categories of products with the same number of attributes, the model introduced by [37] can handle multiple categories, each with different number of common and category-specific attributes. This is achieved by benefiting from factors analysis, which transforms the attribute spaces of the target and sister categories into arbitrary-dimension spaces. Attributes are generally different across different categories, which makes their comparison difficult. Meaning, the factors act as bridges between different categories, and are applicable to all categories. Each factor takes several attributes into account. Thus, one can compare two categories, based on the factor coefficients. If the respective coefficients are similar in sign and magnitude, the underlying attributes are considered correlated and vice versa.

This method, in its general sense, tries to find a utility function

$$U_{hjct} = \beta_{hc}^{\intercal} X_{hjct}, h = 1, ..., H, c = 1, ..., C, j = 1, ..., J_c, t = 1, ..., T_h,$$

where $h$ denotes a household (or a customer segment), $c$ denotes a category of products, $j$ denotes an item within category $c$, and $t$ denotes a shopping instance for customer $h$. The model proposed by Singh to find the optimal values of these coefficients $\beta_{hc}$ is as follows:

$$\beta_{hc} = \Pi_c z_h + \Gamma_c \psi_h + \Lambda_c u_{hc}, c \in 1, .., C$$

This is a mixture model, where $z_h$ defines customer characteristics (e.g., income level, family size, ...), $\psi_h$ is a customer-specific vector of "factors," which is assumed to have a normal distribution $N(0; I_F)$, will generate dependence across preferences within and in between categories, and $u_{hc}$ is composed of i.i.d elements specific to category $c$ and is assumed to have a normal distribution $N(0; I_{K_c})$ (Here $K_c$ is the number of attributes in category $c$). This model can be solved using heuristic approaches such as hierarchical Bayesian models.

## 3.3 Methodology

In this section we demonstrate the terminology used in this chapter, and explain the methodology in details.

### 3.3.1 Terminology

The terms and notations that are used throughout this chapter are defined below:

**Item**: A specific product or SKU that customers may buy, denoted with $p_i$.

**Attribute**: A characteristic of an item, denoted with $a$, for example, brand, size, fat-content, etc.

**Category**: A group of items with the same functionality, denoted with $C = \{p_1, \cdots, p_n\}$. Each product in category $c$ can be explained with a set of attributes $A_c = \{a_1, ..., a_{n_c}\}$ where $n_c$ is the number of attributes.

NOTE: Each item $p_i$ can be represented as a vector in the attribute space $A_c$ as $p_i = [x_{ij}]$, $j = 1, ..., n_c$ where $x_{ij}$ denotes the value of attribute $a_j$ for item $p_i$. The matrix consisting of $x_{ij}$, where each row denotes a product and each column denotes an attribute is called a product attribute matrix, denoted as $P$.

**Target category**: The category under study, denoted with $C_T$.

**Sister category**: A category that shares some attributes with the target category and has at least one or more relevant attributes not in the target category, denoted with $C_S$.

The attributes of target and sister categories can be denoted as

$$A_T = \{a_{c_1}, ..., a_{c_{n_c}}, a_{t_1}, ..., a_{t_{n_t}}\}, A_S = \{a_{c_1}, ..., a_{c_{n_c}}, a_{s_1}, ..., a_{s_{n_s}}\}, \tag{3.1}$$

where $a_{c_1}, ..., a_{c_{n_c}}$ are the common attributes between $C_T$ and $C_S$, $a_{t_1}, ..., a_{t_{n_t}}$ are the attributes unique to $C_T$, and $a_{s_1}, ..., a_{s_{n_s}}$ are the attributes unique to $C_S$. While $n_t$ can be zero, $n_c, n_s$ must be at least one or greater.

### 3.3.2 Method

Using the above, the goal of this research is for a given target category, $C_T$, an attribute set, $A_T$, similar to (3.1), and $N^T$ items $p_1^T, ..., p_N^T$, to provide a method for estimation of attractiveness of every existing and potential new items. Then, we intend to improve the accuracy of these estimates by utilizing the data from another related category (sister category) $C_S$, which will enable us to expand the attribute set of to include new attributes not observed historically in our Target category. This is formally expressed as follows:

We start with a single-category CCM to obtain an initial estimate for the utility of each product based on the attributes that have enough historical data available in the target category. As explained, our choice for this part is based on the methodology presented earlier in Chapter 2 of this dissertation, because of the reasons stated in the mentioned chapter. Assuming that this step (single-category analysis) has provided us with $K$ rankings (heterogeneous customer segments with similar preferences toward products), the output would be $K$ vectors of coefficients $\beta_k^T$, each with a length of $n_C + n_T$, which denote the coefficients of attributes in the target category $C_T$ for each customer segment $k = 1, ..., K$.

Next we need to find a good category of products to use as the Sister category. While this may seem easy to do, not every intuitive choice would be a good candidate. For two categories to be sisters, it is necessary to have a sufficient number of common comparable attributes, which take similar importance levels across both categories. This, however, does not guarantee the usefulness of the sister category. For the sister category to be useful in estimation, one must ensure a meaningful correlation between the common attributes ($a_c$'s) of the two categories exists (This can be achieved by implementing the cross-category analysis). While this will be discussed later in this chapter, it is worth noting that the choice of a good sister category usually requires good domain knowledge accompanied by trial and error.

Then we need to extract heterogeneous customer segments in the sister category as we did for the Target category. The difference is that unlike the target category where segmentation was done using an optimization approach, here the goal is to find similar segments in sister category that are comparable to those in the target category, as much as possible. So we use the following method:

1. Identify the $N^S$ items in sister category, i.e., $p_1^S, ..., p_{N^S}^S$ (including the no purchase option).
2. extract the attributes of these items and divide them into common attributes $a_{c_1}, ..., a_{c_{n_c}}$ and

sister category attributes $a_{s_1}, ..., a_{s_{n_s}}$.

3. Construct the truncated matrix $\beta^T_{trunc}$, i.e., the coefficients $\beta^T_k$ corresponding to the common attributes only,
$$\beta^T_{trunc} = [\beta^T_{i,k}], \quad i \in \{a_{c_1}, ..., a_{c_{n_c}}\}, k = 1, ..., K.$$
4. Let $P^S$ be the attribute matrix of products in the sister category, with columns consisting of $a_{c_1}, ..., a_{c_{n_c}}$ and $a_{s_1}, ..., a_{s_{n_s}}$. Construct the truncated matrix $P^S_{trunc}$ with common attributes only as
$$P^S_{trunc} = [x_{ij}], \quad i \in \{p^S_1, ..., p^S_{n_S}\}, j \in \{a_{c_1}, ..., a_{c_{n_c}}\}.$$
5. Define the matrix $R^S$ for ranking product in each segment, as $R^S = P^S_{trunc} \times \beta^T_{trunc}$, where each element $r_{ik}$ of this matrix corresponds to the "attractiveness" of product $p^S_i$ in "inferred" segment $k$ (We call these segments inferred segments, because they were inferred from the segments of the target category, rather than being directly calculated using the transaction data of the sister category).

   Therefore, the matrix $R^S$ can be used to rank the products of the sister category in each inferred segment, by sorting the products based on their $r_{ik}$ in each segment $k$.
6. The result would be a set of rankings, similar to that of target category, in the form of $\sigma_k : \quad p_x > p_y > ... > p_z$.

Now having the product rankings in each segment of the sister category ready, we can assign each transaction of purchases from the sister category to a specific segment, based on their likelihood (similar to Chapter 2). To do so, for each transaction, observe the presented assortment, and decide which segment would purchase the purchased item. Three possibilities could arise:

1. Only one ranking (segment) corresponds to a transaction (assortment + purchased item combination).
2. Multiple rankings correspond to a transaction.
3. No rankings correspond to a transaction.

In the case of 1, assignment is easy. In the case of 2, the transaction is assigned randomly to one of the possible segments, with probability proportional to $\lambda_k$ obtained from target category segmentation. In the case of 3, the transaction is removed from the analysis. This means that the sister category has potentially more segments than the target one. Since, we are only concerned with improving the accuracy of the segments in the target category, the unique segments of the sister category will not play any role in the calculations.

Next, we use a cross-category CCM method to model the joint choice behavior of customers when considering both categories. Here, our choice is cross-category brand choice model developed in [37] as discussed earlier. The inputs to this model are transactional data from both categories of products consisting of the following columns: category of purchase, customer segment, present assortment at the time of purchase, and the purchased product.

The output of this step (cross-category analysis) would be $K$ vectors of coefficients $\widehat{\beta_{kS}}$, each with length $n_C + n_S$, for Sister category, and $K$ vectors of coefficients $\widehat{\beta_{kT}}$, each with length $n_C + n_T$, for Target category, which denote the coefficients of attributes in each category for each customer segment $k$. Even though [37] can take household demographics variables into account, for the purpose of our study, these variables are ignored. Therefore, our implementation of Singh's model can be simplified to

$$\beta_{hc} = \Gamma_c \psi_h + \Lambda_c u_{hc}.$$

Where $\psi_h$ is a customer-specific vector of "factors," which is assumed to have a normal distribution $N(0; I_F)$, will generate dependence across preferences within and in between categories, and $u_{hc}$ is composed of i.i.d elements specific to category $c$ and is assumed to have a normal distribution $N(0; I_{K_c})$. We've solved this, using hierarchical Bayesian approach. Also, in our implementation, we use the customer segments obtained from first step as households, meaning instead of $h$ we have $k$. This reduces the calculation load significantly, and since the customers in each segment are heterogeneous, does not affect the accuracy. Finally, in our implementation, there are always two categories $C_1 = C_T$ and $C_2 = C_S$. After running the model and obtaining the unknowns $\Gamma_c$ and $\Lambda_c$, one can calculate the coefficients $\beta_{kc}$.

Then, we must calculate the covariances between coefficients from “Target“ and “Sister“ categories to test if the utilized Sister category was a good choice. The assumptions we made in previous step, simplifies the calculation of the covariance matrix. So, using the above assumption, as shown by [37] we have,

$$cov(\beta_h) = \begin{bmatrix} \Gamma_1\Gamma_1^\intercal + \Lambda_1\Lambda_1^\intercal & \\ \Gamma_2\Gamma_1^\intercal & \Gamma_2\Gamma_2^\intercal + \Lambda_2\Lambda_2^\intercal \end{bmatrix}. \tag{3.2}$$

If analysis of all common attributes shows correlatedness in all (or majority of) attributes, we can assume customer behaviors are correlated in the two categories, therefore, we can import the the sister attributes $a_s$ from $C_S$ into $C_T$.

To show cross-category correlatedness, we focus on the term $\Gamma_2\Gamma_1^\intercal$. This term is an $(n_c+n_s) \times (n_c+n_t)$ matrix. Elements $e_{ii}$, $i = 1, ..., n_c$, explain the correlatedness of common attributes across the target and sister categories. Here, the conclusion is somewhat subjective. However, as a rule of thumb, if all $e_{ii}$'s are greater than 0.60, we can consider the two categories as correlated and move on to importing the attribute(s) of $C_S$ into $C_T$.

Finally, we use the result of the cross-category model to estimate the utility of items in $C_T$, by leveraging the calculated importance of attribute(s) in $C_S$. The coefficient of this new attribute in $C_T$ would be the scaled version $\beta_{kC}$. By scaling, we mean that the values of $\beta$ between the two categories for a given customer segment must be comparable. Recall that in utility function, $\beta$'s are only meaningful up to a scale factor. Therefore, when comparing the results of the single-category model with the cross-category model, it is important to make sure $\beta$'s are on the same scale.

To import the $\rho$th unique attribute of sister category (i.e., attribute $a_{s_\rho}$) into the target category, calculate the new coefficient (coefficient of $a_{s_\rho}$ in the target category), denoted as $\beta^T_{new,k}$, for each of the $k = 1, ..., K$ segments of the target category, using the following normalization process:

- Among the common attributes, calculate the following ratios. This gives us the average ratio of coefficients of Target category to those of Sister category.

$$r_k = \left( \sum_{i=1}^{n_C} \frac{\widehat{\beta_{i,k}^T}}{\widehat{\beta_{i,k}^S}} \right) / n_C, \quad k = 1, ..., K$$

- Calculate $\widehat{\beta_{new,k}^T} = r_k \times .\widehat{\beta_{\rho,k}^S}$. This provides us with the normalized coefficient of the Sister category to match coefficients of Target category.
- Convert $\widehat{\beta_{new,k}^T}$ to the original space, i.e., $\beta_{new,k}^T$ according to the formulation below. This ensures the scale of the new attribute that is being added to original model is consistent with existing ones.

$$\beta_{new,k}^T = \frac{\left( \sum_{i=1}^{n_C} \frac{\beta_{i,k}^T}{\widehat{\beta_{i,k}^T}} \right)}{n_C} \cdot \widehat{\beta_{new,k}^T}, \quad k = 1, ..., K.$$

The resulting $\beta_{new,k}^T$ can be used to augment the model of the target category. A summary of the approach presented in this section can be listed as follows:

1. Select a single-category method to model the target category, and Obtain $\vec{\beta_k^T}$s (coefficients that explain relative importance of each attribute for customers in $k$th segment of the model).
2. Find a sister category for the target category, consisting of a set of common attributes, and the new attribute we need to obtain additional information about.
3. Optional (only if single category methods has defined segments of customers): Segment the customers in the sister category into similar segments as the target category using the inferred segmentation logic.
4. Perform a cross-category analysis of the target category and its sister category, and obtain $\widehat{\beta_k^T}, \widehat{\beta_k^S}$.
5. Analyze the effect of the common attributes between the two categories to confirm there is a meaningful correlation between the two categories.
6. Adjust (scale) the coefficient of the imported attribute(s), and add them to the model of Step 1 and use this augmented model to predict the customer choice.

## 3.4 Results

The method described above has been applied to several datasets. All of these datasets were generated synthetically using a comprehensive simulation method, which is explained in data generation process. The reason for favoring synthetic data over real world is two-folds:

1. For the experiment to be evaluated, we need to know the actual values of the parameters, which is almost always impossible to obtain from real world data,
2. We are interested to run several versions of each experiment in a controlled environment.

Nevertheless, we developed an inclusive simulation model to provide the most conformance to real world behavior of customers as much as possible.

### 3.4.1 Data generation

The following procedure is performed for each experiment, to generate (simulate) the required transaction data.

#### 3.4.1.1 Inputs

The inputs of the simulation model are:

- Number of items in target and sister categories ($N^T$, $N^S$),
- Number of common attributes ($n_C$),
- Number of attributes specific to target and sister categories ($n_T, n_S$),
- Correlation value $R > 0$, which explains how correlated are the attractiveness of common attributes in the target category with those of the sister category,
- Number of customer segments to simulate ($K$),
- Number of time slots (e.g., days) to simulate ($T$),
- Number of customers to visit the store per day ($N$),
- probability of an item being present at store, i.e., in the assortment, in each day $P^t_{p_i}$,
- Number of factors (for factor analysis) $F$, and their definitions.

#### 3.4.1.2 Simulation setting

The following assumptions were enforced during the simulation.

- All attributes are binary, i.e., representation of each product in the attribute space is a vector of zeros and ones.
- The number of attributes unique to the target category is the same as that of the sister category, i.e., $n_T = n_S$. Even though it can be relaxed without causing any side-effects to final results, putting this assumption in place makes the results more easily interpretable.
- At each time slot (e.g., day) of simulation, $t = 1, ..., T$, the software decides on the present assortment to the customers randomly. Each item may be in stock with probability $P_{p_i}^t$, or out-of-stock with probability $1 - P_{p_i}^t$.
- Number of factors $F$ is assumed to be the same as number of attributes, i.e., $F = n_C + n_T$. In fact, we let each factor to correspond to one attribute. This does not affect the final result, but makes the results interpretations more meaningful and the math easier.

#### 3.4.1.3 Simulation process

First, the vector representation of each product in the target and sister categories are randomly generated. Then, we randomly generate an attractiveness number for each attribute of the target category ($a_{n_C}$ and $a_{n_T}$) and each customer segment $k$, of the target category ($\beta_k^T$), these numbers are selected independently.

Next, using the correlation value, $R$, randomly generate the $\beta$'s for the common attributes of the sister category. We assume $\beta_k^T$ is correlated with $\beta_k^S$, but we do not impose any correlation assumption among $\beta$'s of different attributes, nor those of different customer segments.

For each day of simulation, first decide on the assortment that will be presented to the customers. Randomly decide which products are present and which ones are out of stock. (If randomly one category is totally empty, recalculate the assortment.)

For each of the $N$ customers visiting on each day: (1) Assign them randomly to a customer segment proportional to $\lambda_k$. (2) Calculate their utility using the MNL model for all of the present products of both the target and sister categories. (3a) If utility of at least one item in each category is positive, generate a transaction consisting of the highest-utility-item of each category. (3b) If

| Parameters | Exp. 1 | Exp. 2 | Exp. 3 |
|---|---|---|---|
| $N^T$ | 4 | 3 | 3 |
| $N^S$ | 4 | 3 | 3 |
| $n_C$ | 3 | 2 | 2 |
| $R$ | 0.8 | 0.8 | 0.8 |
| $K$ | 3 | 3 | 3 |
| $N$ | 2000 | 2000 | 500 |
| $P_{p_i}^t$ | 0.4 | 0.4 | 0.4 |

Table 3.1: Parameters of different experiments carried out in study

one category does not offer any products with positive utility, generate a transaction with only one item (with the highest positive utility). (3c) If neither category offers any item with positive utility, the customer walks out without purchase, and we generate a walk away transaction.

#### 3.4.1.4 Outputs

The first output of the simulation process is the transaction table. Each row consists of the segment of the customer, presented assortments for each category, and the purchased products (including no-purchase option). The second output is the $K$ sets of vectors $\beta_k^T, \beta_k^S$, $k = 1, ..., K$, for each customer segment in each category.

### 3.4.2 Case studies

Using the above data generation method, we performed three different experiments. Experiment 1 is to find how the method works on a problem with medium number of attributes and products, with a reasonable number of transactions. Experiment 2, tests the results when the problem scope is smaller and we have fewer number of products and attributes. And finally, experiment 3 analyzes the results when number of transactions are insufficient. The details of each experiment are shown in Table 3.1.

#### 3.4.2.1 Experiment 1

In this subsection, we thoroughly study experiment 1. A summary of other experiments is provided in the following subsection. The steps that we follow in this experiment, as well as the others, are as described in Section 3.3.

| Customer segment $k$ | $\beta_k^T$ | $\beta_k^S$ |
|---|---|---|
| 1 | $[0.23, 0.68, 0.56, 0.68]$ | $[0.25, 0.65, 0.53, 0.69]$ |
| 2 | $[0.07, 0.74, 0.17, 0.02]$ | $[0.07, 0.82, 0.10, 0.02]$ |
| 3 | $[-0.07, -0.99, 0.03, -0.13]$ | $[-0.09, -0.98, -0.09, -0.15]$ |

Table 3.2: Actual values of $\beta$'s for Experiment 1

The only difference is that in these "controled" experiments, since we have access to the actual original coefficients of models $\beta$'s, we don't need to explicitly run the single-category method to estimate them. Instead, we use the original values. The benefit is that using the real values eliminates any errors from the single-category method, and thus makes the performance evaluation of the proposed method more accurate.

The steps and their outputs for Experiment 1 are as follows:

1. Model inputs
   - Two categories. Target (yogurt), Sister (ice cream).
   - Four SKUs in each category, ($p_i^T, p_i^S$, $i = 1, 2, 3, 4$).
   - Three common attributes ($a_{C_1}$ :fat, $a_{C_2}$ : lactose, $a_{C_3}$ : organic).
   - One attribute unique to the target category ($a_{T_1}$ : probiotic), and one unique to the sister category ($a_{S_1}$ : sugar).
   - The goal is to find the effect of $a_{S_1}$ on customers' decision regarding yogurts.
   - $K = 3$ customer segments, $\lambda^T = [0.2, 0.4, 0.4]$, i.e., 20% of customers belong to Segment 1, 40% to Segment 2, 40% to Segment 3.
   - A correlation of 80% ($R$ = 0.8) is assumed between common attributes of the two categories in all segments.
2. Simulate the data with the given input parameters to obtain $\beta_k^T, \beta_k^S$. Obtained $\beta$'s are shown in Table 3.2.
3. Run the cross-category analysis to obtain $\widehat{\beta_k^T}, \widehat{\beta_k^S}$.
   - The process of getting estimated coefficients $\widehat{\beta}$, is repeated 4 times, because of the

stochastic nature of the cross-category analysis. We need to make sure the results are consistent and close.

- After performing the four runs, we compare $\widehat{\beta}$'s with the real $\beta$'s to measure the variability. Usually, if variability is high, this step is repeated until an acceptable level of variability is achieved. For instance, in Experiment 3, we repeated this step almost 20 times. This is because in that case the solution is stuck in a local optimum, and as high as 20 replications were needed to find the globally optimal solution.
- Figure 3.1 shows the results of this steps. Plots on the left side belong to the target category, and the ones on the right belong to the sister category. Each row of plots corresponds to one customer segment, $k = 1, 2, 3$. In these plots the X-axis shows each of four attributes for each category, and Y-axis shows the values of coefficients for each of the attributes. The light blue chart is for original $\beta_k$s, and the four other charts represent the four $\widehat{\beta_k}$s obtained from four runs of cross category method. This figure shows us good robustness in estimated coefficient values for all attributes in each category and segment of customers.

4. Form the covariance matrix of factors, to make sure the estimated parameters show correlation close to $R$ as shown in (3.2). The diagonal elements of the bottom-left block of this matrix (corresponding to the term $\Gamma_2\Gamma_1^\intercal$) must be evaluated in order to ensure correlatedness. These values for the three segments are shown below. Note that only the first three diagonal elements are of interest (since they correspond to the three common attributes, i.e., *fat*, *lactose*, and *organic*. The fourth diagonal element shows correlation between *probiotic* of the Yogurt category and *Sugar* of the Ice cream category.

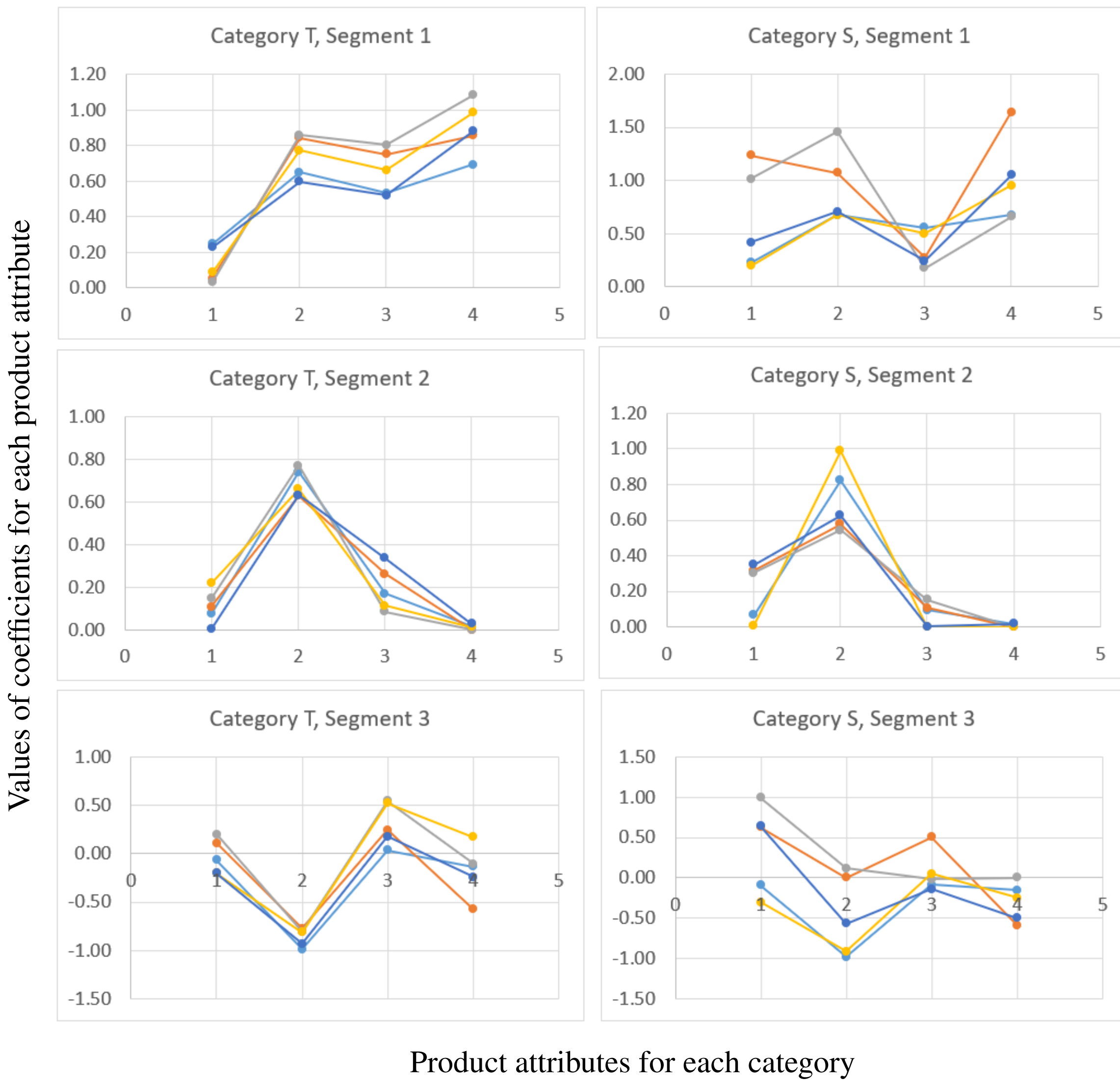



Figure 3.1: $\widehat{\beta}$'s for target and sister categories obtained from cross-category analysis Exp1

$$cov(\beta_1) = \begin{bmatrix} 0.76 & x & x & x \\ x & 0.88 & x & x \\ x & x & 0.86 & x \\ x & x & x & 0.80 \end{bmatrix} cov(\beta_2) = \begin{bmatrix} 0.81 & x & x & x \\ x & 0.89 & x & x \\ x & x & 0.79 & x \\ x & x & x & 0.51 \end{bmatrix} cov(\beta_3) = \begin{bmatrix} 0.79 & x & x & x \\ x & 0.83 & x & x \\ x & x & 0.89 & x \\ x & x & x & -0.30 \end{bmatrix}$$

- As we can see the elements are close to the original correlation parameter $R = 0.80$.

5. Calculate new attribute's coefficient $\beta^T_{new,k}$ using the normalization process explained in Step 6 of Section 3.3. The results for Experiment 1 are shown in Table 3.3.

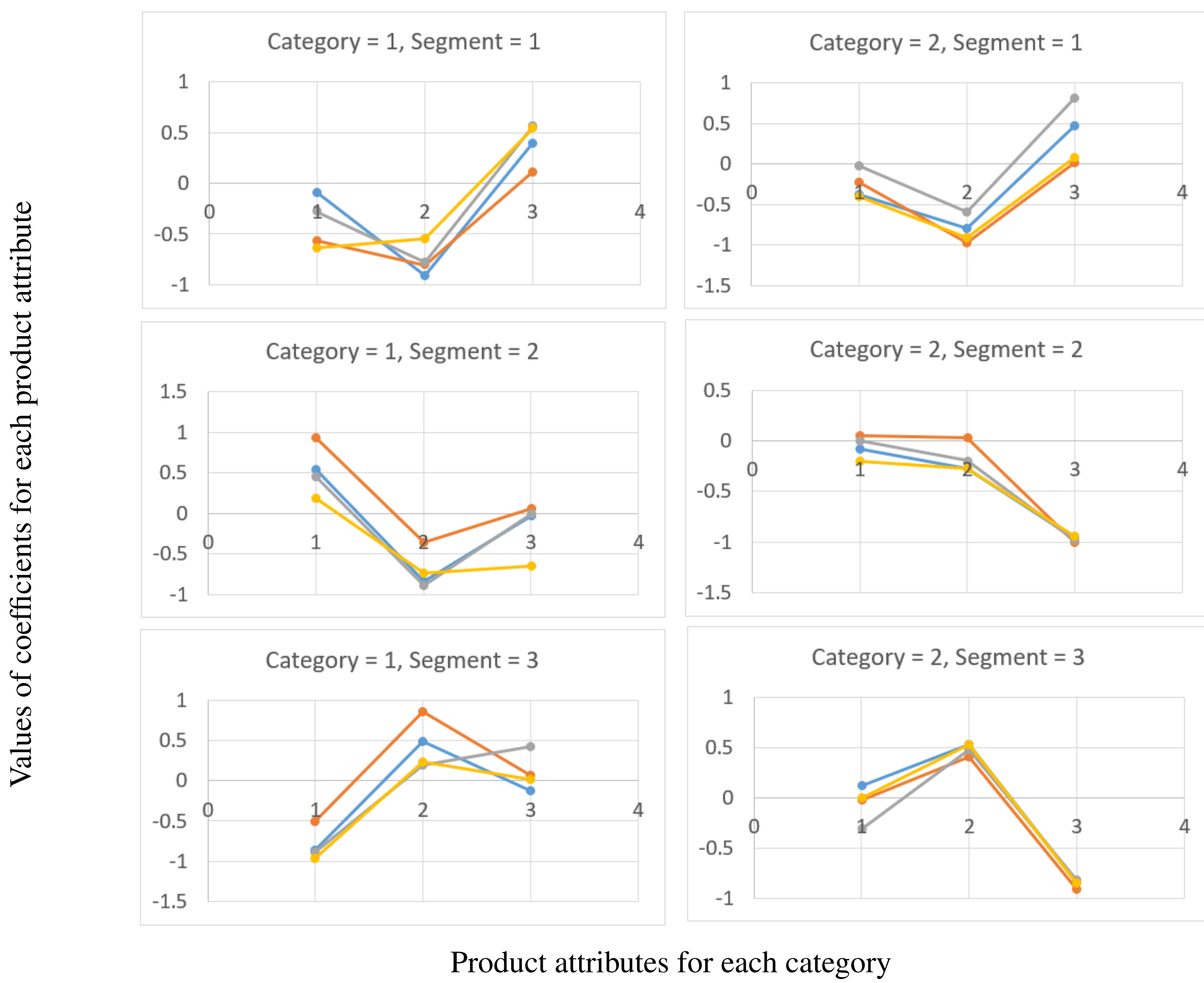
| Segment $k$ | $\widehat{\beta^S_{4,k}}$ | $r_k$ | $\widehat{\beta^T_{5,k}}$ | $\beta^T_{5,k}$ |
|---|---|---|---|---|
| 1 | 0.68 | 1.02 | 0.70 | 0.65 |
| 2 | 0.07 | 0.94 | 0.06 | 0.05 |
| 3 | 0.10 | 1.16 | 0.11 | 0.10 |

Table 3.3: Original $\beta$ of the attribute Sugar for sister category ($\widehat{\beta^S_{4,k}}$), its transformed value when imported to the target category ($\widehat{\beta^T_{5,k}}$), and when imported to the original coordinates ($\beta^T_{5,k}$)

#### 3.4.2.2 Other experiments

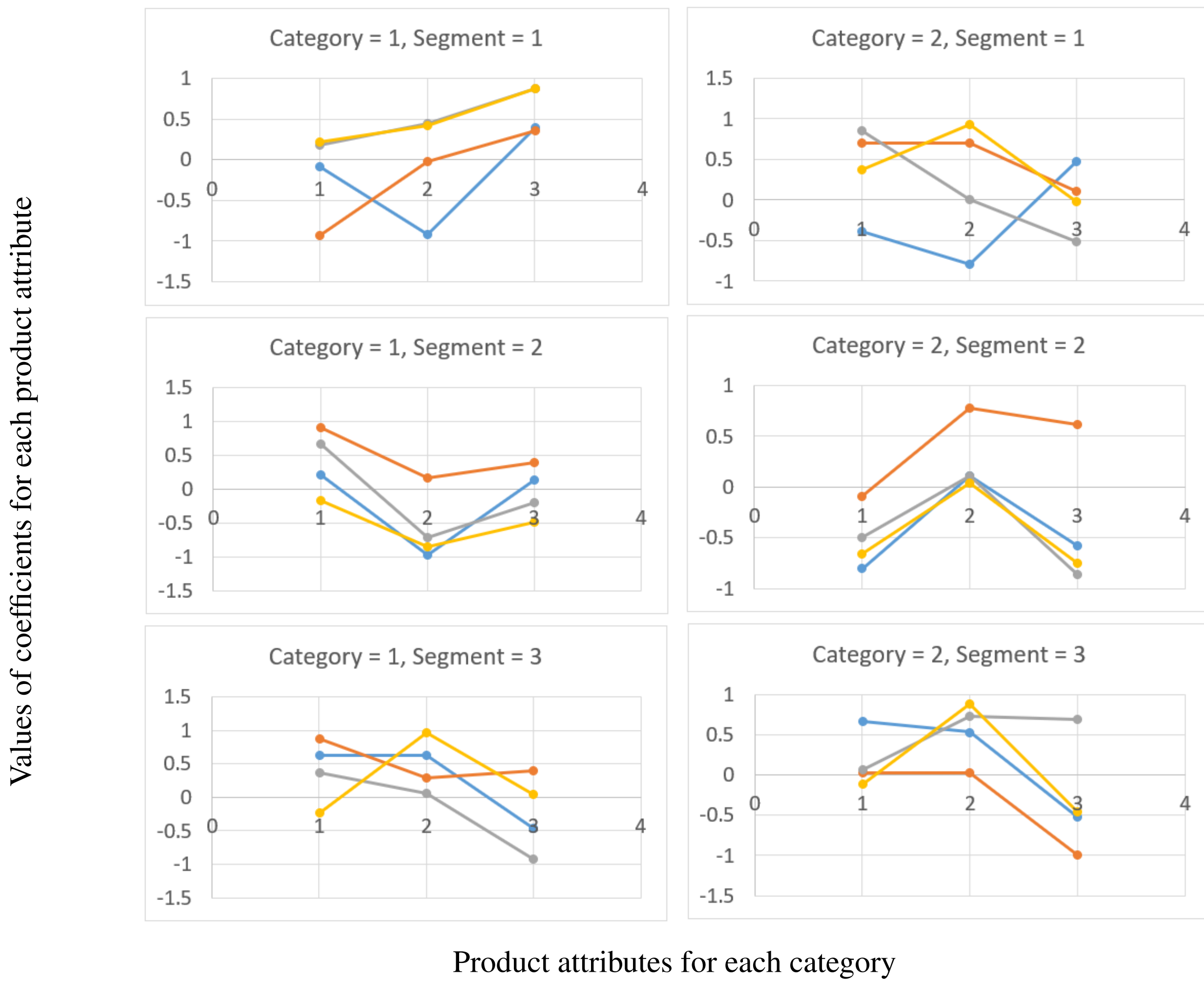
The process is the same for all experiments, here we only mention the outputs of cross-category model for Experiments 2 and 3 the same way we did for Experiment 1. The results for these experiments are shown in Figure 3.2.

These results of Experiment 2 shows that with the same number of transactions, the accuracy of cross category analysis is higher for a smaller problem with fewer number attributes for products, meaning the results show less variability and are more reliable. While Experiment 3 demonstrates how insufficient number of transactions, decreases the accuracy of analysis and bears the risk of the results being stuck in local optima.




(a)

Figure 3.2: $\widehat{\beta}$'s for target and sister categories obtained from cross-category analysis Exp2




(b)

Figure 3.2: $\widehat{\beta}$'s for target and sister categories obtained from cross-category analysis Exp3

### 3.5 Conclusion

The vast literature in marketing recognizes and models the dependence between consumers' choices across multiple categories. This chapter contributes to the literature by offering an approach of leveraging information based on correlatedness of preferences for attributes that are shared by different categories, to be utilized in new product development. Our method decomposes consumer preferences for each product in each category into preferences towards attributes of these products, some of which are common across different categories and at least one of them is not. Thereby we allow similarities in choice behavior across categories to be determined by correlated preferences for the attributes and make the base customer choice model better by using attribute importance from purchasing in existing categories to predict preferences for attributes in new categories.

Our proposed method uses the best practices of single-category and cross-category choice modeling literature and builds on top of them, a framework to extract useful insight from similar categories and feed this information into the existing model of a target category. This additional piece of information will help companies make smarter decisions, when analyzing the market demand for potential new products, for which no or little prior knowledge is available. The outcome will be a more intelligent system to offer new products, which would minimize development, product research, production, and marketing costs.

Even though the results here are promising, the current method still suffers from limitations. The choice of sister category in this research was done manually, using intuition and common sense. While these two are necessary in general, a mathematical formulation to evaluate eligibility of a category as a sister category would be of significant importance. While the current method is limited to ”similar” categories in the sense of actual products (e.g., yogurt and ice cream), applying a mathematical model may propose sister categories that are not intuitive, e.g., yogurt and T-shirt.

Another future venue of improvement would be to apply the methodology discussed here to real-world data. Even though the proposed method was developed with the real-world data in mind, there are simplifications and challenges that need to be addressed in a more coherent way, such as estimation of walk-aways, better estimation of assortments considering the fact that customers may miss some items on the shelf and choose their next priority products, effect of temporary promotions, etc.

# CHAPTER 4

## Conclusion and Directions for Future Research

In this dissertation we our goal was to use customer choice modeling for new product development based on product attributes. As discussed the motivation for this model comes from issues faced by real-world decision makers, and how they can benefit from using CCM models to improve their profitability and performance. First we discussed the analysis of a category of products using single category customer choice modeling. The model's primary goal is to forecast a measure of market share for existing and new products in practice. CCM is challenging in practice due to limitations around the quality of the data available for modeling and potentially complex choice behaviors. We propose a hybrid modeling approach that relies on both parametric and non-parametric methods to derive effective recommendations for product assortment planning and product-line extension. In other words, we show how historical data can be used to obtain potential niche products with promising demand in the products attributes space. Our attribute based estimates are readily interpretable and our overall approach accounts for customer heterogeneity which helps increase accuracy of predictions.

This hybrid method utilizes tow main approaches in CCM and benefits from their strengths while avoiding their shortcomings. We recommend the utilization of non-parametric choice models to first extract an accurate ranking-based product choice model from sales transactions and inventory records. The resulting model is utilized to establish customer segments and derive more actionable product attribute-based parametric models for each segment that can be employed for product assortment optimization as well as product-line extension. For example, a niche products can emerge from identifying an attribute having high level of preference while lacking enough SKUs.

Then, modeling dependence between consumers' choices across multiple categories of products can enable us to leverage from correlatedness of customers' behaviour in different categories.

Our research contributes to the literature by offering an approach of leveraging information based on correlated preferences for attributes that are shared by categories in new product development. Our method decomposes the preference for each product in each category into preferences for attributes, some of which are common across categories and at least one of them is not. Thereby we allow similarities in choice behavior across categories to be determined by correlated preferences for the attributes and make the base customer choice model better by using attribute importance from purchasing in existing categories to predict preferences for attributes in new categories.

Our proposed method uses the best practices of single-category and multi-category choice modeling literature and builds on top of them, a framework to extract useful insight from similar categories and feed this information into the existing model of a target category. This additional piece of information will help companies make smarter decisions, when analyzing the market demand for potential new products, for which no or little prior knowledge is available. The outcome will be a more intelligent system to offer new products, which would minimize development, product research, production, and marketing costs.

In closing, we emphasize that this study can be seen as a base for more comprehensive research on hybrid customer choice models. There are several avenues for future research. First and foremost, the methods should be tested across more application settings (both retail and other) for effectiveness. The employed MNL model only accounted for first order effects of attributes and did not address any potential interactions. The ability to comprehend and leverage these interaction effects and evaluation of other parametric choice models for assortment planning is also worthy of study. Also, most of the non-parametric methods for learning distributions over product 'preference lists' currently cannot scale to large categories (with hundreds of SKUs). Effective heuristic methods need to be developed. Also, one other future research area that could be worth exploring in future research is selecting an informative set of categories for predicting consumer preferences in entirely new categories. The choice of sister category in this research was done manually, using intuition and common sense. While these two are necessary in general, a mathematical formulation to evaluate eligibility of a category as a sister category would be of significant importance. While the current method is limited to ”similar” categories in the sense of actual products (e.g., yogurt and ice cream), applying a mathematical model may propose sister categories that are not intuitive, e.g., yogurt and T-shirt.

**APPENDICES**

Table .1: Customer segments, shares, and MNL model coefficients for attributes when transactions for $\bar{P} = P_5$ are removed from data.

| Segment | $\lambda^i$ | $\sum_{j=1}^{i} \lambda^j$ | MNL Coefficients $\beta_{Material}$ | $\beta_{Length}$ | $\beta_{Style}$ | $\beta_{Fashion}$ | $\beta_{Color}$ |
|---|---|---|---|---|---|---|---|
| $\sigma^1 : P_7 > P_3 > P_6 > P_1 > P_0$ | 0.21 | 0.21 | 0 | -1.13 | -0.99 | 2.24 | -1.25 |
| $\sigma^2 : P_2 > P_0$ | 0.13 | 0.34 | -3.31 | 3.5 | 3.06 | 0 | 0 |
| $\sigma^3 : P_6 > P_3 > P_1 > P_0$ | 0.11 | 0.45 | 2.46 | -1.13 | -0.99 | 0 | 0.99 |
| $\sigma^4 : P_6 > P_4 > P_3 > P_2 > P_0$ | 0.09 | 0.54 | 1.51 | -0.02 | 0.98 | -1.17 | 0 |

Table .2: Demand distributions without and with $\bar{P} = P_5$ for the top four segments.

| | Demand Distribution without $P_5$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | $P_0$ | $P_1$ | $P_2$ | $P_3$ | $P_4$ | $P_6$ | $P_7$ | |
| $\sigma^1 : P_7 > P_3 > P_6 > P_1 > P_0$ | 5.2% | 14.0% | 5.2% | 16.1% | 5.2% | 5.2% | 49.0% | |
| $\sigma^2 : P_2 > P_0$ | 1.0% | 3.8% | 82.0% | 2.5% | 3.8% | 3.0% | 3.8% | |
| $\sigma^3 : P_6 > P_3 > P_1 > P_0$ | 4.7% | 12.5% | 4.7% | 14.4% | 4.7% | 54.5% | 4.7% | |
| $\sigma^4 : P_6 > P_4 > P_3 > P_2 > P_0$ | 4.2% | 4.2% | 11.3% | 11.5% | 13.6% | 51.0% | 4.2% | |
| | Demand Distribution with $P_5$ | | | | | | | |
| | $P_0$ | $P_1$ | $P_2$ | $P_3$ | $P_4$ | $P_6$ | $P_7$ | $P_5$ |
| $\sigma^1 : P_7 > P_3 > P_6 > P_1 > P_0$ | 4.4% | 11.9% | 4.4% | 13.6% | 4.4% | 4.4% | 41.5% | 15.4% |
| $\sigma^2 : P_2 > P_0$ | 1.0% | 3.7% | 79.6% | 2.4% | 3.7% | 2.9% | 3.7% | 2.9% |
| $\sigma^3 : P_6 > P_3 > P_1 > P_0$ | 3.9% | 10.4% | 3.9% | 12.0% | 3.9% | 45.3% | 3.9% | 16.9% |
| $\sigma^4 : P_6 > P_4 > P_3 > P_2 > P_0$ | 2.8% | 2.8% | 7.5% | 7.6% | 9.0% | 33.8% | 2.8% | 33.8% |

Table .3: Customer segments, shares, and MNL model coefficients for attributes when $\bar{P} = P_7$ transactions are removed from data.

| | | | MNL Coefficients | | | | |
|---|---|---|---|---|---|---|---|
| Segment | $\lambda^i$ | $\sum_{j=1}^{i} \lambda^j$ | $\beta_{Material}$ | $\beta_{Length}$ | $\beta_{Style}$ | $\beta_{Fashion}$ | $\beta_{Color}$ |
| $\sigma^1 : P_5 > P_3 > P_1 > P_0$ | 0.23 | 0.23 | 0 | -1.12 | -0.99 | 3.23 | -2.24 |
| $\sigma^2 : P_6 > P_4 > P_3 > P_2 > P_0$ | 0.18 | 0.42 | 1.25 | -0.04 | 0.99 | -3.4 | 2.24 |
| $\sigma^3 : P_2 > P_0$ | 0.15 | 0.56 | -3.24 | 3.16 | 3.42 | 0 | 0 |
| $\sigma^4 : P_6 > P_3 > P_1 > P_0$ | 0.10 | 0.67 | 3.2 | -1.91 | -1.77 | -1.58 | 3.02 |

Table .4: Demand distributions without and with $\bar{P} = P_7$ for the top four segments.

| | Demand Distribution without $P_7$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | $P_0$ | $P_1$ | $P_2$ | $P_3$ | $P_4$ | $P_5$ | $P_6$ | |
| $\sigma^1 : P_5 > P_3 > P_1 > P_0$ | 5.2% | 14.0% | 5.2% | 16.1% | 5.2% | 5.2% | 49.0% | |
| $\sigma^2 : P_6 > P_4 > P_3 > P_2 > P_0$ | 4.8% | 4.8% | 12.8% | 13.3% | 15.1% | 4.7% | 44.6% | |
| $\sigma^3 : P_2 > P_0$ | 0.9% | 2.7% | 83.5% | 3.6% | 2.7% | 3.3% | 3.3% | |
| $\sigma^4 : P_6 > P_3 > P_1 > P_0$ | 2.8% | 14.0% | 2.4% | 16.1% | 3.3% | 2.9% | 58.5% | |
| | Demand Distribution with $P_7$ | | | | | | | |
| | $P_0$ | $P_1$ | $P_2$ | $P_3$ | $P_4$ | $P_5$ | $P_6$ | $P_7$ |
| $\sigma^1 : P_5 > P_3 > P_1 > P_0$ | 2.3% | 6.0% | 2.3% | 6.9% | 2.3% | 21.2% | 2.3% | 56.8% |
| $\sigma^2 : P_6 > P_4 > P_3 > P_2 > P_0$ | 4.7% | 4.7% | 12.7% | 13.2% | 15.1% | 4.7% | 44.4% | 0.5% |
| $\sigma^3 : P_2 > P_0$ | 0.9% | 2.7% | 81.3% | 3.5% | 2.7% | 3.2% | 3.2% | 2.7% |
| $\sigma^4 : P_6 > P_3 > P_1 > P_0$ | 2.8% | 13.9% | 2.4% | 16.0% | 3.3% | 2.8% | 58.1% | 0.7% |

## REFERENCES


[1] mckinsey article. https://www.mckinsey.com/~/media/mckinsey/dotcom/client_service/retail/articles/perspectives\%20book/book_perspectives_2013-03-04_final.ashx, 2013.

[2] Statwizards. http://www.statwizards.com/education/white-papers/discrete-choice-models, 2017.

[3] Andrew Ainslie and Peter Rossi. Similarities in choice behavior across product categories. *Marketing Science*, 17(2):91–106, 1998.

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

[5] Gerardo Berbeglia, Agusttn Garassino, and Gustavo Vulcano. A comparative empirical study of discrete choice models in retail operations. *SSRN Electronic Journal*, 01 2018.

[6] Dimitris Bertsimas and Velibor Misic. Data-driven assortment optimization. Working Paper. MIT. 2015.

[7] Dimitris Bertsimas and Velibor V. Mišić. Exact first-choice product line optimization. *Operations Research*, 67(3):651–670, 2019.

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

[9] Robert C. Blattberg, Peter Peacock, and Subrata K. Sen. Purchasing Strategies across Product Categories. *Journal of Consumer Research*, 1976.

[10] H.D. Block and Jacob Marschak. Random orderings and stochastic theories of response. (66), 1959.

[11] Dieter Brandes and Nils Brandes. In retailing, assortment is job one. https://hbr.org/2011/12/in-retailing-assortment-is-job, 2011. Accessed: 2018-06-06.

[12] by Paul R. La Monica. Cnn article. https://money.cnn.com/2018/05/18/investing/nordstrom-earnings-weak-sales/index.html, 2018.

[13] Ryan Caldbeck. The product launch fallacy of big CPGs. https://medium.com/@ryancaldbeck/the-product-launch-fallacy-of-big-cpgs-434dbd9ed9e4, 2017. Accessed: 2018-06-06.

[14] Pamela N. Danziger. Retail's bankruptcy pandemic may have peaked, but it's not over yet. https://www.forbes.com/sites/pamdanziger/2018/03/14/retails-bankruptcy-pandemic-may-have-peaked-but-its-not-over-yet/#6371ed8d1680, 2018. Accessed: 2018-06-06.

[15] Thomas A. Domencich and Daniel L. McFadden. Urban travel demand: A behavioral analysis. 1975.

[16] Imai Susumu Erdem, Tulin and Michael P. Keane. Brand and quantity choice dynamics under price uncertainty. *Quantitative Marketing and Economics*, 1(1), 2003.

[17] Tülin Erdem and Joffre Swait. Brand equity as a signaling phenomenon. *Journal of Consumer Psychology*, 7(2):131–157, 1998.

[18] Tülin Erdem, Joffre Swait, and Jordan Louviere. The impact of brand credibility on consumer price sensitivity. *International Journal of Research in Marketing*, 19:1–19, 2002.

[19] Peter S Fader and Bruce GS Hardie. Modeling consumer choice among SKUs. *Journal of marketing Research*, 33(4):442–452, 1996.

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

[21] Flávio S Fogliatto and Giovani JC da Silveira. Mass customization: A method for market segmentation and choice menu design. *International Journal of Production Economics*, 111(2):606–622, 2008.

[22] Dennis H Gensch and Wilfred W Recker. The multinomial, multiattribute logit choice model. *Journal of Marketing Research*, 16(1):124–132, 1979.

[23] Peter M Guadagni and John DC Little. A logit model of brand choice calibrated on scanner data. *Marketing science*, 2(3):203–238, 1983.

[24] Vishal Kumar Gupta, QU Ting, and Manoj Kumar Tiwari. Multi-period price optimization problem for omnichannel retailers accounting for customer heterogeneity. *International Journal of Production Economics*, 212:155–167, 2019.

[25] Lauren Hirsch. Cnbc article. https://www.cnbc.com/2019/02/28/l-brands-plans-to-close-roughly-53-victorias-secret-stores-this-year.html, 2019.

[26] Wagner A Kamakura and Gary J Russell. A probabilistic choice model for market segmentation and elasticity structure. *Journal of marketing research*, 26(4):379–390, 1989.

[27] Sang-Won Kim. The impact of customer buying behavior on the optimal allocation decisions. *International Journal of Production Economics*, 163:71–88, 2015.

[28] David Loshin and Abie Reifer. Using information to develop a culture of customer centricity. 2013.

[29] R Duncan Luce. On the possible psychophysical laws. *Psychological review*, 66(2):81, 1959.

[30] Siddharth Mahajan and Garrett Van Ryzin. Stocking retail assortments under dynamic consumer substitution. *Operations Research*, 49(3):334–351, 2001.

[31] Daniel McFadden. The measurement of urban travel demand. *Journal of public economics*, 3(4):303–328, 1974.

[32] Elham Nosrat, Ratna Babu Chinnam, and Evrim Dalkiran. Customer choice modeling in market prediction and product development. 2019.

[33] Chan Su Park and V. Srinivasan. A survey-based method for measuring and understanding brand equity and its extendibility. *Journal of Marketing Research*, 31(2):271–288, 1994.

[34] Girish N Punj and Richard Staelin. The choice process for graduate business schools. *Journal of Marketing Research*, 15(4):588–598, 1978.

[35] P.B. Seetharaman, Andrew Ainslie, and Pradeep K. Chintagunta. Investigating household state dependence effects across categories. *Journal of Marketing Research*, 36(4):488–500, 1999.

[36] Peter Shapiro. The top 6 hurdles to growing assortment-related revenue. http://www.nielsen.com/nz/en/insights/news/2018/perspectives-the-top-6-hurdles-to-growing-assortment-related-revenue.html, 2018. Accessed: 2018-06-06.

[37] Vishal P Singh, Karsten T Hansen, and Sachin Gupta. Modeling preferences for common attributes in multicategory brand choice. *Journal of Marketing Research*, 42(2):195–209, 2005.

[38] Kenneth E. Train. *Discrete Choice Methods with Simulation*. Cambridge University Press, 2 edition, 2009.

[39] Phil Wahba. Inside the fight for survival at Gap, J.Crew and Abercrombie. http://fortune.com/2015/07/24/gap-jcrew-abercrombie/, 2015. Accessed: 2018-06-06.

[40] Xun Xu, Shuo Zeng, and Yuanjie He. The influence of e-services on customer online purchasing behavior toward remanufactured products. *International Journal of Production Economics*, 187:113–125, 2017.

## ABSTRACT

## CUSTOMER CHOICE MODELING FOR RETAIL CATEGORY ASSORTMENT PLANNING AND PRODUCT-LINE EXTENSION

by

**ELHAM NOSRATMIRSHEKARLOU**

**May 2020**

**Advisors:** Dr. Ratna Babu Chinnam, Dr. Evrim Dalkiran

**Major:** Industrial Engineering

**Degree:** Doctor of Philosophy

Growing competitiveness and increasing availability of data is generating great interest in data-driven analytics across industries. One of the areas that has gained a lot of attention is Customer choice modeling, which aims to explain the choices individual customers make in choosing from a set of products based on their preferences. While effective customer choice modeling is essential to a wide variety of application domains, including retail, it is challenging in practice due to limitations around the quality of the data available for modeling and potentially complex choice behaviors. This dissertation presents a hybrid modeling approach that relies on both parametric and non-parametric methods to derive effective recommendations for product development and assortment planning. A generic non-parametric ranking-based choice model is first derived using random utility maximization to best model revealed product-level preferences from sales transactions and inventory records. The resulting product-level ranking-based choice model is utilized to establish customer segments and derive more actionable product attribute-based parametric models that can be employed for product assortment optimization as well as product-line extension. Then, in order to leverage from the correlatedness of customers' preferences toward similar attributes across multiple categories of products, we use cross category customer choice models to make the base predictions more accurate. The proposed modeling approach is validated using data from a leading global apparel retailer as well as synthetic experiments.

Key Words: customer choice modeling, customer segmentation, assortment planning, product-line extension, hybrid choice modeling, cross category choice modeling

## AUTOBIOGRAPHICAL STATEMENT

Elham Nosratmirshekarlou was born in Mashhad, Iran in 1989. She received her B.S. (2012) and M.S. (2014) degrees both in Industrial Engineering from Sharif University of Technology, Tehran, Iran. For her M.S. thesis, she worked on "Modeling Inventory-Queuing system for producer-retailers with $(r, Q)$ inventory control policy". In 2014, she joined the Department of Industrial & Systems Engineering at Wayne State University for her PhD degree under the supervision of Dr. Ratna Babu Chinnam.

Her research interests include Big Data, Choice Modeling, Predictive Analytics, and machine learning algorithms with applications in retail industries.