İPEK DURMUŞ

M.S. Thesis

2017

AN ASSORTMENT PLANNING PROBLEM WITH AN EMPIRICAL DEMAND MODEL

İPEK DURMUŞ

IŞIK UNIVERSITY

2017

AN ASSORTMENT PLANNING PROBLEM WITH AN EMPIRICAL DEMAND MODEL

İPEK DURMUŞ
B.S., Industrial Engineering, IŞIK UNIVERSITY, 2014

Submitted to the Graduate School of Science and Engineering
in partial fulfillment of the requirements for the degree of
Master of Science
in
Industrial Engineering

IŞIK UNIVERSITY
2017

IŞIK UNIVERSITY
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

AN ASSORTMENT PLANNING PROBLEM WITH AN EMPIRICAL DEMAND MODEL

İPEK DURMUŞ

APPROVED BY:

Assoc. Prof. S. Çağlar Aksezer (Thesis Supervisor)	Işık University	
Assist. Prof. Mustafa Hekimoğlu	Işık University	
Assoc. Prof. Gül Tekin Temur	Beykent University	

APPROVAL DATE: 18/01/2017

AN ASSORTMENT PLANNING PROBLEM WITH AN EMPIRICAL DEMAND MODEL

Abstract

In retail industry, although the product range increases with a great pace year by year, the shelf space that products are displayed on does not expand with the same pace. Accordingly, in recent years, placing products efficiently on a limited shelf space become an important problem for maximizing sales. Assortment planning constitutes the product range which is presented for sale in the store and the inventory amounts of these products.

The goal of this study is generating a mathematical programming model that can be solved in a reasonable time and maximizes the profit of a retailer. A regression model for customer demand is formed by using a data which is obtained from a supermarket chain in Turkey. An empirical demand model is used as an input for the nonlinear optimization model. The assortment optimization model that is developed for this study identifies the products which return the maximum profit on a large product range, and it determines the best facing amounts of these products.

Keywords: Assortment planning, nonlinear programming, empirical demand model

DENEYSEL BİR TALEP MODELİ YARDIMIYLA ÜRÜN ÇEŞİDİ PLANLAMA PROBLEMİ

Özet

Parekende sektöründe ürün çeşitliliği yıldan yıla büyük bir hızla artış gösterirken, mağazalarda ürünlerin sergilendiği raf alanlarının, kira ve yeni mağaza tasarımı gibi maliyetlerin fazlalığının etkisiyle, çok az bir miktarda arttığı görülmektedir. Dolayısıyla, ürünlerin sınırlı raf alanına, satışı ençoklayacak en etkili şekilde yerleştirilmesi önemli bir problem haline gelmiştir. Mağazada satılacak ürün çeşitlerinin ve ürünlerin her birinin envanter miktarlarının ayarlanmasını sağlayan ürün çeşidi planlama yöntemleri günümüzde oldukça önem kazanmıştır.

Bu çalışmanın amacı; raf alanı kısıtı altında satıcının kârını maksimize eden ve makul bir sürede çözülebilir olan matematiksel programlama modeli oluşturmaktır. Türkiye'deki bir süpermarket zincirinden alınan veriler kullanılarak müşteri talebi için regresyon modeli oluşturulmuştur. Elde edilen deneysel model, doğrusal olmayan bir programlama modelinde girdi olarak kullanılmıştır. Optimizasyon modeli geniş bir ürün yelpazesi içinden maksimum karlılığı sağlayacak ürünleri ve bu ürünlere atanacak raf önyüz miktarlarını vermektedir.

Anahtar kelimeler: Ürün çeşidi planlama, doğrusal olmayan programlama, ampirik talep modeli

Acknowledgements

First of all, I would like to express my eternal gratitude to my project supervisors, Assoc. Prof. S. Çağlar Aksezer and Assist. Prof. Mustafa Hekimoğlu, for their great interest, patience and profoundly support to complete this study. Working with them was an important opportunity for me. I learned a lot of things thanks to their guidance, and I owe to them for this invaluable knowledge. Also, I would like to thank Assist. Prof. Mehmet Önal for his helps about this project, he always answered to all my questions carefully with a great concern.

Also, I express my gratitude to the foundation, Işık University, that provided scholarship for my master education and Industrial Engineering Department Staff who have presented a large source of information for my vocational development.

Additionally, I am extremely grateful to my family who have an important role on my personal development and my successes. I thank my parents, Mustafa Durmuş and Mukadder Durmuş, and my fiancé, Taner Aslan, for their endless support. They have always trusted and encouraged me. Also, I would like to thank my brother, Orkun Durmuş, who is my idol because of his life which is full of successes. He lost his life five years ago, but I have felt his support and power during my whole education life.

To my family

Table of Contents

List of Tables

List of Figures

List of Abbreviations

KKT	**K**arush **K**uhn **T**ucker
LSL	**L**imit **S**helf **L**ength
MAPE	**M**ean **A**bsolute **P**ercentage **E**rror
SD	**S**helf **D**epth
SKU	**S**tock **K**eeping **U**nit
ULF	**U**pper **L**imit **F**acing

Chapter 1

Introduction

1.1 Problem Statement and Research Objective

It is known that there is huge variety of products in a store that belongs to a supermarket chain which established in market presence. Supermarkets not only offer food need of consumers, and they also present extensive product range for other human needs such as personal care, clothing, technology, hobby, household goods etc. When it is thought about all products in all categories are supplied from a lot of different producers and different enterprises, it can be seen that supermarkets transmit thousands of products to consumers. Also, consumers tastes and necessities are highly differentiated, and supermarkets need detailed product range in order to serve more customers. For instance, in some supermarkets there are product aisles which are prepared for people who have gluten sensitivity. Another example for this support is that there are dozens of varieties of body lotion for consumers who have different skin types in cosmetics section of a supermarket.

In the past, talking about such a huge variety of products was not possible, and during the recent years, the advancement of technology and increasingly globalized world brought a heavy increase of the number of producer firms and the number of products which are produced. It is known fact that supermarket chains succumbed to "more is better" trend and increased the number of goods on their shelves in order to provide service to a wider variety of customers. According to a survey that was done by Food Marketing Institute, in 1998, nearly 47,000 distinct items filled a typical supermarket retailer's shelving, up more than 50% from 1996 [1]. Also, according to the study of Quelch and Kenny, the number of goods in the supermarkets increased by 16% per year between 1985 and 1992, while shelf space expanded by only 1.5% per year in the same period [2]. The pace of extension rate of shelf space remain incapable by comparison with the pace of increasing rate of the number of products, and the shelving of a supermarket are inadequate for containing all goods. Because of this situation, assortment planning for an efficient shelf space usage became unavoidable.

Assortment planning is designating the set of products which are presented at each store and specifying the inventory levels of these products for maximizing profit subject to shelf space and other possible constraints [3]. Retailers are obliged to make assortment planning for their stores and make tiny distinctions for their shelf space management because of the enhancement of products which competes for restricted shelf space. In our assumption, display spaces or facings are highly correlated with the demand rate, and more visibility provides additional demand for each product. More visibility means more attractiveness, and in a supermarket, consumers psychologically go towards the products which are placed with more facings.

In consideration of these observations, the main goal of this thesis is to constitute an assortment optimization model that maximizes revenue for a single store.

In this study, we use real data that belongs to a supermarket chain in Turkey. There are tens of thousands stock keeping units and hierarchically all the products are divided into groups, categories and subcategories in this store. Shelving is divided into facings and in a facing, only one type of product can be placed. The depth of the shelf and the physical size of one unit of product specify the capacity of that facing. In our study, there is no storeroom, and all the products are placed directly on the shelving.

We aim to develop an assortment planning solution which is based on an empirical demand model. Therefore, generating the demand model is the first step of our study, and it can be formed by using regression analysis on real data. After the demand model is constructed and validated, an assortment optimization model which gives the optimal facing amounts for maximum revenue are generated by using the demand model. In order to observe the applicability of the model and determine solution methods, some analysis and numerical experiments are performed.

1.2 Motivational Case Study

Shelving of supermarkets should be organized as convenient with demands of customers for maximizing sales. In retail industry, a product is not positioned behind the other product on shelf, and the maximum product amount on the shelf is the function of facing that is reserved for this product and depth of the shelf. There are thousands of products in every store, so optimizing the facing amounts to maximize profit is a considerable complex problem. The Figure 1.1 provides a better understanding for the importance of assortment optimization problem.

Perishable products are displayed in refrigerators in every supermarket. In this example, a product which is named as "DANONE ACTIVIA KURU KAYISI ($4\times$ 110gr)" is presented for sale on the top shelf with the facing amount approximately fifteen. The average sales of that product was determined as 0.483. The list price of that product is 4.55 TL, and the average endorsement which is provided by this product is 2.06 TL. On the other hand, a product which is named as "MIS AYRAN BARDAK (230ml)" is presented for sale on bottom-right corner of the refrigerator with the facing amount four. The average sales of this product was determined as 36. The list price of that product is 0.5 TL, and the average endorsement which is provided by this product is 18.04 TL. As shown in Figure

Figure 1.1: Assortment arrangement of a supermarket for perishable products

1.1, the area that is reserved for the product which bring 2.06 TL in a day is three times of the area that is reserved for the product which bring 18 TL in a day.

It can be said that this situation can cause some important problems. For instance, “DANONE ACTIVIA KURU KAYISI (4×110gr)” is a perishable product, and it has ten days shelf life. Although this product has large shelf space capacity, it is not demanded sufficiently. Then, the extinguishment probability of this product may increase, and the retailer can lose money because of this situation. Another example for these problems is that “MIS AYRAN BARDAK (230ml)” can be stock-out long before the replenishment time. Because, this product has a small shelf capacity, although the demand of this product is high. These examples show the importance of assortment planning in a supermarket.

Consequently, the assortment optimization is a difficult problem which is vital for maximizing endorsement and profit of retailers. The facing amounts which are assigned to the products directly influence to marginal costs and stock out conditions as well as shaping the demand on products.

1.3 Outline of the Thesis

This thesis is organized as follows. In Chapter 2, we present a review of the existing literature which is relevant to the our study. In Chapter 3, the data that is used in this thesis is described in detail, and the empirical demand model is generated by regression. In Chapter 4, we present the assortment optimization model that maximizes the total revenue under the constraint of shelf space and the application of this model. Finally, we conclude our work in Section 5.

Chapter 2

Literature Survey

In recent years, assortment planning became an important subject for service sector with emergent requirements.

One of the main studies which is about assortment planning was accomplished by Kök and Fisher [3]. They define a specific methodology to estimate the input demand and substitution parameters for the assortment planning problem and assert an optimization algorithm. Their procedure was applied at a leading supermarket chain in the Netherlands. Customers who do not find their favorite product in a store can buy another product which is similar, and this is named as *substitution.* If a product, which is normally in assortment, is stocked out when it is demanded, customers substitute with another similar product. This situation is called *stock-out based substitution.* Also, customers may substitute because of their favorite product is not in assortment of the store, and this substitution type is called *assortment based substitution.* They indicate two methodologies in order to estimate demand and parameters of the substitution model. In the first method, stock-out based substitution is neglected because of high service levels, and only assortment based substitution is considered. Substitution rates and demand are estimated by using sales data from different stores. In the second method, substitution rates and demand are estimated by using inventory transaction data, and it generalizes their approach to the case with stock-out based substitution. Also, they develop an iterative heuristic which can solve a series of separable nonlinear knapsack problems. The method that was developed by Kök and Fisher provides more than a 50% increase in profits.

In general, the same assortment is used in all stores of a retailer, and only some products are eliminated in smaller stores. Fisher and Vaidyanathan [4] state that the assortment should vary according to local tastes for each store. They provide improvement of assortment localization, allowing a constraint on the number of different assortments and quantifying the level of localization effects on revenue. Also, they estimate the demand for new products which have not been presented for sale before in any store, and the demand is estimated by using past sales of products which are currently exist in assortment. They also develop a demand model that suits a case which some products are more preferable substitutes for a given product than others.

According to Gilland and Heese [5], the sequence of customer arrivals has an important effect on profitability. Limited shelf space restricts the number of product in a store, and the customers who arrive to store sequentially have to make an immediate substitution decision based on the current product availability. They also emphasize that stock-out based substitution can cause a hidden cost, although a product is sold. Because, the customers who do not find their favorite product become disgruntled, and they may decide not to come to the store anymore.

Ulu, Honhon and Alptekinoğlu [6] state that better assortment planning can be done, if consumer tastes are known. In order to learn consumer tastes, firms use different methods such as evaluating past sales data and market surveys. Hotelling, that is locational choice model, is used in their study, and a discrete set of consumer locations represents consumer preferences. In order to model data collection process, a Bayesian framework is used, and the firm's knowledge on consumer tastes is updated. In order to show consumer tastes, Honhon et al. [7] use a ranking based consumer choice model. According to this model, every consumers have rankings of the potential products, and they buy their highest product offered in the assortment. In a different study which is performed by

Honhon and Seshadri [8], tastes of customers are characterized by their type, and type is a list of products which the customers are willing to buy in decreasing order of preference. Substitutions are consumer driven, dynamic and stock-out based in this study. They point out that the optimum profit under fixed proportion is always greater or equal to the optimum profit under random proportions. In another study, Golrezaei et al. [9] evaluate the problem that personalizes the assortment of products for every customer who arrives, and the availability of real time data about the characteristic of customer is necessary for this procedure. For example, the product advices, which Amazon.com performs on each customer, dynamically change depending some factors such as the previous purchases of customer, recent reviews, purchases of another customer who has similar tastes. They claim that increment in revenue is possible by personalizing the assortment. However, this technique can only be performed by online retailers.

Multinomial logit model is commonly used consumer choice model in the literature. One of these studies, which considers multinomial logit model as demand model, is completed by Rusmevinhientong et al. [10]. They work on both dynamic and static optimization problems, and they assume that the parameters of multinomial logit model are known for the static problem. However the parameters are not known and should be estimated in a dynamic problem. A joint stocking and product offer problem is studied by Topaloglu [11]. Multinomial logit model is used to determine consumer choice structure. The goal of his study is deciding which product sets to offer and how many unit of each product is stocked in order to maximize profit. Then, he develop a nonlinear programming model, and the decision variables are the stock amount for each product and the duration of time which each product is offered. In the study of Rusmevichientong et al. [12], an assortment optimization problem is considered under the multinomial logit model, and the parameters of the multinomial logit model are random. The reason of randomness is that every consumer has different tastes for the products. Rushmevichientong et al. state that the problem is nonlinear even when two customer segments exist. They point out that assortments are composed of the products which provides highest revenue, and this is called as *revenue-ordered assortments*.

A consumer choice model is developed by using the nested multinomial logit framework with two different hierarchical structures by Kök and Xu [13]. One of these hierarchical structures is *brand-primary model* in which customers initially choose a brand, then a product type is chosen from that brand. The other case is *type-primary model* in which customers initially choose a product type, then a brand is chosen for this product type. They indicate that with the brand-primary model, the most popular product types from the brand constitute the competitive and the optimal assortments for associated brand. However, the competitive and the optimal assortments for each brand may not always compose of the brand's most popular product for the type-primary model. Feldman and Topaloglu [14] study about assortment optimization problem, and the customers demand is learned by the nested logit model. The capacity is limited for the products in assortment. In the nested logit model, the products are grouped and organized in nests. A customer, who comes to store, can decide to make a purchase from one of the nest or leave without purchasing any product. If a nest is chosen, the customer buys one of the product in that nest. They consider two types of capacity constraints for the assortment optimization problem. These are cardinality constraint and space constraint. They develop an algorithm in order to obtain approximate solution. As differently from previous work, Topaloglu et al. [15] use d-level nested logit model to learn consumer preferences, and d-level nested logit model is described as a tree of depth d. They develop an efficient algorithm in order to find the optimal assortment.

Assortment of products and consumer return policies are generally accepted as separate fields of the retailing business. However, Alptekinoğlu and Grasas [16] assert the counter-view by demonstrating that the optimum assortment decisions are different when consumer returns are considered. Optimum assortment consists of a mix of the most popular products which have high attractiveness and the most eccentric products which have low attractiveness, when the consumer returns are considered. The most eccentric products exist in assortment, because they provide sufficiently high restocking fee with the high probability of return.

A product which exists in a store can be stale and lose its attractiveness over time. Caro et al. [17] state that products have different preference weights, profit margins and life cycle patterns. The trade-offs among preference weights, profit margins and limited life cycle are modeled by the formulation of them. The goal of their study is determining the optimum time for introducing each product to store in order to maximize profit.

Chapter 3

Data Analysis and Demand Estimation

Data collection and estimating demand by using the data is the first stage for assortment planning. In this chapter, by making regression analysis, we develop an empirical demand model which will be used in assortment optimization model, and the demand model is analyzed for validation and applicability.

3.1 Data Collection and Analysis

In order to plan assortment, we are allowed to access database of a supermarket in Turkey, and the original data comes from ORACLE database. In this database, we have the remaining stock, the facing quantities, the sales quantities and the prices for every product of all product groups for each day.

In the supermarket, products are divided into different groups such as food, non-food, fruit and vegetable products, personal care products. These product groups branch out categories, subcategories and stock keeping units hierarchically. As might be expected, the data that we work on is enormous and consists of 3,421,500 rows. There are 40,009 types of products and 677 product groups in this supermarket. Therefore, in the first stage of our study, a sample that consists of only ten stock keeping units from four different types of subcategories were used for examination of the optimization model in conjunction with the demand model. In that sample, two SKUs from paprika paste, three SKUs from chopped tomato, two SKUs from tomato puree and three SKUs from tomato paste are used as an example.

In our study, the daily data is grouped as weekly data for convenience in analysis by reducing the data size. Totally, there are twenty-six weeks in the data, but we only use the data of thirteen weeks because of missing data problem of the other thirteen weeks. The weekly averages of remaining stocks and sales quantities for every product are designated, and the facing quantities and the prices of every product remain the same, because they are constant values for each product.

3.2 Demand Estimation by Regression

In order to construct the optimization model that maximizes total revenue, we need to know the prices and demands of the products. The prices of all products are known in the data, but demands of the products are not known, and they should be estimated.

In our study, Minitab, which is a statistics software, is used for estimating demands of all products by conducting a regression analysis. Product IDs, product group IDs, remaining stocks, the facing quantities are accepted as the predictor values for demand of products. The regression model is one of the constraints of the optimization model as the demand equation.

Two type of regressions, which are linear regression and log-linear regression, are compared to understand which one is suitable for this study. Linear regression assumes a linear functional relationship between a dependent variable Y, independent variable $x_i, i \in \{1, \cdots, I\}$ and a random term ε by fitting a linear function.

$$Y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_I x_I + \varepsilon \tag{3.1}$$

where α_0 is the constant term, $\alpha_i, i \in \{1, \cdots, I\}$ are the regression coefficients of independent variables [18].

The other regression technique that is considered is log-linear regression, and it models relationship between a dependent variable Y, independent variable $x_i, i \in \{1, \cdots, I\}$ and a random term ε by fitting a log-linear model.

$$\ln Y = \alpha_0 + \alpha_1 \ln x_1 + \alpha_2 \ln x_2 + \cdots + \alpha_I \ln x_I + \varepsilon \tag{3.2}$$

where α_0 is the constant term, $\alpha_i, i \in \{1, \cdots, I\}$ are the regression coefficients of independent variable [19].

The linear demand model is prevalent due to its simple functional form. Also, it is easy to estimate from data by using linear regression techniques. However, there is a negative feature of the linear regression. It may produce negative demand values, and this situation can cause numerical difficulties when solving an optimization problem. Also, it is unrealistic. Unlike the linear demand model, in the log-linear demand model, demand is always nonnegative. Moreover, it can be recovered to linear form by taking the logarithm of demand, so it is well suited to estimation by using linear regression [20].

The log-linear demand model is generally preferred in the existing researches because of all these reasons, so we decided to use log-linear demand model in this study. At the next step of our study, different regression analyses are conducted by using the predictor variables diversely for accurate assessment of which demand model gives the best fits. In these regression analyses that are performed for deciding demand model, we use the sample data set which is mentioned in Section 3.1. This sample data set only contains ten products for a time period of thirteen weeks.

3.2.1 Analysis of different regression models based on a sample data

In order to determine the demand model which gives the best fits, we try different regressions on a sample data which contains only ten products, and the results of these regressions are compared with the coefficient of determination, that is denoted as R^2, values.

R^2 values and p-values are significant indicators for reliability of regression. While R^2 closes to a hundred percent and p-values become smaller than 0.05, it can be said that the regression becomes more reliable. All regression analyses that are conducted on this sample data are explained in detail in this section.

3.2.1.1 The first regression model for estimating demand

Firstly, a log-linear regression is conducted in order to estimate demand by using *RemainingStock* and *Facing* parameters as predictors. The results of this regression are shown in Table 3.1, Table 3.2 and Equation 3.3.

Table 3.1: Coefficients of the first regression equation

Term	**Coefficient**	**P-value**
Constant	-1.957	0.000
RemainingStock	0.010	0.000
Facing	0.080	0.001

Table 3.2: Model summary of the first regression equation

R-sq	**R-sq(adj)**	**R-sq(pred)**
70.48%	69.75%	66.93%

The first regression equation is

$$\ln Demand = -1.957 + (0.010 \times RemainingStock) + (0.080 \times Facing) \quad (3.3)$$

According to the first regression, p-values of *RemainingStock* and *Facing* are smaller than 0.05, so these predictors are significant for demand estimation. The adjusted R^2 value is 0.6975, and it demonstrates that the regression model explains 69.75% of variance of customer demands.

3.2.1.2 The second regression model for estimating demand

Secondly, a log-linear regression is conducted in order to estimate demand by using *IDProduct*, *IDProductGroup*, *RemainingStock* and *Facing* parameters as predictors. The results of this regression are shown in Table 3.3, Table 3.4 and Equation 3.4.

Table 3.3: Coefficients of the second regression equation

Term	**Coefficient**	**P-value**
Constant	-2.613	0.000
IDProductGroup	1.075	0.002
IDProduct	-0.332	0.014
RemainingStock	0.008	0.000
Facing	0.063	0.004

Table 3.4: Model summary of the second regression equation

R-sq	**R-sq(adj)**	**R-sq(pred)**
74.06%	72.82%	70.36%

The second regression equation is

$$\ln Demand = -2.613 + (1.075 \times IDProductGroup) - (0.332 \times IDProduct) + (0.008 \times RemainingStock) + (0.063 \times Facing) \tag{3.4}$$

According to the second regression, p-values of all predictor variables are smaller than 0.05, so these predictors are significant for demand estimation. The adjusted R^2 value is 0.7282, and it demonstrates that the regression model explains 72.82% of variance of customer demands.

3.2.1.3 The third regression model for estimating demand

A log-linear regression is conducted in order to estimate demand by using *IDProduct*, *IDProductGroup*, $\ln RemainingStock$, $\ln Facing$ and $\ln Price_i$ parameters as predictors, while $Price_i$ represents the price of product i, $i \in \{1, 2, 3, \cdots, 10\}$. The results of this regression are shown in Table 3.5, Table 3.6 and Equation 3.5.

Table 3.5: Coefficients of the third regression equation

Term	**Coefficient**	**P-value**
Constant	-4.940	0.002
IDProductGroup	1.119	0.039
IDProduct	-0.255	0.158
$\ln RemainingStock$	0.191	0.044
$\ln Facing$	1.417	0.000
$\ln Price_1$	-0.365	0.663
$\ln Price_2$	-0.535	0.481
$\ln Price_3$	0.350	0.807
$\ln Price_4$	-1.561	0.034
$\ln Price_5$	0.025	0.954
$\ln Price_6$	0.285	0.672
$\ln Price_7$	-2.750	0.012
$\ln Price_8$	-3.300	0.003

Table 3.6: Model summary of the third regression equation

R-sq	**R-sq(adj)**	**R-sq(pred)**
63.40%	60.33%	56.33%

The third regression equation is

$$\begin{aligned}
\ln Demand &= -4.940 + (1.119 \times IDProductGroup) - (0.255 \times IDProduct) \\
&+ (0.191 \times \ln RemainingStock) + (1.417 \times \ln Facing) \\
&- (0.365 \times \ln Price_1) - (0.535 \times \ln Price_2) + (0.350 \times \ln Price_3) \\
&- (1.561 \times \ln Price_4) + (0.025 \times \ln Price_5) + (0.285 \times \ln Price_6) \\
&- (2.750 \times \ln Price_7) - (3.300 \times \ln Price_8)
\end{aligned} \tag{3.5}$$

According to the third regression, p-values of $IDProductGroup$, $\ln RemainingStock$, $\ln Facing$, $\ln Price_4$, $\ln Price_7$ and $\ln Price_8$ are smaller than 0.05, so these predictors are significant for demand estimation. The adjusted R^2 value is 0.6033, and it demonstrates that the regression model explains 60.33% of variance of customer demands.

3.2.1.4 The fourth regression model for estimating demand

A log-linear regression is conducted to estimate demand by using $\ln IDProduct$, $\ln IDProductGroup$, $\ln RemainingStock$, $\ln Facing$ and $\ln Price_i$ parameters as predictors, while $Price_i$ represents the price of product i, $i \in \{1, 2, 3, \cdots, 10\}$. The results of this regression are shown in Table 3.7, Table 3.8 and Equation 3.6.

Table 3.7: Coefficients of the fourth regression equation

Term	**Coefficient**	**P-value**
Constant	-10.260	0.000
$\ln RemainingStock$	0.080	0.250
$\ln Facing$	1.494	0.000
$\ln IDProductGroup$	4.350	0.000
$\ln IDProduct$	0.503	0.082
$\ln Price_1$	4.122	0.000
$\ln Price_4$	2.006	0.000

Table 3.8: Model summary of the fourth regression equation

R-sq	**R-sq(adj)**	**R-sq(pred)**
53.77%	51.90%	49.19%

The fourth regression equation is

$$\ln Demand = -10.260 + (0.080 \times \ln RemainingStock) + (1.494 \times \ln Facing)$$
$$+ (4.350 \times \ln IDProductGroup) + (0.503 \times \ln IDProduct)$$
$$+ (4.122 \times \ln Price_1) + (2.006 \times \ln Price_4) \tag{3.6}$$

According to the fourth regression, p-values of $\ln IDProductGroup$, $\ln Facing$, $\ln Price_1$ and $\ln Price_4$ are smaller than 0.05, so these predictors are significant for demand estimation. The adjusted R^2 value is 0.5190, and it demonstrates that the regression model explains 51.90% of variance of customer demands.

3.2.1.5 The fifth regression model for estimating demand

A log-linear regression is conducted in order to estimate demand by using *IDProduct*, *IDProductGroup*, *RemainingStock*, *Facing* and $\ln Price_i$ parameters as predictors, while $Price_i$ represents the price of product i, $i \in \{1, 2, 3, \cdots, 10\}$. The results of this regression are shown in Table 3.9, Table 3.10 and Equation 3.7.

Table 3.9: Coefficients of the fifth regression equation

Term	**Coefficient**	**P-value**
Constant	-3.934	0.000
IDProductGroup	0.525	0.115
IDProduct	0.099	0.257
RemainingStock	0.007	0.000
Facing	0.061	0.000
$\ln Price_1$	1.217	0.002
$\ln Price_4$	0.896	0.004

Table 3.10: Model summary of the fifth regression equation

R-sq	**R-sq(adj)**	**R-sq(pred)**
63.12%	61.63%	59.28%

The fifth regression equation is

$$\begin{aligned}
&\ln Demand = -3.934 + (0.525 \times IDProductGroup) + (0.099 \times IDProduct) \\
&+ (0.007 \times RemainingStock) + (0.061 \times Facing) + (1.217 \times \ln Price_1) \\
&+ (0.896 \times \ln Price_4)
\end{aligned} \tag{3.7}$$

According to the fifth regression, p-values of *RemainingStock*, *Facing*, $\ln Price_1$ and $\ln Price_4$ are smaller than 0.05, so these predictors are significant for demand estimation. The adjusted R^2 value is 0.6163, and it demonstrates that the regression model explains 61.63% of variance of customer demands.

3.2.1.6 The sixth regression model for estimating demand

A log-linear regression is conducted in order to estimate demand by using *IDProduct*, *IDProductGroup*, *RemainingStock*, *Facing* and $Price_i$ parameters as predictors, while $Price_i$ represents the price of product i, $i \in \{1, 2, 3, \cdots, 10\}$. The results of this regression are shown in Table 3.11, Table 3.12 and Equation 3.8.

Table 3.11: Coefficients of the sixth regression equation

Term	**Coefficient**	**P-value**
Constant	-3.934	0.000
IDProductGroup	0.525	0.115
IDProduct	0.099	0.257
RemainingStock	0.007	0.000
Facing	0.061	0.000
$\ln Price_1$	0.429	0.002
$\ln Price_4$	0.329	0.004

Table 3.12: Model summary of the sixth regression equation

R-sq	**R-sq(adj)**	**R-sq(pred)**
63.12%	61.63%	59.28%

The sixth regression equation is

$$\begin{aligned}
&\ln Demand = -3.934 + (0.525 \times IDProductGroup) + (0.099 \times IDProduct) \\
&+ (0.007 \times RemainingStock) + (0.061 \times Facing) + (0.429 \times Price_1) \\
&+ (0.329 \times Price_4)
\end{aligned} \tag{3.8}$$

According to the sixth regression, p-values of *RemainingStock*, *Facing*, $Price_1$ and $Price_4$ are smaller than 0.05, so these predictors are significant for demand estimation. The adjusted R^2 value is 0.6163, and it demonstrates that the regression model explains 61.63% of variance of customer demands.

3.2.1.7 The seventh regression model for estimating demand

A log-linear regression is conducted to estimate demand by using $\ln IDProduct$, $\ln IDProductGroup$, $\ln RemainingStock$, $\ln Facing$ and $\ln Price_i$ parameters as predictors, while $Price_i$ represents the price of product i, $i \in \{1, 2, 3, \cdots, 10\}$. At this stage, the interactions of prices are added to the regression as predictor variables. The results of this regression are shown in Table 3.13, Table 3.14 and Equation 3.9.

Table 3.13: Coefficients of the seventh regression equation

Term	**Coefficient**	**P-value**
Constant	-9.000	0.000
$\ln RemaningStock$	0.186	0.024
$\ln Facing$	1.535	0.000
$\ln IDProductGroup$	4.050	0.000
$\ln IDProduct$	-0.063	0.836
$\ln Price_1$	10.510	0.003
$\ln Price_4$	6.820	0.000
$Price_1 * Facing$	-3.100	0.032
$Price_4 * Facing$	-1.749	0.002
$Price_1 * RemainingStock$	0.016	0.947
$Price_4 * RemainingStock$	-0.470	0.001

Table 3.14: Model summary of the seventh regression equation

R-sq	**R-sq(adj)**	**R-sq(pred)**
60.25%	57.51%	54.37%

The seventh regression equation is

$$\begin{aligned}
&\ln Demand = -9.000 + (0.186 \times \ln RemainingStock) + (1.535 \times \ln Facing) \\
&+ (4.050 \times \ln IDProductGroup) - (0.063 \times \ln IDProduct) \\
&+ (10.510 \times \ln Price_1) + (6.820 \times \ln Price_4) \\
&- (3.100 \times (Price_1 * Facing)) - (1.749 \times (Price_4 * Facing)) \\
&+ (0.016 \times (Price_1 * RemainingStock)) - (0.47 \times (Price_4 * RemainingStock))
\end{aligned} \tag{3.9}$$

According to the seventh regression, p-values of $\ln RemainingStock$, $\ln Facing$, $\ln IDProductGroup$, $\ln Price_1$, $\ln Price_4$, $Price_1 * Facing$, $Price_4 * Facing$ and $Price_4 * RemainingStock$ are smaller than 0.05, so these predictors are significant for demand estimation. The adjusted R^2 value is 0.5751, and it demonstrates that the regression model explains 57.51% of variance of customer demands.

3.2.1.8 The eighth regression model for estimating demand

A log-linear regression is conducted in order to estimate demand by using *IDProduct*, *IDProductGroup*, *RemainingStock*, *Facing*, $\ln Price_i$ and the interactions of all price parameters as predictors, while $Price_i$ represents the price of product i, $i \in \{1, 2, 3, \cdots, 10\}$. The results of this regression are shown in Table 3.15, Table 3.16 and Equation 3.10.

Table 3.15: Coefficients of the eighth regression equation

Term	**Coefficient**	**P-value**
Constant	-4.104	0.000
IDProductGroup	1.113	0.002
IDProduct	-0.095	0.325
RemainingStock	0.007	0.000
Facing	0.066	0.000
$\ln Price_1$	3.640	0.007
$\ln Price_4$	2.273	0.001
$Price_1 * Facing$	-0.238	0.039
$Price_4 * Facing$	-0.074	0.144
$Price_1 * RemainingStock$	0.011	0.355
$Price_4 * RemainingStock$	-0.035	0.001

Table 3.16: Model summary of the eighth regression equation

R-sq	**R-sq(adj)**	**R-sq(pred)**
67.52%	65.28%	62.36%

The eighth regression equation is

$$
\begin{aligned}
&\ln Demand = -4.104 + (1.113 \times IDProductGroup) - (0.095 \times IDProduct) \\
&+ (0.007 \times RemainingStock) + (0.066 \times Facing) + (3.640 \times \ln Price_1) \\
&+ (2.273 \times \ln Price_4) - (0.238 \times (Price_1 * Facing)) \\
&- (0.074 \times (Price_4 * Facing)) + (0.011 \times (Price_1 * RemainingStock)) \\
&- (0.035 \times (Price_4 * RemainingStock))
\end{aligned}
\tag{3.10}
$$

According to the eighth regression, p-values of *RemainingStock*, *Facing*, *IDProductGroup*, $\ln Price_1$, $\ln Price_4$, $Price_1 * Facing$ and $Price_4 * RemainingStock$ are smaller than 0.05, so these predictors are significant for demand estimation. The adjusted R^2 value is 0.6528, and it demonstrates that the regression model explains 65.28% of variance of customer demands.

3.2.1.9 The ninth regression model for estimating demand

A log-linear regression is conducted in order to estimate demand by using *IDProduct*, *IDProductGroup*, *RemainingStock*, *Facing*, $Price_i$ and the interactions of all price parameters as predictors, while $Price_i$ represents the price of product i, $i \in \{1, 2, 3, \cdots, 10\}$. The results of this regression are shown in Table 3.17, Table 3.18 and Equation 3.11.

The ninth regression equation is

$$
\begin{aligned}
&\ln Demand = -4.104 + (1.113 \times IDProductGroup) - (0.095 \times IDProduct) \\
&+ (0.007 \times RemainingStock) + (0.066 \times Facing) + (1.286 \times Price_1) \\
&+ (0.836 \times Price_4) - (0.084 \times (Price_1 * Facing)) \\
&- (0.027 \times (Price_4 * Facing)) + (0.004 \times (Price_1 * RemainingStock)) \\
&- (0.013 \times (Price_4 * RemainingStock))
\end{aligned}
\tag{3.11}
$$

Table 3.17: Coefficients of the ninth regression equation

Term	Coefficient	P-value
Constant	-4.104	0.000
IDProductGroup	1.113	0.002
IDProduct	-0.095	0.325
RemainingStock	0.007	0.000
Facing	0.066	0.000
$Price_1$	1.286	0.007
$Price_4$	0.836	0.001
$Price_1 * Facing$	-0.084	0.039
$Price_4 * Facing$	-0.027	0.144
$Price_1 * RemainingStock$	0.004	0.355
$Price_4 * RemainingStock$	-0.013	0.001

Table 3.18: Model summary of the ninth regression equation

R-sq	R-sq(adj)	R-sq(pred)
67.52%	65.28%	62.36%

According to the ninth regression, *p*-values of *RemainingStock*, *Facing*, *IDProductGroup*, $Price_1$, $Price_4$, $Price_1 * Facing$ and $Price_4 * RemainingStock$ are smaller than 0.05, so these predictors are significant for demand estimation. The adjusted R^2 value is 0.6528, and it demonstrates that the regression model explains 65.28% of variance of customer demands.

3.2.1.10 The tenth regression model for estimating demand

A log-linear regression is conducted to estimate demand by using $\ln IDProduct$, $\ln IDProductGroup$, $\ln RemainingStock$, $\ln Facing$, $\ln Price_i$ and the interactions of all price parameters as predictors, while $Price_i$ represents the price of product i, $i \in \{1, 2, 3, \cdots, 10\}$. The results of this regression are shown in Table 3.19, Table 3.20 and Equation 3.12.

Table 3.19: Coefficients of the tenth regression equation

Term	**Coefficient**	**P-value**
Constant	1.260	0.797
$\ln RemainingStock$	0.848	0.000
$\ln Facing$	0.278	0.421
$\ln IDProductGroup$	-2.920	0.423
$\ln IDProduct$	-0.097	0.719
$\ln Price_1$	2.880	0.537
$\ln Price_4$	1.690	0.495
$\ln Price_1 * \ln Facing$	-2.210	0.089
$\ln Price_4 * \ln Facing$	-0.557	0.336
$\ln Price_7 * \ln Facing$	1.220	0.277
$\ln Price_1 * \ln RemainingStock$	-0.483	0.037
$\ln Price_4 * \ln RemainingStock$	-1.110	0.000
$\ln Price_7 * \ln RemainingStock$	-2.756	0.000

Table 3.20: Model summary of the tenth regression equation

R-sq	**R-sq(adj)**	**R-sq(pred)**
69.56%	67.00%	64.10%

The tenth regression equation is

$$\begin{aligned}
&\ln Demand = 1.260 + (0.848 \times \ln RemainingStock) + (0.278 \times \ln Facing) \\
&- (2.920 \times \ln IDProductGroup) - (0.097 \times \ln IDProduct) + (2.880 \times \ln Price_1) \\
&+ (1.690 \times \ln Price_4) - (2.210 \times (\ln Price_1 * \ln Facing)) \\
&- (0.557 \times (\ln Price_4 * \ln Facing)) + (1.220 \times (\ln Price_7 * \ln Facing)) \\
&- (0.483 \times (\ln Price_1 * \ln RemainingStock)) \\
&- (1.110 \times (\ln Price_4 * \ln RemainingStock)) \\
&- (2.756 \times (\ln Price_7 * \ln RemainingStock))
\end{aligned} \tag{3.12}$$

According to the tenth regression, p-values of $\ln RemainingStock$, $\ln Price_1 * \ln RemainingStock$, $\ln Price_4 * \ln RemainingStock$, $\ln Price_7 * \ln RemainingStock$ are smaller than 0.05, so these predictors are significant for demand estimation. The adjusted R^2 value is 0.67, and it demonstrates that the regression model explains 67.00% of variance of customer demands.

3.2.1.11 The eleventh regression model for estimating demand

A log-linear regression is conducted in order to estimate demand by using *IDProduct*, *IDProductGroup*, *RemainingStock*, *Facing*, $\ln Price_i$ and the interactions of all price parameters as predictors, while $Price_i$ represents the price of product i, $i \in \{1, 2, 3, \cdots, 10\}$. The results of this regression are shown in Table 3.21, Table 3.22 and Equation 3.13.

Table 3.21: Coefficients of the eleventh regression equation

Term	**Coefficient**	**P-value**
Constant	-1.040	0.571
IDProductGroup	0.464	0.344
IDProduct	-0.129	0.191
RemainingStock	0.007	0.000
Facing	0.058	0.003
$\ln Price_1$	1.990	0.201
$\ln Price_4$	0.854	0.367
$\ln Price_1 * Facing$	-0.244	0.031
$\ln Price_4 * Facing$	-0.075	0.152
$\ln Price_7 * Facing$	-0.038	0.628
$\ln Price_1 * RemainingStock$	0.012	0.322
$\ln Price_4 * RemainingStock$	-0.036	0.000
$\ln Price_7 * RemainingStock$	-0.206	0.000

Table 3.22: Model summary of the eleventh regression equation

R-sq	**R-sq(adj)**	**R-sq(pred)**
70.25%	67.75%	64.60%

The eleventh regression equation is

$$
\begin{aligned}
&\ln Demand = -1.040 + (0.464 \times IDProductGroup) - (0.129 \times IDProduct) \\
&+ (0.007 \times RemainingStock) + (0.058 \times Facing) + (1.990 \times \ln Price_1) \\
&+ (0.854 \times \ln Price_4) - (0.244 \times (\ln Price_1 * Facing)) \\
&- (0.075 \times (\ln Price_4 * Facing)) - (0.038 \times (\ln Price_7 * Facing)) \\
&+ (0.012 \times (\ln Price_1 * RemainingStock)) \\
&- (0.036 \times (\ln Price_4 * RemainingStock)) \\
&- (0.206 \times (\ln Price_7 * RemainingStock))
\end{aligned}
\tag{3.13}
$$

According to the eleventh regression, p-values of *RemainingStock*, *Facing*, $\ln Price_1 * Facing$, $\ln Price_4 * RemainingStock$ and $\ln Price_7 * RemainingStock$ are smaller than 0.05, so these predictors are significant for demand estimation. The adjusted R^2 value is 0.6775, and it demonstrates that the regression model explains 67.75% of variance of customer demands.

3.2.1.12 The twelfth regression model for estimating demand

A log-linear regression is conducted in order to estimate demand by using *IDProduct*, *IDProductGroup*, *RemainingStock*, *Facing*, $Price_i$ and the interactions of all price parameters as predictors, while $Price_i$ represents the price of product i, $i \in \{1, 2, 3, \cdots, 10\}$. The results of this regression are shown in Table 3.23, Table 3.24 and Equation 3.14.

Table 3.23: Coefficients of the twelfth regression equation

Term	**Coefficient**	**P-value**
Constant	-1.040	0.571
IDProductGroup	0.464	0.344
IDProduct	-0.129	0.191
RemainingStock	0.007	0.000
Facing	0.058	0.003
$Price_1$	0.703	0.201
$Price_4$	0.314	0.367
$Price_1 * Facing$	-0.086	0.031
$Price_4 * Facing$	-0.028	0.152
$Price_7 * Facing$	-0.011	0.628
$Price_1 * RemainingStock$	0.004	0.322
$Price_4 * RemainingStock$	-0.013	0.000
$Price_7 * RemainingStock$	-0.058	0.000

Table 3.24: Model summary of the twelfth regression equation

R-sq	**R-sq(adj)**	**R-sq(pred)**
70.25%	67.75%	64.60%

The twelfth regression equation is

$$\begin{aligned}
&\ln Demand = -1.040 + (0.464 \times IDProductGroup) - (0.129 \times IDProduct) \\
&+ (0.007 \times RemainingStock) + (0.058 \times Facing) + (0.703 \times Price_1) \\
&+ (0.314 \times Price_4) - (0.086 \times (Price_1 * Facing)) \\
&- (0.028 \times (Price_4 * Facing)) - (0.011 \times (Price_7 * Facing)) \\
&+ (0.004 \times (Price_1 * RemainingStock)) \\
&- (0.013 \times (Price_4 * RemainingStock)) \\
&- (0.058 \times (Price_7 * RemainingStock))
\end{aligned} \tag{3.14}$$

According to the twelfth regression, p-values of $RemainingStock$, $Facing$, $Price_1 * Facing$, $Price_4 * RemainingStock$ and $Price_7 * RemainingStock$ are smaller than 0.05, so these predictors are significant for demand estimation. The adjusted R^2 value is 0.6775, and it demonstrates that the regression model explains 67.75% of variance of customer demands.

3.2.1.13 The thirteenth regression model for estimating demand

A log-linear regression is conducted to estimate demand by using $\ln IDProduct$, $\ln IDProductGroup$, $\ln RemainingStock$, $\ln Facing$, $\ln Price$ which is organized data in one column at this time and the interactions of all price parameters as predictors. The results of this regression are shown in Table 3.25, Table 3.26 and Equation 3.15.

Table 3.25: Coefficients of the thirteenth regression equation

Term	**Coefficient**	**P-value**
Constant	2.780	0.570
$\ln RemainingStock$	0.870	0.000
$\ln Facing$	0.505	0.165
$\ln IDProductGroup$	-5.190	0.174
$\ln IDProduct$	-0.057	0.833
$\ln Price$	0.447	0.060
$\ln Price_1 * \ln Facing$	-3.540	0.017
$\ln Price_4 * \ln Facing$	-0.919	0.129
$\ln Price_7 * \ln Facing$	1.140	0.308
$\ln Price_1 * \ln RemainingStock$	-0.489	0.033
$\ln Price_4 * \ln RemainingStock$	-1.064	0.000
$\ln Price_7 * \ln RemainingStock$	-2.655	0.000
$\ln Price_1 * \ln Price_2$	3.030	0.384
$\ln Price_4 * \ln Price_5$	0.980	0.574

Table 3.26: Model summary of the thirteenth regression equation

R-sq	**R-sq(adj)**	**R-sq(pred)**
70.31%	67.59%	64.57%

The thirteenth regression equation is

$$
\begin{aligned}
&\ln Demand = 2.780 + (0.870 \times \ln RemainingStock) + (0.505 \times \ln Facing) \\
&- (5.190 \times \ln IDProductGroup) - (0.057 \times \ln IDProduct) + (0.447 \times \ln Price) \\
&- (3.540 \times (\ln Price_1 * \ln Facing)) - (0.919 \times (\ln Price_4 * \ln Facing)) \\
&+ (1.14 \times (\ln Price_7 * \ln Facing)) - (0.489 \times (\ln Price_1 * \ln RemainingStock)) \\
&- (1.604 \times (\ln Price_4 * \ln RemainingStock)) \\
&- (2.655 \times (\ln Price_7 * \ln RemainingStock)) + (3.030 \times (\ln Price_1 * \ln Price_2)) \\
&+ (0.980 \times (\ln Price_4 * \ln Price_5))
\end{aligned}
\tag{3.15}
$$

According to the thirteenth regression, p-values of $\ln RemainingStock$, $\ln Price_1 * \ln Facing$, $\ln Price_1 * \ln RemainingStock$, $\ln Price_4 * \ln RemainingStock$ and $\ln Price_7 * \ln RemainingStock$ are smaller than 0.05, so these predictors are significant for demand estimation. The adjusted R^2 value is 0.6759, and it demonstrates that the regression model explains 67.59% of variance of customer demands.

3.2.1.14 The fourteenth regression model for estimating demand

A log-linear regression is conducted in order to estimate demand by using *IDProduct*, *IDProductGroup*, *RemainingStock*, *Facing*, $\ln Price$ which is organized data in one column at this time and the interactions of all price parameters as predictors. The results of this regression are shown in Table 3.27, Table 3.28 and Equation 3.16.

Table 3.27: Coefficients of the fourteenth regression equation

Term	**Coefficient**	**P-value**
Constant	0.590	0.770
IDProductGroup	0.063	0.905
IDProduct	-0.241	0.036
RemainingStock	0.007	0.000
Facing	0.074	0.000
$\ln Price$	0.500	0.056
$\ln Price_1 * \ln Price_2$	1.890	0.107
$\ln Price_4 * \ln Price_5$	0.439	0.513
$\ln Price_1 * Facing$	-0.419	0.004
$\ln Price_4 * Facing$	-0.145	0.023
$\ln Price_7 * Facing$	-0.068	0.385
$\ln Price_1 * RemainingStock$	0.016	0.196
$\ln Price_4 * RemainingStock$	-0.032	0.002
$\ln Price_7 * RemainingStock$	-0.179	0.003

Table 3.28: Model summary of the fourteenth regression equation

R-sq	**R-sq(adj)**	**R-sq(pred)**
71.01%	68.35%	65.06%

The fourteenth regression equation is

$$\begin{aligned}
&\ln Demand = 0.590 + (0.063 \times IDProductGroup) - (0.241 \times IDProduct) \\
&+ (0.007 \times RemainingStock) + (0.0734 \times Facing) + (0.500 \times \ln Price) \\
&+ (1.890 \times (\ln Price_1 * \ln Price_2)) + (0.439 \times (\ln Price_4 * \ln Price_5)) \\
&- (0.419 \times (\ln Price_1 * Facing)) - (0.145 \times (\ln Price_4 * Facing)) \\
&- (0.068 \times (\ln Price_7 * Facing)) + (0.016 \times (\ln Price_1 * RemainingStock)) \\
&- (0.032 \times (\ln Price_4 * RemainingStock)) \\
&- (0.179 \times (\ln Price_7 * RemainingStock))
\end{aligned} \tag{3.16}$$

According to the fourteenth regression, p-values of *IDProduct*, *RemainingStock*, *Facing*, $\ln Price_1 * Facing$, $\ln Price_4 * Facing$, $\ln Price_4 * RemainingStock$ and $\ln Price_7 * RemainingStock$ are smaller than 0.05, so these predictors are significant for demand estimation. The adjusted R^2 value is 0.6835, and it demonstrates that the regression model explains 68.35% of variance of customer demands.

3.2.1.15 The fifteenth regression model for estimating demand

A log-linear regression is conducted in order to estimate demand by using *IDProduct*, *IDProductGroup*, *RemainingStock*, *Facing*, *Price* which is organized data in one column at this time and the interactions of all price parameters as predictors. The results of this regression are shown in Table 3.29, Table 3.30 and Equation 3.17.

Table 3.29: Coefficients of the fifteenth regression equation

Term	**Coefficient**	**P-value**
Constant	-1.230	0487
IDProductGroup	0.644	0.176
IDProduct	-0.165	0.086
RemainingStock	0.003	0.015
Facing	0.052	0.005
Price	0.039	0.001
$Price_1 * Facing$	-0.089	0.021
$Price_4 * Facing$	-0.029	0.118
$Price_7 * Facing$	-0.010	0.624
$Price_1 * RemainingStock$	0.005	0.184
$Price_4 * RemainingStock$	-0.013	0.000
$Price_7 * RemainingStock$	-0.057	0.000
$Price_1 * Price_2$	0.200	0.150
$Price_4 * Price_5$	0.088	0.284

Table 3.30: Model summary of the fifteenth regression equation

R-sq	**R-sq(adj)**	**R-sq(pred)**
72.55%	70.03%	67.15%

The fifteenth regression equation is

$$\ln Demand = -1.230 + (0.644 \times IDProductGroup) - (0.165 \times IDProduct)$$
$$+ (0.003 \times RemainingStock) + (0.052 \times Facing) + (0.039 \times Price)$$
$$- (0.089 \times (Price_1 * Facing)) - (0.029 \times (Price_4 * Facing))$$
$$- (0.010 \times (Price_7 * Facing)) + (0.005 \times (Price_1 * RemainingStock))$$
$$- (0.013 \times (Price_4 * RemainingStock))$$
$$- (0.057 \times (Price_7 * RemainingStock)) + (0.200 \times (Price_1 * Price_2))$$
$$+ (0.088 \times (Price_4 * Price_5)) \quad (3.17)$$

According to the fifteenth regression, p-values of $RemainingStock$, $Facing$, $Price$, $Price_1 * Facing$, $Price_4 * RemainingStock$ and $Price_7 * RemainingStock$ are smaller than 0.05, so these predictors are significant for demand estimation. The adjusted R^2 value is 0.7003, and it demonstrates that the regression model explains 70.03% of variance of customer demands.

3.2.1.16 Comparison between regression models

After all of these regressions are conducted on the sample data which consists of only ten products, the outcomes of them are compared in order to decide which demand model gives the best fits. The comparison is made according to R^2 values, and all these R^2 values are shown in Table 3.31.

The R^2 values of the second model is greater than others (Table 3.31). Consequently, we chose the second model as demand model of this sample data, and this type of regression can be performed on all data group for more reliable results. Then, the general demand model of this thesis is shown in Equation 3.18.

$$Demand = e^{-\alpha_0 + \alpha_1 IDProductGroup - \alpha_2 IDProduct + \alpha_3 RemainingStock + \alpha_4 Facing} \quad (3.18)$$

Table 3.31: The R^2 outcomes of the regressions

	R-sq	R-sq(adj)	R-sq(pred)
Model 1	70.48%	69.75%	66.93%
Model 2	74.06%	72.82%	70.36%
Model 3	63.40%	60.33%	56.33%
Model 4	53.77%	51.90%	49.19%
Model 5	63.12%	61.63%	59.28%
Model 6	63.12%	61.63%	59.28%
Model 7	60.25%	57.51%	54.37%
Model 8	67.52%	65.28%	62.36%
Model 9	67.52%	65.28%	62.36%
Model 10	69.56%	67.00%	64.10%
Model 11	70.25%	67.75%	64.60%
Model 12	70.25%	67.75%	64.60%
Model 13	70.31%	67.59%	64.57%
Model 14	71.01%	68.35%	65.06%
Model 15	72.55%	70.03%	67.15%

In Section 3.2.4, demands of all the products which exist in the assortment of the supermarket is estimated according to the demand model (3.18). *IDProduct-Group*, *IDProduct*, *RemainingStock* and *Facing* parameters must be used in the regression analyses as the predictors for more acceptable results.

3.2.2 Analysis of Regression Assumptions for the Selected Demand Model

In the previous sections, we provide discussion for all regression models, and we decided to use the second regression model. Because, R^2 values of the second model is greater than the R^2 values of the other models.

According to the second regression equation (3.4), demand is a function of IDProductGroup, IDProduct, RemainingStock and Facing parameters, and all these parameters are significant to determine the demand because of their small p-values. The coefficients of these parameters show that demand increases in direct proportion to the increase of IDProductGroup, RemainingStock and Facing parameters, and IDProduct affects the demand negatively.

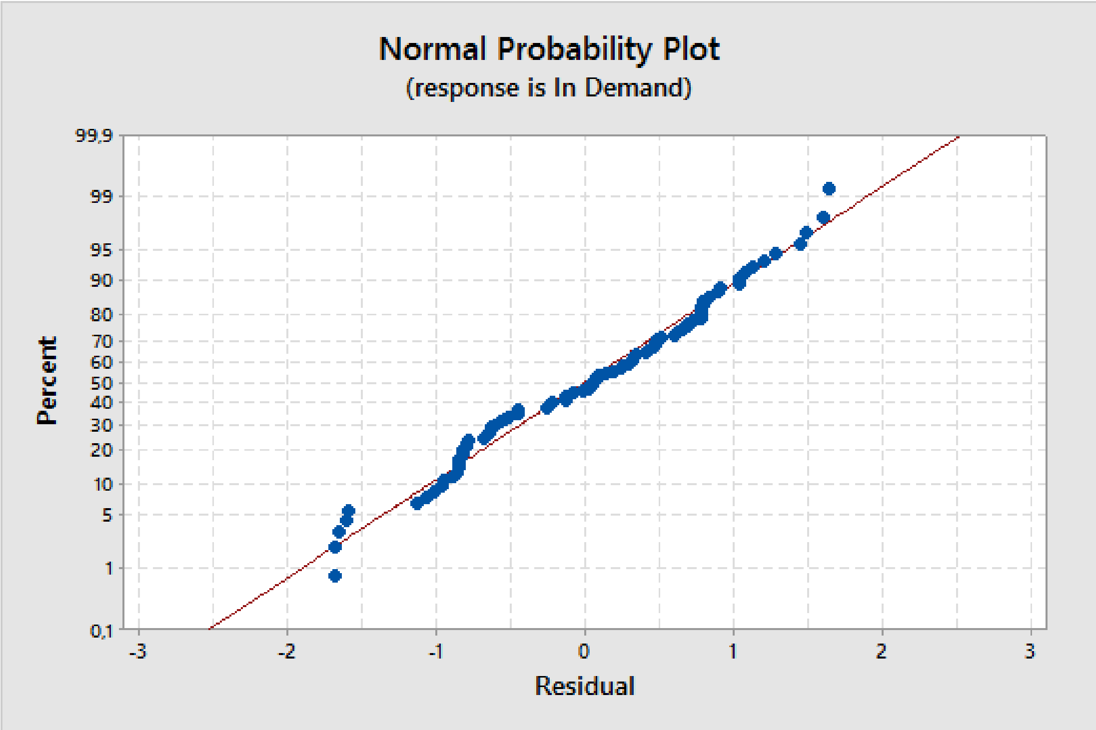

Figure 3.1: The normal probability plot of residuals

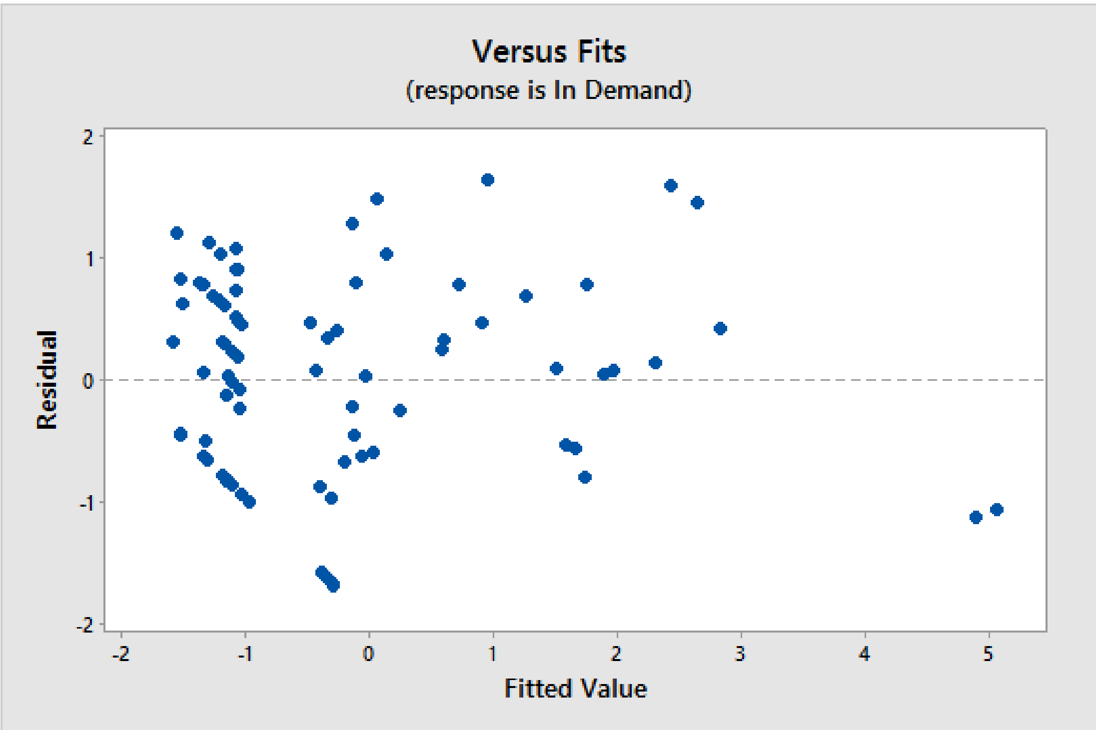

Figure 3.2: The plot of residuals versus fits

Residuals are normally distributed as shown in the Figure 3.1, and the residuals have a nearly constant variance as shown in the Figure 3.2, but there are two outliers. As a result, the residual plots justify our regression assumptions.

3.2.3 Validation of the Regression Model

In order to measure the performance of regression models, estimation of prediction accuracy is necessary. Cross-validation is a model validation technique which is widely used for estimation of the prediction accuracy [21]. In cross-validation test, a *training dataset* is chosen from the whole data in order to run training on that known dataset, and a *testing dataset* is chosen against the model which is tested [22].

In other words, the cross-validation divides a sample of data into complementary subsets, implementing the analysis on one subset which is training set, and validating the analysis on the other subset which is test set. Multiple runs of cross-validation are performed by using different subsets in order to reduce variability, and the cross-validation results are averaged over the runs [22].

The prevalent types of cross-validation are exhaustive and non-exhaustive cross-validation. The type of cross validation is determined by the proportion of data set utilized in the validation test. If the original data is entirely used and divided into training and test sets for cross-validation, exhaustive cross-validation methods are used. If only a sample from the original data is used, non-exhaustive cross-validation methods are used.

The data that we used so far for generating the empirical demand model includes only ten products, and it does not contain the assortment data entirely. The cross-validation test is performed on this sample data, so non-exhaustive cross-validation techniques are used. k-fold cross validation is one of the non-exhaustive cross-validation technique, and we decide to use it for validation of the demand model. All the steps of k-fold cross-validation are described in details in this section.

First of all, data set is divided into k equal parts. In the literature, k is generally chosen as 5 or 10 depending on the size of the data. In our study, the sample data consists of thirteen weeks data of ten products, so there are 130 rows data initially. We decide k to be 5, and it means that the data set which consisted from 130 rows are divided into 5 equal parts. Thus, the test set consists of 26 rows data.

Secondly, the first part is stated as the test set, and the union of the other parts is stated as the training set. A regression is conducted on the training set, and the regression equation and the coefficients are kept in order to estimate demands in testing data. The estimated demand values are calculated according to the regression equation which is formed by using training set. Finally, the MAPE of the first part is calculated by using the Equation 3.19 [23].

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{\mid Y_t - \hat{Y}_t \mid}{\mid Y_t \mid} \tag{3.19}$$

These steps are repeated five times by using a different test set at each time. As a result of the cross-validation test, all MAPE values which are found for five parts are averaged, and the average MAPE value demonstrates estimating accuracy of the demand model. The MAPE values are shown in Table 3.32.

Table 3.32: The result of cross-validation test on a sample data

	MAPE
First part	60.35%
Second part	302.9%
Third part	87.75%
Fourth part	116.4%
Fifth part	210.9%
Average	155.7%

The mean absolute percentage error (MAPE) for cross-validation tests is 155.65%, and the error rate is excessive. Thus, it can be said that the demand model on this data set might estimate demands wrongly. According to our assessments, the high deviation on these data sets that are grouped in the cross-variation test caused to this situation.

3.2.4 Demand Estimation of All Candidate Products with Different Data Sizes

In this study, we aim to suggest an assortment plan of a supermarket. Therefore, an assortment optimization model is set. The assortment optimization model which will be explained in Chapter 4 determines the facing of every product in the assortment. The best demand equation which is an important part of the optimization model is found by conducting different regressions on a sample data. However, we discoursed the demand model for only ten products until now, and the demands of all candidate products should be known in order to solve the model and obtain the optimum solution.

In order to set demand model and estimate the demands of all products which are counted as 5004, a log-linear regression was done by using *IDProduct*, *IDProductGroup*, *RemainingStock* and *Facing* parameters as predictors. The results of this regression are shown in Table 3.33, Table 3.34 and Equation 3.20.

Table 3.33: Coefficients of the regression equation for all products in assorment

Term	Coefficient	P-value
Constant	-0.432	0.000
IDProductGroup	0.025	0.000
IDProduct	-0.001	0.000
RemainingStock	0.001	0.000
Facing	0.002	0.000

Table 3.34: Model summary of the regression equation for all products in assortment

R-sq	R-sq(adj)	R-sq(pred)
19.32%	19.31%	19.06%

The regression equation for all data is

$$\ln Demand = -0.432 + (0.025 \times IDProductGroup) - (0.001 \times IDProduct) + (0.001 \times RemainingStock) + (0.002 \times Facing) \tag{3.20}$$

According to the regression for all product, p-values of *RemainingStock*, *Facing*, *IDProduct* and *IDProductGroup* are smaller than 0.05, so these predictors are significant for demand estimation. The adjusted R^2 value is 19.31 percent, and it demonstrates that the regression model explains 19.31% of data. It can be seen that the R^2 value decreases considerably, but this situation is expected because of increasing the size of data.

The Equation 3.20 is the demand model of all products, and this demand model is used for assortment planning of the supermarket. Additionally, we conduct some more regressions on different size of datasets to use in the computational efficiency analysis that will be explained in Chapter 4. The results of these regressions are shown below.

Table 3.35: Coefficients of the regression equation for 4000 products

Term	**Coefficient**	**P-value**
Constant	-0.331	0.000
IDProductGroup	0.025	0.000
IDProduct	-0.001	0.000
RemainingStock	0.001	0.000
Facing	0.002	0.000

Table 3.36: Model summary of the regression equation for 4000 products

R-sq	**R-sq(adj)**	**R-sq(pred)**
13.39%	13.38%	13.10%

The regression equation for 4000 products

$$\ln Demand = -0.3301 + (0.025 \times IDProductGroup) - (0.001 \times IDProduct) + (0.001 \times RemainingStock) + (0.002 \times Facing) \tag{3.21}$$

Table 3.37: Coefficients of the regression equation for 3000 products

Term	**Coefficient**	**P-value**
Constant	-0.943	0.000
IDProductGroup	0.009	0.000
IDProduct	-0.001	0.000
RemainingStock	0.004	0.000
Facing	0.027	0.000

Table 3.38: Model summary of the regression equation for 3000 products

R-sq	**R-sq(adj)**	**R-sq(pred)**
25.41%	25.40%	25.34%

The regression equation for 3000 products

$$\ln Demand = -0.943 + (0.009 \times IDProductGroup) - (0.001 \times IDProduct) + (0.004 \times RemainingStock) + (0.027 \times Facing) \tag{3.22}$$

Table 3.39: Coefficients of the regression equation for 2000 products

Term	**Coefficient**	**P-value**
Constant	-0.973	0.000
IDProductGroup	0.008	0.000
IDProduct	-0.001	0.000
RemainingStock	0.005	0.000
Facing	0.027	0.000

Table 3.40: Model summary of the regression equation for 2000 products

R-sq	**R-sq(adj)**	**R-sq(pred)**
26.59%	26.57%	26.44%

The regression equation for 2000 products

$$\ln Demand = -0.973 + (0.008 \times IDProductGroup) - (0.001 \times IDProduct) + (0.005 \times RemainingStock) + (0.027 \times Facing) \tag{3.23}$$

Table 3.41: Coefficients of the regression equation for 1000 products

Term	**Coefficient**	**P-value**
Constant	-0.954	0.000
IDProductGroup	-0.024	0.000
IDProduct	0.001	0.000
RemainingStock	0.004	0.000
Facing	0.031	0.000

Table 3.42: Model summary of the regression equation for 1000 products

R-sq	**R-sq(adj)**	**R-sq(pred)**
22.45%	22.42%	22.11%

The regression equation for 1000 products

$$\ln Demand = -0.954 - (0.024 \times IDProductGroup) + (0.001 \times IDProduct) + (0.004 \times RemainingStock) + (0.031 \times Facing) \tag{3.24}$$

To conclude, the log-linear regressions are conducted by using different predictor parameters and different models in order to find the demand equation that gives the best fits. After the best demand model is determined, cross-validation tests are performed for validating the model accuracy. Finally, the demand model is applied by using the all products data for planning assortment of the supermarket entirely. In the next chapter, an assortment optimization model will be formed by consisting of the demand model.

Chapter 4

Assortment Optimization Problem

Assortment optimization problem came into prominence with the claim that is arranging a wide variety of products on shelving which have limited area as the most profitably. In this chapter, an assortment optimization model is developed in order to maximize total revenue of a supermarket under limited shelf space constraint.

The optimization model provides optimum facing amounts for all candidate products of assortment. Facing is defined as the number of products that are visible in front of the shelf, and the facing amounts are directly related to the total shelf space capacity, width and length of the products. The difference between width and length can be explained as that width is the dimension of a product while the products line up side by side along the shelf length, and length is the dimension of a product while the products line up in tandem along the shelf depth. The multiplication of shelf depth and shelf length indicates the total shelf space capacity. The units of shelf depth, shelf length, product width and product length are assumed as centimeters. The Figure 4.1 and the Figure 4.2 illustrate these terms.

We assume that there is no backroom for the products, and all products, which are in the assortment, are displayed on the shelving. The inventory level of a product is fully loaded on the shelf, and the stock of this product decreases when it is sold.

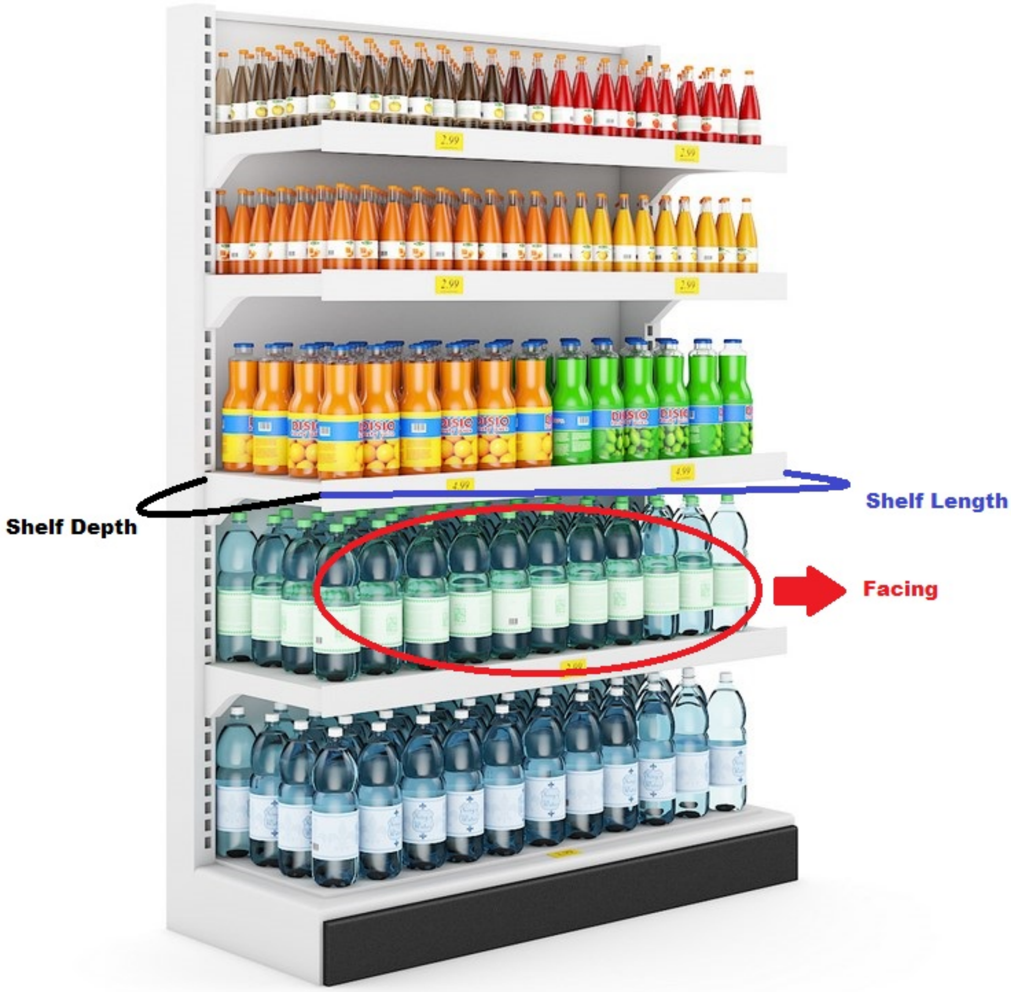

Figure 4.1: The illustration of shelf depth, shelf length and facing on an example shelf arrangement

We also assume that all products are only placed alongside. If the products are placed as one on the top of the other as shown in Figure 4.2, we also need to consider height of the products and the height between two shelving in order to find optimum facing amounts. For instance, in the Figure 4.2, the product which is packaged with a light blue box is placed as one on the top of the other. The height between two shelving is more than the height of the product at least three times, thus three packages of this product can be placed as one on the top of the other.

Figure 4.2: The illustration of product width and product length on an example shelf arrangement

The limit of shelf length (LSL) is determined by taking into consideration of all shelving in the store, and it contains the length of a lot of different shelving which are located alongside or one on the top of the other. We designate the shelf depth (SD) as 60cm, and it is a reasonable number according to shelf standards of supermarkets. We assign artificially large amounts for the upper limit of facing (ULF) such as 50.

In our assumption, the candidate products are not perishable, and it can be stay on the shelf during a long time if it is not sold. The inventory control of perishable products is more complicated, so our optimization model is not suitable for perishable products.

In this chapter, we conduct a convexity analysis of the model, and we prove that the model finds the global optimum solution. Also, we show the solution by using Karush Kuhn Tucker conditions.

As a numerical work, the problem is solved for the data sets which consist of 5004, 4000, 3000, 2000 and 1000 products differently, and GAMS (General Algebraic Modeling System) is used in order to solve the nonlinear model. The solution time for all these data sets is compared for the computational complexity analysis.

This chapter is organized as follows. We first present the mathematical model and explain the constraints in Section 4.1. Secondly, the complexity analysis of the model is made with a proof, and the model is solved by using Karush Kuhn Tucker conditions in Section 4.2. Finally, the model is solved for some data sets, and computational complexity analysis is performed in Section 4.3.

4.1 The Mathematical Model for Assortment Optimization

The mathematical formulation of the model can be written as follows.

$$\max \sum_{j \in J} p_j d_j \tag{4.1a}$$

subject to

$$\ln d_j = -\alpha_0 + \alpha_1 gr_j - \alpha_2 pr_j + \alpha_3 s_j + \alpha_4 f_j, \qquad \forall j \in J, \tag{4.1b}$$

$$\sum_{j \in J} f_j w_j \leq \text{LSL}, \tag{4.1c}$$

$$f_j \leq \text{ULF}, \qquad \forall j \in J, \tag{4.1d}$$

$$(\text{SD}/l_j) f_j - d_j \geq s_j, \qquad \forall j \in J, \tag{4.1e}$$

$$f_j \geq 0, \qquad \forall j \in J, \tag{4.1f}$$

$$s_j \geq 0, \qquad \forall j \in J. \tag{4.1g}$$

J is the set including all candidate products that are evaluated in the model, and $j \in J$ is the index of candidate products that are evaluated for the assortment.

The objective function (4.1a) maximizes the total revenue of the supermarket. The price of product $j \in J$ is defined by the parameter p_j, and the decision variable d_j identifies the demand of product $j \in J$.

The first constraint (4.1b) is the demand function which is developed by log-linear regression in Minitab, and we explained how the demand function is formed in Chapter 3. The coefficient values which are α_0, α_1, α_2, α_3 and α_4 are greater than zero, and they are provided by the regression equation. We define the parameter gr_j which indicates the group ID of the product $j \in J$ and the parameter pr_j which indicates the product ID of the product $j \in J$. We also define the decision variable s_j which states the remaining stock of product $j \in J$ and the decision variable f_j which is the facing quantities of product $j \in J$.

The second constraint (4.1c) limits the shelf length, denoted by LSL, and the products, which are positioned alongside on shelving, are restricted by this constraint. The parameter w_j indicates the width of the product $j \in J$.

The third constraint (4.1d) determines the maximum facing amount which can be assigned to a product, and it is denoted by ULF.

The fourth constraint (4.1e) shows that the value which is remained when the demand is extracted from the total capacity of the product on shelf must always be greater than or equal to the remaining stock of product $j \in J$. Because, we assume that there is no backroom in the store, and all the inventory of products is kept on the shelf. In this constraint, SD states the shelf depth, and it equals to 60. Also, we define the parameter l_j which states the length of product j.

In the constraints (4.1f) and (4.1g), we state that the facing and the remaining stock are always greater than or equal to zero for the product $j \in J$.

Table 4.1: Notation Chart

J	the set of candidate products
j	the index of candidate products in the set J
p_j	the price of the product $j \in J$
d_j	the demand of the product $j \in J$
gr_j	the group ID of the product $j \in J$
pr_j	the product ID of the product $j \in J$
s_j	the remaining stock of the product $j \in J$
f_j	the facing of the product $j \in J$
w_j	the width of the product $j \in J$
l_j	the length of the product $j \in J$

4.2 The Mathematical Analysis of the Model

In order to understand the model and determine the solution method, we make some detailed analyses. Initially, the convexity of the model is investigated in order to understand whether or not the solution of the model is global optimum. After the convexity analysis, the model is solved by using Karush Kuhn Tucker conditions, and the results are evaluated.

4.2.1 Convexity Analysis

A set C is *convex set*, if the line segment which is located between any two points in C lies in C. Thus, for any x_1, $x_2 \in C$ and any θ with $0 \leq \theta \leq 1$,

$$\theta x_1 + (1 - \theta)x_2 \in C.$$

The intersection of two convex sets is convex, and convexity is always preserved under intersection. A is defined as a set of convex sets. While S_α is convex set for every $\alpha \in A$, $\bigcap_{\alpha \in A} S_\alpha$ is convex [24].

In the study of nonlinear programming problems, convex and concave functions have an important place.

$f(x)$ is exemplified as a function which is indicated for all points $(x_1, x_2, \cdots, x_n)$ in a convex set S. The function $f(x)$ is a convex function on S if for any $x' \in S$ and $x'' \in S$,

$$f(cx' + (1-c)x'') \leq cf(x') + (1-c)f(x''), \tag{4.2}$$

holds for $0 \leq c \leq 1$ [25].

The function $f(x)$ is a concave function, if $-f(x)$ is a convex function.

Consider a nonlinear maximization problem. Suppose the feasible region S is a convex set. If the objective function is concave on S, then any local maximum is the global optimum solution to this problem [25]. In this thesis, the assortment optimization model is nonlinear maximization problem. Thus, the solution of this problem is the optimal solution, if the objective function is concave on a convex set.

For multivariate functions, the Hessian matrix can be used in order to determine whether the function is a convex or concave function on a convex set. If this function is convex, all principal minors of the Hessian matrix are nonnegative, and the Hessian matrix should be positive semi-definite [25].

In our study, the objective function of the Model 4.1 is concave function on a convex set.

Lemma 4.1. *If the objective function is concave on a convex set, there exist an unique optimum solution.*

Proof. We know that all constraints except the Constraint 4.1e are convex, and demonstration of convexity for the Constraint 4.1e is sufficient to notice that the solution of this model finds the global optimum.

The Constraint 4.1e is shown below,

$$(\mathrm{SD}/l_j)f_j - d_j \geq s_j.$$

The analysis is made for only one product, and the results are generalized for the model. Because, the Constraint 4.1e is seperable linear constraint, and if the convexity is proved for only one product j, the intersections of all product j is convex anyway. Then, the j indicator can be removed for convenience in analysis.

$$\begin{aligned}
&c_1 f - e^{\ln d} \geq s \\
&c_1 f - e^{\alpha_3 s + \alpha_4 f - c_2} \geq s \\
&c_1 f - \frac{e^{\alpha_3 s} e^{\alpha_4 f}}{e^{c_2}} - s \geq 0
\end{aligned}$$

In these formulations, c_1, c_2, α_3 and α_4 are constant, and there are two variables in this function. Thus, the Hessian matrix can be applied to this function in order to learn the convexity.

$$\begin{aligned}
&g(f, s) = s - c_1 f + \frac{e^{\alpha_3 s} e^{\alpha_4 f}}{e^{c_2}} \\
&\frac{\partial^2 g}{\partial f^2} = \alpha_4^2 e^{\alpha_3 s + \alpha_4 f - c_2} \\
&\frac{\partial^2 g}{\partial s \partial f} = \alpha_4 \alpha_3 e^{\alpha_3 s + \alpha_4 f - c_2} \\
&\frac{\partial^2 g}{\partial f \partial s} = \alpha_4 \alpha_3 e^{\alpha_3 s + \alpha_4 f - c_2} \\
&\frac{\partial^2 g}{\partial s^2} = \alpha_3^2 e^{\alpha_3 s + \alpha_4 f - c_2}
\end{aligned}$$

The first principal minors of the Hessian matrix ($\alpha_4^2 e^{\alpha_3 s+\alpha_4 f-c_2}$ and $\alpha_3^2 e^{\alpha_3 s+\alpha_4 f-c_2}$) are greater than zero, and the Hessian matrix is positive semi-definite. Due to these non-negativities, the function is convex, and the graph of the convex function constitutes a convex set.

■

As a consequence of this analysis, the objective function is concave function on a convex set, and the model finds the global optimum solution.

4.2.2 Analysis of the Model by KKT Conditions

Given the general problem,

$$\text{max f(x)} \tag{4.5a}$$

subject to

$$h_i(x) \leq 0 \text{ , i=1, } \cdots \text{ , m,} \tag{4.5b}$$

$$l_j(x) = 0 \text{ , j=1, } \cdots \text{ , n.} \tag{4.5c}$$

The Karush Kuhn Tucker (KKT) conditions are:

- $\partial f(x) - \sum_{i=1}^{m} \lambda_i \partial h_i(x) + \sum_{j=1}^{n} \upsilon_j \partial l_j(x) = 0$ **(stationarity)**
- $\lambda_i . h_i(x) = 0$ *for all i* **(complementary slackness)**
- $h_i(x) \leq 0$, $l_j(x) = 0$ *for all i,j* **(primal feasibility)**
- $\lambda_i \geq 0$ *for all i* **(dual feasibility)**

For an optimization problem with differentiable and concave objective function on a convex set and constraints for which strong duality exist, any pair of primal and dual optimal points has to satisfy the Karush Kuhn Tucker conditions. Optimization problems can be solved by the KKT conditions analytically [24].

In our study, the KKT analysis is made by using the data of two products which are in assortment, and the consideration of this analysis is generalized for all products. In order to analyze KKT conditions, the mathematical formulation of the optimization model (4.1) is rewritten as shown in Equation 4.6.

$$\max p_1(e^{\beta_0+\beta_1 s_1+\beta_2 f_1}) + p_2(e^{\alpha_0+\alpha_1 s_2+\alpha_2 f_2}) \tag{4.6a}$$

subject to

$$w_1 f_1 + w_2 f_2 \leq LSL, \tag{4.6b}$$

$$f_1 \leq UFL, \tag{4.6c}$$

$$f_2 \leq UFL, \tag{4.6d}$$

$$e^{\beta_0+\beta_1 s_1+\beta_2 f_1} - \gamma_1 f_1 + s_1 \leq 0, \tag{4.6e}$$

$$e^{\alpha_0+\alpha_1 s_2+\alpha_2 f_2} - \gamma_2 f_2 + s_2 \leq 0, \tag{4.6f}$$

$$s_1, s_2, f_1, f_2 \geq 0. \tag{4.6g}$$

The Lagrangian Function (4.7) is written as shown,

$$L = p_1(e^{\beta_0+\beta_1 s_1+\beta_2 f_1}) + p_2(e^{\alpha_0+\alpha_1 s_2+\alpha_2 f_2}) - \lambda_1(w_1 f_1 + w_2 f_2 - LSL) - \lambda_2(f_1 - UFL)$$
$$- \lambda_3(f_2 - UFL) - \lambda_4(e^{\beta_0+\beta_1 s_1+\beta_2 f_1} - \gamma_1 f_1 + s_1) - \lambda_5(e^{\alpha_0+\alpha_1 s_2+\alpha_2 f_2} - \gamma_2 f_2 + s_2). \tag{4.7}$$

Stationarity condition

By taking the first partial derivatives;

$$\frac{\partial L}{\partial s_1} = (p_1 - \lambda_4)(\beta_1 e^{\beta_0+\beta_1 s_1+\beta_2 f_1}) - \lambda_4 = 0, \tag{4.8}$$

$$\frac{\partial L}{\partial s_2} = (p_2 - \lambda_5)(\alpha_1 e^{\alpha_0+\alpha_1 s_2+\alpha_2 f_2}) - \lambda_5 = 0, \tag{4.9}$$

$$\frac{\partial L}{\partial f_1} = (p_1 - \lambda_4)(\beta_2 e^{\beta_0+\beta_1 s_1+\beta_2 f_1}) - \lambda_1 w_1 - \lambda_2 + \lambda_4 \gamma_1 = 0, \tag{4.10}$$

$$\frac{\partial L}{\partial f_2} = (p_2 - \lambda_5)(\alpha_2 e^{\alpha_0+\alpha_1 s_2+\alpha_2 f_2}) - \lambda_1 w_2 - \lambda_3 + \lambda_5 \gamma_2 = 0. \tag{4.11}$$

Complementary slackness conditions

The optimum values must satisfy the complementary slackness conditions.

$$\lambda_1(w_1 f_1 + w_2 f_2 - LSL) = 0, \tag{4.12}$$

$$\lambda_2(f_1 - UFL) = 0, \tag{4.13}$$

$$\lambda_3(f_2 - UFL) = 0, \tag{4.14}$$

$$\lambda_4(s_1 - \gamma_1 f_1 + e^{\beta_0+\beta_1 s_1+\beta_2 f_1}) = 0, \tag{4.15}$$

$$\lambda_5(s_2 - \gamma_2 f_2 + e^{\alpha_0+\alpha_1 s_2+\alpha_2 f_2}) = 0. \tag{4.16}$$

Primal feasibility conditions

The optimum values must satisfy primal feasibility conditions.

$$w_1 f_1 + w_2 f_2 - LSL \leq 0, \tag{4.17}$$

$$f_1 - UFL \leq 0, \tag{4.18}$$

$$f_2 - UFL \leq 0, \tag{4.19}$$

$$e^{\beta_0+\beta_1 s_1+\beta_2 f_1} - \gamma_1 f_1 + s_1 \leq 0, \tag{4.20}$$

$$e^{\alpha_0+\alpha_1 s_2+\alpha_2 f_2} - \gamma_2 f_2 + s_2 \leq 0. \tag{4.21}$$

Dual feasibility conditions

All lagrange multipliers must be greater than zero according to dual feasibility conditions.

$$\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5 \geq 0. \tag{4.22}$$

We solve these set of equations and inequalities in order to find the optimum points which give the maximum solution. The solution of the model is infeasible except two cases, and these cases are shown below.

In the case when λ_1, λ_4 and λ_5 are greater than zero, and λ_2 and λ_3 equal to zero, if the conditions which are $(ULF)w_1 + (ULF)w_2 > LSL$, $\frac{w_1\beta_1}{\beta_2+\gamma_1\beta_1} > 0$ and $\frac{w_2\alpha_1}{\alpha_2+\gamma_2\alpha_1} > 0$ are satisfied, the analytical formulations of facing and remaining stock are shown below for two situations.

If $m(p_1, w_1, \beta_1, \beta_2) > m(p_2, w_2, \alpha_1, \alpha_2)$,

$f_1 = ULF$ and $f_2 = \frac{LSL-(ULF)w_1}{w_2}$.

Then,

$$\ln(\gamma_1(ULF) - s_1) = \beta_0 + \beta_1 s_1 + \beta_2(ULF), \tag{4.23a}$$

$$\ln(\gamma_2(\frac{LSL-(ULF)w_1}{w_2}) - s_2) = \alpha_0 + \alpha_1 s_2 + \alpha_2(\frac{LSL-(ULF)w_1}{w_2}). \tag{4.23b}$$

If $m(p_1, w_1, \beta_1, \beta_2) < m(p_2, w_2, \alpha_1, \alpha_2)$,

$f_2 = ULF$ and $f_1 = \frac{LSL-(ULF)w_2}{w_1}$.

Then,

$$\ln(\gamma_1(\frac{LSL-(ULF)w_2}{w_1})-s_1)=\beta_0+\beta_1 s_1+\beta_2(\frac{LSL-(ULF)w_2}{w_1}), \quad (4.24a)$$

$$\ln(\gamma_2(ULF)-s_2)=\alpha_0+\alpha_1 s_2+\alpha_2(ULF), \quad (4.24b)$$

when the functions $m(p_1, w_1, \beta_1, \beta_2)$ and $m(p_2, w_2, \alpha_1, \alpha_2)$ can be defined from the Equations 4.8 to 4.11.

In the case when λ_2, λ_3, λ_4, λ_5 are greater than zero, and λ_1 equals to zero, if the conditions which are $(ULF)w_1 + (ULF)w_2 \leq LSL$, $\frac{\beta_1}{\beta_2+\gamma_1\beta_1} > 0$ and $\frac{\alpha_1}{\alpha_2+\gamma_2\alpha_1} > 0$ are satisfied, the analytical formulation of facing and remaining stock is shown as,

$$\beta_0+\beta_1 s_1+\beta_2(ULF)=\ln(\gamma_1(ULF)-s_1), \quad (4.25a)$$

$$\alpha_0+\alpha_1 s_2+\alpha_2(ULF)=\ln(\gamma_2(ULF)-s_2). \quad (4.25b)$$

Consequently, the relation between remaining stock and facing is represented. The model has unique optimum solution, because the objective function is differentiable and concave, and the constraints are convex and differentiable.

4.3 Application of the Model and Computational Experiments

Previously, we explained the assortment optimization problem, and the mathematical analyses were conducted. In this section, a numerical experiment will be demonstrated in order to present applicability of the model.

Initially, a product group is chosen, and the total number of products of this group is 5004. We eliminate perishable products, since the Model (4.1) is not suitable for perishable products. Perishable products have limited shelf lives, and this feature is not considered in our study. In this direction, for an exemplary study, see Onal et al. [26].

First, we formed a demand function (3.20) which is given in the Chapter 3. Then, the upper limits which are generalized in the formulation of model (4.1) are set to solve the model for this data group. LSL, ULF and SD are designated respectively as 14000, 50, 60, and these artificial values are assigned according to our approximations.

The model is solved by the BARON solver of GAMS, and the objective value is found as 27901.84. The main goal of this study is finding the optimal facing values for all products in order to plan assortment with maximum total revenue. The Model (4.1) provides achieving to this goal, and the facings of all products are found logically. A small part of the results are shown with the corresponding width and price parameters in Table 4.2.

As is seen in Table 4.2, for some products, the model assigns the maximum limit of facing which is denoted with ULF, and the facing of other products are very small values because of having no more shelf space available. According to our assessments, the reason of this situation is that the model assigns the maximum facing to the products which have less width and high price, and by this way, the model exhibits a behavior as if it targets to earn more revenue on a small shelf area. Consequently, the facing of a product is highly related with price of the product and width of the product, and this relation is shown in the Figure 4.3. When the facing equals to the maximum limit (ULF), that is assigned as 50 for this experiment, price of the products are high and width of the products are small simultaneously.

4.3.1 Computational Complexity Analysis

In this section, numerical experiments are performed with the product groups which have different numbers of products in order to analyze the effect of data

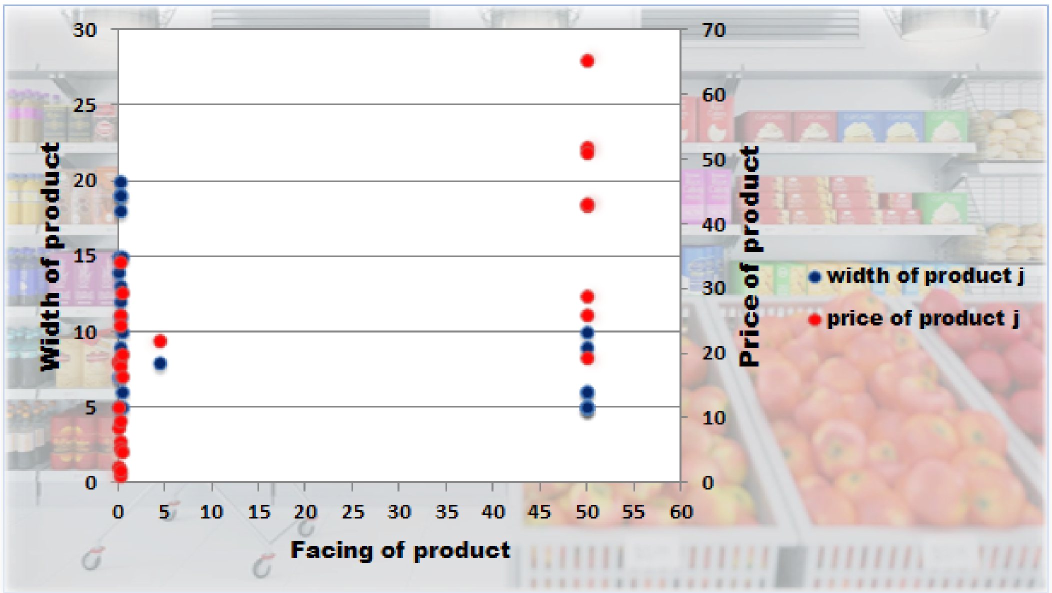

Figure 4.3: The scatter plot of facing with the parameters width and price

size on computational complexity of the model.

All experiments are performed on an Intel(R) Core i5 (4200U) 2.3GHz machine with 8192 MB RAM, and they are performed by using GAMS 24.1.3 that runs with BARON solver. BARON performs deterministic global optimization algorithms of the branch-and-bound type which are guaranteed to find the global optima under general assumption [27].

Previously, we solved the Model (4.1) by using a product group which contains 5004 products, and it was solved in 12 minutes and 35 seconds.

In the other experiments, we obtain new product groups by selecting subsets from the set of all products including 5004 randomly, and the sets of candidate products for these experiments contain 4000, 3000, 2000 and 1000 products. The demand functions for these product groups are conducted by the form of the demand function (3.18).

The upper limits which are generalized in the formulation of the model (4.1) are set to solve the model for all these product groups. The upper limits of shelf length (LSL) are designated as 12000, 10000, 8000, 7000 respectively, and the unit of LSL is centimeter. These artificial values are specified according to data sizes of the experiments. In our assumption, the required shelf length for the products decreases, while the number of candidate products decrease. The upper limit of facing (ULF) is designated as 50 for all experiments, and this is artificially large upper bound for facing amount of a product. The shelf depth is specified as 60cm in all experiments, and this is a reasonable value according to shelf standards of supermarkets.

The objective values are found as 21294.95, 18225.25, 13606.99, 9731.83 respectively, and the model solution time for all product groups is shown in Table 4.3.

Table 4.3 shows that total time for solving the model generally decreases, while the data size decreases. As a consequence, solution time of the model is directly proportional with the size of product group because the model becomes more complex to solve while the data size increases.

In this chapter, the assortment optimization model was conducted, and the model assumptions were explained in detail. We analyzed the convexity of the model, and we proved that the objective function is concave on a convex set. Thus, the model finds the global optimum solution. The model was solved by using Karush Kuhn Tucker conditions, and the relation between facing and remaining stock is presented with analytical formulations. Finally, computational experiments were demonstrated to present applicability of the model, and computational complexity analysis was made on these experiments.

Table 4.2: The outcomes of GAMS for facing of products with the corresponding width and price parameters

# of product	Facing	Width	Price
1	0.122	14	11.60
⋮	⋮	⋮	⋮
369	0.052	7	18.75
370	50	5	28.77
371	0.166	19	19.07
⋮	⋮	⋮	⋮
2911	0.164	19	34.15
2912	50	6	65.37
2913	0.100	5	2.42
⋮	⋮	⋮	⋮
3073	0.281	9	1.94
3074	50	9	42.81
3075	0.198	13	1.23
⋮	⋮	⋮	⋮
3085	0.163	11	25.90
3086	50	5	26.00
3087	0.244	20	24.41
⋮	⋮	⋮	⋮
3534	0.141	12	6.50
3535	50	5	19.37
3536	0.424	5	19.87
⋮	⋮	⋮	⋮
4314	0.115	15	18.52
4315	50	10	51.92
4316	0.306	11	18.17
⋮	⋮	⋮	⋮
4321	0.445	15	29.45
4322	50	5	50.95
4323	0.444	6	4.81
⋮	⋮	⋮	⋮
4876	0.211	18	9.64
4877	50	6	43.25
4878	0.249	19	77.06
⋮	⋮	⋮	⋮
4989	0.174	7	5.23
4990	50	8	21.92
4991	0.109	7	8.59
⋮	⋮	⋮	⋮
5004	0.385	10	16.41

Table 4.3: Total time for solving the model

Number of candidate products	Total solution time
5004	12min, 35sec
4000	4min, 1sec
3000	4min, 7sec
2000	3min, 58sec
1000	47sec

Chapter 5

Conclusion

In recent years, the number of products which are carried by supermarkets increase expeditiously, while the number of shelving and shelf spaces remain almost same. Limited and constricted shelf space necessitates organizing shelving most effectively, and assortment planning becomes necessity for all retailers. Assortment planning is the determination of the set of products which are presented at each store and specification of the inventory levels of these products for maximizing profit subject to shelf space and other possible constraints.

The main goal of this thesis was to plan assortment of a supermarket in order to maximize total revenue under the constraint of limited shelf space. First of all, we collected the data that contains information about remaining stocks, facings, product IDs, group IDs, sales amounts and prices of products for a time period of thirteen weeks. A sample that includes only ten products was chosen in order to determine the best demand model. At this stage, the log-linear regression was conducted because it always provides nonnegative demand values. After fifteen different regressions were completed, we decided that the second regression model is the best demand model because of its higher R^2 values. In order to validate the regression model, cross-validation tests were performed on the sample data, and the mean absolute percentage error (MAPE) of the cross validation tests was calculated as 155.65%. This value shows that the demand model on this dataset might estimates demands wrongly with the high probability. According to our considerations, inaccuracy of dataset that was used in these tests caused to this situation. After determining the demand model and testing it, a regression was conducted on all products in order to plan assortment of the store entirely. At the next stage, the assortment optimization model was developed to maximize total revenue under some constraints, and the solution of this model gives the optimum facing amounts for all products. Convexity analysis was performed for investigating that the solution of model is local optimum or global optimum, and it ensured that the solution is always global optimum. Because, the objective function is concave function on a convex set, and all of the constraints of model are convex. After that, the model was solved by Karush Kuhn Tucker conditions, and the relationship between decision variables was demonstrated with an analytical formulation. Finally, the model was solved for all products in the store, and some computational examples with different sized data sets were performed in order to measure computational complexity of the model. As a result, the model found feasible solutions and assigned the maximum facing to the products which have less width and high price because of claim that is earning more revenue on a small shelf area. The facing of other products became very small values because of having no more shelf space available. Then, facing is highly related with price and width of product simultaneously. Also, computational complexity analysis shows that running time of the model is directly proportional with the size of data group, and solving the model becomes more complex and requires more computational effort while the data size increases.

To conclude, we formed an assortment optimization model which consists of an empirical demand model by using real data. However, the empirical demand model does not correspond to the data completely, because some data can be missing or recorded wrongly. We developed a heuristic approach by solving model as continuous, and we reached the optimum results by this way.

One of the contribution of our study is that the optimization model consists of a justified empirical demand model. We know that the empirical demand model is not used in the most of the studies which are about assortment planning. The other contribution is developing a heuristic approach which can solve the continuous model as optimal. According to our assessments, the continues solution does not deviate too much from the integer solution.

This study provides a large domain for the future studies about the assortment planning. In our study, some data is missing, and a solution approach can be developed for this problem. Also, we neglected the substitution effect. In our regression model, there is no interaction between the products although the probability of having an interaction between the products which are in the same subcategory is high. Thus, in a different work, the interactions between products can be considered in the regression model. Furthermore, we do not consider the height of the products and the height between two shelving, and in this study, the products are only placed alongside on the shelf. A study can be done by considering the effect of the height of products and shelving. Also, the assortment optimization model is suitable for only non-perishable products because the inventory control of perishable products is more complicated. For a future work, the model can be modified and improved for assortment planning of perishable products.

References

[1] S. Pohmer. (2009, Sep.). [Online]. Available: http://www.gpnmag.com/article/somethings-happening-here-0/

[2] J. A. Quelch and D. Kenny, "Extend Profits, Not Product Lines," *Harvard Business Review*, vol. 72, no. 5, pp. 153–160, 1994.

[3] A. G. Kök and M. L. Fisher, "Demand Estimation and Assortment Optimization Under Substitution: Methodology and Application," *Operations Research*, vol. 55, no. 6, pp. 1001–1021, 2007.

[4] 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.

[5] W. G. Gilland and H. S. Heese, "Sequence matters: Shelf-space allocation under dynamic customer-driven substitution," *Production and Operations Management*, vol. 22, no. 4, pp. 875–887, 2013.

[6] C. Ulu, D. Honhon, and A. Alptekinoglu, "Learning Consumer Tastes Through Dynamic Assortments," *Operations Research*, vol. 60, no. 4, pp. 833–849, 2012.

[7] D. Honhon, S. Jonnalagedda, and X. a. Pan, "Optimal Algorithms for Assortment Selection Under Ranking-Based Consumer Choice Models," *Manufacturing & Service Operations Management*, vol. 14, no. 2, pp. 279–289, 2012.

[8] D. Honhon and S. Seshadri, "Fixed versus Random Proportions Demand Models for the Assortment Planning Problem under Stock-out based Substitution," *Manufacturing & Service Operations Management*, vol. 5, no. 4, pp. 7–24, 2014.

[9] N. Golrezaei, H. Nazerzadeh, and P. Rusmevichientong, "Real-Time Optimization of Personalized Assortments Real-Time Optimization of Personalized Assortments," *Management Science*, vol. 60, no. 6, pp. 1532–1551, 2014.

[10] 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.

[11] H. Topaloglu, "Joint stocking and product offer decisions under the multinomial logit model," *Production and Operations Management*, vol. 22, no. 5, pp. 1182–1199, 2013.

[12] P. Rusmevichientong, D. Shmoys, C. Tong, and H. Topaloglu, "Assortment optimization under the multinomial logit model with random choice parameters," *Production and Operations Management*, vol. 23, no. 11, pp. 2023–2039, 2014.

[13] A. G. Kök and Y. Xu, "Optimal and Competitive Assortments with Endogenous Pricing Under Hierarchical Consumer Choice Models," *Management Science*, vol. 57, no. 9, pp. 1546–1563, 2011.

[14] J. B. Feldman and H. Topaloglu, "Capacity Constraints Across Nests in Assortment Optimization Under the Nested Logit Model," *Operations Research*, no. February 2016, pp. 1–22, 2015.

[15] G. Li, P. Rusmevichientong, and H. Topaloglu, "The d -Level Nested Logit Model: Assortment and Price Optimization Problems," *Operations Research*, vol. 63, no. 2, pp. 325–341, 2015.

[16] A. Alptekinoğlu and A. Grasas, "When to Carry Eccentric Products? Optimal Retail Assortment under Consumer Returns," *Production and Operations Management*, vol. 23, no. 5, pp. 877–892, 2013.

[17] F. Caro, V. Martínez-de Albéniz, and P. Rusmevichientong, "The Assortment Packing Problem: Multiperiod Assortment Planning for Short-Lived Products," *Management Science*, vol. 60, no. 11, pp. 2701–2721, 2014.

[18] J. Shen, *Linear Regression.* Boston, MA: Springer US, 2009, p. 1622. [Online]. Available: http://dx.doi.org/10.1007/978-0-387-39940-9{_}542

[19] ——, *Log-Linear Regression.* Boston, MA: Springer US, 2009, pp. 1660–1661. [Online]. Available: http://dx.doi.org/10.1007/978-0-387-39940-9{_}543

[20] K. T. Talluri and G. J. V. Van Ryzin, *The Theory and Practice of Revenue Management.* Springer US, 2005.

[21] T. Fushiki, "Estimation of prediction error by using K-fold cross-validation," *Statistics and Computing*, vol. 21, no. 2, pp. 137–146, 2011.

[22] Wikipedia, "Cross-validation (statistics) — wikipedia, the free encyclopedia," 2016. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Cross-validation_(statistics)&oldid=747269850

[23] J. E. Hanke and D. W. Wichern, *Business Forecasting*, 2009.

[24] S. Boyd and L. Vandenberghe, *Convex optimization.* Cambridge University Press, 2004.

[25] W. L. Winston and J. B. Goldberg, *Operations research: applications and algorithms*, 2004.

[26] M. Onal, A. Yenipazarli, and O. E. Kundakcioglu, "A mathematical model for perishable products with price- and displayed-stock-dependent demand," *Computers and Industrial Engineering*, vol. 102, pp. 246–258, 2016.

[27] N. Sahinidis. (2015, Jun.). [Online]. Available: https://www.gams.com/latest/docs/solvers/baron/index.html

Curriculum Vitae

İpek Durmuş was born on 5 December 1991, in İstanbul. She received her BS degree in Industrial Engineering from Işık University in 2014 with 2^{nd} ranking out of all graduates. She completed her M.S. degree in 2017 in Industrial Engineering-Operations Research from Işık University. She worked as a part-time research assistant at the department of Industrial Engineering of Işık University from 2014 to present. Her research interests include operations research and assortment optimization.