VILNIUS UNIVERSITY
FACULTY OF MATHEMATICS AND INFORMATICS
MODELLING AND DATA ANALYSIS
MASTER'S STUDY PROGRAMME

Modelling competitive facility location problem using Lithuanian census and grocery market data

Master's thesis

Author: Evaldas Jankauskas
VU email address: evaldas.jankauskas@mif.stud.vu.lt
Supervisor: (Dr. Algirdas Lančinskas)

Vilnius

2023

Abstract

Binary, partially binary and proportional (Huff gravity) customer choice rules are widely used in competitive facility location (CFLP) research. Due to limited availability of the necessary data, there's lack of research that validates the goodness of fit of customer choice rules to estimate market share. In this paper, a unique data set of 34 664 demand points and 1 688 facilities were collected and binary, partially binary and proportional customer choice rules were modelled using range of input parameters to estimate market share. Using parameters selected in validation step, competitive facility location model with proportional customer choice rule for existing firm was formulated as mixed binary linear programming problem and solved using commercial solver Gurobi. Results show that proportional customer choice rule is superior in estimating market share, but adjustment in parameters on a facility level is needed in order to achieve decent facility level estimates. For competitive facility location problem a model formulation for existing firm is proposed and implementation using Gurobi solver in Python with satisfactory run time is suggested. Nevertheless, to evaluate capabilities of model implementation, a comparison with other solvers using a more complex problem formulation, is needed.

Keywords: Facility location problem, customer choice rules, Huff gravity model, mixed binary linear programming problem.

1 Introduction

Competitive facility location problem (CFLP) focuses on finding optimal locations for new facilities that have to compete for their market share with other facilities belonging to other firms that are already in the market. Many variations have been considered on how to model and solve this problem. This variety is mainly driven by the multiple elements that needs to be taken into consideration when solving facility location problem. One of the elements is customer choice rules (discussed in section 2.2). It is defined by the rule that customers use when choosing a facility [24]. Many authors use one of the three main customer choice models (binary, partially binary and proportional), but very few tried to validate them. The lack of such research is justified as the necessary data is not readily available. Suhara et al. [33] claimed to be one of the first to validate proportional customer choice rule (also known as Huff gravity model) with a real data set. They used large-scale transactional data set and showed that Huff model fits well for modelling customer behavior in various categories including grocery stores, clothing stores, gas stations, and restaurants. Merino and Ramirez [26] used Huff model in combination with Monte Carlo simulation to estimate sales on a facility level and achieved satisfactory forecasting accuracy. Proportional customer choice rule seems to be the preferred customer behavior model for market share estimation. However, binary and partially binary customer choice rules are still widely used in CFLP research (see [17], [24], [28]). In this paper market share modelling is extended by validating the goodness of fit of binary, partially binary and proportional customer choice rules to estimate market share using range of input parameters (model input parameters are discussed in section 2.1). Moreover, modelling is not only performed on a company level, but also on facility level of one selected company. Results show that proportional customer choice rule is superior in estimating market share, but parameter optimization on a facility level is needed in order to achieve decent facility level estimates (for detailed discussion refer to section 4.1). In addition, both binary and partially binary customer choice rules show significant improvements in market share estimates with parameter adjustment, therefore, both rules can be considered to be valid assumptions about customer behavior.

Attraction function describes how customer's attraction towards facility is gained [29]. Distance has been usually considered as the main criterion to determine attraction between customer and facility. Eventually, many measures have been proposed in the literature, but it is generally agreed that attraction between customer and facility can be considered to be directly proportional to the quality (or attractiveness) of the facility and inversely proportional to the distance between the two ([12], [23], [15]). To define quality measures to include in the model, some researchers employ customer surveys ([3], [10]), others rely on the already available data, such as floor area of the facility ([21]) or the number of cashiers ([27]). To adjust for the unknown attractiveness factors, attractiveness correction parameter α is used. For example, Suhara et al. [33] used facility level revenue data to estimate α. Similarly, Drezner and Drezner [12] used facility level sales volume data and customers' buying power to estimate attractiveness parameter. However, facility level sales data for competitors is rarely available, therefore, in this paper a simple method to estimate chain level attractiveness, that improves market share estimates, is proposed (for method description refer to section 3.2).

Although attraction between customer and facility does not entirely depend on the distance between the two, distance is still an important component in facility location problem. Similar to attractiveness correction parameter α, there is distance correction parameter β, which is used to reflect the decline in probability of a customer to patronize a facility [5]. Default value of the distance decay is 2, which comes from the original gravity model introduced by Reilly [30]. Later, Huff [20] proposed that distance decay depends on retail category and found that for grocery stores distance decay parameter is equal to 3 [21]. To investigate the effect of distance decay parameter a number of values are used when modelling customer choice rules to estimate market share. Results show that, in contrast to Huff [20], increase in the value of distance decay parameter β with proportional customer choice rule results in higher errors in market share estimates.

Fernandez et al. [15] argue that customers do not use the exact distance, but they rather use an approximation to determine their distance to the facility, therefore, all facilities with the same characteristics that are in the same pre-defined area have the same attraction to the customer. In this paper, I try to take into account the idea of distance approximation by using different values of customer catchment distance when modelling customer choice rules. Results show that customer catchment distance is an important parameter when estimating chain level market share, but there is no clear pattern between catchment distance and market share estimate error when evaluating facility level market share.

Results of model validation are used to formulate and solve competitive facility location problem.

To do that, multiple facilities with known market shares are selected as candidate locations and mixed binary linear programming problem is formulated. Many CFLP papers focus on modelling facility location problem for new entrants and in this paper problem formulation for an existing company looking to expand is proposed. This is a slightly different problem as cannibalization of market share for existing own facilities are taken into consideration. Problem solution was implemented using commercial solver Gurobi with academic license and its Python API, and it showed satisfactory run time. Nevertheless, to truly evaluate capabilities of the model implementation, a comparison with other solvers using a more complex problem formulation, is needed.

To estimate market share using customer choice rules and to solve competitive facility location problem, a unique data set with 34 664 demand points of 1 km^2 and with 1 688 facilities was collected. Data set is unique in way of its granularity. Usually, CFLP research is based on a city level or region level data ([15], [17], [26]). Thus, it was interesting to investigate how existing optimization algorithms would perform with data of such granularity when solving competitive facility location problem.

The rest of the paper is organized in the following way. In section 2 literature review discussing main elements of competitive facility location problem is presented. Then, in section 3 data and data processing steps are described together with an experimental setting and formal problem formulation. In section 4 main results are discussed, and in section 5 conclusions and recommendations for future research are presented.

2 Literature review

Constant development of new facilities be it private (plants, shopping malls, logistic centers, bank branches etc.) or public (schools, hospitals, waste facilities etc.) raises the problem of finding optimal location. Establishing new facilities comes with high monetary costs and low tolerance for errors: once facility is established, relocation is the last option. This makes facility location problem a high priority endeavour worth investing in. This is well illustrated by the vast amount of research focused on facility location problem in various fields: vehicle routing [7], chip manufacturing [8], vehicle inspection stations [18], fire stations [32] and so on (for an extensive list refer to [9]).

Facility location problems can be classified into classical and competitive. In competitive facility location problem (CFLP), new facility competes for market share with other facilities belonging to other firms that are already in the market. Competitive facility location problem was first introduced by Hotelling in 1929 [19]. It considered two sellers each placing one facility on a line. Since then, many variations have been considered on how competitors can open their facilities. This variety is driven by the multiple elements that needs to be considered when solving facility location problem. Lančinskas et al. [24] identified five main elements: (1) a facility attraction function, (2) customer behavior rules, (3) decision variables, (4) search space, and (5) objective function. In CFLP research some elements are discussed more extensively than others. For example, search space is often discrete, meaning that there is a finite number of potential locations for new facilities.

It better reflects real world as the firm can rarely select locations from an infinite number of points. Each element of CFLP has its own specifics, which will be discussed next.

2.1 Attraction function

Attraction function describes how customer's attraction towards facility is gained [29]. Distance has been usually considered as the main criterion to determine attraction between customer and facility. Eventually, such view was adjusted to take into consideration that facilities at the same distance may not be equally attractive to the customer because they may be of different quality (e.g. floor area, parking spaces etc.). Huff [21] proposed facility floor area as a proxy for facility attractiveness. Nakanashi and Cooper [27] introduced multiplicative competitive interaction (MCI) model that used a set of facility attributes such as availability of credit card service and the number of cashiers to estimate facility attractiveness. They considered attractiveness of a facility to be a composition of facility attributes rather than single attribute alone. Such attractiveness estimation approach is sometimes implemented using customer surveys. For example, Bell et al. [3] studied attractiveness of grocery stores and over a two-year period interviewed 520 households. Drezner [10] conducted a survey of a shopping mall customers and found that the variety of stores, the mall appearance, and brand names were the most important attractiveness factors. Other authors use existing attributes to estimate facility attractiveness. McGarvey and Cavalier [25] used capacity of a facility as its measure of attractiveness. However, they acknowledged that attractiveness measured by capacity limits the area of application as it may be valid for fast-food restaurants, but not for fine-dining facilities. Although many different attractiveness measures exist, it is generally agreed that attraction between customer and facility can be considered to be directly proportional to the quality of the facility and inversely proportional to the distance between customer and facility ([12]; [23]; [15]). In the next chapter the role of distance between customers and facilities is discussed.

2.1.1 Distance

Distance between demand point and facility has multiple elements that needs to be considered when modelling competitive facility location problem. First, distance measure has to be defined. Multiple authors use Euclidean distance ([16], [1], [2], [22]), which represent the shortest distance between two points. Fernandez et al. [17] use Haversine distance, which is based on the great circle principle. It represents the shortest distance over the earth's surface, therefore, it can be considered as more appropriate for the use in CFLP than Euclidean distance. Merino and Ramirez [26] use Manhattan distance, which is used to calculate the distance between two points in a grid-like path. It can be considered to be a representation of a network type distance. Berman et al. [5] argue that network distances are more suitable for modelling CFLP as customers make their choices which facility to patronize based on time rather than by distance. In addition, customers make their decisions by the perceived distance rather than by real distance ([5], [15]. Based on this distance approximation argument, Fernandez et al. [15] proposed that distance interval could be used instead of actual distance when considering the distance between demand points and facilities.

Then, all facilities with the same characteristics that are in the same pre-defined area have the same attraction to the customer. However, using only one interval size for all facility-demand point pairs may be problematic as the catchment area of a facility in a densely populated area may be smaller than the catchment area of a facility located in an area with a low population density. All in all, as noted by Berman et al. [5], it doesn't matter which distance measure is used because if a facility is "close" or "far" by one measure, it will be close or far by all other measures as well.

Another issue in calculating distance between demand points and facilities is the representation of demand. Demand points usually represent an area and not a mathematical point. Thus, distance between customers located in a demand point and a facility may vary depending on the specific location of a customer within the demand area. To account for this, Drezner [13] proposed using distance correction based on the distance between two disks, where one disk represents a demand point of radius R and another represents a facility of radius r. Author found that the average distance calculated using disks' radii is larger than the distance between the disks' centers. Drezner and Drezner [11] considered the same issue in the subject of competitive location using Huff's gravity model and proposed distance correction formula that only takes into account the area of the demand point, but not the area of the facility. However, distance correction may only be relevant in cases where demand points differ in size as the distance correction formula is based on the radius of a demand point and if the radii of all demand points are the same, then the distance correction is the same for all demand points. In contrast to [13] and [11], Emir and Francis [14] showed that demand points can be aggregated without introducing significant errors in the distance calculations.

Distance decay is part of Huff's gravity model, and specifically the attraction function (see equation (1) in section 2.2). It represents the decline in probability of a customer patronizing a facility as a function of the distance between the customer and the facility [5]. Default value of the distance decay is 2, which comes from the original gravity model introduced by Reilly [30]. Later, Huff [20] proposed that distance decay depends on retail category. For example, he found that for grocery stores distance decay parameter is equal to 3 [21]. There are multiple methods for estimating distance decay parameter. For example, Suhara et al. [33] used particle swarm optimization technique to find optimal values of distance decay and found that the proposed method performed significantly better compared to ordinary least-squares method, which is often used to estimate parameters of the Huff model. Distance decay is not the only parameter in Huff's gravity model. There's also attractiveness adjustment parameter, which is used to adjust the attractiveness of the facility or the company. Despite different techniques proposed to estimate the parameters, few papers discuss the relationship between the two. Specifically, it would be interesting to investigate the importance of each parameter. This would help to set the priorities when parameter estimation is considered.

2.2 Customer choice rules

In order to define what market share a facility captures it is necessary to determine what part of customers' demand will be gained by each of the competing facility. In other words, customer behavior for patronizing facilities must be known. Customer behavior is defined by the rule that customers use when choosing the facility [24]. The most common customer choice rule is binary or deterministic, by which the full demand of a customer is satisfied by one facility to which it is attracted the most [29]. Customer choice rule in which customer demand is split between all facilities proportionally to the attraction it feels towards each facility is called probabilistic or proportional and was first introduced by Huff in 1964 [20]. The third choice rule which is a combination of the first two is called partially binary or multi-deterministic ([6], [15]). Based on this rule, full demand of a customer is split among all the firms, but patronizing only one facility from each firm, the one with the maximum attraction. Then the demand is allocated among those facilities in proportion to their attraction. Biesinger et al. [6] notes that research that use binary choice rule is the richest. Few recent papers shows that it is still widely used. For example, Lančinskas et al. [24] proposed a solution for asymmetric discrete competitive facility location problem using ranking of candidate locations and binary customer choice rule. Atta et al. [1] proposed new model for multi-objective uncapacitated facility location problem. Pelegrin et al. [28] proposed ties breaking rule model in competitive location problem. Literature about partially binary customer behavior is rare ([6], [15]). One of the recent papers that employs partially binary customer choice rule [6] considers a discrete competitive facility location model with minimal market share constraints and equity-based ties breaking rule. Authors argue that partially binary customer choice rule may explain customer behavior better. For example, when doing grocery shopping, customers most likely choose various supermarkets as there may be products that are only available in one firm, or there may be price differences for those products [6]. However, it is unclear why customer would only choose one facility per firm. For example, if the attraction towards facilities of the same firm is similar, one could argue that customer may still split its demand towards both of these facilities and not necessarily choose one with a slightly higher attraction. In such cases proportional customer choice rule may be more appropriate. There are also multiple papers that considered proportional customers choice rule. Kucukaydin et al. [22] employed it for competitive facility location problem with attractiveness adjustment of the follower. Benati and Hansen [4] also argued that firm cannot forecast customer behavior in a deterministic way, but rather it can be embedded by a probability distribution. They used logit model to model customer decision, which is often employed in marketing analysis. Biesinger et al. [6] considered all three customer choice rules and devised 6 scenarios with binary, partially binary and proportional choice rules all combined with essential and unessential demand in a leader-follower setting.

Although many papers employ different customer choice rules, very few tried to validate them by fitting with the real-world data [33]. This is not surprising as it requires a unique and rich data set, which is not readily available [25]. Suhara et al. [33] claimed to be one of the first to validate Huff gravity model with a real data set. They used large-scale transactional data to estimate market share and showed that Huff (proportional) model fits well for modelling customer behavior in various categories including grocery stores, clothing stores, gas stations, and restaurants. Merino and Ramirez [26] used Huff model in combination with Monte Carlo simulation to estimate sales on a facility level and achieved satisfactory forecasting accuracy. Although authors claim that such results have been achieved with little information about retail stores and consumers, rather detailed information have been used. For example, number of households, population size, household income and estimated expenditures for different product groups have been used to describe consumers. For Huff model, consumer survey data and behavioral elements, such as site-specific and retail-specific characteristics were included. Nevertheless, it is evident that researchers prefer Huff's gravity model for market share estimation. However, in facility location literature many papers use binary or partially binary customer behavior rules. Therefore, it would be interesting to understand how these customer choice models compare to each other.

2.3 Objective function

Barbati [2] distinguish between push and pull objectives in the facility location problem. An example of a push objective is the maximization of the distance between facilities and customers. Such problems are usually formulated in applications for public sector, for example, when considering location of a waste plant. Pull objective, on the other hand, is defined as the minimization of the distance between facilities and customers. In competitive facility location problem pull objectives are the most prevalent. Many papers focus on market share maximization ([15], [23], [28], [25], [4]). Market share is defined as the total demand captured by facility ([29]). Some authors introduce multi-objective functions. For example, Atta et al. [1] focused on finding facilities that (1) minimize the sum of the opening costs of the opened facilities and minimize the service cost to serve all customers, and (2) maximize sum of customer preferences. Lančinskas et al. [23] developed a general parallel algorithm for an existing firm to solve facility location problem with objectives to (1) maximize market share for new facilities, and (2) minimize the loss of market share of the existing ones. Interestingly, the proposed model is independent of the customers behavior rules, but, according to authors, its complexity may be further reduced by optimizing based on the certain customer behavior model. Some authors also include establishment costs in the objective function making it about maximizing profit instead of market share [22]. Another unique objective function in competitive facility location problem was proposed by Drezner and Drezner [12], who focused on the minimization of the probability that revenue would fall below the threshold. They found that optimal location is different to the location at which the expected revenue is maximized. Another interesting approach was proposed by Lančinskas et al. [24], who introduced a solution to competitive facility location problem with an asymmetric objective function. They argued that the real-world facility location problems usually have asymmetric objective function as facilities with different qualities such as size, parking, services etc., located at different sets of points produce different market share values. Regardless of the variable to be optimized, an important element in formulating objective function is constraint. Fernandez et al. [15] introduced minimal market share constraint when solving a discrete competitive facility location problem with equity-based ties breaking rule. Inclusion of such constraint better reflects real-world facility location problem as it is usual for a new facility to be opened only if it reaches a specific target. McGarvey and Cavalier [15] set multiple constraints. One for customer demand so that it won't exceed capacity of the facility, for specific forbidden regions, and for budget to limit total expenditures on the construction of a new facility. Although constraints may help better reflect real-world scenarios and make models more precise, there are some situations where the same results can be achieved without introducing a constraint. For example, when solving a facility location problem for opening a new facility, market share can be calculated after selecting optimal location.

2.4 Solving facility location problem

Discrete facility location problems with a discrete set of potential locations and a discrete set of demand points, are often formulated as integer or mixed integer programming problems ([24], [6], [28], [22]). ReVelle et al. [31] notes that there are also countless number of hybrid models. For example, Fernandez et al. [15] formulated their problem as a binary nonlinear programming problem and introduced equivalent formulation as a binary linear programming problem. Non-linearity was a result of including ties breaking rule, which was introduced in order to deal with instances when multiple facilities have the same level of attractiveness to the customer. Benati and Hansen [4] introduced competitive facility location model under the hypothesis that customers' behavior can be modeled by random utility functions and formulated it as hyperbolic sum integer programming problem, whose terms are ratios with a particular structure. This is a new model in combinatorial optimization and authors note that to find the optimal solution implicit enumeration or cutting plane techniques must be applied. Complex objective functions and the need for complex analysis of large amounts of data usually make solving real-world competitive facility location problems computationally expensive. In addition, it can be impossible to find optimal solutions within acceptable time frame, therefore, heuristic methods, that can help approximate the optimal solutions, are often used to solve these problems. Here is a short overview of algorithms and methods used for solving CFLP. Fernandez et al. [15] proposed heuristic algorithm for random search based on ranking of location candidates and geographical distance and showed that for small size problems proposed algorithm always obtains the same solutions as a standard optimizer (Xpress) and for more complex cases it either obtains the best solution or converges to its approximation. Lančinskas et al. [24] and Fernandez et al. [15] were able to determine the optimal solution of different instances of facility location problems using similar heuristic algorithms based on ranking of candidate locations and distance, which also outperformed genetic algorithm. Biesinger et al. [6] proposed evolutionary algorithm for competitive facility location problems with different customer behavior models. This algorithm outperformed both tabu search and memetic algorithms. They also proposed a complete solution archive, which is used to detect already visited candidate solutions and convert them efficiently into similar, but not yet considered ones. Kucukaydin et al. [22] proposed a modified version of GMIN-aBB algorithm to solve bi-level CFL problem with attractiveness adjustment of the following firm. Benati and Hansen [4] who formulated their CFL problem as a hyperbolic sum integer programming problem developed branch and bound method and proposed variable neighborhood search heuristic algorithm to solve larger instances. They noted that most of the time it is rather easy to find optimal solution and that the main difficulty is to prove its optimality.

3 Research

In this section data and data processing steps are described. Then, formal description of customer choice rules and formulation of competitive facility location problem for an existing firm is provided. Finally, experimental setting for validating customer choice rules and for solving competitive facility location problem is given.

3.1 Data

To model facility location problem, multiple data sets from different data sources were collected. For demand information, population data from Lithuanian Statistics' department was collected. It consists of 34.664 1x1 km. grids with demographic data such as population size of the grid, average age etc., and grid coordinates expressed in LKS-94 coordinates. Population data is based on census of 2021. For facilities, Lithuanian grocery market data was used. It was collected using internal resources of one of the grocery market chains. Data set consists of 1.688 grocery stores belonging to both major and minor chains. Each store has coordinates expressed in WGS84 coordinates, opening date (when the store was first opened), closing date, which chain it belongs to, and sales area in square meters. Market share is expressed as sales share from total market sales. Data on a chain level was collected using publicly available sources. Sales data is for 2021. Facility level market share data for one of the chains operating in the market was collected using internal resources. Facility level market share is expressed as monthly sales average from total monthly sales for the chain.

3.1.1 Data processing

Demand point coordinates were originally expressed in LKS-94 coordinates and were converted to WGS84 coordinates using publicly available PHP code which was adapted for R programming language by the author. For facility data, stores that were closed before 2021.12.31 were removed from the data set. There were 188 such stores. For distance between demand points and facilities, Haversine distance was calculated.

3.2 Customer choice rules

In this section, a formal description of customer choice rules is presented. Existing facilities of the retail chain and those of competitors are considered. Binary, partially binary and proportional customer choice rules define how customers patronize facilities (refer to section 2.2 for details). All those rules are based on attraction (or utility) that demand point i feels towards facility j. Formally, attraction is defined as shown in (1):

$$a_{ij} = \frac{A_j^{\alpha}}{d_{ij}^{\beta}} \tag{1}$$

Here A_j is the attractiveness of the facility j, d_{ij} is the distance between demand point i and facility j, and α and β parameters are used to adjust effect of attractiveness and distance factors (for discussion on distance adjustment refer to section 2.1.1). Attractiveness adjustment parameter α may be used to account for various factors that influence differences between facilities. For example, Suhara et al. [33] used it to account for differences between urban and rural regions. In this paper, attractiveness adjustment parameter is used to factor in the differences between the retail chains. A simple method that considers chain's efficiency is proposed. Efficiency is defined as sales per square meter and it is considered to reflect differences in how effective each chain is in utilizing its sales area and, in turn, it indicates how attractive it is. Attractiveness parameter α for chain k is calculated as shown in (2):

$$\beta_k = \log_{S_t} S_k \tag{2}$$

Here S_k is sales per square meter for chain k and S_t is average sales per square meter for total market. For distance correction parameter β different values proposed in the literature (refer to section 2.1.1) are modeled and compared (for details about experimental setting refer to section 3.4).

Market share captured with binary customer choice rule by the retail chain k can then be defined as shown in (3):

$$M_{bk} = \sum_{i \in I^>} w_i \tag{3}$$

Here w_i is demand at demand point i and $I^>$ is defined as shown in (4):

$$I^> = \{i \in I : a_{ij} > max\, \{a_i(J_k) : k \in K\}\} \tag{4}$$

Here $a_i(J_k)$ is the maximum attraction that demand point i feels for facilities of the chain k, $a_i(J_k) = max\, \{a_{ij} : j \in J_k\}$

Market share captured with partially binary customer choice rule by the retail chain k can be defined as shown in 5:

$$M_{pbk} = \sum_{i \in I} w_i \frac{a_i(J_k)}{a_i(J_k) + \sum_{k \in K} a_i(J_k)} \tag{5}$$

Market share captured with proportional (Huff) binary customer choice rule by the retail chain k can be defined as shown in (6):

$$M_{prk} = \sum_{i \in I} w_i \frac{\sum_{j \in J_k} a_{ij}}{\sum_{j \in J} a_{ij}} \tag{6}$$

Here $\sum_{j \in J_k} a_{ij}$ is total attraction demand point i feels towards facilities of chain k, and $\sum_{j \in J} a_{ij}$ is total attraction demand point i feels towards facilities of all chains.

3.3 Competitive facility location problem for an existing firm

An existing retail chain wants to choose location for multiple new facilities from a set of candidate locations in order to maximize its total captured market share. There already exist k chains operating j facilities in the market. Customer demand is served proportionally by all facilities that fall within the pre-specified catchment area and demand is then split between those facilities proportionally to their attraction. Due to proportional customer choice rule, model is a nonlinear programming problem. Linearization method is based on [15] and [17], where linear reformulation for partially binary customer choice rule for a new firm is proposed. Here linearization method is adapted for model with proportional customer choice rule and existing firm already operating in the market and looking to expand.

3.3.1 Notation

The following general notation is used:

- i, I index and set of demand points (customers),
- k, K index and set of retail chains,
- j, J index and set of facilities (stores),
- O set of existing facilities that are part of one's own chain, $O \subset J$
- N set of existing facilities that belong to competitors, $N \subset J$
- w_i demand at i,
- A_j attractiveness of facility j,
- d_{ij} distance between demand point i and facility j,
- D_i maximum distance between facility and demand point i (customer catchment distance),
- L set of candidate locations for the new facilities,
- s a number of new facilities to be located,
- X set of new facilities locations, $X \subset L$, $|X| = s$

Due to the maximum distance condition:

$$a_{ij} = \begin{cases} \frac{A_j^{\alpha}}{d_{ij}^{\beta}} & \text{if } d_{ij} \leq D_i \\ 0 & \text{otherwise.} \end{cases}, \forall i \in I, \forall j \in J. \tag{7}$$

$M(X)$ denotes total market share captured by the existing firm with new facilities at X:

$$M(X)=\sum_{i\in I} w_i \frac{a_i(X)+\sum_{j\in O} a_{ij}}{a_i(X)+\sum_{j\in O} a_{ij}+\sum_{j\in N} a_{ij}}. \tag{8}$$

The problem is then:

$$\operatorname{Max}\left\{M(X):|X|=s, X\subset L\right\}. \tag{9}$$

If the following variables are considered:

$$\begin{aligned} x_j &= \begin{cases} 1 & \text{if a new facility is located at } j, \\ 0 & \text{otherwise} \end{cases} \\ y_{ij} &= \begin{cases} 1 & \text{if demand } i \text{ is allocated to a new facility } j, \\ 0 & \text{otherwise} \end{cases} \qquad j\in L \end{aligned} \tag{10}$$

The model can be formulated as binary nonlinear programming problem:

$$(\mathrm{PM})=\begin{cases} \operatorname{Max} \sum_{i\in I} w_i \frac{\sum_{j\in L} a_{ij}y_{ij}+\sum_{j\in O} a_{ij}}{\sum_{j\in L} a_{ij}y_{ij}+\sum_{j\in O} a_{ij}+\sum_{j\in N} a_{ij}} \\ \text{s.t.} \quad \sum_{j\in L} x_j = s, \\ \qquad y_{ij}\le x_j, \forall i\in I, \forall j\in L, \\ \qquad \sum_{j\in L} y_{ij}=1, \forall i\in I, \\ \qquad y_{ij}\in\{0,1\}, \forall i\in I, \forall j\in L, \\ \qquad x_j\in\{0,1\}, \forall j\in L. \end{cases} \tag{11}$$

Linearization of the objective function (PM). Formulate auxiliary variables:

$$\begin{aligned} q_i &= \frac{1}{\sum_{j\in L} a_{ij}y_{ij}+\sum_{j\in O} a_{ij}+\sum_{j\in N} a_{ij}}, \quad \forall i\in I, \\ z_i &= \sum_{j\in L} a_{ij}y_{ij}+\sum_{j\in O} a_{ij}, \forall i\in I, \\ v_{ij} &= w_i q_i z_i, \forall i\in I, \forall J\in L. \end{aligned} \tag{12}$$

Here, $0\leqslant q_i\leqslant M$, and $M=\max_i \frac{1}{\sum_{j\in J} a_{ij}}$.

Then, competitive facility location model with proportional customer choice rule can be formulated as a mixed binary linear programming problem:

$$
(\text{LPM}) = \begin{cases} \text{Max} \sum_{i \in I} \sum_{j \in L} v_{ij} \\ \text{s.t.} \sum_{j \in L} v_{ij} + w_i q_i \sum_{j \in N} a_{ij} = w_i, \forall i \in I, \\ \qquad v_{ij} \leq w_i z_i q_i, \forall i \in I, \forall J \in L, \\ \qquad v_{ij} \geqslant 0, \forall i \in I, \forall J \in L, \\ \qquad y_{ij} \leq x_j, \forall i \in I, \forall J \in L, \\ \qquad \sum_{j \in L} y_{ij} = 1, \forall i \in I, \\ \qquad \sum_{j \in L} x_j = s, \\ \qquad x_j \in \{0, 1\}, \forall j \in L, \\ \qquad y_{ij} \in \{0, 1\}, \; \forall i \in I, \forall j \in L. \end{cases} \tag{13}
$$

3.4 Experimental setting

As a first step, customer choice rules described in 3.2 were modelled and compared. Then, using results from step 1, mixed binary linear programming problem was formulated and solved (step 2). In the following sections each step is described in detail.

3.4.1 Validating customer choice rules

Customer choice rules were used to allocate demand to facilities, then demand was aggregated on a chain level and corresponding market shares were compared with actual market shares. Due to limited availability of facility level data for competing chains, market shares on a facility level were only compared for one chain. Due to computational complexity, small number of parameters were explored using grid search method. For distance correction parameter β, three different values were used $\beta = \{1, 2, 3\}$. These are based on findings from previous research (refer to section 2.1.1). For attractiveness correction parameter α, default value of 1 and chain-level attractiveness parameter calculated based on sales efficiency (refer to section 3.2) were used. To reduce computational complexity of facility location problem in step 2, multiple catchment distance values were explored. This was done in order to reduce the number of valid demand-facility pairings. In other words, it does not make sense to consider demand-facility pairs that are, for example, 300 km. away from each other. Maximum value was set to 100 km., minimum value to 2 km. and the step was set to 5 km. To compare customer choice rules and different parameter settings, multiple types of errors were calculated. Different types of errors have different properties. For example, $RMSE$ is sensitive to outliers, but is easy to interpret as its value is presented in original units. Mean percentage error (MPE) use actual values rather than absolute, therefore, it can be used as a measure of the bias. Types of errors used are presented in (14):

$$
\begin{aligned}
MAE &= \frac{\sum_{i=1}^{n} |\widehat{y}_i - y_i|}{n} \\
MPE &= \frac{1}{n} \sum_{t=1}^{n} \frac{\widehat{y}_i - y_i}{y_i} \\
MAPE &= \frac{1}{n} \sum_{t=1}^{n} \left| \frac{\widehat{y}_i - y_i}{y_i} \right| \\
RMSE &= \sqrt{\sum_{i=1}^{n} \frac{(\widehat{y}_i - y_i)^2}{n}}
\end{aligned}
\tag{14}
$$

3.4.2 Solving competitive facility location problem

Using results from section 3.4.1, competitive facility location model with proportional customer choice rule was formulated and solved with different number of new facilities to open. To do that, 10 facilities with known market shares that were opened in 2021, were selected and ranked based on their market share. Then facility location problem described in section 3.3 was formulated and solved using multiple values for new facilities to open. In this way, predicted ranks of candidate facilities were compared with actual ranks and model accuracy was evaluated. To examine if there is any cannibalization effect, the same procedure was repeated while excluding existing own facilities from numerator to model competitive facility location problem for new firm instead of existing one (refer to equation 7). This is an interesting exploration because the only difference between facility location problem for existing firm and the new firm is that in the former new facilities may cannibalize existing facilities resulting in different results. Despite new player entering the market being a rare event, large body of research focus on modelling facility location problem for new firms. Thus, it is interesting to explore the difference between the two models in a practical setting. In addition, such comparison between the models serves as a verification that the model is implemented properly. To solve the problem, commercial solver Gurobi with academic license and its Python API was used (for implementation details refer to Jupyter notebook stored in Github repository. Link provided in Appendix A).

4 Results and discussion

In this section results are presented and discussed. First, results for estimating market share on a chain level and on a facility level of one chain using binary, partially binary and proportional (Huff) customer choice rules are presented. Then, results for competitive facility location model with proportional customer choice rule are discussed. Conclusions and implications of the results are discussed within each subsection.

4.1 Customer choice rules

4.1.1 Chain level

Customer choice rules were modelled using different values of distance correction parameter β, attractiveness parameter α and different values for customer catchment distance (refer to section 3.4.1). Demand was then aggregated on a chain level and predicted market shares were compared with actual market shares. Multiple error calculations were used (see section 3.4.1), but all show the sames trends, therefore, only RMSE is presented (see Figure 1).

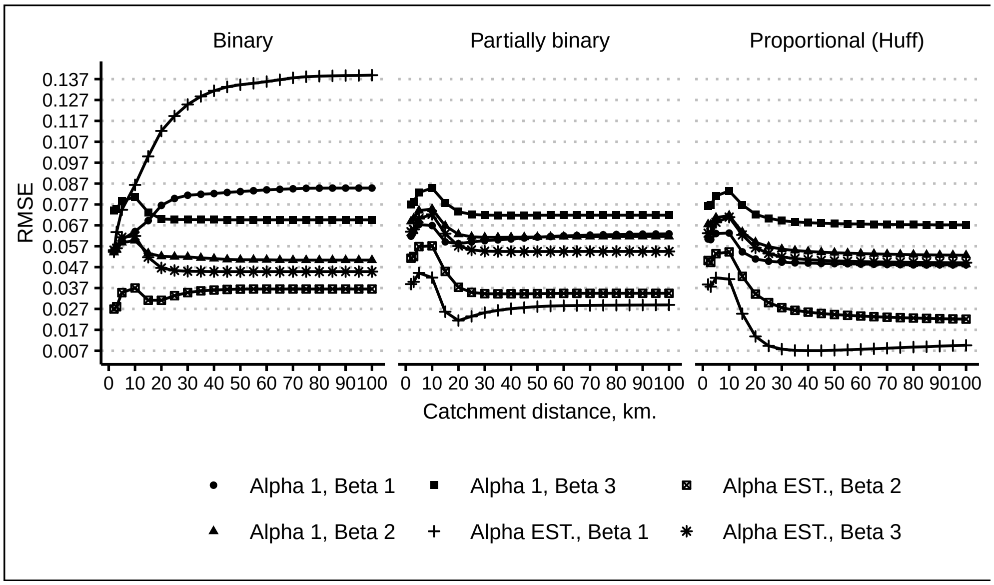

Figure 1: Market share RMSE on a chain level

Proportional (Huff) customer choice rule with distance correction parameter $\beta = 1$ and attractiveness correction parameter α, calculated based on chain's efficiency, and customer catchment distance of 45 km. produced the lowest RMSE of 0.007056. Interestingly, proportional customer choice rule with distance correction parameter $\beta = \{2, 3\}$ showed worse results than $\beta = 1$. This is in contrast to Huff [21], who found that for grocery stores distance decay parameter is equal to 3. In proportional customer choice rule, higher values of β showed worse results with different α values, which indicates that attractiveness correction parameter α may be a better target for further optimization. As for customer catchment distance, RMSE stays at the similar level until catchment distance reaches 10 km. and then starts to decrease at an increasing rate until it plateaus at 25 km. and reaches its minimum value at 40 km. After that it starts to increase steadily. The same trend is present for all other β and α values in proportional customer choice rule except that RMSE keeps decreasing with an increasing value of customer catchment distance.

Similar to proportional customer choice rule, partially binary customer choice rule also showed the lowest $RMSE = 0.021435$ with $\beta = 1$ and α estimated based on chain's efficiency. However, it reached its lowest value at around 20 km. of customer catchment distance. As for the trends, it also showed similar patterns as proportional customer choice rule (reaching a plateau with an increase in customer catchment distance).

In contrast to proportional and partially binary customer choice rules, binary customer choice rule showed the worst results with distance correction parameter $\beta = 1$ and estimated attractiveness parameter α and showed the lowest $RMSE = 0.02694$ with $\beta = 2$ and estimated α value. All in all, proportional customer choice rule showed superior results when evaluating market share on a chain level by a factor of 3.04 compared to partially binary customer choice rule and by a factor of 3.82 compared to binary choice rule, nevertheless, both binary and partially binary customer choice rules showed improved results with some parameter adjustment.

4.1.2 Facility level

Figure 2 shows results of the same experiment as described in 4.1.1, but instead of aggregating demand on a chain level, it is aggregated on a facility level for one chain and predicted market shares are compared with actual market shares. Results are similar to 4.1.1 in a way that proportional customer choice rule showed the smallest errors and binary choice rule showed the highest. In addition, proportional customer choice rule showed higher errors for higher distance correction parameter β, which is again in contrast to findings in [21]. Otherwise, results are rather different. First, there is virtually no difference between errors for different values of attractiveness correction parameter α. This is not truly surprising as the parameter is estimated on a chain level, so stores evaluated in this step all have the same α value. There is also no clear pattern in relation to customer catchment distance. All customer choice rules show lower errors with smaller customer catchment distance when distance correction parameter $\beta = 1$. Error slightly increases with an increase in catchment distance and then remains at the similar levels. This is in contrast to $\beta > 1$, where the opposite pattern is evident: errors are higher with smaller customer catchment distance values, starts decreasing and reaches a plateau.

For facility level market share estimation, it is interesting to explore mean absolute percentage error (MAPE) as in practice it is more common to compare actual store results with comparable figures in percentage terms. Results are presented in Table 2 (see Appendix B). General patterns are the same as in case of RMSE and proportional customer behavior rule with $\beta = 1$ and estimated α shows the lowest error (0.302) at customer catchment distance of 3 km. This is in contrast to chain level market share estimation, where lowest errors were reached at around 30 km. catchment distance. Nevertheless, 30.2% prediction error is rather poor result, which indicates that model parameters needs to be optimized not only on a chain level, but on a facility level as well. Figure 3 further illustrates this point as it shows that model tends to overestimate market share. Therefore, further investigations are needed in order achieve satisfactory results on facility level.

4.2 Competitive facility location problem

Using results from section 4.1.2, competitive facility location model with proportional customer choice rule, distance correction parameter $\beta = 1$ and estimated attractiveness correction parameter α was formulated and solved with different number of new facilities to open. This was done in

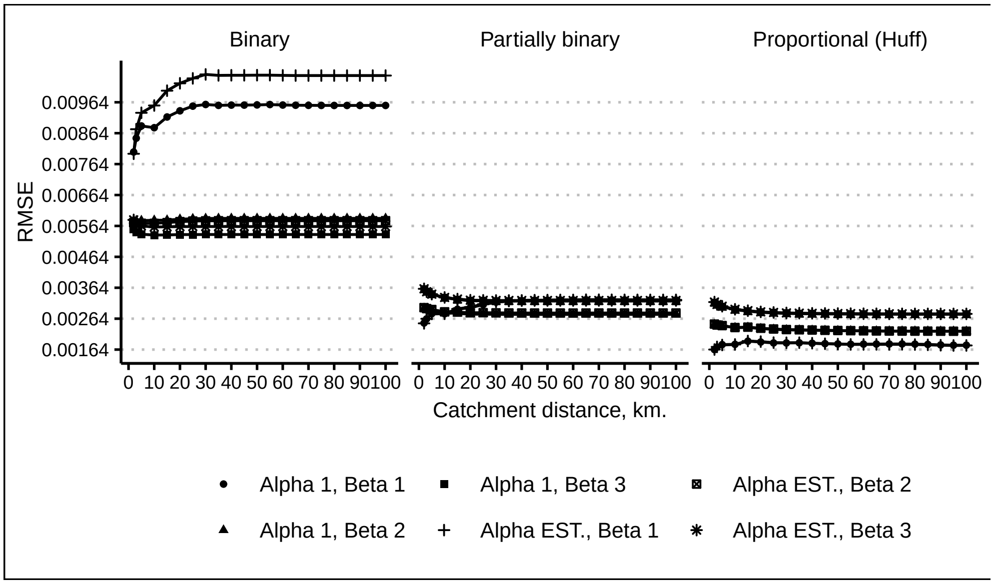

Figure 2: Market share RMSE on a facility level

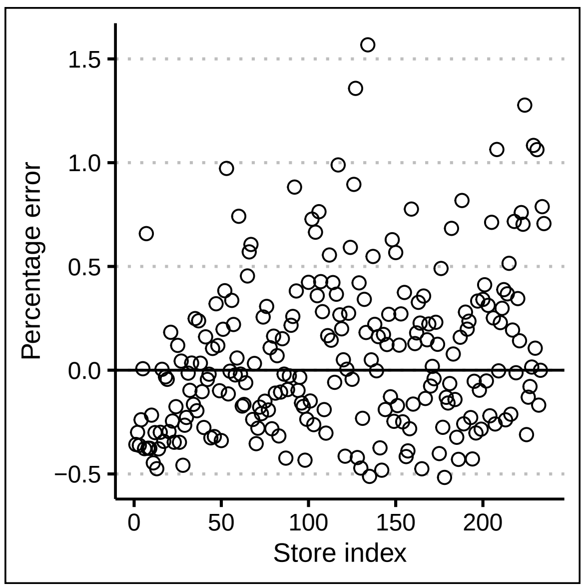

Figure 3: Market share percentage error with proportional choice rule

order to be able to make ranking of candidate facilities. Predicted ranking was then compared with actual ranking in order to evaluate accuracy. Furthermore, effect of cannibalization was examined by formulating second model without existing own facilities (for methodology description refer to section 3.4.2).

Problem was implemented and solved using commercial solver Gurobi with academic license and its Python API (for implementation details refer to Jupyter notebook stored in Github repository. Link provided in Appendix A). Model with one new facility to find had 10 167 quadratic objective terms, same number of quadratic constraints, 30 501 continuous and 10 binary variables. Solver

Candidate location no.	Observed rank	Predicted rank model 1	Predicted vs obs. model 1	Predicted rank model 2	Predicted vs obs. model 2
I1/00168	1	1	0	2	1
I1/00172	2	5	3	5	3
I1/00234	3	2	-1	1	-2
I1/00178	4	6	2	6	2
I1/00209	5	4	-1	4	-1
I1/00215	6	7	1	7	1
I1/00235	7	8	1	8	1
I1/00171	8	3	-5	3	-5
I1/00227	9	9	0	9	0
I1/00222	10	10	0	10	0

Table 1: Observed and predicted candidate location rankings

explored 402 simplex iterations in 0.23 seconds and found optimal solution with tolerance of $1.00e^{-4}$ and 0% gap. Total model run time, including variable and constraints setup, was 2 min. 30 s.

Predicted and observed rankings of candidate locations are presented in Table 1. Model 1, which included both new and existing own facilities, managed to accurately predict ranking for candidate location with highest actual market share. However, for other facilities the error varies up to 5 positions. This is in line with findings in section 4.1.2 and again indicates the need for further refinement of the allocation model parameters. Interestingly, rankings of model 2, where existing own facilities where excluded to represent new firm entering the market, slightly differs from the rankings of model 1. Specifically, model 2 was not able to correctly rank facility with the highest ranking. It turns out that facility I1/00234 for which model 2 predicted rank 1, is located 200 m. away from other facility owned by the same chain, therefore. This indicates that it is worth considering cannibalization effect when formulating competitive facility location problem for existing firm. In addition, it also validates that model 1 was implemented correctly.

5 Conclusions

This paper makes use of a unique data set of 34 664 demand points and 1 688 facilities to model competitive facility location problem. Binary, partially binary and proportional (Huff gravity model) customer choice rules were used to estimate market share using range of input parameters. Results show that proportional customer choice rule outperforms other customer choice models based on comparison of different types of errors (for detailed discussion refer to section 4.1). However, further adjustments in model parameters are needed to achieve satisfactory facility level market share estimates. In addition, both binary and partially binary customer choice rules show significant improvements in market share estimates with parameter adjustment, therefore, both rules can be considered to be valid assumptions about customer behavior. Interestingly, results show that, in contrast to Huff [20], increase in the value of distance correction parameter β with proportional customer behavior rule results in higher errors in market share estimates. This is true for different attractiveness parameter values, which may be an indication that for Lithuanian grocery market customers, distance to the store is as important as other attractiveness factors.

Different values of customer catchment distance were modelled with a catchment distance of around 40 km. showing the smallest market share error on a chain level with proportional customer choice rule. Interestingly, catchment areas from 25 km. to 100 km. show very similar results, but error starts increasing rapidly going from catchment area of 25 km. to 10 km. and then reaches a plateau. Relationship between catchment area and accuracy of the model is different when looking on a facility level. There are no clear patterns or fluctuations with error staying at the similar level throughout the whole range of catchment distance values. This points to the need to focus on optimizing attractiveness correction parameter when facility level estimates are the target.

In addition, a simple method for calculating attractiveness parameter, which improves accuracy of the model, is proposed. It is based on the efficiency of the chain, which is reflected in the average sales per square meter. With this estimated attractiveness parameter error for proportional customer choice model shows seven-fold improvement when compared to the base model. Nevertheless, it shows almost no improvement when estimating market shares on a facility level, which indicates that attractiveness within a chain is not homogeneous and to improve the accuracy of the model attractiveness parameter should be optimized on a facility level.

Using parameters selected in validation step, competitive facility location model with proportional customer choice rule for existing firm was formulated as mixed binary linear programming problem. This was done in order to simplify the process of selecting the best location as solving it exactly is not efficient when the number of candidate locations is high. Commercial solver Gurobi and its Python API was used to formulate and solve the optimization problem of finding the best location out of multiple candidate locations and actual market shares were used to evaluate the results (for setting description refer to section 3.4.2). When asked to select only one new facility to build, model was able to select the one that actually has the highest market share, however, if the number of new facilities were increased, model ranking was not entirely in line with actual rankings of those facilities (for results refer to section 4.2). In addition, two models, one for existing firm looking to expand and one for new entrant, were formulated and solved showing slightly different results. This points to the need to take into consideration existing facilities when solving competitive facility location problem for an existing firm. This may not always be evident as usually candidate facilities are selected in a way so that cannibalization effect is minimized. Finally, implementation in Python showed satisfactory run time, nevertheless, to truly evaluate capabilities of model implementation, a comparison with other solvers using a more complex problem formulation, is needed.

Proposed modelling methods show good results on a total chain level, however, the accuracy suffers when models are used on a facility level. This indicates that there's a need for further improvement of these models with a focus on facility level parameters. For example, including income information may reduce facility level error as this would help differentiate between stores located in richer urban areas from stores located in poorer rural areas. In addition, using travel time or Manhattan distance as a distance measure may also help improve the model as this would better reflect ease of access to the stores.

As for competitive facility location problem, formulation of the mixed binary linear programming problem could be further improved, for example, by introducing market share constraints for new facilities. This would be a useful addition as firms usually aim to open new facilities only if they can reach a certain market share. Finally, implementation of the problem solution using open source solvers could help understand the true efficiency of Gurobi solver and it could also make solution accessible to wider audience.

References

[1] Soumen Atta, Priya Ranjan Sinha Mahapatra, and Anirban Mukhopadhyay. Multi-objective uncapacitated facility location problem with customers' preferences: Pareto-based and weighted sum ga-based approaches. *Soft Computing*, 23(23):12347–12362, 12/01 2019. ID: Atta2019.

[2] Maria Barbati. Models and algorithms for facility location problems with equity considerations. 2013.

[3] David R. Bell, Teck-Hua Ho, and Christopher S. Tang. Determining where to shop: Fixed and variable costs of shopping. *Journal of Marketing Research*, 35(3):352–369, 2022/12 1998. 31.

[4] Stefano Benati and Pierre Hansen. The maximum capture problem with random utilities: Problem formulation and algorithms. *European Journal of Operational Research*, 143(3):518–530, 16 December 2002 2002. ID: 271700.

[5] Oded Berman, Tammy Drezner, Zvi Drezner, and Dmitry Krass. *Modeling Competitive Facility Location Problems: New Approaches and Results*, pages 156–181. 09 2009.

[6] Benjamin Biesinger, Bin Hu, and Günther Raidl. Models and algorithms for competitive facility location problems with different customer behavior. *Annals of Mathematics and Artificial Intelligence*, 76(1):93–119, 02/01 2016. ID: Biesinger2016.

[7] Julien Bramel and David Simchi-Levi. A location based heuristic for general routing problems. *Operations research*, 43(4):649–660, 08/01; 2022/05 1995. doi: 10.1287/opre.43.4.649; 29.

[8] Jun Dong Cho and Majid Sarrafzadeh. The pin redistribution problem in multi-chip modules. *Mathematical Programming*, 63(1):297–330, 01/01 1994. ID: Cho1994.

[9] John Current, Mark Daskin, and David Schilling. *Discrete Network Location Problems*. 01/01 2001.

[10] Tammy Drezner. Derived attractiveness of shopping malls. *IMA Journal of Management Mathematics*, 17, 05 2006.

[11] Tammy Drezner and Zvi Drezner. Replacing continuous demand with discrete demand in a competitive location model. *Naval Research Logistics (NRL)*, 44(1):81–95, 02/01; 2023/01 1997. https://doi.org/10.1002/(SICI)1520-6750(199702)44:13.0.CO;2-I; 05.

[12] Tammy Drezner and Zvi Drezner. Validating the gravity-based competitive location model using inferred attractiveness. *Annals of Operations Research*, 111(1):227–237, 03/01 2002. ID: Drezner2002.

[13] Zvi Drezner. Location of regional facilities. *Naval Research Logistics Quarterly*, 33(3):523–529, 08/01; 2023/01 1986. https://doi.org/10.1002/nav.3800330316; 05.

[14] Hülya Emir-Farinas and Richard L. Francis. Demand point aggregation for planar covering location models. *Annals of Operations Research*, 136:175–192, 2005.

[15] Pascual Fernandez, Algirdas Lančinskas, Blas Pelegrín, and Julius Žilinskas. A discrete competitive facility location model with minimal market share constraints and equity-based ties breaking rule. *Informatica*, pages 1–20, 01/01 2020.

[16] José Fernández, Blas Pelegrı´n, Frank Plastria, and Boglárka Tóth. Solving a huff-like competitive location and design model for profit maximization in the plane. *European Journal of Operational Research*, 179(3):1274–1287, 16 June 2007 2007. ID: 271700.

[17] Pascual Fernández, Blas Pelegrín, Algirdas Lančinskas, and Julius Žilinskas. New heuristic algorithms for discrete competitive location problems with binary and partially binary customer behavior. *Computers Operations Research*, 79:12–18, March 2017 2017. ID: 271709.

[18] M. J. Hodgson, Richard T. Wong, and John Honsaker. The p-centroid problem on an inclined plane. *Operations research*, 35(2):221–233, 2022/05 1987. 29.

[19] Harold Hotelling. Stability in competition. *The Economic Journal*, 39(153):41–57, 2022/12 1929. 31.

[20] David L. Huff. Defining and estimating a trading area. *Journal of Marketing*, 28(3):34–38, 2022/12 1964. 31.

[21] David L. Huff. A programmed solution for approximating an optimum retail location. *Land Economics*, 42(3):293–303, 2022/12 1966. 31.

[22] Hande Küçükaydin, Necati Aras, and I. Kuban Altınel. Competitive facility location problem with attractiveness adjustment of the follower: A bilevel programming model and its solution. *European Journal of Operational Research*, 208(3):206–220, 1 February 2011 2011. ID: 271700.

[23] Algirdas Lančinskas, Pilar Martinez Ortigosa, and Julius Žilinskas. Parallel optimization algorithm for competitive facility location. *Mathematical Modelling and Analysis*, 20(5):619–640, 2015.

[24] Algirdas Lančinskas, Julius Žilinskas, Pascual Fernández, and Blas Pelegrín. Solution of asymmetric discrete competitive facility location problems using ranking of candidate locations. *Soft Computing*, 24(23):17705–17713, 12/01 2020. ID: Lančinskas2020.

[25] Ronald G. McGarvey and Tom M. Cavalier. Constrained location of competitive facilities in the plane. *Computers Operations Research*, 32(2):359–378, February 2005 2005. ID: 271709.

[26] María Merino and Adrian Ramirez-Nafarrate. Estimation of retail sales under competitive location in mexico. *Journal of Business Research*, 69(2):445–451, February 2016 2016. ID: 271680.

[27] Masao Nakanishi and Lee G. Cooper. Parameter estimation for a multiplicative competitive interaction model: Least squares approach. *Journal of Marketing Research*, 11(3):303–311, 2022/12 1974. 31.

[28] Blas Pelegrín, Pascual Fernández, and María Dolores García Pérez. On tie breaking in competitive location under binary customer behavior. *Omega*, 52:156–167, April 2015 2015. ID: 271690.

[29] Frank Plastria. Static competitive facility location: An overview of optimisation approaches. *European Journal of Operational Research*, 129(3):461–470, 16 March 2001 2001. ID: 271700.

[30] W.J. Reilly. *The Law of Retail Gravitation*. W.J. Reilly, 1931.

[31] C. S. ReVelle, H. A. Eiselt, and Mark Daskin. A bibliography for some fundamental problem categories in discrete location science. *European Journal of Operational Research*, 184:817–848, 02/01 2008.

[32] David Schilling. Dynamic location modeling for public-sector facilities: A multicriteria approach. *Decision Sciences*, 11:714–724, 06/07 2007.

[33] Yoshihiko Suhara, Mohsen Bahrami, Burçin Bozkaya, and Alex 'Sandy' Pentland. Validating gravity-based market share models using large-scale transactional data, 2019.

6 Appendices

6.1 Appendix A

Computer code is stored in the publicly available Github repository and can be accessed via link below. Facility level market share data and store location data is property of one retail chain, therefore, only computer code is made publicly available. To request access to the data, please, contact author of the thesis via email: `jankauskas.ev@gmail.com`.

`https://github.com/EvaldasJ/cflp`

6.2 Appendix B

Catchment	B, α 1, β 1	B, α EST, β 1	B, α 1, β 2	B, α EST, β 2	B, α 1, β 3	B, α EST, β 3	PB, α 1, β 1	PB, α EST, β 1	PB, α 1, β 2	PB, α EST, β 2	PB, α 1, β 3	PB, α EST, β 3	PR, α 1, β 1	PR, α EST, β 1	PR, α 1, β 2	PR, α EST, β 2	PR, α 1, β 3	PR, α EST, β 3
2	1.295	1.369	1.167	1.151	1.151	1.227	0.562	0.560	0.622	0.616	0.738	0.734	0.308	0.305	0.433	0.428	0.592	0.591
3	1.352	1.468	1.166	1.150	1.144	1.213	0.586	0.587	0.618	0.613	0.718	0.713	0.305	0.302	0.420	0.416	0.573	0.570
5	1.400	1.451	1.172	1.155	1.141	1.206	0.613	0.610	0.614	0.612	0.705	0.703	0.311	0.308	0.412	0.409	0.563	0.561
10	1.425	1.498	1.187	1.170	1.150	1.211	0.652	0.650	0.619	0.622	0.700	0.699	0.335	0.334	0.415	0.413	0.552	0.554
15	1.452	1.570	1.189	1.177	1.153	1.216	0.698	0.699	0.629	0.634	0.701	0.701	0.359	0.362	0.424	0.424	0.552	0.555
20	1.436	1.557	1.189	1.174	1.153	1.216	0.698	0.700	0.625	0.630	0.692	0.695	0.355	0.359	0.414	0.414	0.543	0.546
25	1.430	1.546	1.188	1.173	1.153	1.215	0.687	0.689	0.620	0.624	0.689	0.692	0.344	0.347	0.407	0.405	0.535	0.538
30	1.419	1.541	1.184	1.170	1.154	1.216	0.675	0.676	0.613	0.617	0.686	0.689	0.347	0.348	0.398	0.397	0.527	0.530
35	1.419	1.541	1.184	1.169	1.154	1.216	0.671	0.671	0.612	0.615	0.685	0.688	0.352	0.353	0.394	0.393	0.523	0.526
40	1.417	1.540	1.184	1.169	1.154	1.215	0.667	0.666	0.611	0.614	0.685	0.688	0.347	0.348	0.391	0.389	0.521	0.523
45	1.417	1.540	1.184	1.169	1.154	1.215	0.661	0.660	0.611	0.614	0.685	0.688	0.344	0.343	0.388	0.386	0.518	0.521
50	1.416	1.540	1.184	1.169	1.154	1.215	0.657	0.655	0.610	0.614	0.685	0.688	0.341	0.340	0.386	0.383	0.517	0.519
55	1.414	1.540	1.184	1.169	1.154	1.215	0.655	0.653	0.610	0.614	0.685	0.688	0.338	0.337	0.385	0.382	0.516	0.518
60	1.414	1.540	1.184	1.169	1.154	1.215	0.654	0.651	0.610	0.614	0.685	0.687	0.337	0.335	0.384	0.381	0.515	0.517
65	1.414	1.540	1.184	1.169	1.154	1.215	0.653	0.651	0.610	0.614	0.685	0.687	0.336	0.333	0.383	0.380	0.514	0.516
70	1.414	1.540	1.184	1.169	1.154	1.215	0.652	0.650	0.610	0.614	0.685	0.687	0.337	0.334	0.382	0.379	0.513	0.516
75	1.414	1.540	1.184	1.169	1.154	1.215	0.652	0.649	0.610	0.613	0.685	0.687	0.339	0.336	0.381	0.378	0.513	0.515
80	1.414	1.540	1.184	1.169	1.154	1.215	0.651	0.649	0.610	0.613	0.685	0.687	0.338	0.335	0.381	0.378	0.513	0.515
85	1.414	1.540	1.184	1.169	1.154	1.215	0.651	0.649	0.610	0.613	0.685	0.687	0.337	0.334	0.380	0.378	0.513	0.515
90	1.414	1.540	1.184	1.169	1.154	1.215	0.651	0.649	0.610	0.613	0.685	0.687	0.336	0.334	0.380	0.378	0.512	0.515
95	1.414	1.540	1.184	1.169	1.154	1.215	0.651	0.649	0.610	0.613	0.685	0.687	0.337	0.334	0.380	0.377	0.512	0.515
100	1.414	1.540	1.184	1.169	1.154	1.215	0.651	0.649	0.610	0.613	0.685	0.687	0.338	0.335	0.380	0.377	0.512	0.515

Table 2: Facility level customer choice rules MAPE