A NEW MATHEMATICAL PROGRAMMING FORMULATION FOR MULTIVARIATE REGRESSION CLUSTERING WITH A STORE CLUSTERING APPLICATION IN RETAIL SECTOR

(ÇOKLU ÇIKTI DEĞİŞKENLİ REGRESYON TABANLI KÜMELEME İÇİN YENİ BİR MATEMATİKSEL PROGRAMLAMA FORMÜLASYONU VE PERAKENDE SEKTÖRÜNDE MAĞAZA KÜMELEME UYGULAMASI)

by

Alper ALTINOY

Thesis

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

in

LOGISTICS AND FINANCIAL MANAGEMENT

in the

GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

of

GALATASARAY UNIVERSITY

Supervisor: Assoc. Prof. Dr. Ebru Angün

June 2019

This is to certify that the thesis entitled

A NEW MATHEMATICAL PROGRAMMING FORMULATION FOR MULTIVARIATE REGRESSION CLUSTERING WITH A STORE CLUSTERING APPLICATION IN RETAIL SECTOR

prepared by **Alper ALTINOY** in partial fulfillment of the requirements for the degree of **Master of Science in Logistics and Financial Management** at the **Galatasaray University** is approved by the

Examining Committee:

Assoc. Prof. Dr. Ebru ANGÜN (Supervisor)
Department of Industrial Engineering
Galatasaray University -------------------------

Prof. Dr. Temel ÖNCAN
Department of Industrial Engineering
Galatasaray University -------------------------

Prof. Dr. Semra AĞRALI
Department of Industrial Engineering
Mef University -------------------------

Date: -------------------------

ACKNOWLEDGEMENTS

First and foremost, I would like to sincere gratitude to my advisor, Assoc. Prof. Dr. M. Ebru Angün for her invaluable support and guidance. I have learned a lot from her. I owe many thanks to her advice and ongoing support for my career.

I am especially indebted to my parents and my brother, Binnas Altınoy, Halil Altınoy and Atilla Altınoy. I dedicate this thesis to them.

Special thanks go to İbrahim for helping me a lot by sharing his valuable knowledge and experience in retail sector.

I am also grateful to my dear friend Aylin for her advices during my thesis journey.

Lastly, I would like to thanks to my friends, Ürem, Yasin and Fatih for their friendship and assistance.

To all other people that their names are not mentioned here who contribute to this master study; I would like to thank them all.

June 2019
Alper ALTINOY

TABLE OF CONTENTS

ACKNOWLEDGEMENTS .... **iii**

TABLE OF CONTENTS .... **iv**

LIST OF SYMBOLS .... **vi**

LIST OF FIGURES .... **vii**

LIST OF TABLES .... **viii**

ABSTRACT .... **ix**

ÖZET .... **xi**

1. INTRODUCTION **1**

2. LITERATURE REVIEW **4**

2.1 Store Clustering or Segmentation 4

2.2 Retail Store Performance 8

2.3 Clusterwise Linear Regression 31

2.3.1 Clusterwise Linear Regression Problem and Classical Methods........ 31

2.3.2 Mathematical Programming Methods 38

3. METHODOLOGY **44**

3.1 Mathematical Model Formulations 45

3.1.1 Big-M Formulation 45

3.1.2 SOS Type-1 Formulation 47

3.2 Determination of Cluster Number Q 49

4. EXPERIMENTS AND RESULTS **51**

4.1 Proposed Model Application on a Problem from the Literature 51

4.1.1 Objective and Scope of Application 51

4.1.2 Data and Data Preprocessing 51

4.1.3 Results and Discussions 53

4.2 Proposed Model Application on the Store Clustering Problem 57

4.2.1 Objective and Scope of Application 57

4.2.2 Data and Data Preprocessing 57

4.2.3 Results and Discussions 60

5. CONCLUSIONS 63

REFERENCES 64

BIOGRAPHICAL SKETCH 75

LIST OF SYMBOLS

AIC : Akaike Information Criterion
BIC : Bayesian Information Criterion
CRM : Customer Relationship Management
DC : Difference of Convex
EBIT : Earnings Before Interest and Taxes
EBITDA : Earnings Before Interest, Taxes, Depreciation and Amortization
E-M : Expectation Maximization
ERP : Enterprise Resource Planning
FPC : Fixed Point Cluster
GA : Genetic Algorithm
GCLR : Generalized Clusterwise Linear Regression
MAE : Mean Absolute Error
MILP : Mixed-Integer Linear Programming
MIQP : Mixed-Integer Quadratic Programming
MLQP : Mixed Logical Quadratic Programming
PLS : Partial Least Squares
POI : Point of Interests
S&L : Savings and Loan
SA : Simulated Annealing
SKU : Stock Keeping Unit
SOS : Special Ordered Set
TUIK : Turkish Statistical Institute
VIF : Variance Inflation Factor
VNS : Variable Neighborhood Search

LIST OF FIGURES

Figure 4.1: VIF values and correlation matrix of explanatory variables in JURA data 53
Figure 4.2: Explanatory variables used in GCLR for store clustering 58
Figure 4.3: Brand positioning study in fashion retail industry 59
Figure 4.4: VIF values and correlation matrix of explanatory variables in store clustering data 60
Figure 4.5: VIF values and correlation matrix of explanatory variables except HE variable.. 60

LIST OF TABLES

Table 2.1: Performance measures in the literature.. 9
Table 2.2: Elements influencing store performance 10
Table 2.3: Comprehensive literature review on store performance, store clustering and selection of store location 11
Table 4.1: Descriptive statistics of explanatory and response variables in JURA dataset 52
Table 4.2: Optimal assignments of 30 locations to two and three clusters and their optimal objective value 54
Table 4.3: Regression parameter values for all clusters and responses 55
Table 4.4: Optimal assignments of 30 locations based on different approaches 56
Table 4.5: Descriptive statistics of explanatory variables in store clustering dataset.... 59
Table 4.6: Comparison of running times of formulations for different number of clusters for store clustering 61
Table 4.7: Optimal assignments of stores into clusters... 61
Table 4.8: Regression parameter values for all clusters and responses for store clustering... 62

ABSTRACT

Clustering in machine learning is an unsupervised learning technique, and it is the study of grouping similar entities together. In particular, clustering via regression groups entities such that the fit of a linear-in-parameter model is "best" with respect to a norm function of residuals---usually an L^2-norm (Euclidean norm) of residuals, which gives the least-squares estimates of the unknown regression parameters. This study considers linear-in-parameters regression model-based clustering when there are multiple correlated responses (i.e., dependent variables). Furthermore, this study proposes two new mixed-integer linear programming formulations, where the first formulation is a simple extension of the classic formulations with the big-M constraints in the literature for multiple responses. Besides, the objective function minimizes L^∞-norm of residuals instead of L^1-norm or L^2-norm. However, the big-M in the formulations depends on the unknown clusters and regression parameters, and hence, it is impossible to know a "good" value for the big-M a priori. A too big value for the big-M results in numerical instabilities, and a too small value for the big-M cuts the true optimal solution. Therefore, this study proposes to replace the classic formulations with the big-M constraints with formulations that have special ordered sets of type one (SOS Type-1) variables.

The proposed formulation for multiple responses with the big-M constraints is applied to the JURA dataset, which clusters locations to predict the concentration of some metals that are more expensive to measure using measurements of metals that are cheaper to sample. For this numerical example, this study illustrates that the classic approach in the literature, which considers one response at-a-time, usually assigns the same entity to different clusters with respect to different responses; hence, eventually, it is not evident to which cluster that entity belongs to. Also, even though the classic approach assigns that entity to the same cluster with respect to all responses, this

assignment can be different than the one obtained when all responses are considered simultaneously; hence, the classic approach can result in a false clustering when multiple corelated responses have to be considered at the same time.

The proposed formulations and algorithms are also applied for store clustering problem of a fashion retailer in Turkey. The store clustering problem simply means grouping of the stores based on some predetermined characteristics so that those stores within the same group are more similar to each other than the other stores. The purpose of such a clustering can be to detect and analyze the effects of location, customer and store characteristics on the store performance, which can depend on different clusters. The proposed models are solved using the branch-and-cut algorithm using optimization software Gurobi 8.1.0 C API.

ÖZET

Kümeleme, birbirine benzer elemanların benzer gruplara ayrılmasını amaçlayan, gözetimsiz öğrenme tekniklerinden biridir. Doğrusal regresyon tabanlı kümeleme ise benzer elemanları, her bir grup için uyarlanan doğrusal regresyon model hatalarının belirli bir normda en iyilecek şekilde gruplamasıdır---literatürde genellikle L^2-normu (Öklid normu) minimize eden en küçük kareler yöntemi kullanılarak regresyon parametreleri tahmin edilir. Bu çalışma yüksek korelasyona sahip birden fazla bağımlı değişkenin varlığında kullanılabilecek doğrusal regresyon tabanlı kümeleme modelini dikkate almaktadır. Ayrıca iki adet yeni karmaşık tamsayılı doğrusal programlama formülasyonu önerilmiştir. İlk formülasyon literatürde bulunan klasik büyük M kısıtlı modelin birden fazla bağımlı değişkeni dikkate alacak şekilde genişletilmiş halidir. Bununla birlikte amaç fonksiyonunda L^1 ve L^2 norm yerine L^∞ normu minimize edilmektedir. Ancak büyük M değeri bilinmeyen kümeleme ve regresyon parametrelerine bağımlı olduğundan dolayı, uygun bir büyük M değerini önceden bilmek mümkün değildir. Çok büyük bir büyük M değeri çözümde sayısal dengesizliklere, çok küçük bir büyük M değeri ise optimal sonuca ulaşamamaya neden olmaktadır. Bu yüzden çalışmamızda büyük M kısıtlarını içeren klasik formülasyon yerine tip 1 özel sıralı set değişkenlerini içeren formülasyon önerilmiştir.

Birden fazla çıktı değişkeni için önerilen yeni büyük M kısıtlı formülasyon literatürde bulunan JURA veri setine uygulanmıştır. JURA veri setindeki farklı bölgelerden alınan toprak örneklerinde ölçüm maliyeti düşük olan metallerin konsantrasyonu ile ölçüm maliyeti yüksek olan metallerin konsantrasyonu tahmin edilmeye çalışılmaktadır. Bu uygulamada, her bir çıktı değişkenini ayrı ayrı dikkate alarak gerçekleştirilen kümeleme sonuçlarının birbirinden farklı olduğu ve bu nedenle hangi örneğin hangi kümeye ait olduğu bilgisinin belirli olmadığı gösterilmiştir. Her çıktı değişkenini ayrı ayrı dikkate alarak gerçekleştirilen kümeleme sonuçları birbiri ile aynı olsa bile, bütün değişkenleri

eş zamanlı değerlendirerek yapılan kümeleme sonuçlarından farklı olabilecektir. Bu nedenle birden fazla yüksek korelasyona sahip değişkenin bulunduğu durumlarda bu klasik yaklaşım yanlış kümeleme sonuçlarına neden olabilir.

Son olarak, önerilen yeni matematiksel formülasyonlar Türkiye'de faaliyet gösteren bir moda perakendecisine ait mağazaların kümelenmesi problemine uygulanmıştır. Mağaza kümelenmesi kısaca önceden belirlenen özelliklere dayanarak benzer mağazaların gruplara ayrılmasını ifade eder. Bu kümeleme çalışmasının amacı küme bazında değişiklik gösterecek olan lokasyon, müşteri ve mağazaya ait özelliklerin ilgili mağaza performansı üzerindeki etkilerini belirlemek ve analiz etmektir. Önerilen matematiksel modeller Gurobi 8.1.0 çözücüsünde dal kesme algoritması kullanılarak çözümlenmiştir.

1. INTRODUCTION

Today, the amount of data produced and stored at the global level is unimaginable and is growing day by day. Therefore, new terms such as big data, data mining have become buzzwords in recent times. Clustering and regression are among the most popular data mining techniques. The objective of clustering is to divide entities into similar groups based on some features. Clustering is considered as an unsupervised learning method in data mining literature, which means different clusters have no labels. Clustering is usually related to the proximity, samples that are closer to each other should be considered as belonging to the same clusters. On the other hand, regression analysis relies on fitting a function (mostly linear) to the data to discover how one or more variables vary as a function of another. In addition, regression is considered as supervised learning method in data mining.

Clusterwise linear regression (CLR) is an approach that performs the clustering and regression simultaneously and is developed by Späth (1979). More precisely, CLR problem is defined to find clusters of entities such that a norm function of residuals of the fitted regression models over all clusters is minimized, where a residual is defined to be the difference between the observed value of the dependent variable and its estimated value from the fitted regression function. Although the CLR problem can be solved with a two-step approach, which first executes the clustering and then fits linear functions for each obtained cluster separately, this approach has some drawbacks like it is unclear which group of variables (explanatory or response) should be considered in clustering process. More importantly, clustering and regression try to optimize different criteria, which are irrelevant.

In the CLR problem, the assignment variables and the regression parameters are decision variables. The CLR problem can be formulated as a mixed-integer linear programming (MILP) or mixed-integer quadratic programming (MIQP) depending on the choice of the norm function of residuals. Because of the combinatorial nature of the CLR problem researchers generally have been working on algorithmic aspect of the problem. There were different approaches in the literature to solve the CLR problem based on as K-means (Späth, 1979; Zhang, 2003), Expectation-Maximization (E-M) (DeSarbo and Cron, 1988; Preda and Saporta, 2005), and optimization (Lau et al., 1999; Bertsimas and Shioda, 2007; Carbonneau et al., 2011; Bagirov et al., 2013; Park et al., 2017).

As a result of globalization and increasing competition, store clustering or segmentation has become popular and recommended to retailers by the institutions that provide consulting services. Store clustering simply means grouping of stores based on predetermined characteristics; therefore, stores within a group are more similar to each other. Different possible objectives for store segmentation can be summarized as follows:

- Assortment Planning
- Space Planning
- Promotion Planning
- Pricing Strategies
- Inventory and Labor Management
- Performance Monitoring

In this study, we consider the last objective, which is monitoring the performance of stores by segment type. Retailers will be able to identify the stores that are under-performing against potential and allocate right resources or make right decisions on each of these stores by understanding the factors affecting the performance.

It is possible to group the stores into several groups that are supposed to be homogenous to support performance evaluation by considering just one performance criterion like sales, demand, or profit. However, criterion that is used to define homogeneous groups remains too macro because the stores within a cluster are still dissimilar in terms of competitive conditions, customer characteristics, location characteristics, and product variety. Therefore, it would be appropriate to employ clusterwise linear regression for this problem, we propose new mixed-integer linear programming formulations with and without the big-M constraints.

The contributions of this study to the literature can be summarized as follows:

- Proposed mathematical formulation with big-M constraints considers multiple responses for which regression models have to be fitted simultaneously.
- Instead of minimizing the sum of absolute or squared residuals, our formulation minimizes the maximum of the sum of residuals.
- A new mixed-integer optimization formulation that overcomes the problem of the determination of the big-M value is proposed using special ordered sets of type one (SOS Type-1) variables.

The remainder of the thesis is organized as follows. In Section 2, firstly, the previous studies about store clustering will be introduced. Secondly, store performance will be reviewed for the purpose of determining explanatory and response variables that will be used in store clustering application. Finally, a detailed clusterwise linear regression literature will be presented in two main categories as classical methods and mathematical programming methods. Section 3 introduces two mathematical formulations that are applied in this thesis. Section 4 provides two applications of proposed clusterwise linear regression formulations on a dataset from literature and the store clustering problem. Finally, in Section 5 we give conclusions and a few issues that will be considered as future research areas.

2. LITERATURE REVIEW

Three sets of literature are relevant for this thesis: store clustering / segmentation, retail store performance and clusterwise linear regression. First, the literature regarding the store clustering is discussed in detail to provide (1) motivation behind these researches and (2) methods and variables used for clustering and regression. Following this discussion, a brief review of the retail store performance literature is provided to understand both the elements influencing store performance and the performance criteria considered for performance evaluation. Finally, clusterwise linear regression literature will be presented.

2.1 Store Clustering or Segmentation

In the marketing literature, customer segmentation for variety of sectors has been studied by using customer data (demographic data, socioeconomic data, sales data etc.). However, according to our literature review, it was found that studies related to store clustering/segmentation are very rare.

Davies R.L. (1973) considered three different techniques for assessing store attributes and sales performance. In total, 35 variables were collected from 72 different retail stores. These variables were categorized into three different classes: financial and physical characteristics of the stores, locational circumstances and competitive position involved, nature of the urban markets and consumer populations. Firstly, principal components analysis and factor analysis were performed on the 35 variables and results show that only six of these attributes were able to explain %72 of the total variance. Secondly, these six factors (namely trade potential, store characteristics, site and location, character of town, type of consumers, and the level of selling efficiency) were involved in the clustering process. Ward's (1963) technique was used for clustering purpose. Thirdly and finally, the paper performed stepwise multiple regression analysis for forecasting purposes and assessing the impact on sales of a change in the variables. Hawkes and McLaughlin (1994) clustered 78 stores in six supermarket chains across the United States according to the sales potential for different consumer segments. They used geo-demographic techniques for segmenting each supermarket. The aim of their study was to demonstrate that geo-demographic clustering of stores can be used to isolate consumer segments with targeted merchandising programs to increase product-based (margarine) sales.

Kumar and Karande (2000) found that internal and external environment for grocery stores facilitated store performance, but that this varied by the socioeconomic characteristics in which the store was placed. They tested different model forms (linear, exponential, and multiplicative) for explaining the effect of retail environment on sales and sales per square foot and find that the linear form offered the best fit according to the error sums of squares in predicting. They considered both internal environment variables such as the number of checkout counters per 10,000 square foot of selling area, number of nongrocery products sold, whether the store at least doubles manufacturers coupons, whether there is a banking facility, and whether the store is open for 24 hours and external environment variables such as the total number of households in the trade area, and the geographical region where the store is located in their linear model. After they fitted the linear model for whole data and determined the effects of each variables on store performance, they segmented 460 stores using cluster analysis based on the socioeconomic characteristics of the trade area such as percentage of households with annual income greater than $30,000, percentage of households with four or more members, and percentage of households owning their residences. They conducted cluster analysis using Ward’s method. Finally, they analyzed how well the parameter estimates predict sales and sales per square foot for assessing the consistency of the segments. For this purpose, they used the data from 160 stores in the test sample.

Clarke et al. (2003) and Bermingham et al. (2013) clustered existing retail stores to decide on the locations of new stores. In both studies, a special package program which is called as Modelling Intuition in Retail Site Assessment (MIRSA) groups the stores into clusters according to given data. Cognitive mapping techniques were used to extract a set of critical variables that influence stores' sales performance, so that they might contribute to the store clustering process. In addition, the firm and the variables used in the application section were kept confidential. However, the variables were generally explained as demographic characteristics, competition characteristics and economic situation of the region where the store belongs. Their MIRSA program uses the K-means clustering technique for segmentation purposes.

Mendes and Cardoso (2006) clustered existent outlets in order to support evaluation of relative performance of different locations and identification of potential site locations. They divided the variables which are considered in clustering framework into three classes as location and outlet attributes, influence area characterization and client characteristics. Since a small number of supermarket stores (25 stores) were clustered according to a large number of variables (250 variables), experts' knowledge was also considered in the clustering process. They collected data from shop surveys, mystery shopping program and national geographical census. They compared three different approaches; (1) a priori: provides values for perceived dissimilarities between pairs of outlets; (2) a posteriori: evaluates results from multiple regression trees; (3) interactively: helps to select base variables and evaluate results from alternative dendrograms. They employed Ward's technique for clustering purposes in the first and third approach. Therefore, only input variables were taken into account in these applications. However, in the second approach, they considered both input and output variables together via using regression trees. They determined the annual outlet turnover as target variable.

Kargari and Sepehri (2012) separated the stores owned by a company that producing automotive parts into clusters to reduce transportation costs. In this study, K-means technique was used for clustering. The data of the last three years were examined for 815 stores and stores were divided into homogenous groups according to sales of 75 critical products. After stores were assigned to clusters, they proposed an appropriate distribution policies for every clusters. The results of the study indicated significant cost reduction in distribution and transportation costs.

Bilgic et al. (2015) aimed to segment retailers' stores for generating disparate marketing strategies for each segment rather than mass marketing. They analyzed a supermarket chain, which owns 73 stores in Turkey. Firstly, they applied Ward's hierarchical clustering algorithm to assign stores in five clusters by using the R programming language. After, they employed the Apriori algorithm for mining frequent item sets and association rules. For clustering purposes, they used three different groups of variables. The first group was related to store such as store size, whether there is a university, a factory, a trade center, a car park near the store, and competitors close to the store's location. The second group was related to trade area demographics such as age groups, marital status, and educational level. The third group involved average apartment rentals, in order to presume the wealth level of people who live in the trade area.

Rooij (2017) proposed a two-stage clustering framework based on price elasticity to define segments of stores. The aim of the study was to prove that there is an opportunity to increase revenue and profit by segmentation. Price elasticity was used as an indicator of customer behavior. Application of the method was conducted in a supermarket chain which has 305 stores in the Netherlands. They separated framework into two parts as mentioned before, in the first part they recommended K-means++ clustering algorithm to establish clusters of stores for each product group separately. In the second stage, they combined the information comprised in these separate clusterings into a general clustering. The data used in this study were roughly split into two sets: scanner data and store specific characteristics and trading area data. The scanner data included prices for considered products, prices for substitutes and price discounts. The data on store specifics and trading area are composed of competition, socio-demographics, and socio-economic variables. In addition, they evaluated the robustness of the clustering method using the Jaccard dissimilarity measure.

Almohri et al. (2019) recommended a model based clustering technique for segmenting the stores into homogeneous groups for internal benchmarking. It is possible to take into account both internal (e.g. inventory, advertising) and external data sources (e.g., demographics and local competition) in their proposed method. They proposed a Finite Mixture of Regressions technique based on competitive learning and called Mixture

Model with Competitive Learning (MMCL) to their algorithm. For more detail, Finite Mixture of Regression was used for modeling a continuous output as a function of inputs via different linear regression models. In this application, they considered an automotive dealership across the U.S. for a particular original equipment manufacturer (OEM). The data include both variables that represent financial statements of OEMs and demographic information. Although each cluster yielded specific component models depending on certain inputs for each of output variables as a result of MMCL algorithm, they recommended a multi-objective optimization model to derive suggestions for stores that takes into account two different objectives. They split the data into two part as training and test data to find best values of parameters that need to be selected prior to applying their technique.

2.2 Retail Store Performance

The purpose of this part is to determine inputs (explanatory variables) and outputs (response variables) that have been used in literature. In line with this objective, in addition to store clustering literature, past studies about measuring the performance of retail stores, selection of retail store location, evaluation of the customer choices within retail stores and frequency of visits (retail patronage) have been analyzed. Performance measures (outputs) can be classified into two categories: financial or economic outcomes and behavioral outcomes. Table 2.1 summarizes most commonly appeared performance measures in analyzed studies.

Although there are several classifications of factors (explanatory variables) which affect the success of individual stores in the literature (Mendes and Cardoso, 2006; Vyt, 2008; Turhan et al., 2013), we prefer consider three main categories as follows:

- Store related characteristics, such as age of the store, rent of the store, etc. In addition to those features, most of management decisions are made by both senior management and store manager belongs to this category. With the assumption that the same decisions are applied in all stores, these decision variables are often ignored in the literature. However, the increasing competition and customer-oriented management style have necessitated the implementation of store-specific decisions in different stores. Also, employee characteristics play an important role in this category.

- Location characteristics, are split into two sub-categories as environmental and customer characteristics. Customer characteristics have significant effects on retail store performance as mentioned in the past studies (Dhar and Hoch, 1997; Pauler et al., 2009). Moreover, the demographics and socio-economic variables of the people who live in the proximity of the store the are most commonly used factors for clustering purposes in marketing literature (Wedel and Kamakura, 2000). Magnets indicate crowded points like hospital, cinema, school, hotel, etc. Magnets are mentioned as the point of interests (POI's) in the retail industry.
- The product characteristics can be taken into account as the third category of explanatory variables, however, they are ignored most of the studies.

In Table 2.2, the explanatory variables which have been used to explain store performance in the literature are presented.

Table 2.1: Performance measures in the literature

Outputs	
Financial or economic outcomes	**Behavioral outcomes**
Sales or demand volume	Retail patronage or brand loyalty
Profits	Customer satisfaction
Market share	
Number of customers	
Number of visitors	
Price elasticity	
EBIT or EBITDA	

Table 2.3 provides a comprehensive summary of the papers that have analyzed inputs and outputs of the store performance measures. Application areas of the papers are also mentioned in the table.

Table 2.2: Elements influencing store performance

Inputs				
STORE	**LOCATION**			**PRODUCT**
Store characteristics	**Environmental characteristics**	**Customer characteristics**		Product attributes
Store size and age	**Competitive conditions**	**Demographic variables**	**Socio-economic variables**	Design / Style
Number of employees	Numbers and size of competitors			Color
Employee turnover	Distance to competitors	Gender	Income level	Material
Atmospherics	Competition intensity	Age	Total disposable income	Comfort
Number of cash desk	Competitors' sales volume	Education level	Purchasing power index	Durability
Assortments of product	**Magnets (POI)**	Marital status	House ownership	Pricing Strategy
Inventory	Crowd points	Occupation	House value	Promotion
Working hours of store	Culture and education institution	Population	House rental prices	Country of origin
Cost of rent	Relaxation	Population density	Homeowners	Quality
Employee characteristics	Government organization	Population growth rate	Householders	
Hours worked	Vehicle maintenance	Household size	Autos owned	
Education and experience	**Enviromental conditions**	Social classes & subcultures	Unemployment	
Training level	Store location (mall or street)		Employment rate	
Wage rate	Climate	Purchasing habits	Working women rate	
Ethnicity of employee	Holiday	Media usage		
Absenteeism	Parking adequacy	**Psychographic variables**		
Managerial efforts	Pedestrian crossing			
Ownership type	Sidewalk and road width	Traditionalist		
Store service level	Vehicle traffic density	Outgoing/Individualist		
Store layout	Passenger traffic	Quality/Service		
Store promotions	Store visibility	Socially Conscious		
Store advertising	Touristic area			
Employee monitoring	Distance to customers			

Table 2.3: Comprehensive literature review on store performance, store clustering and selection of store location

Authors	**Inputs**	**Outputs**	**Application Area**
Reilly (1931)	Population Distance	Proportion of retail trade from intermediate towns	Theoratical
Huff (1964)	Size of shopping center Consumer's travel time Population	Number of customers	Theoratical
Applebaum (1966)	Distance Purchases per capita	Store sales	Supermarkets
Martin (1967)	**Economic and demographic variables** Population over 65 Social rank Population of minorities Household size Population of white collar workers Household income Home value Population of immigrants **Population variables** Population Population growth **Competitive and penetration variables** Number of competitors Number of customers of competitors	Sales	Branches of banks
Davies (1973)	Gross selling area Rents and rates Distance to nearest major car park The number of other branch stores found in the same town Index of store accessibility	Major product sales	Durable goods company stores

Authors	Inputs		Outputs	Application Area
Cottrell (1972)	Sales area Number of checkout counters Price level Family income Family size Changes in consumer price index for food Number of competitors	Discount competition percentage Population in trade area Square miles in trading area Population density per square mile Neighborhood type (Blue collar, Black or Small town neighborhood) Location type (Shopping center or Street)	Total sales Product group sales Direct expenses Gross margin percentages	Supermarkets
Clawson (1974)	**Population variables** Renter-occupied dwellings S&L savings per capita Income per capita Value of owner-occupied homes Persons age 45-64 Persons age 65 and over **Branch variables** Location of branch (shopping center, business district, etc.) Age of branch Exterior attractiveness Interior decor Parking adequacy Branch advertising and promotion cost	**Competition variables** Competing S&L facilities Population per S&L facility Commercial bank facilities Net savings gain of local S&L competitors Share of total local S&L savings held by local main and executive offices of competitors Total assets of competing S&L associations having local branches	Net savings gain in branch	Saving & Loans association branches

Authors	Inputs		Outputs	Application Area
Bearden et al. (1978)	**Demographic variables** Gender Race Age Residence (Urban or Rural) Total family income **Psychographic Dimensions** Traditionalist Outgoing/Individualist Quality/Service Socially Conscious Other-Directed	**Media Usage** Television Radio Magazine Newspaper Education	Retail patronage	Convenience stores Shopping centers Discount stores Fast food franchises
Ingene and Lusch (1980)	**Enviromental variables** Household income Household size Mobility (automobiles per 1000 households) Density of population (people per square mile)	**Managerial variables** Wage rate Growth rate of population Store area per customer Employees per store area	Household expenditure at stores	Shopping centers
Gautschi (1981)	Assortment of product Center design Price levels Working hours Availability of car park	Travel time Travel and parking costs Travel safety Travel performance	Number of subjects patronizing the stores	Shopping centers

Authors	Inputs		Outputs	Application Area
Ingene (1982)	Store size Wage rate Growth of population Number of competitive stores	Household income Household size Automobiles per household	Sales per employee	Supermarkets
Ghosh and McLafferty (1982)	Store size Distance to trade area		Market share	Supermarkets
Achabal et al. (1982)	Store size Distance		Retail patronage	Theoratical
Arnold et al. (1983)	Locational convenience Lowest prices Friendly staff Pleasant environment	Weekly specials Fastest checkout Assortment	Retail patronage	Supermarkets
Hise et al. (1983)	**<u>Store manager variables</u>** Age Annual income Marital status Number of children Educational level Hours worked per week Experience in retailing Experience in present position Experience with present employer **<u>Location variables</u>** Mall size Market size	**<u>Store variables</u>** Store size Fixed assets Inventory level Years store opened Employees per store **<u>Competitive variables</u>** Primary competitors Secondary competitors	Sales Gross margin Return assets	

Authors	Inputs		Outputs	Application Area
Ghosh and Craig (1983)	Distance Number of competitors Population growth Customer expenditure		Profit	Supermarkets
Ingene (1983)	**Socioeconomic and environmental variables** Household income Automobiles per household Automobiles per square mile (congestion) Household size Percentages of males Percentages of whites Firm concentration ratio (number of stores/trade area)	**Marketing mix variables** Wage rate Store size Number of employees Population growth rate	Sales	Restaurants Grocery stores
Walters and Rinne (1986)	Holiday weeks Double-couponing promotions Loss leader portfolios		Store traffic Grocery sales Gross grocery profits Sales of promoted products Sales of non promoted products	Grocery departments of stores
Banker and Morey (1986)	Total labor cost Other operating costs Average size of inventory Population being served		Number of prescriptions Dollars of sales	Pharmacies

Authors	Inputs	Outputs	Application Area
Ingene and Brown (1987)	**<u>Demographic and environmental variables</u>** Income level Population age Household age Household size Labor force Population density Population growth Number of automobiles per household Unemployment rate **<u>Marketing mix variables</u>** Number of pumps per station Number of employees Wage rate	Sales	Gasoline stations
Walters and MacKenzie (1988)	Promotions Holidays	Sales Store traffic Store profit	Supermarkets
Durvasula et al. (1992)	Competitor's market share Growth rate of competitor's sales Aggressiveness of competitor Age of competitor's branch	Market share	Financial institution branches
Hawkes and McLaughlin (1994)	Consumer demographics Space allocation of products	Major product sales	Supermarkets

Authors	Inputs	Outputs	Application Area
Hoch et al. (1995)	**Consumer characteristics** Population over 60 years of age Population with a college education Percentage of households with five or more members Income Sales price of houses Percentage of women who work Consumers who were black and Hispanic **Competitive characteristics** Distance to nearest competitors Sales volume of nearest competitors	Price elasticity	Supermarkets
Athanassopoulos (1995)	Drinking area Number of covers Population between 20-45 ages Population of social classes Number of competitors	Sales	Restaurants
Athanassopoulos and Ballantine (1995)	Capital employed Fixed assets Number of employees Number of outlets Sales area	Sales	Supermarkets

Authors	Inputs		Outputs	Application Area
Boufounou (1995)	**Location features** Population Number of households Household size Number of firms in manufacturing-handicraft sectors Employment Employment in manufacturing-handicraft sectors Employment in trade sector Number of employees Number of employees in trade-handicraft sectors Number of employees in agriculture-lifestock sectors Number of rentiers Number of pensioners Number of self-employed professionals Number of tax-payers **Competitive situation features** Number of own bank branches Number of commercial bank affiliates' branches Number of competitor banks' branches	**Trade area characteristics** Number of firms Employees' income Income of employees in trade-handicraft sectors Income of employees in agriculture-lifestock sectors Income of self-employed professionals Rentiers' income Pensioners' income **Internal branch characteristics** Branch age Number of branch employees Night deposit facility New look	Volume of savings deposits Volume of sight deposits Volume of time deposits Volume of total deposits	Bank branches

Authors	**Inputs**		**Outputs**	**Application Area**
Kamakura et al. (1996)	Total operational costs Wage rate Man-hours of labor Floor area		Cash deposits Checks and money order Funds in transit Service fees	Bank branches
Dhar and Hoch (1997)	**Consumer factors** Wealthy people in customer base Elderly people in customer base Higher education level in customer base Ethnic composition of customer base **Brand factors** Number of competitive brands Heterogeneity in market shares across brands Promotions	**Retailer factors** Number of competitors Heterogeneity in market shares across competitors Quality of products Promotion strategy Depth of assortment Category expertise Promotion intensity for private label	Market share	Supermarkets
Thomas et al. (1998)	**Labor** Full-time employees per square foot Full-to-Part-time employees Salaries **Location related costs** Occupancy costs Operating expenses Population Households Household income Proximity to nearest company store	**Internal processes** Inventory Average value of transactions Employee turnover **Experience** Employee tenure Store manager tenure Store age	Sales revenue Profit	Supermarkets

Authors	Inputs		Outputs	Application Area
Donthu and Yoo (1998)	Store size Manager experience Store location Promotions expenses		Sales Customer satisfaction	Fast food restaurants
Grewal et al (1999)	Operating expenses Size of store Level of product availability		Sales	Automobile parts retail stores
Reinartz and Kumar (1999)	**<u>Market potential</u>** Households with 4 or more persons Number of households Households with children **<u>Socio-economic status</u>** Household income Households with 3 or more cars Head of household with college education	**<u>Store attractiveness</u>** Grocery Scrambling Non-Grocery Scrambling Open 24 Hours Double Couponing Store New or Renovated	Sales Sales per square feet	Supermarkets
Kumar and Karande (2000)	Number of checkout counters Number of nongrocery products sold Promotions Working hours	Presence of bank Number of households Geographical region information	Sales Productivity (Sales per square feet)	Supermarkets

Authors	Inputs		Outputs	Application Area
Campo et al (2000)	**Store variable** Store size **Category variables** Product space allocation Number of competitive stores	**Location variables** Store size of competitors Areas with young families with child Multiracial areas with single and low income families Areas with upper middle class families Areas with mature adults and single child families Number of people working but not living in the area	Store sales Category share in store sales	Supermarkets
Kuo et al. (2002)	**Competition characteristics** Competition type Number of competitive stores Area of competitive stores **Magnet points** Crowd points Culture and education organizations Relaxation centers Government & business organizations Vehicle maintenance **Convenience characteristics** Parking convenience Pedestrian crossing Sidewalk width Road width	**Availability characteristics** Crowd Stations Bus stop Car flow **Store characteristics** Visibility Store front area **Population characteristics** Population Population density Population growth rate Income level Consumption level	Number of visiting customers	Convenience stores

Authors	Inputs		Outputs	Application Area
Barros and Alves (2003)	Full-time employees Part-time employees Cost of labour Absenteeism Area of outlets	Number of points of sales Age of the outlet Inventory Other costs	Sales Operational results	Supermarkets
Clarke et al. (2003)	Competition Site / store configuration Retail composition Site access	Catchment access Market size Catchment quality	Sales	Retail stores
Keh and Chu (2003)	**Labour variables** Floor staff Management wages Benefits controlling for the number of hours worked **Capital variables** Occupancy Utilities Maintenance and general expenses	**Distribution service variables** Accessibility Assortment Assurance of product delivery Availability of information Ambience	Sales	Supermarkets

Authors	Inputs	Outputs	Application Area
Gijsbrechts et al. (2003)	**<u>Location factor</u>** Age Low income Employment Single person households Competitive strength Store size Transient shoppers **<u>Store flyer characteristics</u>** Flyer size Discount size Food promotions space in flyer Private label promotions space in flyer Special promotions on cover page	Store sales Store traffic	Supermarkets
Barros and Alves (2004)	Number of full-time employees Cost of labour Number of cash-out points Stock Other costs	Sales Operational results	Supermarkets
Korhonen and Syrjänen (2004)	Labor hours Store size	Sales Store Profit	Supermarkets

Authors	Inputs	Outputs	Application Area
Campo and Gijsbrechts (2004)	**Category variables** Category share in store area **Local competition** Competitive stores' area Number of stores selling the same products **Additional covariates** Number of people working but not living in the area Degree of urbanization Number of households Number of own chain stores **Store variables** Store area Total attraction of store Total attraction of other outlets of chain in trading zone **Population characteristics** Areas with young families with children Multiracial areas with single and low income families Areas with upper middle class families Areas with mature adults and single child families	Store sales Category share in store sales	Supermarkets
Donthu et al. (2005)	Advertising and promotions expenses Manager experience Number of employees	Store sales Customer satisfaction	Fast food outlets
Barros (2005)	Price of labour Price of capital Population of trade area Area of competitor's outlets Purchasing power in trade area Rate of temporary workers in the total work-force Staff absenteeism	Sales Earnings before taxes	Supermarkets

Authors	Inputs	Outputs	Application Area
Mendes and Cardoso (2006)	**Location and outlet attributes** Sales area Retail composition Chain image / services Site accesibility Site configuration **Client characteristics** Average buy Client preferences Client demography Buying power **Influence area characterization** Competition sales area Competition quality Income distribution Demographic data	Store turnover	Supermarkets
Sellers-Rubio and Mas-Ruiz (2006)	Employees Outlets Capital	Sales Profits	Supermarkets
Barros (2006)	Number of employees Capital (value of assets) Market share Number of outlets Ownership Regulation Location	Sales Operational results Value added	Supermarkets
De Mateo et al. (2006)	Salesperson labour Cashier labour Sales and administration expenses Marketing expenses Store floor surface	Gross sales	Department stores
Barth (2007)	Labour hours Liters of inventory depletion	Sales	Wine stores

Authors	Inputs		Outputs	Application Area
Vyt (2008)	**Store variables** Sales area Number of full-time employees Space allocated to category product **Competition variables** Competition index for supermarkets Competition index for convenience stores Competition index for hard discounters	**Socio-demographics variables** Percentage of seniors Unemployment rate Percentage of second homes Percentage of individual houses Number of inhabitants	Sales for liquid products	Supermarkets
Pauler et al. (2009)	**Store features** Sales area Presence of ATM machines Presence of banking office Selling beer Presence of film developing Presence of film lab Presence of floral center Service delivery Presence of food service Store remodeled Presence of everyday low price In store bakery Presence of restaurant snack bar Service pharmacy Located within a plaza	**Socio-demographic features** Population density Household size Income Education Median age Competitive threat index	Sales Market share	Supermarkets

Authors	Inputs		Outputs	Application Area
Mishra (2009)	Value of stock Recurrent costs mainly in the form of wages Floor space		Sales Customer satisfaction	Confidential information
Camanho et al. (2009)	Stock Operational costs Staff costs Products spoiled	Area Population Competition	Sales	Supermarkets and hypermarkets
Joo et al. (2009)	**Model 1** Cost of sales Wages and benefits Other expenses Occupancy expenses	**Model 2** Cost of sales Wages Other expenses	**Model 1** Total sales **Model 2** Restaurant sales Retail sales	Coffee stores
Banker et al. (2010)	Total selling hours Store size Average inventory Support activities Monitoring index (number of managers / number of employees) Household income	Age of population Family size Population with college education Population Urban or rural area Competition intensity	Store sales	High-end retail outlets
Sharma and Choudhary (2010)	Store area Experience of manager Location of store		Sales Customer satisfaction	Department stores
Gupta and Mittal (2010)	Total square feet area of store Total number of stock keeping units Number of point of sales machines	Labour cost/wages of employees Number of employees Working hours of employees	Sales Customer conversion ratio (no. of transactions/customer traffic)	Grocery retail stores

Authors	Inputs	Outputs	Application Area
Vaz et al. (2010)	Floor area of section Value of products in stock Diversity of products available in section Value of products stolen, spoiled or whose validity expired	Section sales	Supermarket sections
Sellers and Nicolau (2011)	**<u>Internal factors</u>** Employees Selling area Points of sale **<u>Environmental factors</u>** Degree of competition Population of city Purchasing power index	Sales revenue	Supermarkets
Li and Liu (2012)	Store accessibility (distance) Store attractiveness (store size) Demographic characteristics of customers Size and number of agglomeration stores Size and number of competition stores	Sales	Discount stores
Kargari and Sepehri (2012)	Distance between stores Order concurrency Lot size	-	Spare parts distribution stores
Gauri (2013)	**<u>Format strategy</u>** Supermarket Supercenter Limited assortment **<u>Exogenous factors</u>** Household size Median age Average income Population per square mile Distance to competition Number of competitors **<u>Store specific factors</u>** Selling area Check-out counters Number of employees Number of store features **<u>Pricing strategy</u>** High-Low pricing Everyday low pricing Hybrid pricing	Sales	Grocery selling stores

Authors	Inputs		Outputs	Application Area
Uyar et al. (2013)	Store area Population Inventory value Number of employees Total salaries of employees	Operating expenses Education of store manager Experience of store manager Experience of staff Age of store	Sales Profit	Bookshops
Mishra and Ansari (2013)	**<u>Input factors</u>** Labour Capital (infrastructure) Retail merchandise Store interiors Systems and processes IT and point of sales	**<u>Control variables</u>** Type of retail store Size of retail store Characteristics of consumers Income level of consumers	Quantity of fashion category products in store Quantity of core category products in store Quantity of accessory products in store Value added to products	Stores in apparel and lifestyle sector
Xavier el al. (2015)	Size of stores Investments Personnel costs		Sales EBITDA	Women's clothing stores
Xavier et al. (2015)	Number of workers Years of experience of store manager Purchasing power per capita index	Employees average level of education Population density Whether store located in shopping center or urban streets	Technical efficiency (Sales volume & EBIT / Rent costs & Salaries and wages & Investments in assets)	Fast fashion retail stores
Bilgic et al. (2015)	Store size Average apartment rental Number of competitor stores Population by marital status Population by age groups Population by completed education level	Presence of university in trade area Presence of factory in trade area Presence of car park Presence of bus service Whether trade area is touristic/business or not	-	Supermarkets

Authors	Inputs	Outputs	Application Area
Trivedi et al. (2016)	**<u>Promotional input variables</u>** Feature Average discount Proportion of SKUs on discount Proportion of high price tier SKUs on deep discount Proportion of private label SKUs on discount Dummy variable (any of top 3 SKUs on deep discount) **<u>Category characteristics</u>** Number of SKUs Sales share Interpurchase time Average purchase amount Stockpilability Impulse **<u>Control variables</u>** Relative advertising Temperature Holiday **<u>Store characteristics</u>** Total selling area Number of employees Service level Store age Distance to nearest competitor Household size Per-capita income	Sales of category	Grocery stores
Kahraman et al. (2018)	Size of land Number of employees Number of deliveries Total cost	Number of customers Sales Store evaluation of customers	Supermarket chain stores
Chang et al. (2018)	Area Labor cost Rent cost	Total sales revenue Gross margin Net profit	Fast food stores
Almohri et al. (2019)	Key performance indicators	Sales Profits	Automotive dealerships

2.3 Clusterwise Linear Regression

Clusterwise linear regression first appeared in the literature in the late 1970s. In accordance with the thesis subject, the literature of CLR is divided into two subheadings. Primarily, the first studies of CLR and model-based methods CLR will be reviewed. Secondly, because of the aim of the thesis is to introduce several mathematical programming methodology, the literature about the mathematical programming methods for solving CLR problems will be reviewed in detail.

2.3.1 Clusterwise Linear Regression Problem and Classical Methods

Before the literature review, we give the notation used in mathematical formulations in the literature review as follows. Any additional notation will be listed separately with the respective formulation.

i	the entity subscript ranged from $1\ to\ I$
I	the number of entities
k	the entity subscript ranged from $1\ to\ K$
K	the number of explanatory variables
q	the clusters subscript ranged from $1\ to\ Q$
Q	the number of clusters
C_q	q^{th} cluster
y_i	the value of response variable for entity i
x_{ik}	the value of the k-th explanatory variable for entity i
β_{qk}	the k-th regression coefficient for cluster q
β_{q0}	the intercept coefficient for cluster q

The problem and terminology of clusterwise linear regression first appeared in the literature with the work of Späth (1979, 1981). Späth created an exchange algorithm that is conceptually similar to the popular K-means algorithm to implement on the CLR problem. Späth's exchange algorithm tries to find a given number Q of clusters for entities $i = 1, \dots, I$, such that $C_q \cap C_k = \emptyset\ q \neq k$, $C_1 \cup \dots \cup C_Q = I'$, where $I' = \{1, \dots, I\}$

and corresponding regression coefficients $(\beta_{q1}, \dots, \beta_{qK})$ such that the sum of the residuals sums of squares over all clusters is minimized:

$$min \sum_{q=1}^{Q} \min_{\beta_q} \sum_{i \in C_q} (y_i - \beta_{q0} - \sum_{k=1}^{K} \beta_{qk} x_{ik})^2 \quad (2.3.1.1)$$

Späth reported that this formulation is plausible when the number of entities is relatively large as compared to the number of explanatory variables and when the entities derived from diversified groups. Furthermore, in order to have a solution for (2.3.1.1), the number of entities in each cluster should be greater than or equal to the number of explanatory variables $|C_q| \geq K$. The procedure of Späth's exchange algorithm which is stepwise optimal but not globally optimal can be summarized as follows:

Step 1: Choose a random initial partition that is feasible. Set i := 0.
Step 2: Set $i \coloneqq i + 1$ and reset $i \coloneqq 1$ if $i > I$. For $i \in C_q$ and $|C_q| \geq K$ examine whether there are clusters C_p with $p \neq q$ such that replacement entity i from C_q to C_p improves the objective function. If this criterion is satisfied, choose C_r such that the reduction becomes maximal and redefine $C_q \coloneqq C_q - \{i\}$ and $C_r \coloneqq C_r \cup \{i\}$. Otherwise return step 2.
Step 3 : Repeat step 2 as long as an improvement in the objective function is observed, else stop.

Späth indicated that final result of the exchange algorithm is highly dependent on the initial partition. Therefore, it is recommended to run the algorithm with multiple starting partitions.

Späth (1982) also upgraded the exchange algorithm to an incremental version, which allows to add or remove new entities into the dataset. Späth (1986) and Meier (1987) customized the procedure to clusterwise linear least absolute deviations regression:

$$min \sum_{q=1}^{Q} \min_{\beta_q} \sum_{i \in C_q} \left| y_i - \beta_{q0} - \sum_{k=1}^{K} \beta_{qk} x_{ik} \right| \quad (2.3.1.2)$$

Meier indicated that L^1-norm minimization in (2.3.1.2) is often superior to L^2-norm minimization in (2.3.1.1), because of the robustness properties (extreme values do not affect the parameters so much).

Desarbo and Cron (1988) proposed a maximum likelihood methodology for fuzzy clusterwise regression. The authors assumed that the dependent variable, y_i, is distributed as a finite sum or mixture of conditional univariate normal densities:

$$y_i \sim \sum_{q=1}^{Q} \lambda_q f_{iq}(y_i | x_{ik}, \sigma_q^2, \beta_{qk}) \tag{2.3.1.3}$$

$$= \sum_{q=1}^{Q} \lambda_q (2\pi\sigma_q^2)^{-1/2} \; exp\left[\frac{-(y_i - x_{ik}\beta_{qk})^2}{2\sigma_q^2}\right] \tag{2.3.1.4}$$

where:

λ_q the mixing proportion for the q-th cluster

σ_q^2 the variance term for the q-th cluster

Given a dataset of I independent entities, Q, y, and x, the authors found the estimates of λ_q, σ_q^2, and β_{kq} using the Expectation-Maximization (E-M) algorithm, where their problem is given by

$$ln\,L = \sum_{i=1}^{I} ln\left[\sum_{q=1}^{Q} \lambda_q (2\pi\sigma_q^2)^{-1/2} exp\left[\frac{-(y_i - x_{ik}\beta_{qk})^2}{2\sigma_q^2}\right]\right] \tag{2.3.1.5}$$

subject to:

$$0 \le \lambda_q \le 1 \tag{2.3.1.5a}$$

$$\sum_{q=1}^{Q} \lambda_q = 1 \tag{2.3.1.5b}$$

$$\sigma_q^2 > 0 \tag{2.3.1.5c}$$

Their EM algorithm, however, requires the knowledge of the distribution of the dependent variable, and hence it cannot be applied if there are only observations of the dependent variable. Authors assigned fuzzy cluster membership parameters via posterior probabilities, using Bayes' theorem conditioned on the estimates of λ_q, σ_q^2, and β_{kq} which are obtained within any iteration. Finally, in order to select the most appropriate number of clusters, Q, they employed the Akaike information criterion

(AIC) (Akeike, 1974). They applied the methodology in both a Monte Carlo analysis and a real situation about evaluation of marketing managers' trade show performance.

DeSarbo et al. (1989) introduced an alternative approach for clusterwise linear regression. A modified simulated annealing (SA) procedure is utilized for clustering entities into groups and estimates corresponding regression coefficients simultaneously. The SA metaheuristic starts from a random and feasible initial partition and recursively specifies steps in a random direction in the possible solution space. The new solution is accepted if it develops the criterion; if not, it is rejected with a probability proportional to the increase of the criterion value. Furthermore, their model deals with replicated observations per entity, multiple dependent variables and overlapping clusters. The methodology proposed in this paper is applied to determine consumer satisfaction determinants.

Wedel and Kistemaker (1989) introduced a generalization of clusterwise linear regression for benefit segmentation. This model is suited for finding customer clusters especially if collinearity plays a role in fitting preference models at the individual level. Their model is also able to handle more than one observation per entity and multiple dependent variables similar to the study of DeSarbo et al. (1989). Their algorithm shows some similarities to Späth's exchange algorithm. However, their procedure has an advantage over previous studies, as it can also be applied in circumstances when the number of observations for each subject is smaller than the number of explanatory variables.

Wedel and Steenkamp (1989, 1991) developed a fuzzy clusterwise linear regression method that estimates corresponding preferences to products within each number of segments and the degree of membership of subjects in those segments. Therefore, in addition to previous studies, their model handles partial membership of entities in clusters.

DeSarbo and Edwards (1996) modified the study of Desarbo and Cron (1988) by adding a new constraint to the model. Because of the hypothesized direction of the positive relation between each independent variable and dependent variable known, they added the following constraint to the formulation (2.3.1.5):

$$\beta_{qk} \geq 0 \tag{2.3.1.5d}$$

Hennig (1997, 1998, 2002, 2003) proposed an alternative stochastic approach called Fixed Point Cluster (FPC) analysis which is applied to clusterwise linear regression. In contrast to previous concepts, the FPC analysis searches for a single linear regression cluster at a time. All the remaining data are assumed to be composed of outliers with respect to the cluster. Therefore, not every data point needs to belong to linear regression clusters in FPC analysis. According to that, FPC analysis is not a new method for estimating mixture parameters, but it provides an alternative description of a cluster. Moreover, the other advantage of this framework is that the number of clusters does not need to be identified in advance. Hennig (2000), in addition to his works about CLR, investigated the identifiability of the parameters of clusterwise regression models which was described as being difficult and partially identifiable.

Aurifeille (2000) applied a genetic algorithm (GA), which combines both random changes and crossovers of the better solutions in a way similar to mechanisms of genetics to the problem of clusterwise linear regression.

Zhang (2003) presented an algorithm for solving the clusterwise linear regression problem using the K-harmonic means clustering algorithm. The aim of this paper was to eliminate the dependency of algorithms (such as Späth's algorithm) to the initial clustering. For this purpose, the paper employed the K-harmonic algorithm which is less sensitive to the initial clusters.

Brusco et al. (2003) proposed multi-criterion clusterwise linear regression (MCLR) model that can be applied for the joint segmentation problem. The recommended model considers both homogeneity of clusters and multiple dependent variables in a weighted objective function. Simulated annealing heuristic was used to provide solutions. They also implemented the MCLR to a local exchange provider which sought to segment a potential long-distance market in anticipation of new market entry.

Preda and Saporta (2005) offered an interesting alternative to classical methods of clusterwise linear regression analysis. They recommended using the partial least square regression estimators for regression coefficients of each cluster in the particular case where the set of independent variables forms an L_2-continuous stochastic process. Furthermore, the Partial Least Squares (PLS) regression method is particularly adapted to solve multicollinearity. Additionally, it is adapted to solve the CLR problem when the number of entities per cluster is smaller than the number of explanatory variables. For evaluation of the method, an application on stock-exchange data was performed.

Caporossi and Hansen (2007) recommended the use of the Variable Neighborhood Search (VNS) metaheuristics to get a better solution for the clusterwise linear regression problem. They demonstrated that the VNS performs better than multi-start Späth (1979) algorithm via both the dataset created by Monte Carlo simulation and a real dataset from the literature.

Luo and Chou (2006) developed a fuzzy clusterwise linear regression methodology where the membership of an entity depends on the proportional relative distance to a particular cluster. Specifically, an entity has a larger membership to a closer cluster, but a smaller membership to a further cluster. It is assumed that an entity can belong to more than one cluster. This simplification called soft clustering in literature which the membership parameters are not forced to be 0 or 1 when the objective function is optimized. Therefore, this simplification makes it easy to solve the clusterwise linear regression model. A possible application of the modified model is also proposed to predict the pavement condition in the future on the basis of the present conditions. The application showed that the new model's results are more accurate predictions than the ordinary least squares regression method which is most widely used for the prediction of the pavement condition.

Brusco et al. (2008) reported that the objective of the clusterwise linear regression which is minimization of the sum of squares of residuals or minimization of the sum of absolute residuals for the within-cluster regression makes no effort to separate the error explained by clustering from the error explained by the regression. Therefore, to guard against the misuse of CLR, they divide the variation of response variable into two part as: variation explained by within-cluster regression models and variation explained by clustering process. Subsequently, they propose a benchmarking procedure that compares the observed response variable with the obtained response from using new developed CLR algorithm that considers both within cluster variation and between cluster variation.

Poggi and Portier (2011) used clusterwise linear regression model for forecasting of the daily mean of particles whose diameter is less than 10mm (PM10) atmosphere. They estimate model parameters for each cluster using the E-M algorithm. Clustering is done by using maximum a posteriori principle.

Hsu (2015) employed a mixture model clusterwise regression for modeling building energy consumption. Furthermore, he used two-stage framework which includes K-means clustering for separating buildings in similar groups in the first stage and ordinary least squares for estimating regression model parameters in the second stage. He prefered to use Bayesian information criterion (BIC) for selecting the best fitting mixture model for clusterwise regression. In addition, for evaluation of the model results, the author measured the stability of clusters by using the Jaccard coefficient, and evaluate predictive accuracy by using 20-fold cross validation. These results show that there seems to be an inherent tradeoff between cluster stability and predictive accuracy. Another crucial consequence of his study is that clusterwise regression provides extremely accurate predictions but unstable clusters, while the two-stage K-means clustering derives more stable clusters.

Di Mari et al. (2017) proposed a fully data-dependent soft constrained method for maximum likelihood estimation of clusterwise linear regression. Specifically, they recommended a method that incorporates constraints on the eigenvalues of the component covariances of Gaussian mixtures that are tuned on the data to clusterwise linear regression.

2.3.2 Mathematical Programming Methods

Lau et al. (1999) formulated the clusterwise linear regression problem as a mathematical programming problem for both parametric and nonparametric procedure. They also stated that the CLR problem is a hard combinatorial optimization problem. For the parametric procedure, they imposed a distributional assumption on the density function of the error terms and maximized the log-likelihood function to determine the clusters and the regression parameters of two clusters simultaneously, which resulted in a nonlinear programming model with linear constraints. On the other hand, for the nonparametric procedure, when the distribution of the error terms is not known, the objective function becomes minimization of sum of squared errors. Because the second approach is more related to this thesis, we provide the mathematical model in detail as follows:

$$min \sum_{i=1}^{I} \sum_{i \in C_q} z_{iq} \left(y_i - \beta_{q0} - \sum_{k=1}^{K} \beta_{qk} x_{ik}\right)^2 \tag{2.3.2.1}$$

subject to

$$\sum_{q=1}^{Q} z_{iq} = 1 \quad \forall i \tag{2.3.2.1a}$$

$$z_{iq} \in \{0,1\} \quad \forall (i,q) \tag{2.3.2.1b}$$

$$\beta_{qk}\ free\ \forall (q,k) \tag{2.3.2.1c}$$

where:

z_{iq} the binary variable that equals 1 if and only if entity i belongs to cluster q

We provide a general formulation of their model. However, their analysis was restricted to a single dependent variable and two clusters.

Bertsimas and Shioda (2007) considered a single dependent variable. They formulated CLR problem as mixed-integer linear programming through a big-M formulation, which means the minimization of the total absolute residuals over all clusters subject to the constraints that each entity belongs to exactly one cluster and the big-M constraints; the latter constraints provide the values of the residuals, and they are activate only if the entity belongs to that cluster through 0-1 assignment variables. Constraint (2.3.2.2c) requires that each entity be assigned to one cluster. Their mixed-integer optimization model is as follows:

$$min \sum_{i=1}^{I} t_i \tag{2.3.2.2}$$

subject to

$$t_i \geq \left(y_i - \beta_{q0} - \sum_{k=1}^{K} \beta_{qk} x_{ik}\right) - M\left(1 - z_{iq}\right) \; \forall (i, q, k) \tag{2.3.2.2a}$$

$$t_i \geq -\left(y_i - \beta_{q0} - \sum_{k=1}^{K} \beta_{qk} x_{ik}\right) - M\left(1 - z_{iq}\right) \; \forall (i, q, k) \tag{2.3.2.2b}$$

$$\sum_{q=1}^{Q} z_{iq} = 1 \; \forall i \tag{2.3.2.2c}$$

$$z_{iq} \in \{0,1\} \;\; \forall (i, q) \tag{2.3.2.2d}$$

$$t_i \geq 0 \;\; \forall i \tag{2.3.2.2e}$$

where:

t_i the absolute error for the corresponding entity

Because of the huge number of binary variables arising from memberships of data points to clusters, they provided a simple heuristic method which applies a nearest-neighbor clustering algorithm in the combined (x, y) space to form clusters, and then they provided a similar mixed-integer linear optimization problem which determines the clusters to be merged and the resulting regression coefficients; because the predetermined number of clusters is very small compared to the number of data points, the latter problem can be solved in a reasonable amount of time. They further provided polyhedral set representations of clusters by solving two linear optimization problems.

Carbonneau et al. (2011) showed numerically that the big-M formulation can sometimes result in non-optimal solutions, and they proposed a mixed logical quadratic optimization problem, which eliminates the big-M constraints from the formulation. This is achieved by activating the constraints on residuals only if the entity belongs to that cluster through logical constraints. The logical propositions stand in their raw formulation while they also benefit from both the logic processing and the power of linear or quadratic programming. The commercial CPLEX software was used to optimize both of these models. In conclusion, the mixed logical quadratic programming (MLQP) formulation is observed to be numerically stable and guaranteed to provide exact global optimal solutions for the CLR problem.

Carbonneau et al. (2012) proposed a new method based on branch and bound algorithm with iterative heuristics, observation sequencing and ending subset optimization to solve the mixed logical quadratic optimization problem in Carbonneau et al. (2011). Although this proposed approach provides significant performance against to CPLEX for problems with a few clusters, it becomes quickly intractable as the number of clusters rises. Therefore, Carbonneau et al. (2014) proposed a column generation based approach coupled with several efficient heuristic strategies, which were used to insert new columns into the restricted master problem and to optimize with incrementally larger ending subsets. In other words, the column generation divides the original CLR problem into two parts. The master problem which is a binary set partitioning problem finds the set of clusters (columns) that minimize the error while considering the predefined number of clusters and the constraint that an entity must be in only one cluster at a time. The subproblem generates a column (cluster) that can enhance the master problem's solution or it demonstrates that it cannot generate an improving column, thus providing the master problem is globally optimized.

Zhu et al. (2012) formulated the CLR problem with the least sum of absolute deviations. They reformulated the Bertsimas and Shioda (2007)'s model which is including big-M constraints for integrating outlier detection into the CLR analysis framework. They tested the proposed approach on both generated instances and real instances using the CPLEX solver with default settings.

Bagirov et al. (2013) formulated CLR problems as the minimum of Q quadratic functions which were given by the sum of squared residuals, and Q was the predetermined number of clusters. Their problem is known to be non-smooth and non-convex in general. To solve this problem, they proposed an incremental algorithm, which starts with one linear function and summarizes the underlying structure of the data by dynamically adding one linear function at each iteration. Bagirov et al. (2015a) enhanced the algorithm in Bagirov et al. (2013) with the use of the discrete gradient method. Furthermore, Bagirov et al. (2015b) improved the algorithm in Bagirov et al. (2013) by approximating the originally non-smooth functions through smooth ones, which enabled them to use very efficient smooth optimization algorithms. Finally, Bagirov and Ugon (2018) reformulated the problem in Bagirov et al. (2013) as the differences of convex functions, and hence obtained a DC representation of the problem. They derived optimality conditions and proposed an algorithm, which exploits this DC representation to find stationary points of the CLR problem.

Bagirov et al. (2017) examined the application of the CLR method introduced in Bagirov et al. (2015a) for monthly rainfall prediction in different locations in Australia. In addition, they compared the performance of the proposed method with the CLR using the maximum likelihood framework by the E-M algorithm, multiple linear regression, artificial neural networks and the support vector machines for regression models. The results showed that the proposed algorithm outperforms other methods in most locations.

Park et al. (2017) considered the CLR problem where each entity has more than one observation and called their problem generalized CLR (GCLR). They proposed both a mixed-integer quadratic program (MIQP) formulation and a set partitioning formulation for GCLR. They formulated MIQP formulation for GCLR as follows:

$$min \sum_{i=1}^{I} \sum_{r}^{R} t_{ir}^2 \tag{2.3.2.3}$$

subject to

$$t_{ir} - \left(y_{ir} - \beta_{q0} - \sum_{k=1}^{K} \beta_{qk} x_{ikr}\right) + M\left(1 - z_{iq}\right) \geq 0 \; \forall (i,q,r) \tag{2.3.2.3a}$$

$$t_{ir} + \left(y_{ir} - \beta_{q0} - \sum_{k=1}^{K} \beta_{qk} x_{ikr}\right) + M\left(1 - z_{iq}\right) \geq 0 \; \forall (i,q,r) \tag{2.3.2.3b}$$

$$\sum_{q=1}^{Q} z_{iq} = 1 \;\; \forall i \tag{2.3.2.3c}$$

$$\sum_{i=1}^{I} z_{iq} \geq n \;\; \forall q \tag{2.3.2.3d}$$

$$z_{iq} \in \{0,1\} \;\; \forall (i,q) \tag{2.3.2.3e}$$

$$t_{ir} \geq 0 \;\; \forall (i,r) \tag{2.3.2.3f}$$

$$\beta_{qk} \; free \;\; \forall (q,k) \tag{2.3.2.3g}$$

where:

r observations, $r = 1, \ldots, R$

The (2.3.2.3a) and (2.3.2.3b) constraints are, also known as big-M constraints, make t_{ir} equal to absolute residual for corresponding observation y_{ir} in the optimal solution when they are active. Constraint (2.3.2.3c) requires that every entity be assigned to one cluster, and (2.3.2.3d) imposes the limit on the cardinality of each cluster. Consequently, their MIQP formulation is more general than one proposed by Bertsimas and Shioda (2007) because they considered more than one observation per entity. They also proved NP-hardness of the generalized CLR problem.

The notation in their set-partitioning formulation is ψ, denote the set of all clusters on entities with the cardinality equal to or greater than i, c_S denote the cost of clusters S, which is equal to the sum of squared residuals when performing the regression over cluster S and a_{iS} equal one if entity i belongs to cluster S. In addition, z_S equal to 1 if cluster S is selected. They proposed the master set-partitioning problem as follows:

$$min \sum_{S \in \psi} c_S z_S \tag{2.3.2.4}$$

subject to

$$\sum_{S \in \psi} z_S = Q \tag{2.3.2.4a}$$

$$\sum_{S \in \psi} a_{iS} z_S = 1 \; \forall i \tag{2.3.2.4b}$$

$$z_S \in \{0, 1\}, S \in \psi. \tag{2.3.2.4c}$$

Their pricing problem can be stated as follows. Let w and π_i be the dual variable for constraints (2.3.2.4a) and (2.3.2.4b), respectively. In addition, z_i equal to 1 if $i \epsilon S$.

$$min \sum_{i=1}^{I} \sum_{r}^{R} t_{ir}^2 - \sum_{i=1}^{I} \pi_i - z_i \tag{2.3.2.5}$$

subject to

$$t_{ir} - \left(y_{ir} - \beta_{q0} - \sum_{k=1}^{K} \beta_k x_{ikr}\right) + M(1 - z_i) \; \geq 0 \; \forall (i, r) \tag{2.3.2.5a}$$

$$t_{ir} + \left(y_{ir} - \beta_{q0} - \sum_{k=1}^{K} \beta_k x_{ikr}\right) + M(1 - z_i) \; \geq 0 \; \forall (i, r) \tag{2.3.2.5b}$$

$$\sum_{i=1}^{I} z_i \geq n \tag{2.3.2.5c}$$

$$z_i \in \{0,1\} \ \ \forall i \tag{2.3.2.5d}$$

$$t_{ir} \geq 0 \ \ \forall (i,r) \tag{2.3.2.5e}$$

$$\beta_k \, free \ \ \forall k \tag{2.3.2.5f}$$

They employed the column generation based heuristic for large-scale problems, a genetic algorithm with Lloyd’s (1982) algorithm, a two-stage heuristic method, and a modified Späth (1979) algorithm. They further compared these methods over on synthetic and real-world data for clustering SKUs. Consequently, they reported that the genetic algorithm with Lloyd’s algorithm provides a good balance between solution quality and solution time compared to the other algorithms.

Khadka et al. (2017, 2018) proposed a mathematical programming framework within the CLR approach to determine simultaneously optimum number of clusters, cluster memberships, cluster-specific independent variables, and regression coefficients. Bayesian information criterion (BIC) (Schwarz 1978) was used as the objective function to achieve models that balance the goodness of fit and complexity. They proposed an algorithm based on simulated annealing (SA) with all subset regression to solve the problem. This algorithm can manage to identify and adress multicollinearity issues between the independent variables. They applied the proposed method for modeling of pavement performance. They showed that the models provide few prediction errors without any overfitting issue.

3. METHODOLOGY

There could be multiple correlated responses for which linear regression models have to be fitted simultaneously, for a single entity, in some real world problems. For example, suppose that I entities $(i = 1, \ldots, I)$ are customers, J responses $(y_{i1}, \ldots, y_{iJ})$ are the amounts of money spent on the J products by customer i, and the K components of x_i $(x_{i1}, \ldots, x_{iK})$ represent the characteristics of both customer i and the J products. These products can be unrelated (i.e., neither complementary nor substitute), yet the components in $(y_{i1}, \ldots, y_{iJ})$ are in general negatively correlated because customer i has a limited budget. In such a case, the J linear regression models have to be fitted simultaneously to obtain a partition of customers that minimizes the total sum of squared or absolute residuals over Q clusters.

In general, to partition I entities with respect to J responses, one can apply the CLR by accounting for a single response each time, and then repeat the same approach for all responses; however, such clustering usually results in a different cluster for the same entity, i, so that it is not known which cluster the entity i belongs to when all J responses are considered simultaneously (Angun and Altınoy, 2019).

In this section, firstly we mention about determining number Q of clusters. Secondly, we present a general mathematical formulation for the GCLR problem which considers multiple responses and fits regression models simultaneously. Furthermore, we propose a new mixed-integer optimization formulation, which overcomes the difficulties of the previous formulations due to the indicator (big-M) constraints.

3.1 Mathematical Model Formulations

Suppose that we have data of the form $\{x_i, y_i\}$ for entities $i = 1, \ldots, I$, where x_i and y_i are $R_i \; x \; K$ and $R_i \; x \; J$ matrices of explanatory variables and multiple correlated response variables, respectively; i.e.,

$$x_i = \begin{bmatrix} x_{i11} & \cdots & x_{i1K} \\ \cdots & \cdots & \cdots \\ x_{iR_i1} & \cdots & x_{iR_iK} \end{bmatrix}, y_i = \begin{bmatrix} y_{i11} & \cdots & y_{i1J} \\ \cdots & \cdots & \cdots \\ y_{iR_i1} & \cdots & y_{iR_iJ} \end{bmatrix}$$

where R_i is the number of replications for entity i, K is the number of explanatory variables, and J is the number of response variables. Given the number Q of clusters, our aim is to find a partition of I entities into Q clusters such that $C_q \cap C_{q'} = \emptyset$ for all $q, q = 1, \ldots, Q$ and $q \neq q'$, and $\cup_{q=1}^{Q} C_q = \{1, \ldots, I\}$, where C_q is the qth cluster. This partitioning is to be found by minimizing p-norm of residuals, which are defined by $t_{ijr_i} = y_{ir_ij} - \beta_{j0} - \beta_j^T x_{ir_i}$ for all $i \in C_q$, where y_{ir_ij} is the r_ith observation of the jth dependent variable in the matrix y_i for $r_i = 1, \ldots, R_i$ and $j = 1, \ldots, J$, x_{ir_i} is the r_ith row of the matrix x_i, β_{j0} and $\beta_j = (\beta_{j1}, \ldots, \beta_{jk})^T$ are the unknown coefficients of the regression model for response j, and T is the transpose of a vector or a matrix.

3.1.1 Big-M Formulation

Before presenting our model, we give below the notation used in the mathematical formulation:

Q	given number of clusters
y_{ir_ij}	r_ith observation of response j for entity i, $r_i = 1, \ldots, R_i$, $j = 1, \ldots, J$; i.e., (r_i, j)th entry of the matrix y_i
x_{ir_ik}	r_ith observation of independent variable k for entity i, $k = 1, \ldots, K$; i.e.,(r_i, k)th entry of the matrix x_i
β_{qjk}	regression coefficient of independent variable k for response j and cluster $q, q = 1, \ldots, Q$

β_{qj0}	intercept of the regression model for response j and cluster q
t_{qijr_i}	r_ith absolute residual of response j, entity i, and cluster q
t_q	sum of absolute residuals over a cluster q; $t_q = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{r_i=1}^{R_i} t_{qijr_i}$
t	Q-dimensional vector of the total absolute residuals; i.e., $t = (t_1, \dots, t_Q)$
z_{qi}	decision variable which shows that entity i belongs to cluster q if $z_{qi} = 1$, and $z_{qi} = 0$ otherwise
M	A big positive number

A slightly more general form of the GCLR problem for multiple responses can be formulated as follows:

$$min\|t\|_p \tag{3.1.1.1}$$

subject to

$$t_{qijr_i} - [y_{ir_ij} - (\beta_{qj0} + (\textstyle\sum_{k=1}^{K} \beta_{qjk} x_{ir_ik})] + M(1 - z_{qi}) \geq 0 \quad \forall(i,q,j,r_i) \tag{3.1.1.1a}$$

$$t_{qijr_i} + [y_{ir_ij} - (\beta_{qj0} + (\textstyle\sum_{k=1}^{K} \beta_{qjk} x_{ir_ik})] + M(1 - z_{qi}) \geq 0 \quad \forall(i,q,j,r_i) \tag{3.1.1.1b}$$

$$\textstyle\sum_{q=1}^{Q} z_{qi} = 1 \ \forall i \tag{3.1.1.1c}$$

$$\textstyle\sum_{i=1}^{I} z_{qi} \geq K + 1 \ \forall q \tag{3.1.1.1d}$$

$$z_{qi} \in \{0,1\} \ \forall(q,i) \tag{3.1.1.1e}$$

$$\beta_{qj0}, \beta_{qjk} \ free \ \forall(q,j,r_i) \tag{3.1.1.1f}$$

$$t_{qijr_i} \geq 0 \ \forall(i,q,j,r_i) \tag{3.1.1.1g}$$

where for any integer $i, i \in I$ means that $i \in \{1, \dots, I\}$. In (3.1.1.1), the objective function is p-norm of the vector of the total absolute residuals defined by

$$\|t\|_p = \left(\textstyle\sum_{q=1}^{Q} |t_q|^p\right)^{1/p} \tag{3.1.1.2}$$

where $|\,.\,|$ is the absolute value of a scalar. Furthermore, the two sets of constraints in (3.1.1.1a) and (3.1.1.1b) are indicator constraints; i.e., if $z_{qi} = 1$, the two corresponding constraints together with the minimization in the objective function make $t_{qijr_i} =$

$\left|y_{ir_ij} - (\beta_{qj0} + \left(\sum_{k=1}^{K} \beta_{qjk} x_{ir_ik}\right)\right|$. If $z_{qi} = 0$, these constraints provide negative lower bounds for t_{qijr_i}, and the nonnegativity conditions on t_{qijr_i} together with the minimization make $t_{qijr_i} = 0$ for that cluster. The set of constraints (3.1.1.1c) assigns each entity i to exactly one cluster, and the set of constraints (3.1.1.1d) is required for the identifiability of regression coefficients β_{qj0} and β_{qjk} per cluster. Finally, (3.1.1.1e), (3.1.1.1f) and (3.1.1.1g) require the assignment variables z_{qi}, the regression coefficients β_{qj0} and β_{qjk}, and the absolute residuals t_{qijr_i} to be binary, unrestricted in sign, and nonnegative, respectively.

In this study, we consider $p = \infty$, where the ∞-norm defined as:

$$\|t\|_\infty = \max_{1 \le q \le Q} \{|t_q|\} \tag{3.1.1.3}$$

The proposed GCLR problem is solved using the branch-and-cut algorithm. Before starting the algorithm, an initial value for the M in the (3.1.1.1a) and (3.1.1.1b) has to be determined, and that value has to be bigger than any absolute residual. To find an appropriate initial M value, we initially randomly assign entities to clusters and solve a linear programming problem over each cluster independently, which minimizes the maximum sum of residuals given the random assignments of entities to that cluster. After solving these linear programming problems, the biggest objective value over all clusters is used as the initial M value in (3.1.1.1).

3.1.2 SOS Type-1 Formulation

The indicator constraints (3.1.1.1a) and (3.1.1.1b) are known to work well if a big enough value for the M is known prior to solve the problem. However, having a tight value for the M is impossible because this value depends on the currently unknown optimal clustering of all entities and the unknown regression coefficients for all clusters. If a small value is used for the M, the constraints (3.1.1.1a) and (3.1.1.1b) can cut the globally optimal solution; hence, solving (3.1.1.1) results only in a sub-optimal clustering (Carbonneau et al., 2011). Using a too big value for the M can lead to numerical stability problems. Therefore, for the GCLR problem with multiple responses, we propose the following formulation, which avoids using the big-M value.

$$min\|t\|_p \tag{3.1.1.4}$$

subject to

$$t_{qijr_i} - \left[y_{ir_ij} - (\beta_{qj0} + \left(\sum_{k=1}^{K} \beta_{qjk} x_{ir_ik}\right)\right] + s_{qijr_i} \geq 0 \quad \forall(i,q,j,r_i) \tag{3.1.1.4a}$$

$$t_{qijr_i} + \left[y_{ijr_i} - (\beta_{qj0} + \left(\sum_{k=1}^{K} \beta_{qjk} x_{ijk}\right)\right] + s_{qijr_i} \geq 0 \quad \forall(i,q,j,r_i) \tag{3.1.1.4b}$$

$$\sum_{q=1}^{Q} z_{qi} = 1 \ \forall i \tag{3.1.1.4c}$$

$$\sum_{i=1}^{I} z_{qi} \geq K + 1 \ \forall q \tag{3.1.1.4d}$$

$$z_{qi} \in \{0,1\} \ \forall(q,i) \tag{3.1.1.4e}$$

$$\beta_{qj0}, \beta_{qjk} \ free \ \ \forall(q,j,r_i) \tag{3.1.1.4f}$$

$$t_{qijr_i} \geq 0, s_{qijr_i} \geq 0 \ \ \forall(i,q,j,r_i) \tag{3.1.1.4g}$$

where for a fixed q and i, the sets $\{z_{qi}, s_{qijr_i}\}$ form special ordered sets of type 1 (SOS Type-1) for all $j \in J$ and $r_i \in R_i$; i.e., at most one of the variables z_{qi} and s_{qijr_i} can take a value bigger than zero. These SOS Type-1 conditions are also defined for all $q \in Q$ and $i \in I$. In addition to SOS Type-1 conditions, the proposed formulation differs from the old formulation in (3.1.1.1) in that we define new nonnegative variables s_{qijr_i}, and replace the values $M\left(1 - z_{qi}\right)$ in (3.1.1.1a) and (3.1.1.1b) by s_{qijr_i} to obtain (3.1.1.4a) and (3.1.1.4b); this way, we avoid using the big-M, yet we have $QIJR_i$ more continuous variables in the proposed formulation than the old formulation.

The proposed formulation works exactly the same as the old formulation, which can be explained as follows. Because of the assignment constraints (3.1.1.4c), for each entity i, exactly one z_{qi} is equal to one, say $z_{q'i} = 1$. Then, for this specific i, SOS Type-1 conditions make $s_{q'ijr_i} = 0$ for all $j \in J$ and $r_i \in R_i$, and the constraints (3.1.1.4a) and (3.1.2.1b) together with the minimization makes $t_{q'ijr_i}$ equal to the absolute residuals for all $j \in J$ and $r_i \in R_i$; i.e., $t_{q'ijr_i} = \left|y_{ir_ij} - (\beta_{q'j0} + \left(\sum_{k=1}^{K} \beta_{q'jk} x_{ir_ik}\right)\right|$. For the remaining clusters $q \in Q$, $q \neq q'$, s_{qijr_i} take nonnegative values to satisfy the constraints (3.1.1.4a) and (3.1.1.4b) and to force the corresponding t_{qijr_i} to be zero due to the minimization. The remaining constraints and the objective function are exactly the same as the old formulation.

3.2 Determination of Cluster Number Q

The selection of the number Q of clusters is critical to the CLR. However, there are very few papers that deal with determining the number of clusters. The Akaike information criterion (AIC) and the Bayesian information criterion (BIC) can be employed when the residuals have a certain distribution like normal distribution. Furthermore, they are reported to overestimate the number of clusters in the literature. Some papers suggest to determine a fixed number Q of clusters, solve the CLR problem for this Q, and repeat the same procedure for a few different values of Q. Among these Q values, one approach is to choose the one which gives the smallest optimal objective value as the optimal number of clusters.

Moreover, we propose another practical approach to determine a good value for Q based on cross-validation. This procedure is applied after the optimization, and hence, the optimal assignments z_{qi}^* are given. We follow the steps below for each cluster.

Step 1: Delete the date of one of the entity in the cluster, say entity κ.

Step 2: Solve a linear programming problem with the data of the remaining entities in that cluster to find the new $\beta_{qj-\kappa}$ for all responses simultaneously.

Step 3: Compute the predicted response value $\beta_{qj-\kappa}x'_\kappa$ for each response j, and repeat the steps 1 through 3, each time selecting a new entity κ, until all entities in that cluster are selected.

Step 4: Compute the mean absolute error $MAE_{j,q}$ per response j and per cluster q by

$$MAE_{j,q} = \frac{1}{|C_q|} \sum_{\kappa \in C_q} \sum_{r_\kappa=1}^{R_\kappa} \left| y_{\kappa r_i j} - \beta_{qj-\kappa} x'_\kappa \right| \tag{3.2.1}$$

where $\left|C_q\right|$ is the cardinality of cluster q and $\left|y_{\kappa r_\kappa j} - \beta_{qj-\kappa} x'_\kappa\right|$ is the absolute deviation between the observed response value $y_{\kappa r_\kappa j}$ and its prediction $\beta_{qj-\kappa} x'_\kappa$. Smaller values of $MAE_{j,q}$ indicate better prediction accuracies of models, and therefore, they are preferable. Select the number Q of clusters which gives the minimum $MAE_{j,q}$ in (3.2.1) over all responses and over all clusters as the optimal number of clusters.

4. EXPERIMENTS AND RESULTS

The optimization software Gurobi 8.1.0 C API is used for solving optimization problems that are explained in this section. The computational platform has Windows Server 2008 R2 as the operating system, and 32 GB RAM, and 9 threads of 1.7 Ghz each.

4.1 Proposed Model Application on a Problem from the Literature

4.1.1 Objective and Scope of Application

This application consists of two sections, where the first section introduce the dataset and data preprocessing. The second section presents the optimal clustering based on (3.1.1.1) and the corresponding regression parameters. In the second section, we further apply the two procedures to determine an optimal Q and compare the results of our approach which considers all responses simultaneously, with the approach in the literature which considers one response at-a-time.

4.1.2 Data and Data Preprocessing

We use the JURA dataset in (Goovaerts, 1997) which consists of measurements of concentrations of seven heavy metals, namely cadmium, cobalt, chromium, copper, nickel, lead, and zinc, recorded at 359 locations in the topsoil of a region of the Swiss Jura. The type of land use (forest, pasture, meadow, tillage) and rock type (Argovian, Kimmeridgian, Sequanian, Portlandian, Quaternary) are also recorded for each location. We are interested in the prediction of the concentration of metals that are more expensive to measure using measurements of metals that are cheaper to sample. In this study, cadmium, copper and lead are treated as responses while the remaining metals along with land use type, rock type and the coordinates of each location are used as explanatory variables. We experiment with 30 randomly selected locations from the JURA data, and we call these locations as L1, L2,..., L30. The summary and descriptive statistics of the explanatory variables and the responses are given below in Table 4.1.

Table 4.1: Descriptive statistics of explanatory and response variables in JURA dataset

Variable	**Count**	**Mean**	**Min**	**Max**	**Standard Deviation**	**Explanation**
$\boldsymbol{X_1}$	30	3.18	1.69	4.92	0.93	Abscissa of location
$\boldsymbol{X_2}$	30	2.76	0.99	5.12	1.33	Ordinate of location
$\boldsymbol{X_3}$	30	2.60	1.00	3.00	0.72	Land type
$\boldsymbol{X_4}$	30	2.50	1.00	5.00	1.53	Rock type
$\boldsymbol{X_5}$	30	9.14	2.36	16.32	3.90	Cobalt concentration
$\boldsymbol{X_6}$	30	33.96	18.40	52.00	9.02	Chromium concentration
$\boldsymbol{X_7}$	30	20.31	6.64	35.12	8.43	Nickel concentration
$\boldsymbol{X_8}$	30	76.55	25.20	145.60	30.34	Zinc concentration
$\boldsymbol{Y_1}$	30	1.28	0.24	4.19	0.91	Cadmium concentration
$\boldsymbol{Y_2}$	30	23.36	3.96	66.12	14.67	Copper concentration
$\boldsymbol{Y_3}$	30	55.71	21.48	141.00	24.34	Lead concentration

In multiple linear regression analysis, multiple binding or linear dependence between vectors of regressor variables can have serious effects on the estimation of parameters (Gunst and Webster, 1975). In detail, multicollinearity inflates the variance of regression parameters. Therefore, we employed the variance inflation factor (VIF) to examine potential multicollinearity issues (Mansfield and Helms, 1982). In general, a VIF that is greater than ten is considered unacceptable. Figure 4.1 shows the VIF values and correlations between eight explanatory variables. Since there are no values greater than ten, all explanatory variables are included in the GCLR model. Consequently, there are $K = 8$ explanatory variables for which the regression parameters β_{qjk} are to be found in addition to the intercept β_{qj0}. Therefore, the right-hand-side of the (3.1.1.1d) constraint is nine; i.e., each cluster has to have at least nine entities for the identifiability of the β_{qjk}'s. This gives $Q_{max} = \left[{}^{30}/_{9}\right] = 3$; hence, in both subsections, we consider grouping the 30 locations into $Q = 2$ and $Q = 3$ clusters.

	VIF		X1	X2	X3	X4	X5	X6	X7	X8
X1	1.22	X1	1	-0.087	-0.08	0.18	0.34	0.11	0.28	0.2
X2	1.25	X2	-0.087	1	-0.027	0.0077	-0.34	-0.027	-0.28	-0.021
X3	1.19	X3	-0.08	-0.027	1	0.31	0.094	0.02	0.077	0.12
X4	1.30	X4	0.18	0.0077	0.31	1	0.21	-0.045	0.038	-0.049
X5	3.13	X5	0.34	-0.34	0.094	0.21	1	0.53	0.78	0.56
X6	2.95	X6	0.11	-0.027	0.02	-0.045	0.53	1	0.68	0.76
X7	3.72	X7	0.28	-0.28	0.077	0.038	0.78	0.68	1	0.66
X8	2.90	X8	0.2	-0.021	0.12	-0.049	0.56	0.76	0.66	1

Figure 4.1: VIF values and correlation matrix of explanatory variables in JURA data

4.1.3 Results and Discussions

We start by fixing the number of clusters to $Q = 2$ in (3.1.1.1), and solve the resulting MILP. Then, we do the same for $Q = 3$. For $Q = 2$, the M value is fixed to $M = 260$, and for $Q = 3$, to $M = 70$. These values of M are found by randomly assigning entities to clusters and solving the resulting linear programming problems a few number of times. The clock solution times for $Q = 2$ and $Q = 3$ are 139.24 seconds and 528.66 seconds, respectively. For the solution of the MILP problem, the gap tolerance between the lower and upper bounds is set to 10^{-4}. The optimal cluster assignments of 30 locations as well as the optimal objective value are given in Table 4.2.

To determine Q, we first choose the optimal number of cluster is given by $Q^* = 3$, because according to (3.1.1.1) the optimal objective value $w^* = 9.4592$ for $Q = 3$ is less than the optimal objective value $w^* = 59.5728$ for $Q = 2$, the optimal number of clusters is given by $Q^* = 3$. Before introducing our results of the cross-validation procedure, we give below in Table 4.3 the regression parameters β_{qj} for all clusters and all responses for $Q^* = 3$.

Table 4.2: Optimal assignments of 30 locations to two and three clusters and their optimal objective value

Clusters	Optimal Objective	Optimal Assignments
$Q = 2$	$w^* = 59.5728$	C_1 = {L1, L3, L6, L8, L10, L15, L17, L22, L23, L26, L28, L30} C_2 = {L2, L4, L5, L7, L9, L11, L12, L13, L14, L16, L18, L19, L20, L21, L24, L25, L27, L29}
$Q = 3$	$w^* = 9.4592$	C_1 = {L1, L2, L5, L11, L12, L14, L16, L24, L28, L29, L30} C_2 = {L2, L4, L5, L7, L9, L11, L12, L13, L14, L16, L18, L19, L20, L21, L24, L25, L27, L29} C_3 = {L3, L6, L7, L8, L15, L20, L21, L22, L27}

We apply the cross-validation for both $Q = 2$ and $Q = 3$, using the optimal assignments in Table 4.2. For example, for $Q = 2$, we first delete L1 from C_1 and recompute the regression parameters $\beta_{1j,-L1}$ for $j = 1,2,3$. We repeat this procedure for all locations in C_1 and then C_2, and compute $MAE_{j,q}$ in (3.2.1) for $j = 1,2,3$ and $q = 1,2$. We follow the same steps for $Q = 3$. According to results, the sum of total $MAE_{j,q}$ over all responses and clusters for $Q = 2$ is 73.8650, whereas $Q = 3$, it is 92.2763. Hence, for prediction accuracies of models, $Q = 2$ is a better choice.

We compare the results of our approach, which considers all three responses simultaneously in solving problem (3.1.1.1) with the approach in the literature, which considers a single response at-a-time in solving (3.1.1.1). We adapt our objective function to the sum of L^1-norm residuals. We consider the same 30 randomly selected locations and cluster number $Q = 2$. We resolve the problem with the new objective function, considering all three responses simultaneously. The optimal assignments given in the first row Table 4.4 are the results of this optimization. Then we consider only response 1 (cadmium concentration) in Table 4.1, and solve problem (3.1.1.1) for that response. We repeat the same procedure for response 2 (copper concentration) and response 3 (lead concentration); i.e., the locations are clustered each time for that response. The optimal assignments results for the underlying response are in Table 4.4.

Table 4.3: Regression parameter values for all clusters and responses

Cluster	Regression Parameter	Response j = 1	Response j = 2	Response j = 3
Cluster 1	β_{1j0}	-2.4512	10.3735	156.775
	β_{1j1}	0.4582	-5.8365	-15.565
	β_{1j2}	0.075	-1.0991	-9.5503
	β_{1j3}	-0.1098	8.1412	-0.0759
	β_{1j4}	0.1385	-5.5809	2.7142
	β_{1j5}	-0.1701	4.1292	-1.2278
	β_{1j6}	0.0496	-0.6562	-2.5881
	β_{1j7}	0.0128	0.1659	1.5027
	β_{1j8}	0.0237	0.0874	0.6244
Cluster 2	β_{2j0}	0.4746	-74.235	-48.445
	β_{2j1}	0.1279	5.1319	6.2178
	β_{2j2}	0.0588	0.2402	-2.2873
	β_{2j3}	-0.0431	26.8436	23.7311
	β_{2j4}	-0.044	-1.4961	-2.4029
	β_{2j5}	0.0502	-0.9788	-2.0614
	β_{2j6}	-0.0437	-1.1756	-0.2311
	β_{2j7}	0.0451	-1.0492	-1.3972
	β_{2j8}	0.0023	1.0106	1.0031
Cluster 3	β_{2j0}	0.3169	-31.099	61.7788
	β_{2j1}	0.3021	-7.9282	-16.379
	β_{2j2}	0.1694	5.3172	8.2055
	β_{2j3}	-1.1382	29.2985	27.2979
	β_{2j4}	-0.0074	-4.6247	-4.7941
	β_{2j5}	-0.0131	0.4684	-0.7217
	β_{2j6}	0.0095	-0.0609	-1.4955
	β_{2j7}	-0.0095	-1.0344	-0.8527
	β_{2j8}	0.032	0.3047	0.6714

Table 4.4: Optimal assignments of 30 locations based on different approaches

All vs. One At-a-time	Optimal Assignments
Three responses simultaneously	C_1 = {L1, L2, L4, L5, L7, L9, L11, L12, L13, L14, L15, L16, L18, L19, L20, L21, L24, L25, L27, L29} C_2 = {L3, L6, L8, L10, L17, L22, L23, L26, L28, L30}
With respect to response 1 only	C_1 = {L1, L2, L4, L5, L6, L11, L12,L13, L14, L19, L20, L23, L24, L25} C_2 = {L3, L7, L8, L9, L10, L15, L16, L17, L18, L21, L22, L26, L27, L28, L29, L30}
With respect to response 2 only	C_1 = {L1, L2, L3, L4, L6, L9, L10, L11, L12, L13, L14,L18, L19, L20, L21, L27, L29} C_2 = {L5, L7, L8, L15, L16, L17, L22, L23, L24, L25, L26, L28, L30}
With respect to response 3 only	C_1 = {L1, L5, L7, L8, L12, L14, L18, L20, L21, L22, L23, L26, L27} C_2 = {L2, L3, L4, L6, L9, L10, L11, L13, L15, L16, L17, L19, L24, L25, L28, L29, L30}

Now, suppose that we want to cluster these 30 locations with respect to three responses and suppose that we apply the procedure in the literature, which considers one response at-a-time. For example, consider location 2 (L2). This location is clustered in C_1 with respect to both cadmium and copper concentrations, but it is clustered in C_2 with respect to lead concentration. As a result, it is not evident where to cluster L2. Our approach suggests to cluster L2 in C_1. Examining further all optimal assignments in the last three rows of Table 4.4, only eight locations out of 30 are grouped in the same cluster with respect to all three responses, using the approach in the literature. These locations are L12, L14, and L20 in C_1, and L15, L16, L17, L28, and L30 in C_2. For the remaining 22 locations, it is not clear where to cluster with this approach. Furthermore, our approach suggests to cluster L15, L16, and L28 in C_1; i.e., even though one response at-a-time approach suggests to group locations in the same cluster C_2 for all responses, this can still result in a wrong clustering when all responses are considered simultaneously. Consequently, current approach in the literature can give false clustering when there are multiple responses for which entities have to be clustered, which makes our approach even more valuable.

4.2 Proposed Model Application on the Store Clustering Problem

4.2.1 Objective and Scope of Application

We employ the proposed new generalized clusterwise linear regression formulation to provide comparable store segments for a fashion retail company operating in Turkey by considering both the factors that affect performance and performance criteria as explanatory variables and response variables, respectively. We determine the variables that are involved in the model with the contributions of the executives of the fashion retailer. However, we are not able to disclose most of the data obtained from the company and the name of the company due to confidentiality. The rest of the study consists of two sections, where the first section introduce the dataset and data preprocessing. The second section presents the optimal clustering based on (3.1.1.1) and the corresponding regression parameters.

4.2.2 Data and Data Preprocessing

The most commonly used performance measures appear in the literature review are sales amount, store profits, market share, customer satisfaction, and price elasticity. Company executives considered the sales amount in Turkish Lira as the output variable because it is the most used performance criterion in the company's reporting system. The gross sales amounts data of three substitutable sub-categories of products in the most sold main-category are collected from the enterprise resource planning (ERP) system. In addition, there are two observations, which are the sales during the first (January, February, March) and the fourth quarter (October, November, December) for each sub-category. Specifically, we consider partitioning 21 stores located in Istanbul, and the whole data used in this study belongs to 2018.

The five explanatory variables that are used to predict each store's performance can be grouped into two main categories as location related and store related characteristics which are illustrated in Figure 4.2. In addition, the descriptive statistics of the input variables are shown in Table 4.5.

The only characteristic that is related to store is store age (S_AGE) which represents the period that the store has been serving in the same location. In our model, this period is specified as years and obtained from the company's database.

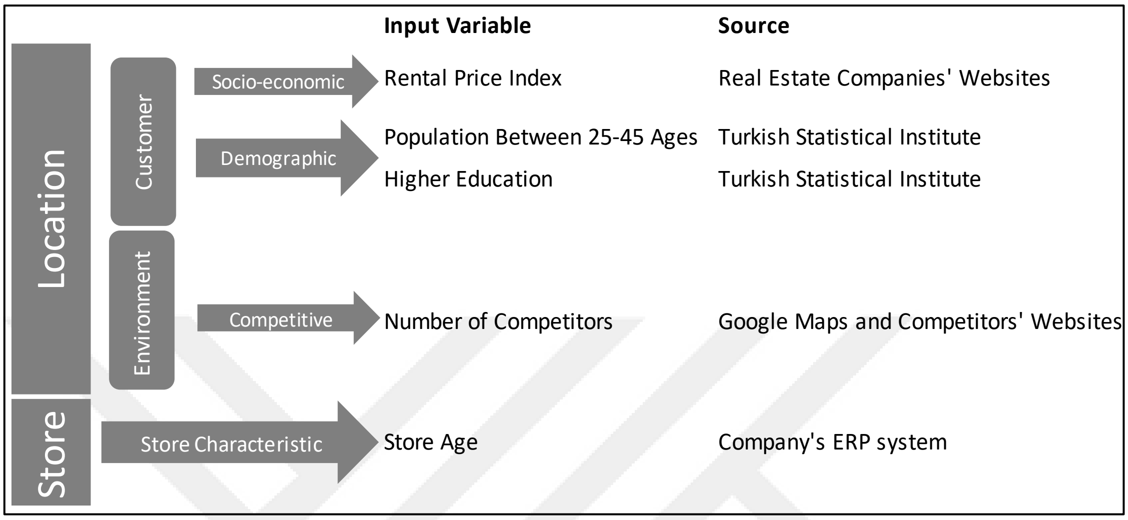

Figure 4.2: Explanatory variables used in GCLR for store clustering

With increasing competition, the information about customer characteristics has become vital for all companies in retail industry. Most companies establish customer relationship management (CRM) department and try to collect customer data with full effort. Although a customer loyalty program has created for the company that is analyzed, there is not adequate information about its customers' characteristics for each store. Consequently, the data of customers demographics are obtained from the Turkish Statistical Institute (TUIK, 2018) on the assumption that stores serve the district where they are located. Because the main target market of the company is composed of women, 25-45 years of age and with high-level education; population of women (POP) who are between 25-45 ages and percentage of higher educated women (HE) in total inhabitants are selected as demographic variables.

The competitive environment is another factor that affects store performance directly. Therefore, the number of competing stores (NoC) in the same shopping center or in the proximity of 500 meters is determined by using Google Maps (2018), and websites of competing companies. Moreover, the result of the brand positioning study is taken into consideration in the selection of competing companies. As illustrated in Figure 4.3 Brand A, B, C, D, E, F, G are selected as competitors to the analyzed company.

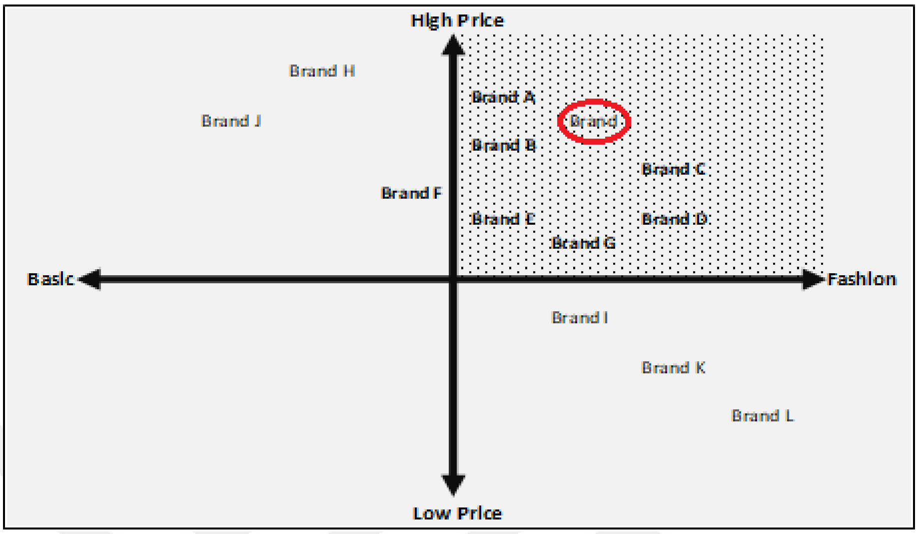

Figure 4.3: Brand positioning study in fashion retail industry

Finally, the socio-economic characteristics of the population are closely related to the total sales potential in the trade area where the store is located. For evaluation of this factor on the store performance, we collected rental price index (RPI) data for each district from real estate firms' websites (sahibinden.com, 2018). The rental price index is an indicator of average rental prices for apartments per square meter.

Table 4.5: Descriptive statistics of explanatory variables in store clustering dataset

Variable	Count	Mean	Min	Max	Standard Deviation	Units of measure
RPI	42	18.38	9.00	27.00	4.83	Turkish Lira/Square Meter
NoC	42	4.14	0.00	9.00	2.58	Number
S_AGE	42	6.97	2.58	11.28	2.43	Years
POP	42	71,671	34,084	161,949	32,969	Number
HE	42	0.70	0.39	0.90	0.14	Percentage

Figure 4.4 shows the VIF values and correlations between explanatory variables. In general, a VIF that is greater than 10 is considered unacceptable. The HE variable has the highest VIF value and the correlation between RPI and HE is over than 0.80.

	VIF		RPI	NoC	S_AGE	POP	HE
RPI	7.70	RPI	1	0.29	0.31	-0.59	0.92
NoC	1.49	NoC	0.29	1	0.42	-0.31	0.44
S_AGE	1.36	S_AGE	0.31	0.42	1	-0.28	0.42
POP	1.83	POP	-0.59	-0.31	-0.28	1	-0.67
HE	10.44	HE	0.92	0.44	0.42	-0.67	1

Figure 4.4: VIF values and correlation matrix of explanatory variables in store clustering data

Therefore, for overcoming possible multicollinearity problems, the HE variable is removed from the inputs. After HE is removed, VIF values and correlations of the remained variables are illustrated in Figure 4.5. Because there is not any variable to be removed when considering VIF and correlations, these variables included in the GCLR model as input variables.

	VIF		RPI	NoC	S_AGE	POP
RPI	2.16	RPI	1	0.44	0.42	-0.67
NoC	1.36	NoC	0.44	1	0.42	-0.31
S_AGE	1.33	S_AGE	0.42	0.42	1	-0.28
POP	1.82	POP	-0.67	-0.31	-0.28	1

Figure 4.5: VIF values and correlation matrix of explanatory variables except HE variable

4.2.3 Results and Discussions

We call the stores as S1, S2, …, S31. Because there are $K = 4$ independent variables, the right handside of the (3.1.1.1d) and (3.1.2.1d) constraints are five. This gives $Q_{max} = \left[{}^{21}/_{5} \right] \cong 4$. We solve the problem with the formulations of (3.1.1.1) and (3.1.2.1) for $Q = 2,3,4$. Then, we do the same for $Q = 3$. For all Q values, the big-M

value determined as the 30 times larger than average response values. For the solution of the problems, the gap tolerance between the lower and upper bounds is set to 10^{-4}. The clock solution times with both formulations are mentioned in Table 4.6. First outcome of this section is clock solution times are increasing through the number Q of clusters because of the combinatorial nature of the GCLR problem. Secondly, although, the formulation in (3.1.1.4) eliminates the problem of setting the big-M value, the solution times are longer than (3.1.1.1). The optimal cluster assignments of the stores are in Table 4.5.

Table 4.6: Comparison of running times of formulations for different number of clusters for store clustering

Entity_Cluster	Big-M	SOS Type-1	SOS Type-1 / Big-M
21_2	11.80 seconds	25.82 seconds	2.2
21_3	457.07 seconds	3,579.41 seconds	7.8
21_4	1,027.71 seconds	18,804.74 seconds	18.3

Table 4.7: Optimal assignments of stores into clusters

Cluster	Optimal Objective	Optimal Assignments
$Q=2$	$w^*=$ 735,781.4	C_1={S1, S5, S6, S7, S9, S10, S14, S15, S16, S17,S21} C_2={S2, S3, S4, S8, S11, S12, S13, S18, S19, S20}
$Q=3$	$w^*=$ 402,412.0	C_1={S7, S10, S11, S12, S14, S15, S20} C_2={S5, S6, S9, S13, S16, S17, S21} C_3={S1, S2, S3, S4, S8, S18, S19}
$Q=4$	$w^*=$ 262,464.1	C_1={S1, S6, S10, S14, S18} C_2={S3, S4, S5, S7, S8} C_3={S11, S15, S16, S17, S19, S21} C_4={S2, S9, S12, S13, S20}

We apply the cross-validation for both $Q=2$, $Q=3$ and $Q=4$ using the optimal assignments in Table 4.7. In conclusion, according to prediction accuracies of models, $Q=3$ is a better choice. We give below in Table 4.8 the regression parameters β_{qj} for all clusters and all responses for $Q^*=3$. According to the regression parameters the sales of analyzed product categories of the stores in the first cluster is inversely proportional to S_AGE and directly proportional with RPI, NoC and POP. On the other hand, the sales of the stores in the second cluster is inversely proportional to S_AGE and POP, while directly proportional to RPI and NoC. Lastly, the sales of stores in the third cluster is only inversely proportional to NoC, while directly proportional to other explanatory variables.

Table 4.8: Regression parameter values for all clusters and responses for store clustering

Cluster	Regression Parameters	Responses		
		j = 1	j = 2	j = 3
Cluster 1	β_{1j0}	-440,636.3	-534,632.0	-140,812.6
	β_{1j1}	33,750.7	33,129.0	8,548.0
	β_{1j2}	12,923.6	14,062.8	5,198.4
	β_{1j3}	-4,416.9	-3,603.5	-884.7
	β_{1j4}	2.5	2.6	0.6
Cluster 2	β_{2j0}	17,629.0	-680.8	-8,740.3
	β_{2j1}	2,817.3	1,786.8	1,224.0
	β_{2j2}	3,131.6	3,309.9	1,115.5
	β_{2j3}	-2,172.0	-2,197.8	-599.5
	β_{2j4}	-0.1	-0.1	0.0
Cluster 3	β_{2j0}	-146,545.9	-94,638.6	-35,288.7
	β_{2j1}	5,102.5	3,865.9	1,372.1
	β_{2j2}	-2,167.1	-4.,012.7	-1,178.7
	β_{2j3}	113.2	291.3	141.0
	β_{2j4}	9.5	6.5	2.4

5. CONCLUSIONS

In this thesis, we derive new mixed-integer linear programming formulation with big-M and SOS Type-1 variables for multiple responses for clusterwise linear regression problem. Moreover, our formulation consider more than one observation per entity. In addition, we define a slightly different objective function by minimizing L^{∞}-norm of residuals. We also present a reasonably practicable process for estimating appropriate initial big-M value. Moreover, the formulation with SOS Type-1 variables eliminates the needs of big-M in the modeling although the solution time lasts longer than previous model.

We apply our model formulations to the JURA dataset and illustrate the benefits of taking into account multiple responses at the same time. Furthermore, we employ the proposed clusterwise linear regression models to cluster the stores of a fashion retailer in Turkey. For determining the variables that are possibly included in models we present a brief literature review about store performance and selection of store location studies. Both the explanatory and response variables that are used in literature are classified in categories detailed. The model results show that there are different negative or positive relationships between response variables and explanatory variables for different clusters. Retail company could benefit from these results while making critical decision about stores. For instance, the decision of open a new store or close an existing store could be evaluated in detail.

In conclusion, future research on tailor-made algorithms is needed to decrease the solving time for the problem and to be able to work on large-scale problems. Furthermore, a new mathematical programming formulation can be edited as set-partitioning problem. In addition, column generation algortihm would be appropriate to solve that new formulation.

REFERENCES

Achabal, D., Gorr, W. L., & Mahajan, V. (1982). MULTILOC: A multiple store location model. *Journal of Retailing,* 58, 5-25.

Almohri, H., Chinnam, R. B., & Colosimo, M. (2019). Data-driven analytics for benchmarking and optimizing the performance of automotive dealerships. *International Journal of Production Economics,* 213, 69-80.

Angun, E., Altınoy, A., 2019. “A new mixed-integer linear programming formulation for multiple responses regression clustering”, *in Proceedings of the 6th International Conference on Control, Decision and Information Technologies*.

Applebaum, W. (1966). Methods for determining store trade areas, market penetration, and potential sales. *Journal of marketing Research,* 3(2), 127-141.

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

Aurifeille, J.M. (2000). A Bio-Mimetic Approach to Marketing Segmentation: Principles and Comparative Analysis. *European Journal of Economic and Social Systems,* 14(1), 93–108.

Athanassopoulos, A. D. (1995). Performance improvement decision aid systems (PIDAS) in retailing organizations using data envelopment analysis. *Journal of Productivity Analysis,* 6(2), 153-170.

Athanassopoulos, A. D., & Ballantine, J. A. (1995). Ratio and frontier analysis for assessing corporate performance: evidence from the grocery industry in the UK. *Journal of the Operational Research Society*, 46(4), 427-440.

Bagirov, A. M. & Ugon, J. (2018). Nonsmooth DC programming approach to clusterwise linear regression: optimality conditions and algorithms. *Optimization methods and software,* 33(1), 194-219.

Bagirov, A. M., Mahmood, A., & Barton, A. (2017). Prediction of monthly rainfall in Victoria, Australia: Clusterwise linear regression approach. *Atmospheric research*, 188, 20-29.

Bagirov, A. M., Ugon, J., & Mirzayeva, H. (2013). Nonsmooth nonconvex optimization approach to clusterwise linear regression problems. *European Journal of Operational Research*, 229(1), 132-142.

Bagirov, A. M., Ugon, J., & Mirzayeva, H. G. (2015a). Nonsmooth optimization algorithm for solving clusterwise linear regression problems. *Journal of Optimization Theory and Applications*, 164(3), 755-780.

Bagirov, A. M., Ugon, J., & Mirzayeva, H. G. (2015b). An algorithm for clusterwise linear regression based on smoothing techniques. *Optimization letters*, 9(2), 375-390.

Banker, R. D. & Morey, R. C. (1986). The use of categorical variables in data envelopment analysis. *Management Science*, 32(12), 1613-1627.

Banker, R. D., Lee, S. Y., Potter, G., & Srinivasan, D. (2010). The impact of supervisory monitoring on high-end retail sales productivity. *Annals of Operations Research*, 173(1), 25-37.

Barros, C. P. (2005). Efficiency in hypermarket retailing: A stochastic frontier model. *The International Review of Retail, Distribution and Consumer Research*, 15(2), 171-189.

Barros, C. P. (2006). Efficiency measurement among hypermarkets and supermarkets and the identification of the efficiency drivers: A case study. *International Journal of Retail & Distribution Management*, 34(2), 135-154.

Barros, C. P. & Alves, C. (2004). An empirical analysis of productivity growth in a Portuguese retail chain using Malmquist productivity index. *Journal of Retailing and Consumer Services*, 11(5), 269-278.

Barros, C. P. & Alves, C. A. (2003). Hypermarket retail store efficiency in Portugal. *International Journal of Retail & Distribution Management*, 31(11), 549-560.

Barth, J. E. (2007). Customer engagement and the operational efficiency of wine retail stores. *International Journal of Wine Business Research*, 19(3), 207-215.

Bearden, W. O., Teel, J. E., & Durand, R. M. (1978). Media usage, psychographic, and demographic dimensions of retail shoppers. *Journal of retailing*, 54(1), 65-74.

Bermingham, P., Hernandez T., & Clarke, I. (2013). Network planning and retail store segmentation: A spatial clustering approach. *International Journal of Applied Geospatial Research*, 4(1), 67-79.

Bertsimas, D. & Shioda, R. (2007). Classification and regression via integer optimization. *Operations Research*, 55(2), 252-271.

Bilgic, E., Kantardzic M., & Cakir O. (2015). Retail store segmentation for target marketing. *Advances in Data Mining: Applications and Theoretical Aspects*, Vol. 9165 of *Lecture Notes in Computer Science*, Springer Cham, pp. 32-44.

Boufounou, P. V. (1995). Evaluating bank branch location and performance: A case study. *European Journal of Operational Research*, 87(2), 389-402.

Brusco, M. J., Cradit, J. D., Steinley, D., & Fox, G. L. (2008). Cautionary remarks on the use of clusterwise regression. *Multivariate Behavioral Research*, 43(1), 29-49.

Brusco, M. J., Cradit, J. D., & Tashchian, A. (2003). Multicriterion clusterwise regression for joint segmentation settings: An application to customer value. *Journal of Marketing Research*, 40(2), 225-234.

Camanho, A. S., Portela, M. C., & Vaz, C. B. (2009). Efficiency analysis accounting for internal and external non-discretionary factors. *Computers & Operations Research*, 36(5), 1591-1601.

Campo, K. & Gijsbrechts, E. (2004). Should retailers adjust their micro-marketing strategies to type of outlet? An application to location-based store space allocation in limited and full-service grocery stores. *Journal of Retailing and Consumer Services*, 11(6), 369-383.

Campo, K., Gijsbrechts, E., Goossens, T., & Verhetsel, A. (2000). The impact of location factors on the attractiveness and optimal space shares of product categories. *International Journal of Research in Marketing*, 17(4), 255-279.

Caporossi, G. & Hansen, P. (2007). Variable neighborhood search for least squares clusterwise regression. Technical report, Les Cahiers du GERAD, 18.

Carbonneau, R. A., Caporossi, G., & Hansen, P. (2011). Globally optimal clusterwise regression by mixed logical-quadratic programming. *European Journal of Operational Research*, 212(1), 213-222.

Carbonneau, R. A., Caporossi, G., & Hansen, P. (2012). Extensions to the repetitive branch and bound algorithm for globally optimal clusterwise regression. *Computers & Operations Research*, 39(11), 2748-2762.

Carbonneau, R. A., Caporossi, G., & Hansen, P. (2014). Globally optimal clusterwise regression by column generation enhanced with heuristics, sequencing and ending subset optimization. *Journal of classification*, 31(2), 219-241.

Chang, T. H., Kao, L. J., Ou, T. Y., & Fu, H. P. (2018). A hybrid method to measure the operational performance of fast food chain stores. *International Journal of Information Technology & Decision Making*, 17(04), 1269-1298.

Clarke, I., Mackaness W., & Ball, B. (2003). Modelling intuition in retail site assessment (MIRSA): making sense of retail location using retailers' intuitive judgements as a support for decision making. *The International Review of Retail, Distribution and Consumer Research*, 13(2), 175-193.

Clawson, C. J. (1974). Fitting branch locations, performance standards, and marketing strategies to local conditions: A straightforward approach to establishing goals and reevaluating marketing strategies for individual branch locations. *Journal of Marketing*, 38(1), 8-14.

Cottrell, J. L. (1972). An environmental model for performance measurement in a chain of supermarkets. *Journal of Retailing*, 49(3), 51-63.

Davies, R. L. (1973). Evaluation of retail store attributes and sales performance. *European Journal of Marketing*, 7(2), 89-102.

De Mateo, F., Coelli, T., & O'Donnell, C. (2006). Optimal paths and costs of adjustment in dynamic DEA models: with application to Chilean department stores. *Annals of Operations Research*, 145(1), 211-227.

DeSarbo, W. S. & Cron, W. L. (1988). A Maximum Likelihood Methodology for Clusterwise Linear Regression. *Journal of Classification*, 5(2), 249-282.

DeSarbo, W. S. & Edwards, E. A. (1996). Typologies of compulsive buying behavior: A constrained clusterwise regression approach. *Journal of Consumer Psychology*, 5(3), 231-262.

DeSarbo, W. S., Oliver, R. L., & Rangaswamy, A. (1989). A simulated annealing methodology for clusterwise linear regression. *Psychometrika*, 54(4), 707-736.

Dhar, S. K. & Hoch, S. J. (1997). Why store brand penetration varies by retailer. *Marketing Science*, 16(3), 208-227.

Di Mari, R., Rocci, R., & Gattone, S. A. (2017). Clusterwise linear regression modeling with soft scale constraints. *International Journal of Approximate Reasoning*, 91, 160-178.

Donthu, N. & Yoo, B. (1998). Retail productivity assessment using data envelopment analysis. *Journal of retailing*, 74(1), 89-105.

Donthu, N., Hershberger, E. K., & Osmonbekov, T. (2005). Benchmarking marketing productivity using data envelopment analysis. *Journal of Business Research,* 58(11), 1474-1482.

Durvasula, S., Sharma, S., & Andrews, J. C. (1992). STORELOC: A retail store location model based on managerial judgments. *Journal of Retailing*, 68(4), 420-444.

Gauri, D. K. (2013). Benchmarking retail productivity considering retail pricing and format strategy. *Journal of Retailing*, 89(1), 1-14.

Gautsch, D. A. (1981). Specification of patronage models for retail center choice. *Journal of marketing research*, 18(2), 162-174.

Ghosh, A. & Craig, C. S. (1983). Formulating retail location strategy in a changing environment. *Journal of Marketing,* 47(3), 56-68.

Ghosh, A. & McLafferty S.L. (1982). Locating stores in uncertain environments: A scenario planning approach. *Journal of Retailing*, 58(4), 5-22.

Gijsbrechts, E., Campo, K., & Goossens, T. (2003). The impact of store flyers on store traffic and store sales: A geo-marketing approach. *Journal of Retailing*, 79(1), 1-16.

Google Maps (2018), Location of competitors stores. Available online: https://www.google.com/maps

Goovaerts, P. (1997). Geostatistics for Natural Resources Evaluation. Oxford University Press.

Grewal, D., Levy, M., Mehrotra, A., & Sharma, A. (1999). Planning merchandising decisions to account for regional and product assortment differences. *Journal of Retailing*, 75(3), 405-424.

Gunst, R. F., & Webster, J. T. (1975). Regression analysis and problems of multicollinearity. *Communications in Statistics-Theory and Methods,* 4(3), 277-292.

Gupta, A. ,& Mittal, S. (2010). Measuring retail productivity of food & grocery retail outlets using the DEA technique. *Journal of Strategic Marketing*, 18(4), 277-289.

Gurobi Optimizer Reference Manual (2018). Available online: http://www.gurobi.com

Hawkes, G. F. & McLaughin E. W. (1994). STARS: Segment targeting at retail stores. *Staff Papers*, Department of Applied Economics and Management, Cornell Univesity.

Hennig, C. (1997). Fixed Point Clusters and their relation to stochastic models, *Classification and Knowledge Organization*, Book series of *Studies in Classification, Data Analysis, and Knowledge Organization*, Springer Berlin, pp. 20-27.

Hennig, C. (1998). Clustering and Outlier Identification: Fixed Point Cluster Analysis, *Advances in Data Science and Classification*, Book series of *Studies in Classification*, *Data Analysis, and Knowledge Organization*, Springer Berlin, pp. 37-42.

Hennig, C. (2000). Identifiablity of models for clusterwise linear regression. *Journal of Classification*, 17(2), 273-296.

Hennig, C. (2002). Fixed point clusters for linear regression: computation and comparison. *Journal of Classification*, 19(2), 249-276.

Hennig, C. (2003). Clusters, outliers, and regression: fixed point clusters. *Journal of Multivariate Analysis,* 86(1), 183-212.

Hise, R. T., Kelly, J. P., Gable, M., & McDonald, J. B. (1983). Factors affecting the performance of individual chain store units: An empirical analysis. *Journal of Retailing*, 59(2), 22-39.

Hoch, S. J., Kim, B. D., Montgomery, A. L., & Rossi, P. E. (1995). Determinants of store-level price elasticity. *Journal of Marketing Research*, 32(1), 17-29.

Hsu, D. (2015). Comparison of integrated clustering methods for accurate and stable prediction of building energy consumption data. *Applied energy*, 160, 153-163.

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

Ingene, C. A. (1982). Labor productivity in retailing. *Journal of Marketing*, 46(4), 75-90.

Ingene, C. A. (1983). Intertype competition: Restaurants versus grocery stores. *Journal of Retailing*, 59(3), 49-75.

Ingene, C. A. & Brown, J. R. (1987). The structure of gasoline retailing. *Journal of Retailing*, 63(4), 365-392.

Ingene, C. A. & Lusch, R. F. (1980). Market selection decisions for department stores. *Journal of Retailing*, 56(3), 21-40.

Joo, S. J., Stoeberl, P. A., & Fitzer, K. (2009). Measuring and benchmarking the performance of coffee stores for retail operations. *Benchmarking: An International Journal*, 16(6), 741-753.

Kahraman, C., Uluğ, İ., Othan, C. B., Özkan-Özen, Y. D., & Kazançoğlu, Y. (2018). Efficiency analysis in retail sector: Implementation of data envelopment analysis in a local supermarket chain. *The International Symposium for Production Research*, Springer Cham, pp. 884-897.

Kamakura, W. A., Lenartowicz, T., & Ratchfrord, B. T. (1996). Productivity assessment of multiple retail outlets. *Journal of retailing*, 72(4), 333-356.

Kargari, M. & Sepehri, M. M. (2012). Stores clustering using a data mining approach for distributing automotive spare-parts to reduce transportation costs. *Expert Systems with Applications*, 39(5), 4740-4748.

Keh, H. T. & Chu, S. (2003). Retail productivity and scale economies at the firm level: A DEA approach. *Omega*, 31(2), 75-82.

Khadka, M. & Paz, A. (2017). Comprehensive clusterwise linear regression for pavement management systems. *Journal of Transportation Engineering, Part B: Pavements*, 143(4), 04017014.

Khadka, M., Paz, A., & Singh, A. (2018). Generalised clusterwise regression for simultaneous estimation of optimal pavement clusters and performance models. *International Journal of Pavement Engineering*, 1-13.

Korhonen, P. & Syrjänen, M. (2004). Resource allocation based on efficiency analysis. *Management Science*, 50(8), 1134-1144.

Kumar, V. & Karande, K. (2000). The effect of retail store environment on retailer performance. *Journal of Business Research*, 49(2), 167-181.

Kuo, R. J., Chi, S. C., & Kao, S. S. (2002). A decision support system for selecting convenience store location through integration of fuzzy AHP and artificial neural network. *Computers in Industry*, 47(2), 199-214.

Lau, K. N., Leung, P. L., & Tse, K. K. (1999). A mathematical programming approach to clusterwise regression model and its extensions. *European Journal of Operational Research*, 116(3), 640-652.

Li, Y. & Liu, L. (2012). Assessing the impact of retail location on store performance: A comparison of Wal-Mart and Kmart stores in Cincinnati. *Applied Geography*, 32(2), 591-600.

Lloyd, S. (1982). Least squares quantization in PCM. *IEEE transactions on information theory*, 28(2), 129-137.

Luo, Z. & Chou, E. Y. (2006). Pavement condition prediction using clusterwise regression. *Transportation Research Record*, 1974(1), 70-77.

Mansfield, E. R., & Helms, B. P. (1982). Detecting multicollinearity. *The American Statistician*, 36(3a), 158-160.

Martin, P. (1967). Savings and loans in new submarkets: Search for predictive factors. *Journal of Marketing Research*, 4(2), 163-166.

Meier, J. (1987). A fast algorithm for clusterwise linear absolute deviations regression. *OR Spectrum*, 9(3), 187-189.

Mendes, A. B. & Cardoso, M. G. (2006). Clustering Supermarkets: The role of experts. *Journal of Retailing and Consumer Services*, 13(4), 231-247.

Mishra, A. & Ansari, J. (2013). A conceptual model for retail productivity. *International Journal of Retail & Distribution Management*, 41(5), 348-379.

Mishra, R. K. (2009). Benchmarking scheme for retail stores efficiency. *International Journal of Marketing Studies*, 1(2), 131.

Park, Y. W., Jiang, Y., Klabjan, D., & Williams, L. (2017). Algorithms for generalized clusterwise linear regression. *INFORMS Journal on Computing*, 29(2), 301-317.

Pauler, G., Trivedi, M., & Gauri, D. K. (2009). Assessing store performance models. *European Journal of Operational Research*, 197(1), 349-359.

Poggi, J. M. & Portier, B. (2011). PM10 forecasting using clusterwise regression. *Atmospheric Environment*, 45(38), 7005-7014.

Preda, C. & Saporta, G. (2005). Clusterwise PLS regression on a stochastic process. *Computational Statistics & Data Analysis*, 49(1), 99-108.

Reilly, W. J. (1931). The law of retail gravitation. Knickerbocker Press New York.

Reinartz, W. J. & Kumar, V. (1999). Store-, market-, and consumer-characteristics: The drivers of store performance. *Marketing Letters*, 10(1), 5-22.

Rooij, V.G. (2017). *Clustering stores of retailers via consumer behavior*, Master's thesis, Erasmus University Rotterdam.

Sahibinden.com (2018), Rental Price Index Data. Available online: https://www.sahibinden.com/emlak-endeksi

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461-464.

Sellers, R. & Nicolau, J. L. (2011). Analysing the effect of environmental factors on service performance. *The Service Industries Journal*, 31(9), 1471-1488.

Sellers-Rubio, R. & Mas-Ruiz, F. (2006). Economic efficiency in supermarkets: Evidences in Spain. *International Journal of Retail & Distribution Management*, 34(2), 155-171.

Sharma, V. & Choudhary, H. (2010). Measuring operational efficiency of retail stores in Chandigarh Tri-city using DEA. *Journal of Services Research*, 10(2), 99-115.

Späth, H. (1979). Algorithm 39: Clusterwise Linear Regression. *Computing*, 22(4), 367-373.

Späth, H. (1981). Correction to Algorithm 39: Clusterwise Linear Regression. *Computing*, 26(3), 275.

Späth, H. (1982). Algorithm 48: A Fast Algorithm for Clusterwise Linear Regression., *Computing*, 29(2), 175-181.

Späth, H. (1986). Clusterwise linear least absolute deviations regression. *Computing*, 37(4), 371-377.

Thomas, R. R., Barr, R. S., Cron, W. L., & Slocum Jr, J. W. (1998). A process for evaluating retail store efficiency: A restricted DEA approach. *International Journal of Research in Marketing*, 15(5), 487-503.

Trivedi, M., Gauri, D. K., & Ma, Y. (2016). Measuring the efficiency of category-level sales response to promotions. *Management Science*, 63(10), 3473-3488.

TUIK (2018), Address Based Population Registration System Results. Available online: http://www.turkstat.gov.tr/PreTabloArama.do?metod=search&araType=vt

Turhan, G., Akalın, M., & Zehir, C. (2013). Literature review on selection criteria of store location based on performance measures. *Procedia-Social and Behavioral Sciences,* 99, 391-402.

Uyar, A., Bayyurt, N., Dilber, M., & Karaca, V. (2013). Evaluating operational efficiency of a bookshop chain in Turkey and identifying efficiency drivers. *International Journal of Retail & Distribution Management,* 41(5), 331-347.

Vaz, C. B., Camanho, A. S., & Guimarães, R. C. (2010). The assessment of retailing efficiency using network data envelopment analysis. *Annals of Operations Research,* 173(1), 5-24.

Vyt, D. (2008). Retail network performance evaluation: a DEA approach considering retailers' geomarketing. *International Review of Retail, Distribution and Consumer Research*, 18(2), 235-253.

Walters, R. G. & MacKenzie, S. B. (1988). A structural equations analysis of the impact of price promotions on store performance. *Journal of Marketing Research*, 25(1), 51-63.

Walters, R. G. & Rinne, H. J. (1986). An empirical-investigation into the impact of price promotions on retail store performance. *Journal of Retailing*, 62(3), 237-266.

Ward, J.H. (1963). Hierarchical grouping to optimise an objective function. *Journal of the American Statistical Association*, 58(301), 236-244.

Wedel, M. & Kamakura, W. A. (2000). Market segmentation: Conceptual and methodological foundations. Vol. 2, Springer Science & Business Media.

Wedel, M. & Kistemaker, C. (1989). Consumer benefit segmentation using clusterwise linear regression. *International Journal of Research in Marketing*, 6(1), 45-59.

Wedel, M. & Steenkamp, J. B. E. (1989). A fuzzy clusterwise regression approach to benefit segmentation. *International Journal of Research in Marketing*, 6(4), 241-258.

Wedel, M. & Steenkamp, J. B. E. (1991). A clusterwise regression method for simultaneous fuzzy market structuring and benefit segmentation. *Journal of Marketing Research*, 28(4), 385-396.

Xavier, J. M., Moutinho, V. F., & Moreira, A. C. (2015). An empirical examination of performance in the clothing retailing industry: A case study. *Journal of Retailing and Consumer Services*, 25, 96-105.

Xavier, J. M., Moutinho, V. M., & Moreira, A. C. (2015). Efficiency and convergence analysis in a women's clothing retail store chain: Evidence from Portugal. *International Journal of Retail & Distribution Management,* 43(9), 796-814.

Zhang, B. (2003). Regression clustering. *In Proceedings of the Third IEEE International Conference on Data Mining* (ICDM'03), IEEE, pp. 451-458.

Zhu, Z., Li, Y., & Kong, N. (2012). Clusterwise linear regression with the least sum of absolute deviations–An MIP approach. *International Journal of Operations Research*, 9(3), 162-172.

BIOGRAPHICAL SKETCH

Alper ALTINOY was born in Istanbul on February 27, 1990. He studied at Ihsan Mermerci Anatolian High School where he was graduated in 2008. He attended the undergraduate program of Industrial Engineering at Yıldız Technical University. He received his B.S. degree in the Industrial Engineering in 2013. He started his professional career as Production Engineer at Kelebek Mobilya in 2012. Afterwards he worked at Samet Kalıp for 3.5 years as Planning Chief. Currently, he is working towards master's degree in Logistics and Financial Management under the supervision of Assoc. Prof. Dr. Ebru Angun the Institute of Science and Engineering, Galatasaray University.

The papers that have appeared as full text in referred proceedings of international conferences:

Angun, E., Altınoy, A., 2019. "A new mixed-integer linear programming formulation for multiple responses regression clustering", *in Proceedings of the 6th International Conference on Control, Decision and Information Technologies*.