ASSORTMENT PLANNING FRAMEWORK WITH SUBSTITUTION AND COMPLEXITY COST

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE

OF BILKENT UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF

MASTER OF SCIENCE

IN

INDUSTRIAL ENGINEERING

By

Dilara Sönmez

August 2021

ASSORTMENT PLANNING FRAMEWORK WITH SUBSTITUTION AND COMPLEXITY COST
By Dilara Sönmez
August 2021

We certify that we have read this thesis and that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Alper Şen(Advisor)

Savaş Dayanık

Özgen Karaer

Approved for the Graduate School of Engineering and Science:

Ezhan Karaşan
Director of the Graduate School

ABSTRACT

ASSORTMENT PLANNING FRAMEWORK WITH SUBSTITUTION AND COMPLEXITY COST

Dilara Sönmez
M.S. in Industrial Engineering
Advisor: Alper Şen
August 2021

While increasing product variety may have a positive effect on market share, it may lead to difficulties in product management and forecasting, leading to increase in inventory holding costs. In addition, as a result of increased setup times, efficiency decreases. In this study, we propose two methods to help a multi-national tire manufacturer manage their assortment to find a balance between variety and sales. The first method evaluates the marginal complexity cost of a set of new products so that a data-driven decision can be made to introduce or not to introduce new products. The second method determines the set of tires to be included in the assortment that increases the total profit. To account correctly for the partial sales revenue losses due to discontinued products, we estimate the fractions of demands substituted by the products left in the assortment. Customer no-purchase information were unobserved, as well as the exact timings of stock-outs and sales up to those times. Also, market share information for a particular group of tires for the company were unavailable. Based on these incomplete data, the substitution probabilities were estimated using an iterative method. The solutions were iterated on to find the set of substitution probabilities that best fit the data. Discontinued products are expected to save from complexity costs due to production capacity losses because of frequent break times. The complexity cost is a set function that accounts for the interactions between the products during manufacturing processes, as well as the variety in the product portfolio. In order to estimate the break time savings, a machine learning model was used, and an algorithm was designed to measure the effect of a set of products being discontinued. For the periods in which the machines work in full capacity, the additional profit due to potential new sales was also considered. This value and the conversion cost saved due to discontinued products were added to the profit function. The predictions from the machine learning model and the other costs are used to formulate a large-scale assortment optimization problem with a complex objective function. The assortment problem is solved using genetic algorithm. The results show that the new assortment obtained through our analysis has between 3.8% and 15.2% less products than the initial assortment. The new assortment leads to additional profit between 0.6% and 4.6% of the company's annual income. The results also show that considering complexity costs in assortment decisions leads to substantially different assortments and additional savings in comparison to those obtained without their considerations.

Keywords: assortment planning, stock-out based substitution, complexity cost, data driven decision making.

ÖZET

ÜRÜN İKAMESİ VE ÜRETİM KARMAŞIKLIĞI MALİYETİ GÖZETEN ÜRÜN GAMI PLANLANMASI

Dilara Sönmez
Endüstri Mühendisliği, Yüksek Lisans
Tez Danışmanı: Alper Şen
Ağustos 2021

Bu çalışmada, ürün çeşitliliğini artırarak satışlar ve pazar payı artırılabilse de, bu artış ile birlikte ürün yönetimi ve satış tahminleri zorlaşmakta, envanter maliyetleri yükselmektedir. Ayrıca, artan değişim süreleri sonucu verimlilik azalmaktadır. Uluslararası faaliyet gösteren bir lastik imalatçısının ürün çeşitliliği ve satış miktarları arasındaki dengeyi sağlayan ürün gamını oluşturabilmesi için iki çözüm yaklaşımı önerilmiştir. Bu yaklaşımlardan ilki yeni ürünlerin marjinal üretim karmaşıklığı maliyetini hesaplamakta, böylece ürün gamına yeni ürünler eklenirken veriye dayalı kararlar alınmasına yardımcı olmaktadır. İkinci yaklaşım; satış gelirleri, envanter ve üretim karmaşıklığı maliyetlerinden oluşan kâr fonksiyonunu iyileştiren ürün gamını bulmaktadır. Ürün gamından çıkarılan ürünlerin yol açacağı potansiyel satış kaybının bulunabilmesi için taleplerinin ürün gamında kalan ürünler tarafından ikame edilme oranı hesaplanmaktadır. Bazı periyotlarda bazı ürünlerin stok dışı olduğu ve müşterilerin satın almama kararlarının gözlemlenmediği periyodik satış verisi kullanılmaktadır. Periyot içerisinde ürünlerin stok dışı olduğu zaman ve o zamana kadarki satışlar da gözlemlenmemektedir. Ayrıca, imalatçının lastik gruplarının pazar payları bilinmemektedir. Bu durumda ikame olasılıklarının hesaplanabilmesi için yinelemeli bir yöntem sunulmaktadır. Üretim karmaşıklığı maliyetlerinin ürün gamının fonksiyonu olarak ölçülmesi, çeşitliliğe bağlı duruş sürelerini tahmin eden bir makine öğrenmesi modeli ile sağlanmaktadır. Kapasitenin tamamının kullanıldığı periyotlarda potansiyel yeni satışlardan gelecek kâr da hesaplanmaktadır. Bu değer, tüm periyotlardaki dönüşüm maliyeti kazancı ile birlikte kâr fonksiyonuna eklenmektedir. Derin öğrenme modeli ile bulunan tahminler ve diğer maliyetler ile karışık bir amaç fonksiyonu olan, büyük ölçekli, birleşimsel bir problem elde edilmektedir. Genetik algoritma kullanılarak bulunan çözüme göre ürün sayısında %3.8 ve %15.2 arasında bir düşüş önerilmektedir. Yeni ürün gamı ile elde edilecek ilave kârın, şirketin yıllık gelirlerinin %0.6 ve %4.6'si arasında bir değer olması beklenmektedir. Ayrıca, sonuçlar karmaşıklık maliyeti gözetildiğinde, önemli ölçüde farklı ürün gamı ve kâr değerleri bulunacağını gösteriyor.

Anahtar sözcükler: ürün gamı planlama, ikame olasılığı, karmaşıklık maliyeti.

Acknowledgement

I would like to express my deepest gratitude to Alper Şen and Savaş Dayanık for their support, patience and for going above and beyond to teach me and help me do my best for this project. I feel lucky to have their invaluable guidance that helped me make important decisions in life.

I am in debt to BriSA, especially, Dr. Mustafa Tacettin, Aytaç Alkan, Hürdoğan Güneş for allowing me to participate in the project and for their invaluable comments and suggestions on my work. I also want to thank Özgen Karaer for her review and suggestions.

I want to thank my family for making my time during my master's degree easier and my friends Aleyna, Damla, Ege, Tuğba for making it much more fun. They have been the best thing this program has given me. Thank you to Giovanni for being there for me during my difficult time preceding my decision to get a master's degree.

Finally, I want to thank TUBITAK for supporting my studies as part of the 2210/A postgraduate scholarship program.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Assortment planning is a relatively new area, but there is a great deal of research about it because it implies significant improvements in decision making process of retailers. Assortment planning deals with the problem of selecting the optimal subset of products to be offered to the customer so that the retailer's profit or sales are maximized. This decision is mainly dependent on the customer choice process. Hence there is a focus on research aiming to model the stochastic customer behavior, mainly the substitution behavior. The most frequently used demand model for this purpose is the multinomial logit model (MNL) [1]. Assortment planning problems can also be jointly dealt with the problem of determining the optimal prices and inventory levels of the products in the assortment. Therefore, this area of research deals with a great variety of optimization models with different decisions to be made, objectives and constraints depending on the setting of the retailer or the market. The most frequently studied type of constraints are the capacity constraints, which are the reason why we need to select a subset of products to be included in the assortment arises. Another reason is the tradeoff between complexity costs and sales revenues, both increasing with richer assortment.

Some retailers may fail to recognize the importance of assortment planning on their profitability. Mantrala et al. [2] argue that failing to offer an assortment that is balanced in terms of variety, depth, service levels and costs, lose current and potential sales. Fisher [3] presents three real examples of poor assortment planning decisions, and discusses its consequences. In an attempt to declutter their stores, Walmart reduced their assortment size by 15% and faced an estimated $1.85 billion loss in sales [4]. After removing the low-selling 20% of its dry-grocery assortment, a grocery retailer lost 40% percent of its sales and faced bankruptcy. A home furnishing retailer localized its assortments by store which proved successful for some of its product categories. However, it failed for other categories and showed that an assortment planning formula cannot be applied without thorough assessments. H&M, a large fashion retailer has been struggling with growing unsold stock partially due to faulty assortment decisions [5]. An academic study of a Dutch supermarket chain's assortment problem resulted in an estimated increase of more than 50% in profits [6]. More success stories will be presented in Section 3.

Failing to consider the tradeoff between product variety and complexity costs also has consequences. Philips' revenues decreased by 40% due to excessive operational complexity caused by a large volume of innovations [7]. Anderson [8] analyzed three textile weaving plants of a single firm and empirically showed that there is a relationship between product variety and complexity costs. Gourville and Soman [9] provides many examples from the industry and literature that indicates that high product variety affects the customer behavior by increasing the no-purchase rate due to indecisiveness. Data from an experiment done by an online grocery store show that a decrease in product variety led to an average increase of 11% in sales [10].

Many papers in literature deal with modeling the substitution behavior of customers in relation to assortment planning. Modeling the substitution behavior in assortment planning problems is important because it significantly improves the effectiveness of retailer decisions. There are two types of customer choice models: static and dynamic. In static models, the customer buys his/her favorite product if it is in stock and does not make a purchase if it is out of stock. In dynamic models, the customer can buy another product in stock; namely, substitutes his/her favorite product with another one in stock [11]. Kök et al. [12] define two types of customer substitutions related with the method in this thesis: stock-out based and assortment-based substitution. In the stock-out based substitution, the customer cannot find his/her favorite product in stock and chooses to buy a similar product. In assortment-based substitution, a customer's favorite product is not offered in the assortment, so he/she chooses to buy a similar product. In this thesis, stock-out based substitution probabilities were found and used as estimates of assortment-based substitution probabilities.

The complexity of the assortment planning problems increases with a potential correlation between product demands included in the assortment. Demand for each product as well as a product's contribution to the cost of complexity is a function of the assortment set. Cost of complexity can be defined as the direct or indirect impact of product diversity on the costs related with many components of the supply chain. Measuring this impact is not straightforward since it involves interactions between various cost items that are difficult to foresee.

Though diversity may promote sales up to a point by reducing the risk of not meeting the demand of a diverse set of customers with different preferences, having too much of it might confuse customers while hindering an effective sales strategy. In a manufacturing environment, as the product diversity increases, complexity in process design and management also increases whereas utilization decreases due to frequent process changeovers. It becomes harder to forecast demand. It also increases overhead costs and lead times. It becomes necessary to carry more inventory. With poor management of product variety, increased costs of operation, product development, marketing, and administration and decreased customer satisfaction may undo the benefits of product variety [13].

In this thesis, we study the assortment problem of a leading multi-national tire manufacturer. The company manufactures three tire brands and has about 2000 different tires in its assortment. Scientific methods were sought by the managers to reduce the assortment variety in order to increase profits. They also wanted a data-driven method to assess the complexity cost of a new product, not only by estimating the cost implications of the new product's interaction with other products in the assortment, but also its design and material costs as they currently do.

The company believed that estimating the effect of variety was difficult because of many hard-to-predict factors. To help the company find better assortments that balance sales and complexity costs, we propose two methods. The first method calculates a new product's (or a set of new products') marginal contribution to the complexity cost of the current assortment so that decision to introduce a new product can be made based on statistical evidence. The second method finds a subset of the current assortment that improves the profit function consisting of profits from sales, savings from inventory holding and complexity cost. Our complexity model captures that synergy between products with similar attributes and helps lower complexity cost. By finding the possible set of attributes that affect the efficiency of manufacturing processes, our model can be customized to every company's data.

We defined a profit function with three components while determining the set of tires to include in the assortment. The first component is the potential sales impact of the discontinued products. We formed clusters of products and assumed that if one of the products in a cluster is discontinued, a fraction of its demand is substituted to the products in the same cluster. Our data posed two challenges: The customer no-purchase option was not observed in the sales data and because the sales records were kept monthly, the stock-out times were not observed. With these incomplete data, we estimated the substitution probabilities using an iterative method. The second component of the profit function captures the inventory costs.

The third component estimates the reduction in the complexity costs, defined as the monetary implications of the break time savings during three major manufacturing processes when a set of products are discontinued. The complexity cost was calculated as a set function so that the interactions between the products during manufacturing processes, as well as the variety in the product portfolio were considered. In order to estimate the break time savings, a machine learning model was used, and an algorithm was designed to measure the effect of the discontinuation of a set of products on break time costs.

Using these three profit function components, we formulated a large-scale assortment optimization problem with a complex objective function due to the complexity cost component calculated using predictions from the machine learning model. The genetic algorithm with the island evolution approach was used to find good solutions.

Based on different scenarios, our analysis suggests a reduction in the size of the current assortment between 3.8 and 15.2. The profits with these reductions are expected to be between 0.6 and 4.6 of the annual sales revenue of the company. We showed that the impact of complexity cost savings on profits can be as significant as the profit loss from sales, making up between 0.81 and 58.65 of the net savings, when the assortment size is reduced using our framework. This establishes the benefits of finding a balance between product variety and complexity costs for companies. Also, we propose a method to estimate substitution probabilities using a small size of periodic sales data. Since the data are periodic, time of sales transactions are unobserved which means that time of occurrence of stockouts and sales up to those times are unknown. Also, the accuracy of predictions increases as the number of data points increase but our data do not have enough number of periods that would yield a reasonable level of accuracy. We suggest methods to overcome the problem of lack of enough data by combining data from multiple product groups.

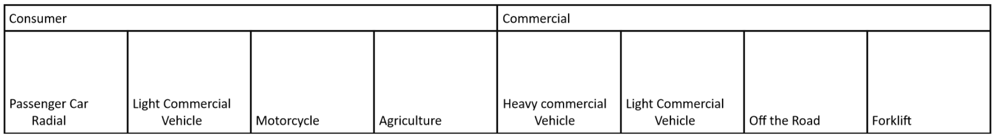
The rest of this thesis is organized as follows:

Problem is defined, and the data were described in Section 2. A literature review about customer demand models, substitution probabilities, complexity cost and assortment planning frameworks are provided in Section 3. Section 4 outlines three methods of finding the maximum likelihood estimators of substitution probabilities using incomplete data and provides a discussion on their performance under different problem settings. Section 5 details our machine learning model to quantify complexity cost as a function of the assortment. Section 6 describes our method of using the model from Section 5 to estimate the contribution of a new product to the complexity cost, as a function of the current assortment. Section 7 presents the large-scale combinatorial optimization problem incorporating the substitution probabilities estimated in Section 4 and the measure of complexity defined in Section 5 and its solution by means of genetic algorithm. A summary of results, final insights and future research directions are given in Section 8.

Chapter 2

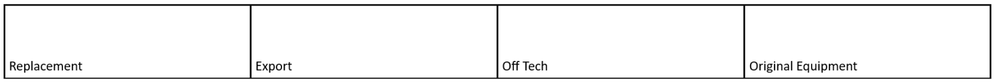
Description of the Problem and the Data

The complaint of the company is that the variety in the product portfolio creates problems in the production environment. Currently, the company holds 2116 products in its assortment. The assortment includes products from different categories and these products are sold through different channels, as shown in Figures 2.1 and 2.2. The scope of this thesis is the PSR (passenger car radial) tire category consisting of 368 products and their sales in replacement and export channels.



Consumer				Commercial			
Passenger Car Radial	Light Commercial Vehicle	Motorcycle	Agriculture	Heavy commercial Vehicle	Light Commercial Vehicle	Off the Road	Forklift

Figure 2.1: Product Categories



Replacement	Export	Off Tech	Original Equipment

Figure 2.2: Sales Channel Categories

The data on sales (monthly sales quantity, revenue, cost, sales channel, customer information), inventory level and cumulative backlogged demand for the most recent 2.5 years; product attributes and manufacturing data including production plans (monthly production quantity and time of each part produced and production machine information), bill of materials, duration of break times and reasons for breaks, monthly number of shifts available for each machine of 10 years were provided by the company. We should note that we have the data on sales to the dealers. We do not have the store level sales data. Therefore, we can only observe preferences of the dealers, not the preferences of the customers directly. We assume that the dealer preferences is a representation of customer preferences.

There are also data showing the list of products most suitable for each geographical market. The company stated that it is not possible to offer some products in some geographical markets even if our analysis might suggest that discontinuing an alternative product might result in some fraction of its demand being substituted by that product. Therefore, those data were used when a product is discontinued and there are no other products left in the assortment with similar attributes that might capture a fraction of the demand of the discontinued, product and is eligible to be offered in a certain geographical market. In such cases, substitution probability for the sales of the discontinued product to customers in that geographical market is assumed to be zero without further analysis. This was referred as *country-based business* rules in the rest of the thesis.

As shown in Table 2.1, each product has a unique product code and there are nine attributes that define a product: Season, brand, load index, speed index, size, run-flat-technology (rft), wet performance index, noise index, rolling resistance coefficient (rrc) and detailed size. There are two types of tires available regarding the season in which the tire is suitable to use: summer and winter. There are three types of brands available. Load index of a tire indicates its maximum carrying capacity, and it might range from 60 to 200. Speed index indicates the maximum speed capacity, and it might range from lowest speed E to highest speed Z. Size of a tire provides information about its width, aspect ratio, method of construction and the diameter of the wheel rim. rft indicates whether the tire has the Run-Flat Technology, which enables the driver to continue to drive safely even if the

Table 2.1: A sample of the product attributes

product code	load	speed	size	d_size	season	brand	rrc	wet performance	noise	rft
1	121/119	Q	235/65R16	design 1	summer	brand 2	C	C	71	TRUE
2	97	V	245/40R18	design 2	winter	brand 1	E	B	70	TRUE
3	94	V	225/45R17	design 3	winter	brand 1	E	B	70	TRUE
4	113/111	R	215/75R16	design 4	summer	brand 3	C	A	71	TRUE
5	110/108	R	205/75R16	design 5	summer	brand 3	C	A	71	TRUE
6	110/108	R	205/75R16	design 6	summer	brand 3	C	A	71	TRUE

tire is punctured while driving [14]. Wet performance index depends on braking distance of tires on wet roads. It might range from the highest braking distance G to the shortest braking distance A. Noise index is a measure of the exterior noise a tire generates while driving because of its contact with the road when it is in motion. rrc (rolling resistance coefficient) index is a measure of a tire's fuel efficiency, and it ranges from G (least efficient) to A (most efficient) [15]. Finally, the detailed size indicates the details about the design of the tire.

Since there are tires suitable to use in different seasons, the sales data have seasonality, and the nature of this seasonality depends on the season and brands attributes, as explored in Section 4.1.

Inventory levels and replacement balance data are used to create a stock-out criterion. Replacement balance is the cumulative backlogged demand. If the replacement balance is greater than the beginning inventory, that product was labeled as stocked-out in that month. The company experiences stock-outs frequently due to low accuracy in demand forecasting and low efficiency in manufacturing. Most machines usually work in full capacity which suggests that there is a problem with capacity management, partially due to high variety in their product portfolio.

The stock-out value of a product only tells us whether the product was out of stock at some point during the period, it does not tell us when it stocked out. As the assumption is that stocking out of a product affects the sales rate of the other product in its group, this information is crucial. Also, most consumer choice

Table 2.2: A sample of the sales data used to estimate substitution probabilities where I is the beginning inventory level, y is the sales quantity and s is a binary value indicating whether the product stocks out during the period t.

t	product_label	I	y	s
1	A	32	0	0
1	B	491	212	0
2	A	32	0	0
2	B	1860	610	0
3	A	32	0	0
3	B	1836	1539	0

models like MNL require data regarding the no-purchase option; i.e., demand of customers that chose not to buy any of the products. This is also unknown, as is the case in most real-life applications.

Chapter 3

Literature Review

Modelling customer behavior to estimate demand under different conditions is an essential part of assortment optimization. There are applications of different customer choice models used in assortment optimization problems in literature, the most common one being the Multinomial Logit model (MNL) [16]. Mixtures of MNL [17], nested logit model [18], locational choice model or an exogenous demand model were also used in successful applications. Conjoint analysis is another method used to model customer behavior [19]-[20]. See [12] for a more detailed review.

Most of the existing literature assumes no-purchase data are available, or the market share of the seller is known. However, in our problem, no-purchase decisions are not observed, and the market share of the seller is unknown. Also, data are censored due to stock-outs, meaning that the time of occurrence of stock-outs during a period and the sales up to that time are not known.

There are many papers that jointly estimate the parameters of the MNL model and the customer arrival rates with unobservable no-purchase decisions [21]-[22]. Talluri and van Ryzin [21] also deal with the problem of incomplete data; i.e., purchase transaction data without no-purchase decisions using the expectation-maximization (EM) method. In their formulation of the problem, there is at most one arrival in each period. The incomplete information that is estimated in the Expectation stage of the EM method is the number of periods without an arrival and the number of periods in which there was an arrival but no purchase has been made. They jointly estimate the parameters of the MNL model and the arrival probabilities. Newman et al. [23] address the same problem with Talluri and Van Ryzin, but aim to reduce estimation times without sacrificing from the number of parameters that correspond to utilities of the product attribute levels. Though Vulcano et al.'s [24] method does not require no-purchase data, it does require perfect knowledge of the market share. Abdallah and Vulcano [25] propose a new minorization-maximization algorithm instead of EM to maximize the incomplete log-likelihood function, that is faster and more accurate. They also outline the conditions under which the MLE estimates are unique. Li and Talluri [26] propose a two-step GMM (Generalized Method of Moments) procedure that not only addresses the issues of unobservability of no-purchase decisions, censored demand data due to stock-outs and the market share information being impossible or irrelevant to obtain but also is robust to frequent change of firms' policies as a revenue management tool, which they call "optimization-induced endogeneity".

There are some papers in literature that provide a holistic methodology that makes necessary consumer choice estimations and integrates the estimation results into a combinatorial optimization problem to find the best assortment. Fisher and Vaidyanathan [27] present a framework for the problem of finding the optimal assortment by maximizing revenue with a capacity constraint for three retail categories: snack cakes, tires and appearance chemicals. They define a product as a set of attribute levels so that the substitution probability between two products is the multiplication of the substitution probabilities between each of their attribute levels. This approach also allows the possibility of introducing new products. They use maximum likelihood estimation to find the substitution probabilities between attribute levels and the demand estimations and use hedonic regression for the pricing of new products. They use these as inputs for the optimization problem for which they outline greedy and interchange heuristic methods. The application of their framework for a tire retailer with an initial assortment of 105 SKUs resulted in the discontinuation of 24 SKUs and insertion of 11 new SKUs with 5.8% increase in revenue with the new assortment. Though this paper provides a good guideline if the attributes influence substitution behavior, there is no attempt to quantify and incorporate the cost of complexity into the optimization model. Also, managerial insight of the company's we work with was that most of the attributes of tires do not affect the purchase decision of customers; and the ones that do affect, has no substitution between them. Therefore, our method to estimate the substitution probabilities can be adopted for applications on products with no common attributes or ones whose attributes cannot be easily identified.

Yunes et al. [28] use randomized part-worth utilities to create a set of migration lists, lists that include a set of products in the decreasing order of probability of purchase, to capture the customer behavior. They use a previously calibrated piece-wise linear complexity cost function. Using these inputs, they solve a mixed-integer programming model to find the optimal assortment of configurations that reduces the number of configurations by 20% to 50% in different product lines. Their discussion regarding quantifying the complexity cost is limited.

Similar to our work, Shunko et al. [29] formulate a line reduction problem by incorporating the assortment's effect on demand and cost of complexity. To model customer preferences and substitution behaviors, they use similarity and ranking functions to obtain migration lists for each customer or each market segment, as Yunes et al. [28] did. Their contribution is the identification of the many elements of complexity cost, which were threefold: i) optional effects that include variety-based and attribute-based complexity, ii) temporal effects, iii) volume effects including the impact of production volume and number of unique configurations. They use linear regression to estimate the parameters of their cost of complexity function. At the optimization stage, they find the optimal assortment and prices by maximizing the profits from sales and decrease in complexity costs. Number of configurations decreased from 37920 to 135 and sales increased by 7% in their real-life application with Caterpillar, using this methodology. Their complexity cost approach is similar to ours since we also model optional and volume effects however, instead of a linear model, we use a more sophisticated machine learning model, and we measure complexity cost by variety-based break times.

Ward et al. [13] provide Hewlett-Packard (HP) with two tools to manage their variety in product portfolio: The first tool calculates a return-on-investment (ROI) of a new product that compares its contribution to the complexity costs and its expected sales revenues. They propose a breakup of variable and fixed costs that are possibly a result of high variety to quantify complexity costs. The second tool that is to be used after the introduction of new products, finds the optimal assortment by maximizing fraction of the revenue of its covered orders to the total revenue, with a capacity constraint. Applications in various business units of HP resulted in 40% SKU reduction in LaserJet unit, elimination of 3300 out of 10000 units in enterprise servers and $500 million increase in profits in three years [13]. Ward et al.'s two tools work interdependently since the second tool does not take account of the complexity cost. Also, there is no attempt to model substitution behavior of customers, unlike our methodology.

Chong et al. [30] model consumer behavior, purchase incidences and brand share in the context of a product category consisting of different brands and products, extending the work of Guadagni and Little [31]. They then formulate a profit maximization model with a category constraint. There is no attempt to measure the complexity cost of the assortment.

Rash and Kempf [32] offer a holistic model that combines assortment planning, product design and production scheduling in one optimization problem and a solution algorithm based on a genetic algorithm.

Chapter 4

Estimating Stock-Out Based Substitution Probabilities

The company's insight was that there is no substitution between products with different season, brand, size, speed index and load index attributes. Therefore, these products were assigned to different groups so that products in each group have common season, brand, size, speed and load attributes. 146 of such groups were obtained. 96 out of 146 groups had two members. The focus of this chapter will be those 96 groups.

The aim of this chapter is to estimate the stock-out based substitution probabilities in between members of a group so that the expected decrease in sales after discontinuing a product can be calculated. It was assumed that the demand of a product can only be substituted by another product in its group. Two products in a group were labeled as A and B, and the substitution rate from stocked-out product j to in-stock product i was denoted by $p_{i\bar{j}}$, where $(i, j) \in \{(A, B), (B, A)\}$.

We identified 87 products that can potentially be discontinued. We observed that 62 of them belongs to a group with two members. These are the groups whose substitution probabilities are actually needed. However, we propose two methods in the following sections that use the data from other groups with some common attributes to estimate the substitution probabilities in between members of a group.

While identifying the 87 products that can potentially be discontinued, a systematic approach was employed. The entropy of sales quantity distribution of all products within each group with common season, brand, size, speed, load attributes were calculated. Large entropy indicates severe imbalance in sales quantities of different products, which we have as a proxy for production quantities. Groups with large production quantity imbalance are undesirable because of frequent setup or break times. Therefore, the aim was to find the groups of products with the same season, brand, size, speed, load attributes, but having imbalanced sales quantities in 2019. We used the normalized entropy to find the product groups with the most imbalanced sales quantities in 2019 because entropies of groups with different number of products are not on the same scale. The entropy of a probability distribution $(p_i, i = 1, \ldots, n)$ is defined

$$\text{Entropy}(p_i; i = 1, \ldots, n) = \sum_{i=1}^{n} p_i \log \frac{1}{p_i} = E \log \frac{1}{p_N},$$

where N is a random variable taking values $1, \ldots, n$ with probabilities $p_1, \ldots, p_n$. By Jensen's inequality

$$\text{Entropy}(p_i; i = 1, \ldots, n) = E \log \frac{1}{p_N} \leq \log E \frac{1}{p_N} = \log \sum_{i=1}^{n} p_i \frac{1}{p_i} = \log n.$$

The inequality holds tightly for $\overline{p}_i = 1/n, i = 1, \ldots, n$. Indeed,

$$\text{Entropy}(\overline{p}_i; i = 1, \ldots, n) = \sum_{i=1}^{n} \frac{1}{n} \log n = \log n.$$

Namely, the entropy takes its largest value for the uniform distribution and decreases as the uncertainty represented by the distribution decays. The smaller the entropy is, the more distant is the distribution from the uniform distribution.

In our study, the production shares of individual products in the same segment is a probability distribution. We are interested in product groups whose production share distributions have smallest entropies. However, the maximum entropy of a distribution on n distinct mass points grows with n. Therefore, the entropies of sales share distributions of product groups with different number of products are not directly comparable. As a remedy, we define the normalized entropy as the entropy of a product group divided by the maximum entropy that any group with the same product number can have, which is the log of the product number:

$$\begin{aligned}\text{Normalized entropy}(p_i; i = 1, \ldots, n) &= \frac{\text{Entropy}(p_i; i = 1, \ldots, n)}{\text{Entropy}(\overline{p}_i; i = 1, \ldots, n)} \\ &= \frac{\text{Entropy}(p_i; i = 1, \ldots, n)}{\log n}\end{aligned}$$

We use normalized entropy to compare the sales quantity imbalances in product groups with different number of products. Note that this metric will also be used in Section 5 to relate variety-based break times by means of a machine learning model to production quantity or time imbalances experienced on machines. We expect that in the groups with the highest imbalanced sales quantities, eliminating the low selling products will decrease complexity costs. We also use the order of release date of the lowest selling product in a group as another metric. If a group has high entropy and if the low selling product in that group is also the product with the earliest release date, this is an indication that the product has started being replaced by other products in its group with more up-to-date technologies. After taking into account the insights from company's many departments, such products are expected to have high substitution probabilities and are labelled as products that can be discontinued. However, this approach only helps us find a short list of products that have the highest probability to be discontinued. The final decision of discontinuation will be determined after we can acknowledge with a separate statistical analysis that the unmet demand will be substituted to a satisfactorily large extend by the products left in the same pool. We undertake that statistical analysis and estimate the substitution probabilities in this section. Having such a short list also reduced the complexity of the optimization problem that is given in Section 7.

Anupindi et al. [33] propose a customer choice model to estimate stock-out based substitution patterns between items in a category, by modeling the sales in different stock-out cases as homogeneous Poisson processes and estimating their rates. They use EM method to deal with unobserved stock-out times and to find the maximum likelihood estimates of sales rates which allow for the calculation of substitution probability estimates. Our method of estimating substitution probabilities is based on their work since two problems are very similar in the sense that there is no data regarding the no-purchase decisions or the market share and that the sales data are censored. In Section 4.2, we explain the method of Anupindi et al. and provide the derivations of their maximum likelihood problem. In Sections 4.3 and 4.4, we propose two methods to use the Anupindi model with limited data by making assumptions that allow us to combine the sales data of multiple product groups to find a common substitution probability between the old and new products in each group. Since the estimation methods proposed in these sections assume stationary demand of products, estimating the seasonality effect in sales data of all tires at all periods and deseasonalizing the data was a necessary step before estimating the substitution probabilities. Therefore, a method to deseasonalize the sales data was explained in Section 4.1.

4.1 Deseasonalizing the Sales Data

Since consumers demand tires with different attributes on different seasons, the sales data had seasonality. After comparing many models using different combinations of possible explanatory variables, our conclusion was that this seasonality effect depends not only on the month and year of sales, but also on season and brand attributes of the products.

Let B, S, Y and M be the set of all brands, seasons, years and months present in the data, indexed by b, s, y and m. Let w_{ym} be the number of working days in year m, month m. Sum of the sales of all products with brand b and season s in year y and month m is shown as y_{bsym}.

Firstly, the monthly sales data were normalized based on the number of working days in a month as $z_{bsym} = y_{bsym}\frac{22}{w_{ym}}$. Then, a linear regression model was fit where the response variable is the log normalized sales z_{bsym}.

Table 4.1: Definitions of variables used in Section 4.1

Variable	Definition
w_{ym}	Number of working days in year m, month m
y_{bsym}	Sum of the sales of all products with brand b and season s in year y and month m
z_{bsym}	Normalized sum of the sales of all products with brand b and season s in year y and month m
β_i^I	Regression coefficient of the effect of i
d_{bsym}	Deseasonalization factor of a product with brand b, season s at year y, month m

Let β_i^I be the regression coefficient of the effect of i where i can be brand, season, month, year or the interaction of month and year with brand and season. The log normalized sale in season s for brand b in year y and in month m is modelled as normally distributed with variance σ^2 and mean $\beta_0 + \beta_b^B + \beta_s^S + \beta_y^Y + \beta_m^M + \beta_{b:s}^{B:S} + \beta_{b:y}^{B:Y} + \beta_{b:m}^{B:M} + \beta_{s:y}^{S:Y} + \beta_{s:m}^{S:M} + \beta_{b:s:y}^{B:S:Y} + \beta_{b:s:m}^{B:S:M}$.

In this model, season and brand as well as year and month interact on their effects on log sales, but year and month do not interact.

The vertical axis and horizontal axes of the plot in Figure 4.1 show the observed and fitted log normalized sales, respectively. The points clustering about the identity line suggest that the model fits the data well. Also, variance is not quite constant which suggests that there is no room for improvement.

The value $\widehat{\log z_{bsym}} - \beta_0 = \beta_b^B + \beta_s^S + \beta_y^Y + \beta_m^M + \beta_{b:s}^{B:S} + \beta_{b:y}^{B:Y} + \beta_{b:m}^{B:M} + \beta_{s:y}^{S:Y} + \beta_{s:m}^{S:M} + \beta_{b:s:y}^{B:S:Y} + \beta_{b:s:m}^{B:S:M}$ is the measure of additional sales in log scale relative to the grand mean. This value was normalized again with respect to the first month of the first year present in the data, which is January 2018.

The deseasonalization factor of a product with brand b, season s at year y, month m is

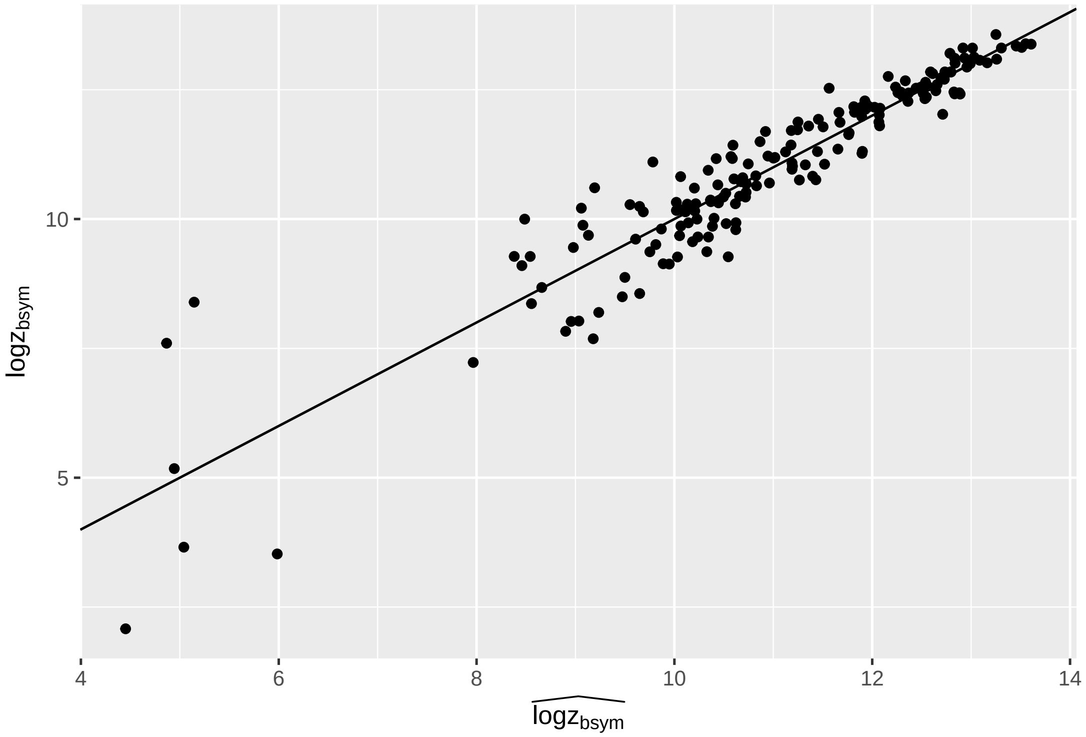


Figure 4.1: Plot of the fitted vs. observed values of the log normalized sales

$$d_{bsym} = \frac{1}{\frac{e^{\widehat{\log z_{bsym}} - \beta_0}}{e^{\log \hat{z}_{bs\ \ 2018\ \ Jan} - \beta_0}}} \times \frac{22}{w_{ym}}.$$

We find y_i^t, deseasonalized sales of a product i in period t by multiplying the observed sales by the d_{bsym}, where b and s are the brand and season attributes of product i, and y and m are the year and month corresponding to period t. These deseasonalized sales values were used to estimate substitution probabilities with the methods that will be explained in the following sections.

4.2 λ-Method: Estimating Sales Rates

Anupindi et al. [33] model a customer choice process where customers occasionally face stock-outs and can make substitutions depending on the availability of their first choice. However, the first choice of the customers cannot be observed. The only information available is the total sales of products in the same category and their beginning inventory levels. Therefore, the sales of a product is known, but the demand or the time of each sales transaction is unknown. In other words, if a product stocks out during a period, the exact time of stock-out and sales up to that point are unknown. In a group of two products, A and B, there are four possible stock-out cases: both products are in stock; A is in stock, B stocks out; B is in stock, A stocks out, and both products stock out. These cases will be referred to as case 1, case 2, case 3 and case 4, respectively. Anupindi et al. assume that sales rate of each product depends on whether the other product in its group is in stock by modelling the sales process of products in different stock-out cases as independent homogeneous Poisson processes. They estimate the parameters of these Poisson processes, and it becomes straightforward to find the substitution probabilities using these sales rates. Let $\lambda_{A\bar{B}}$ be the sales rates of product A when B is out of stock and let λ_A and λ_B be the sales rates of products A and B when both are in stock. Then, we find the fraction of demand that can be substituted by product A when B is out of stock, $p_{A\bar{B}}$, using (4.1).

$$p_{A\bar{B}} = \frac{\lambda_{A\bar{B}} - \lambda_A}{\lambda_B} \tag{4.1}$$

The method will be described focusing on its application to one product group. However, this method should be applied to each product group separately using its own data.

4.2.1 Derivations for the λ-Method

Let D, E, F, G be the set of period indices corresponding to case 1, case 2, case 3 and case 4 observed in the data respectively, indexed by d, e, f and g. Superscripts of the variables used in the derivations indicate the stock-out case that the variable is related to. L_c^t is the likelihood function corresponding to case c and period t, where $c \in \{1, 2, 3, 4\}$ and $t \in \{1, \ldots, |D| + |E| + |F| + |G|\}$. The Poisson random

Table 4.2: Definitions of variables used in Chapter 4.2

Variable	Definition
Sets	
D	Set of period indices corresponding to case 1
E	Set of period indices corresponding to case 2
F	Set of period indices corresponding to case 3
G	Set of period indices corresponding to case 4
Variables	
$p_{i\bar{j}}$	Substitution rate from stocked-out product j to in-stock product i
L_c^t	Likelihood function corresponding to case c and period t
ℓ_c^t	Log-likelihood function corresponding to case c and period t
Y_i	Poisson random variable corresponding to the sales of product i when both products are in stock
λ_i	Rate of Y_i
$Y_{i\bar{j}}$	Poisson random variable corresponding to the sales of of product i when product j is out of stock
$\lambda_{i\bar{j}}$	Rate of $Y_{i\bar{j}}$
y_i^t	Observed sales of product i in period t
T_i^t	Time of stock-out of product i in period t

variable corresponding to the sales of product i when both products are in stock was denoted by Y_i, and its rate is λ_i, where $i \in \{A, B\}$. The Poisson random variable corresponding to the sales of of product i when product j is out of stock was denoted by $Y_{i\bar{j}}$ and its rate is $\lambda_{i\bar{j}}$, where $(i, j) \in \{(A, B), (B, A)\}$.

Derivations for cases 2 and 3 are symmetrical, therefore only the derivations for case 2 will be shown.

Case 1 : Both products are in stock

Since both products are in stock; i.e., the demand is less than the inventory available for both products, sales rates when both products are in stock, (λ_A and λ_B) will be used for the whole period. This also means that the demands Y_A and Y_B, which have Poisson distributions, are equal to the observed sales y_A^d and y_B^d. The length of a period was assumed to be 1, which can be scaled up or down to any value.

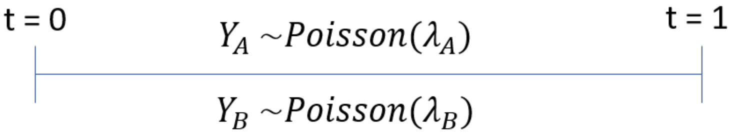


Figure 4.2: Description of Case 1

The contribution to likelihood and log-likelihood functions, respectively, become

$$
\begin{aligned}
L_1^d &= P\{Y_A = y_A^d, Y_B = y_B^d\} \propto \frac{\lambda_A^{y_A^d} e^{-\lambda_A}}{y_A^d!} \frac{\lambda_B^{y_B^d} e^{-\lambda_B}}{y_B^d!}, \\
\ell_1^d &= \log L_1^d = y_A^d \log \lambda_A - \lambda_A + y_B^d \log(\lambda_B) - \lambda_B.
\end{aligned}
$$

Note that the division by $y_A^d! y_B^d!$ in the likelihood function can be omitted in the maximum log-likelihood problem since that term becomes a constant in ℓ_1^d.

Case 2: A is in stock, B stocks out

In this case, demand of product B in period e is greater than its available inventory. This also means that the demand of product B, Y_B is greater than or equal to the observed sales of product B, y_B^e, whereas the demand of product A, Y_A equals its observed sales y_A^e. Let $T_B^e \sim \text{Gamma}(y_B^e, \lambda_B)$ be the time of stock-out of product B in period e, $e \in E$. Before $t = T_B^e$, sales rates when both products are in stock, λ_A and λ_B are used. After $t = T_B^e$, sales of product B is zero and the sales rate of product A when B is out of stock, $\lambda_{A\bar{B}}$ is used for product A.

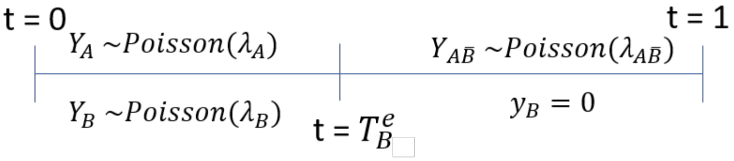

t = 0
$Y_A \sim Poisson(\lambda_A)$
t = 1
$Y_{A\bar{B}} \sim Poisson(\lambda_{A\bar{B}})$
$Y_B \sim Poisson(\lambda_B)$
$t = T_B^e$
$y_B = 0$

Figure 4.3: Description of Case 2

The contribution to the likelihood function becomes

$$L_2^e = P\{Y_B \geq y_B^e, Y_A = y_A^e\} = P\{T_B^e < 1, Y_A(T_B^e) + Y_A(1) - Y_A(T_B^e) = y_A^e\}.$$

Note that, given T_B^e, the random variables $Y_A(T_B^e)$ and $Y_A(1) - Y_A(T_B^e)$ are conditionally independent and have Poisson distribution with means $\lambda_A T_B^e$ and $\lambda_A(1 - T_B^e)$, respectively. Therefore, given T_B^e, the random variable $Y_A = Y_A(T_{B^e}) + Y_A(1) - Y_A(T_{B^e})$ has Poisson distribution with mean $\lambda_A T_B^e + \lambda_{A\bar{B}}(1 - T_B^e)$ on the event $\{T_B^e < 1\}$.

Since $T_B^e \sim \text{Gamma}(y_B^e, \lambda_B)$; i.e., $P\{T_B^e \in d\tau\} = \frac{\lambda_B^{y_B^e} \tau^{y_B^e - 1}}{(y_B^e - 1)!} e^{-\lambda_B \tau} d\tau$, the contribution to likelihood function is

$$\begin{aligned}
L_2^e &= \int_0^1 P\{T_B^e \in d\tau, Y_A(T_B^e) + Y_A(1) - Y_A(T_B^e) = y_A^e\} \\
&= \int_0^1 P\{Y_A(T_B^e) + Y_A(1) - Y_A(T_B^e) = y_A^e \mid T_B^e \in d\tau\} P\{T_B^e \in d\tau\} \\
&= \int_0^1 P\{Y_A(T_B^e) + Y_A(1) - Y_A(T_B^e) = y_A^e\} P\{T_B^e \in d\tau\} \\
&= \int_0^1 \frac{[\lambda_A \tau + \lambda_{A\bar{B}}(1-\tau)]^{y_A^e}}{y_A^e!} e^{-[\lambda_A \tau + \lambda_{A\bar{B}}(1-\tau)]} \frac{\lambda_B^{y_B^e} \tau^{y_B^e - 1}}{(y_B^e - 1)!} e^{-\lambda_B \tau} d\tau \\
&\propto \int_0^1 [\lambda_{A\bar{B}} + (\lambda_A - \lambda_{A\bar{B}})\tau]^{y_A^e} \lambda_B^{y_B^e} \tau^{y_B^e - 1} e^{-[\lambda_{A\bar{B}} + (\lambda_A + \lambda_B - \lambda_{A\bar{B}})\tau]} d\tau \\
&= \int_0^1 e^{y_A^e \log[\lambda_{A\bar{B}} + (\lambda_A - \lambda_{A\bar{B}})\tau] + y_B^e \log \lambda_B + (y_B^e - 1)\log \tau - \lambda_{A\bar{B}} - [\lambda_{A\bar{B}} + (\lambda_A + \lambda_B - \lambda_{A\bar{B}})\tau]} d\tau \\
&= \int_0^1 e^{f(\tau)} d\tau, \quad \text{where} \\
f(\tau) &= y_A^e \log[\lambda_{A\bar{B}} + (\lambda_A - \lambda_{A\bar{B}})\tau] + y_B^e \log \lambda_B + (y_B^e - 1)\log \tau - \lambda_{A\bar{B}} \\
&\quad - [\lambda_{A\bar{B}} + (\lambda_A + \lambda_B - \lambda_{A\bar{B}})\tau]
\end{aligned}$$

for which we define

$$M_2 := \sup_{\tau \in (0,1]} f(\tau)$$

and write

$$L_2^e \propto \int_0^1 e^{M_2-(M_2-f(\tau))}d\tau = e^{M_2}\int_0^1 e^{-(M_2-f(\tau))}d\tau.$$

Therefore, the contribution to log-likelihood function becomes

$$\ell_2^e = M_2 - \log\int_0^1 e^{-(M_2-f(\tau))}d\tau.$$

Case 4: Both products stock out

In this case, demand of each product in period g is greater than its available inventory. Let $T_A^g \sim \text{Gamma}(y_A^g, \lambda_A)$ be the time of stock out of product A in period $g \in G$. Let $T_B^g \sim \text{Gamma}(y_B^g, \lambda_B)$ be the time of stock out of product B in period $g \in G$. Depending on which product stocks out before, there are two depictions of this case in Figures 4.4 and 4.5. Suppose that product A stocks out before product B. Before $t = T_A^g$, both products are in stock, and A and B sell at λ_A and λ_B rates. After $t = T_A^g$, sales of product A is zero and product B sells at rate $\lambda_{B\bar{A}}$. After $t = T_B^g$, sales of both products are zero.

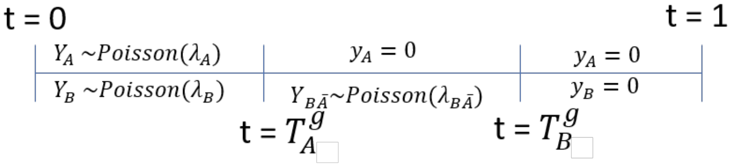

t = 0
t = 1
$Y_A \sim Poisson(\lambda_A)$
$y_A = 0$
$y_A = 0$
$Y_B \sim Poisson(\lambda_B)$
$Y_{B\bar{A}} \sim Poisson(\lambda_{B\bar{A}})$
$y_B = 0$
t = T_A^g
t = T_B^g

Figure 4.4: Description of case 4 when product A stocks out before product B

The contribution to likelihood function becomes

$$L_4^g := P\{T_A^g < 1,\ Y_B(T_A^g) < y_B^g,\ Y_B(1) \geq y_B^g\} + P\{T_B^g < 1,\ Y_A(T_B^g) < y_B^g,\ Y_A(1) \geq y_B^g\}.$$

$$\begin{aligned}L_{41}^g &:= P\{T_A^g < 1,\ Y_B(T_A^g) < y_B^g,\ Y_B(1) \geq y_B^g\}\\ &= P\{T_A^g < 1, Y_B(1) \geq y_B^g\} - P\{T_A^g < 1, Y_B(T_A^g) \geq y_B^g, Y_B(1) \geq y_B^g\}\end{aligned}$$

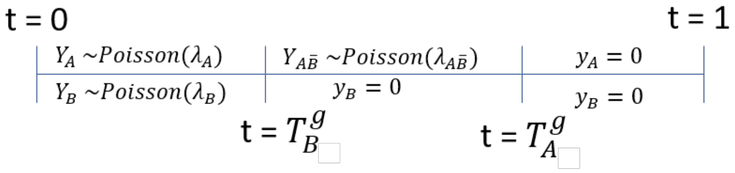

t = 0
t = 1
$Y_A \sim Poisson(\lambda_A)$
$Y_{A\bar{B}} \sim Poisson(\lambda_{A\bar{B}})$
$y_A = 0$
$Y_B \sim Poisson(\lambda_B)$
$y_B = 0$
$y_B = 0$
t = T_B^g
t = T_A^g

Figure 4.5: Description of case 4 when product B stocks out before product A

Since $y_B^g \leq Y_B(T_A^g) \leq Y_B(1)$ on $\{T_A^g < 1\}$, we have

$$\begin{aligned}
L_{41}^g &= P\{T_A^g < 1, Y_B(1) \geq y_B^g\} - P\{T_A^g < 1, Y_B(T_A^g) \geq y_B^g\}, \quad \text{and} \\
L_{411}^g &:= P\{T_A^g < 1, Y_B(1) \geq y_B^g\} \\
&= \int_0^1 P\{Y_B(T_A^g) + Y_B(1) - Y_B(T_A^g) \geq y_B^g \mid T_A^g = \tau_A\} P\{T_A^g \in d\tau_A\}
\end{aligned}$$

where, given T_A^g, the sum $Y_B(T_A^g) + Y_B(1) - Y_B(T_A^g)$ has conditionally Poisson distribution with mean $\lambda_B T_A^g + \lambda_{B\bar{A}}(1 - T_A^g)$. Therefore,

$$L_{411}^g = \int_0^1 \frac{[\lambda_B \tau_A + \lambda_{B\bar{A}}(1-\tau_A)]^{y_B^g}}{y_B^g} e^{-[\lambda_B \tau_A + \lambda_{B\bar{A}}(1-\tau_A)]} \, \frac{\lambda_A^{y_A^g} \tau^{y_A^g - 1}}{(y_A^g - 1)!} e^{-\lambda_A \tau_A} d\tau_A.$$

Similarly,

$$\begin{aligned}
L_{412}^g &:= P\{T_A^g < 1, Y_B(T_A^g) \geq y_B^g\} \\
&= \int_0^1 P\{Y_B(\tau_A) \geq y_B^g \mid T_A^g = \tau_A\} P\{T_A^g \in d\tau_A\}.
\end{aligned}$$

Since $Y_B(\tau_A) \sim Poisson(\lambda_B \tau_A)$, we have

$$L_{412}^g = \int_0^1 \frac{(\lambda_B \tau_A)^{y_B^g}}{y_B^g!} e^{-(\lambda_B \tau_A)} \, \frac{\lambda_A^{y_A^g} \tau_A^{y_A^g - 1}}{(y_A^g - 1)!} e^{-\lambda_A \tau_A} d\tau_A.$$

For simplicity in notation, we define

$$
\begin{aligned}
f_{41}(\tau_A) &:= \frac{[\lambda_B\tau_A + \lambda_{B\bar{A}}(1-\tau_A)]^{y_B^g}}{y_B^g} e^{-[\lambda_B\tau_A+\lambda_{B\bar{A}}(1-\tau_A)]}, \\
g_{41}(\tau_A) &:= \frac{(\lambda_B\tau_A)^{y_B^g}}{y_B^g!} e^{-(\lambda_B\tau_A)}, \\
h_{41}(\tau_A) &:= \frac{\lambda_A^{y_A^g}\tau_A^{y_A^g-1}}{(y_A^g-1)!} e^{-\lambda_A\tau_A}.
\end{aligned}
$$

Using the derivations for L_{411} and L_{412}, we have

$$
\begin{aligned}
L_{41}^g &= L_{411}^g - L_{412}^g \\
&= \int_0^1 [f_{41}(\tau_A) - g_{41}(\tau_A)]h_{41}(\tau_A)d\tau_A \\
&= \int_0^1 f_{41}(\tau_A)\Big[1 - \frac{g_{41}(\tau_A)}{f_{41}(\tau_A)}\Big]h_{41}(\tau_A)d\tau_A \\
&= \int_0^1 e^{\log f_{41}(\tau_A)+\log[1-\frac{g_{41}(\tau_A)}{f_{41}(\tau_A)}]+\log h_{41}(\tau_A)} d\tau_A
\end{aligned}
$$

where $0 < g_{41}(\tau_A) < f_{41}(\tau_A) < 1, \quad \forall\tau_A \in (0,1)$. For a more numerically stable calculation of L_{41}, we define M_{41}.

$$
\begin{aligned}
M_{41} &:= \max_{\tau_A\in[0,1]} \Big\{ \log f_{41}(\tau_A) + \log\Big[1 - \frac{g_{41}(\tau_A)}{f_{41}(\tau_A)}\Big] + \log h_{41}(\tau_A) \Big\} \\
\ell_{41}^g &= \log L_{41}^g = M_{41} + \log \int_0^1 e^{\log f_{41}(\tau_A)+\log[1-\frac{g_{41}(\tau_A)}{f_{41}(\tau_A)}]+\log h_{41}(\tau_A)-M_{41}} d\tau_A
\end{aligned}
$$

Calculation of L_{42}^g is symmetric with the calculation of L_{41}^g. Using the derivations for L_{41} and L_{42}, we have

$$
\begin{aligned}
L_4^g &= L_{41}^g + L_{42}^g \\
&= M_{41} + \log \int_0^1 e^{\log f_{41}(\tau_A) + \log[1 - \frac{g_{41}(\tau_A)}{f_{41}(\tau_A)}] + \log h_{41}(\tau_A) - M_{41}} d\tau_A \\
&+ M_{42} + \log \int_0^1 e^{\log f_{42}(\tau_A) + \log[1 - \frac{g_{42}(\tau_A)}{f_{42}(\tau_A)}] + \log h_{42}(\tau_A) - M_{42}} d\tau_A \\
, \ell_4^g &= \log(L_{41}^g + L_{42}^g) = \log(e^{\ell_{41}^g} + e^{\ell_{42}^g}) = \max(\ell_{41}^g, \ell_{42}^g) \log\left(1 + e^{-|\ell_{41}^g - \ell_{42}^g|}\right).
\end{aligned}
$$

Bringing the contributions of all likelihood together gives the likelihood function as in

$$
L = \prod_{d=1}^{D} L_1^d \prod_{e=1}^{E} L_2^e \prod_{f=1}^{F} L_3^f \prod_{g=1}^{G} L_4^g,
$$

and likelihood as in

$$
\ell = \sum_{d=1}^{D} \ell_1^d + \sum_{e=1}^{E} \ell_2^e \sum_{f=1}^{F} \ell_3^f + \sum_{g=1}^{G} \ell_4^g.
$$

The maximization problem to find the MLE estimators of the sales rates becomes

$$
\begin{aligned}
\max{}_{\lambda_A, \lambda_B, \lambda_{A\bar{B}}, \lambda_{B\bar{A}}} \quad & \ell(\lambda_A, \lambda_B, \lambda_{A\bar{B}}, \lambda_{B\bar{A}}) \\
\text{s.t.} \quad & 0 < \lambda_A \leq \lambda_{A\bar{B}} \leq \lambda_A + \lambda_B \\
& 0 < \lambda_B \leq \lambda_{B\bar{A}} \leq \lambda_A + \lambda_B.
\end{aligned}
\tag{4.2}
$$

After estimating the sales rates, substitution probabilities are given by

$$
p_{A\bar{B}=\frac{\lambda_{A\bar{B}}-\lambda_A}{\lambda_B}} \quad \text{and} \quad p_{B\bar{A}=\frac{\lambda_{B\bar{A}}-\lambda_B}{\lambda_A}}.
$$

Table 4.3: Definitions of variables used in Section 4.2.2

Variable	Definition
I_i	Beginning inventory of product i
$S_i = \begin{cases} 0, & \text{if product } i \text{ stocks out,} \\ 1, & \text{otherwise.} \end{cases}$	Binary variables showing the stock-out cases
L_i	Quantity of unmet demand of product i
$X_{ik} \sim Exponential(\lambda_i)$	Interarrival time of the k^{th} arrival for product i
M	A real number large enough that $\sum_{k=1}^{M} X_{Ak} \geq 1$

4.2.2 Simulation

In order to test the accuracy of this method and, others that follow, and to determine the amount of data required for accurate estimates, we use a simulation approach. This simulation method is the base method used by all other simulations methods that will be described in the following sections to test all three methods, two of which will be described later.

Beginning inventories of the products A and B, denoted by I_A and I_B, and order arrival rates λ_A, λ_B, $\lambda_{A\bar{B}}$ and $\lambda_{B\bar{A}}$ are the required inputs for this method. Sales quantities y_A, y_B, binary variables showing the stock-out cases S_A, S_B and the number of unmet demands L_A, L_B to be used in the fill rate calculations are the outputs generated for one period. For notational convenience, the period subscript t for all variables were suppressed.

A pseudocode for this method is provided below.

```
        $T_A := \sum_{k=1}^{I_A} X_{Ak}, \quad T_B := \sum_{k=1}^{I_B} X_{Bk}$
        $T_1 := \min\{T_A, T_B, 1\}$
        $z_A := \min\Big\{j \in \{1, \dots, I_A\} \mid \sum_{k=1}^{j} X_{Ak} < T_1\Big\}$
        $z_B := \min\Big\{j \in \{1, \dots, I_B\} \mid \sum_{k=1}^{j} X_{Bk} < T_1\Big\}$
    If  $T_A < T_B$,
        $S_A := 1$
        $w_B = Exp(\lambda_{B\bar{A}}(1 - T_1))$
        $y_B := \min\{z_B + w_B, I_B\}$
        $y_A := z_A$
        $S_B := 1_{z_B + w_B > I_B}$
        $L_A := min\Big\{j \in \{1, \dots, M\} \mid \sum_{k=1}^{j} X_{Ak} < 1\Big\} - y_A$
        $L_B := (z_B + w_B - I_B)S_B$
    else if  $T_B < T_A$
        Symmetrical with the case $T_A < T_B$
    else
        $S_A := 0, S_B := 0$
        $L_A := 0, L_B := 0$
        $y_A := z_A, y_B := z_B$
```

4.2.2.1 Simulation for a Single Group

λ-method was designed to be applied using the data of one group to find that group's substitution probabilities. Since we have multiple groups, λ-method should be applied to each group separately. The simulation method in this section was used to test the performance of λ-method when applied to a single group,

Table 4.4: Definitions of the parameters used in Section 4.2.2.1

Variable	Definition
N_{00}	Number of periods in which both products are in stock
N_{10}	Number of periods in which only product A is out of stock
N_{01}	Number of periods in which only product B is out of stock
N_{11}	Number of periods in which both products are out of stock
(l, u)	Multipliers determining the ratio of sales rate to the lower and upper limits of the distribution of beginning inventories of both products
$p_{i\bar{j}}$	True substitution rate from product j to product i

unlike the one explained in Section 4.2.2.2 below, which combines the data of multiple groups to estimate common substitution probabilities.

It was discovered that the number of periods with stock-outs and the average fill rate affect the accuracy of the sales rate estimations. In order to control these values, the parameters of the simulation were defined in Table 4.4

We draw $p_{i\bar{j}}$ values from a uniform distribution between 0 and 1. Inputs for the simulation method described in Section 4.2.2, which were used to generate necessary data for one period and one product group, were generated as follows:

$$\begin{aligned}
I_i &\sim \text{Unif}(\lambda_i l, \lambda_i u), \quad i \in \{A, B\}, \\
\lambda_i &\sim \text{Unif}(100, 10000), \quad i \in \{A, B\}, \\
\lambda_{i\bar{j}} &= \lambda_i + \lambda_j \times p_{i\bar{j}}, \quad i \in \{A, B\} \quad j \in \{A, B\}.
\end{aligned}$$

To show that the average fill rate affects the accuracy of the results as a justification for keeping it as a parameter of the simulation, thirty experiments with fill rates ranging from 0.8 to 0.99 for each different N_{01} value were performed, and the results were summarized in Figure 4.7. As fill rate increases, stock-outs become rare, and without stock-out data, it is harder to learn substitution probabilities. Therefore, it can be observed that the absolute errors increase as fill rate increases.

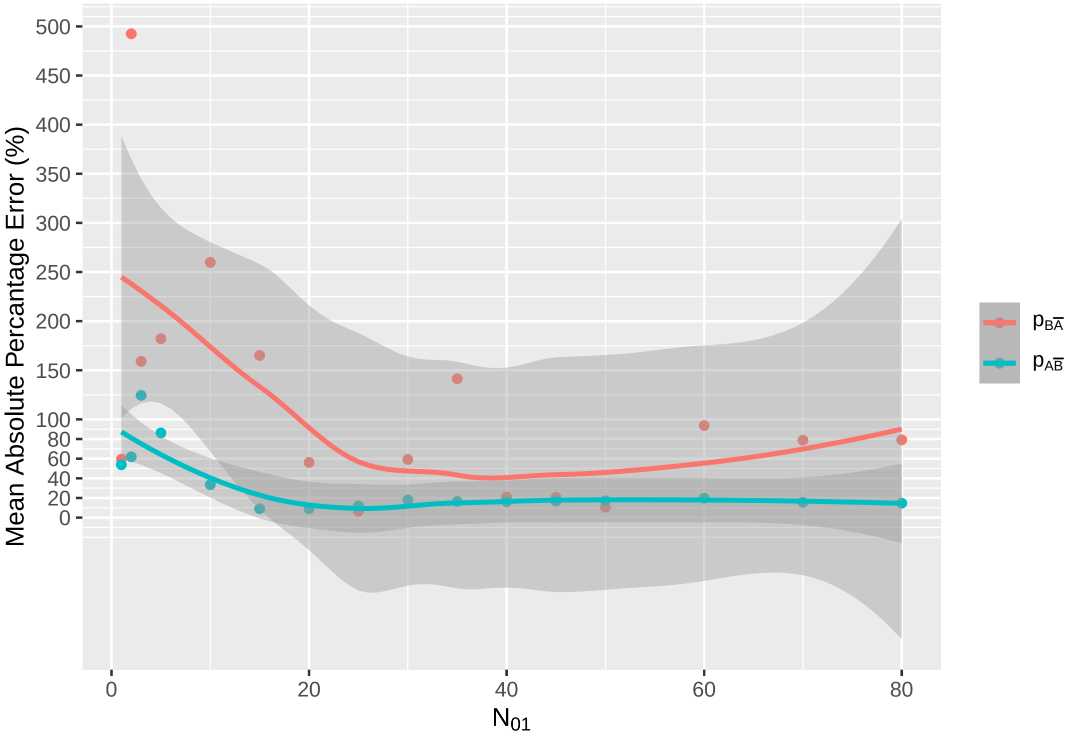


Figure 4.6: Mean absolute percantage errors of substitution probabilities when λ-method was used with varying number of stock-out observations

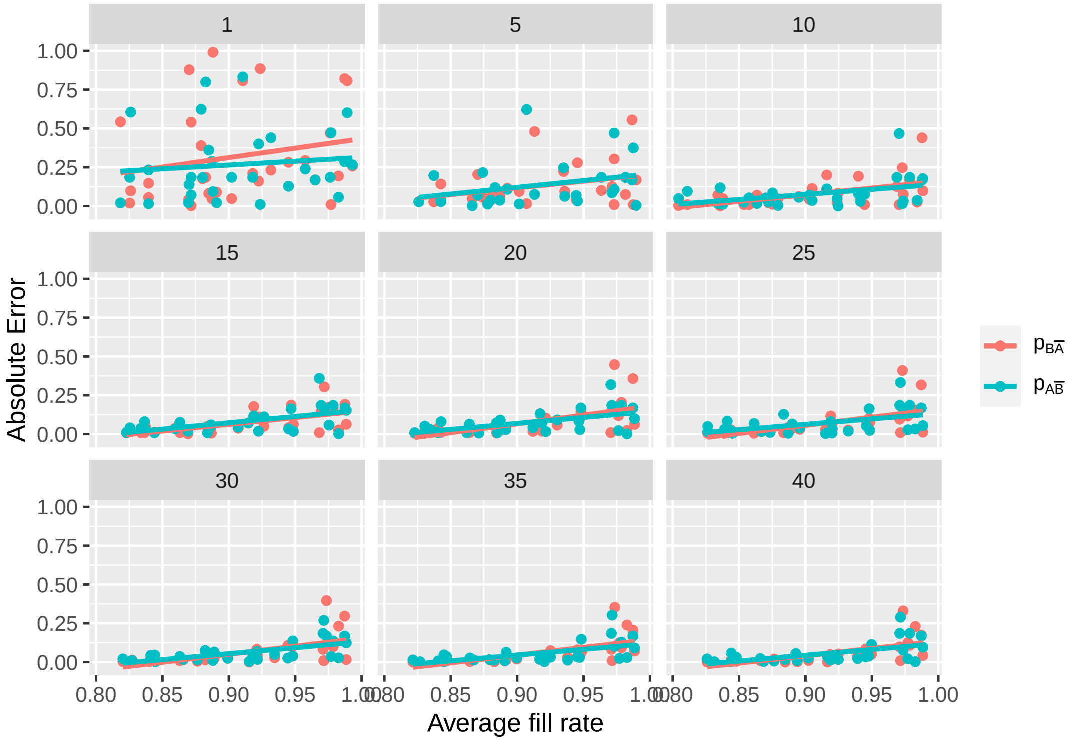


Figure 4.7: Absolute errors of substitution probabilities when λ-method was used with varying number of data and fill rates. Numbers in the panel headers correspond to N_{01} values where N_{10}, N_{11} and N_{00} were set equal to N_{01}.

Also, five experiments for each N_{01} value were performed. Other parameters N_{10}, N_{11} and N_{00} were set equal to N_{01}. For example, if N_{01} is 20, this means a data set with a total of 80 observations includes 60 observations where at least one product stocks out during the period. The values of (l, u) were set as $(0.9, 3)$ to obtain mean fill rates that vary between 0.899 and 0.903. Results were summarized in Figure 4.6. According to the simulation results, assuming an average fill rate around 0.9; N_{10} should be at least 10 in order to obtain an average absolute error of $p_{B\bar{A}}$ that is less than 0.05. Similarly, N_{01} should be at least 10 in order to obtain an absolute error of $p_{A\bar{B}}$ less than 0.05. However, there are 2.52 periods on average with the required type of stock-out case, for the groups including a product that can potentially be discontinued. This lack of data leads to the assumption that groups with certain attributes can form a cluster and a cluster of groups have a common substitution probability between the old and new products that can be estimated using the data of all groups in a cluster. The performance of λ-method in case of combining the data of multiple groups to estimate a common substitution probability was tested in Section 4.2.2.2.

4.2.2.2 Simulation by Combining Multiple Groups

This simulation method was designed to study the statistical properties of the mean absolute errors resulting in using λ-method in case of combining the data of multiple groups to estimate a common substitution probability. For example, let there be two groups with products (A_1, B_1) and (A_2, B_2). Let the release date of products A_1 and A_2 be before the release date of products B_1 and B_2. We assume that $p_{A_1\bar{B}_1} = p_{A_2\bar{B}_2}$ and $p_{B_1\bar{A}_1} = p_{B_2\bar{A}_2}$. Since λ-method uses sales rates to find substitution probabilities, this also means that we assume $\frac{\lambda_{A_1\bar{B}_1} - \lambda_{A_1}}{\lambda_{B_1}} = \frac{\lambda_{A_2\bar{B}_2} - \lambda_{A_2}}{\lambda_{B_2}}$. Since the sales rates of different groups can be at different scales, we need to scale the sales values of different groups so that the assumption of $\frac{\lambda_{A_1\bar{B}_1} - \lambda_{A_1}}{\lambda_{B_1}} = \frac{\lambda_{A_2\bar{B}_2} - \lambda_{A_2}}{\lambda_{B_2}}$ can be valid.

In a group of two products, the product that was released earlier was labeled as A and the other one as B. Each group's sales and inventory data for one period

Table 4.5: Definitions of sets and variables used in Chapter 4.2.2.2

Variable	Definition
J	Set of group indices = 1,. . . ,N_{groups}
T	Set of period indices = 1,. . . , $N_{00} + N_{10} + N_{01} + N_{11}$
y^t_{ij}	Sales of product i of group j at period t, $i \in \{A, B\}$, $j \in J$, $t \in T$

were simulated using the method in Section 4.2.2.1. Number of groups to be combined (N_{groups}) is an additional parameter of this simulation.

To scale the sales data of different groups, the scaling method in (4.3) was used.

$$\begin{aligned} y_{max} &= \max_j\{\sum_{t\in T}(y^t_{Aj} + y^t_{Bj})\}, \\ y^t_{Aj} &:= y^t_{Aj}\frac{y_{max}}{\sum_{t\in T}(y^t_{Aj} + y^t_{Bj})}, \quad \forall j \in J, t \in T, \\ y^t_{Bj} &:= y^t_{Bj}\frac{y_{max}}{\sum_{t\in T}(y^t_{Aj} + y^t_{Bj})}, \quad \forall j \in J, t \in T. \end{aligned} \tag{4.3}$$

λ-method performed poorly when this simulation method was used. The mean absolute errors were high due unreasonable assumptions of this simulation method. This led to the necessity of a modification in the method, described in Section 4.3, so that substitution probabilities are estimated directly, without the need of estimating sales rates at a common scale for multiple groups.

4.3 p-Method: Estimating Common Substitution Probabilities

The purpose of this method is to estimate the common substitution probabilities between old and new products of a cluster of groups so that the size of the data used for estimation is greater when compared with the case of estimating the substitution probabilities of each group separately. The λ-method that estimates sales rates was modified so that $\lambda_{A\bar{B}}$ and $\lambda_{B\bar{A}}$ are no longer estimated. Instead, they were written in terms of the substitution probabilities $p_{A\bar{B}}$ and $p_{B\bar{A}}$.

Let K be the set of indices of all clusters of groups. Let J_k be the set of indices of groups that belong to cluster $k \in K$.

We obtain λ_A and λ_B values using the average sales of periods with no stock-outs. Remember that those values, being decision variables, were estimated in the maximum likelihood problem of the λ-method. Namely,

$$\begin{aligned}\lambda^k_{A\bar{B}}(p^k_{A\bar{B}}) &= \lambda^k_A + \lambda^k_B \times p^k_{A\bar{B}},\\ \lambda^k_{B\bar{A}}(p^k_{B\bar{A}}) &= \lambda^k_B + \lambda^k_A \times p^k_{B\bar{A}}.\end{aligned}$$

The log-likelihood is now a function of the substitution probabilities which are the only decision variables.

$$\begin{aligned}\max_{p^k_{A\bar{B}},p^k_{B\bar{A}}} \quad & \prod_{j\in J_k} \ell_j(\lambda^k_A, \lambda^k_B, \lambda^k_{A\bar{B}}(p^k_{A\bar{B}}), \lambda^k_{B\bar{A}}(p^k_{B\bar{A}})), \qquad \forall k \in K,\\ \text{s.t.} \quad & 0 \le p^k_{A\bar{B}} \le 1,\\ & 0 \le p^k_{B\bar{A}} \le 1.\end{aligned} \tag{4.4}$$

The problem in (4.4) is solved for all clusters of groups, whereas the problem in λ-method, (4.2), is solved for all clusters of groups separately.

4.3.1 Simulation for the p-Method:

The same simulation method described in Section 4.2.2.2 was used to test the performance of p-method that estimates the common substitution probabilities since $\lambda_{A\bar{B}}$ and $\lambda_{B\bar{A}}$ were already written in terms of the substitution probabilities in those methods. However, in the simulation combining multiple groups, the step of scaling the sales data is no longer necessary since now the common substitution probabilities of groups are being estimated, not the sales rates in case of stock-outs. The results of this simulation and its comparison with other methods will be discussed in Section 4.5.

4.4 β-Method: Logistic Regression Model for Substitution Probabilities

Another approach to overcome the problem of lack of enough data to use λ-method is that each attribute of a group has its own contribution to the substitution probabilities. The level of this contribution is common among all groups and is a linear predictor of the substitution probabilities. The magnitude of these contributions were expressed in the vectors $\boldsymbol{\beta}_{\mathbf{A\bar{B}}}$, $\boldsymbol{\beta}_{\mathbf{B\bar{A}}}$ and they became the new decision variables of the maximum likelihood problem. Note that there are two different $\boldsymbol{\beta}$ vectors subscripted with $A\bar{B}$ and $B\bar{A}$ since the magnitude of contributions of group attributes to $p_{A\bar{B}}$ and $p_{B\bar{A}}$ might be different. In this structure, the substitution probability of a group is determined using the data of all groups having at least one common group attribute with it.

There were seven group attributes that were defined for our data with the number of levels being 2, 3, 3, 2, 3, 3, 3 respectively. This means there are 13 regression parameters to be estimated for each substitution probability, $p_{A\bar{B}}$ and $p_{B\bar{A}}$. Let $\boldsymbol{\beta}_{\mathbf{A\bar{B}}}$ and $\boldsymbol{\beta}_{\mathbf{B\bar{A}}}$ be vectors of regression parameters ($|\boldsymbol{\beta}_{\mathbf{A\bar{B}}}| = |\boldsymbol{\beta}_{\mathbf{B\bar{A}}}| = 13$), and let $\mathbf{x}$ be a vector of binary values indicating if a product has a certain level of an attribute or not. We obtain λ_A and λ_B values using the average sales of periods with no stock-outs as we did in the p-method.

We modified the p-method so that $p_{A\bar{B}}$ and $p_{B\bar{A}}$ are expressed in terms of the regression parameters corresponding the group attributes as in

$$\begin{aligned} p^j_{A\bar{B}} &= \sigma(\mathbf{x}^{j^T}\boldsymbol{\beta}_{\mathbf{A}\bar{\mathbf{B}}}), \\ p^j_{B\bar{A}} &= \sigma(\mathbf{x}^{j^T}\boldsymbol{\beta}_{\mathbf{B}\bar{\mathbf{A}}}), \\ \lambda^j_{A\bar{B}}(\boldsymbol{\beta}_{\mathbf{A}\bar{\mathbf{B}}}) &= \lambda^j_A + \lambda^j_B \times p^j_{A\bar{B}}(\boldsymbol{\beta}_{\mathbf{A}\bar{\mathbf{B}}}), \\ \lambda^j_{B\bar{A}}(\boldsymbol{\beta}_{\mathbf{B}\bar{\mathbf{A}}}) &= \lambda^j_B + \lambda^j_A \times p^j_{B\bar{A}}(\boldsymbol{\beta}_{\mathbf{B}\bar{\mathbf{A}}}), \end{aligned}$$

where $\sigma(y) = (1 + e^{-y})^{-1}$.

The log-likelihood is now a function of the regression parameters $\boldsymbol{\beta}_{\mathbf{A}\bar{\mathbf{B}}}$ and $\boldsymbol{\beta}_{\mathbf{B}\bar{\mathbf{A}}}$, which are the only decision variables in this problem:

$$\max_{\boldsymbol{\beta}_{\mathbf{A}\bar{\mathbf{B}}},\boldsymbol{\beta}_{\mathbf{B}\bar{\mathbf{A}}}} \quad \prod_{j\in J} \ell_j(\lambda^j_A, \lambda^j_B, \lambda^j_{A\bar{B}}(\boldsymbol{\beta}_{\mathbf{A}\bar{\mathbf{B}}}), \lambda^j_{B\bar{A}}(\boldsymbol{\beta}_{\mathbf{B}\bar{\mathbf{A}}})). \tag{4.5}$$

Note that the problem in (4.5) is solved once, using the data of all groups.

4.4.1 Simulation for the β-Method

To test the performance of β-method, we formulated two simulation methods. One of them sets the value of some parameters from random distribution, whereas the second calibrates them using our data set.

4.4.1.1 With Random Parameters

Let A be the set of attributes observed in the data. Parameters of the simulation are:

- N_{00}, N_{10}, N_{01} and N_{11},
- λ^j_A, $\lambda^j_B \quad \forall j \in J$,
- (l, u),

- N_{groups},
- $R = \{a_1, \ldots, a_{|R|}\} \subseteq A$: The set of attributes that determine the substitution probabilities $p_{A\bar{B}}$ and $p_{B\bar{A}}$.

Note that the set R can also include interactions between attributes. λ_A^j and λ_B^j values are drawn from a uniform distribution between 100 and 10000. The true values of the regression parameters $\boldsymbol{\beta_{A\bar{B}}}$ and $\boldsymbol{\beta_{B\bar{A}}}$ are also drawn from a uniform distribution, between -2 and 2.

Inputs of the simulation method described in Section 4.2.2 for each group j were generated as follows:

$$\begin{aligned}
p_{A\bar{B}}^j &= \sigma(\mathbf{x}^{j^T} \boldsymbol{\beta_{A\bar{B}}}), \\
p_{B\bar{A}}^j &= \sigma(\mathbf{x}^{j^T} \boldsymbol{\beta_{B\bar{A}}}), \\
\lambda_{A\bar{B}j} &= \lambda_{Aj} + \lambda_{Bj} \times p_{A\bar{B}}^j, \\
\lambda_{B\bar{A}j} &= \lambda_{Bj} + \lambda_{Aj} \times p_{B\bar{A}}^j, \\
I_{Aj} &\sim \text{Unif}(\lambda_{Aj} l, \lambda_{Aj} u), \\
I_{Bj} &\sim \text{Unif}(\lambda_{Bj} l, \lambda_{Bj} u).
\end{aligned}$$

4.4.1.2 With Parameters as Calibrated Using Real Data from the Tire Manufacturer

In this method, some of the parameters of the method in Section 4.4.1.1, namely N_{00}, N_{10}, N_{01}, N_{11}, N_{groups} and λ_A and λ_B were calibrated using real data from the tire manufacturer so that the performance of the β-method on our data can be measured. Values of λ_A and λ_B of each group are the average sales of periods with no stock-outs in the data.

There are 96 groups observed in the data. Therefore, we set $N_{groups} = 96$. As an example, values of the simulation parameters calibrated with data of six groups were given in Table 4.6.

Table 4.6: Values of the simulation parameters calibrated with data of six groups

Group Index	N_{00}	N_{10}	N_{01}	N_{11}	λ_A	λ_B
1	4	0	25	1	1231	397
4	13	4	2	3	7565	1000
6	5	0	23	2	133	83
7	8	0	19	3	126	100
8	4	1	21	3	43	89
9	16	2	8	0	1349	56

4.5 Comparison of Methods

We needed to select which method to use depending on the size of our data and whether the assumptions made for the p- and β-methods were valid. For p- and β-methods, we also needed to select the set of attributes that determine the substitution probabilities, R. We used mean absolute error and AIC criteria for these purposes.

4.5.1 Mean Absolute Error Comparison

Figure 4.8 shows that when data are created using the simulation method in Section 4.4.1.1 and the same data are used to apply three methods estimating sales rates, common substitution probabilities and regression parameters, denoted by λ, p and β, respectively, estimating regression parameters yield the lowest mean absolute error. As expected, when the substitution probabilities of different groups are common, combining their data and estimating common substitution probabilities yield lower mean absolute errors than estimating sales rates of each group separately. This is true for both simulation methods using random and observed parameters. However, due to the lack of enough periods with stock-outs and lack of groups with certain combinations of attributes, the mean absolute error (MAE) values of the substitution probabilities increase as the number of estimated parameters increases when the simulation method using observed parameters is used.

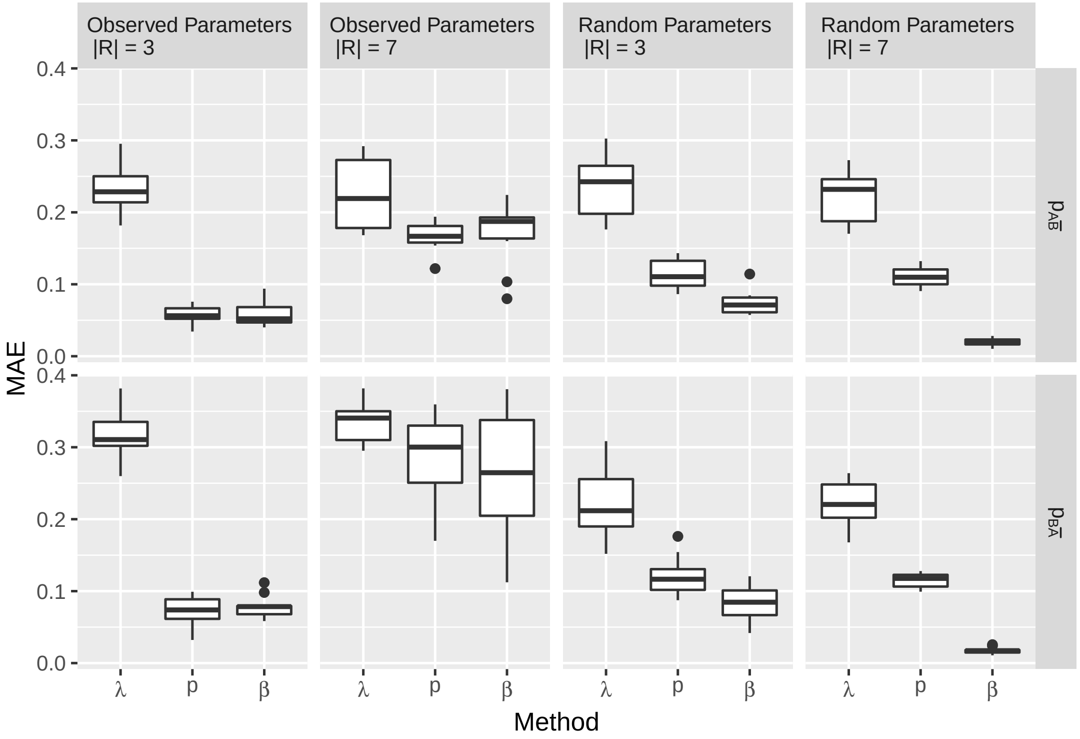


Figure 4.8: MAE comparison of all three methods

4.5.2 AIC and BIC Comparison

Simulating sales data and substitution probabilities were necessary to compare the MSE values of the three methods with the underlying assumption that substitution probabilities depend on the attributes of the groups. In the simulation method used for the p-method, it was assumed that groups with the same levels of certain attributes have the same substitution probabilities; whereas in the simulation method used for the β-method, certain attributes of each group have a fixed contribution to the group's substitution probabilities. However, in the real data, the dependency of substitution probabilities on the attributes of the groups is unknown. Estimations using a smaller data set rather than a larger data set with a faulty assumption might be more favorable. Therefore, AIC and BIC criteria that can calculate the relative performance of different models on a certain data set were used to find the best method for our data set.

The selection of attributes that affect the substitution rate was another decision to be made for the methods estimating substitution probabilities and regression parameters.

Performance of p-method was tested by combining the data from groups with common attributes in R, and the set R was set as all possible combinations of attributes in A. Figure 4.9 shows that as the cardinality of R increase; namely, as the number of attributes that determine the substitution probabilities of a group increase, AIC and BIC values decrease, and that they are consistent with each other. This result was an indication that some groups might favor λ-method over other methods when AIC and BIC criteria were used. Therefore, the method and set of attributes R yielding the lowest BIC value for each group were selected. The selection was made among the λ-method, p-method with the set of all possible combinations of the seven attributes including the empty set, β-method with all seven attributes and β-method with three attributes which were selected using the likelihood ratio (LR) test, as described in (4.6).

$$\begin{aligned}
&B = \text{log-likelihood of the large model,}\\
&b = \text{number of parameters of the large model,}\\
&S = \text{log-likelihood of the small model,}\\
&s = \text{number of parameters of the small model,}\\
&LR = 2(B - S) \sim \chi^2(b - s) \text{ if small model is correct,}\\
&p - value = P\{\chi^2(b - s) > LR\},\\
&\text{If } p\text{-value} < 0.05, \text{ we reject the small model.}
\end{aligned} \tag{4.6}$$

In Table 4.7, number of groups that prefer different methods with different cardinalities of the attribute set R based on the value AIC criterion are given. The cardinality of R, rather than its elements is provided. Results show that although many groups favored λ-method that estimates sales rates possibly because the underlying assumptions of other methods were not valid for them, there still a

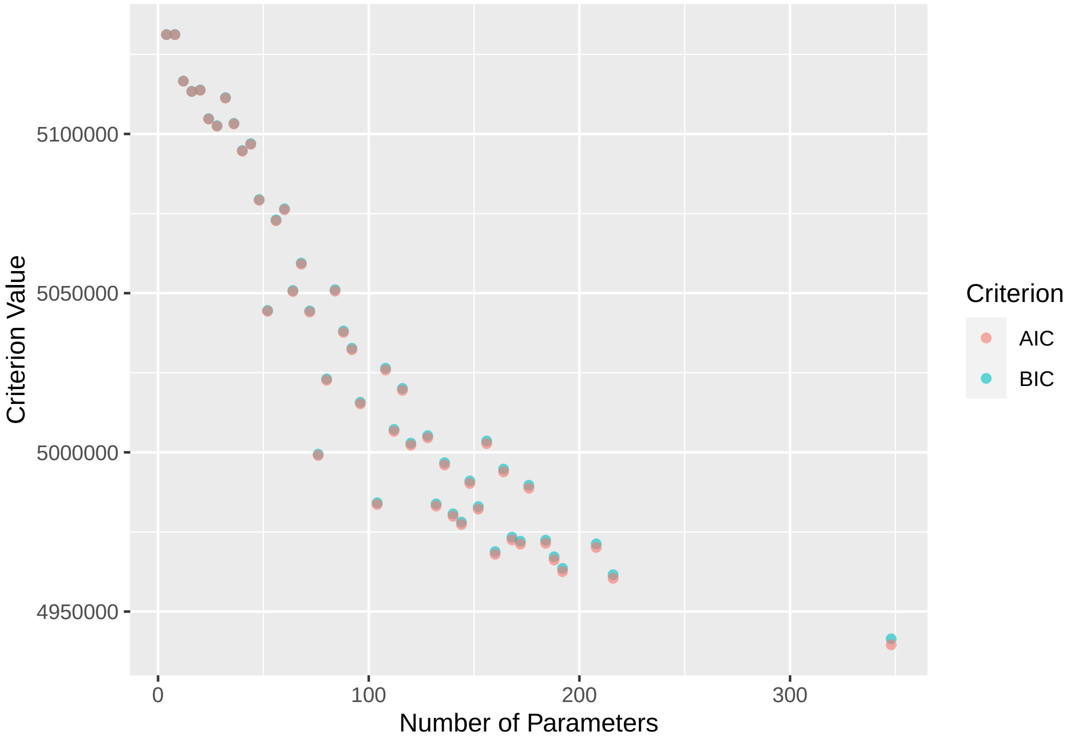


Figure 4.9: AIC and BIC criteria values using p-method, for different number of estimated parameters which depend on the cardinality of the attribute set R

Table 4.7: Number of groups that selected different methods with different cardinality of the attribute set R

Method	Cardinality of the Attribute Set	Number of Groups
p-method	3	1
p-method	5	1
p-method	6	1
p-method	6	1
p-method	6	1
p-method	7	15
λ-method	NA	39

significant portion (33.9%) that favors the other two methods.

To find the substitution probabilities of the groups with more than two members, the method estimating sales rates were used. Two dummy products labeled as A and B were created by combining the sales data of the ones that are discontinued, and the ones that are kept in the assortment. While combining products, the sum of the sales quantities and the minimum of S_A or S_B values of the products being combined were used as the sales quantity and stock-out case of the new dummy products.

Chapter 5

Modeling the Cost of Complexity

Complexity of a set of products was defined as the monetary implications of the sum of the variety-based break times that occur during the three major manufacturing processes: curing, tire building and extruding. Variety-based break times are the duration of the breaks during the production caused by the variety in the product portfolio, such as set up times and delay in the arrival of materials unique for some products.

In this section, we will formulate a machine learning model that estimates the duration of the variety-based break times as a function of the assortment. This is a major component of estimating the complexity cost savings, as will be explained in Section 7.

Variety-based break times that occur during the three major manufacturing processes were estimated using the entropy and fraction of the attributes of the tires assigned to each machine that might have an effect on the duration of break times. However, our model is flexible. One can incorporate other factors as long as data are available.

The attributes of the products assigned to each machine in each month for 10 years, as well as the production and variety-based break times were known. We assume that machines in the same manufacturing processes are identical. Then,

Table 5.1: Definitions of variables used in Section 5

Variable	Definition
E_n^{time}	Time entropy for an attribute with n categories
pt_i	Total production time of the products from category i in a machine in a month
NE_n^{time}	Normalized time entropy for an attribute with n categories
$f_n^{time,i}$	Fraction of the production time of products from category i to the total production time of the products from n categories
$E_n^{time,k}$	Time entropy for the attribute k with n categories
$NE_n^{time,k}$	Normalized time entropy for the attribute k with n categories
E_n^{count}	Count entropy for an attribute with n categories
η_i	Number of products from category i in a machine in a month
NE_n^{count}	Normalized count entropy for an attribute with n categories
$f_n^{count,i}$	Fraction of the number of products from category i to the total number of the products from n categories
$E_n^{count,k}$	Count entropy for the attribute k with n categories
$NE_n^{count,k}$	Normalized count entropy for the attribute k with n categories
vt	Total variety-based break times in a machine in a month
pt	Total production time in a machine in a month

we estimate the fraction of variety-based break time vt in the total machine times as a function of various observed attributes of machine and the product bundles planned to be produced on the same machines. An ensemble of deep learning, random forest, gradient boosting machine, xgboost, and generalized linear models with their best parameters were selected after a five-fold cross-validation over a suitable grid space, using the H20 automatic machine learning function accessible from R [34, 35].

The variety-based break time typically increases with the increasing variety of product numbers, attributes, their production times for the products planned to be produced on the same machine. To measure the divergence between those numbers, we calculate the entropies of product attributes assigned for production to the same machine. Scaled entropies became the predictors in the machine learning models for variety-based break time predictions.

Entropies of product attributes corresponding to each machine, month and year were calculated with the entropy formula applied to fractions of the production times of all products with the same attribute values. Entropy for an attribute with n categories, each allocated production time pt_i, $i = 1, \ldots, n$ was denoted as E_n^{time} and given by

$$\begin{aligned} E_n^{time} &= \sum_{i=1}^{n} \frac{pt_i}{\sum_{j=1}^{n} pt_j} \log \frac{\sum_{j=1}^{n} pt_j}{pt_i} \\ &= \frac{1}{\sum_{j=1}^{n} pt_j} \sum_{i=1}^{n} pt_i \left[\log \sum_{j=1}^{n} pt_j - \log pt_i \right] \\ &= \log \sum_{j=1}^{n} pt_j - \frac{\sum_{i=1}^{n} pt_i \log pt_i}{\sum_{j=1}^{n} pt_j}. \end{aligned}$$

The maximum value of E_n^{time} depends on n;

$$0 \leq \sum_{i=1}^{n} \frac{pt_i}{\sum_{j=1}^{n} pt_j} \log \left(\frac{\sum_{j=1}^{n} pt_j}{pt_i} \right) \leq \log \sum_{i=1}^{n} \frac{pt_i}{\sum_{j=1}^{n} pt_j} \frac{\sum_{j=1}^{n} pt_j}{pt_i} = \log n.$$

This means that the entropy values of the machines with different number of products assigned to them are not directly comparable. However, $\log n$ bound is tight and can be attained if the pt_i are the same. Therefore, entropy values were normalized as $NE_n^{time} = E_n^{time} / \log(n)$.

Another type of entropy value for an attribute with n categories, each common to η_i, $i = 1, \ldots, n$ products assigned to the same machine on a given production planning month is denoted as E_n^{count}.

$$E_n^{count} = \sum_{i=1}^{n} \frac{\eta_i}{\sum_{j=1}^{n} \eta_j} \log \frac{\sum_{j=1}^{n} \eta_j}{\eta_i}$$

Like E_n^{time}, we use normalized count-based entropies, $NE_n^{count} = E_n^{count} / \log(n)$.

A list of all attributes and their respective number of levels are given in Table 5.2. We can illustrate the process of generating inputs for the machine learning model on a small example. Let there be a machine with 5 products assigned to it during a month. Two attributes of those products and their production times are given in Table 5.3. Let $k = 1$ be the attribute index of rim size and $k = 2$ be the attribute index of rag code and let $E_n^{time,k}$ be the time entropy of attribute k with n levels. There are four distinct values of rim sizes observed in this machine, which means $n = 4$ for rim size. We find the number of products for each value of rim size and their total production times in Table 5.4. Let $f_n^{time,i} = \frac{pt_i}{\sum_{j=1}^{n} pt_j}$. Then, the time entropy of the attribute rim size on this machine during this month is $E_n^{time} = \sum_{i=1}^{n} f_n^{time,i} \log(f_n^{time,i})^{-1}$. This means that the sum of the $f_n^{time,i}$ columns multiplied by -1 gives us the values in column $E_n^{count,k}$ of Table 5.5. To obtain the log normalized values $NE_n^{time,k}$, we divide $E_n^{time,k}$ by $\log n$. Same calculations are done to find the time entropies of both attributes. Final calculations for both attributes were provided in Table 5.5. Notice that while there are four distinct values of rim sizes, there is only one distinct value of rag codes. Therefore, heterogeneity measured by the normalized entropy of rag code is as low as it can be, that is 0, while the entropies of rim sizes are close to 1.

A linear regression model assumes that all the observations come from a normal distribution with the same variance. Since the original data do not satisfy these assumptions, data transformation was necessary. Therefore, a Box-Cox transformation was applied to the response variable which was the ratio of the total variety-based break times to the production time on a machine [36]. After the transformation, the new response variable becomes:

$$y = \operatorname{sgn}(\lambda) \left(\frac{vt}{pt} \right)^{\lambda} + \varepsilon$$

for some random ε. The power λ was found using Box-Cox method, with *boxcox* function of the *MASS* package in R [37].

After using the machine learning model to estimate the transformed ratio

Table 5.2: Table of all attributes and corresponding number of levels

Attribute	Number of levels
rim size	10
rag code	12
green case commonality	9
compound comm	30
tandem	2
filler width	29
aspect ratio	11
original equipment	2
feg application	2
ply construction	5
silica tread	4
vacuum container	2
compact pt	2
envelope	2
cat compound	48
cubic root of volume to mold ratio	32

Table 5.3: Rim sizes, rag codes and production times of a machine during a month

product_code	production_time	rim size	rag code
1	265	16	LP008
2	372	16	LP008
3	572	15	LP008
4	908	14	LP008
5	201	13	LP008

Table 5.4: Counts and production times corresponding to each unique rim size level

Attribute level index (i)	rim size	η_i	pt_i	$f_n^{time,i}$	$f_n^{count,i}$
1	13	1	201	-0.2120260	-0.3218876
2	14	1	908	-0.3671233	-0.3218876
3	15	1	572	-0.3453027	-0.3218876
4	16	2	637	-0.3549641	-0.3665163

Table 5.5: Calculation of time and count entropies for the attributes rim size and rag code

Attribute index (k)	attribute	n	$E_n^{count,k}$	$E_n^{time,k}$	$NE_n^{count,k}$	$NE_n^{time,k}$
1	rim size	4	1.332179	1.279416	0.960964	0.9229036
2	rag code	1	0.000000	0.000000	0.000000	0.0000000

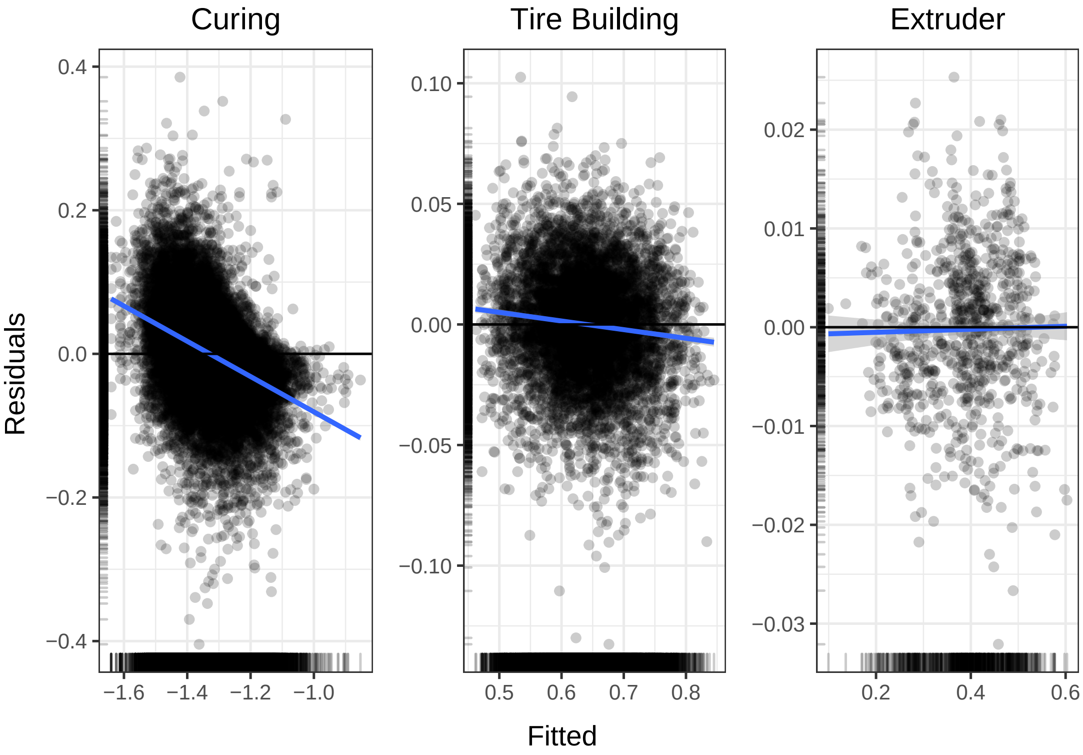


Figure 5.1: Residual Analysis plots of the machine learning models for the three machine groups

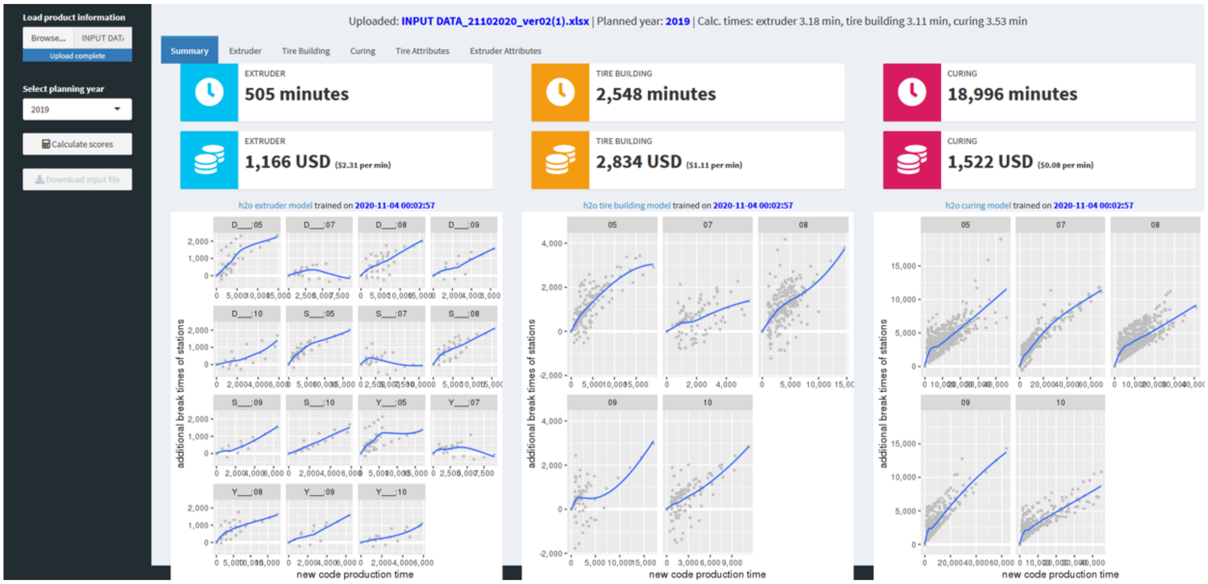
of the total variety-based break times to the total production time, the back-transformation in (5.1) is necessary to obtain the variety-based break times in minutes.

$$\hat{vt} = (sign\,(\lambda)\ \ \hat{y})^{\lambda} \times pt. \tag{5.1}$$

Chapter 6

Complexity Score Calculator

After modelling the variety-based break times as a function of the assortment, we designed an interface that calculates the marginal complexity cost of a new product. The interface requires an input of the attributes of the new product used as an explanatory variable in the model and the production plan of the new product. An example of these inputs can be found in Tables 6.2 and 6.1.



Uploaded: INPUT DATA_21102020_ver02(1).xlsx | Planned year: 2019 | Calc. times: extruder 3.18 min, tire building 3.11 min, curing 3.53 min

Load product information: INPUT DAT/ — Upload complete

Select planning year: 2019

EXTRUDER 505 minutes

TIRE BUILDING 2,548 minutes

CURING 18,996 minutes

EXTRUDER 1,166 USD ($2.31 per min)

TIRE BUILDING 2,834 USD ($1.11 per min)

CURING 1,522 USD ($0.08 per min)

h2o extruder model trained on 2020-11-04 00:02:57

h2o tire building model trained on 2020-11-04 00:02:57

h2o curing model trained on 2020-11-04 00:02:57

Figure 6.1: Complexity score calculator interface

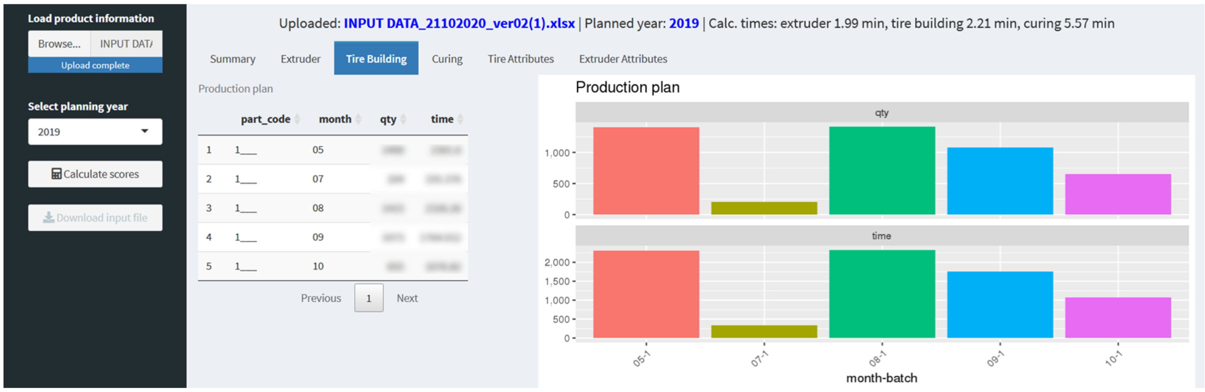
Figure 6.1 shows the main outputs of the interface, which are the total increase

Table 6.1: An example of an input table containing the production plan of the new product for the complexity score calculator interface

Part Code	Machine Group	Month	Quantity	Time (min)
X	Tire Building	03	100	1000
X	Tire Building	05	200	2000
X	Tire Building	07	300	3000
X	Tire Building	09	400	4000
X	Curing	03	100	1000
X	Curing	05	200	2000
X	Curing	07	300	3000
X	Curing	09	400	4000

Table 6.2: An example of an input table containing the attributes of the new product for the complexity score calculator interface

Tire Attributes	Data Type	Similar Code	New Code
Product Code	-	Y	X
Rim Size	past data	16	19
Rag Code (LV001, LR014, LR018)	past data	RC1	RC2
Green case commonality index	new data	0.01	0.01
Compound commonality index	new data	0.01	0.01
Tandem compound (1,0)	new data	0	0
High thin bead filler	new data	0.1	0.1
Volume/ mold ratio	new data	50	50
Aspect ratio	new data	55	55
OE (0/1)	-	0	0
Feg application (0/1)	-	0	0
Ply construction (1P, 2P (1+1,2+0))	past data	1+0	1+0
Silica tread (%50, %90,0)	new data	50	90
Vacuum container (Yes/No)	-	0	0
Compact pt (0/1)	-	0	0
Envelope (0/1)	-	0	0
Cap Comp' (Tread)	past data	CC1	CC2



Load product information

Browse... INPUT DAT/

Upload complete

Select planning year

2019

Calculate scores

Download input file

Uploaded: INPUT DATA_21102020_ver02(1).xlsx | Planned year: 2019 | Calc. times: extruder 1.99 min, tire building 2.21 min, curing 5.57 min

Summary | Extruder | Tire Building | Curing | Tire Attributes | Extruder Attributes

Production plan

	part_code	month	qty	time
1	1___	05	[redacted]	[redacted]
2	1___	07	[redacted]	[redacted]
3	1___	08	[redacted]	[redacted]
4	1___	09	[redacted]	[redacted]
5	1___	10	[redacted]	[redacted]

Previous 1 Next

Figure 6.2: Input to the complexity score calculator: attributes of product in question on the left and production plan for each month on the right.

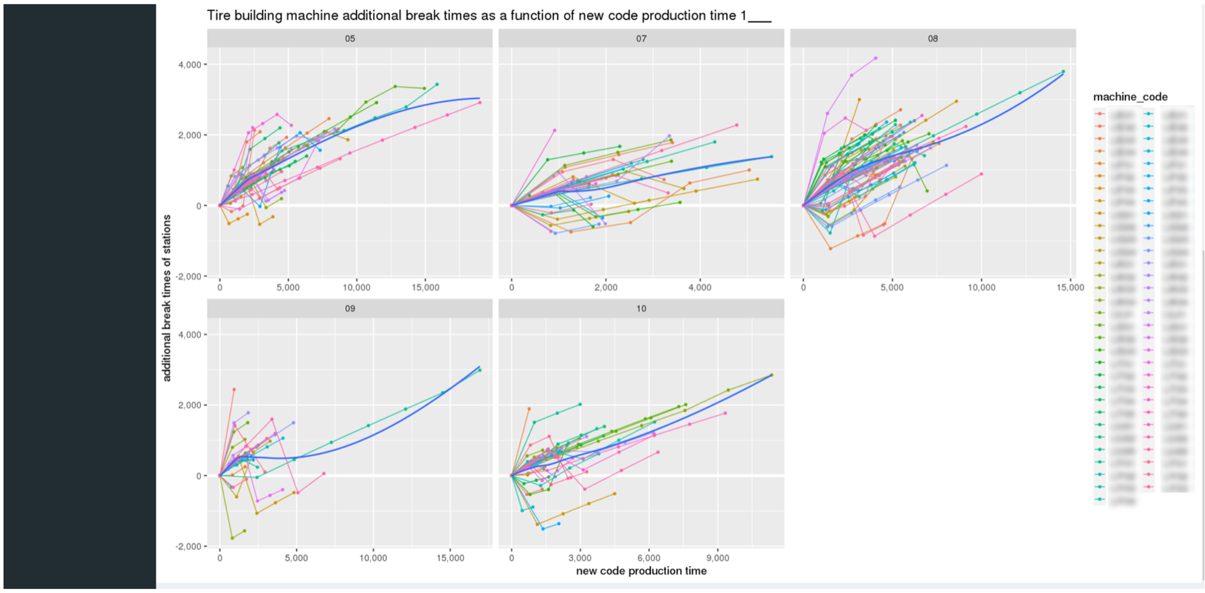


Figure 6.3: Machine specific estimated additional break times in each month 05-10 for different possible production times of the product in question, but at most up to unused production time left on the machine.

in variety-based break times in all machine groups after a new product is added to the assortment in 2019, in minutes, and the expected conversion cost during the same period, in US dollars. The plots show the increase in variety-based break times as a function of the production time of the new product, for all machine groups and all months in the production plan. Above are the predicted total additional break times and costs. Below a point in a panel represents a machine in curing, tire building or extruder workshop in every production planning month. The horizontal axis shows the maximum available production time on the machine to produce the new product in question. The vertical axis shows the additional break time estimated with machine learning algorithm if product in question is produced as much as the available production time on the machine is used. The solid smoothing curve in each panel is the estimated additional break times as a function of new product's production time averaged over all machines with some idle production time in that production planning month.

A summary of the production plan can be found in Figure 6.2. We use the production quantities and months in the production plan of 2020 in this example. Since our model of complexity cost uses machine-level data, adding a new product to the assortment requires the selection of machines from each machine group that the new product will be assigned to. Depending on the attributes of the products that are already assigned to machines, this selection affects the variety-based break time savings after the addition of the new products. The plots in Figure 6.3 shows the increase in variety-based break times as a function of the production time of the new product to every available machine from all machine groups, for all months in the production plan. The smoother line on each plot gives the average value across all machines after excluding the outlier values. These average values were used to estimate the total increase in variety-based break times after a new product is added to the assortment that were shown in Figure 6.1.

We can illustrate the process of finding the complexity scores of new products in each machine in terms of minutes and dollars by providing a numerical example. Let the production plan of the new product X in the curing machines be as given in Table 6.3. We see that product X is planned to be produced in three batches

Table 6.3: An example of an input table containing the production plan of the new product for the complexity score calculator interface

Part Code	Machine Group	Month	Quantity	Time(min)
X	Curing	05	100	200
X	Curing	08	50	100
X	Curing	10	20	40

in May, August and October in the current year 2020. We will start by doing the calculations of the first batch in May for the machine group curing. Production time required in May is 200 minutes. We will refer to the production plan of the same month in the previous year, May 2019. We see the available times in each curing machine for the production of the new product in Table 6.4. There are only two machines with adequate time to produce the new product: CC3 ad CC5. We use our machine learning model to estimate the increase in variety-based break times in those machines as a result of adding this new batch to the machines. Let these values be 50 and 20 minutes for machines CC3 and CC5, respectively. Then, the increase in variety-based break times in May in curing machines are the average of the two values, 35 minutes. Let the values in Table 6.5 be the increase in variety-based break times corresponding to all batches as a result of the production of the new product. Sum of those values, 170 minutes is the total increase in variety-based break times after the new product is added to the assortment in the curing machine group. If the conversion cost in curing machines per minute is $2, the conversion costs savings become $340. These calculations are repeated for every batch of every machine group to obtain the results as in Figure 6.1.

Table 6.4: Available times in each curing machine in May 2019 for the production of the new product

Curing Machine	Available Time (min)
CC1	100
CC2	120
CC3	205
CC4	195
CC5	400

Table 6.5: Increase in variety-based break times corresponding to all batches as a result of the production of the new product

Month	Increase in Variety-Based Break Times (min)
05	35
08	45
10	90

Chapter 7

Assortment Optimization

7.1 Optimization Model

We can finally integrate the results from Sections 4 and 5 to find the optimal assortment by maximizing the total savings from discontinuing a set of products. The necessary set, variable and function definitions used in this Chapter are given in Table 7.1.

Since the company identified 87 products that can potentially be discontinued, we have set the D as the set of indices of these products that are not allowed to be discontinued. The optimization model corresponding to this problem is

$$\begin{aligned} \max_x \quad & inv(\mathbf{x}) - rl(\mathbf{x}) + cc(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{x}_i = 0, \quad i \in D, \\ & \mathbf{x}_i \in \{0,1\}, \quad i \in I. \end{aligned} \tag{7.1}$$

where $inv(\mathbf{x})$ is the savings from inventory holding cost as a function of $\mathbf{x}$, $rl(\mathbf{x})$ is the profit loss from sales as a function of $\mathbf{x}$ and $cc(\mathbf{x})$ is the complexity cost savings as a function of $\mathbf{x}$. The first constraint set ensures that only the 87

Table 7.1: Definitions of variables used in Section 7

	Definition
Sets	
I	Set of all products, indexed by i
K	Set of group indices, indexed by k
D	Set of indices of products that are not allowed to be discontinued
$J_k = \{i \in I \mid g_i = k\}$	Set of products that belong to group k, $k \in K$
$J = \{J_k \mid k \in K\}$	Set of all groups
M	Set of machine groups, indexed by m. This set consists of curing, tire building and extruding machine groups
T	Set of all periods in 2019, indexed by t
B	Set of periods in 2019 that has at least one machine group with utilization greater than or equal to the utilization threshold u
Variables	
$\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{\lvert I \rvert})$	The solution vector
$\mathbf{x}_i = \begin{cases} 1, & \text{if product } i \text{ is discontinued,} \\ 0, & \text{otherwise.} \end{cases}$	Binary variable indicating whether product i is discontinued
$g_i \in K$	Index of the group that contains product i
$p_{j\bar{i}}$	Substitution probability from product j to i; namely, fraction of product i's demand that will be substituted by j if product i is out of stock
$p_{\bar{i}} = \sum_{\{j \in J_{g_i} \mid j \neq i\}} p_{j\bar{i}}$	Fraction of product i's demand that will be substituted by the other products in its group if product i is out of stock
pr_i	Total profit from product i in 2019
y_i	Total sales quantity of product i in 2019
cbl_i	Total demand loss from product i due to country-based business rules
up_i	Unit profit from product i
h_i	Inventory holding cost of product i
c^m	Value of variable cost per minute of production time in each machine group
cr	Sales creation rate
u	Utilization threshold
at_t^m	Total capacity of machine group m in period t, in minutes
u_t^m	Used capacity of machine group m in period t, in minutes
et_t^m	Capacity of machine group m that can be used for extra production machine group m in period t, in minutes
pt_{it}^m	Production time spent in machine group m for the production of product i in period t
at_{it}^m	Production time already available in machine group m that have utilization less than the utilization threshold before discontinuing any products
q_{it}	Production quantity of product i in period t
Functions	
$bt^m(\mathbf{x}, t)$	Break time savings from machine group m
$spt^m(\mathbf{x})$	Production time that should be spent for the production of the substituted demand in machine group m
$inv(\mathbf{x})$	Savings from inventory holding cost
$rl(\mathbf{x})$	Profit loss from sales
$cc(\mathbf{x})$	Complexity cost savings

products that can potentially be discontinued can actually be discontinued and the second constraint set defines the decision variables as binary variables.

The inventory holding cost function, $inv(\mathbf{x})$ is a linear function of $\mathbf{x}$. The values of the parameters h_i, $\forall i \in I$ was calculated by the company. Since the weight of inventory holding cost savings in the objective values are insignificant when compared with the value of profit loss from sales and the complexity cost savings, we do not go into detail of the calculation of h_i.

$$inv(\mathbf{x}) = \sum_{\{i \in I | \mathbf{x}_i = 1\}} h_i$$

$$rl(\mathbf{x}) = \sum_{k \in K} \Big\{ \sum_{i \in J_k} pr_i - \Big[\sum_{\{i \in J_k | \mathbf{x}_i = 0\}} y_i + \sum_{\{i \in J_k | \mathbf{x}_i = 1\}} (y_i - cbl_i) p_{\bar{i}} \Big] \\ \frac{\sum_{\{i \in J_k | \mathbf{x}_i = 0\}} up_i y_i}{\sum_{\{i \in J_k | \mathbf{x}_i = 0\}} y_i} \Big\}$$

We had company data to find the values of the parameters used in the $rl(\mathbf{x})$ function such as the unit profits and production quantities. The estimation of substitution probabilities, denoted by $p_{j\bar{i}}$, was explained in Section 4. Note that there is no substitution between products that belong to different groups; namely, $p_{j\bar{i}} = 0, \quad if \ g_i \neq g_j$.

To estimate the break time savings when a set of products are discontinued, the most recent year in the production plan data was chosen as the planning year: 2019. The difference between the estimation of the break times with and without the products that will be discontinued in the production plan of 2019 was calculated to find the break time savings corresponding to $\mathbf{x}$. The machine learning model explained in Section 5 was used to find the necessary estimations.

When a set of products are discontinued, their production time and break time savings due to the reduction in the variety of assortment indicate a freed-up capacity. However, since some fraction of the demand of the discontinued products shift to the products in their groups, some of the freed-up capacity is used for this additional demand for products in the assortment. Figure 7.1 shows the nonhomogeneous nature of the production times of different products. Production times and attributes of products, hence their contributions to the complexity can be different which might lead to the favoring of some products over others as a result of our algorithm. If there is still time left after the production time needed for the substituted demand, it was assumed that this excess time can be used to produce an "average" product, with production times and unit profit being equal the average of those for all products in the current assortment (before any products being discontinued), and the expected profit from it contributes to the complexity cost savings term of the objective function. However, it only makes sense to assume that this extra production can be converted into sales in the periods that have at least one machine group working in full capacity, referred to as *busy periods*, since otherwise the company would have already used the existing free capacity. Direct savings from complexity cost also occur by means of the savings from the variable costs such as energy and labor cost that would be used for the production of products removed from the assortment. The value of variable cost per minute of production time in each machine group, $c^m,\ m \in M$ was calculated, and the net additional time of each machine group during the free periods were multiplied with this value in order to find the monetary implication of the net additional time. We defined the sum of savings from the variable costs and profit from the extra production as the monetary value of the break time savings during the three major manufacturing processes when a set of products are discontinued; namely, complexity cost savings. The complexity cost savings were calculated as a set function so that the interactions between the products during manufacturing processes, as well as the variety in the product portfolio were taken into consideration.

7.2 Scenario Generation

The rate at which extra production of an average product can be turned into sales, called the *sales creation rate* and denoted by cr, and the *utilization threshold* u that defines a machine as working in full capacity are two parameters of this

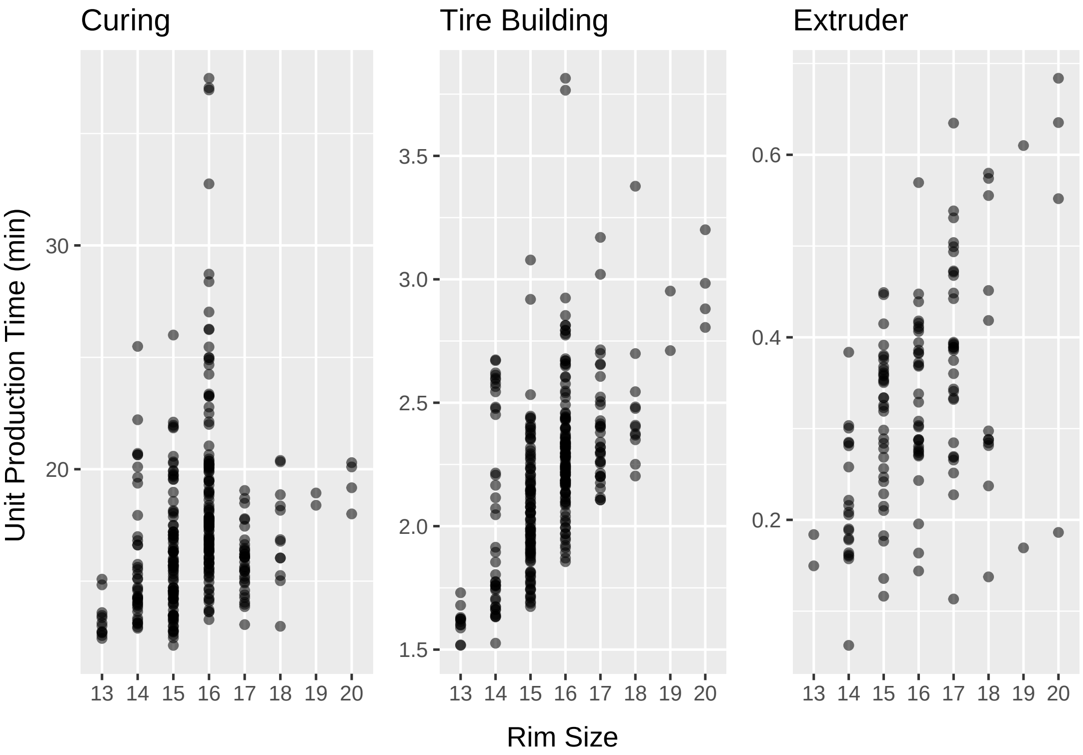


Figure 7.1: Production times of products with different rim sizes three machine groups

method that generate different scenarios. If a machine group m has a total capacity of at_t^m minutes in period t and uses u_t^m minutes, it is assumed that only $[(at_t^m - u_t^m) - (at_t^m(1-u))]^+$ minutes can be used for extra production and this time was denoted as et_t^m. As seen in Figure 7.2, if we set the utilization threshold as 100%, there are only two busy periods with the curing and extruder machines being the bottleneck machine groups. This means that there are only two periods in which the profit from the production of an average production can be considered as complexity savings. However, there are some periods with utilizations above 98% in which machines might have worked in full capacity but are not considered as such when utilization threshold is 100% possibly because the remaining 2% is fragmented into so many small pieces that we cannot use those times to produce anything useful. Depending on the utilization threshold, the number of periods in which the profit from the production of an average production can be considered as complexity savings vary from 2 to 12.

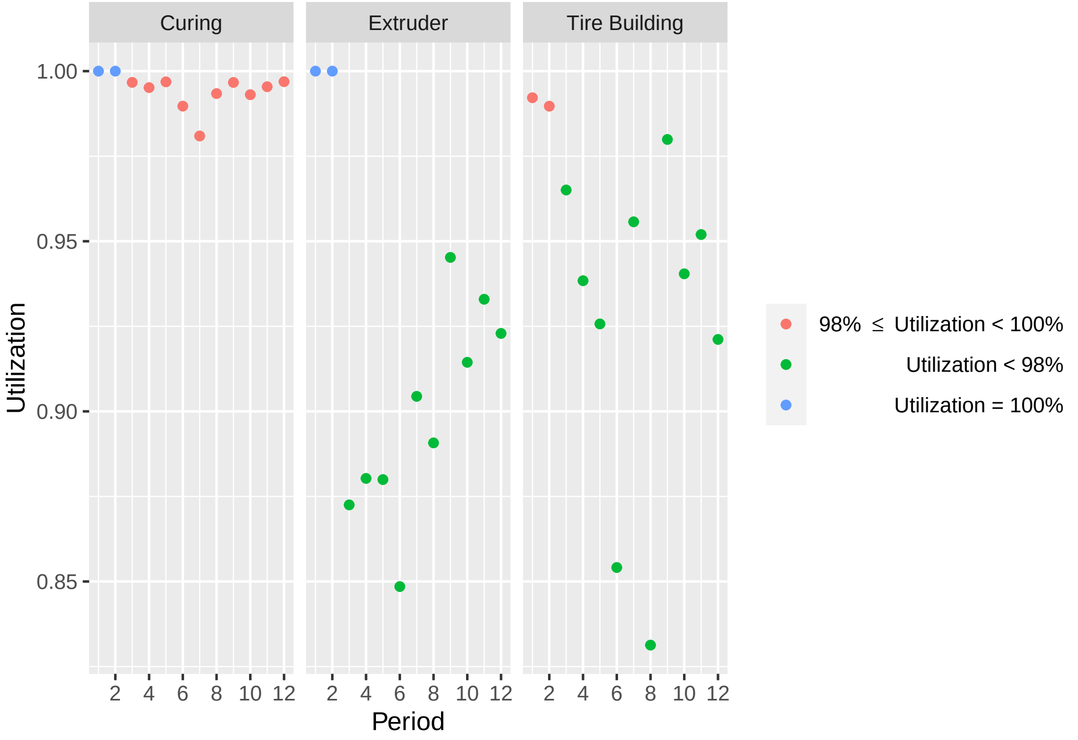


Figure 7.2: Utilizations of the three major machine groups in each month of the planning year, 2019.

We define the complexity cost function as follows:

$$\begin{aligned}
spt^m(\mathbf{x},t) &= \sum_{k\in K}\Big\{\Big[\frac{\sum_{\{i\in J_k|\mathbf{x}_i=0,\ t\in T\}} pt_{it}^m}{\sum_{\{i\in J_k|\mathbf{x}_i=0,\ t\in T\}} q_{it}} \sum_{\{i\in J_k|\mathbf{x}_i=1\}} (q_{it} - cbl_{it})p_{\bar{i}}\Big] \sum_{i\in J_k} \mathbf{x}_i\Big\},\\
adt^m(\mathbf{x},t) &= \sum_{\{i\in I|\mathbf{x}_i=1\}} pt_{it}^m + bt^m(\mathbf{x},t) - spt^m(\mathbf{x},t),\\
eq^m(\mathbf{x},t) &= \frac{adt^m(\mathbf{x},t) + et_t^m}{\frac{\sum_{i\in I,\ t\in T} pt_{it}^m}{\sum_{i\in I,\ t\in T} q_{it}}},\\
cc(\mathbf{x}) &= \sum_{m\in M,\ t\in T} [c^m adt^m(\mathbf{x},t)] + \sum_{t\in B}\Big[\min_{m\in M}\{eq^m(\mathbf{x},t)\} \sum_{i\in T} \frac{pr_i}{y_i} cr\Big].
\end{aligned}$$

7.3 Genetic Algorithm

Using the three profit components, we formulated a large-scale assortment optimization problem with a complex objective function due to the break time savings component, $bt^m(\mathbf{x},t)$, calculated using predictions from the machine learning model. Since the objective function cannot be expressed as a closed form equation, a genetic algorithm was used to find a good solution. The island evolution approach was incorporated to prevent the algorithm from being stuck in local optima.

The genetic algorithm with the island evolution approach requires the input of some parameters that determine the population size of each island, number of islands, the proportion of individuals that should migrate between the islands, the number of iterations at which exchange of individuals takes place, the probability of crossover between pairs of chromosomes, the probability of mutation in a parent chromosome, and the maximum number of iterations to run before the GA search is halted [38]. A grid search was done to determine the set of values for each parameter, by defining a set of values for each parameter and each combination of the selected values in these sets was used to run the algorithm for 40 iterations. The set of parameters yielding the highest objective value was selected.

The algorithm also requires the input of an initial population; namely, a set of solutions in the size of the population. Based on the hypothesis that products with similar attributes create a synergy and result in decreased break times during production, clusters of products were formed using hierarchical clustering. The details of this hierarchical clustering are given in the next subsection. The solutions that correspond to one cluster of products being discontinued at a time were fed to the genetic algorithm as a part of the initial population, along with the *initial solution* that corresponds to discontinuation of products that has a positive objective function value when they are the only product that is being discontinued. Some randomly generated solutions were also fed to the algorithm so that other regions in the solution space could also be explored.

7.3.1 Hierarchical Clustering

Hierarchical clustering was used to find clusters of products with similar attributes so that the solution that corresponds to discontinuation of products in one cluster altogether can be used as an initial solution while determining the initial population used by the genetic algorithm. The idea here is to use the remove a product along with other products similar attributes so that the heterogeneity of the products in the assortment can decrease. See James et al. [39] for detailed information on hierarchical clustering. There are categorical product attributes such as Rag code and cat compound code that show the part code of some of the main components of the tires. Therefore, Gower's distance was used as a distance metric to calculate the dissimilarity between products and a visualization of this distance matrix is given in Figure 7.3. Darker red regions around the right diagonal line of the plot suggests that there are there are three natural clusters.

After performing hierarchical clustering with complete linkage that was used to calculate the dissimilarity between clusters, we use Silhouette and Elbow methods to determine the optimal number of clusters. These methods suggest that the optimal number of clusters are 2 and 3 respectively, as seen in Figures 7.4 and 7.5.

A dendrogram plot is given in Figure 7.6 to visualize the dissimilarity between

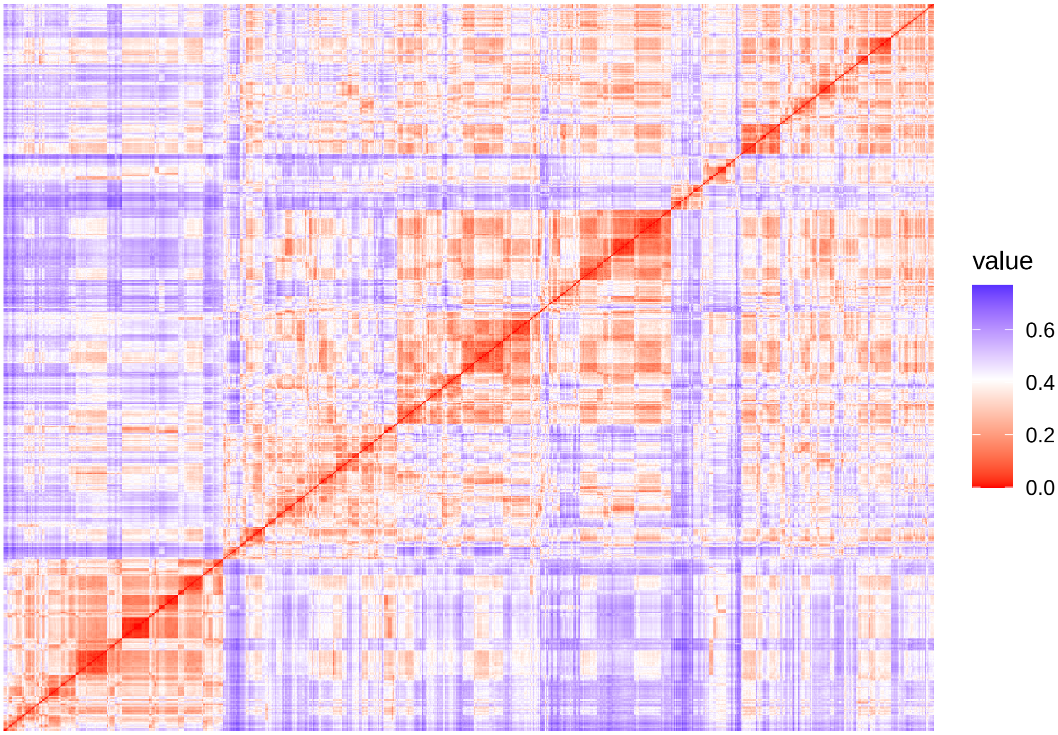


Figure 7.3: Visualization of the distance matrix used in hierarchical clustering where both axes correspond to products, created using Gower's distance metric [40, 41]

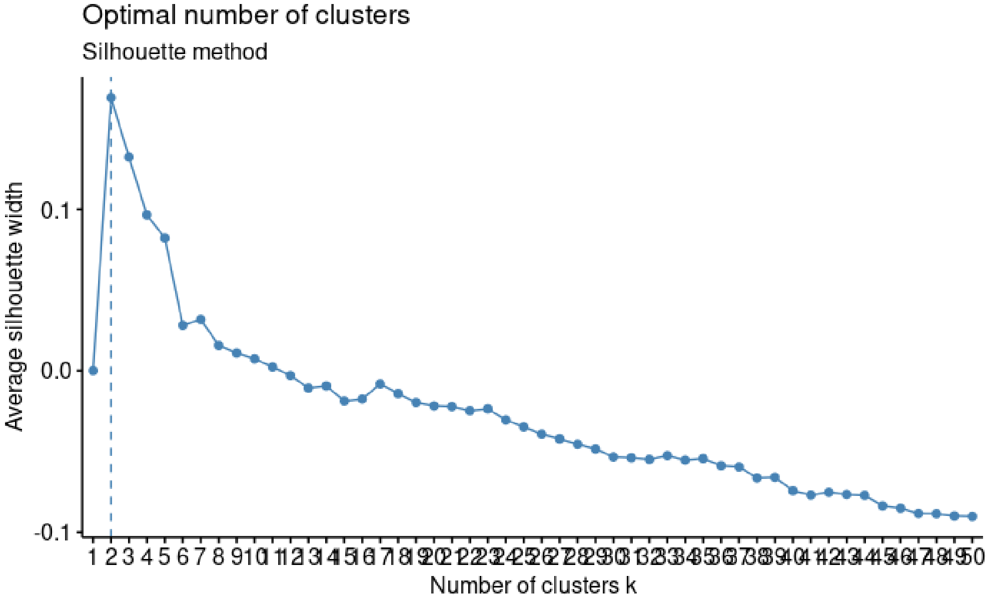


Figure 7.4: Optimal number of clusters using the Silhouette method

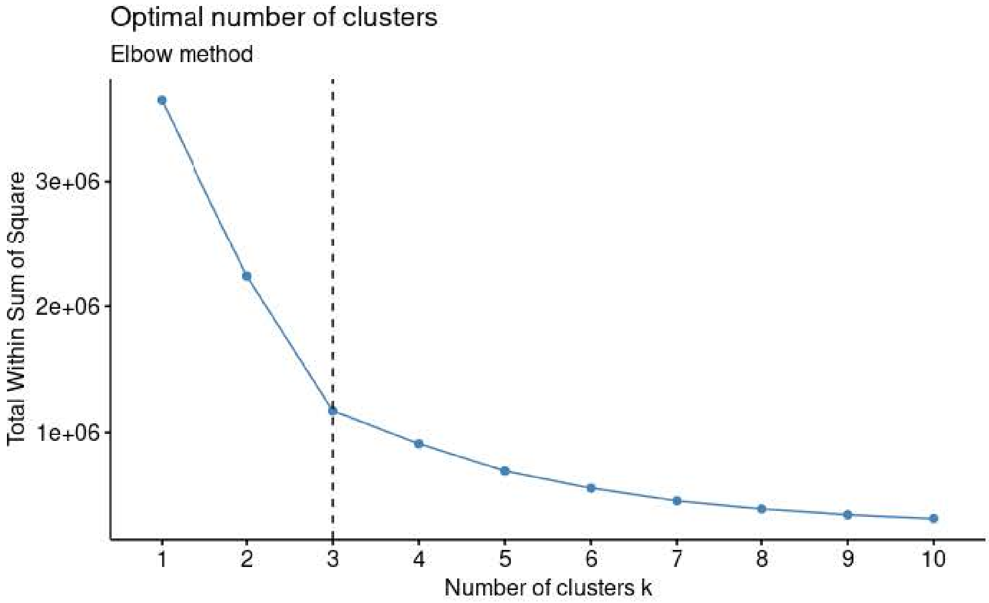


Figure 7.5: Optimal number of clusters using the Elbow method

the three clusters. Each leaf represents a product and leaves are labelled with different colors to represent the three clusters of products. Even though the optimal number of clusters were determined to be three, number of clusters were set as all values between two to twenty to obtain multiple clustering results to be used as initial solutions in the genetic algorithm. In cases where we had more initial solutions than the size of the required initial population, solutions were sorted based on their objective values and only the best solutions were fed to the genetic algorithm.

7.4 Results

Thirteen scenarios were created by setting the value of the utilization threshold as 1, 0.99 or 0.98 and the value of sales creation rate as 1, 0.75, 0.5, 0.25 or 0. A summary of the results corresponding to the best and initial solutions are given in Tables 7.2 and 7.3.

Our complexity model takes into account not only the variety effect of the

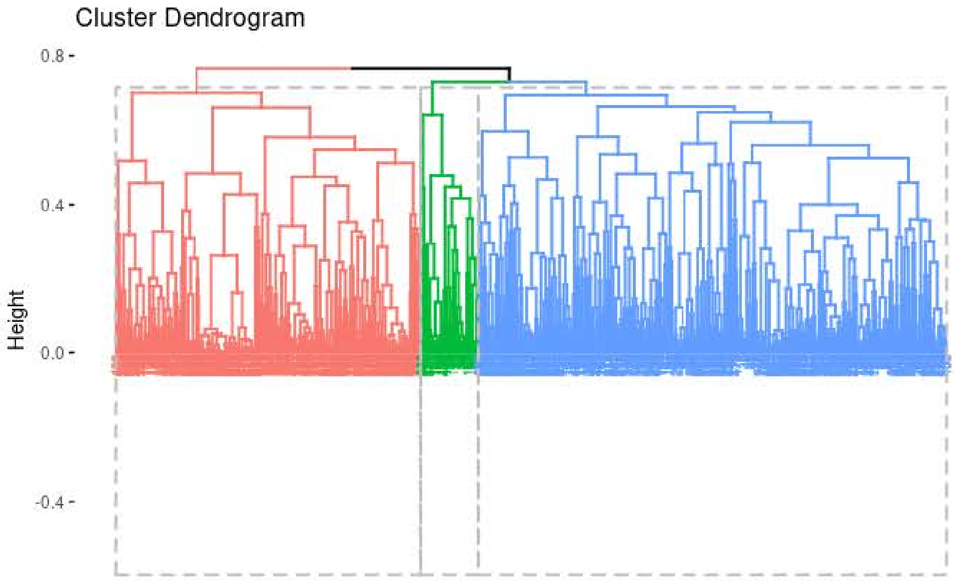


Figure 7.6: Dendrogram plot of the hierarchical clustering results with three clusters

assortment by incorporating the entropy values of many attributes, but also the size of the assortment by incorporating the number of distinct products assigned to each machine and their production times. However, it can be observed that in some scenarios, savings due to additional sales and savings due to conversion cost savings, whose sum gives the value of the complexity cost function, is higher in the best solution despite the smaller number of discontinued products. The reason for this counter-intuitive behavior of the model will be explained using scenario 2 as an example.

In the initial solution, the estimated variety-based break time after 46 products are discontinued in the curing machine PPD11 in October 2019 was 10973.93 minutes, which is more than the estimated variety-based break time after 43 products are discontinued in the best solution, which is 10009.51 minutes. There are seven products assigned to PPD11, and the initial solution suggests discontinuing products 30 and 28, whereas the best solution suggests discontinuing only product 30. When the attributes of the products assigned to PPD11 are examined, it was observed that products 30 and 28 belong to a group with similar attributes, and that group is dominant in number. Discontinuing more products from the dominant group makes the assortment more heterogeneous, resulting in higher entropy values. Entropy alone does not consider the size of the assortment, which is measured by the total production time in the machine. However, the production time of product 28, which is only 33.24 minutes and is small compared to the production time of other products assigned to PPD11 ranging from 4000 to 36000 minutes, is not enough to make up for the increase in entropy when it is discontinued. As a result, a smaller assortment can result in higher variety-based break time estimates.

Some values of the profit loss from sales column in Tables 7.2, 7.3 and 7.4 being negative is a result of some of the discontinued products having high substitution probabilities and lower unit profit compared with other products in its group. In such case, the profits from those groups increase when that product is discontinued. There are also columns that break down the contribution of four terms to the savings from complexity component of the objective function: conversion cost savings due to break time savings, conversion cost savings due to production time savings, additional sales from break time savings and additional sales from production time savings. Negative values that appear in the contribution of additional sales from production time savings column can be explained by discontinued products having high substitution probabilities and lower unit production times compared with other products in their groups.

Overall, the results show that as the utilization threshold decreases or the sales creation rate increases, more products are being discontinued since the weight of break time savings in the objective value increases, as seen in Figure 7.8. Also, the synergy between products becomes more important, and products that cause higher profit loss from sales are being discontinued as this loss is partially reimbursed by the savings from complexity. Note that even if the number of discontinued products is the same across different scenarios and different types of solutions, the products that are being discontinued may be different. Since the initial solution does not favor synergy between products, the difference between the best and the initial solutions in terms of the products that are being discontinued also increase as the utilization threshold decreases or the sales creation rate increases, as Figure 7.7 suggests. As a proof that the contribution of break time savings in the solution is significant, percentage contribution to the complexity savings term of the objective function of break time savings while calculating profit from extra production were shown and compared with the contribution of the production time savings. Extra profit from additional sales makes up most of the savings from complexity cost term and the contribution of the break time savings vary between 0% to 53.8%. Also, the optimal solution of all scenarios when the contribution of break time savings are removed from the objective function are given in Table 7.4. When the break time savings are removed, the problem is no longer a combinatorial problem therefore the initial solution becomes the optimal solution. The change in the number of discontinued products as well as the set of products that are being discontinued substantiates the importance of modelling the complexity costs as a function of the assortment.

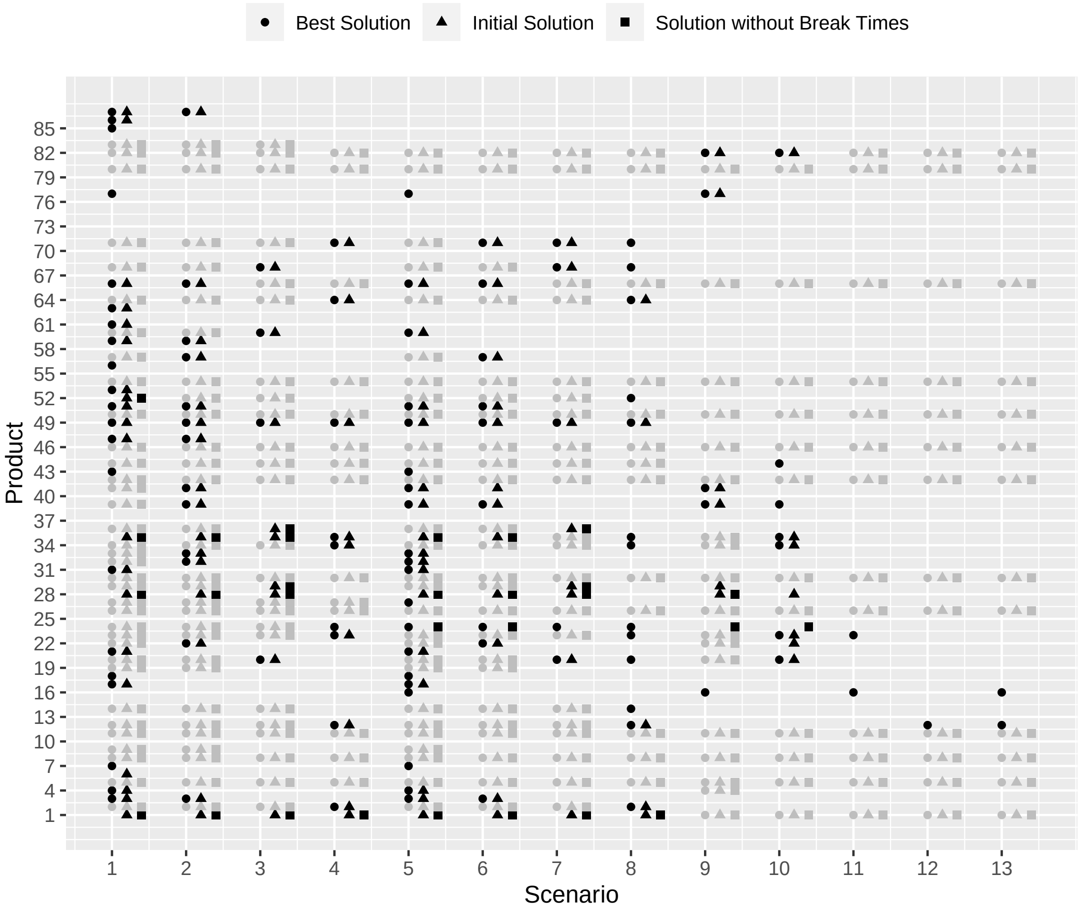


Figure 7.7: Discontinued products for different scenarios

Table 7.2: Summary of the Best Solutions of Different Scenarios

					Objective Value Components			Savings from Complexity Components			Percentage Contribution to Savings from Complexity(%)				
											Conversion Cost Savings due to		Additional Sales from		
Scenario	Utilization Threshold	Sales Creation Rate	Number of Periods With Additional Sales	Number of Discontinued Products	Savings from Inventory Holding Cost*	Profit Loss from Sales*	Savings from Complexity*	Savings due to Time Saved*	Savings due to Additional Sales*	Objective Value*	Break Time Savings	Production Time Savings	Break Time Savings	Production Time Savings	Run Time(h)
1	0.98	1.00	12	56	0.59	758.44	1757.86	60.72	1697.14	1000.00	0.81	2.64	23.25	73.30	1164.19
2	0.98	0.75	12	43	0.68	466.53	1081.66	48.69	1032.97	615.81	1.08	3.42	23.47	72.03	1044.79
3	0.98	0.50	12	28	0.40	80.34	399.69	26.16	373.53	319.74	2.03	4.51	29.11	64.35	992.68
4	0.98	0.25	12	23	0.48	-45.43	130.17	16.28	113.89	176.07	5.43	7.08	36.98	50.51	1157.33
5	0.99	1.00	10	48	0.84	516.78	1290.15	49.72	1240.42	774.20	0.84	3.01	20.73	75.41	1029.02
6	0.99	0.75	10	33	0.41	378.23	857.14	42.27	814.87	479.31	1.03	3.90	19.89	75.18	1016.98
7	0.99	0.50	10	26	0.31	9.07	278.38	21.02	257.36	269.62	2.84	4.71	33.63	58.82	1062.62
8	0.99	0.25	10	18	0.18	-106.74	56.14	8.47	47.67	163.06	9.90	5.18	53.76	31.15	929.36
9	1.00	1.00	2	23	0.10	-19.21	181.36	16.82	164.54	200.68	3.84	5.44	23.32	67.40	1065.90
10	1.00	0.75	2	19	0.14	-45.60	115.09	15.49	99.60	160.82	5.78	7.67	24.18	62.36	1033.31
11	1.00	0.50	2	15	0.15	-111.72	25.24	7.54	17.71	137.12	20.79	9.07	29.93	40.21	954.08
12	1.00	0.25	2	14	0.14	-122.58	8.22	5.80	2.42	130.94	58.65	11.90	146.90	-117.45	1394.46
13	NA	0.00	NA	15	0.14	-122.58	5.80	5.80	0.00	128.52	83.13	16.87	0.00	0.00	948.72

*

*Scaled values

Table 7.3: Summary of the Initial Solutions of Different Scenarios

					Objective Value Components			Savings from Complexity Components			Percentage Contribution to Savings from Complexity(%)			
											Conversion Cost Savings due to		Additional Sales from	
Scenario	Utilization Threshold	Sales Creation Rate	Number of Periods With Additional Sales	Number of Discontinued Products	Savings from Inventory Holding Cost*	Profit Loss from Sales*	Savings from Complexity*	Savings due to Time Saved*	Savings due to Additional Sales*	Objective Value*	Break Time Savings	Production Time Savings	Break Time Savings	Production Time Savings
1	0.98	1.00	12	55	0.64	576.36	1371.97	48.31	1323.66	796.25	1.02	2.50	28.81	67.67
2	0.98	0.75	12	46	0.77	383.80	870.56	39.88	830.69	487.54	1.38	3.20	29.83	65.59
3	0.98	0.50	12	33	0.52	250.62	483.23	31.79	451.44	233.13	1.82	4.76	28.25	65.17
4	0.98	0.25	12	23	0.48	-46.52	128.98	16.13	112.85	175.97	5.47	7.04	37.20	50.29
5	0.99	1.00	10	44	0.74	399.37	1007.66	39.53	968.13	609.03	1.08	2.84	26.45	69.63
6	0.99	0.75	10	36	0.47	302.28	677.09	33.92	643.17	375.28	1.39	3.62	26.31	68.68
7	0.99	0.50	10	29	0.39	233.55	410.45	30.25	380.20	177.29	2.06	5.31	27.56	65.07
8	0.99	0.25	10	18	0.18	-117.84	43.62	6.68	36.93	161.64	11.75	3.57	61.62	23.06
9	1.00	1.00	2	24	0.17	171.77	273.40	25.51	247.89	101.81	2.68	6.66	16.45	74.22
10	1.00	0.75	2	19	0.15	129.76	197.39	24.08	173.31	67.78	3.51	8.69	15.52	72.28
11	1.00	0.50	2	13	0.13	-122.60	10.63	5.79	4.84	133.36	45.24	9.20	227.28	-181.72
12	1.00	0.25	2	13	0.13	-122.60	8.21	5.79	2.42	130.94	58.58	11.91	147.17	-117.66
13	NA	0.00	NA	13	0.13	-122.60	5.79	5.79	0.00	128.52	83.10	16.90	0.00	0.00

*

*Scaled values

Table 7.4: Summary of the Solutions without Break Time Savings

					Objective Value Components			Savings from Complexity Components			Percentage Contribution to Savings from Complexity(%)			
											Conversion Cost Savings due to		Additional Sales from	
Scenario	Utilization Threshold	Sales Creation Rate	Number of Periods With Additional Sales	Number of Discontinued Products	Savings from Inventory Holding Cost*	Profit Loss from Sales*	Savings from Complexity*	Savings due to Conversion Cost Savings*	Savings due to Additional Sales*	Objective Value*	Break Time Savings	Production Time Savings	Break Time Savings	Production Time Savings
1	0.98	1.00	12	39	0.69	344.95	1075.07	36.97	1038.10	730.81	0.92	2.52	26.63	69.93
2	0.98	0.75	12	33	0.57	262.30	717.24	32.44	684.80	455.51	1.22	3.30	27.76	67.72
3	0.98	0.50	12	29	0.59	224.38	450.64	29.65	420.99	226.85	1.82	4.76	27.97	65.45
4	0.99	1.00	10	32	0.47	298.94	879.02	33.69	845.33	580.55	1.02	2.81	25.34	70.83
5	0.99	0.75	10	28	0.27	232.57	590.31	29.68	560.63	358.01	1.35	3.68	26.89	68.08
6	0.99	0.50	10	25	0.42	216.14	387.43	28.53	358.90	171.72	2.02	5.35	26.98	65.65
7	0.99	0.25	10	14	0.15	-122.05	37.86	5.98	31.88	160.05	12.82	2.97	64.54	19.67
8	1.00	1.00	2	20	-0.15	147.01	246.22	22.49	223.73	99.06	2.19	6.95	14.35	76.52
9	1.00	0.75	2	13	-0.16	-110.08	13.11	3.99	9.12	123.02	24.69	5.76	58.14	11.41
10	1.00	0.50	2	13	0.13	-122.60	10.63	5.79	4.84	133.36	45.24	9.20	227.28	-181.72
11	1.00	0.25	2	13	0.13	-122.60	8.21	5.79	2.42	130.94	58.58	11.91	147.17	-117.66
12	NA	0.00	NA	13	0.13	-122.60	5.79	5.79	0.00	128.52	83.10	16.90	0.00	0.00
13	NA	0.00	NA	13	0.13	-122.60	5.79	5.79	0.00	128.52	83.10	16.90	0.00	0.00

*

*Scaled values

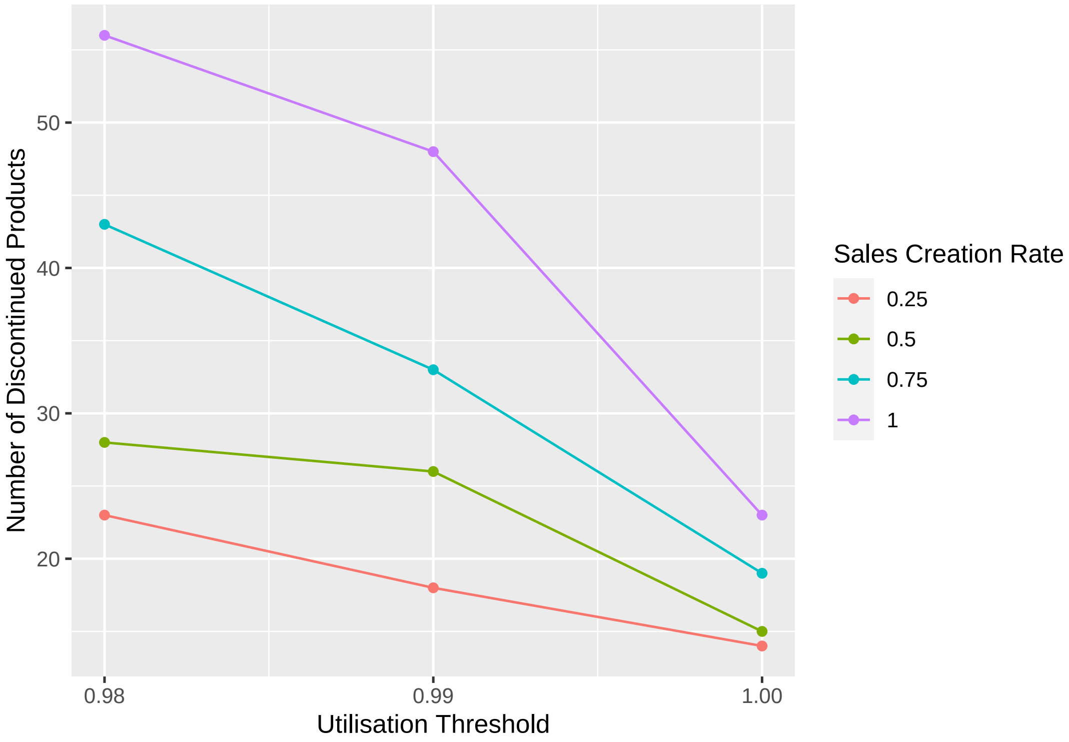


Figure 7.8: Number of discontinued products for different scenarios

Chapter 8

Conclusion

We achieve a reduction in the current assortment by 3.8% - 15.22%. The profit with this reduction is expected to be between 0.59% and 4.59% of the annual sales revenue of the company. We show that the impact of savings from complexity on profits can be as important as the profit loss from sales when the assortment size is reduced using our framework, establishing the benefits of finding a balance between variety and complexity for companies. This thesis is a work that combines statistical analysis with company's insights to find a solution that is evidence based and that best fits to the company's needs. The foundation of our method is to change the product mix by identifying products with low profitability that are difficult to produce in terms of the duration of break times caused by their production, and to eliminate them if it is possible to find evidence from the data that their demand will be substituted by similar products left in the assortment.

We also propose a method to estimate substitution probabilities using a small size of periodic data. We contribute to the literature by showing that estimations can be done using standard methods with aggregated data when needed. We also make a novel definition of complexity cost while estimating the variety-based break time savings using a machine learning model that considers the synergy between products. We show that finding the monetary impact of the capacity savings provided by the elimination of some products by exploiting that capacity to make additional production was the key to quantify the importance of complexity cost savings.

Our framework has a top-down design. It breaks down the assortment planning problem, which has a large scope, into sub-modules and deals with each module separately before integrating the results in an optimization model. We employ a diverse set of methodologies including statistical tools, machine learning, regression models, iterative estimation methods and genetic algorithm. The richness in the variety of methodologies used is what makes this work valuable.

Some limitations of this thesis were identified regarding the estimation process of the substitution probabilities, which can be directions for future research. Sales are modeled as random variables coming from homogeneous Poisson processes and the timings of the stock-outs were modeled as random variables coming from gamma distributions. Different distributions can be tried to model these variables if a better fit can be obtained. A limitation of the β-method is its run time. A faster algorithm to estimate substitution probabilities with limited and censored data can be designed as a future research topic. We have also worked on employing the Bass diffusion model [42] and incorporating substitution into it while estimating substitution probabilities, as in the work of Norton and Bass [43] but we failed to get a good fit. This is possibly due to the nature of our periodic, company level data since the Bass model was usually employed for product categories with industry level data collected in many years. However, adaptations of this model and its applications on larger data sets might be promising. Another improvement to this thesis can be done by considering different optimization methods other than genetic algorithm, such as simulated annealing, that are capable of evaluating a complex objective function, and are possibly faster and more robust.

Bibliography

[1] M. E. Ben-Akiva and S. R. Lerman, “Discrete choice analysis: Theory and application to travel demand,” *MIT Press*, vol. 38, no. 7, pp. 964 – 972, 1985.

[2] M. K. Mantrala, M. Levy, B. E. Kahn, E. J. Fox, P. Gaidarev, B. Dankworth, and D. Shah, “Why is assortment planning so difficult for retailers: A framework and research agenda,” *Journal of Retailing*, vol. 1, pp. 71 – 83, 2009.

[3] M. Fisher, “Don’t trust your gut with assortment planning.” https://hbr.org/2011/11/dont-trust-your-gut-with-assortment-planning, 2011. Accessed: 2021-07-09.

[4] “Assortment planning: Optimizing the right product mix for retail channels.” https://www.riversand.com/blog/assortment-planning-optimizing-product-mix-for-retail-channels. Accessed: 2021-07-09.

[5] S. Chaudhuri, “H&M profit is squeezed as pile of unsold stock grows.” https://www.wsj.com/articles/h-m-profit-is-squeezed-as-pile-of-unsold-stock-grows-1530176871?mod=searchresults_pos8&page=, 2018. Accessed: 2021-07-09.

[6] A. G. Kok and M. L. Fisher, “Demand estimation and assortment optimization under substitution: Methodology and application,” *Operations Research*, vol. 55, pp. 1001 – 1021, 2007.

[7] W. Ross, “The problem with product proliferation.” https://hbr.org/2017/05/the-problem-with-product-proliferation, 2018. Accessed: 2021-07-14.

[8] S. W. Anderson, "Measuring the impact of product mix heterogeneity on manufacturing overhead cost," *The Accounting Review*, vol. 70, no. 3, pp. 363 – 387, 1995.

[9] J. T. Gourville and D. Soman, "Overchoice and assortment type: When and why variety backfires," *Marketing Science*, vol. 24, no. 3, pp. 382 – 395, 2005.

[10] P. Boatwright and J. C. Nunes, "Reducing assortment: An attribute-based approach," *Journal of Marketing*, vol. 65, pp. 50 – 63, 2001.

[11] H. Shin, S. Park, E. Lee, and W. C. Benton, "A classification of the literature on the planning of substitutable products," *European Journal of Operational Research*, vol. 246, no. 3, pp. 689 – 699, 2015.

[12] A. G. Kok, M. L. Fisher, and R. Vaidyanathan, "Assortment planning: Review of literature and industry practice," *Retail Supply Chain Management*, vol. 122, pp. 99 – 153, 2009.

[13] J. Ward, B. Zhang, S. Jain, C. Fry, T. Olavson, H. Mishal, J. Amaral, D. Beyer, A. Brecht, B. Cargille, R. Chadinha, K. Chou, G. DeNyse, Q. Feng, C. Padovani, S. Raj, K. Sunderbruch, R. Tarjan, K. Venkatraman, J. Woods, and J. Zhou, "HP transforms products portfolio management with operations research," *Interfaces*, vol. 40, no. 1, pp. 17 –32, 2010.

[14] "Run flat tyres: Everything you need to know." https://www.holtsauto.com/holts/news/run-flat-tyres-everything-need-know/. Accessed: 2021-07-29.

[15] "What's new with eu tyre labelling?." https://www.goodyear.eu/en_gb/consumer/learn/eu-tire-label-explained.html, 2021. Accessed: 2021-07-29.

[16] P. Rusmevichientong, Z. J. M. Shen, and D. B. Shmoys, "Dynamic assortment optimization with a multinomial logit choice model and capacity constraint," *Operations Research*, vol. 58, no. 6, pp. 1666 – 1680, 2010.

[17] A. Şen, A. Atamtürk, and P. Kaminsky, "A conic integer optimization approach to the constrained assortment problem under the mixed multinomial logit model," *Operations Research*, vol. 66, no. 4, pp. 994 – 1003, 2018.

[18] J. M. Davis, G. Gallego, and H. Topaloglu, “Assortment optimization under variants of the nested logit model,” *Operations Research*, vol. 62, no. 2, pp. 250 – 273, 2014.

[19] X. Wang, J. D. Camn, and D. J. Curry, “A branch-and-price approach to the share-of-choice product line design problem,” *Management Science*, vol. 55, no. 10, pp. 1718 – 1728, 2009.

[20] A. Belloni, R. Freund, M. Selove, and D. Simester, “Optimizing product line designs: Efficient methods and comparisons,” *Management Science*, vol. 54, no. 9, pp. 1544 – 1552, 2008.

[21] K. Talluri and G. V. Ryzin, “Revenue management under a general discrete choice model of consumer behavior,” *Marketing Science*, vol. 50, no. 1, pp. 15 – 33, 2004.

[22] S. Subramanian and P. Harsha, “Demand modeling in the presence of unobserved lost sales,” tech. rep., IBM T. J. Watson Research Center, 2017.

[23] J. P. Newman, M. E. Ferguson, L. A. Garrow, and T. L. Jacobs, “Estimation of choice-based models using sales data from a single firm,” *Manufacturing & Service Operations Management*, vol. 16, no. 2, pp. 184 –197, 2014.

[24] G. Vulcano, G. van Ryzin, and R. Ratliff, “Estimating primary demand for substitutable products from sales transaction data,” *Operations Research*, vol. 60, no. 2, pp. 313 – 334, 2012.

[25] T. Abdallah and G. Vulcano, “Demand estimation under the multinomial logit model from sales transaction data,” *Manufacturing & Service Operations Management*, 2020. Published Online.

[26] A. Li and K. Talluri, “Estimating demand with unobserved no-purchases on revenue-managed data,” 2020. Working Paper.

[27] M. Fisher and R. Vaidyanathan, “A demand estimation procedure for retail assortment optimization with results from implementations,” *Management Science*, vol. 60, no. 10, pp. 2401 – 2415, 2014.

[28] T. H. Yunes, D. Napolitano, A. Scheller-Wolf, and S. Tayur, "Building efficient product portfolios at John Deere and Company," *Operations Research*, vol. 55, no. 4, pp. 615–629, 2007.

[29] M. Shunko, T. Yunes, G. Fenu, A. Scheller-Wolf, V. Tardif, and S. Tayur, "Product portfolio restructuring: Methodology and application at Caterpillar," *Production and Operations Management*, vol. 27, no. 1, pp. 100 –120, 2017.

[30] J. Chong, T. Ho, and C. S. Tang, "A modelling framework for category assortment planning," *Manufacturing & Service Operations Management*, vol. 3, no. 3, pp. 191 – 210, 2001.

[31] P. Guadagni and J. Little, "A logit model of brand choice calibrated on scanner data," *Marketing Science*, vol. 2, no. 3, pp. 203 – 238, 1983.

[32] E. Rash and K. Kempf, "Product line design and scheduling at Intel," *Interfaces*, vol. 42, no. 5, pp. 425 – 436, 2012.

[33] R. Anupindi, M. Dada, and S. Gupta, "Estimation of consumer demand with stock-out based substitution: An application to vending machine products," *Marketing Science*, vol. 17, no. 4, pp. 406 – 423, 1998.

[34] E. LeDell, N. Gill, S. Aiello, A. Fu, A. Candel, C. Click, T. Kraljevic, T. Nykodym, P. Aboyoun, M. Kurka, and M. Malohlava, *h2o: R Interface for the 'H2O' Scalable Machine Learning Platform*, 2020. R package version 3.32.0.1.

[35] A. Boulangé, *automl: Deep Learning with Metaheuristic*, 2020. R package version 1.3.2.

[36] G. E. P. Box and D. R. Cox, "An analysis of transformations," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 26, no. 2, pp. 211 – 252, 1964.

[37] W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S.* New York: Springer, fourth ed., 2002. ISBN 0-387-95457-0.

[38] L. Scrucca, "On some extensions to GA package: hybrid optimisation, parallelisation and islands evolution," *The R Journal*, vol. 9, no. 1, pp. 187–206, 2017.

[39] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning with Applications in R.* Springer, 2013. ISBN 978-1-4614-7138-7.

[40] M. Maechler, P. Rousseeuw, A. Struyf, M. Hubert, and K. Hornik, *cluster: Cluster Analysis Basics and Extensions*, 2021. R package version 2.1.2 — For new features, see the 'Changelog' file (in the package source).

[41] A. Kassambara and F. Mundt, *factoextra: Extract and Visualize the Results of Multivariate Data Analyses*, 2020. R package version 1.0.7.

[42] F. M. Bass, "A new product growth for model consumer durables," *Management Science*, vol. 15, no. 5, pp. 215 – 227, 1969.

[43] J. A. Norton and F. M. Bass, "A diffusion theory model of adoption and substitution for successive generations of high-technology products," *Management Science*, vol. 33, no. 9, pp. 1069 – 1086, 1987.