MASTER

Insights in customer store choice behavior and competitive facility location strategies in the Dutch supermarket industry

Wijsman, W.G. (Gideon)

Award date:
2023

Department of Industrial Engineering and Innovation Sciences
Operation Management & Logistics (OML)
Operations, Planning, Accounting and Control Group (OPAC)

Insights in customer store choice behavior and competitive facility location strategies in the Dutch supermarket industry.

Wednesday 29th March, 2023, Eindhoven

Master Thesis Project

Supervisors
dr. Layla Martin
dr. Nevin Mutlu

Author: W.G. Wijsman
Student number: 0873677

TUE. Department of Industrial Engineering & Innovation Sciences
Series Master Theses Operations Management and Logistics

Keywords: Inverse Optimization, Parameter Estimation, Facility Location Problem, Consumer Choice Behavior, Grocery Retail Industry

Abstract

Writing Abstract

Executive Summary

Introduction This is the executive summary for the report

Preface

This Thesis concludes my master Operations Management and Logistics at the Eindhoven University of Technology and my years as a student. This thesis would not have been possible without the help of some wonderful people and I would like to take a moment to express my sincerest gratitude and appreciation toward them.

First and foremost, my first supervisor Layla Martin. Secondly, Tobias Crönert for helping me with the model.

Special thanks to Edyta Hop-Deminska from Distrifood Dynamics for providing me with data about supermarket locations info in The Netherlands.

Finally my parents and friends

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Opening a new retail store is an expensive project, as property acquisition and facility construction require large investments. Retail chains which open new stores are therefore exposed to large financial risk. Should a newly opened shop fail, the parent company loses the invested money and is also subjected to damage to its brand image. Facility location analysis is therefore a vital aspect of a retail firm's long-term strategy (Baviera-Puig et al., 2016).

Customer store choice behavior is a key factor for competing retailers to determine their store locations, as the location of the new retail store affects the consumer experience, which influences consumer patronage (Kim and Choi, 2013). Scientific literature pays considerable attention to research in product and store criteria that affect consumers' choice of retail patronage (Briesch et al., 2009), (Uncles and Hammond, 1995). It is observed that factors like distance, assortment, price, and relative distance to other points of interest can play a key role in customers' preference for certain retail chains or stores (Statista, 2022d). Since this research builds on previous studies that assume existing retail chains choose new facility locations based on these factors, the main problem lies in determining the relative importance of these factors as they can change over time, per geographical location, or even between retailers.

Market incumbents have access to vast amounts of historical sales data which they can use to incur information about the values of certain customer attraction factors. Accordingly, they use this information to determine their facility location strategy for opening new stores. As opposed to existing retailers, market entrants do not possess granular sales data and therefore it is much harder for them to retrieve information on consumer store choice behavior. As a retailer's facility location strategy depends on consumer store choice behavior, it would be much more difficult for new market entrants to choose optimal facility locations.

However, if the existing location structure of incumbent retailers is (close to) optimal, information about the values of the customer attraction parameters can be derived (Crönert et al., 2022). Crönert et al. (2022) develop a novel inverse optimization approach in their study which can identify a feasible set of values for the customer attraction parameters that best describe the observed solutions. New market entrants are then able to retrieve information on the customer attraction parameters which they can use to improve their facility location strategy, negating the data access advantage of market incumbents.

Being able to infer a retail firm's underlying facility location strategy is interesting from a variety of viewpoints. Firstly, new market entrants are able to derive information on customer store choice behavior from the market incumbents, which lowers the barriers to market entrance. Secondly, competing operating retailers could identify other underlying decision parameters used by their competitors. They could adapt their own strategy based on information about the strategy used by other players in the market. Thirdly, the derived results could be interesting from the viewpoint of confirming economic theory. And finally, it can be of practical importance in legal proceedings where spatial competition analysis can be used to provide evidence in cases in which a firm's location strategy and its competitive implications are at issue such as collusion or predation (West, 1989). This research tries to understand the market structure for the supermarket industry based on the customer store choice variables and their relative importance using the inverse optimization model proposed by Crönert et al. (2022). Through gaining insights into the form of the utility function and the relative valuations of its parameters, this study aims to obtain information on the facility location strategy of grocery retailers and provide managerial advice.

1.1 Contribution and Research Questions

Crönert et al. (2022) model the facility location strategy of grocery retailers as described in the previous section, as an integer programming game (IPG). An IPG is a simulated game where multiple decision makers simultaneously decide where to open stores, considering customers choose locations based on a parameterized utility function. The developed approach identifies the customer choice parameters of the utility function, that best explain the observed behavior of multiple competitors in a Nash equilibrium (Crönert et al., 2022). The study conducted by Crönert et al. (2022) is however limited to numerical experiments only and the developed algorithm is yet to be applied in real-life scenarios. They propose future research could extend the analysis using real-world datasets in order to gain additional insights into managerial applications of inverse optimization in practice.

The objective of this research is therefore aimed at applying the inverse optimization approach by Crönert et al. (2022) in practice. Different models are developed and tested to incur information on the underlying consumer store choice parameters. The project is executed using publicly available data, as to mimic a party without access to granular sales data. This research tries to bridge the gap between theory and practice and is, therefore, an extension of the research of Crönert et al. (2022). Theory is applied to practice in multiple case studies using real-life datasets in order to provide an explanation of facility location decisions of market incumbents in the supermarket retail industry. In order to achieve the research objective, multiple research questions are formulated that support the journey to ultimately complete the research goal. The research questions are categorized into two different segments. The first segment considers the market and retailers' perspective, whereas the second segment is concerned with the consumer perspective and decision-making.

1.1.1 Market Structure and Retailers Analysis

This section covers the research question on the market structure and retailer specifications. In order to apply the theoretical model in practice, a market analysis is necessary of the candidate geographical area, in order to verify that it is a suitable area for the inverse optimization method. In the numerical experiments of the inverse optimization algorithm performed by Crönert et al. (2022), it is assumed that all observations belong to similar regions and hence, the different case studies in this research should also show similarities.

1.a - Which geographical areas are best suited for the analysis using the inverse optimization method?

First, this study performs a literature study on the topic of game theory, applied to facility location decisions in retail industries. After this literature search, this research conducts a small market analysis that contains information on the Dutch grocery retail market. This analysis is the starting point that helps in the case studies to determine the relevant players that compete in the market area. This leads to the next topic. Since Crönert et al. (2022) assume that retail chains are in perfect competition with each other and satisfy the exact same market segments, in practice, stores have possible complementary factors such as discounters having a complementary effect with organic supermarkets, and both retail chains focus on different customer segments. Although they are both categorized as supermarket retailers, these branches actually reinforce each other as customers benefit from multi-purpose shopping. This leads to the following research question.

1.b - Which retail chains compete over the same customer segments and which retail chains play a complementary role?

For this research question, the industry analysis developed for the previous research question is also used. This analysis shows the specifics of the retail chains in the supermarket industry in each case to answer this question.

The next research question in this subsection is related to the geographic representation of the market in the case studies. The topic of this research question is quite extensive as it entails everything related to building the best representation of the digital map of each case study. A real-life geographical dimension is very different than a theoretical spatial dimension. In the numerical study, Crönert et al. (2022) assumes randomly uniform distributed customer locations and store locations. Here the distances are covered through a straight line. In practice, the geographical representation of a city, with customer and facility locations is much more difficult to determine. A straight line distance is different than a road route to the destination. Traffic throughout the day, one-way streets, and other factors could influence the geographical dimension that a simple spatial model does not incorporate. Another directly related geographical problem is that of retail locations. In practice, opening new stores on a particular location will also depend on other variables like the costs of opening a new facility at said location. Besides, in most city districts it is not possible to open a store at any possible location, as only certain spots are available for retailers to open a new facility location.

1.c - What points of interest affect customers and increase store convenience?

1.d - What are the most important transportation modes for customers?

1.e - Which measure of distance best describes the routes between customers and stores?

This research develops a geospatial map for these questions that best represents the locations of all the different grocery store locations. Besides the grocery stores, this map includes nodes for customer locations with demographic characteristics, candidate locations for opening new stores, and other points of interest nodes that affect the convenience factor of a retail location.

1.1.2 Consumer details and store choice decision making.

This section proposes and discusses research questions on consumer store choice behavior. The inverse optimization approach relies on a parameterized utility function, consisting of multiple weighted variables that describe consumer store preferences. Determining the form of the utility function thus presumes a prior general knowledge of the decision-making process of the customer and the market incumbents, as the variables to the utility function must be defined in advance (Crönert et al., 2022). This leads directly to the first research question.

2.a - Which important factors in consumer store choice behavior in the grocery retailer industry should be included in the model?

Question [2.a] will be answered through a literature study. Ample previous studies exist on this topic, especially in the behavioral and marketing sciences that try to find answers to the motivations behind consumers' choice. These studies have determined the most important factors related to consumer store choice specified for grocery retailing. The following research questions are questions that are answered by applying the inverse optimization model in practice. The values for the customer store choice parameters are examined in different scenarios in order to gain insights into parameter changes over time, in different locations, and for different retailers. Through answering these research questions, managerial advice is provided which serves the research goal of this project. The next research questions are as follows.

2.b - How does consumer behavior change over time?

2.c - Do the consumer store choice parameter values differ across geographical locations?

2.d - What is the form of the utility function for different supermarket formats?

The scope of this research will consist of choosing multiple different cities, divided into city districts, which will be studied in depth on the market structure of the incumbent grocery retailers. The algorithm is used to determine the parameter values of the utility function for different cities and different time intervals. For time-series analysis, a geographic location (city), is examined during a longer time-span, e.g. every year for 10 years, to determine how the market is behaving in that time-frame. This is especially relevant with respect to new innovations such as online shopping and new market entrants that solely focus on at-home delivery. These new innovations disrupt the current market because customers' preferences change. Grocery stores need to react accordingly to not lose any market share and retain their customers. In order to determine these changes over time, more literature could be conducted on these topics. However, because these changes are quite recent, there may not be much literature available. Research question [2.c] is of a similar nature as research question [2.b] and is concerned with the differences in parameters across different geographical locations. Research question [2.d] shows which different models perform best for different retailers. This provides information on how different retailers value certain parameters for consumer store choice behavior.

1.2 Outline and Limiting Assumptions

This research contributes to the scientific literature by using a novel approach to determining customer store choice parameter values in practice. Previous literature on customer store choice modeling mainly focuses on extracting information through maximum likelihood estimations (Zhu and Singh, 2009) and (Seim, 2006), or through regression models (Shriver and Bollinger, 2022). However, these methods all require access to vast and detailed amounts of consumer preference data using customer surveys or granular sales data, but outside parties and new market entrants often lack access to such data. With the inverse optimization model, having access to this data is no longer required to extract information about consumer preferences. Information on the parameters of the consumer store choice variables can be achieved by observing the facility locations of the current market incumbents and simulating the current market structure through an integer programming game.

As stated, the main goal of this research is to gain as much practical insight into consumer store choice behavior as possible and provide managerial results, using the inverse optimization method and using limited, publicly available data. However, conducting this research has some limitations.

The first limitation is that this research is solely focused on the Dutch grocery retail industry. This is a significant market in itself and interesting to analyze. Also, obtaining geographical location data from other countries, especially outside Western Europe, is more difficult to obtain since this research partly relies on OpenStreetMap as a data source for determining retail store locations. Lacking central supervision, data uploaded to volunteered geospatial information (VGI) types of services is collected by voluntary individuals who might not have any professional background. This means that data quality and validity assurance of such information systems has always been an area of concern (Teimoory et al., 2021). Since the data quality of VGI systems is to be assumed lower in poorer regions, these regions are not of interest to this research.

Another reason for choosing only a single country scope is that markets can differ significantly between countries. Kilroy et al. (2015) discuss in their article certain barriers for modern retailers in emerging markets as success in such markets is not guaranteed. They argue how international retailers that wish to succeed in foreign markets need to become experts at local tailoring (Kilroy et al., 2015).

Besides the geospatial scope of this research, the scope is also only limited to the supermarket industry. Other industries, e.g. clothing retailers, benefit from conglomerating in the same vicinity.

Another limitation in this study lies in the application of the inverse optimization algorithm is only limited to the market leaders and to large-sized cities. As retail chains should have enough facilities present in the designated area for the algorithm to properly work. In other words, a minimum threshold exists for the number of data points as the algorithm needs enough data points to yield statistically relevant results. As retail chains operating in the area require a certain number of observations, other parties in the market which do not own the minimum amount of facilities play a non-active role in the IPG. They own store locations and these locations are filled in the game but the retail chains do not make decisions or adapt their strategy.

The final limiting assumption in this research states the model assumes homogeneity across revenue and costs for each retailer. The costs and profit functions are assumed to be equal for every player participating in the IPG. In practice costs and revenue may vary due a variety of factors.

Chapter 2

Literature Background

This chapter of the report provides the necessary background information to create a good understanding of the research process and the modeling techniques. This chapter is divided into four sections where the first section provides an introduction to game theory and discusses concepts such as Nash equilibria. The second section covers facility location strategy and concepts such as Huff models. The third section discusses consumer store choice factors which is needed for the formulation of the utility function in the models described in chapter 3.2.1.

The final section briefly covers inverse optimization and discusses the parameter estimation method developed by Crönert et al. (2022) in more detail.

2.1 Introduction to Game Theory

This section briefly covers the basics of game theory, specifically, it introduces the concepts of Nash equilibria (Nash, 1951) and ϵ-equilibria (Daskalakis et al., 2006) which are important concepts for this research. Game theory is the mathematical discipline that studies the behavior of (rational) decision-makers called players, whose decisions affect themselves as well as others (Aumann, 2016). Game theory can be used to explain past events, predict future actions by the involved players, and can be used to model a variety of real-world scenarios like negotiations, pricing strategies, and new product decisions. The most important assumption in game theory is that players are rational actors that act self-interested and utility-maximizing. Although this assumption is one of the foundational assumptions of many economic models, it is however directly one of the key limitations since people in the 'real world' are emotional beings who will not always make rational decisions (Akerlof and Yellen, 1987). Game theory has different forms as one can differentiate between cooperative and non-cooperative games, where the former deals with how groups of people interact to achieve a common purpose and the latter describes how individual players interact in a situation where they act in their own interest and try to achieve their own goals. Another important classification is simultaneous games vs. sequential games. In simultaneous games, all players take decisions at the same time and are unaware of the decisions of other players (Aumann, 2016), which is in contrast with sequential games where players make their decision one after another. Players in a sequential have information on the moves of players who have already adopted a strategy.

Nash Equilibrium

Nash (1951) defines in his article a mixed-strategy equilibrium for any game with a finite set of actions and he proves at least one equilibrium point must exist. This equilibrium point, a Nash equilibrium, is achieved in a game if no player has any incentive for deviating from their own strategy. In other words, no player receives any benefit from changing their actions, assuming other players remain constant in their strategies. A game can have a single, multiple, or no Nash equilibrium at all (Nash, 1951).

Consider a game with n players. Let s_i, be a strategy of player i and let $S_i = \{s_i^1, \ldots, s_i^m\}$ be the strategy set of player i, and player i has m possible strategies. Then, let $s = (s_1, \ldots, s_n)$ be the strategy profile of all n players, which represents the outcome of the game depending on all chosen strategies by the individual players. Consider $s_{-i} = (s_1, \ldots, s_{i-1}, s_{i+1}, \ldots, s_n)$ as the strategy profile of the other $n-1$ players. For convenience, $s = (s_i, s_{-i})$. The payoff of player i is written as $u_i(s_i, s_{-i})$, a function of the strategy profile played by the n players in the game. Finally, S describes the set of all possible strategy profiles.

Consider a strategy s_i for player i that is a best response to the strategy profile s_{-i}, where it is important to note that there can exist multiple 'best response' options at the same time. This holds if $u_i(s_i, s_{-i}) \geq u_i(s_i', s_{-i})$ for all $s_i' \in S_i$. Note that s_{-i} denotes one of many different strategy profiles that could be played by the other players. As there can be multiple best responses to s_{-i}, the set of best responses is denoted by $BR_i(s_{-i})$ for player i to s_{-i}. Also, note that $s_i \in BR_i(s_{-i})$.

The strategy profile $s^* = (s_1^*, \ldots s_n^*)$ is a Nash Equilibrium if each player's strategy is a best response to the strategy profile played by the other players in the game. I.e., s^* is a Nash equilibrium if $s_i^* \in BR_i\left(s_{-i}^*\right)$ for all players i, or equivalently, if $u_i\left(s_i^*, s_{-i}^*\right) \geq u_i\left(s_i, s_{-i}^*\right)$ for all $s_i \in S_i$ and for all players i.

Example 2.1.1. Consider a duopoly market with 2 players, Company A and Company B. Both companies have the option to expand their product line and they want to determine whether they should make use of the expansion option (differentiation) or keep the product line as it is (homogeneous). If the two companies differentiate their products, both increase their revenue by 100. If only one company decides to expand its product line, it will increase its revenue by 200, while the other company has no increased revenue. If both Company A and B decide not to differentiate, neither company will increase its revenue. This situation is modeled with a payoff matrix shown in Figure 2.1.

Company A \ Company B	Differentiation	Homogeneous
Differentiation	100 \ 100	200 \ 0
Homogeneous	0 \ 200	0 \ 0

Figure 2.1: Payoff Matrix with Nash Equilibrium - Differentiation vs. Homogeneous.

Consider the situation in the bottom right corner of Figure 2.1 where both companies do not differentiate their products. Company A can improve its payoff from 0 to 200 by choosing to differentiate its product line and moving to the top right corner. Likewise, Company B considers this strategy from Company A. Company B will also choose differentiation because it can improve its payoff from 0 to 100 by moving from the top right to the top left corner. Hence, both companies will always choose the differentiation option. No matter in which of the 4 scenarios they start, they will always end up in the top left corner. So, the scenario in which both companies differentiate their products is a Nash equilibrium.

Simultaneous and Sequential Games

The previous example 2.1.1 shows a game following a simultaneous decision-making process. In example 2.2 the previous situation now follows a sequential decision-making process and is rewritten to extensive form. Extensive-form games are strategic situations in which the players take sequential actions, e.g. a game of chess. They are represented using decision trees which describe the sequence of decisions, the information players have when making decisions, and the associated output or payoff of each decision path.

Example 2.1.2. Consider example 2.1.1 in the previous section in which two companies in a duopoly market have the opportunity to expand their product line through differentiation. In this slightly adapted example the game is represented in extensive form, shown in figure 2.2. In this situation of sequential decision-making, Company A as the market leader chooses its strategy first and Company B makes a decision based on the strategy chosen by A. The associated payoffs with each decision are as follows: If both firms choose to differentiate their products, Company A receives a payoff of 200, and Company B receives a payoff of 100. If only one company chooses to differentiate they capture the entire market and receive a payoff of 300 whereas the other receives nothing. If both companies opt for the homogeneous option neither receives any payoff.

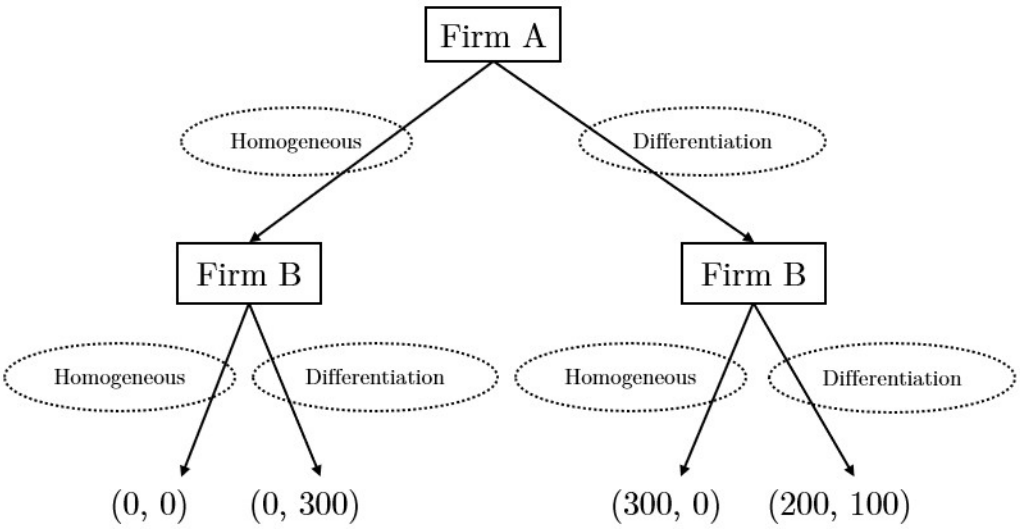

Figure 2.2: Sequential Game in Extensive Form - Differentiation vs. Homogeneous.

Through the process of backward induction, the Nash equilibrium of a sequential game can be obtained. Firstly, identify the player's best response at the final subtrees. The best response for every subtree is called a subgame perfect Nash Equilibrium (SPNE). A strategy profile is a subgame perfect equilibrium if it represents a Nash equilibrium of every subgame of the original game. (Guo, 2021). The associated payoff with the best response at each subtree (the observed SPNE) then becomes the payoff of the parent subtree. This process is then repeated until the first decision node is reached. In this example, the observed SPNE is achieved if Company B chooses the Differentiation option in both subtrees, as the payoff corresponds to (0,300) and (200,100). The observed SPNE in the parent tree is that Company A also chooses Differentiation over the Homogeneous option. Since the decision node of Firm A is the final node in the decision tree, it is also the Nash equilibrium solution to the initial problem.

In this example, when both companies choose to differentiate their product lines Company A expects a higher payoff than Company B (200 vs. 100). The difference in payoff is attributed to the First-Mover Advantage, as Company A decides first to differentiate the product line and Company B follows later.

ϵ-Equilibrium

Nash (1951) proves the existence of at least one equilibrium in any finite game with mixed strategies, which is useful in practice as most real-life applications of strategic game theory are modeled as situations with a finite set of n players, each with a finite set of actions. However, finding the Nash equilibrium is computationally particularly difficult for large problem instances, as it requires modeling every different scenario to determine the payoff for each player (Ceppi et al., 2010). However, as is also the case with many general complex optimization problems, finding a "good enough" solution that is near the optimal solution is often sufficient and common practice. Daskalakis et al. (2006) discuss the concept of Approximate Nash Equilibria which is a sub-optimal solution to a Nash Equilibrium. Recall the definition of a Nash Equilibrium meaning "no incentive to deviate from strategy", then the definition of Approximate Nash Equilibrium is somewhat as "low incentive to deviate". This definition of an approximate Nash equilibrium is captured in the form of an epsilon-equilibrium (ϵ-Equilibrium) by Daskalakis et al. (2006). An ϵ-equilibrium or near-Nash equilibrium, is a strategy profile that approximately satisfies the condition of a Nash equilibrium in the sense that no player has any incentive to switch strategies. In an ϵ-equilibrium, this condition is weakened so that players may have a small incentive to deviate from their strategy. The absolute advantage that players may obtain by deviating is measured with ϵ. Approximating a Nash Equilibrium using this technique is considered to be an adequate solution concept and may sometimes be preferred over finding the actual Nash Equilibrium due to it being relatively easier to compute.

Recall the formal mathematical notation of a Nash equilibrium in an N player game where each player chooses their strategy s_i to maximize their payoff or utility $u_i\left(s_i^*, s_{-i}^*\right)$ which depends on their own decision of playing strategy s_i^* and the decision of their $N-1$ competitors playing s_{-i}^*. A Nash equilibrium in which no player benefits from deviating from their strategy is written as follows (Nash, 1951).

$$u_i\left(s_i^*, s_{-i}^*\right) \geq u_i\left(s_i, s_{-i}^*\right) \quad \forall s_i \in S_i, \forall i \in I \tag{2.1}$$

As discussed, the ϵ-equilibrium approximately satisfies the condition above with the addition that no player can gain more than ϵ from unilateral deviation (Daskalakis et al., 2006). The Nash Equilibrium equation changes to the ϵ-equilibrium equation shown below.

$$u_i\left(s_i^*, s_{-i}^*\right) + \epsilon \geq u_i\left(s_i, s_{-i}^*\right) \quad \forall s_i \in S_i, \forall i \in I \tag{2.2}$$

2.2 Competitive Facility Location Problem

The Competitive Facility Location Problem is an adaptation of the general Facility Location Problem, that focuses on retail and other facilities that operate in a competitive environment where the objective is to maximize profit or market share. The basic concept is to determine the optimal location of one or more new facilities in a market area where competition, and/or one's own chain facilities, already exist or will exist in the future. In this scenario, it is assumed that profit increases when market share increases, and thus the objective of the location problem is to maximize the market share (Eiselt and Marianov, 2017).

Since multiple retail chains operate in the same market area, the location decision of a firm or retail chain does not only affect its own market share but also its competitors' market shares, and hence, the competitive facility location problem is modeled with concepts from game theory discussed in section 2.1. The game is of a non-cooperative form and can either be modeled as a simultaneous or sequential game, depending on the available information.

History of research on competitive facility location problems goes back to Hotelling (1929), who models competition in a spatial context where firms share the market using a simple model, illustrated in Example 2.2.1. He considers duopolists, who each locate their firm on a fixed point along a line segment, referred to as "main street." Each duopolist aims to maximize its profit by increasing its consumer pool or market share. This results in the two firms choosing to locate at the mid-point of the line. A firm that unilaterally moves away from the mid-point loses customers and market share.

Example 2.2.1. Consider the two duopolists from example 2.1.1, Company A and Company B. They both operate in a market represented as a straight line that stretches on the 0-1 interval. Customers are uniformly distributed along that interval. Customers choose the vendor closest to their location and split themselves evenly if both vendors choose the same location. Each vendor wants to maximize its number of customers to maximize its market share. The initial situation in Figure 2.3 depicts firm A choosing an arbitrary position x_1 along the line. Since he is currently the only player, the entire market share is captured by Company A. Figure 2.4 shows the next situation in which a new player, Company B, enters the market at position x_2. Since customers choose the company closest to their location the market share by Company B is given through $M_B = \frac{x_2+x_1}{2}$, and the market share for company A is $M_A = 1 - M_B$ since the entire market is equal to 1.

Figure 2.3: Hotelling's Law: Initial situation with only 1 company

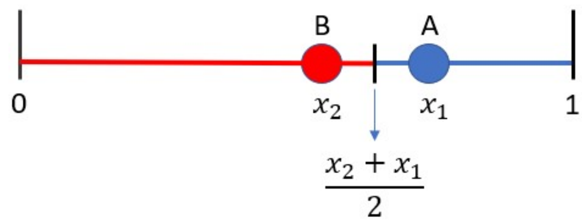

Figure 2.4: Hotelling's Law: Situation with a new market entrant

Since Company B will now have a much larger market share than Company A, Company A in turn will move right next to the other side of Company B, in order to gain Company B's market share. In response, Company B will to the exact same to re-establish its loss. This cycle of moving will continue until an equilibrium is reached when both firms locate arbitrarily close at the center of the line. See Figure 2.5.

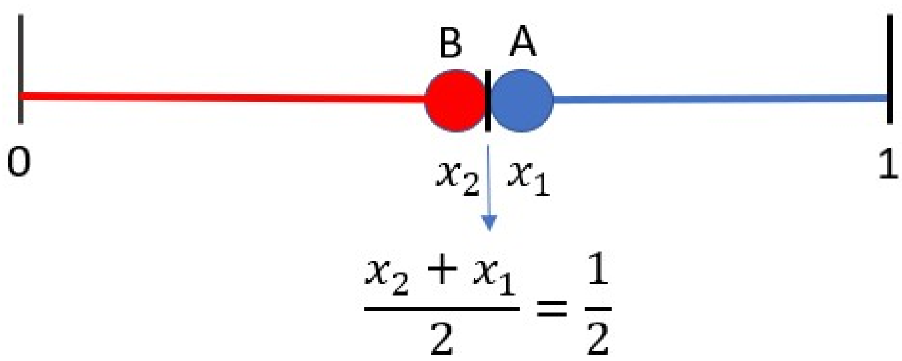

Figure 2.5: Hotellings's Law: Final Situation - Nash Equilibrium

It is not for much later that Reilly (1931) proposed his retail gravitation theory. This model explains and predicts consumer shopping patterns using Newton's law of gravitation. The latter describes how particles in the universe attract each other with a force depending on their mass and relative distance, which forms the basis of the formulation of Reilly (1931), which states that large retail zones that have high levels of attractiveness, capture a larger market share. This is attributed to the assumption that customers are willing to travel longer distances to these retail centers. This gravitational model is later extended by David L. Huff (1964) who transformed the deterministic models of retail gravitation into statistical explanations as shown in equation 2.3. Huff-like gravity models operate on the principle that the probability of customers visiting a certain retail zone is a function of its attractiveness and distance to customer locations, as well as the distance and attractiveness of competing sites in the vicinity.

$$P_{jk} = \frac{A_k / D_{jk}^{\beta}}{\sum_{k=1}^{n} \left(A_k / D_{jk}^{\beta} \right)} \tag{2.3}$$

Where:

- P_{jk} = The probability of consumer j shopping at store k.
- A_k = The measure of the attractiveness of each retail store or zone k.
- D_{jk} = The distance from consumer j to store or zone k.
- β = A distance decay parameter

Solving facility location problems as an Integer Programming Game

The facility location problem is formulated as a mathematical optimization problem and solved using integer programming (IP). Integer programming, which is of special interest to this research, is a form of linear programming (LP) where some of the variables are restricted to be integers. The facility location problem is a classical optimization problem where the objective of the problem, captured in the objective function, depends on the problem context but often relates to minimizing costs or maximizing profit or service. Problem constraints impose certain limitations or conditions on the variables that are required to be satisfied. These often relate to budget limitations, production output, and capacity. Optimization problems are usually expressed in matrix form shown in equation 2.4, or in canonical form shown in equation 2.5. This research builds on the study by Crönert et al. (2022) who model the retail store location decisions as an Integer Programming Game (IPG), combining game theory in section 2.1 with the facility location strategy described in this section. An IPG models a situation where multiple decision-makers interact with each other by simultaneously choosing strategies, in this case choosing facility locations, that impact themselves and the other players. Section 2.4.2 explains the IPG used for this research in more detail

$$\max \left\{ \mathbf{c}^{\mathrm{T}} \mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n \wedge A\mathbf{x} \leq \mathbf{b} \wedge \mathbf{x} \geq 0 \right\} \tag{2.4}$$

$$\begin{array}{ll} \text{Maximize} & c^T x \quad \text{(w.r.t. } x\text{)} \\ \text{subject to:} & Ax \leq b \\ \text{and:} & x \geq 0 \end{array} \tag{2.5}$$

2.3 Consumer Store Choice Behavior

Consumer store choice behavior is the study field concerned with research on how and why consumers choose certain stores over others, which is particularly interesting for this research as the goal is to identify customer behavior in this industry. This section summarizes past research on this topic in an effort to build a basis on which this thesis can build. One key finding is that not only does the consumer's perception of directly store-related attributes like pricing, product range and quality, or market positioning strategy have a significant impact on consumer behavior, but also the geographical location of the store itself affects consumer choice substantially (Erath et al., 2007). This section starts by explaining general utility theory and utility functions which is a necessary theory for the development of different models in section The rest of this section is to find the motivations for supermarket store choice and to identify the most important store attributes observed by previous studies.

2.3.1 Utility Functions

Utility theory is an important concept in economics and game theory and is used to describe or model value or attractiveness. It is based on the notion that each rational consumer $j \in J$ chooses the option which yields the highest utility, out of the set of all available options $k \in K$ (Reutterer and Teller, 2009). The utility function, denoted in 2.6, is a central concept in utility theory and represents the obtainable utility as a function of all variables for each specific set of choices available to the consumer.

$$U(x_1, x_2, \ldots, x_n) \tag{2.6}$$

In most practical cases, the objective is to find the best alternative out of a set of choices that yields the highest amount of utility for the consumers. Another objective that is more relevant to this thesis, is maximizing the utility under certain constraints where no clear set of choices is defined, which means the problem can be formulated as a linear programming problem as described in section 2.2. A short example described in 2.3.1 shows the working of a choice problem using a utility function.

Example 2.3.1. Suppose an arbitrary consumer has a utility function of the form $U = x^2 + 3y$. Where x represents the number of apples and y represents the number of oranges. The consumer has the choice between two alternatives A and B. Alternative A consists of 10 apples and 5 oranges, and alternative B has 8 apples and 12 oranges. Inserting the values for x and y in the utility function yields for A: $U = 10^2 + 3 \cdot 5 = 115$, and for alternative B: $U = 8^2 + 3 \cdot 12 = 100$. So, alternative A yields the highest utility and a rational consumer would choose this option.

Simplistic utility models assume that goods are an object that directly yields utility, see example 2.3.1 which shows how utility is obtained through apples and oranges as consumable goods. In the example, for an arbitrary consumer apples have a higher utility than oranges and therefore the consumer will prefer to consume apples. However, this raises the question of why the consumer prefers apples over oranges. To address this issue, Lancaster (1966) developed an approach that assumes utility is achieved not through goods in themselves but through the properties or characteristics such goods possess. For example, a certain consumer may prefer price over the nutritional value of a product, and if apples are cheaper than oranges, even though they possess the same nutritional value, the consumer will receive more utility from consuming apples than oranges. This theory is applied throughout this thesis as it is assumed consumer store choice behavior is based on a utility function. Rational customers choose the store that yields the highest utility and this thesis tries to understand which store attributes are present it the utility function.

2.3.2 Consumer store choice factors

The study conducted by Nilsson et al. (2015) provides a small overview of previous literature that studies the importance of retail store attributes and their effect on consumer store choice. This thesis extends the overview provided by Nilsson et al. (2015) by adding new studies on this topic and adapting their overview by sorting previous literature according to the geographical location where the study was conducted. The reason for sorting between different geographical zones is that demographic, cultural, and economic differences between countries could significantly affect store choice behavior as the consumers' perception of relative store attribute importance changes (Uusitalo, 2001). This assumption holds as research conducted by Arnold et al. (1983) shows that supermarket choice attributes are indeed different across US cities (Cleveland, St. Louis, and Tampa), as well as cities in different countries (US, UK, and The Netherlands). More recent studies conducted by e.g. Nilsson et al. (2015) show how the relative importance of store attributes changes for consumers across countries as they found attributes related to accessibility (e.g. accessibility by car and parking options) are more important to Swedish consumers. Contrary to studies conducted in the United States (e.g. Carpenter and Moore (2006) which found store attributes linked to product range and quality are more important to US citizens. The adapted overview of Nilsson et al. (2015) is located in A. The bulk of studies on this topic shows that several grocery store attributes have a significant effect on consumer store choice behavior, e.g. Wong and Dean (2009) shows product quality is a very important attribute and the impact of price level is examined by Mitchell and Harris (2005) and Baltas and Papastathopoulou (2003). Product and service quality are the most important factors according to Reutterer and Teller (2009) but according to studies conducted by Carpenter and Moore (2006) and Uusitalo (2001), product supply is the most important factor. The aforementioned examples show that even though a vast amount of literature exists on this topic, it is apparent there is a certain lack of consistency among the studies on the relative importance of store attributes. Since the list of possible relevant store attributes is extensive (e.g. price, product and service quality, opening hours, fast-checkouts), Nilsson et al. (2015) have also provided a table with the most common main store attributes and elements that could be included in each main attribute concept, see table 2.1. In this table, the attributes correspond to a number and the literature overview table in Appendix A shows which attributes are examined by each previous study by providing the corresponding numbers.

Accessibility and attractiveness

In order to create a better understanding of the attributes, Nilsson et al. (2015) categorizes store attributes into two distinct classes, accessibility, and attractiveness. Roughly speaking, this distinction is based on the controllability of the retailer on the attributes. Retailers can influence the attractiveness attributes, which consist of attributes like product price, range, and quality, but also attributes such as service quality and store layout. Unlike the controllable elements of attractiveness attributes, retailers can much less exert control over attributes associated with the accessibility category, which includes the relative location to other shops, public transport availability, and accessibility by car like the number of parking spaces in the vicinity. Table 2.1 shows how Nilsson et al. (2015) organizes store attributes in ten main categories. The list of included elements is still far from being complete, e.g. number of available shopping carts, paying for grocery bags, or having an in-store bakery or hard-liquor section are not included in the list but are just a few of the elements that could be added. It would take an extensive list to include all possible elements, which would not be feasible, and naming every possible alternative would not be adding much value as most of the elements can be categorized in one of the ten main attributes. The main attribute categories are divided into the two classes of accessibility and attractiveness. Attribute categories 1, 2, 3, 4, 5, and 7 are best described as attractiveness attributes and categories 6, 8, 9, and 10 are more related to accessibility. In practice, the intrinsic relationships

Attribute	Elements	Attribute	Elements
(1) Product Quality	Product quality Organic products Exotic products	(2) Product Range	Available products Alternative products Alternative brands of same product
(3) Price	Product price level Promotions and discount Loyalty programs	(4) Service Quality	Overall service quality Information services Self-scanning
(5) Storescape Quality	Store layout Cleanliness Navigation easiness	(6) Closeness other facilities	Complementary stores Food services ,Liquor stores
(7) Secondary facilities	Child-friendliness facilities Restrooms Handicap-friendliness	(8) Availability	Opening hours Closeness to home On work/home route
(9) Accessibility motor vehicles	Easy to reach with car Parking spaces Free parking	(10) Accessibility other modes	Public Transport Bike foot

Table 2.1: List of main retail store attributes and elements included in each conceptual attribute. Copied from Nilsson et al. (2015)

and effects of certain attributes on consumer behavior are more complicated but combining attributes with similar characteristics is very useful for modeling purposes, for example, Crönert et al. (2022) uses three variables in the utility function to explain consumer store choice behavior where store attributes are generalized under a 'brand' (attractiveness) factor, and a 'convenience' (accessibility) factor. Section 2.4.2 explains the exact construction of the utility function and variables used in their study. This research also generalizes multiple store attributes under single variables for modeling purposes. The construction of different utility functions and variables is further explained in the Methodology section 3.

2.4 Parameter estimation through Inverse Optimization

This section combines the elements of the previous sections and covers the study conducted by Crönert et al. (2022) in more detail. But first, this section explains the concept of inverse optimization which is the most important modeling technique that this thesis uses. Then, the study of Crönert et al. (2022) models the competitive facility location problem as described in 2.2 as an integer programming game (IPG). But, instead of determining the equilibrium of the IPG, also referred to as the forward problem, the study focuses on deriving the parameter set that best describes the observed equilibrium. Through extending prior inverse optimization methods for integer programming games, Crönert et al. (2022) can estimate the parameters that best explain the observed behavior of multiple competitors in a Nash equilibrium.

2.4.1 Inverse Optimization

As the name indicates, inverse optimization is a process that is the opposite or "reverse" of traditional mathematical optimization. But instead of computing the optimal solution from a given objective function and set of constraints, inverse optimization takes decisions as input and determines the objective and/or constraints that would make the given set of decisions optimal. Based on the knowledge of the observed solution, inverse optimization makes it possible to infer information about the unknown parameters that are present in the constraints and objective function of an optimization problem (Chan et al., 2021).

Recall the basic concepts of Linear programming from section 2.2, where the main goal is to find the optimal value of the decision variable, by maximizing or minimizing the objective function. Accordingly, the solution must satisfy certain conditions, which are a set of equations named constraints. The general notation for LP problems is as follows:

$$\begin{array}{ll} \text{Maximize} & c^T x \quad \text{(w.r.t. } x) \\ \text{subject to:} & Ax \leq b \end{array} \tag{2.7}$$

Here x corresponds to the set of decision variables and A, b, and c are the unknown parameters. In an inverse optimization framework, the solution to the problem x^* is known, and the model parameters are not. The notation for the inverse optimization problem is quite similar to the forward problem equation.

$$\begin{array}{ll} \text{Maximize} & c^T x^* \quad \text{(w.r.t. } A, b, c) \\ \text{subject to:} & Ax^* \leq b \end{array} \tag{2.8}$$

A, b, and c have become the decision variables in the new problem, whereas x^* is now a known parameter. The notation in the above-shown equations shows the general idea behind inverse optimization, or more specifically, inverse linear programming. Considering a feasible solution x^* to an optimization problem with unknown parameters A, b, and c, inverse optimization is used to find the values of the parameters that would lead to the optimal solution. It is important to note that in inverse optimization literature it is common practice to keep referring to parameters of the forward problem, here A, b, c as parameters even though they represent decision variables in the inverse problem. The (decision) variables in the forward problem, here denoted as x, are continued to be called variables in the inverse problem although they are now parameters. This report will consistently hold to this form of notation to avoid confusion. For further convenience, a list of all symbols, parameters, and variables is provided in Appendix C.

This research is particularly interested in inverse Mixed Integer Linear Programming (invMILP), which is a more complex form of inverse optimization than inverse optimization of Linear Programming (LP) problems as it involves both continuous and integer variables. Even more importantly, the duality theory for MILPs is much more sophisticated than duality in LPs. In linear programming, the primal problem and the dual problem, which are both linear programs, are related by strong duality. Strong duality states that if there exists an optimal solution to the primal problem, then the dual problem also has an optimal solution with the same optimal objective value. However, in MILP, the primal problem involves both continuous and integer variables while the dual problem is a linear programming problem, not restricted to integers. Because of the extra constraint of integer variables in the primal problem, solutions to the primal and dual problems may not have the same optimal value (the duality gap) and thus, strong duality does not hold. The fact that strong duality does not hold makes the MILP duality theory far more complex and finding solutions is more difficult since strong duality problems cannot be exploited. Finding a feasible solution for a mixed integer problem can be difficult since solving the integrality constraint is not always straightforward and may require additional techniques such as branch-and-bound or cutting plane methods.

2.4.2 Estimating parameters through Inverse Optimization

Consider a network of potential store locations $k \in K$ and customer locations $j \in J$, using a data set of multiple observations $o \in O$ of store locations of incumbent retail chains $i \in I$ in said given network. Observation o represents the network for a single snapshot in a time frame or for one out of multiple (similar) network regions. For each observation o, $\hat{x}_{ik}^{o} = 1$ represents a decision variable that indicates if retail chain i operates in location $k \in K$. Variable $x_{ik}^{o} \in (0, 1)$ represents an alternative location decision available to retail chain i.

This research assumes that retailers use customer store choice factors for determining store locations. The customer store choice factors are represented as parameterized variables in a utility function. The utility function used for the inverse optimization approach developed by Crönert et al. (2022) is of the form:

$$u_{ijk}^{o} = \beta_i + \alpha \tilde{d}_{jk}^{o} + (1 - \alpha)\tilde{g}_k^{o} \tag{2.9}$$

Where u_{ijk} is the utility of customers in location j receive from patronizing facility k from retail chain i in observation o. The form of this utility function is further covered in section 2.3.1 as it is not important for the remainder of this section.

Forward Problem: Simultaneous Location Selection The goal of each retail chain i is to maximize its own profit. The profit $\Pi_i^o\left(\mathrm{x}_i^o, \hat{\mathrm{x}}_{-i}^o\right)$ retail chain i can attain in observation o depends on both its own strategy, as well as the chosen strategies of all other players except i. The strategy of retailer i in observation o is given by $\mathbf{x}_i^o := (x_{ik}^o)_{k \in K} \in S_i$. And $\hat{\mathrm{x}}_{-i}^o$ denotes a combination of strategies for all players except i, during observation o. The profit function of player i, depends on both its own strategy as well as the combined strategies of its competitors.

Inverse Problem: Parameter Estimation In the forward problem, the goal for each retailer is to maximize their profit by choosing the best set of locations for opening retail facilities. The goal of the inverse problem is to identify a set of parameters (α, β) that best explains the currently observed location structure of all retail chains as the (near) optimal outcome of the forward problem. Consider the situation where a retail chain i could increase its profit by opening a facility at a new location, then through switching strategies, the retail could improve on the initial solution. Since adopting this new strategy holds for every player i, the unilateral improvement potential variable δ_i^o is introduced in equation 2.10.

$$\delta_i^o = \max_{\tilde{\mathbf{x}}_i^o \in S_i} \Pi\left(\tilde{\mathbf{x}}_i^o, \hat{\mathbf{x}}_{-i}^o\right) - \Pi_i^o\left(\mathrm{x}_i^o, \hat{\mathbf{x}}_{-i}^o\right) \quad \forall i \in I, \forall o \in O \tag{2.10}$$

The unilateral improvement potential is equal to the profit function depending on the best deviating strategy chosen by player i, minus the profit value of the current set of strategies of all players including i. In this inverse optimization problem it is chosen to minimize the cumulative deviations over all observations, $\epsilon = \min_{\alpha,\beta} \|\delta\|$. Here, the ϵ refers to the 'noisy observations' or error term in the inverse optimization problem. As it is assumed that the existing infrastructure of store locations is near-optimal, solving equation 2.10 would yield the decision parameter set $(\alpha, \boldsymbol{\beta})$ that could be represented as the input parameters for the forward problem.

Solution Approach Since full enumeration of all possible alternative strategy combinations is unmanageable in realistically sized problems, a cut-generation algorithm is proposed for practical research. The goal of the algorithm is to determine a feasible parameter set $(\alpha, \boldsymbol{\beta})$ that solves the inverse optimization problem within a manageable time frame. This approach is derived from the paper by Wang (2009) where the problem is split into a master problem, which identifies a suitable set of parameters $(\alpha, \boldsymbol{\beta})$ through minimizing the unilateral improvement potential $\boldsymbol{\delta}$ across all players, and multiple sub-problems that generate cuts in the solution space to identify the best set of parameters that define an (approximate) equilibrium in the inverse programming game. The master problem represents a relaxed version of the inverse problem 2.10. In more detail: Minimize the unilateral improvement potential $\boldsymbol{\delta} = (\delta_i^o)_{i \in I, o \in O}$ between the observed solution $\hat{x}_i^o$ and the optimal solution $\mathbf{x}_i^o$ for all given observations $\hat{\mathbf{x}}$ and across all observations $o \in O$ and players $i \in I$. This is achieved through choosing the estimation parameters (α, β) such that: $\min_{\alpha,\beta} ||\delta||$. However, instead of ensuring that δ_i^o are minimal for all alternative strategies $\bar{x}_i \in S_i$, we relax 2.10 to apply to the enumerated subset $\bar{S}_i$ of S_i only $\left(\bar{S}_i \subseteq S_i\right)$:

$$\delta_i^o \geq \Pi_i^0 \left(\tilde{\mathbf{x}}_{\mathrm{i}}^0, \hat{\mathbf{x}}_{-1}^0\right) - \Pi_1^0 \left(\mathbf{x}_{\mathbf{i}}^0, \hat{\mathbf{x}}_{-1}^0\right) \quad \forall i \in I, \forall o \in O, \overline{\mathbf{x}}_{\mathrm{i}}^o \in \bar{S}_{\mathrm{i}} \tag{2.11}$$

Through solving the master problem an initial feasible parameter set $(\alpha, \boldsymbol{\beta})$ is obtained. The initial relaxed optimality conditions are then iteratively reconstructed through the forward problem. This is the sub-problem and works as follows: The parameter set obtained by solving the master problem is now used as input to the initial forward problem and is solved to optimality for every player and observation which returns an integer solution $\mathbf{x}_i^o$. This enumerated solution from the sub-problem is now added as a cutting plane to the master problem:

$$\bar{S}_i = \bar{S}_i \cup \{\mathbf{x}_i^o\} \quad \forall i \in I, \forall o \in O \tag{2.12}$$

All parameter sets obtained through solving the master problem must ensure an optimal solution when compared with the obtained enumerated sub-problem solutions. When the sub-problem does not yield in any new cuts for the master problem (no new enumerated solutions can be added to the set $\bar{S}_i$, the algorithm converges. This means that the current solution, parameter set $(\alpha, \boldsymbol{\beta})$, is the optimal solution to the master problem. Figure B.1 depicts the flowchart of the proposed inverse optimization algorithm.

Chapter 3

Methodology

This section explains the methods used to answer the research questions formulated in the introduction chapter 1. This chapter begins with a market analysis of the Dutch grocery retail industry in section 3.1.2. This market analysis is necessary to determine the specifics of the case studies that are later performed in chapter 5. The market analysis starts with a short description and key figures on the supermarket industry followed by a Porters' fiver forces analysis. This analysis provides insights into the market structure which is used to answer the research questions on this topic accordingly. Section 3.2.1 describes the different models that are developed. These models are then tested using different scenarios that are described in the case studies chapter 5.

3.1 Theoretical Market Analysis

This section covers the theoretical analysis of the Dutch supermarket industry, where the purpose of this analysis is to gain information on important aspects such as the market structure and the dominant players. First, a global view of the market is formed by gathering data from websites such as Statista and industry reports that help in best describing the current market structure. Grocery stores in the Netherlands are sorted into three main categories which are listed below and each category represents a different format.

- Traditional premium supermarket chain stores: Stores operating under a large corporate umbrella such as Albert Heijn, Jumbo, Coop, and Plus. These chains have multiple stores operating in many different cities.
- Discounters: There exist some grocery retail chains in The Netherlands focusing on low prices. They possess certain characteristics that differentiate them from traditional supermarket chain stores in price, product range, and service quality. The largest discount chains are Lidl and Aldi both of German origin. Many shoppers both shop at discounters and traditional supermarkets for a mix of both cheaper essentials and more premium products.
- Organic grocery retailers: These high-profile specialty retailers such as Ekoplaza and Marqt are more expensive than traditional supermarkets and prize themselves for having a wide assortment of fresh, organic, and vegan products. They target a young and rich customer segment that is very environmentally conscious.

Since the last decade, traditional stores have come under pressure from online store formats but the Dutch supermarket industry is still very strong. According to data obtained from Statista on the industry, it is evident that the supermarket industry is still growing. Figure 3.2 shows the total revenue generated in the market, which has increased from 33.6 billion euros in 2015, to 46.9 billion euros in 2022, which relates to 39.5% growth in just seven years. The growth in the market is further proven by Figure 3.3, showing the number of stores has increased with nearly 800 supermarkets in 8 years.

Although the market is strong, competition is even stronger as the industry is dominated by some large corporations. Albert Heijn is the market leader with over 1100 operating stores (ranging from large AH XL superstores to small AH to-go shops) and achieved a revenue of 12.72 billion euros as of 2022. Following Albert Heijn is the Jumbo with nearly 700 stores and a revenue of 7.38 billion euros. Behind these two market giants of Dutch origin come a number of other retail chains such as Lidl, Plus, and Aldi who round the top 5. These 5 retail chains roughly generate 60% of the total revenue from the industry and own nearly half of the total number of stores. Besides these industry titans, there are a number of smaller retail chains who are mostly focused on specific regions such as Jan Linders and Coop (Statista, 2022h), (Statista, 2022c). Figure 3.1 shows the combined revenue of the top 5 grocery retailers against the total market revenue in 2022 and Figure 3.2 shows the combined revenue from supermarkets in the range 2015 to 2022.

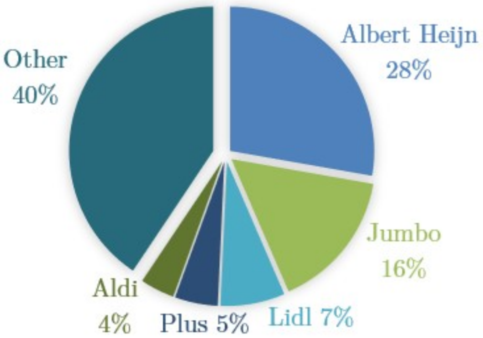

Figure 3.1: Percentage of total revenue of the top 5 supermarkets in the market in The Netherlands in 2022, adapted from Statista (2022c).

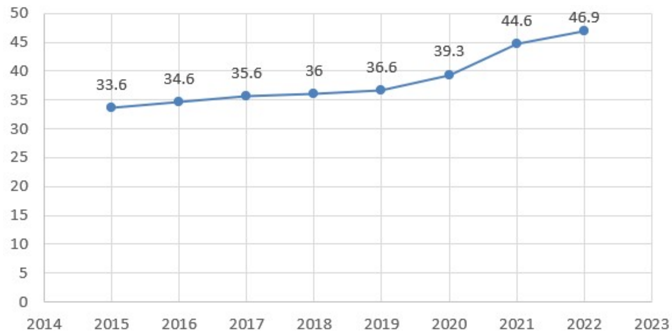

Figure 3.2: Total market revenue of supermarkets in the Netherlands from 2015 to 2022, adapted from Statista (2022c).

Figure 3.3: Number of operating stores per retail chain compared to the total number of supermarkets from 2014 - 2022. Adapted from Statista (2022f).

3.1.1 Porter's five forces analysis

Porter (1979) developed a method for analyzing the market structure of a business or industry. He argues how competition for profits in an industry is not solely dependent on the direct competitors that are present, but rather is a combination of factors that are categorized into five competitive forces. The five categories are: Bargaining power of suppliers, bargaining power of buyers, threats of potential entrants, threats of substitute products, and finally, rivalry among existing competitors (Porter, 1979). The competition that results from all five forces defines the structure of the industry and the intensity of competition. This provides valuable information for managers and analysts as it determines the attractiveness of an industry in terms of profitability and if it is wise to stay or enter in the market. Although industries can differ significantly in terms of structure and appearance, Porter (1979) argues that the underlying factors that determine profitability are the same. Figure 3.4 shows a graphical representation of Porter's five forces model with each of the 5 main forces and possible sub-factors for each of the forces.

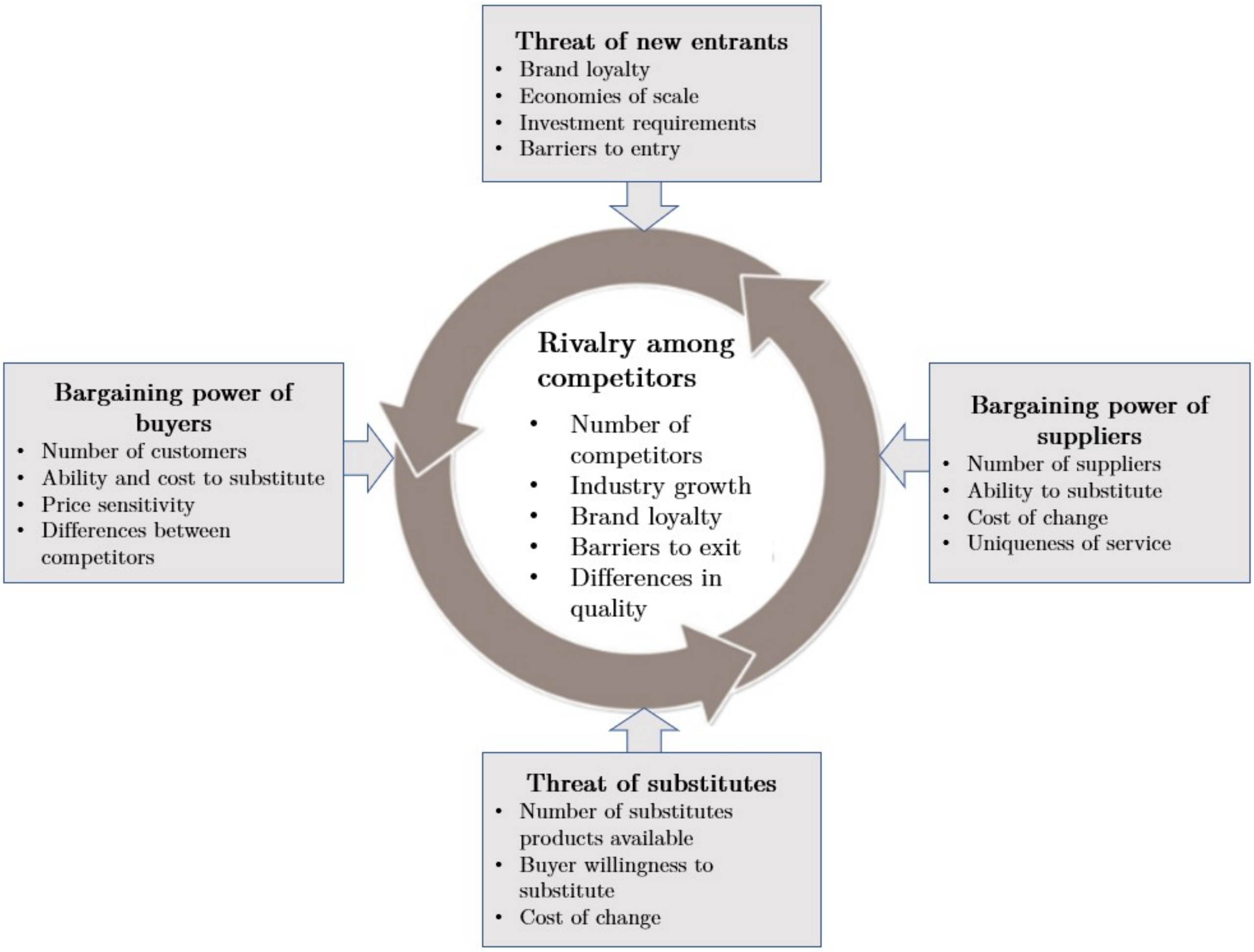

Figure 3.4: Porter (1979) five forces model. Copied from Business-to-you (2016)

Following the general concept of Porter's five forces model, this research applies the model to the Dutch supermarket industry. The model was constructed based on observations by the researcher, combined with internet research into the market using sources such as Statista, Cbs and other market analysis reports about the industry. Table 3.1 provides an overview of each of the five competitive forces for the industry.

Bargaining power of suppliers	Power level
Costs of suppliers to switch to competitors	high
Easiness of switching to competitors for suppliers	low
Number of alternative suppliers available	high
Bargaining power of buyers	**Power level**
Easiness and costs for customers to switch to competitors	low
Availability of alternatives	high
Consumer demand for product standards	medium
Necessity of products to consumers	high
Threats of new entrants	**Threat level**
Loyalty of customers to market incumbents	medium
Investment requirements to enter the market	medium
Specific technologies required for market entrance	low
Effect of economies of scale and distribution network	high
Threats of substitution	**Threat level**
Substitution through environmental and sustainable consumer alternatives	medium
Digital substitution through home delivery	high
Substitutability of products	low
Competition among existing firms	**Competition level**
Number of competitors in the market	high
Customer loyalty towards retailer brand	medium
Existence of price-cutting races	high
Average growth of the industry in recent years	medium

Table 3.1: Influence factors of Porters' five forces model in the Dutch supermarket industry.

Competition among existing firms

In general, the Dutch grocery store industry is a highly competitive market that is dominated by multiple large retail chains as is evident from figure 3.3 and figure 3.1. Retail chains focus on driving down prices and often advertise by offering the best quality for the lowest price. Some retailers such as Lidl and Aldi are more focused on 'Everyday Low Pricing' (EDLP) strategies, whereas Albert Heijn offers weekly promotions and discounts on various products using a 'High-Low' pricing (HL) strategy. Supermarkets often try to undercut each other with these strategies and retain customers with loyalty programs.

Bargaining power of suppliers

The bargaining power of suppliers is relatively low in the industry as there exist many national and international suppliers that are ready to cater to the needs of the retailers, this means supermarkets are in a strong position to negotiate deals that favor them. Stichele and Young (2009) shows how large retail corporations are in a position to abuse their power when it comes to purchasing practices. This is especially true for food suppliers at the bottom of the supply chain such as exotic product farmers in developing countries. The fierce competition and growing dominance of a number of retail chains in the Dutch food sector has a negative effect on these food suppliers as they can enforce strict contracts with negative implications for farmers and workers worldwide (Oxfam, 2018).

Bargaining power of buyers

The bargaining power of consumers is relatively high in the industry as the easiness of switching to competitors or substitute products is very low. Especially in large cities, the number of available stores of different retailers is very high so consumers can easily choose to patronize a different store if they have a bad experience. This is harder for consumers in rural areas where only a few options are available. Consumers also develop an increasing preference for organic and sustainable products. According to Research (2021), spending on sustainable products in the food sector has increased with more than 7% in 2020, compared to the previous year. Supermarkets need to adapt their strategy to fulfill these customer demands or they risk losing them to competitors or substitutes that are willing to satisfy the needs of the customers, which means that customers have a high demand power over supermarkets. Although the previously mentioned arguments suggest buyers are in an excellent position to make demands, the products that are offered by supermarkets are critical to the consumers' existence. This negates some of the buyer power and increases the power of the firms.

Threats of new entrants

The threat level of new market entrants is determined by the barriers new companies face when they wish to enter the market. These include physical barriers that represent monetary investments or sunk costs that new market entrants face. Here, economies of scale play a large role in the restriction of opportunities for entry. As competition is high and firms drive down the prices to gain a competitive advantage, the firms require large output to operate at minimum efficient scale. Non-physical entry barriers refer to building up and establishing a brand name and raising customer awareness (CPB, 2008). Both these barriers are relatively high as market incumbents have access to vast distribution networks that help them achieve economies of scale to help them remain competitive. They also benefit from established brand names and try to retain their customers through loyalty programs such as bonus cards and customer-specific promotions.

Threats of substitution

Substitution in this industry can be split into two categories, namely, the substitution of products and the substitution of grocery store format. The threat of the former is relatively low as a direct replacement for food is unthinkable, but the type of food definitely is substitutable. As pointed out by Research (2021) in recent years consumers have shifted towards demanding more organic and sustainable product options, thus as alternative options become available, the threat of substitution rises. With respect to substitution of the supermarket format itself, this threat is slightly higher. In the last decade, online grocery delivery formats emerged to challenge the traditional brick-and-mortar store format. Data obtained from Statista shows how the market penetration of online delivery formats has increased to 16.4% in 2022 shown in figure 3.5 (Statista, 2023), and new market entrants solely focused on online delivery such as "Picnic" have established themselves in the Dutch online grocery industry, almost quadrupling their market share in just 6 years as shown in figure 3.6 (Statista, 2022g). According to Company (2020) the grocery retail industry is under pressure from innovations such as online delivery which promise alternatives for supermarkets. This means the threat of substitutions for supermarket store formats is significantly higher.

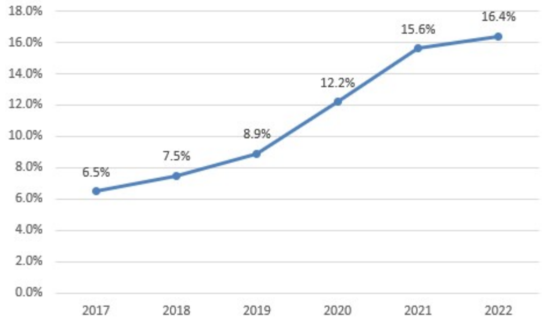

Figure 3.5: Penetration rate of online grocery delivery market in the Netherlands, copied from Statista (2023).

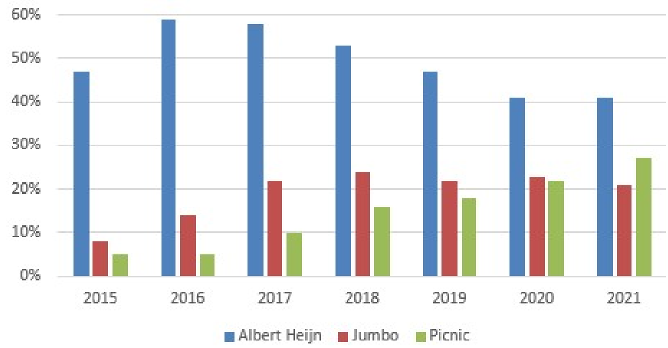

Figure 3.6: Retailers with the highest market share for online grocery shopping, copied from Statista (2022g).

3.1.2 Spatial distance between supermarkets

This research considers discounters and premium supermarkets as complementary factors although the relationship between the two is difficult to define. Discounters such as Aldi and Lidl are known for having cheap product lines whereas traditional supermarkets such as Albert Heijn and Jumbo are known for having more premium product brands. Consumer research such as Consumentenbond (2022) compares the average prices of supermarkets in The Netherlands and finds indeed large differences between premium brands and budget alternatives. However, most traditional supermarkets also have their own budget lines which are very price competitive with discounter stores such as Aldi and Lidl. Still, discounters and premium supermarkets often opt to locate in each other's vicinity which raises the belief that there exists a certain symbiotic relationship. For example, this research analyzed the average minimum distances between retail chains in Eindhoven for the years 2013 and 2022.

Figures 3.7 and 3.8 show for each Aldi store, the distance to the nearest Albert Heijn and Jumbo store. These figures show that for every existing Aldi, there is a store of one of the premium retailers located within an average distance of 100 meters in 2013 and 300 meters in 2022. This observation shows the same for the analysis of Lidl stores in figures 3.9 and 3.10. The average minimum distance to the nearest premium retailer for each Lidl store was 470 meters in 2013 and just 260 meters in 2022. Observing these retailers are always closely located in different time periods does indeed give reason to believe a certain positive relationship exists between these formats as it is assumed stores that are trying to locate close to each other receive some form of benefit. Also, Albert Heijn and Jumbo are analyzed together to gain insight into the distance between the stores of both retailers. Table 3.2 shows for both retail chains, the average distance to the nearest store of its own brand as well as its competitor. The results show this distance is always approximately a kilometer and so it suggests that stores that do not have any complimentary effect are located further from each other.

	2013		2022	
	Albert Heijn	**Jumbo**	**Albert Heijn**	**Jumbo**
Albert Heijn	1.22	1.24	1.12	0.97
Jumbo	0.96	2.48	0.90	1.52

Table 3.2: Average minimum distance in kilometers between stores for the Albert Heijn and Jumbo in 2013 and 2022

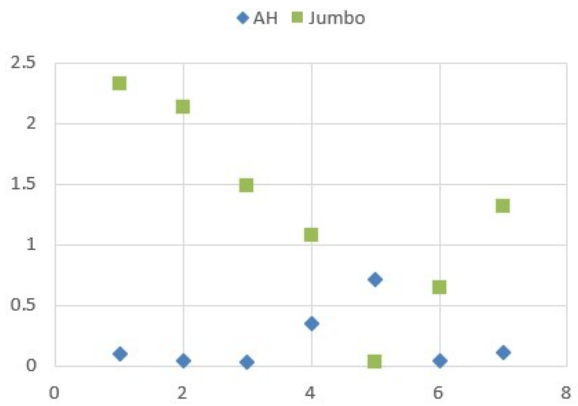

Figure 3.7: Minimum distance in kilometers to nearest premium supermarket per Aldi store in 2013.

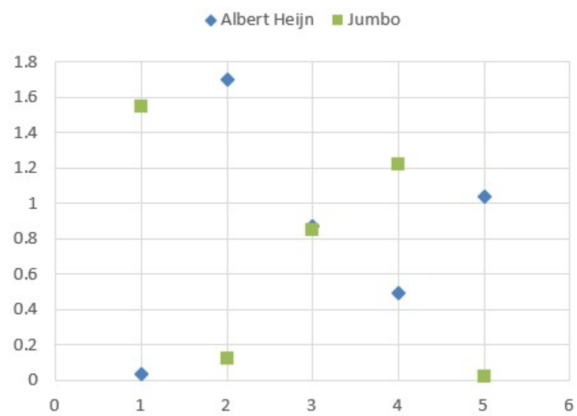

Figure 3.8: Minimum distance in kilometers to nearest premium supermarket per Aldi store in 2022.

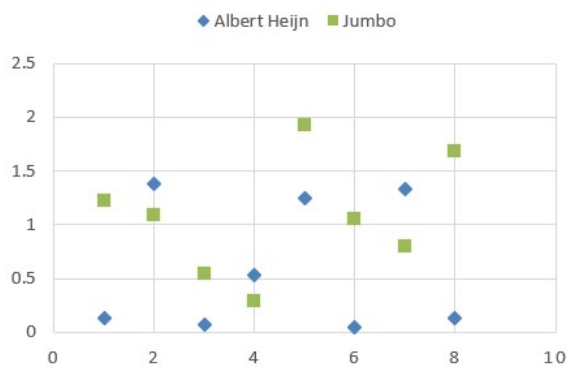

Figure 3.9: Minimum distance in kilometers to nearest premium supermarket per Lidl store in 2013.

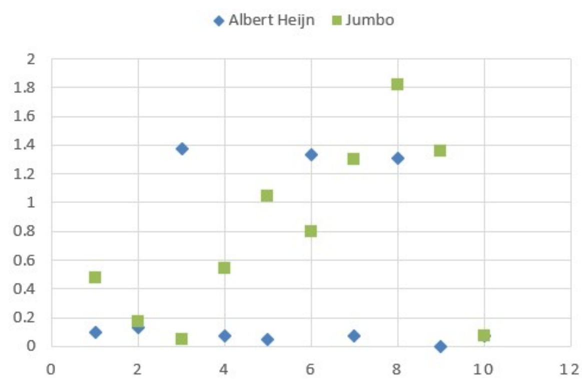

Figure 3.10: Minimum distance in kilometers to nearest premium supermarket per Lidl store in 2022.

3.2 Proposed Models

This section describes the different models that are created for the research in order to answer the research questions formulated in section 1.1.2. The goal of these research questions is to determine customer store choice parameters and how they change over time, per geographical location and for different retailers. This thesis develops different models by modifying the utility function through adding, removing, or changing the parameters in the function. These models are then applied to real data sets described in chapter 5 to find which utility function best describes the observed equilibrium in practice.

3.2.1 Construction of the Utility functions

Section 2.3.1 discusses the basic concept of the utility function and this research assumes rational customers base their store choice on the one that yields the highest utility. The utility function defined by Crönert et al. (2022) formulated in equation 3.1, forms the basis on which the other utility functions in this thesis are defined using additional parameters. Assume a network of customer locations J and shop locations K in observation o. Then the utility obtained by customers in location j, patronizing facility k from retail chain i is defined as u_{ijk}.

$$u^o_{ijk} = \alpha_0\beta_i + \alpha_1\tilde{d}^o_{jk} + \alpha_2\tilde{g}^o_k \tag{3.1}$$

With $\beta_i, \tilde{d}^o_{jk}, \tilde{g}^o_k \in [0,1]$ and $\alpha_0, \alpha_1, \alpha_2 \in (0,1]$.

Equation 3.1 is an additive utility function based on three weighted parameters. β_i represents the aggregate measure of retail chain brand attractiveness. This includes attributes that are difficult to quantify such as product range, price, and quality but are assumed to be equal across all stores k of the same retailer i. $\tilde{g}^o_k$ represents a convenience factor of store k in observation o, which captures synergies with other points of interest in close proximity to the store (e.g., public transport facilities or complementary stores). Finally, $\tilde{d}^o_{jk}$ represents the normalized distance for customers in location j to store k. Let d^o_{jk} be the distance between customer j and store k and let the maximum distance a customer is willing to travel be $\bar{d}$, then $\tilde{d}^o_{jk} = \frac{\bar{d}-d^o_{jk}}{\bar{d}}$. This means that if the distance of a customer location j to a store k becomes smaller, (the store is more accessible to this customer), the closer $\tilde{d}^o_{jk}$ is to 1. The normalization of d^o_{jk} is necessary as it is intuitive to assume an increasing distance should have a decreasing effect on the utility which is achieved through this formulation. Furthermore, the parameter weights α_0, α_1, and α_2 describe the relative importance between the three parameters. Assuming the weights are strictly positive should be a valid assumption as it seems logical that a decreasing accessibility or convenience should not have a positive effect on the utility of the customers.

Furthermore, the approach by Crönert et al. (2022) divides all u^o_{ijk} by α_0 and constrains $\frac{\alpha_1}{\alpha_0}$ and $\frac{\alpha_2}{\alpha_0}$ to sum up to 1. This sets a limit to which extent consumers can value the brand of a chain compared to locational factors. In doing so auxiliary variable α replaces α_1 and α_2, since $\frac{\alpha_1}{\alpha_0} + \frac{\alpha_2}{\alpha_0} = 1$

$$\alpha = \frac{\alpha_1}{\alpha_1 + \alpha_2}$$
$$\alpha = 1 - \frac{\alpha_2}{\alpha_1 + \alpha_2}$$

Resulting in the final utility function represented in equation 3.2

$$u^o_{ijk} = \beta_i + \alpha\tilde{d}^o_{jk} + (1-\alpha)\tilde{g}^o_k \tag{3.2}$$

Utility Function model 2

The utility functions used for this model are adapted from the utility function developed by Crönert et al. (2022) which serves as a base model. By adding new variables to the equation the utility functions are modified and the different utility models are tested against each other. The base utility function assumes three components, whereas the first new utility function consists of four components listed below.

- β_i : Representing the aggregate measure of retail chain brand attractiveness, which is homogeneous for all stores of said brand across the market.
- $\tilde{d}^o_{jk}$: Representing the normalized distance of customer to store.
- $\tilde{g}\_multi^o_k$: Representing the normalized convenience factor of store k with respect to multi-purpose shopping.
- $\tilde{g}\_con^o_k$: Representing the normalized convenience factor of store k with respect to accessibility.

The new additive utility function then becomes:

$$u^o_{ijk} = \alpha_0\beta_i + \alpha_1\tilde{d}^o_{jk} + \alpha_2\tilde{g}\_multi^o_k + \alpha_3\tilde{g}\_con^o_k \tag{3.3}$$

With $\beta_i, \tilde{d}^o_{jk}, \tilde{g}_multi^o_k, \tilde{g}_con^o_k \in [0,1]$ and $\alpha_0, \alpha_1, \alpha_2, \alpha_3 \in (0,1]$.

Breaking the standard convenience factor g into two components is justified as previous literature categorizes store attributes that have an impact on consumer store choice in 10 main attribute categories (Nilsson et al., 2015). These categories can then again be divided in two categories, attractiveness, and accessibility which is equivalent to convenience. Crönert et al. (2022) then assumes that all attributes related to convenience can be described by the same variable. However, one could argue that multiple factors should be included as previous literature such as Nilsson et al. (2015) shows how store attributes such as complementary stores and availability of parking spaces are categorized differently, as shown in table 2.1.

For the remainder of the utility function the same assumptions hold as are explained by Crönert et al. (2022). All factors are divided by α_0 and $\frac{\alpha_1}{\alpha_0}$, $\frac{\alpha_2}{\alpha_0}$and $\frac{\alpha_3}{\alpha_0}$ sum up to 1. Just as the utility function with three factors, the described approach with 4 factors implies a limit on how much a customer can value the brand of a chain (β_i) in comparison to store-specific properties $\tilde{d}^o_{jk}$, $\tilde{g}_multi^o_k$ and $\tilde{g}_con^o_k$. Assume a store with maximal factors $\tilde{d}^o_{jk} = 1$, $\tilde{g}_multi^o_k = 1$, $\tilde{g}_con^o_k = 1$ and with $\beta_i = 1$, the maximal relative weight of this brand for the utility of the customer is still 50%, compared with all other store properties which are jointly also valued at 50%. Resulting in utility function represented in equation 3.4

$$u^o_{ijk} = \beta_i + \alpha_1\tilde{d}^o_{jk} + \alpha_2\tilde{g}\_multi^o_k + \alpha_3\tilde{g}\_con^o_k \tag{3.4}$$

Utility Function model 3

This next utility function is an adaption of the previous utility function and another new variable is added. The utility function for the second model consists of five components listed below.

- β_i : Representing the aggregate measure of retail chain brand attractiveness, which is homogeneous for all stores of said brand across the market.

- $\tilde{d}^o_{jk}$: Representing the normalized distance of customer to store.
- $\tilde{g}\_multi^o_k$: Representing the normalized convenience factor of store k with respect to multi-purpose shopping.
- $\tilde{g}\_con1^o_k$: Representing the normalized convenience factor of store k with respect to accessibility by motor vehicles.
- $\tilde{g}\_con2^o_k$: Representing the normalized convenience factor of store k with respect to accessibility by other modes of transportation.

The new additive utility function then becomes:

$$u^o_{ijk} = \beta_i + \alpha_1 \tilde{d}^o_{jk} + \alpha_2 \tilde{g}\_multi^o_k + \alpha_3 \tilde{g}\_con1^o_k + \alpha_4 \tilde{g}\_con2^o_k \tag{3.5}$$

With $\beta_i, \tilde{d}^o_{jk}, \tilde{g}\_multi^o_k, \tilde{g}\_con1^o_k$ and $\tilde{g}\_con2^o_k \in [0,1]$ and $\alpha_1, \alpha_2, \alpha_3, \alpha_4 \in (0,1]$.

The convenience factor that relates to the accessibility of a store g_con is now broken into two components. The reason for splitting the convenience factor for accessibility is that previous research shows different store attributes for different modes of transportation. Table 2.1 shows how store attribute 9 is related to accessibility by motor vehicles only, whereas attribute 10 is related to the accessibility of other modes of transportation such as public transport.

Chapter 4

Model Requirements

This section describes the necessary steps taken to build the models described in section 2.3.1. First, the chapter starts with an overview of the required elements that are included in the model in section 4.1.1. Section 4.2 describes the process of collecting the necessary data for the elements listed in the previous section. This chapter concludes with an overview of all the necessary modeling steps in section 4.3.5. Finally, the solution approach is described with respect to the evaluation of the models. A general description is provided in section 4.4 and the in-depth evaluation of the outcome of each model and case is discussed in the corresponding case chapter.

4.1 Required model elements

This section covers all necessary data elements that are required for each model and case study in order to obtain the needed results.

Network elements The goal of the most basic model is to create a location network for customers and retail stores with other locations that have an effect on the convenience factor of the supermarkets, these locations are called Points Of Interest (POI) throughout the report. The elements required in this network are listed below.

- 1. A set of customer locations $j \in J$.
- 2. A set of supermarket locations $k \in K$.
- 3. A set of retail chains operating the supermarkets $i \in I$.
- 4. A list of Points of Interest (POIs) that affect the convenience score of the supermarket. A description of the different types of POIs that are included in the model is covered in subsection 4.1.1.
- 5. A pairwise distance matrix between each supermarket location and each customer location.
- 6. A pairwise distance matrix between each supermarket location and each POI.

Besides the required network elements there exist a few other elements that are required for the model. The forward problem in the inverse optimization model requires a profit function. According to this profit function, denoted in equation 4.1, each retailer tries to maximize their own profit, depending on their own strategy as well as the strategies employed by other retailers denoted as $\Pi_i^o\left(\mathbf{x}_i^o, \hat{\mathbf{x}}_{-i}^o\right)$. In this equation, x_{ik}^o is a binary variable representing the decision if a store k is opened by retailer i in observation o. f_{ij}^o represents the fraction of customers in location j that patronize retailer i. p_j^o is the total population in customer location j. m_{ij}^o is represented by Crönert et al. (2022) as the cumulative contribution margin retailer i receives per customer in location j. It is important to note that this name can be confusing since margins are always represented as percentages, which is not the case for this factor. Section 4.3.3 presents a full analysis of the cumulative contribution margin m developed for this model. Finally, c_{ik}^o refers to the annualized costs required for operating store k.

$$\Pi_i^o\left(\mathbf{x}_i^o, \hat{\mathbf{x}}_{-i}^o\right) = \underbrace{\sum_{j \in J} f_{ij}^o m_{ij}^o p_j^o}_{\text{annualized operating margin}} - \underbrace{\sum_{k \in K} x_{ik}^o c_{ik}^o}_{\text{annualized fixed costs}} \tag{4.1}$$

In order to formulate this profit function, additional data is required for the model which is listed below. The continuous decision variable f_{ij}^o is obtained through a Huff-like gravity model (Huff, 1964), as shown in equation 4.2. More general information on Huff gravity models is explained in section 2.2.

$$f_{ij}^o = \frac{\sum_{k \in K | d_{jk} \leq \bar{d}} x_{ik} u_{ijk}^o}{\sum_{k \in K | d_{jk} \leq \bar{d}} \sum_{\tilde{i} \in I} x_{\tilde{i}k} u_{\tilde{i}\tilde{j}k}^o} \tag{4.2}$$

The additional required modeling parameters for which data needs to be gathered are listed below.

- 1. The cumulative contribution margin per customer for each customer location.
- 2. The total population for each customer location.
- 3. The total yearly costs for operating a supermarket at a given location.

4.1.1 Points of Interest

In this section, the POIs that are expected to influence the convenience score of supermarkets are discussed. A list is generated of other shops that have a complementary effect on supermarkets (e.g. pharmacies, liquor stores), and other locations that otherwise have a positive impact on the accessibility of the store such as public transportation stations and parking areas.

The following items, when in the vicinity of grocery retailers, are expected to positively affect the convenience score of a supermarket as they improve the accessibility of supermarkets for different transportation modes.

- Gas stations: Improve accessibility of motor vehicles
- Parking lots: Improve accessibility of motor vehicles
- Bus stop: Improve accessibility of public transport
- Tram stop: Improve accessibility of public transport
- Bicycle parking spaces: Improve accessibility of bicycle/walking mode

The following POIs are also expected to positively influence the convenience factor when located near supermarkets, as this table includes stores that have complementary effects on supermarkets to improve multi-purpose shopping.

Pharmacies	Drug Stores	Cosmetics stores
Grocery Discounters	Organic Supermarkets	Health food shops
Liquor Stores	Wine Stores	Cheese shops
Bakeries	Butcher Shops	Pastry shops
Seafood shops	Spice shops	Pet specialty shops
Book Stores	Hardware stores	Gift stores
Florists	Tobacco shops	Electronic stores

Table 4.1: Table of all Points of Interest that are expected to influence the convenience factor of supermarkets related to multi-purpose shopping.

4.2 Data Collection

This section covers all the methods that were used for collecting the necessary data for the required model elements described in the previous section.

Supermarket Locations
To obtain supermarket location data in The Netherlands, an independent industry news organization named 'Distrifood' (https://www.distrifood.nl/), has agreed to help with this research and Distrifood (2022) from Distrifood provided a partially complete dataset with supermarket location info in The Netherlands. The necessary location info consists of the following elements: "Store_ID", "Retail_chain", "City", "Province", "Street", "Zip_Code", "Latitude_Coordinate" and a "Longitude_Coordinate".

The dataset in excel contained this information for all supermarkets in The Netherlands over a time span of 10 years, ranging from 2013 until 2022. Distrifood began only to store the Latitude and Longitude coordinates from supermarkets in 2019 and thus a solution was required to obtain the coordinates for each supermarket in the years before 2019. Since the original dataset is in excel format, a method was developed in excel to solve this problem. The formula in 4.1 compares the values of the same cell in different sheets. Applying this method shows which data entries are different. The sheets are then cleaned so that every entry corresponds to the same store and then Latitude and Longitude coordinates can be copied to obtain the coordinates for each year.

```
=IF('2017'!A63 <> '2018'!A63, "2017:"&'2018'!A63&" vs 2018:"&'2017'!A63, "")
```

Figure 4.1: Formula for excel operations, comparing the data entries of different sheets.

Customer Locations
The customer locations should be represented as nodes or points in the network preferably on a small scale to represent customer locations as accurately as possible. In order to achieve this, the study uses information available on the municipality websites in The Netherlands to obtain a shapefile that describes all neighborhoods in the city as a Polygon using x- and y-coordinates. The reason for neighborhood data is that it has the highest granularity that is available. From each polygon, the center of the polygon was extracted as a single point which represents the customer node in the network graph.

Population and consumer margin
Just as with the data for customer locations, the municipality websites offer detailed information on demographic data such as population, age, and income. For each neighborhood, the total population

for each year is extracted, as well as income and household composition which are necessary to obtain the customer margin parameter which is calculated in section 4.3.5.

Data collection Points of Interest
To determine the locations of the POIs listed in table 4.1, this research uses OpenStreetMap (https://www.openstreetmap.org). OpenStreetMap (OSM), is a free open geographic database. Just like Wikipedia, it is maintained and updated by members of the community. It serves as a good alternative to Google Maps which charges users if they want to scrape data. In order to obtain a list of the required POIs, a python script is built which makes use of the "Overpass Turbo" data mining tool designed for OSM (https://overpass-turbo.eu/). By including the Overpass Turbo API in the python script, a search query can be constructed to retrieve the data from the OSM website and store it in a dataframe for further analysis.

Another explored method is by building the query directly in the Overpass Turbo tool and exporting the results to a GeoJSON file which can then be uploaded in python for further analysis. Appendix D shows the OSM query which was used to obtain all elements related to parking spaces in Eindhoven using the Overpass Turbo API.

4.3 Modeling Steps

This section describes the necessary steps to build the model from the required data elements. The first section describes the process of creating both the customer and store locations. Section 4.3.2 describes the process of constructing the distance matrices for both the shop locations - customer locations, as well as the shop locations - POI locations. Section 4.3.5 describes how the value for the convenience scores is determined.

4.3.1 Consumer- and Store Locations

The first step in the model is creating a network by representing customer- and store locations on a geographical map. The network is modeled using the Folium library package in the open-source programming language Python. Folium allows the visualization of data on an interactive geographic leaflet map. Representation of the networks in the case studies is further described in section 5. In the network, the customer- and store locations are represented as single-point nodes using a latitude and longitude coordinate set and contain node-specific information.

Creation of 'Candidate' store locations
Besides existing store locations, empty candidate store locations are created. This is necessary for the forward problem part of the algorithm. The forward problem uses simultaneous location selection for all retailers to determine the optimal set of locations that maximizes the profit for each retailer. In order to achieve this, many candidate locations, which are alternatives to the existing store locations need to be created. This leads to the question of what is the best method for creating these candidate store locations that would best represent real-life scenarios. In order to answer this question, an interview with the department of economic affairs of the municipality of Eindhoven was conducted. A transcription of the interview is provided in Appendix E. This interview was conducted to gain more insight into the process where retailers open new supermarkets and how they coordinate their decisions with the municipality. The conclusion of this interview is that retailers have the option to open a new supermarket almost anywhere they want if they believe that this candidate location has the highest economic upside. On another note, the urban planning department does maintain close contact with retailers on where they plan to build small or large shopping areas so supermarkets are encouraged to factor this into their decisions. But ultimately the decision is up to the retailers themselves. Another

result from this interview was that the municipality of Eindhoven remains in close contact on this topic with other major cities in the Netherlands, and the decision-making process is roughly the same in other cities. For this reason, it is assumed the decision-making process in determining candidate store locations is the same for every case presented in chapter 5.

To create 'empty' candidate locations the shapefile of the neighborhoods of the corresponding city is used to concatenate the polygon of every neighborhood into one large polygon that outlines the entire city. Then, a grid is spanned across the polygon of the city using latitude and longitude coordinates with steps of 500 meters. Because there exists a trade-off in model speed vs accuracy, the distance of 500 meters is chosen as it is assumed this is the highest granularity possible for the model. Since retailers can choose to build a new store at any given location, the model should present candidate store locations as close to each other as possible. But decreasing the spacing between stores means the number of stores K and the solution space greatly increases, which has a negative effect on the model runtime.

The polygon of the city outline serves as the grid boundary and each grid point represents a candidate location. This means the end result is a square-form grid with a starting node in the most South-West corner and a set of nodes each 500 meters North of each other and a set of nodes 500 meters East of each other. A grid visualization can be found for each case in section 5.

4.3.2 Pairwise Distance Matrix

The next step is creating a pairwise distance matrix between every customer location and every (candidate) store location to obtain the distance d_{jk} between customer j and store k. According to Shih (2015), a good method for calculating the distance between two geographical coordinates is using the Great Circle Distance (GCD) method. The GCD is an accurate method of computing the distance between two points on a spherical surface using latitude and longitude coordinates. As the earth is a sphere this method is applied in this model to construct the distance matrix. The GCD uses the haversine formula shown in equation 4.3, which is a re-formulation of the spherical law of cosines but is better applicable in cases with small angles and distances.

$$d_{jk} = 2 \cdot R \cdot \arcsin\left(\sqrt{\sin^2\left(\frac{\mathrm{lat}\,2 - \mathrm{lat}\,1}{2}\right) + \cos(\mathrm{lat}\,1) \cdot \cos(\mathrm{lat}\,2) \cdot \sin^2\left(\frac{\mathrm{lon}\,2 - \mathrm{lon}\,1}{2}\right)}\right) \tag{4.3}$$

Where d_{jk} is the distance between the two points and $R = 6371$ is the radius of the earth in kilometers. $(lat1, lon1)$ and $(lat2, lon2)$ respectively represent the coordinate couples of location 1 and location 2.

Elimination of arcs

After creating the pairwise distance matrix the maximum willingness to travel parameter $\bar{d}$ is defined which is used to eliminate all customer-store pairs for which the distance d_{jk} exceeds the value of $\bar{d}$. Determining this parameter is quite tricky as it requires aggregated data from multiple sources. Data by Access Development (2016) shows that people are less willing to travel further distances for more common purchases. They find that customers are only willing to travel 8 minutes on average to a grocery store. Combining this with data from Statista (2021) which shows the main mode of transportation for grocery trips in Amsterdam are walking (40%) and biking (30%). This indicates that customers who use walking as their main mode of transportation are only willing to travel a maximum of $\frac{8}{60} \cdot 5 = 0.667$ kilometers on average (assuming an average walking distance of 5 km/h). Customers who use biking as their main mode of transport have on average a maximum travel distance of $\frac{8}{60} \cdot 18 = 2.4km$, assuming an average speed of 18 km/h for biking. On average, the customers' willingness to travel $\bar{d}$ is then estimated at 1.5 kilometers. CBS (2010) shows that the concentration

	1-Person Household	2-Person Household	Household with Children
Total expenditure	€ 2500	€ 4500	€ 6500

Table 4.2: Total expenditures per household type on grocery shopping in Euros per year.

of supermarkets is very high in large cities as consumers in Amsterdam, Rotterdam, and The Hague on average have the choice of three supermarkets within a kilometer radius. Since this research only focuses on large cities, the estimation of $\bar{d}$ is believed to be quite solid as customers have more than three supermarket options on average.

4.3.3 Cumulative Contribution Margin

This section describes the process for obtaining the yearly cumulative contribution margin per person represented as m in the profit function 4.1. In economics, the contribution margin also known as Gross Profit Margin is defined as the selling price per unit minus the variable costs of the unit or Costs of Goods Sold (COGS), shown in equation 4.4.

$$Gross\,Profit\,Margin = \frac{Total\,Revenue - COGS}{Total\,Revenue} \tag{4.4}$$

In this research, the cumulative contribution margin means the total yearly expenditures on groceries per person, accounting for the gross profit margin of the supermarket industry. Although the name suggests otherwise, this parameter is not represented as a fractional number. Obtaining the parameter requires multiple calculation steps which are fully explained in Appendix F and summarized below.

First, the average gross profit margin in the supermarket industry in the Netherlands is determined. A thorough internet search finds this to be around 30%. Sources such as Statista and the CBS confirm this and show the gross profit margin for the supermarket industry has somewhat increased over the years, ranging between 26% and 30% (Statista, 2022b). The gross profit margin is then set to 30% as this seems a decent estimation for further calculations. The next steps involve population statistics per neighborhood obtained from municipality websites. The information required is

- Total Population
- Number of 1-Person Households
- Number of 2-Person Households
- Number of Households with children
- Average net income

Furthermore, NIBUD (National Institute for Family Finance Information) has conducted research and estimated the average yearly expenditure on groceries for each household type, shown in table 4.2 (NIBUD, 2023).

Using this data, the gross profit expenditure per person is calculated for each customer location. This research then factors in the weight of the difference in yearly income for each neighborhood, to obtain the final value of the cumulative contribution margin. This parameter differs per customer location and per year to accurately describe a real-life scenario.

	Avg. Price per m^2	Housing	Personnel	Other	Total Costs c_k^o
Amsterdam	475	452150	891000	276850	1620000
Eindhoven	200	180000	630000	315000	1125000

Table 4.3: Cost components for each case study

4.3.4 Operating Costs

This subsection describes the process for obtaining the yearly operating costs for a supermarket, represented as c_k^o in the profit function described in equation 4.1. This parameter captures all expenses related to the operation of the store such as rent, utilities, personnel costs, and equipment. Note that the costs of goods sold are already captured in section 4.4 and are not covered by c_k^o.

According to data from the central bureau of statistics, most costs are related to personnel costs and housing costs (CBS, 2023). The rest of the costs comprise many different smaller components which are together viewed as 'other costs' as can be seen in figure 4.2. This research assumes the total operating costs c_k^o are a combination of personnel, housing and other costs respectively valued at 56%, 16%, and 28%.

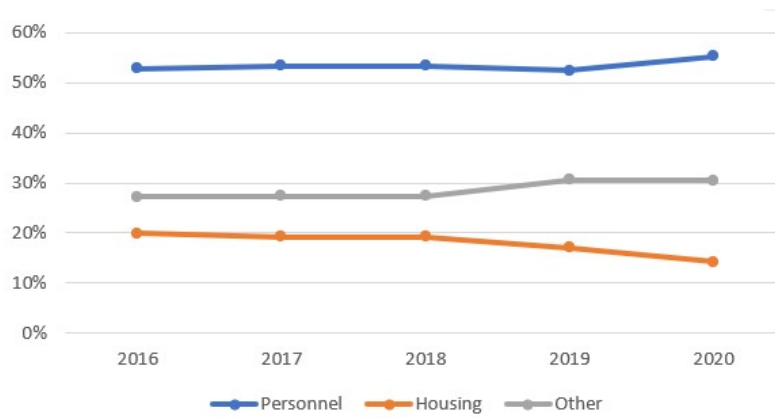

Figure 4.2: Fraction of yearly operating costs for supermarkets divided into three main components, adapted from Retail Insiders (2023) and CBS (2023)

The next step is to find a method to quantify these costs. Supermarket chains are reluctant to share information about their costs so a calculation is developed to estimate these costs, which is based on the portion of the cost allocated to housing. First, the average floorspace of grocery stores in the Netherlands is found to be around $950m^2$ in 2019. (Locatus, 2020) and (Statista, 2022a). To approximate the Housing costs, the floorspace area is then multiplied by the average retail rental price in the city per m^2 which is obtained through internet searches on commercial real estate prices and sources such as Statista (2022e). From the housing costs, the rest of the costs can be derived using figure 4.2 and the combined costs yield an approximation for c_k^o. Table 4.3 provides the cost approximations for each component and for each individual case as described in chapter 5.

4.3.5 Convenience scores

This section describes the process of obtaining the convenience parameters, represented by $\tilde{g}_k^o$ or as variation, such as $\tilde{g}_{multi_k}^o$ or $\tilde{g}_{con_k}^o$ in the different models described in section 3.2.1. Each step with an explanation is listed below.

- 1: Retrieve the coordinates for all points of interest defined in section 4.1.1 using the OpenStreetMap data.
- 2: Create a pairwise distance matrix for all supermarkets and all points of interest. For parking areas, the center point is extracted from the polygon shape and used for the matrix calculation.
- 3: All store-POI distance pairs which exceed the threshold are eliminated. It is assumed that POIs with a higher distance do not have any effect on the convenience of the store as they are too far away, especially in an urban setting.
- 4: For each store k the total convenience score is calculated using the previously described formulas for each POI. Note that the calculation of the convenience score is model specific as $\tilde{g}_k, w^o$ is broken into different components per model.
- 5: The convenience score is normalized $\tilde{g}_k, w^o = \frac{g_k, w^o}{\bar{g}}$ which is necessary to not over-emphasize the weight of this factor in the utility function with respect to the other factors.

The first step explained in section 4.1.1, is to identify possible POIs that may have a positive effect on supermarkets. The second and third steps are to determine at which distance, these POIs actually have a significant impact on the convenience score of a supermarket. When a complementary store such as a bakery is right next door, it is expected to have a more significant effect than when this store would be located more than a kilometer away. This research captures this effect in the form of a threshold factor that states that every POI within a certain radius has a positive effect, but outside of this radius, the POI is considered too far away and is not included in the convenience factor of nearby supermarkets $\tilde{g}_k^o$. The fourth step in this process is to determine the "impact" of each POI on the convenience factor. For each different type of POI, an explanation is provided below and calculations where necessary are provided in more detail in Appendix G.

Multi-purpose shopping

First, the Threshold for multi-purpose shopping stores is set to 500 meters. Since the maximum willingness to travel $\bar{d}$ is assumed to be 1.5 kilometers setting the threshold to a higher value could lead to a location-specific factor bias as the POI affects all supermarkets in a large area and the resulting convenience factors for the supermarkets only differ slightly. However, if the threshold is too low, it becomes possible that an existing effect is not taken into account which leads to the same problem of a biased convenience score. Next, it is assumed that all complementary shops related to multi-purpose shopping in the vicinity have the same impact. This means a bakery, pharmacy, or florist all have the same impact on the convenience score of supermarkets nearby. Secondly, it is assumed that when multiples of the same POI are connected to a supermarket, it does not have a linear effect on the convenience score. In other words, the effect of the first bakery has a significantly higher effect than the second, third, etc. The mathematical formulation of the convenience score calculation is explained in Appendix G.

Fuel Stations

The method described for multi-purpose shopping is also used for determining the convenience score with respect to fuel stations. Both the threshold and calculation of the convenience score are the same. But it is important to note that gas stations do not influence multi-purpose shopping but rather influence accessibility (by motor vehicles).

Parking

van der Waerden et al. (2017) show that the most important factor for consumers is 'walking distance' when it comes to parking choice. Their article also discusses the maximum willingness customers want to travel from a parking space to their destination. One of the key findings that is of interest to this research is that customers are only willing to walk a short distance when doing short trips such as weekly shopping (approximately 100 meters). More than 80% of respondents are not willing to walk more than 200 meters between their parking spot and the weekly shopping destination. For this reason, this research sets the threshold for parking spaces at 200 meters. Parking spaces with a higher distance to the supermarket are not used in calculating the convenience score.

Furthermore, it is assumed that large parking areas have a stronger effect on the convenience score than smaller parking spaces as more people can park their cars to do groceries at the same time. For this reason, the area in square meters is calculated for all parking lots. Then, the parking lots are divided into 4 categories each with its own impact values listed in table 4.4.

Area size (A) of parking lot in m^2	**Impact score**
$A < 500$	0
$500 \leq A < 1000$	0.5
$1000 \leq A < 2000$	0.75
$A \geq 2000$	1.0

Table 4.4: Impact score for parking lot area size

The parking convenience score for a supermarket is equal to the sum of all impact scores of all connected parking lots. Since OSM differentiates in parking lot areas and designated parking garages, an additional 0.5 is added for each connected parking garage.

Public Transportation
The threshold for public transportation POIs is set to 300 meters. The reason is this value is closely related to the threshold for parking by car since customers using public transportation also need to travel the last meters on foot. The threshold value was 200 meters for car users, this value is set slightly higher to 300 meters to include a decent amount of options for customers in different directions. Unlike the convenience factor calculation for multi-purpose shopping, it is assumed that the score linearly increases with the number of public transport connections. The formulation is described in Appendix G.

Bicylce parking The last POI is the parking space designated for bicycles. As Statista (2021) shows that almost 30% of customers use cycling as their preferred transportation mode when doing groceries, having enough space at a supermarket for customers to park their bikes seems a relatively important factor. The threshold for bikes is set to be 200 meters just as for car parking. The calculation of the convenience factor linearly increases with the number of connection points just as in the calculation of the public transportation score. However, the score per connection is different as described in G.

4.4 Model Evaluation

This section describes the general approach and the necessary steps to be taken to evaluate the models and to test the values for the obtained alpha and beta parameters.

We want to know if the obtained values for alpha and beta (or other relevant weights included in the utility function) make sense. We do this by testing them using the sub-problem part of the model.

First, we run the model for train dataset to obtain the values for alpha and beta and the objective value that is associated with these values (Master problem). We feed these values for alpha and beta into the forward problem for each observation and retailer in the test dataset, to obtain the objective value. The objective values of all different models are analyzed and compared to determine the model performance.

Computer properties

Home PC CPU model: Intel(R) Core(TM) i7-9700F CPU @ 3.00GHz Thread count: 8 physical cores, 8 logical processors, using up to 6 threads

Laptop

Chapter 5

Case Studies

This chapter starts with a short introduction to the case studies that were performed and a brief motivation is provided for the choice of the case studies. The other sections describe each case in more detail using the model requirements and the modeling steps discussed in chapter 4.

For this study (2-3) cities in The Netherlands were selected to serve as case studies. These form real-life situations to which the developed models are applied and obtain results. Due to the project duration, this was the maximum amount of possible case studies. But it is expected that examining (2-3) cities in-depth yields significant interesting results and is enough to provide (cautious) managerial advice. Crönert et al. (2022) describes in his article how the set of observations O refers to a geographical location in different time steps or to a set of similar geographical regions. For this reason, this thesis uses multiple different cities in The Netherlands as these are considered to be similar and do not have significantly large differences. Another reason for only researching cases located in The Netherlands is related to data collection. The researcher is a native Dutch person and is familiar with general knowledge about the country's culture and language which significantly improves the data collection process. Research in other countries would be substantially more difficult as the data collection process would require more time and energy and is more prone to errors. This research also uses a time horizon of 10 years with steps of 1 year as Distrifood (2022) could provide statistics on supermarket locations up to 10 years in the past whereas population statistics can often be retrieved from around 15 to 20 years back. For this reason, the time span of 10 years is chosen as a suitable range.

5.1 Eindhoven

The city of Eindhoven serves as the first case to be analyzed. First, a short industry analysis is performed to obtain some key insights into the market statistics for this city. Table 5.1 shows the number of supermarkets in the city per retail chain between 2013 and 2022. These supermarkets are then visualized in figures 5.1 and 5.2. From this table, it can be observed that Albert Heijn, Jumbo, Aldi and, Lidl all have a significant market share and are thus the chosen retailers for further analysis. Chapter 3.1.2 shows how premium supermarkets differ from discounter supermarkets and so the Integer Programming Game is split into two distinct cases, Albert Heijn vs. Jumbo in the premium retail analysis, and Aldi vs. Lidl in the Discounter supermarket analysis. Then the case is further split as in order to analyze Albert Heijn and Jumbo in the Integer Programming Game the dataset is split into two regions. The solution space and runtime of the model greatly increase for larger instances but a dataset with only a few initial data points would reduce accuracy. And so a balance needs to

	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Albert Heijn	16	16	16	16	16	17	18	18	18	18
Aldi	7	9	7	7	7	7	9	9	6	5
Jumbo	6	7	11	11	11	11	11	12	12	12
Lidl	8	6	8	8	8	10	10	10	10	10
Plus	3	2	2	2	2	2	2	2	2	1
Other	10	9	7	7	9	9	8	6	6	7
Total	50	49	51	51	53	56	58	57	54	53

Table 5.1: Evolution of the number of supermarkets in Eindhoven since 2013.

be found in the number of initial existing supermarket locations. From table 5.1 it can be observed the total number of initial locations k for the Aldi and Lidl lies between 15 and 19 which is a good sample size. The number of initial store locations for the Albert Heijn and Jumbo lies between 22 and 30 which is considered too much. Therefore, the city is split into two regions, North and South, each having around 15 data points. After the split, the data is gathered and the models are built using the processes described in chapter 4. Figures 5.1, 5.2, 5.3, and 5.4 represent different steps in the network creation to give an understanding of the network and how the nodes relate with each other. Next the cases are prepared by splitting the data into the three distinct cases described earlier. After this split, the data is again split into two components along the time horizon for each separate case, a training set and a testing set dataset. The training datasets consist of the years 2013 up to 2017, and the test dataset is comprised of the years 2018 up to 2022. The proposed models described in 3.2.1 are individually applied to each training dataset using the inverse optimization model developed by Crönert et al. (2022) in order to obtain the parameters that best describe the observed situation. The obtained parameters for each model are then used as input to the forward problem which is applied to the test datasets for each year and each individual retailer to compare the performance of the models and draw conclusions.

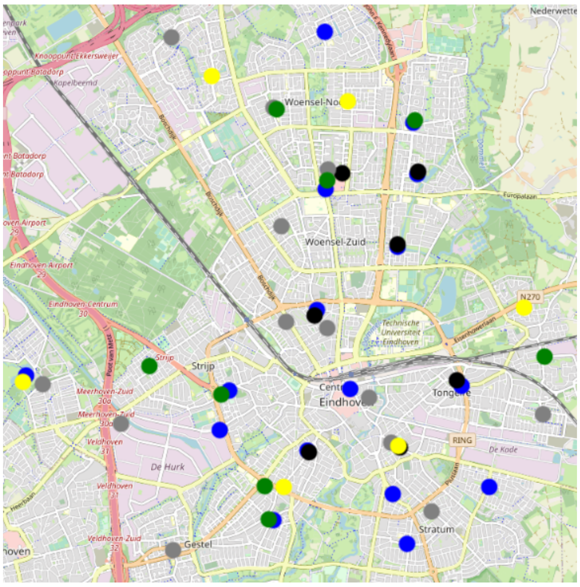

Figure 5.1: Supermarket Locations in Eindhoven in 2013 where Albert Heijn = blue, Jumbo = yellow, Aldi = black, Lidl = green, Other = grey.

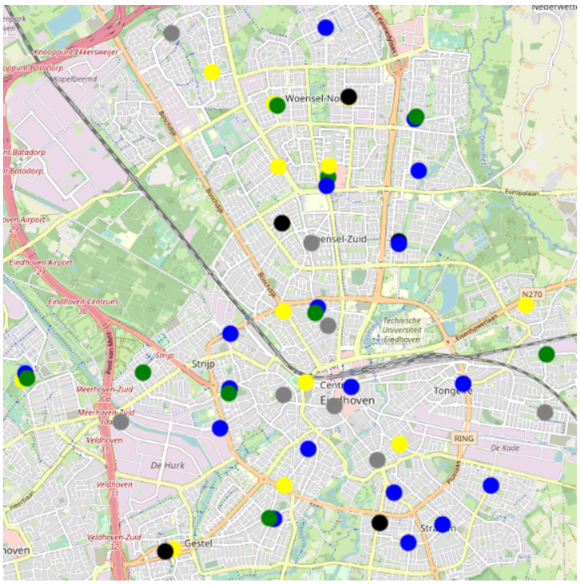

Figure 5.2: Supermarket Locations in Eindhoven in 2022 where Albert Heijn = blue, Jumbo = yellow, Aldi = black, Lidl = green, Other = grey.

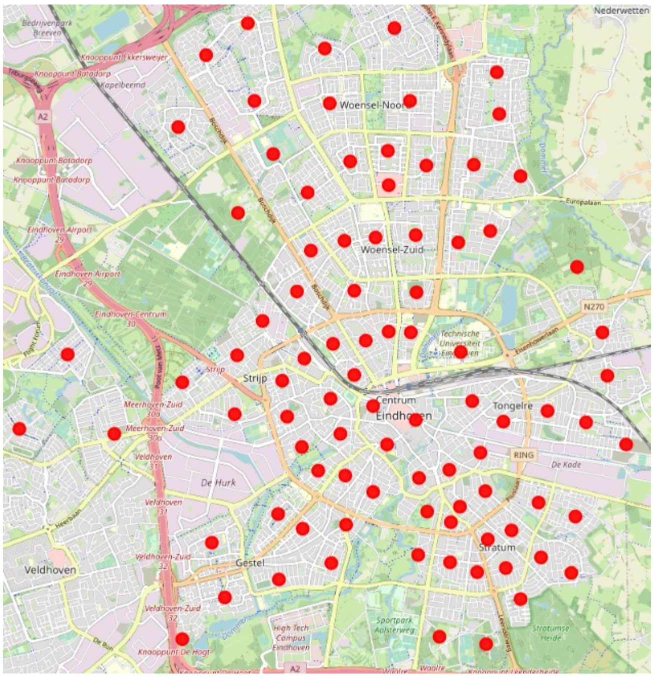

Figure 5.3: Customer locations in Eindhoven represented by each individual neighborhood as a node.

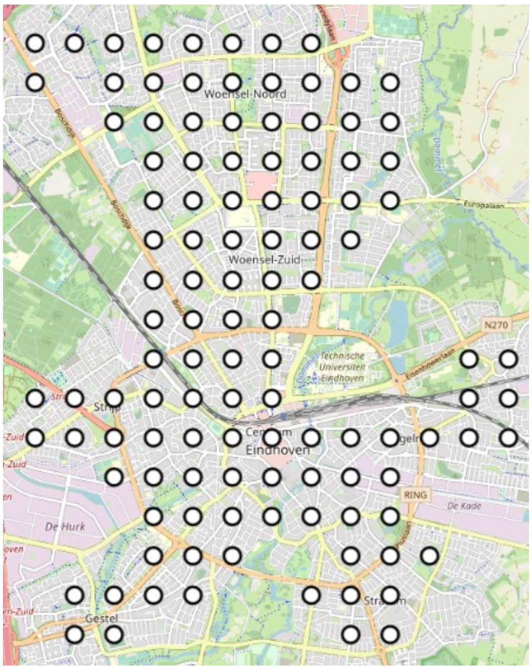

Figure 5.4: Candidate supermarket locations in Eindhoven in grid form with steps of 500 meters between nodes.

5.1.1 Results and implications

Aldi and Lidl

This section covers the results obtained by testing the different models in each scenario. Starting with the analysis of Aldi and Lidl, the results of the inverse optimization method are shown in Table 5.2. For each different model trained on the data between 2013 and 2017, it shows the parameter values that best describe the observed situation. The objective value is equal to 0 in all models, which means Gurobi was able to identify an exact optimal solution to the equilibrium in practice (i.e., $\min_{\alpha,\beta} ||\delta|| = 0$). The second step is to solve the forward problem in the test datasets for all observations $o = [2018, 2022]$ and for both retailers $i = (Aldi, Lidl)$, for each of the models with the corresponding parameters presented in table 5.2. As the forward problem is an optimization problem where the objective is to maximize profit, this research is interested in comparing the objective values for each model, and the one with the highest value indicates the superior model. The table with the objective values for the test datasets is presented in Table 5.3. This table shows model 3 as the superior model for Lidl in all observations. For Aldi, the base model (model 1), is the superior model in years 2018, 2019, and 2022, and in the years 2020 and 2021, model 2 performs the best. However, the difference in performance between models is very small as is evident from the results. For Lidl, model 3 performs on average only 2.40% better than model 2 and 4.91% better than the base model. For Aldi, the results are even more closely together as the base model performs on average only 0.07% better than model 2 and 0.27% better than model 3. In the case of Aldi, gradually changing the model by splitting the convenience factor g_k^o in $g_{multi_k}^o$, $g_{con1_k}^o$ and $g_{con2_k}^o$ decreases model performance, albeit very slightly. But for Lidl, adding these extra factors improves model performance.

Albert Heijn and Jumbo Northern District

The training results for the Jumbo and Albert Heijn for each different model trained on the data between 2013 and 2017 is shown in Table 5.4. The objective value is equal to 0 in all models, but many disparities can be observed from the table as the values for α and β vary greatly between models.

Base Model		Model 2		Model 3	
Obj. Value	0	Obj. Value	0	Obj. Value	0
α	0.57	α_1	0.41	α_1	0.39
β_1	0.16	α_2	0.14	α_2	0.13
β_2	0.14	α_3	0.45	α_3	0
Time	33755.77	β_1	0	α_4	0.48
		β_2	0	β_1	0
		Time	42065.55	β_2	0
				Time	20885.9

Table 5.2: Training simulation results for Aldi and Lidl, showing the retrieved parameter values for each different model.

		Objective Value			Difference in Objective Value					
		Model 1	Model 2	Model 3	Model 2-1		Model 3-1		Model 3-2	
2018	Aldi	83486823	83361304	83405400	-125519	-0.15%	-81423	-0.10%	44096	0.05%
	Lidl	97391906	99593929	101694986	2202023	2.26%	4303080	4.42%	2101057	2.11%
2019	Aldi	84416632	84316414	84181582	-100218	-0.12%	-235050	-0.28%	-134832	-0.16%
	Lidl	89098984	91860554	94187461	2761570	3.10%	5088477	5.71%	2326907	2.53%
2020	Aldi	85842980	85928310	85862250	85330	0.10%	19270	0.02%	-66060	-0.08%
	Lidl	90859566	93520351	96584535	2660785	2.93%	5724969	6.30%	3064184	3.28%
2021	Aldi	86818869	86863871	86466268	45002	0.05%	-352601	-0.41%	-397603	-0.46%
	Lidl	104749405	107019933	109292815	2270528	2.17%	4543410	4.34%	2272882	2.12%
2022	Aldi	88617107	88431542	88086496	-185565	-0.21%	-530611	-0.60%	-345046	-0.39%
	Lidl	112451179	114464291	116700738	2013112	1.79%	4249559	3.78%	2236447	1.95%

Table 5.3: Comparison of the objective value for all the test datasets 2018-2022 for Aldi and Lidl.

Base Model		Model 2		Model 3	
Obj. Value	0	Obj. Value	0	Obj. Value	0
α	0.64	α_1	0.86	α_1	0.27
β_1	0.26	α_2	0	α_2	0
β_2	0.18	α_3	0.14	α_3	0.1
Time	12509.83	β_1	0.97	α_4	0.63
		β_2	0	β_1	0.99
		Time	44121.82	β_2	0.22
				Time	15973.75

Table 5.4: Training simulation results for Jumbo and Albert Heijn in the Northern sector of Eindhoven, showing the retrieved parameter values for each different model.

		Objective Value			Difference in Objective Value					
		Model 1	Model 2	Model 3	Model 2-1		Model 3-1		Model 3-2	
2018	Jumbo	60258112	71386356	67200591	11128244	18.47%	6942479	11.52%	-4185765	-5.86%
	A. H.	53689521	35401998	42816158	-1.8E+07	-34.06%	-10873363	-20.25%	7414160	20.94%
2019	Jumbo	56598938	69334443	64848545	12735505	22.50%	8249607	14.58%	-4485898	-6.47%
	A. H.	54088101	35839311	43268142	-1.8E+07	-33.74%	-10819959	-20.00%	7428831	20.73%
2020	Jumbo	57315759	70170721	65675381	12854962	22.43%	8359622	14.59%	-4495340	-6.41%
	A. H.	54858822	36348478	43847376	-1.9E+07	-33.74%	-11011446	-20.07%	7498898	20.63%
2021	Jumbo	58063363	70960637	66324096	12897274	22.21%	8260733	14.23%	-4636541	-6.53%
	A. H.	55557168	36967880	44468046	-1.9E+07	-33.46%	-11089122	-19.96%	7500166	20.29%
2022	Jumbo	59269988	72269362	67826304	12999374	21.93%	8556316	14.44%	-4443058	-6.15%
	A. H.	56911518	37989912	45613931	-1.9E+07	-33.25%	-11297587	-19.85%	7624019	20.07%

Table 5.5: Comparison of the objective value for all the test datasets 2018-2022 for Jumbo and Albert Heijn in Eindhoven Northern District

Table 5.5 shows again the comparison of the objective values of the different models applied to the test datasets. From this table it is observed that for the Albert Heijn, the base model performs much better than both other models as the objective value decreases on average with 33.65% for model 2 and with 20.03% for model 3. This would suggest that splitting the convenience factor g_k^o into more components leads to drastic overfitting. However, for the Jumbo it is observed the second model is the best performer and so, through extending the base model by splitting the convenience factor into two components, the objective value of the model improves significantly with 21.51% on average.

5.1.2 Evaluation

The results obtained from the first two cases lead to some interesting observations. First, the Aldi and Lidl analysis in Table 5.2 shows that the beta parameters β_1 and β_2 are very low in all three models, which indicates customers are brand insensitive when it comes to these two supermarkets. In practice, Aldi and Lidl do not differ significantly as a brand. They are both discounter supermarkets from Germany that have roughly the same strategy concerning price and quality. This could indeed lead to customers being brand insensitive when choosing between these two retailers and valuing locational factors such as distance and convenience per store very highly. This is also evident from the values for α and α_1 which correspond to the relative importance of distance, and the values of α_2 (multi-purpose shopping), α_3, and α_4 which relate to accessibility. What is interesting to note is that the parameter relating to accessibility by motor vehicles, which is only present in the third model, has a value of 0 whereas the accessibility by other modes is valued quite highly. A reason for this might be that relatively few customers travel to supermarkets using a car in the city, which is logical since the maximum willingness to travel is assumed to be only 1.5 kilometers and there exist many store options in only a small radius of the customer CBS (2010). Indeed, Statista (2021) also shows the primary travel modes to supermarkets in Amsterdam for cars is only 18%, and so a low valuation of accessibility by motor vehicles parameter $g^{o}_{con1_k}$ seems reasonable.

The analysis of Albert Heijn and Jumbo shows certain similarities but also differences with the results obtained from Aldi and Lidl. First, the values for the β parameters vary across the different models.

5.2 Amsterdam

5.3 Evaluation of Results

Chapter 6

Conclusions, Limitations and Recommendations

This chapter concludes the thesis and presents the main findings of the research in section 6.1. A summary and answers are provided on the research questions and the research objective. Section 6.2 summarizes the limitations of the research, from research boundaries to limiting model assumptions. Finally, section 6.3 outlines recommendations and suggestions for future research.

6.1 Conclusions and Research Objective

6.2 Research Limitations

6.3 Directions for Future Researchs

Chapter 7

Bibliography

Access Development (2016). The impact of retail proximity on consumer purchases. https://cdn2.hubspot.net/hubfs/263750/Access_Consumer_Spend_Study_2016.pdf. [Online; Last accessed 03-March-2023].

Akerlof, G. A. and Yellen, J. L. (1987). Rational models of irrational behavior. *The American Economic Review*, 77(2):137–142.

Allaway, A. W., Huddleston, P., Whipple, J., and Ellinger, A. E. (2011). Customer-based brand equity, equity drivers, and customer loyalty in the supermarket industry. *Journal of Product and Brand Management*, 20(3).

Amine, A. and Cadenat, S. (2003). Efficient retailer assortment: a consumer choice evaluation perspective. *International Journal of Retail and Distribution Management*, 31(10).

Ann, P. and Koenraad, V. C. (2010). Designing a retail store environment for the mature market: A european perspective. *Journal of Interior Design*, 35(2):21–36.

Arnold, S. J., Oum, T. H., and Tigert, D. J. (1983). Determinant attributes in retail patronage: Seasonal, temporal, regional, and international comparisons. *Journal of Marketing Research*, 20(2):149–157.

Aumann, R. J. (2016). *Game Theory*, pages 1–40. Palgrave Macmillan UK, London.

Baltas, G. and Papastathopoulou, P. (2003). Shopper characteristics, product and store choice criteria: A survey in the greek grocery sector. *International Journal of Retail and Distribution Management*, 31:498–507.

Baviera-Puig, A., Buitrago-Vera, J., and Escriba-Perez, C. (2016). Geomarketing models in supermarket location strategies. *Journal of Business Economics and Management*, 17(6):1205–1221.

Bellizzi, J. A. and Bristol, T. (2004). An assessment of supermarket loyalty cards in one major us market. *Journal of Consumer Marketing*, 21(2).

Bridson, K., Evans, J., and Hickman, M. (2008). Assessing the relationship between loyalty program attributes, store satisfaction and store loyalty. *Journal of Retailing and consumer Services*, 15(5):364–374.

Briesch, R., Chintagunta, P., and Fox, E. (2009). How does assortment affect grocery store choice? *Journal of Marketing Research - J MARKET RES-CHICAGO*, 46:176–189.

Business-to-you (2016). Porter's five forces. https://www.business-to-you.com/porters-five-forces/. [Online; Last accessed 18-November-2022].

Carpenter, J. M. and Moore, M. (2006). Consumer demographics, store attributes, and retail format choice in the us grocery market. *International Journal of Retail and Distribution Management*, 34(6):434–452.

CBS (2010). Supermarket within walking distance for most dutch people. https://www.cbs.nl/en-gb/news/2010/36/supermarket-within-walking-distance-for-most-dutch-people. [Online; Last accessed 02-March-2023].

CBS (2023). https://www.cbs.nl/nl-nl/visualisaties/dashboard-economie/bedrijven#G+47+471. [Online; Last accessed 27-February-2023].

Ceppi, S., Gatti, N., Patrini, G., and Rocco, M. (2010). Local search methods for finding a nash equilibrium in two-player games. volume 2, pages 335–342.

Chan, T. C. Y., Mahmood, R., and Zhu, I. Y. (2021). Inverse optimization: Theory and applications.

Company, M. . (2020). Digital disruption at the grocery store. https://www.mckinsey.com/industries/retail/our-insights/digital-disruption-at-the-grocery-store#/. [Online; Last accessed 24-November-2022].

Consumentenbond (2022). Budgetmerken gemiddeld 50% goedkoper dan a-merken. https://www.consumentenbond.nl/nieuws/2022/budgetmerken-gemiddeld-50-goedkoper-dan-a-merken. [Online; Last accessed 25-March-2023].

CPB (2008). Static efficiency in dutch supermarket chain. https://www.cpb.nl/sites/default/files/publicaties/download/static-efficiency-dutch-supermarket-chain.pdf. [Online; Last accessed 02-December-2022].

Crönert, T., Martin, L., Minner, S., and Tang, C. S. (2022). Estimating customer attraction parameters for competitive retail location selection using inverse optimization.

Daskalakis, C., Mehta, A., and Papadimitriou, C. (2006). *A note on approximate Nash equilibria.* Springer.

Demoulin, N. T. and Zidda, P. (2009). Drivers of customers' adoption and adoption timing of a new loyalty card in the grocery retail market. *Journal of Retailing*, 85(3):391–405.

Distrifood (2022). Supermarket location dataset for the netherlands.

Eiselt, H. A. and Marianov, V. (2017). *Asymmetries in Competitive Location Models on the Line*, pages 105–128. Springer International Publishing, Cham.

Erath, A., Frank, N., Lademann, R., and Axhausen, K. W. (2007). The impact of travel time savings on shopping location choice or how far do people go to the shop cheaply. *Arbeitsberichte Verkerhrs- und Raumplanung*, 445.

Falk, T. and Julander, C.-R. (1983). Research on consumer goods distribution in sweden. *International Journal of Physical Distribution and Materials Management*, 13(5).

Fernandes, T. and Pedroso, R. (2017). The effect of self-checkout quality on customer satisfaction and repatronage in a retail context. *Service Business*, 11(1):69–92.

Guo, R. (2021). *Cross-Border Resource Management*. Elsevier.

Hansen, R. A. and Deutscher, T. (1978). Empirical-investigation of attribute importance in retail store selection. *Journal of Retailing*, 53(4):59.

Hotelling, H. (1929). Stability in competition. reprinted in spatial economic theory.

Hsu, M. K., Huang, Y., and Swanson, S. (2010). Grocery store image, travel distance, satisfaction and behavioral intentions: Evidence from a midwest college town. *International Journal of Retail & Distribution Management*, 38(2).

Huff, D. L. (1964). Defining and estimating a trading area. *Journal of Marketing*, 28(3):34–38.

Hutcheson, G. D. and Moutinho, L. (1998). Measuring preferred store satisfaction using consumer choice criteria as a mediating factor. *Journal of Marketing Management*, 14(7):705–720.

Kelly, R. F. and Stephenson, R. (1967). The semantic differential: an information source for designing retail patronage appeals. *Journal of Marketing*, 31(4):43–47.

Kerin, R. A., Jain, A., and Howard, D. J. (1992). Store shopping experience and consumer price-quality-value perceptions. *Journal of Retailing*, 68(4):376–397.

Kilroy, T., Child, P., and Naylor, J. (2015). Modern grocery and the emerging-market consumer: A complicated courtship. *Perspectives on retail and consumer goods*, 4:4–11.

Kim, H. and Choi, B. (2013). The influence of customer experience quality on customers' behavioral intentions. *Services Marketing Quarterly*, 34(4):322–338.

Kim, J. O. and Jin, B. (2001). Korean consumers' patronage of discount stores: domestic vs multinational discount store shoppers' profiles. *Journal of consumer marketing*, 18(3):236–255.

Koistinen, K. and Järvinen, R. (2009). Consumer observations on channel choices—competitive strategies in finnish grocery retailing. *Journal of Retailing and Consumer Services*, 16(4):260–270.

Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of Political Economy*, 74(2):132–157.

Locatus (2020). Als elk huishouden met één persoon boodschappen doet is er meer dan genoeg ruimte voor iedereen. `https://locatus.com/blog/als-elk-huishouden-met-een-persoon-boodschappen-doet-is-er-meer-dan-genoeg-ruimte-voor-iede` [Online; Last accessed 02-March-2023].

Marjanen, H. (1997). *Distance and store choice: with special reference to out-of-town shopping*. Turku School of Economics and Business Administration.

Mitchell, V.-W. and Harris, G. (2005). The importance of consumers' perceived risk in retail strategy. *European Journal of Marketing*, 39:821–837.

Morschett, D. D., Swoboda, B., and Foscht, T. (2005). Perception of store attributes and overall attitude towards grocery retailers: The role of shopping motives. *The International Review of Retail, Distribution and Consumer Research*, 15(4):423–447.

Moutinho, L. and Hutcheson, G. (2007). Store choice and patronage: a predictive modelling approach. *International Journal of Business Innovation and Research*, 1.

Myers, H. and Lumbers, M. (2008). Understanding older shoppers: a phenomenological investigation. *Journal of Consumer Marketing*, 25(5):294–301.

Nair, S. R. and Shams, S. R. (2020). Impact of store-attributes on food and grocery shopping behavior: insights from an emerging market context. *EuroMed Journal of Business*, 16(3):324–343.

Nash, J. (1951). Non-cooperative games. *Annals of Mathematics*, 54(2):286–295.

NIBUD (2023). Huishoudelijke uitgaven. https://www.nibud.nl/onderwerpen/uitgaven/huishoudelijke-uitgaven/. [Online; Last accessed 20-February-2023].

Nilsson, E., Gärling, T., Marell, A., and Nordvall, A.-C. (2015). Importance ratings of grocery store attributes. *International Journal of Retail and Distribution Management*, 43:63–91.

Orel, F. D. and Kara, A. (2014). Supermarket self-checkout service quality, customer satisfaction, and loyalty: Empirical evidence from an emerging market. *Journal of Retailing and Consumer services*, 21(2):118–129.

Oxfam (2018). Ripe for change: Ending human suffering in supermarket supply chains. https://policy-practice.oxfam.org/resources/ripe-for-change-ending-human-suffering-in-supermarket-supply-chains-620418/. [Online; Last accessed 22-November-2022].

Pan, Y. and Zinkhan, G. M. (2006). Determinants of retail patronage: a meta-analytical perspective. *Journal of retailing*, 82(3):229–243.

Paul, J. and Rana, J. (2012). Consumer behavior and purchase intention for organic food. *Journal of consumer Marketing*, 29(6).

Piha, S. and Räikkönen, J. (2017). When nature calls: The role of customer toilets in retail stores. *Journal of Retailing and Consumer Services*, 36:33–38.

Porter, M. E. (1979). How competitive forces shape strategy. *Harvard Business Review*, 57:137–145.

Prasad, C. J. and Aryasri, A. R. (2011). Effect of shopper attributes on retail format choice behaviour for food and grocery retailing in india. *International Journal of Retail and Distribution Management*, 39:68–86.

Reilly, W. J. (1931). *The law of retail gravitation*. WJ Reilly.

Research, W. U. . (2021). Demand for sustainable products rising; supply increasing too. https://www.wur.nl/en/newsarticle/demand-for-sustainable-products-rising-supply-increasing-too.htm. [Online; Last accessed 27-November-2022].

Retail Insiders (2023). https://www.retailinsiders.nl/branches/levensmiddelenzaken/supermarkten/. [Online; Last accessed 03-March-2023].

Reutterer, T. and Teller, C. (2009). Store format choice and shopping trip types. *International Journal of Retail and Distribution Management*, 37.

Seim, K. (2006). An empirical model of firm entry with endogenous product-type choices. *The RAND Journal of Economics*, 37(3):619–640.

Severin, V., Louviere, J. J., and Finn, A. (2001). The stability of retail shopping choices over time and across countries. *Journal of Retailing*, 77(2):185–202.

Shih, H. (2015). Facility location decisions based on driving distances on spherical surface. *American Journal of Operations Research*, 05:450–492.

Shriver, S. K. and Bollinger, B. (2022). Demand expansion and cannibalization effects from retail store entry: A structural analysis of multichannel demand. *Management Science*, 0(0).

Sinha, P. K. and Banerjee, A. (2004). Store choice behaviour in an evolving market. *International Journal of Retail and Distribution Management*, 32(10).

Statista (2021). Grocery store visits in amsterdam (the netherlands) in 2016 and 2018, by mode of transport*. `https://www.statista.com/statistics/701953/grocery-store-visits-in-amsterdam-by-mode-of-transport-and-type-of-visitor/`. [Online; Last accessed 12-March-2023].

Statista (2022a). Average shop floor area of supermarkets in the netherlands from 2005 to 2019. `https://www.statista.com/statistics/1082559/average-shop-floor-area-of-supermarkets-in-the-netherlands/`. [Online; Last accessed 02-March-2023].

Statista (2022b). Gross profit margin of supermarkets in the netherlands from 2014 to 2020. `https://www.statista.com/statistics/756139/gross-profit-margin-of-supermarkets-in-the-netherlands/`. [Online; Last accessed 25-March-2023].

Statista (2022c). Leading companies in the food industry in the netherlands in 2021, by revenue. `https://www.statista.com/statistics/629898/top-25-companies-in-the-food-industry-in-the-netherlands-by-turnover/`. [Online; Last accessed 23-November-2022].

Statista (2022d). Netherlands: Customer binding with primary supermarket, by reason 2021. `https://www.statista.com/statistics/684121/customer-binding-with-primary-supermarket-in-the-netherlands-by-reason/`. [Online; Last accessed 29-March-2022].

Statista (2022e). Prime rent for office real estate in amsterdam from 2013 to 2021. `https://www.statista.com/statistics/530076/office-real-estate-prime-rent-amsterdam-netherlands-europe/`. [Online; Last accessed 20-February-2023].

Statista (2022f). Ranking of operating supermarket chains in the netherlands in 2022, by number of stores. `https://www.statista.com/statistics/589809/ranking-of-supermarkets-in-the-netherlands-by-number-of-stores/`. [Online; Last accessed 25-November-2022].

Statista (2022g). Supermarkets with the highest market share for online grocery shopping in the netherlands from 2015 to 2021. `https://www.statista.com/statistics/659373/leading-online-supermarkets-based-on-share-of-shoppers-in-the-netherlands/`. [Online; Last accessed 24-November-2022].

Statista (2022h). Total number of supermarkets in the netherlands from 2014 to 2022. https://www.statista.com/statistics/614743/total-number-of-supermarkets-in-the-netherlands/. [Online; Last accessed 23-November-2022].

Statista (2023). Penetration rate of the online food delivery market in the netherlnds from 2017 to 2027 by segment. https://www.statista.com/forecasts/1265532/penetration-rate-segment-online-food-delivery-netherlands. [Online; Last accessed 22-November-2022].

Stichele, M. V. and Young, B. (2009). The abuse of supermarket buyer power in the eu food retail sector. https://www.somo.nl/the-abuse-of-supermarket-buyers-power-in-the-eu-food-sector/. [Online; Last accessed 23-November-2022].

Teimoory, N., Abbaspour, R., and Chehreghan, A. (2021). Reliability extracted from the history file as an intrinsic indicator for assessing the quality of openstreetmap. *Earth Science Informatics*, 14.

Theodoridis, P. K. and Chatzipanagiotou, K. C. (2009). Store image attributes and customer satisfaction across different customer profiles within the supermarket sector in greece. *European Journal of Marketing*, 43(5/6):708–734.

Uncles, M. and Hammond, K. (1995). Grocery store patronage. *The International Review of Retail, Distribution and Consumer Research*, 5(3):287–302.

Uusitalo, O. (2001). Consumer perceptions of grocery retail formats and brands. *International Journal of Retail and Distribution Management*.

van der Waerden, P., Timmermans, H., and de Bruin-Verhoeven, M. (2017). Car drivers' characteristics and the maximum walking distance between parking facility and final destination. *Journal of Transport and Land Use*, 10(1):1–11.

von Freymann, J. (2002). Grocery store pricing and its effect on initial and ongoing store choice. *Marketing Management Journal*, 12(1).

Wang, L. (2009). Cutting plane algorithms for the inverse mixed integer linear programming problem. *Operations Research Letters*, 37.

West, D. S. (1989). Evaluating a supermarket relocation strategy using spatial competition analysis. *The Annals of Regional Science*, 23:137–154.

Williams, J., Memery, J., Megicks, P., and Morrison, M. (2010). Ethics and social responsibility in australian grocery shopping. *International Journal of Retail and Distribution Management*, 38(4).

Wong, A. and Dean, A. (2009). Enhancing value for chinese shoppers: The contribution of store and customer characteristics. *Journal of Retailing and Consumer Services*, 16(2):123–134.

Woodside, A. and Trappey, R. (1992). Finding out why customers shop your store and buy your brand: Automatic cognitive processing models of primary choice. *Journal of Advertising Research*, 32:59–78.

Zhu, T. and Singh, V. (2009). Spatial competition with endogenous location choices: An application to discount retailing. *Quantitative Marketing and Economics*, 7:1–35.

Zielke, S. (2010). How price image dimensions influence shopping intentions for different store formats. *European Journal of Marketing*, 44(6).

Appendix A

Overview of Previous Research

Geographical Zone	General Retail Industry	Attributes	Grocery Retail Industry	Attributes
General	Hansen and Deutscher (1978)	1,2,3,4,5,6,8,9	Kelly and Stephenson (1967)	1,2,3,4,5,6,8,9
	Pan and Zinkhan (2006)	1,2,3,4,5,8,9		
Western Europe	Amine and Cadenat (2003)	2	Ann and Koenraad (2010)	4,5
	Myers and Lumbers (2008)	5	Demoulin and Zidda (2009)	3
			Fernandes and Pedroso (2017)	4
			Hutcheson and Moutinho (1998)	1,3,4,7,9
			Moutinho and Hutcheson (2007)	1,3,4,7,9
			Mitchell and Harris (2005)	2,3,5,7
			Morschett et al. (2005)	1,2,3,4,5
			Reutterer and Teller (2009)	1,2,3,4,5,9
			Zielke (2010)	3
Scandinavia	Falk and Julander (1983)	2,3,4,5,8,9,10	Koistinen and Järvinen (2009)	1,2,3,4,5,7,8,9,10
	Marjanen (1997)	6,9	Uusitalo (2001)	2,8,9
	Piha and Räikkönen (2017)	7	Nilsson et al. (2015)	1,2,3,4,5,6,7,8,9,10
	Severin et al. (2001)	1,2,3,4,5,8		
Southern Europe			Baltas and Papastathopoulou (2003)	1,2,3,4,5
			Theodoridis and Chatzipanagiotou (2009)	1,2,3,4,5
China			Wong and Dean (2009)	1,2,3,4,6
Korea and Japan	Kim and Jin (2001)	3,4,5,7,9,10		
India	Sinha and Banerjee (2004)	1,4,8,9	Nair and Shams (2020)	1,2,3,4,5
			Paul and Rana (2012)	1
			Prasad and Aryasri (2011)	8
Middle-East			Orel and Kara (2014)	4
North-America	Severin et al. (2001)	1,2,3,4,5,8	Briesch et al. (2009)	2,8
	Woodside and Trappey (1992)	1,2,3,4	Carpenter and Moore (2006)	2,3,4,5,6,7,9
			Allaway et al. (2011)	1,2,3,4
			Bellizzi and Bristol (2004)	3
			Kerin et al. (1992)	1,3,4
			Hsu et al. (2010)	1,2,3,5,8,9
			von Freymann (2002)	3
Australia	Bridson et al. (2008)	3	Williams et al. (2010)	1

Table A.1: Table of previous literature sorted by geographical zone where the study has been conducted and the examined retail store attributes. Adapted from Nilsson et al. (2015).

Appendix B

IO Algorithm

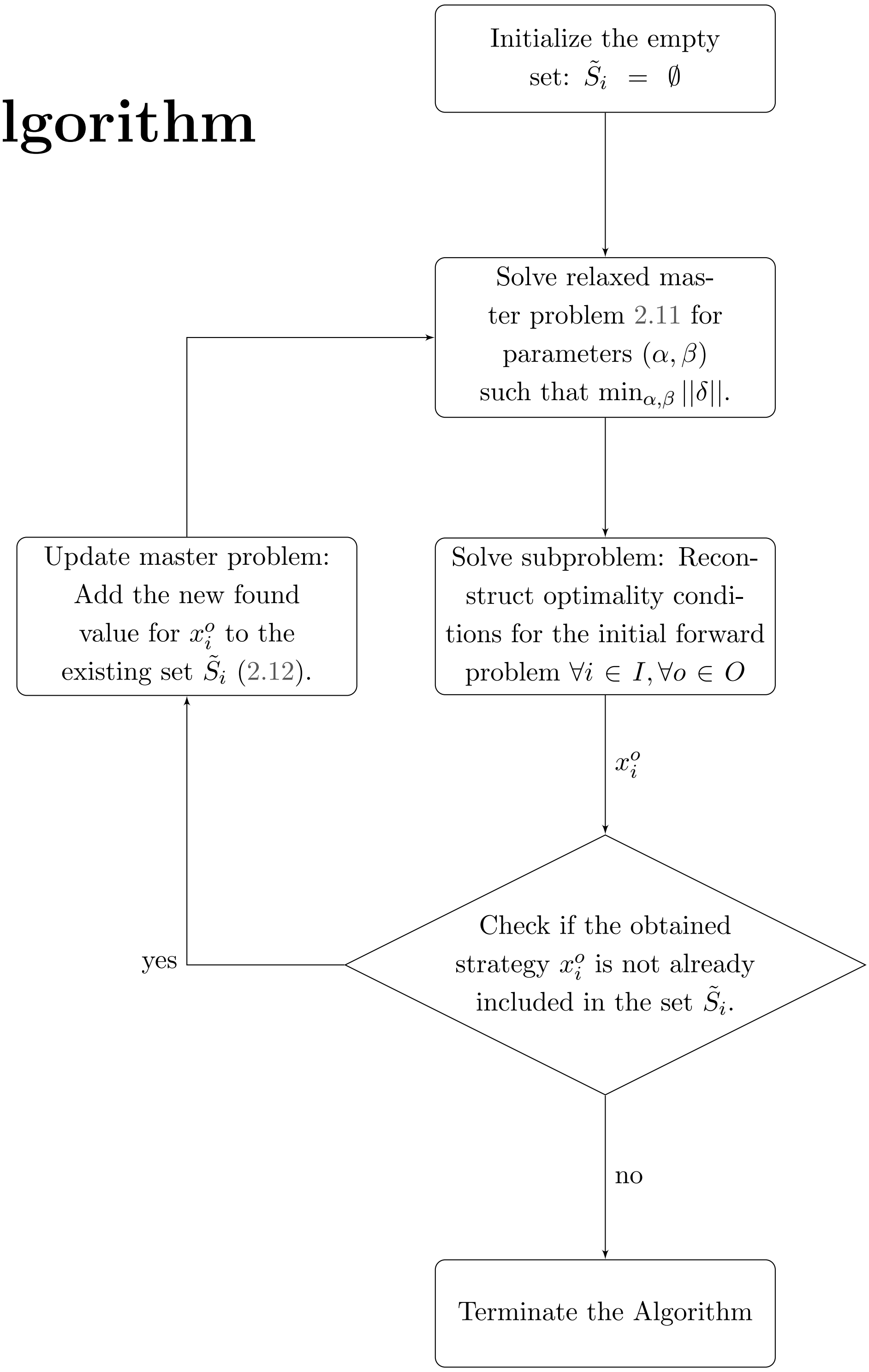

Figure B.1: Flowchart Representation of the inverse optimization Algorithm. Copied from Crönert et al. (2022).

Appendix C

Overview of Notation

Game Theory

I	Set of n players
S_i	Set of all possible strategies for player i
m_i	Number of possible strategies for player i
s_{-i}	The strategy profile of all n players except i

Sets

S_i	Set of all possible strategies for retailer i
I	Set of incumbent retail chains
J	Set of customer locations
K_i	Set of potential facility locations per retailer i
O	Set of observations

Endogenous Variables

u^o_{ijk}	Utility gained for customers $j \in J$ of visiting facility $k \in K$ of retail chain $i \in I$ in observation $o \in O$
f^o_{ij}	Fraction of customers in j patronizing retailer i
x^o_{ik}	Binary variable $x^o_{ik} \in (0,1)$ and is equal to 1 only if retailer $i \in I$ opens facility $k \in K$
α	Normalized sensitivity towards distance
$\beta = (\beta_i)_{i \in I}$	Vector of brand attractiveness for retail chains $i \in I$
$\delta := (\delta^o_i)_{i \in I, o \in O}$	Vector of unilateral improvement potentials

Appendix D

OSM Search Querry

```
<osm-script output="json" timeout="200">
 <id-query {{nominatimArea:Eindhoven}} into="area"/>
 <union>
      <query type="node">
           <has-kv k="amenity" v="parking"/>
           <area-query from="area"/>
      </query>
         <query type="node">
           <has-kv k="amenity" v="parking_entrance"/>
           <area-query from="area"/>
      </query>
            <query type="way">
           <has-kv k="amenity" v="parking"/>
           <area-query from="area"/>
      </query>
 </union>
 <union>
      <item />
      <recurse type="down"/>
 </union>
 <print mode="body" />
</osm-script>
```

Listing D.1: OSM Querry for searching parking spaces

This search query is used for the Overpass Turbo API to retrieve the locations of parking spaces in Eindhoven. For other cities, change "nominatimArea" to the specific location. And for different types of POIs, the "amenity" has to be changed to the desired POI code. These descriptions can be found at the OpenStreetMap wiki page (https://wiki.openstreetmap.org/wiki/Key:amenity).

Appendix E

Interview Transcription

Interview with Mats Frijters of the economic affairs department of the municipality of Eindhoven.

Transcription:

Appendix F

Calculation of the cumulative distribution margin

The factors required to obtain the cumulative contribution margin (CCM) parameter m for the profit function are listed in table F.1. The formula is given in Equation F.1 to calculate the cumulative contribution margin per customer in location j and observation o.

Parameter name	Parameter Symbol	Value if applicable
Gross margin factor	GMF	30%
Yearly expenditure 1-Person Households	YE_{one}	€ 2500
Yearly expenditure 2-Person Households	YE_{two}	€ 4500
Yearly expenditure Households with Children	YE_{child}	€ 6500
1-Person Households	HH_{one}	
2-Person Households	HH_{two}	
Households with children	HH_{child}	
Total population	p	
Income	I	

Table F.1: Required parameters for the calculation of the CCM

$$m_j^o = \frac{\left(HH^o_{one_j} \cdot YE^o_{one_j} + HH^o_{two_j} \cdot YE^o_{two_j} + HH^o_{child_j} \cdot YE^o_{child_j}\right) \cdot GMF}{p_j^o} \cdot \frac{I_j^o}{\frac{\sum_{j \in J} I_j^o}{J}} \tag{F.1}$$

Appendix G

Calculation of the Convenience Scores

The calculation for the convenience factor related to multi-purpose shopping for store k in o before normalization is shown in equation G.1. The set Z corresponds to the set of all POIs mentioned in table 4.1. Equation G.2 holds also for determining the convenience factor for fuel stations.

$$g^o_{multi_k} = \sum_{z \in Z} y^o_{zk} \tag{G.1}$$

where y is defined by the following discrete equation and x represents the number of "identical" connections:

$$y = \begin{cases} 0 & if \quad x = 0 \\ 1 & if \quad x = 1 \\ 1 + (0.2 \cdot (x - 1)) & if \quad x > 1 \end{cases} \tag{G.2}$$

Equation G.3 shows the calculation for the convenience score for public transportation before normalization where x represents the number of public transportation nodes. The upper-bound exists because in OSM data some stations are tagged with many different (bus)stops because of the different routes. This leads to some stores having extreme outliers.

$$g^o_{public_k} = \begin{cases} \frac{x}{2} & if \quad 0 \leq x \leq 8 \\ 4 & if \quad x > 8 \end{cases} \tag{G.3}$$

Equation G.4 shows the calculation for the convenience score for the bicycle mode where x is the number of parking areas dedicated to bicycle parking.

$$g^o_{bicylce_k} = 0.2 \cdot x \tag{G.4}$$