Advances in Methods to Support Store Location and Design Decisions

Auke Hunneman

Publisher:	University of Groningen Groningen The Netherlands	
Printer:	Ipskamp Drukkers B.V.	
ISBN:	978-90-367-4749-3 978-90-367-4758-5	 (e-book)

Advances in Methods to Support Store Location and Design Decisions

Proefschrift

ter verkrijging van het doctoraat in de
Economie en Bedrijfskunde
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus, dr. F. Zwarts,
in het openbaar te verdedigen op
maandag 7 februari 2011
om 14:45 uur

door

Auke Hunneman

geboren op 1 maart 1981
te Drachten

Promotor:	Prof. dr. T.H.A. Bijmolt
Copromotor:	Dr. J.P. Elhorst
Beoordelingscommissie:	Prof. dr. K. Campo Prof. dr. P.S.H. Leeflang Prof. dr. F. Ter Hofstede

Acknowledgements

This dissertation has benefitted significantly from the help and support I received from colleagues, friends, and family members while I worked on this project. In this section, I express my sincere gratitude to them.

Foremost, I am deeply indebted to my promoter Tammo Bijmolt. I could not have wished for a better mentor and coach to guide me through the process of writing a dissertation. Tammo: I thank you for your willingness to be my supervisor, for sharing your knowledge and passion, for your advice on how to play in the "Champions League," for helping me obtain data, and for asking questions such as, "Why are we actually doing this?" Above all, I very much appreciate you as person. Your accessibility, enthusiasm, and open-mindedness made it a real pleasure to work with you; I also look forward to working together in the future.

My co-promoter, Paul Elhorst, became involved in my dissertation project when we decided to apply spatial econometric methods. Paul is an expert in this field, and it is hard to describe his passion for the subject. He has always been available to answer (difficult) questions and help me understand the technical details of an econometric model. Paul: Thank you for sharing your knowledge and enthusiasm, as well as for your unlimited patience in explaining the details of the econometric models. I remember meetings in your office, writing "Greek" on the whiteboard, as well as moments of joy when the Matlab routines finally started working.

Furthermore, I thank the other members of the graduation committee, Katia Campo (Catholic University of Leuven), Peter Leeflang (University of Groningen), and Frenkel ter Hofstede (University of Texas at Austin) for evaluating my dissertation and giving me useful feedback. I especially thank Peter for all his support and guidance, as well as the pleasure of collaborating on projects not included in this dissertation.

The three empirical studies in this dissertation rely on high-quality data. In Chapters 3 and 4, we used data from the Dutch clothing chain Duthler, provided by Martijn Akkermans and Margaret Morren from Brova Groep. Marloes Grift (WDM Nederland) generously offered the geodemographic data for these studies. Erjen van Nierop wrote a program to obtain travel distances between zip codes from an online route planner. I sincerely thank each of them for making these data available for this academic research, as well as for their practical perspectives.

I owe many thanks to Matthijs Streutker for being my co-author on Chapter 2, which started as a joint project in the Essay Competition of the SOM Research Institute and Graduate School. The text has substantially improved in response to comments by Matilda Dorotic (Chapter 1) and Maarten van der Vlerk (Chapter 2). Chapter 3 benefitted greatly from comments made by the editor and reviewers of *Marketing Science*. I appreciate all of their contributions.

Another person who deserves special mention is Eelko Huizingh, who detected my interest in research when I assisted him in writing a Dutch marketing textbook. He and Maryse Brand immediately involved me in a joint research project about the drivers of ICT adoption by small- and medium-sized enterprises. Eelko also introduced me to Tammo, who became my advisor. In retrospect, these were my first steps into the world of “Marketing Science,” and I am grateful to Eelko for giving me these opportunities and for his help at the start of my academic career in Groningen.

Special thanks go to my former colleagues and friends at the Marketing Department in Groningen. It has been a great privilege to work in an environment that is so stimulating and sociable at the same time. Because of these surroundings, I actually enjoyed working on my dissertation and teaching all my courses, which is also the reason I look forward to my future visits to Groningen. I particularly thank my doctoral colleagues over the years, with whom I have enjoyed “small” informal chats at the coffee corner, interesting academic debates and discussions, and participations in various social events. Although I owe thanks to all of them, I especially mention Aljar, Matilda, Peter van Eck, Jacob and Alessio, Ernst, Hans, Sara, Stanislav, Stefanie, and Umut, with whom I also had the pleasure of sharing an office. I am also grateful to Marjolein, who took on many organizational issues associated with teaching large classes, which allowed me to concentrate on my teaching and finishing my dissertation. Finally, I express gratitude to the (former) secretaries of the Marketing Department, Hanneke, Jeannette, and Frederika, who were always available for administrative support and small chats.

Special thanks go to my friends and paranymphs Ernst and Stefanie. Ernst and I shared an office for a long time, along with interests and spare time: watching soccer matches, having dinner and drinks together, travelling in New York and Malaysia, and conversing about virtually everything. Although I have known Stefanie for a much shorter time, we became friends nearly immediately after meeting for the first time. I could not have wished for a better roommate while writing the last part of my dissertation. Ernst and Stefanie: Thank you for the nice times we have had together and for your help with preparing my defense. I hope we will be friends forever!

During my doctoral years, I also had the pleasure of meeting doctoral candidates from other universities who are now not just colleagues but also friends. I especially mention Carlos, with whom I enjoyed travelling in Arizona, and Isabel, for her enduring support and friendship over the years.

I also note the warm welcome from my new colleagues at the Marketing Department of BI Handelshøyskolen. Tor Andreassen and Kenneth Wathne gave me the opportunity to finish my dissertation and pursue my academic career in Oslo. Giulia involved me in many social activities outside of work, which has made it much easier for me to feel at home in Norway.

Finally, I thank my family. My parents and my brother Henk have always believed in me, supported me, shown great interest in my work, and motivated me to pursue my goals, even when I decided to move to Norway. Heit, mem, and Henk: I cannot thank you enough for what you have done for me. Your love and support is invaluable!

Oslo, November 2010

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation and Background

Since its foundation in 1965, Walmart has grown into one of the world's largest retailers, operating more than 8,000 stores and club locations in 15 countries. By the end of 2005, 46 percent of U.S. households lived within five miles of a Walmart store, and 88 percent lived within 15 miles of the nearest outlet (Basker 2007). Yet the company continues to formulate ambitious expansion plans, including its recent announcement that it would add approximately 38 million square feet of store space globally in 2010, which would require capital expenditures of $12.5–13.1 billion (Wal-Mart Stores Inc. 2009). The success of Walmart has and continues to depend heavily on its careful selection of geographical markets for new stores; Walmart's entry to a market also has significant implications for incumbent retailers, because it effectively lures away some of their best customers (Ailawadi et al. 2009; Basker 2007; Gielens et al. 2008; Singh, Hansen, and Blattberg 2006).

This example illustrates some important trends in the modern retail industries of Western economies. Many retailers have increased the number of their outlets to reach more consumers (Table 1.1), resulting in high levels of retail store concentration (Bucklin, Siddarth, and Silva-Risso 2008; Pal and Sarkar 2002). One of the implications of this development is saturation in many markets, which makes good locations ever more scarce and hard to obtain. Competitive pressures and rising land and property costs make location decisions critical to a retailer's financial strategy as well (Levy and Weitz 2004; McGoldrick 1990). Recently, General Motors, Ford, and Chrysler all decided to minimize their dealer networks to reduce costs (Bucklin, Siddarth, and Silva-Risso 2008), even as companies such as Walmart and Toyota, whose sales in the United States are still growing, might add new outlets to existing distribution networks. With their significant profit implications, both types of decisions (i.e., locations to open or to close) should be based on an analysis of the relationship between distribution and sales (Haans and Gijsbrechts 2010).

Table 1.1: Number of stores operated by the 10 largest U.S. retailers (annual revenues, 2009)
Source: www.stores.org.

Company	Revenue	Yearly Revenue Growth	Number of Stores	Growth in the Number of Stores
Walmart	$405,607,000	7.2%	7873	8.4%
Kroger	$76,000,000	8.2%	3654	-0.2%
Costco	$72,483,020	12.6%	544	5.0%
Home Depot	$71,288,000	-7.8%	2274	1.8%
Target	$64,948,000	2.5%	1682	5.7%
Walgreen	$59,034,000	9.8%	6934	15.6%
CVS Caremark	$48,989,900	8.7%	6981	10.8%
Lowe’s	$48,230,000	-0.1%	1649	7.5%
Sears Holdings	$46,770,000	-7.8%	3918	1.8%
Best Buy	$45,015,000	12.5%	3942	200.0%

Despite the commonly held belief that distribution intensity is a key marketing contributor to a company’s sales and market share, empirical support for this notion remains scarce (c.f. Bucklin, Siddarth, and Silva-Risso 2008). The effect sizes for other elements of the marketing mix, such as price (Bijmolt, Van Heerde, and Pieters 2005) and advertising (Assmus, Farley, and Lehmann 1984), are much better documented. Yet the importance of quantifying the performance implications of each element in the marketing mix, typically referred to as marketing accountability, has been emphasized recently by Verhoef and Leeflang (2009), who show that if marketing’s contribution to firm performance cannot be made clear, other departments take decision-making roles formerly performed by marketing.

Although they argue this shift is not necessarily negative, efforts to coordinate marketing activities may be negatively affected by it.

In turn, the coordination of marketing activities is of crucial importance, particularly for location decisions, because the location of a store determines its trade area and consequently the type of consumers most likely to visit it (Briesch, Chintagunta, and Fox 2009; Chan, Padmanabhan, and Seetharaman 2007; Pan and Zinkhan 2006). Without sufficient numbers of consumers in the trade area and a good match with the store's positioning, profitable sales levels are hard to obtain (González-Benito, Bustos-Reyes, and Muñoz-Gallego 2007; Inman, Shankar, and Ferraro 2004). Differences in product category appeal across retail stores result from variations in consumer and competitor characteristics in local markets, such that each store creates opportunities for the location-specific allocation of store space. Campo and Gijsbrechts (2004) and Campo et al. (2000) show that retailers tailoring their assortments to the needs and wants of local consumer groups can improve chain profits dramatically. Thus, the best store locations are not necessarily those with the highest population densities but those frequented by consumers who find the store attractive. Retailers thus must consider consumer profiles and spatial locations when selecting new store locations and entering new geographical markets.

For a long time, early methodologies developed by Applebaum (1966) and Huff (1964) provided the basis for academic research into store location evaluations. Yet these approaches do not account for consumer heterogeneity. Recent developments in methodology and computing power have helped overcome these shortcomings and offer new opportunities for store location research. First, the widespread adoption of customer loyalty cards enables retailers to gain insights into the spatial distribution of their stores' sales. Second, the accessibility of geodemographic information through (commercially available) databases has increased. These two developments, coupled with advances in methodology, including the application of spatial econometrics to the marketing field and improvements in geographic information system (GIS) technology, allow retailers to investigate who buys their merchandise and how these decisions depend on consumer characteristics and their location relative to the store.

This dissertation builds on these developments by proposing and testing new models for store location evaluations. Using state-of-the-art methodologies, we link customer-level sales data from customer databases with commercially available information about consumer demographics at the zip code level, to build models that (1) assess the impact of various drivers on store sales; (2) evaluate the expected performance of stores; and (3) predict sales impacts of changes in assortment, location changes, and new store openings. Therefore, in this introductory chapter, we begin in Section 1.2 by presenting the central research problem of this dissertation, as well as a brief review of location models used in practice. Section 1.3 offers an outline of this dissertation. This chapter ends with a discussion of how the different chapters relate to one another.

1.2 Research Questions

This dissertation investigates the central research problem of evaluating locations for opening new retail outlets and aims to answer the practical question of where a retailer should open new stores. As such, the central question is:

> *What is the impact of store location and assortment design on store performance, and through which methods can insights about these relationships be used to support store location and design decisions?*

This central question comprises several research questions addressed by the various chapters of this dissertation:

1. *What is the optimal number of stores for a particular market?*
2. *What are the best store characteristics (assortment, store size) for each particular site?*
3. *Which factors drive the performance of individual retail stores?*
    - *Do these drivers of store performance differentially affect each component of store sales?*
    - *What are the performance implications of changes in the retail environment of a store?*
4. *How do existing stores perform in comparison with the sales potential of a particular location?*

To answer these questions, we use econometric models, applied in three empirical studies presented in Chapters 2, 3, and 4. In the current section, we define the optimal store location in the context of this dissertation, followed by a more detailed discussion of the research questions addressed in each of the subsequent chapters.

Location Research

Location research in a strict sense pertains to the identification and selection of store locations that optimize certain objectives, such as the minimization of transportation costs, optimization of consumer service, or profit maximization (Shang et al. 2009). In addition to literature on site selection, store performance research offers another domain of interest that focuses on identifying the drivers of store performance. Finally, literature on local marketing, which belongs to the latter stream of research, explores the potential for tailoring a store's marketing mix to local conditions, thereby exploiting spatial variation in consumer characteristics and the competitive environment. The contribution of this dissertation lies at the intersection of these three streams. The proposed models can identify and evaluate new store locations, but they have greater applicability as well. The models also can help evaluate store performance, find a store's optimal assortment for a particular location, and assess the sales impacts of changes in the retail environment of individual stores.

Specifically, in Chapter 2 we propose a model to support the location decisions for multi-unit firms—firms that consist of two or more retail outlets. Location decisions for such networks can be rather complicated because of the multitude of potential interactions among individual stores, which have performance implications for the whole network. The model in Chapter 2 is an optimization model, based on the mixed integer linear programming paradigm, which determines the optimal number of stores, store locations, and store sizes to maximize overall firm performance. Chapter 2 thus (mainly) addresses research questions 1 and 2 with an empirical study on the location of new health clubs in the Rotterdam area. The model assumes that consumers choose among alternative stores on the basis of the travel distance to the store and store size. This rather simplistic assumption suggests that this model should be used in combination with the richer models proposed in Chapters 3 and 4. Nevertheless, even this relatively simple approach can generate better solutions than some benchmark models. This property of the model, together with its specification flexibility, effectively supports its implementation in marketing practice.

In Chapter 3, we adopt a modeling framework to explain several components of store sales. In particular, we decompose sales to loyalty program members within a particular zip code and specify separate models for, among other things, the penetration rate of the loyalty program, the number of visits, and average expenditures per visit. By relating these variables to store, competitor, and consumer characteristics, we provide more detailed insights into the drivers of store sales relative to commonly used sales models. Because it uses geodemographic data at the zip code level, this model also captures the sales effects of observed spatial differences in store, competitor, and consumer characteristics better than existing models, even as it accounts for unobserved sources of spatial heterogeneity in store sales. The high predictive performance of this decomposition model underlines its value for store location and evaluation decisions. Chapter 3 thus mainly addresses research questions 3 and 4 with an empirical study of a Dutch clothing chain.

Chapter 4 extends that model by simultaneously considering store location and assortment composition, two important strategic decisions rarely considered simultaneously in existing research. We propose several models that explain not only explain a store's overall performance but the performance of each individual store department. In this empirical study, we show that the proposed models can forecast the performance of new stores reasonably well, which makes it a useful tool for store location evaluation (research question 4). The performance of the individual store departments also is affected differentially by consumer and competitor characteristics, such that a retailer that wants to improve the performance of individual stores can do so by assigning more space to departments that are more appealing to the local market (research question 2). As a final application of the model, we evaluate the performance of existing stores (research question 4).

Demarcations

Although the focus of this dissertation is on the Dutch retail industry, location decisions occur in a vast variety of settings (e.g., health clubs, restaurants, banks), and the application of the proposed models is thus not limited to the retail sector. Each time a model is applied in a new setting, the researcher must decide about the composition of the explanatory variables. Whereas some variables may be common to all businesses, others more likely reflect the firm's particular type of industry, which may call for the introduction of additional variables not used in the presented studies (e.g., Kumar and Karande 2000; Pan and Zinkhan 2006). However, including additional variables in the proposed models would be straightforward and easy to accommodate. Furthermore, the location models are intended as tools to support retailers' decisions about where to locate new stores. As such, the model offers no clear-cut answer to the question of whether a company should enter a particular market. Rather, the retailer must decide whether it wants to invest in a particular location, after considering the model's predictions about its potential sales and making managerial judgments about the costs.

Managerial Implications

This dissertation thus has clear managerial relevance. Although most retailers acknowledge the importance of location decisions (Clarkson, Clarke-Hill, and Robinson 1996), few use the more advanced techniques, such as multiple regression and location allocation models, that are at their disposal (Hernandez, Bennison, and Cornelius 1998). The complexity of existing models and their requirements for significant quantitative modeling knowledge hinders retailers' efforts to apply such models. A survey about whether retailers use location models, by Hernandez and Bennison (2000), indicates that approximately 40 percent of retailers use regression models, 39 percent use gravity models, and only 16 percent employ neural networks. Prior experience seems the most important factor in location decisions for a new store. Although this approach may have been sufficient in the past, the highly dynamic retail environment in which today's retailers operate calls for rigorous and systematic approaches to location decisions. This point is not to say that we no longer need managerial judgements; even using a model, the final decision still must depend on whether the retail manager believes the model predictions are accurate. However, the use of multiple site evaluation methods, in which complementary models mitigate the weaknesses of any one approach, typically results in the best estimate of a site's sales potential.

In this dissertation, we develop several models to support location decisions, all of which can be adjusted easily to a wide variety of settings, which enhances their applicability in practice. We hope this dissertation and developments such as the greater supply of high quality data and increased user-friendliness of GIS, further stimulate the application of location models in marketing practice.

1.3 Data

Central to the development and use of location models is the availability of high quality data. This dissertation combines a rich variety of data sets and thus exploits recent advances in data gathering by retailers (customer loyalty cards), private vendors (demographics at the zip code level), and government data providers (Census data). In Chapters 3 and 4, we use data on Dutch clothing stores from one chain to calibrate the market response models, which explain different components of store sales. The data used in these chapters therefore include:

1. Point of sale (POS) scanning data, at both the individual customer and store levels. These data provide addresses of individual customers and detailed information about all their purchases (when, what, and where they purchased).
2. Geodemographic data at the zip code level. This source of data includes information on consumer socio- and demographics and household composition, among other things, for all four-digit zip codes in the Netherlands.
3. Survey data on the retail environment of each individual store, including information about store ages, store sizes, and competitors.

In Chapter 2, we use a different data set that combines information about the locations and sizes of the individual health clubs (from the company's database), financial performance indicators at the company level (Amadeus database), rental prices in the Rotterdam area (www.realnext.nl), and population sizes for each four-digit zip code in the Rotterdam area (Statistics Netherlands).

1.4 Methods

Whereas Chapters 3 and 4 use spatial econometric techniques to model (components of) store performance, Chapter 2 takes a different perspective. For the study reported in Chapter 2, we specify and test a "competitive facility location" model (Drezner 1995; Plastria 2001) that can be solved as a mixed integer linear programming (MILP) model. To the best of our knowledge, this study is one of the earliest applications of the MILP paradigm to competitive facility location problems; it performs better than similar benchmark models, which makes it a useful tool for supporting location and design decisions in practice.

Chapter 3 uses spatial random effects models extended with a spatial error autocorrelation specification to analyze the relationship among several components of store sales and their drivers. Compared with the ordinary random effects model, these models better capture the effects of potential omitted variables that might drive the response variables. Consumers living in close proximity often share some unobserved characteristics, so the error terms of neighboring observations are likely to be spatially correlated. In these situations, spatial error models obtain more efficient and reliable parameter estimates than do models that ignore these effects (Anselin 1988). The use of seemingly unrelated regressions accounts for the potential simultaneity between a particular location's sales potential and the size of its trade area, as well as the local number of competitors.

Finally, in Chapter 4, we consider a store as a composite of several departments and therefore use a multivariate spatial model to explain each department's share of sales. Specifically, a market share model specification determines the attractiveness of each department in a particular zip code. In addition, a Tobit model explains overall sales per zip code to account for censoring of sales data. Decisions about the optimal (relative) size of each department and its share of sales are unlikely to be completely independent; therefore, we use separate equations to disentangle these effects.

1.5 Synthesis: How the Chapters Relate to One Another

Several authors have emphasized that a retailer's location strategy typically involves decisions at three different stages, which differ in the level of aggregation at which they approach the problem (Levy and Weitz 2004; McGoldrick 1990; Mendes and Themido 2004). The first stage pertains to the identification of geographical areas, such as a country, region, or a particular city, that have location potential for new outlets. At this stage, the retailer also determine the (optimal) number of stores to open in the selected regions by weighting the potential benefits from economies of scale versus the risk of cannibalization (Kalnins 2004). The second step is to define and evaluate the trade area, which refers to the contiguous geographic area from which the store(s) will get the majority of their sales and customers. Although these two stages are separate conceptually, the factors affecting the attractiveness of a particular region and trade area are the same. To evaluate overall demand in a particular region/trade area, retailers consider the size and profile of the local population and the (anticipated) level of competition. Finally, the last stage of the decision hierarchy involves the evaluation and selection of individual sites within the chosen region(s). The most important factor in choosing a particular site is the amount of sales it can generate; therefore, the retailer estimates the sales potential of each site. Several methods attempt to predict the potential sales for (new) store locations, including analog, gravity, and regression models, as well as spatial allocation models (Buckner 1998; Craig, Ghosh, and McLafferty 1984). The third level of the location strategy also includes decisions about store characteristics for the chosen sites. It becomes increasingly difficult for retailers to differentiate themselves just geographically (Iyer and Seetharaman 2008; Thomadsen 2007), so they attempt to gain distinctiveness by finding the right match between their stores' positioning and the demographic profile of the trade areas in which that are located. We illustrate this three-stage decision-making process in Figure 1.2.

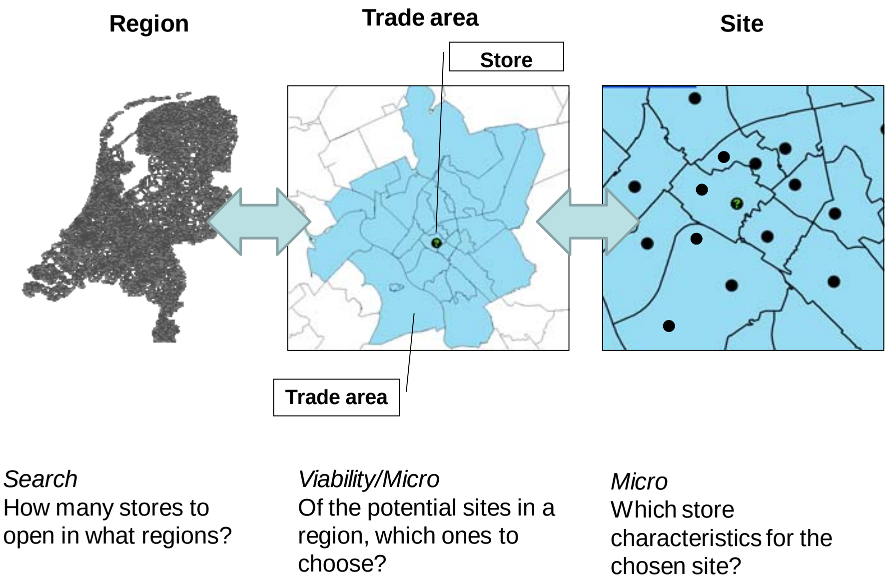


Figure 1.1: The store location decision process.
Adapted from Levy and Weitz (2004); McGoldrick (1990); Mendes and Themido (2004).

A similar decision sequence, though with different terminology, has been proposed by Bowlby, Breheny, and Foot (1984), who argue that retailers proceed through the (1) search, (2) viability, and (3) micro stages. The first decision in an "ideal" retail location strategy identifies geographical regions for locating new outlets (search), followed by finding the best sites in these areas based on forecasts of the amount of sales available (viability). The last step (micro) examines all factors that drive potential sales at a particular site; it thus relates closely to the previous two stages.

Although these subdivisions of location decisions are appealing from a theoretical perspective, the problems addressed at each level of the decision process are not independent. Levy and Weitz (2004) therefore rightfully conclude that retailers should always consider these issues together when designing their location strategies. For example, a region at first sight may seem very attractive because of the large turnover it generates, but if suitable sites are lacking due to strict government planning policies, the retailer must locate its stores elsewhere.

Although the models proposed in the next chapters all deal with the same general problem—finding the best store locations from a retailer perspective—they differ in a few important ways (Table 1.1). A certain specification makes a model more (or less) appropriate for particular stages of the decision process. In the remainder of this section, we therefore discuss the differences and similarities of the proposed models, along with suggest the level(s) of the decision process at which each model can be applied successfully.

Table 1.2: Differences between models presented in this dissertation

	Location Allocation Model Ch. 2	**Regression Model Ch. 3**	**Regression Model Ch. 4**
Intended use	Prescriptive/ Predictive	Descriptive/ Predictive	Descriptive/ Predictive
Objective Function	Chain Profits	Store Sales Components	Department-Level Store Sales
Number of Stores	Endogenously Determined	1- Few	1-Few
Marketing Problems Considered	Site Selection & Outlet Size	Site Selection	Site Selection & Assortment Composition
Consumer Heterogeneity	Small	Large	Large
Demand/Supply	Demand & Supply	Demand	Demand
Decision Stage	Region/Trade area	Site/Trade area	Site/Trade area

The models in the subsequent chapters can be classified along several dimensions. One dimension is the methodology involved. We adopt the typology used by, among others, Buckner (1998) and Craig, Ghosh, and McLafferty (1984) and thus classify the models in the first two chapters as (advanced) regression models. The model in Chapter 2 is a more recent innovation, known as a location allocation model, which has roots in operations research. Specifically, the competitive facility location model is based on (assumed) consumer preferences and simultaneously considers the location and design of multiple stores belonging to the same company.

Location allocation models take a rather different focus than regression models, which appear in Chapters 3 and 4. Whereas regression models are particularly useful for determining the impact of several drivers of store performance and thus evaluating individual sites (stage 3), location allocation models typically help assess locations at the region or trade area level. As a major advantage of the latter models, they enable the researcher to evaluate systematically many alternative configurations of store networks (Achabal, Gorr, and Mahajan 1982; Ghosh and McLafferty 1982; Ghosh and Craig 1983) and select the one that maximizes a specified objective at the company level. This capability is particularly important when a chain aims to open multiple stores in the same market—an increasingly common strategy in the U.S. market, in which companies with four or more stores account for much of the total retail industry (Pal and Sarkar 2002). The model in Chapter 2 also can be adjusted easily to apply to franchisors' decision-making problems, whose goal of maximizing system-wide profits often is at odds with the individual franchisees' goal of maximizing their own revenues and profits (Ghosh, McLafferty, and Craig 1995). Because sales through franchise systems have grown to nearly 3.5 percent of United States gross domestic product (GDP) (Kalnins 2004; Lafontaine and Shaw 1999), it has become increasingly important to develop a model to support franchise expansion decisions.

Another important distinction between the models is that those in Chapters 3 and 4 can evaluate only a limited number of candidate locations, whereas the model in Chapter 2 endogenously determines the optimal number of stores and their locations. This model is particularly useful for a retailer that must decide about the optimal number of stores to build and that has access to many desirable sites in a particular region. Because the optimal locations of n stores are not necessarily a subset of the optimal $n+1$ stores, it is necessary to account for the impact of each store on the entire network. Although location allocation models have increased in richness and sophistication, their treatment of consumer patronage decisions remains quite limited; sometimes, they simply assume that consumers visit the nearest store, as early full-capture models did (Serra and Revelle 1995). However, with these simplifying assumptions, existing location allocation models largely ignore the heterogeneity of consumers who constitute the store's trade area. Conversely, in Chapter 2 we assume that consumers make a trade-off between travel distance to the store and store size, which provides a proxy for assortment size. Regression models enable the retailer to identify relationships between store sales and a large set of predictor variables, including those that measure consumer characteristics and competition. These models thus are ideal for heterogeneous markets in which each retailer (or store) has its own positioning to target a particular consumer group.

Chapters 3 and 4 examine the location problem from the demand side, which means that these models can assist retailers in predicting the amount of sales an individual site can generate, but they do not consider the costs associated with building and operating stores at these locations. The ultimate aim of a retailer is to maximize the company's overall profit; therefore, costs should be taken into account. The model in Chapter 2 incorporates personnel and housing costs; the objective function there thus maximizes overall chain profit. Although cost data might be a useful addition to the models in Chapters 3 and 4, information on costs is relatively easy to obtain, which obviously does not hold for sales estimates for store locations.

The models of Chapters 2 and 4 also are similar in that they both consider store design, in addition to store location. As it becomes increasingly difficult for retailers to distinguish themselves from competitors by store location, they look for other ways to do so. In Chapter 4, we show that retailers can enhance store-level sales by adjusting their assortment to local conditions, whereas in Chapter 2, we find that changes in consumers' sensitivity to store size relative to travel distance significantly affects the spatial configuration of the optimal store network and total chain profits.

In summary, despite their similarities, the models in this dissertation differ in some important respects, including the level of the store location decision sequence to which they apply. The location allocation model (Chapter 2) is useful for a retailer that wants to open more than one store in a market but does not know how many stores to open and where to locate them. In general, this model addresses the search phase or *identification* of potential new store locations, whereas the regression models (Chapters 3 and 4) deal with the (detailed) *evaluation* of particular sites. We propose the use of both types of models sequentially, to determine the optimal number of stores and potential locations using the location allocation model, then evaluate each individual site carefully with the regression model of Chapter 3. Finally, the assortment can be adjusted to fit local market conditions according to the model developed in Chapter 4.

Chapter 2

Location and Design of Multiple Stores in a Competitive Environment: A Mixed Integer Linear Programming Approach[1]

2.1 Introduction

One of the fundamental decisions a retailer must make is the location of new stores. The importance and complexity of such decisions have increased, due to the rapid growth of multistore networks (Ghosh, McLafferty, and Craig 1995). Today's retail industries are characterized by high levels of concentration. In the United States for example, retailers with four or more stores account for more than half of total retail business (Pal and Sarkar 2002). Location decisions are highly complicated for networks of stores operated by a single firm, because they require the systematic evaluation of the impact of each store on the entire network and the consideration of interactions among stores (Ghosh, McLafferty, and Craig 1995). From a consumer perspective, distance to the store is not the only factor driving store choice; differences in designs across stores play an important role as well. Consumers who choose among various competitors offering a more or less identical products do not automatically visit the nearest store (see, e.g., Briesch, Chintagunta, and Fox 2009; Chernev and Hamilton 2009); they also consider, for example, multipurpose traveling, comparison shopping, and differences in prices, goods, assortments, and images among stores (for an overview, see Pan and Zinkhan 2006). Consequently,

[1] This chapter is based on a working paper with the same title, co-authored by Matthijs Streutker.

configuring multistore networks is a nontrivial task that requires the simultaneous optimization of store location and design in a competitive environment. In this chapter, we propose a model to support these decisions for a chain of stores. The model contributes to extant literature because it can be solved as a mixed integer linear programming (MILP) model, which is very flexible in its specification and can be solved effectively for instances with realistic sizes.

The remainder of this chapter is organized as follows: First, we summarize the relevant literature (section 2.2). In section 2.3, we formulate our general competitive facility location and design problem model. Some extensions of the model, which explain the behavior of consumers or the preferences of the company in more detail, also appear in this section. In section 2.4, we illustrate the model with a case study that describes the location and design of luxury leisure clubs in the Rotterdam area of the Netherlands. We also provide intuitions about model behavior through a comparison with the classical full capture model[2] and a sensitivity analysis of the findings for the store size sensitivity parameter. The concluding remarks and directions for further research are in section 2.5.

2.2 Literature Overview

This chapter considers the competitive facility location problem: Finding the optimal sites for several retail outlets operating in a competitive environment. In principle, it is a multi-agent optimization problem, where the agents are the retailer, the consumers, and competitors. All agents have their own (conflicting) goals (e.g., profit maximization for the retailer and competitors, utility maximization for consumers) and make their own decisions (e.g., store location/size, store choice). This complicated problem cannot be solved easily, so simplifying assumptions about the behavior of consumers and competitors are necessary. Many models

[2] This model also has been referred to as the maximum capture (MaxCap) model. We prefer the term full capture model.

assume competitors are already present in the market, and their locations and other characteristics remain fixed for the time period considered. We restrict ourselves to this static setting as well (for an overview of similar competitive facility location models, see Plastria 2001). Thus, there are no competitor decisions, and the only two actors are the retailer and consumers.

The question then becomes how to incorporate consumer decisions into models that optimize the locations and design of new stores. Consumers usually concentrate in a finite set of points, such that each point represents a geographical entity, such as a region, city, or zip code. Patronage likely depends on utility functions that measure the attractiveness of a store to a consumer, generally based on travel distance and perhaps other attraction factors, such as store size and (perceptions of) the store's price level. In full capture models, as initiated by Revelle (2006), each consumer patronizes the most attractive store for him or her. Therefore, all consumers concentrated in a demand point visit the same store. This so-called deterministic patronizing behavior represents a strong assumption; alternatively, in spatial interaction or gravity models, all stores are visited by the group of consumers represented by a particular demand point in proportion to their relative attractiveness. These models, introduced by Huff (1964), feature probabilistic patronizing behavior, and Santos-Peñate, Suárez-Vega, and Dorta-González (2007) and Serra and Colomé (2001) describe the differences in the two classes. In particular, both store choice, given attractiveness to consumers, and the attraction function itself can be specified in several ways. In optimization models, the most widely exploited formulations are additive and multiplicative, for which the attractiveness of a store to a consumer depends on several factors. As the names suggest, a multiplicative attraction function uses the product of weighted factors, whereas the additive version calculates the sum. Discussions of these options appear in Drezner, Drezner, and Eiselt (1996) and Drezner and Eiselt (2002).

Important developments on the retailer side of the problem include the less spatially restricted location decisions and endogenous determination of the number of stores to be located. Originally, competitive facility location problems were formulated in discrete space, such that the number of candidate sites were determined in advance, and the store(s) were located in a subset of these sites. The network approach is more flexible; stores can be located at a node or along an edge of a predefined network (e.g., Suárez-Vega, Santos-Peñate, and Dorta-González (2004). More recent models propose a continuous planar space, such that the facilities can appear anywhere on the plane defined by the demand points. Fernández et al. (2007) offer a spatial interaction continuous model; Plastria and Carrizosa (2004) outline a full capture version. Introducing design as a decision variable in competitive facility location models also changed the model set up, because the number of facilities became endogeneous instead of exogeneous. Without design questions, the user specifies the number of facilities, with the objective of maximizing revenues. Without any costs involved, the endogeneous determination of the optimal number of facilities results in a facility at each potential location. However, because design induces costs, the objective changes to maximizing profits for multiple facilities, determined endogeneously.

The competitive facility location and design problem (CFLDP) thus is difficult to solve, regardless of the specific formulations—deterministic or probabilistic, discrete, network, or planar. Until recently, models could evaluate only the performance implications of a (finite) number of location and design alternatives. Enumeration schemes can be effective for the location and design of a single facility when the number of possible locations is limited and the optimal size for a certain location can be computed effectively (Eiselt and Laporte 2006). Furthermore, recent effective optimization methods support specific formulations of planar single-facility location and design problems, such as the maxcovering-minquantile problem reformulation by Plastria and Carrizosa (2004), disk covering reformulation by Zhang (2001), Weiszfeld-like algorithm and interval branch-and-bound algorithm by Fernández et al. (2007), and universal evolutionary global optimizer by Redondo et al. (2009). Multiple facilities also complicate the planar case; to the best of our knowledge, Tóth et al. (2009) are the only ones to consider this problem for two facilities. They apply the interval branch-and-bound algorithm, but when demand points increase in number to 50–100, this algorithm takes many hours to solve. Moreover, discrete competitive facility models also are complicated by the introduction of design variables. Therefore, many authors consider design fixed (e.g., Benati 1999; Benati and Hansen 2002; Colomé, Lourenço, and Serra 2003), and contributions that pursue "true" design and location optimization for multiple stores are rare. Zhang and Rushton (2008) suggest a genetic algorithm to solve the problem, but they cannot provide numerical results. Suárez-Vega, Santos-Peñate, and Dorta-González (2004) solve a network CFLDP for multiple facilities by combining a global search procedure with three combinatorial heuristics. Unfortunately, the optimality of their solutions is not clear. The tangent-line approximation algorithm by Aboolian, Berman, and Krass (2007), with finite design alternatives, provides this information; specifically, they find optimal solutions for instances with 60–80 nodes and two design alternatives within 90 minutes.

We propose in turn a full capture model for locating and designing multiple stores in a competitive environment, which can be solved as a mixed integer linear programming (MILP) model. This approach not only allows for a continuum of design alternatives but also can include all kinds of constraints, which explain either the behavior of consumers or the preferences of the retailer. We have sufficient knowledge about solving MILP models, such that instances of considerable sizes can be solved effectively. Moreover, this well-known paradigm is widespread in decision-oriented applications, which eases model adoption. Nevertheless, we break in some sense with current research traditions in CFLDP, because we model deterministic, instead of probabilistic patronizing behavior. Furthermore, the attraction function is additive instead of multiplicative, and our location space is discrete instead of planar. A probabilistic model for store patronage seems realistic, but in the end, what matters most is whether the model provides good location and design suggestions. The objective of CFLDP models is not to forecast store patronage, which instead is the next step in a location study that uses detailed (regression) models and can include many (explanatory) variables. Store patronage obviously differs between probabilistic and deterministic models. For example, in a location-only study, Serra and Colomé (2001) show that with uncertainty about true patronage behavior, the deterministic model provides the most robust solutions. In addition, regarding the specification of the attraction function, classic utility theory suggests a multiplicative function is more appealing than an additive one. However, Drezner, Drezner, and Eiselt (1996) argue that the additive rule is more consistent, because with the additive rule, preferences do not change en route. Although planar CFLDP thus is an intriguing mathematical problem, we believe that for practical implementations, the discrete model suffices. We therefore discuss the specifications of a general CFLDP model in more detail.

2.3 Model

Consider a retailer that wants to open several new stores in a certain region. This retailer needs to know (1) how many stores to open, (2) where to locate these stores, and (3) the optimal size for each store. These new stores will compete for consumers with the stores of competitors that already operate in this region. We assume competition is static and known in advance. The retailer aims to maximize profit with these store openings; we assume it does not have any incumbent stores in the region yet, though this assumption easily can be relaxed. Demand and potential store locations may coincide and are represented by points. We use travel distance to calculate distances between each pair of points, which indicates the actual distances consumers travel by car to get from one point to another.

A demand point represents a populated area, such as a village or city district. Demand for products sold at these stores should be inelastic, such that the volume of demand does not depend on product quality. Consumers exhibit deterministic patronage behavior and choose the store most attractive to them. We model the attractiveness of a store with an additive utility function that consists of two explanatory variables: travel distance and store size.

Store points can be population areas as well, but other strategic locations in the region to be specified by the retailer. Each store point is as a candidate site for a new store, and they differ in costs. Some store points are occupied by stores of competitors; no store point can support more than one store, so they comprise potential locations for new stores and locations of competitors. Relaxing this condition would lead to a novelty-oriented tie-breaking rule (for a discussion on tie-breaking rules in full capture models see, e.g., Plastria 2001), because building a slightly bigger store at the same point as a competitor means capturing all the clientele from the competitor's store.

Notation

For the formulation of the model we use the following notation:

Sets

P set of all points, with the following subsets:

P^d set of demand points,

P^p set of possible store points, and

P^c set of competitor store points.

Note that $P^p \cap P^c = \varnothing$ by the one store per store point assumption.

Variables

x_j = 1 if a store opens at possible store point j, and 0 otherwise, $j \in P^p$; for

$j \in P^c$, x_j is a parameter with $x_j = 1$.

s_j = size of the store at possible store point j, $j \in P^p$; for $j \in P^c$, s_j is a parameter.

x_{ij} = 1 if a consumer of demand point i shops at store point j, and 0 otherwise, so $i \in P^d$, $j \in P^p \cup P^c$.

z_{ij} = size of the store where a consumer of demand point i shops, $i \in P^d$, and $j \in P^p \cup P^c$.

y_j = revenues in store point j, $j \in P^p \cup P^c$.

Parameters

c_j = variable cost per unit size for store at store point j, $j \in P^p$.

s = minimum size of a store.

d_{ij} = distance between demand point i and store point j, $i \in P^d$, and $j \in P^p \cup P^c$.

β = size sensitivity parameter.

A_i = decisive attraction for consumers at demand point i, $i \in P^d$.

δ_i = demand of consumers in demand point i, $i \in P^d$.

M = a large number (maximum realistic size of a store).

Formulation

Before we can formulate the objective and constraints, we need to know consumers' attractions to the incumbent stores before the store opens. First, for each consumer, attraction to each store is calculated using the auxiliary parameter A_{ij}:

$$A_{ij} = -d_{ij} + \beta s_j,$$

for all $i \in P^d$ and $j \in P^c$. Note that for $j \in P^c$, s_j is a parameter. The attraction a consumer has toward a certain store depends linearly on the distance required to travel to it and its size. Note that β relates store size to travel distance. In most CFLDP models, it is the other way around: distance gets discounted by design attractiveness. We choose this formulation instead, because with it we can easily turn off the size effect, which facilitates the comparison of the results with the location-only full capture model when we set β to 0. Second, the consumer visits the store with the highest attraction. Following Plastria and Carrizosa (2004), we call this attraction the decisive attraction of a consumer, given by

$$A_i = \max_{j \in P^c} A_{ij}, \tag{2.1}$$

The location constraints are as follows:

$$x_j \in \{0,1\}, \text{ and} \tag{2.2}$$

$$s_j \geq s x_j, \tag{2.3}$$

for all $j \in P^p$. Constraint 2.2 requires that at a certain point j, either a new store opens ($x_j = 1$) or not ($x_j = 0$). We also specify a minimum size for new stores. If a store opens at a certain location ($x_j = 1$) its size must be at least s (Constraint 2.3).

The locations of new stores together with the configuration of incumbent stores determine shopping possibilities, which submit to the following constraints:

$$x_{ij} \in \{0,1\}, \tag{2.4}$$

$$x_{ij} \leq x_j, \tag{2.5}$$

$$z_{ij} \leq s_j, \text{ and} \tag{2.6}$$

$$z_{ij} \leq M x_{ij}, \tag{2.7}$$

for all $i \in P^d$ and $j \in P^p \cup P^c$. Constraint 2.4 requires that a consumer spends his or her entire budget in a certain store or nothing at all. A consumer can only shop at a certain location if there is a store located there, as noted in Constraint 2.5. The size of the store where a consumer shops cannot be larger than the actual size of the shop at that location (Constraint 2.6), and store choice and size choice are linked (Constraint 2.7).

Next, the following constraints ensure that consumers are allocated to stores only if they are at least as attractive as the currently most attractive store:

$$\sum_{j \in P^p \cup P^c} x_{ij} = 1, \text{ and} \tag{2.8}$$

$$\sum_{j \in P^p \cup P^c} \left(-d_{ij} x_{ij} + \beta z_j \right) \geq A_i , \tag{2.9}$$

for all $i \in P^d$. Equation 2.8 requires that each consumer is assigned to exactly one store location, and Equation 2.9 ensures that consumers can be assigned to a new store only if their attraction for it is greater than that for the old situation (without new stores).

The choices of the consumers determine the revenues generated at each store location, given by:

$$y_j = \sum_{i \in P^d} \delta_i x_{ij} ,$$

for all $j \in P^p \cup P^c$.

The objective function the retailer wants to maximize is:

$$\sum_{j \in P^p} \left(y_j - c_j s_j \right).$$

Profits are equal to total chain revenues minus costs, which depend on store sizes.

Remarks

This model maximizes the retailer's profits through a systematic evaluation of a number of potential locations in a particular market; in this process, it allocates consumers to stores. The stores to which a consumer can be allocated are determined by Equation 2.9. It is easy to see that a consumer is not necessarily allocated to the store for which he or she has the highest attraction. However, if consumers visit the stores they like most, chain profits remain the same. This assertion follows from two observations: First, the retailer is only concerned about whether a consumer visits a store of its chain or a competitor. From a chain perspective, it is not important which store in the network the consumer visits.

Second, if a consumer moves to a new store, the store for which he or she has the highest attraction must belong to this particular retail chain as well. A consumer can only be allocated to a new store if the attraction level for this store is higher than the decisive attraction. Therefore, the store the consumer likes most has an attraction level higher than the decisive attraction and must be one of the new stores.

The parameter of interest is β, or the size sensitivity parameter, and this quantity (provided it is unequal to 0), differentiates the proposed model from the location-only full capture model. The beta measures indicate the relative importance that consumers attach to store size compared with travel distance when deciding where to shop. Low values of β indicate that consumers find store size less important than travel distance; in the extreme case that β is 0, travel distance is the only factor considered. Therefore, the proposed model becomes nearly a classical location-only full capture model, except that costs enter the newly proposed models. In contrast, high values of β make distance less important. Ultimately, if beta is very large, the retailer opens only one store that is slightly larger than the largest competitive store at the location with the lowest costs. In the empirical application, we vary the value of β to investigate how it affects the sizes and locations of the stores to be opened. It is difficult to judge in advance how the objective function changes in response to different values of β, which appears on both the left- and right-hand sides of the model (Equations 2.9 and 2.1).

Extensions

Because we use the MILP framework to solve the CFLDP, the preceding general model can be extended in various ways; we discuss two. The first extension pertains to the level of behavioral detail of the utility functions, and the second deals with possible decision-making preferences from a retailer's perspective.

Consumer Behavior. In the current model, travel distance and store size explain the attraction of consumers to each store, though it also is possible to include other variables in the utility functions, such as product assortment, prices, and opening times. Evidence also indicates that consumers differ in their responses to marketing activities (Pauwels et al. 2011). Retailers seek to exploit these differences by tailoring their marketing mix to the specific needs of local consumer groups. From a marketing perspective, it therefore is reasonable to consider multiple consumer types, which can be defined based on sociodemographic variables and differ in the numerical values of the attraction parameters. In the empirical study, we elaborate on and illustrate this extension.

In the previous model, we implicitly assumed that consumers were willing to travel any distance to satisfy their demand. Prior literature instead indicates that some consumers only will travel a limited distance for certain products, known as the range of a product (e.g., Craig, Ghosh, and McLafferty 1984). If there is no store within this distance, consumers switch to substitutes. If we thus include a maximum travel distance in the model, we must add the following constraint:

$$\sum_{j \in P^P \cup P^C} d_{ij} x_{ij} \leq d_i ,$$

in which the maximum distance a consumer from demand area i wants to travel is denoted by d_i. Therefore, a lost sales variable also must be added for each demand area.

Decision Making Preferences. The current model specification determines the optimal number of new stores endogeneously, though it is possible to set a bound on this number, whether an exact number or a minimum or maximum value. If we denote the number of new stores to be located by n, we add the following constraint to the general model to ensure that exactly n new stores are located:

$$\sum_{j \in P^p} x_j = n. \tag{2.10}$$

Profit maximization may be one of multiple objectives a retailer pursues in a particular market. A retailer also might consider it important to limit the market shares of competitors, which we can model by a multiple goal objective. If we define w^p and w^c as weights for the sales of new stores and competitor stores, respectively, the multiple goal objective function takes the following form:

$$-\sum_{j \in P^p} c_j s_j + w^p \sum_{j \in P^p} y_j - w^c \sum_{j \in P^c} y_j .$$

2.4 Empirical Application

We test the model using semi-realistic data from an international chain of luxury leisure clubs that wants to extend its business in the Netherlands. The issues faced by its management are typical for the expansion of multistore networks, which boil down to the determination of the appropriate number of new stores to open in a particular market, their optimal locations, and ideal sizes.

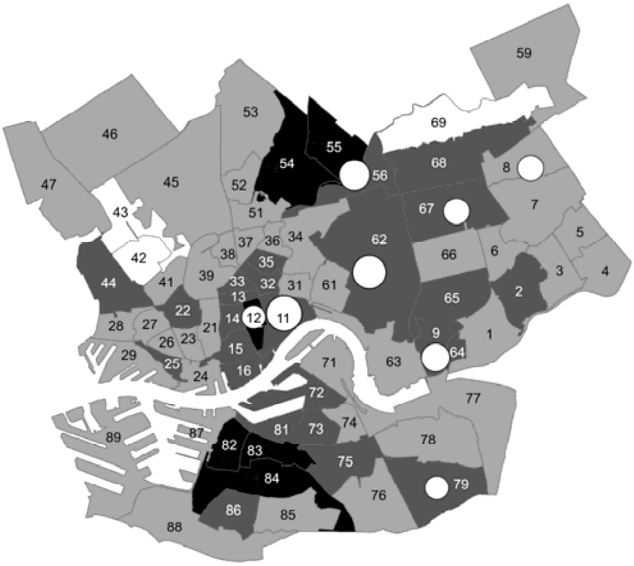
We compare the outcomes of the proposed model with those of several benchmark models. First, we investigate whether the optimal locations of the proposed model differ from those of a location-only full capture model. We also contrast our findings with those of an extended full capture model that includes costs. Second, we evaluate the performance of the proposed model, in which club locations and sizes are optimized simultaneously against an approach that first locates clubs using the extended full capture model and then identifies the optimal sizes for each club. Third, we assess the sensitivity of the location and size results to different values of the size sensitivity parameter through a simulation study. This section ends with a discussion of the results of one of the model extensions, namely, the inclusion of multiple consumer types.

2.4.1 Data

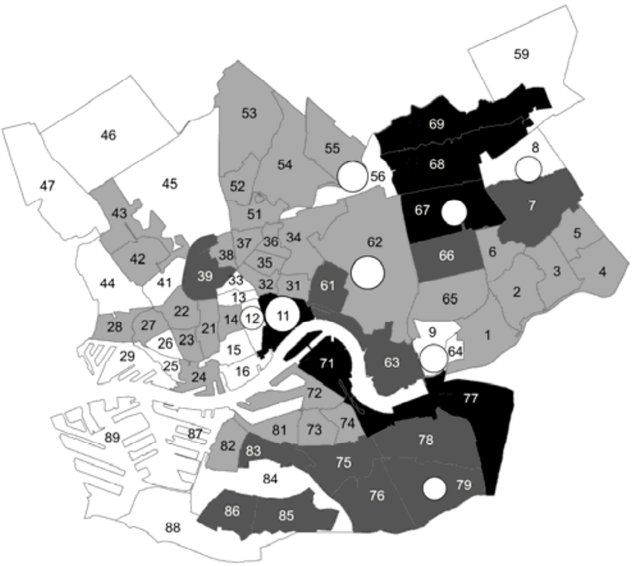
We assume that the management of this health club chain has already decided on the region in which to locate new club(s), namely, the Rotterdam area in the Netherlands, which includes Rotterdam and the neighboring village Capelle aan de IJssel. Decision makers typically separate the identification of the most appropriate region from the choice of particular sites within this region (Levy and Weitz 2004; Mendes and Themido 2004). We subdivide the Rotterdam area into four-digit zip codes, which results in 74 polygons (see Figure 2.1), or "store points" in our model. (In the figures, we omit the first two digits of each zip code, because only the last two digits are unique for all zip codes considered.) We further assume that no clubs of this particular chain are present in the Rotterdam area yet (an assumption that can be easily relaxed). The locations and sizes of competitors appear in Table 2.1 and Figure 2.1. We also use the zip codes as demand points, and for each combination, we obtain travel distances in miles from an online route planner (http://www.mapquest.com).

Table 2.1: Competitors

Location	08	09	11	12	56	62	67	79
Size (m^2)	1000	1200	2000	800	1700	1800	1000	800



[Costs in Euros per year per m^2 with categories, from light to dark: < 200, 200–300, 300–400, and > 400.]



[Demand in thousand Euros per year with categories, from light to dark: < 25, 25–50, 50–75, and > 75.]

Figure 2.1: Costs and demand for each zip code and location of competitors.

Notes: Dot diameter denotes relative size of club.

The first step is to determine the spatial distribution of demand for fitness, which is not readily observed. We use the Statline database of Statistics Netherlands (http://statline.cbs.nl) to determine the number of inhabitants of each zip code. Because the chain offers high-quality facilities, people with higher incomes are its target customers.[3] Of this group, 25% are likely to join a fitness club, and the average membership fee is approximately 40 Euros (IHRSA 2006). Unfortunately, we do not have data to estimate the size sensitivity parameter but instead solve the model for a range of values for this parameter, measured in miles/1000m^2.

We assume the costs of each leisure club consist of those associated with employees (wages) and housing (rent). We obtained information on rental prices in the Rotterdam area from the Realnext Web site (http://www.realnext.com), on which brokers provide information about the (regional) supply of commercial real estate. We then calculated average rental prices for commercial real estate in each zip code. Costs of employees are based on an analysis of company reports from this chain (available in the AMADEUS database[4]). We divided total employee costs by the chain's overall floor space and obtained an amount of 120 euros per m^2.

2.4.2 Results

The model was written in the algebraic modeling language Mosel and solved by Xpress 19 of Dash Optimization (http://www.dashoptimization.com) on a 2.33 GHz Intel Core2Duo machine with 2 GB of RAM. The automatic tuning facilities helped find good values for the many MILP settings.

Results of the Benchmark Models

In this section, we first discuss the results of the location-only full capture model. The standard full capture model assumes that consumers visit the nearest club. This model specification can be obtained from the model proposed in the previous

[3]Consistent with the definition of Statistics Netherlands, we define higher incomes as incomes above the 40 percent point of the Dutch income distribution

[4]http://www.bvdep.com/en/amadeus.html.

section by setting costs and the size sensitivity parameter to 0: $c_j = 0$ for all $j \in P^p$, $\beta = 0$, and we include the number of stores constraint from Equation 2.10. For a number of stores ranging from 1 to 6, the outcomes are listed in Table 2.2 (all solutions found within one second). The first column shows the number of stores considered, whereas the subsequent dots denote the zip codes for new clubs. The last column expresses the overall chain revenues for each configuration (the objective function in this setting). The dominant locations for new clubs are 15 and 35: Location 15 is slightly to the west of the most western competitor and obtains revenues from consumers living in the western part of the region, whereas location 35 serves the northern part of the market.

Table 2.2: Full capture results: Store locations and corresponding revenues
Notes: A dot indicates the presence of a store in a particular zip code.

# Stores	Zip code 02	15	35	68	73	78	81	Revenues
1		●						668,158
2			●				●	1,148,470
3		●	●				●	1,389,666
4	●	●	●		●			1,611,632
5	●	●	●	●			●	1,824,786
6	●	●	●	●	●	●		1,953,950

We also consider the extended full capture model, which includes costs, so the c_i's are restored to their original values, but distance remains the only factor driving consumers' store choice. Because size does not affect consumers' choices of a particular club but does affect costs, we only locate clubs of a minimum size of 1000 m^2. As the results of this model in Table 2.3 show, none of the locations found for the basic full capture model is part of the solution for the new model. Frequently occuring zip codes are 21, 36, and 74. Zip codes 21 and 36 are the low-cost equivalents of zip codes 15 and 35 that dominate the standard full capture model. Location 74 is a centrally located, low-cost zip code in the southern part of the region in which competition is low. Opening three clubs at these locations is optimal from a chain profit perspective.

Table 2.3: Results for full capture with costs
Notes: A dot indicates the presence of a store in a particular store.

# Stores	Zip code 06	21	36	39	69	74	77	Profit
1		●						363,505
2				●		●		557,733
3		●	●			●		598,756
4		●	●		●	●		577,910
5	●	●	●		●	●		501,765
6	●	●	●		●	●	●	363,168

Results of the Proposed Model

This section discusses the results of the proposed model, which assumes that consumers decide to visit a particular leisure club according to its size and the distance required to travel to the club. Therefore, the size sensitivity parameter (β), measuring consumers' trade-off between club size and travel distance, must be different from 0. To investigate the effect of this parameter on model outcomes, we vary it and solve the model for each value of the parameter. Specifically, we varied β from 0 to 0.6 in steps of 0.001, so we solved the model 600 times for different values of β . Table 2.4 shows the optimal locations and sizes of clubs for several focal values; the two righmost columns contain the profits and solution times for each situation. In Figure 2.2, we display these results graphically. The horizontal axis of this graph indicates different values of the size sensitivity parameter, whereas the cumulative size of all clubs opened appears on the vertical axis. The shaded areas represent the zip codes in which new clubs are opened and their respective sizes in each situation. These findings confirm the "extreme" solutions we anticipated in section 2.3. For small values of β , the results are in line with the distance-only full capture with costs model, whereas high values result in a single large club. Moreover, the results indicate nonmonotonic behavior of the objective value as a function of β, which means that in the interval 0–0.6, profits both increase and decrease over the range of β values. This result emerges because β is part of both the left- and right-hand sides of the model. Regarding club locations, we find that zip codes 21, 36, and 74 are still dominant, just as in the full capture with costs model. However, for β s in the range of 0.15–0.55, location 69 also comes into play. It offers low costs and is located in the high-demand northeastern part of the region. For β s in [0.35,0.5], zip code 21 gets replaced by 23, which is slightly less expensive.

Table 2.4: Results for location and design model

β	21	23	36	61	69	74	Profit	Sol. time
0.00	1000	0	1000	0	0	1000	598,757	< 1
0.05	1000	0	1000	0	0	1000	591,264	1
0.10	1000	0	1000	0	0	1000	587,102	2
0.15	1000	0	1000	0	1600	1000	590,872	4
0.20	1000	0	1000	0	1450	1000	584,116	7
0.25	1000	0	1000	0	1360	1000	596,760	12
0.30	1000	0	1000	0	1300	1000	605,189	18
0.35	0	1029	1000	0	1257	1000	612,133	27
0.40	0	1000	1000	0	1225	1000	624,236	37
0.45	0	1000	1044	0	1200	1000	617,645	75
0.50	0	1020	1140	0	1180	1000	593,422	202
0.55	1000	0	1218	0	1164	1000	574,750	787
0.60	0	0	0	6667	0	0	638,151	375

Results of Two-Stage Procedure

Because the optimal club locations of the distance-only full capture model with costs appear in the solutions to our proposed model as well, we might question whether it is necessary to perform an analysis that identifies optimal club locations and designs simultaneously. Seemingly, the less time-demanding two-stage procedure suffices, because we can first solve the full capture with costs model to identify optimal club locations and then determine the best club sizes for each location. Therefore, to investigate how this two-stage model performs compared with our approach, we conduct both analyses and plot the results in Figure 2.3.

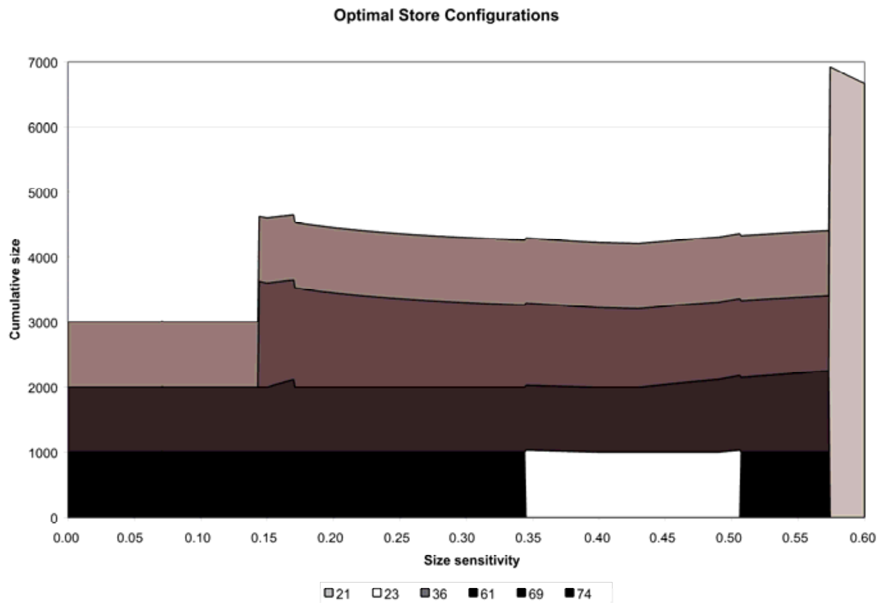


Figure 2.2: Cumulative sizes in the proposed model.

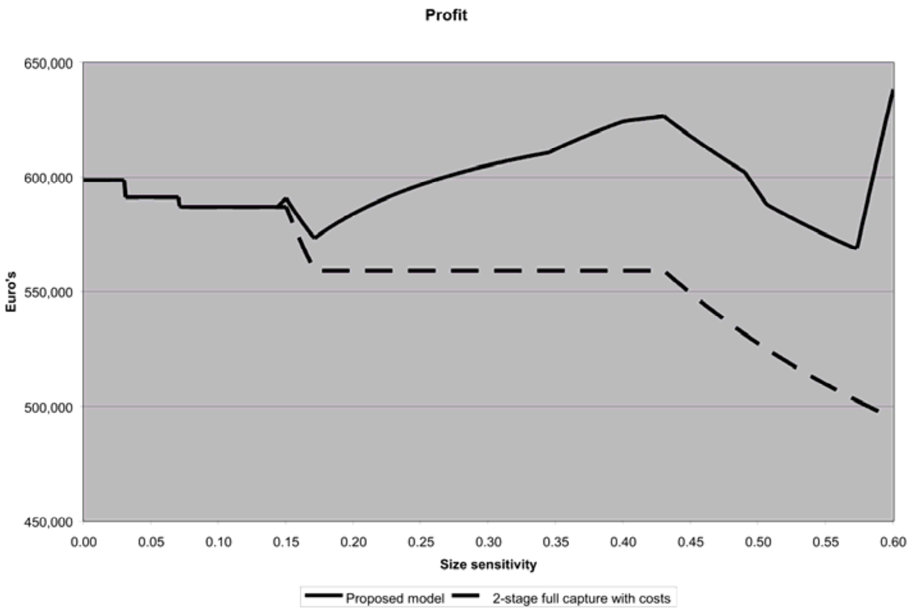


Figure 2.3: Profits of the proposed and two-stage full capture with costs models.

The overall chain profits of the CFLDP model and two-stage model refer to β s in the range [0,0.60]. For low values of β, profits do not differ between the two models, because they result in equal store configurations. Medium values of β result in a CFLDP model profit level approximately 10% higher than that of the two-stage model. This considerable difference becomes much larger for high beta values.

Multiple Consumer Types

The proposed CFLDP model can be extended to include multiple consumer types. In this section, we therefore investigate the effects of distinguishing between two consumer groups: young professionals and families. Consistent with the customer profiles developed by this particular fitness chain, we define these groups as people aged 25 to 35 years (young professionals) and people 35 years and older (families). These two groups also differ in the value of their size sensitivity parameter. Thus, we split all sets, parameters, variables, and equations for the consumers into two subsets. Demand for fitness is determined using demographic data (i.e., population sizes of the two groups) from Statistics Netherlands.

Table 2.5: Results for two consumer types

β^{yp}	β^{f}	21	23	36	39	69	74	Profit	Sol. time
0.300	0.300	1000	0	1000	0	1300	1000	605,189	59
0.354	0.275	1000	0	1000	0	1327	1000	601,358	60
0.408	0.250	0	1000	1000	0	1360	1000	598,750	66
0.462	0.225	0	1000	1070	0	1400	1000	577,223	72
0.516	0.200	1013	0	1000	0	1450	1000	547,568	98
0.571	0.175	1000	0	1246	0	1514	1000	535,243	87
0.625	0.150	1000	0	1000	0	1600	1000	521,273	90
0.679	0.125	1013	0	1000	0	0	1000	513,937	90
0.733	0.100	0	1000	1072	0	0	1000	486,313	111
0.787	0.075	0	1000	1136	0	0	1000	471,812	129
0.841	0.050	0	0	0	1000	0	1000	459,992	189
0.895	0.025	0	0	0	1000	0	1039	449,977	209
0.949	0.000	0	1000	0	0	0	1094	449,564	183

In Table 2.5, β^{yp} and β^{f} are the size sensitivity parameters for young professionals and families, respectively. We assume that families are less likely to travel long distances if they visit a leisure club, and young professionals pay more attention to the size of the club they visit. The first row of Table 2.5 provides the results for a situation in which there is no difference in club size sensitivity between the two groups; in each subsequent row, heterogeneity increases. We lower the size sensitivity of the families in small steps while increasing the sensitivity of young professionals. In each case, the value of β^{yp} is such that the average size sensitivity of the whole consumer population equals the starting value of 0.3.

The results of this extended model are very similar to those of the general model with no consumer heterogeneity. As we show in Table 2.5, the same zip codes are part of the solutions to both models. The only exception is zip code 39, which is a more central location than zip codes 21 and 23, so clubs in these latter zip codes are replaced by that in zip code 39 when travel distance becomes more important to families. Demand by families in each zip code is generally higher than that of young professionals; the configuration of new stores thus closely follows the spatial distribution of demand for this group if β^{f} is very small. Another important observation is that the club located in zip code 69 gets bigger when consumers are more hetereogeneous in their preferences. Therefore, the club opened at this location effectively lures young professionals away from (smaller) competing clubs nearby. With maximum consumer heterogeneity, only two clubs with (almost) minimum sizes open in zip codes 23 and 74, which represent areas with low competition and costs, where family demand for fitness is relatively high. Another notable observation is that the chain’s profits decrease monotonically as consumer heterogeneity increases, perhaps because heterogeneous consumer preferences mean each store can only satisfy the needs of one consumer group: either large (young professionals) or close (families), which by definition results in less sales to the other group.

2.5 Conclusions and Further Research

The key objective of this chapter has been to provide a modeling approach to the competitive facility location and design (CFDLP) problem, when the goal is to determine simultaneously the optimal locations and sizes of multiple stores in a competitive environment. The proposed model contributes to (static) competitive facility literature in several important ways. First, it enables us to generate an optimal solution rather than evaluate a fixed number of predefined store configurations. Second, our model offers a continuum of design alternatives and can include all kinds of constraints. Therefore, a new store can be of any size (instead of a limited number of size alternatives), and the behavior of consumers and/or company preferences can be represented realistically. Mathematically, we formulate the model as a mixed integer linear programming (MILP) model, a well-established paradigm that eases the model adoption and implementation. Moreover, knowledge about solving MILP models is sufficient, such that we can solve instances of considerable sizes effectively.

In the empirical study, we illustrate this model by applying it to a realistic case study of the location of new health clubs in the Rotterdam area. The solution of the full model with consumers who are not very sensitive to store size is similar to that of the full capture (with costs) model, whereas for extremely size-sensitive consumers, a single large club is optimal (from the company perspective). Moreover, profits depend nonmonotonically on the value of the store size sensitivity parameter. We also evaluate whether a model with sequential location and design, which demands less time to compute, can generate similar solutions. Consistent with Tóth et al. (2009), we find that the difference in profits between sequential and simultaneous approaches is considerable, depending on the store size sensitivity parameter. Finally, we consider a multiple consumer type extension.

The reasonable solution times for problems of considerable sizes underline the model's usefulness for supporting location and design decisions in practice. Retailers that want to expand their operations to a particular region with several stores opening simultaneously can use the model to find an optimal number of stores for a particular area, identify suitable locations for each store, and determine their optimal designs. In a later stage, a more detailed analysis performed for each site can evaluate its attractiveness and sales potential.

We acknowledge several limitations of our study that suggest directions for further research. In particular, an important and significant challenge will be to find an empirically validated estimate of the store size sensitivity parameter, which we believe is worth the effort. Validating this parameter would make it possible to use it for simulation and prediction purposes. Because we do not know the exact attitudes of individual consumers toward travel distance and store size, we assume the parameters measuring the effects of these variables on store attraction vary across consumers and follow some general distribution (e.g., normal). Another area for research therefore would include constraints that limit the amount of cannibalization—that is, the loss of sales in a particular store due to the overlap of its trade area with that of another store in the same chain. Such modeling additions are crucial if a franchisor wants to add new stores in an area in which it already operates. Sales through franchise systems have grown rapidly in the United States in recent decades (Kalnins 2004; Lafontaine and Shaw 1999), so we consider this extention highly relevant.

Chapter 3

Store Location Evaluation Based on Geographical Consumer Information[1]

3.1 Introduction

The old adage regarding the three most important things in retailing—"location, location, location" (Jones and Simmons 1987)—still holds, because store location remains a crucial driver of store performance in modern retail environments (Levy and Weitz 2004; McGoldrick 1990; Pan and Zinkhan 2006). From a consumer perspective, travel distance to the store strongly affects the store's attractiveness, and from a retailer perspective, store location decisions involve massive and almost irreversible capital investments that, given the behavior of consumers, largely determine the trade area of the store (Achabal, Gorr, and Mahajan 1982; Ailawadi and Keller 2004; Briesch, Chintagunta, and Fox 2009; Craig, Ghosh, and McLafferty 1984). Therefore, successful retailers routinely evaluate the performance of their current stores and predict the sales impact of potential location changes or new store openings (Gauri, Pauler, and Trivedi 2009; Ghosh and McLafferty 1982; Ghosh and Craig 1983; Kumar and Karande 2000).

These observations have given rise to several quantitative approaches to decision making about store locations, such as gravity and regression models (Buckner 1998; Craig, Ghosh, and McLafferty 1984). Two recent developments

[1] This chapter is based on a working paper with the same title co-authored by Tammo H.A. Bijmolt and J. Paul Elhorst, which is currently under publication review.

offer additional opportunities for store location evaluation. First, the widespread adoption of customer loyalty cards has resulted in detailed data about customer purchase behavior (Leenheer and Bijmolt 2008). Second, developments in spatial econometrics provide a means to account for spatial autocorrelation among the error terms of neighboring observations and thereby increase the efficiency of parameter estimates (e.g., Bradlow et al. 2005; Bronnenberg and Mahajan 2001; Bronnenberg 2005). We exploit these developments by proposing and testing a new model for store location evaluation. By combining data about store, competitor, and consumer characteristics, we build a model that can be used to (1) assess the impact of various drivers on store sales, (2) evaluate the expected performance of stores, and (3) predict sales impacts of future changes in competition, changing demographics within markets, location changes, and new store openings.

Expenditures by loyalty program members generally exceed those of nonmembers (Drèze and Hoch 1998; Van Heerde and Bijmolt 2005). Therefore, to understand the mechanisms that drive store sales, we adopt a decomposition framework to investigate how loyalty program members and nonmembers contribute to sales (Farris, Parry, and Ailawadi 1992; Van Heerde and Bijmolt 2005). We further decompose sales to loyalty program members within a particular zip code and specify the penetration rate of the loyalty program, the number of visits, and the average expenditures per visit. This decomposition framework and knowledge of the location or residence of loyalty program members offer a means to relate these variables to store, competitor, and consumer characteristics and thus obtain more detailed insights in drivers of store sales (Van Heerde, Leeflang, and Wittink 2004). Moreover, the collection of zip code-level data is relatively unobtrusive for customers. Thus, the proposed methods are not only managerially useful but also appropriate for the modern retail environment, in which many consumers believe their privacy is at stake. We use the number of competitors as one of the explanatory variables of the different sales components, as well as a variable that needs to be explained. Competition may have a notable effect on sales, and simultaneously, potential sales can attract competitors.

The error terms in model equations that aim to explain store sales variables are likely to be spatially correlated when they include data at the zip code level, because zip codes in close proximity often share unobserved characteristics, such as climate, resources, sociodemographic factors, and economic circumstances. Therefore, unobserved explanatory variables pertaining to consumer lifestyles, attitudes, preferences, and choices within zip codes close to one another cannot be considered completely independent (Steenburgh, Ainslie, and Engebretson 2003; Ter Hofstede, Wedel, and Steenkamp 2002; Yang and Allenby 2003). Thus, failing to account for spatial error autocorrelation when it exists causes inefficiency (Anselin 1988). Models that take spatial dependence into account, among observations in general and the error term in particular, recently have received attention in marketing literature from, for example, Bronnenberg and Sismeiro (2002) and Yang and Allenby (2003). Building on this stream of research, we use random effects models extended with spatial error autocorrelation to account for spatial dependencies in sales components across zip codes.

The remainder of this chapter is organized as follows: First, we provide an overview of existing store location evaluation models that serves as a starting point for further model development. Second, we present the decomposition framework for modeling store sales and introduce spatial-error random-effects models that we use to explain different store sales variables. Third, with data from 28 stores of a Dutch retail clothing chain, we test our framework in an empirical setting. Fourth, we discuss the results, demonstrate how to evaluate the performance of stores and how to predict (potential) sales of new store locations, provide managerial implications of the environmental changes in some explanatory variables, and suggest directions for further research.

3.2 Previous Literature

Various analytical tools attempt to evaluate store location decisions (Buckner 1998; Craig, Ghosh, and McLafferty 1984; Levy and Weitz 2004), mostly by considering the amount of sales each location can generate in a certain period, given the current spatial distribution of demand and competition (Ghosh and McLafferty 1982). However, changes in the retail environment may have significant impacts on store sales. These changes might include shifts in the spatial distribution of demand caused by population or income dynamics (Ghosh and Craig 1983), as well as altered activities by competing firms (Ghosh and McLafferty 1982; Singh, Hansen, and Blattberg 2006). Store location evaluation models therefore should accommodate the effects of such changes.

Huff's gravity model and its extensions (Gautschi 1981; Huff 1964; Stanley and Sewall 1976) provides one of the earliest applications of spatial models in marketing. Such models predict the geographical extent of store trade areas on the basis of a negative relation between store patronage and distance to consumers. Thus, they can explain the proportion of visits from a certain area to the store but not any changes in consumers' expenditures. To predict sales, the probability that a consumer will visit the store from a particular location is multiplied by an estimate of the (average) expenditures at that location and by population size or, alternatively, by (average) expenditures per household and number of households. As another important limitation, these models do not include consumer characteristics; they assume store patronage depends only on store size and distance to the store. Some authors include factors other than distance and store size that may affect consumers' store choice, such as retail center characteristics (e.g., Gautschi 1981; Stanley and Sewall 1976), but in general, the number of variables describing the store environment remains limited.

Regression models enable analysts to identify several factors associated with different levels of sales from stores at different sites. However, existing studies typically do not account for the spatial heterogeneity of the store's trade area, even though retailers in most Western countries serve trade areas with a rather heterogeneous population (Campo and Gijsbrechts 2004). That is, previous studies use aggregate measures of consumer demographics for the entire trade area to predict sales at a particular store location. As such, these models ignore how consumers, competitors, and demographics are distributed throughout the trade area. Recent studies reveal that the geographic location can be an important variable for predicting consumer behavior (e.g., Bronnenberg and Mahajan 2001; Bronnenberg and Sismeiro 2002; Steenburgh, Ainslie, and Engebretson 2003; Yang and Allenby 2003). Bronnenberg and Mahajan (2001), for example, find that 66–95% of the variation in market shares for two homogenous product categories are due to spatial rather than temporal variation. In particular, growing literature on spatial models in marketing (for overviews, see Bradlow et al. 2005; Bronnenberg 2005) reveals how spatial covariation in sales can be exploited to gain better insights into the effectiveness of marketing activities across markets. In other words, despite reasons to believe that these spatial effects can be substantial, most of these effects heretofore have been ignored in store location literature (Duan and Mela 2009).

Similar to Bronnenberg and Mahajan (2001), we adopt a sales model with a spatial autocorrelation component. However, our model differs from theirs in that we decompose sales into various components. Most studies consider only a few components of store sales (e.g., Bawa and Ghosh 1999; Inman, Shankar, and Ferraro 2004) or sales in general (Kumar and Karande 2000) and offer no insights in the underlying mechanisms causing changes in store sales components. Pan and Zinkhan (2006), however, show that various regressors can have different effects across sales components, which suggests that decomposing sales (effects) into constituent parts may offer richer insights than a model of total store sales only. As another difference, Bronnenberg and Mahajan's (2001) model only captures spatial heterogeneity in (secondary) demand arising from unobserved and observed differences in retailer behavior across markets. In contrast, our model accommodates the sales effects of observed spatial differences for a broad range of variables, such as store, competitor, and consumer characteristics, while also accounting for unobserved sources of spatial heterogeneity in store sales.

Another stream of research that we consider belongs to empirical economics literature (Chintagunta et al. 2006). Some examples from this stream include recent studies by Chan and colleagues (2007), Duan and Mela (2009), and Thomadsen (2007), who determine equilibrium prices or sales conditional on outlet location and capacity. These studies show that location competition affects sales, whether positively or negatively, while the reverse can be true as well; that is, (potential) sales may attract competitors. To separate these alternative explanations, we employ the number of competitors as a variable to explain sales components, and also include an equation to explain the number of competitors in a particular location.

3.3 Model for Store Location Evaluation

1. Determine perimeters of store trade areas and allocate zip codes to each store.

2. Estimate the store location evaluation model:
   - Decompose total sales into different components;
   - Identify potential drivers of store sales;
   - Assess the impact of drivers of store sales;
   - Explain number of competitors.

3. Use the model for:
   - Explaining sales (components) for previous periods;
   - Predicting sales (components) for current and future periods;
   - Evaluating effects of changes in drivers of store performance;
   - Predicting sales (components) for new store locations.

Figure 3.1: The store location evaluation process.

In Figure 3.1, we propose a three-stage decision model for evaluating current and future store performance. The first stage involves defining the trade area on the basis of detailed customer information derived from loyalty cards. Using this definition, we can select zip codes that belong to the stores' trade areas. The second stage requires analyzing store sales using regression-type models that relate data on store sales to data on store and competitor characteristics, as well as consumer characteristics observed at the zip code level within the trade area. Finally, in the third stage, the estimated coefficients of these relations serve to (1) assess the impact of various drivers of store sales, (2) evaluate the overall performance of stores, and (3) predict sales impacts of future changes in competition and demographics within markets and of location changes and new store openings. We discuss these three stages of the decision process in the following subsections.

3.3.1 Trade Area Definition

The first stage of the store location evaluation process (Figure 3.1) involves defining the trade area. Knowledge about the store's trade area is essential, because it allows retailers to identify and serve the consumers who are most likely to purchase at that particular store. The various approaches for trade area delineation (O'Kelly and Miller 1989) generally fall into two broad research streams: studies based on the Reilly model and its extensions (e.g., Huff 1964) and those based on the Applebaum (1966) procedure. These model types differ in the kind of questions they address. Whereas the Reilly model attempts to determine the percentage of consumers in a given area who patronize a particular store, the Applebaum procedure tries to find the fraction of a store's customers from a given area. Because these approaches are highly related, some studies integrate them (O'Kelly and Miller 1989).

Although these approaches establish the spatial distribution of consumers, they typically fail to allow for heterogeneity across space (e.g., Donthu and Rust 1989; Gonzalez-Benito, Munoz-Gallego, and Kopalle 2005; González-Benito and González-Benito 2005). That is, they assume trade areas are homogeneous, such that neighborhoods have similar characteristics and spending patterns. However, in most Western countries, such areas typically consist of a mosaic of small zip codes with specific sociodemographic and lifestyle characteristics (Campo and Gijsbrechts 2004). Therefore, researchers must consider differences in consumer characteristics across the trade area. Information about (changes in) the demographic and competitive characteristics of the store's trade area can help predict how local market potential may evolve over time and indicate whether a store currently is over- or underperforming (Montgomery 1997; Putler, Kalyanam, and Hodges 1996).

Because customers who sign up for loyalty programs must provide the retailer with their addresses and because the system registers their subsequent purchases, we can use this combined information to define the trade area of a store. To this end, we first split sales into those to loyalty program members whose addresses are known and those to nonmembers whose addresses are not known (Van Heerde and Bijmolt 2005), such that

$$S_{it} = SL_{it} + SN_{it}, \tag{3.1}$$

where i refers to the store ($i = 1, \ldots, I$, where I is the total number of stores), t represents a given time period ($t = 1, \ldots, T$, where T is the number of time periods), S_{it} is total sales of store i at time t, SL_{it} is sales to loyalty program members of store i at time t, and SN_{it} is sales to nonmembers of store i at time t.

Trade areas typically consist of two or three zones (Applebaum 1966; Levy and Weitz 2004), depending on the amount of sales generated in each area. A store's primary trade area is the zone from which it gets most of its sales—usually about 65% of total sales. The secondary zone generates the next 20% of total sales, whereas the tertiary zone captures sales from non-regular visitors (i.e., the remaining 15–20%). Because we model the purchase behavior of loyalty card holders who should be regular visitors (Allaway, Berkowitz, and D'Souza 2003; Van Heerde and Bijmolt 2005), we focus on the primary and secondary zones and consider zip codes belonging to those two zones part of the store’s trade area. The number of zip codes in the trade area depends on sales, so trade area boundaries should be considered endogenous (Greene 2003). Furthermore, trade areas might be subject to spatial variation. Differences in product assortments, competition, and consumer shopping habits (Huff and Rust 1984) can create variations in trade area boundaries across stores and over time. We therefore specify a model for trade area delineation, in which we model trade area sizes as a function of store and competitor characteristics.

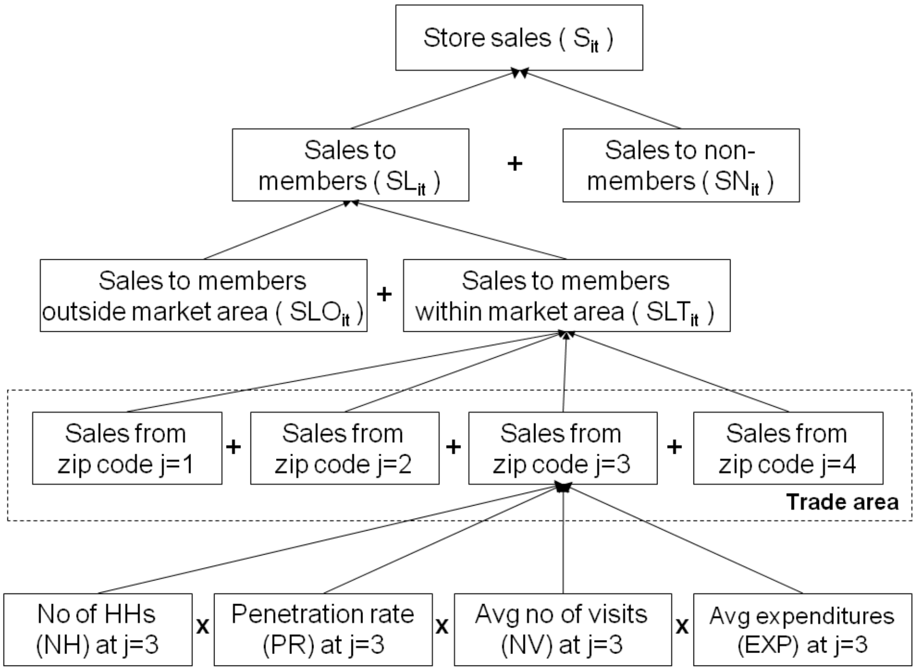


Figure 3.2: Decomposition framework for store sales.

Even though members may be responsible for a large proportion of total sales (Singh, Hansen, and Blattberg 2006; Van Heerde and Bijmolt 2005), their spatial distribution does not necessarily coincide with the store's trade area. That is, the primary and secondary zones contain many loyalty program members but not all of them. Therefore, we further decompose sales to members into those sales to members living inside the store's trade area and those to members living outside the trade area, such that

$$SL_{it} = SLO_{it} + SLT_{it}, \tag{3.2}$$

where SLO_{it} is sales to members living outside the trade area of store i at time t and SLT_{it} is sales to members living within the trade area of store i at time t. We illustrate both decompositions in Figure 3.2.

3.3.2 Components of Store Sales

The second stage of the store location evaluation process (Figure 3.1) pertains to assessing the impact of store, competitor, and consumer characteristics. The viability of a store depends largely on its capability to satisfy the needs of consumers who live within the trade area. Successful retailers have information about the sociodemographic profiles of consumers who live in the trade areas of their stores and develop strategies to influence responses to the store's marketing activities (Campo et al. 2000; Campo and Gijsbrechts 2004; Hoch et al. 1995; Montgomery 1997; Mulhern 1997). Therefore, store location evaluation models should determine the impact of (changes in) trade area demographics on components of store sales and capture these effects over both space and time.

To obtain detailed insights into the drivers of store sales, we decompose sales into different components and allow the explanatory variables to possess different parameters across these components. From the customer database, we can obtain measures of the membership rate, visit frequency, and average amount spent per visit for each zip code. Thus, sales to loyalty program members living in the trade area of store i at time t can be modeled as follows (see also Figure 3.2):

$$SLT_{it} = \sum_{j=1}^{J_{it}} SL_{ijt} \equiv \sum_{j=1}^{J_{it}} NH_{jt} \times PR_{jt} \times NV_{ijt} \times EXP_{ijt}, \tag{3.3}$$

where j refers to zip codes ($j = 1,\ldots,J_{it}$, where J_{it} is the number of zip codes belonging to the trade area of store i at time t), SL_{ijt} is sales to members of store i living in zip code j at time t, NH_{jt} is the number of households living in zip code j at time t, PR_{jt} is the penetration rate of the loyalty card in zip code j at time t, NV_{ijt} is the average number of visits to store i at time t among members living in zip code j, and EXP_{ijt} is the average expenditures per visit of members from zip code j at store i during time t.

In summary, we decompose sales into six variables: SN_{it}, SLO_{it}, NH_{jt}, PR_{jt}, NV_{ijt}, and EXP_{ijt}. The number of households living in a particular zip code, NH_{jt}, may be treated as an exogenous variable, but we model the other five sales components as functions of store, competitor, and consumer characteristics. In addition, we model the trade area perimeter (TAP_{it}) and the number of competitors (NC_{it}).

3.3.3 Drivers of Store Sales

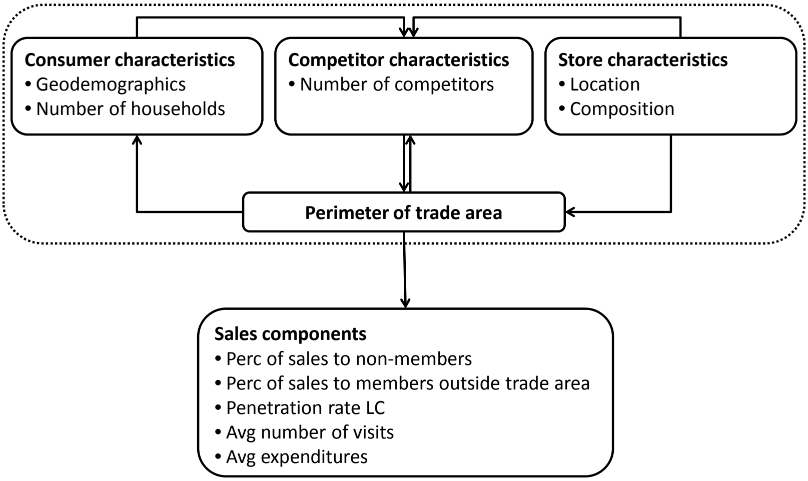


Figure 3.3: Conceptual model of potential drivers of store sales.

Following extant literature, we employ both store and trade area (i.e., consumer and competitor) characteristics to explain store sales components, as illustrated in Figure 3.3.

Many studies document possible relationships between consumer demographics and various components of store sales (e.g., Kumar and Karande

2000; Pan and Zinkhan 2006; Reinartz and Kumar 1999), but no consensus exists about how consumer demographic variables may relate to store sales. Reinartz and Kumar (1999) find that the number of households living in the store's trade area has the largest impact on store performance, followed by store attractiveness and socioeconomic status. The theory of time allocation between different activities, as used by Hoch and colleagues (1995) and Kumar and Karande (2000), suggests that store performance relies, among other things, on household income and size. Because high-income households have higher opportunity costs for their time, they tend to visit stores less frequently but spend more per visit. Pan and Zinkhan (2006) indicate that consumer demographics have the greatest impact on visit frequency, whereas product and market-relevant factors have more influence on store choice. Specifically, gender represents an important predictor of visit frequency, whereas store characteristics (e.g., service quality, store atmosphere) and product attributes (e.g., product selection, quality) determine store choice.

With respect to competitor characteristics, Singh and colleagues (2006) find that the entrance of a large competitor has a significant effect on the number of visits of loyalty program members to an incumbent store, though the residence location of customers appears to moderate this effect. Moreover customer locations, as Allaway and colleagues (2003) show, influence a customer's likelihood of adopting a new loyalty program, according to distance from the store. The number of nearby adopters at a particular location also influences the decision to join a new program. Thomadsen (2007) shows that locations with a large number of people matching the firm's target customer profile typically attract a large number of competitors as well. Finally, Seim (2006) finds that undifferentiated firms avoid direct competition by locating their stores far from those of competitors.

The results of these studies imply that the impact of environmental changes vary across different components of store performance. Therefore, we next model the components of store sales (SN_{it}, SLO_{it}, PR_{jt}, NV_{ijt}, and EXP_{ijt}), as well as the trade area perimeter (TAP_{it}) and the number of competitors (NC_{it}), as a function of store and trade area characteristics and allow the coefficients of explanatory variables to differ across the components. We first present model specifications of variables explained at the zip code level and then at the store level.

3.3.4 Specification of Zip Code-Level Models

To explain the average number of visits, NV_{ijt}, and the average expenditures per visit, EXP_{ijt}, we use a hierarchical random effects model, because we have data at two levels. Zip codes, which we call level 1 units, group within the trade areas of stores, which are the level 2 units. Consequently, each store has its own intercept, which is the sum of a general mean (γ^{NV}) for all stores and a normally distributed error term with zero mean, $E\left(\nu_i^{NV}\right) = 0$, and constant variance, $Var\left(\nu_i^{NV}\right) = \phi_{NV}^2$. The set of explanatory variables includes variables observed at the store level (X) and zip code level (Z). The X variables are indexed by i (store), t (time), and k ($k = 1,\ldots,K$) to denote their index within the total number of explanatory variables. The Z variables are indexed by j (zip code) too, while k is replaced by n ($n = 1,\ldots,N$). We take the natural logarithm (ln) of both store performance variables, because they are bounded by zero and skewed to the right. The model for NV_{ijt} therefore can be written as:

$$\begin{aligned} ln\left(NV_{ijt}\right) &= \gamma_i^{NV} + \sum_{k=1}^{K}\alpha_k^{NV} X_{itk} + \sum_{n=1}^{N}\beta_n^{NV} Z_{ijtn} + \varepsilon_{ijt}^{NV}, \\ \gamma_i^{NV} &= \gamma^{NV} + \nu_i^{NV}. \end{aligned} \tag{3.4}$$

If we substitute EXP_{ijt} for NV_{ijt}, we obtain the model for the average expenditures per visit.

As we noted in the Introduction section, customers from different zip codes who live in close proximity may share the same unobservable characteristics. To deal with such spatial dependence, we must extend the model. We turn to a well-known model in spatial econometrics literature that corrects for spatial dependence with a first-order spatial autoregressive process that generates the error terms

$$\varepsilon_{it}^{NV} = \lambda^{NV} W_{it} \varepsilon_{it}^{NV} + \xi_{it}^{NV}, \tag{3.5}$$

where ε_{it}^{NV} and ξ_{it}^{NV} are written in vector form for each cross-section of zip codes ($j = 1, \ldots, J_{it}$) in the trade area of store i at time t, such that $E\left(\xi_{ijt}^{NV}\right) = 0$ and $Var\left(\xi_{ijt}^{NV}\right) = \sigma_{NV}^2 I_{J_{it}}$ (Anselin 1988). In addition, W_{it} is a $\left(J_{it} \times J_{it}\right)$ non-negative matrix with zeros on the diagonal that describes the spatial arrangement of the zip codes within the trade area of store i at time t. Specifically, each zip code j appears as both a row and a column in W_{it}, and the weights $w_{jj'}$ indicate the relation between zip codes j and j'. The weights are based on first-order contiguity, meaning that they are set to 1 when zip codes share a common border and 0 otherwise, and row standardized, such that $w_{jj'}^s = w_{jj'} / \sum_{j'} w_{jj'}$ if j and j' are direct neighbors and 0 otherwise. The spatial econometrics literature points out that empirical findings might depend on the operationalization of the spatial weights matrix (Anselin 2002; Leenders 2002; Pace and LeSage 2004). We use first-order contiguity weights matrices, because Stakhovych and Bijmolt (2009) show that these models perform better than those that employ other weights matrix specifications. Finally, λ^{NV} is the spatial autocorrelation coefficient, assumed to be fixed. We provide an explanation of how to estimate the parameters of this model in the first part of Appendix A.

Next we specify a model for the penetration rate of the loyalty program, PR_{jt}, which is defined as the ratio between the number of loyalty card holders and the number of households in a particular zip code. Even though loyalty cards may be issued at different stores, a loyalty card adopted by a customer is valid for all outlets of the chain. We cannot allocate loyalty program members in a zip code to a particular store, so we lack the hierarchical structure in the data that we attain for visit frequency and average expenditures. We therefore model the penetration rate at the chain level rather than the store level and explain loyalty card penetration for the whole region in which the chain operates, not only the zip codes belonging to the stores' trade areas. The model also includes a random intercept at the zip code level rather than at the store level. We measure the store-specific explanatory variables so that they refer to the nearest store. By applying a logit transformation to the loyalty card penetration rate, we ensure that this performance variable follows a normal distribution by approximation. In summary, the model for the penetration rate can be written as follows:

$$\begin{aligned} logit(PR_{jt}) &= \gamma_j^{PR} + \sum_{k=1}^{K} \alpha_k^{PR} X_{jtk} + \sum_{n=1}^{N} \beta_n^{PR} Z_{jtn} + \varepsilon_{jt}^{PR}, \\ \gamma_j^{PR} &= \gamma^{PR} + \nu_j^{PR}, \end{aligned} \tag{3.6}$$

where $E\left(\nu_j^{PR}\right) = 0$ and $Var\left(\nu_j^{PR}\right) = \sigma_\nu^2$. If the observations are stacked in a vector for each cross-section of zip codes at time t $\left(j = 1,\ldots,J\right)$, we take spatial error autocorrelation into account with:

$$\varepsilon_t^{PR} = \lambda^{PR} W \varepsilon_t^{PR} + \xi_t^{PR}, \tag{3.7}$$

where $E\left(\xi_t^{PR}\right) = 0$, $Var\left(\xi_t^{PR}\right) = \sigma_{PR}^2 I_J$, W is a row-standardized first-order contiguity weight matrix of size $\left(J \times J\right)$ that describes the spatial arrangement of zip codes, and J is the total number of zip codes within the country. We explain how to estimate the parameters of this model in the second part of Appendix A.

Finally, it is likely that a zip code that has high unobserved variables for one sales component also has them for another. As a result, the disturbance terms for the sales components will not be independent. The seemingly unrelated regressions (SUR) model allows for correlation among the error terms of a set of model equations. However, if all equations have the same explanatory variables, as is the case for the model equations estimated at zip code level, then estimating each equation separately yields identical results as estimating all equations simultaneously (Greene 2003).

3.3.5 Specification of Store-Level Models

The percentage of sales to nonmembers $\left(SN_{it} / S_{it}\right)$, the percentage of sales to members living outside the store's trade area $\left(SLO_{it} / SL_{it}\right)$, the trade area perimeter (TAP_{it}), and the number of competitors (NC_{it}) can be explained using an ordinary random effects model at the store level. We take the logit of the first two variables and the natural logarithm of the last two variables to ensure that they follow a normal distribution by approximation. The model for the first variable can be written as:

$$\begin{aligned} logit\left(\frac{SN_{it}}{S_{it}}\right) &= \gamma_i^{SN} + \sum_{q=1}^{Q} \alpha_q^{SN} X_{itq} + \sum_{n=1}^{N} \beta_n^{SN} \bar{Z}_{itn} + \varepsilon_{it}^{SN}, \\ \gamma_i^{SN} &= \gamma^{SN} + \nu_i^{SN}, \end{aligned} \tag{3.8}$$

in which the $\bar{Z}_{itn}$ variables are the averages for each zip code-level variable Z_{ijtn} for all zip codes belonging to the store’s trade area in a particular time period. Furthermore, $E\left(\nu_i^{SN}\right) = 0$, $Var\left(\nu_i^{SN}\right) = \eta_{SN}^2$, $E\left(\varepsilon_{it}^{SN}\right) = 0$, and $Var\left(\varepsilon_{it}^{SN}\right) = \sigma_{SN}^2$. If we replace $\left(SN_{it} / S_{it}\right)$ by $\left(SLO_{it} / SL_{it}\right)$ and logit$\left(SN_{it} / S_{it}\right)$ by log(TAP_{it}) and log(NC_{it}), we obtain the model for the other variables.

Just as in the previous section, it is likely that a store that has high unobserved variables for one sales component has them for another, as a result of which the disturbance terms for the sales components will not be independent. The SUR model allows for correlation among the error terms of a set of model equations. Because the equations estimated at the store level will have different sets of explanatory variables, estimating each equation separately does not yield identical results as estimating all equations simultaneously. We therefore adopt the estimation procedure spelled out in Magnus (1982). However, generalized least squares (GLS) offers no improvement in efficiency if the variables in one equation are a subset of those in another block of equations (Greene 2003). This warning applies to the model for the number of competitors in a particular area, which shares all variables with the other store model equations.

3.3.6 Prediction in Spatial Random Effects Model

An important element of store location evaluation involves predicting sales, which constitutes the third stage of our store location evaluation process (Figure 3.1). In this section, we discuss the properties of prediction in random effects models with spatial autocorrelation. Baltagi and Li (2004) provide more information on prediction in spatial models based on panel data.

Goldberger (1962) shows that for the error variance–covariance matrix Ψ, the best linear unbiased predictor (BLUP) for the jth area at a future period $T+C$ is given by:

$$\hat{y}_{j,T+C} = U'_{j,T+C}\hat{\theta}_{GLS} + \psi'\Psi^{-1}\hat{\varepsilon}_{GLS}, \tag{3.9}$$

where $\psi = E\left(\varepsilon_{j,T+C}\varepsilon\right)$ is the covariance between the future disturbance $\varepsilon_{j,T+C}$ and the sample disturbances ε, U' covers the explanatory variables of the model, θ_{GLS} corresponds to the GLS estimator of θ, and $\hat{\varepsilon}_{GLS}$ denotes the corresponding GLS residual vector.

The inverse of the variance–covariance matrix of the ordinary random effects model, such as in Equation (3.8), is:

$$\Psi_{RE}^{-1} = \frac{\sigma_{SN}^{2}}{T\eta_{SN}^{2} + \sigma_{SN}^{2}} \left(\frac{1}{T} \iota_T \iota_T^{'} \otimes I_I \right) + \left(I_T - \frac{1}{T} \iota_T \iota_T^{'} \right) \otimes I_I , \tag{3.10}$$

where ι_T is a vector of dimension T, and I_I is an identity matrix of dimension I. If the observations are stacked in a vector for each cross-section of stores $(i = 1,\dots,I)$, the corresponding BLUP correction term is (Baltagi and Li 2004):

$$\psi^{'} \Psi_{RE}^{-1} \hat{\varepsilon}_{GLS} = \frac{T\eta_{SN}^{2}}{T\eta_{SN}^{2} + \sigma_{SN}^{2}} \frac{1}{T} \sum_{t=1}^{T} \hat{\varepsilon}_{t,GLS} . \tag{3.11}$$

In words, we first average the residuals of each store over the sample period and then multiply them by a factor that can take values between 0 and 1. We combine this term with Equation (3.8) and thereby predict the percentage of sales to nonmembers at the store level. An equivalent procedure predicts the percentage of sales to members living outside the store's trade area, the trade area perimeter, and the number of competitors.

For a random effects model with spatial autocorrelation, such as in Equation (3.6), Baltagi and Li (2004) demonstrate that the BLUP correction term for each cross-section of zip codes ($j = 1,\dots,J$) is:

$$\frac{\sigma_{\nu}^{2}}{\sigma_{PR}^{2}} V^{-1} \sum_{t=1}^{T} \hat{\varepsilon}_{t,GLS} , \tag{3.12}$$

where $V = T \frac{\sigma_{\nu}^{2}}{\sigma_{PR}^{2}} I_J + \left(\left(I_J - \lambda^{PR} W \right)^{'} \left(I_J - \lambda^{PR} W \right) \right)^{-1}$. In other words, the BLUP correction term to $U_{j,T+C}^{'} \hat{\theta}_{GLS}$ is a weighted average of the GLS residuals for the J zip codes. The weights depend not only on $(1/T)$ but also on the spatial weight matrix W and the spatial autocorrelation coefficient λ^{PR}. We combine this formula with Equation (3.6) and thereby predict the penetration rate of the loyalty card at the zip code level.

For a hierarchical random effects model with spatial autocorrelation, such as in Equation (3.4), one may use the BLUP correction term of a random effects model in combination with the Kelejian and Prucha (2007) correction for spatially autocorrelated errors. For a cross-section of zip codes ($j = 1,..., J_{it}$) of a particular store i ($i = 1,\ldots, I$) this takes the form

$$\frac{\tilde{T}_i \phi_{NV}^2}{\tilde{T}_i \phi_{NV}^2 + \sigma_{NV}^2} \frac{1}{\tilde{T}_i} \sum_{t=1}^{T} \sum_{j=1}^{J_{it}} \hat{\varepsilon}_{ijt,GLS}^{NV} + \lambda^{NV} \frac{1}{T} \sum_{t=1}^{T} W_{it} \hat{\varepsilon}_{it,GLS}^{NV}, \tag{3.13}$$

where $\tilde{T}_i = \sum_t J_{it}$.[2]

When taking the exponent of the log number of visits, the log expenditures per visit, the log trade area perimeter, and the log number of competitors, we obtain biased predicted values due to Jensen's Inequality. That is, the error term when $\ln(\hat{y})$ gets transformed into $\hat{y}$ will follow a log-normal instead of a normal distribution, which has a mean greater than 0. Therefore, the detransformed predictor systematically underestimates the true values of $\hat{y}$. A remedy suggested by Miller (1984) takes the following form for the random effects model of the log perimeter of the trade area:

$$\hat{E}(y \mid U) = e^{U' \theta_{GLS} + \text{BLUP correction}} e^{1/2\left(\sigma_{TAP}^2 + \eta_{TAP}^2\right)}, \tag{3.14}$$

A similar procedure applies to *NV*, *EXP*, and *NC*. The antilogit of the predicted values of the penetration rate, percentage of sales to nonmembers, and percentage of sales to members living outside the store's trade area are not biased, because the transformation achieves a mean of 0.

Finally, when new stores open, the BLUP correction terms are set to 0, because the error terms on which they depend cannot be observed. To assess the quality of the prediction, we follow Verbeek (2000) and define the R^2 for panel data applications in terms of the squared correlation coefficient between actual and fitted values, instead of using the usual R^2 in terms of the sums of squares.

[2] Alternative forms but without spatial error correlation are discussed in Frees (2004). Kelejian and Prucha (2007) consider predictors based on cross-sectional data. We adopt its space-time counterpart. The advantage is that this correction term does not involve observations at T+C, as a result of which it is feasible.

3.4 An Empirical Analysis

We structure this section around potential applications of the proposed store evaluation model. Specifically, after discussing the empirical data in section 3.4.1, we analyse the determinants of the trade area perimeter and the number of competitors of a particular store in section 3.4.2. Next, we examine the impact of store, competitor, and consumer characteristics on store sales in section 3.4.3. Because the second goal of our model is to evaluate the performance of stores, we discuss this point in section 3.4.4. Finally, section 3.4.5 deals with answering "what-if" questions, such as determining the effect on total sales if we were to change the relative sizes of certain departments of the store.

3.4.1 Empirical Setting

We use data from 28 Dutch clothing stores that belong to a single chain to test our conceptual framework and methodological model in an empirical setting. The stores offer a medium-quality assortment and are mostly located in medium-sized towns. The company's loyalty program attempts to strengthen its relationship with customers. Those who participate in the loyalty program obtain a 5% cash reward on every purchase, which is credited to their loyalty card and can be spent two times a year.

The customer database contains personal data in addition to purchase data. We use the addresses of members to overlay several sociodemographic variables for the zip code in which each address appears. In addition, we supplement these data with information from a chain-wide survey of outlet managers that provides, for each store, information about the store itself and the competitive environment.

We use information gathered from store managers to identify (direct) competitors, who can be defined as clothing stores targeting the same customer segment. For the store characteristics, we include store size (in 10,000 m^2), the relative size of the various departments (women's, men's, and children's assortments), the number of months the store is open in a particular year, and the year of establishment of that particular outlet (1932 = 0). Unfortunately, the data base does not contain information about marketing variables such as price and quantity, but the lack of these variables is not a severe limitation of the model. First, these variables do not differ to a great extent among the stores, because outlet managers must follow the marketing activities dictated by the head office. Second, we study yearly data, whereas marketing mix variables generally exhibit most of their variance and effects at a much shorter temporal rate. We do account for the effect of assortment composition of the store, which is a key strategic variable for retailers. Finally, to the extent that they are spatially correlated across zip codes, the marketing activities of the chain and its competitors are covered by the spatially autocorrelated disturbance terms.

We have data from a period of five years, 2002–2006, of which we use the first four years for estimation and the last year for validation. Withholding the last year for validation enables us to verify whether the results apply to a new time period. In addition, the chain opened two new stores in 2006, so we can assess whether the results apply to new stores as well by comparing their actual sales figures with the model predictions.

To determine the geographical extent of markets, we first sort all zip codes in descending order of travel distance to the stores and then selecting, for each store and each year, the (first) zip codes responsible for 85% of total sales (we also consider 75% and 95% of total sales). For these zip codes, we determine the perimeter of the trade area according to the maximum travel distance to the store. Travel distance to the store is hereby defined as the shortest time distance (in miles) a car can travel from (the centroid of) a four-digit zip code to the store under consideration. Next, we model the perimeter of store trade areas using an ordinary random effects model as a function of store and competitor characteristics based on 102 (24 stores in 2002 + 26 stores × 3 years) observations. Even if a store does not exist yet, we can use the estimated coefficients to predict the (potential) extent of its trade area at a particular point in time by including all zip codes that reside closer to the store than the predicted trade area perimeter. In this way we obtain an endogenously determined trade area perimeter for each store. The number of zip codes belonging to a store's trade area varies across stores: from 44 to 307, with an average of 110 zip codes.[3] We explain the loyalty card penetration for all zip codes in the Netherlands (N = 4008) in four successive years (2002–2005), which is how we obtain 16,032 (4008 zip codes × 4 years) observations. Since the number of zip codes belonging to one of the trade areas is smaller than 4008, the number of observations on the number of visits is 11,140 and the expenditures per visit 9,884. The latter number is lower than the former, because the number of visits in some zip codes equals zero.

The average perimeter of a store's trade area is 15 miles and the high standard deviation in the third column of Table 3.1 indicates substantial variation in trade area sizes. Table 3.1 also shows that, on average, nonmembers account for 27 percent of store sales. The number of competitors is large, ranging from 22 to 78 across stores. We also see that approximately 10% of the households in the zip codes belonging to a store's trade area have a loyalty card, that they, on average, visit the store once every two years, and that the average cardholder spends €66 per visit (Table 3.2).

[3] The trade areas of stores show little overlap. Only about 15 percent of the zip codes in the Netherlands belong to two or more trade areas, with a maximum of four. To test whether the results were affected by trade area overlap, we included a dummy variable that indicates whether a zip code is assigned to one (0) or multiple stores (1) in the models for visit frequency and average expenditures per visit. In neither of these models, this dummy is significant.

Table 3.1: Descriptive statistics of models estimated at the store level

Variable	**Mean**	**St.dev**
Dependent variables		
Trade area perimeter (in miles)	15.86	5.54
% Unscanned sales	0.27	0.09
% Outside trade area	0.15	0.06
Number of competitors	42.30	12.64
Store characteristics		
Size (in 10,000 m^2)	0.07	0.01
% female assortment	0.44	0.05
% children's assortment	0.20	0.05
Proportion of the year the store is open (in months)	0.99	0.07
Year of establishment (/100)	0.23	0.10
Distance to the nearest store (in miles)	19.55	8.85
Competitor characteristics		
Number of competitors (/100)	0.42	0.13
Consumer characteristics		
Population size (/100,000)	0.80	0.40
% households with children	0.43	0.04
% couples without children	0.39	0.02
% households with high SES	0.21	0.01
% households with above-average SES	0.08	0.02
% households with average SES	0.32	0.03
% households with low SES	0.42	0.04
% of double-income families	0.14	0.02
Average number of low-educated	0.34	0.10
> average number of low-educated	0.47	0.19
Average number of middle-educated	0.78	0.08
> average number of middle-educated	0.10	0.06
Average number of high-educated	0.31	0.08
> average number of high-educated	0.22	0.12
Time trend	2.53	1.11
Number of observations	102	

Table 3.2: Descriptive statistics of models estimated at the zip code level

	Dependent Variable					
	Loyalty card penetration		**Visits**		**Expenditures**	
	Mean	**St.dev**	**Mean**	**St.dev**	**Mean**	**St.dev**
Dependent variables						
Loyalty card penetration	0.10	0.13				
Visits			0.40	0.44		
Expenditures					66.33	29.90
Store characteristics						
Size (in 10,000 m^2)	0.08	0.01	0.07	0.01	0.07	0.01
% female assortment	0.42	0.04	0.43	0.04	0.43	0.04
% children's assortment	0.20	0.04	0.20	0.04	0.20	0.05
Proportion of the year the store is open (in months)	0.99	0.05	1.00	0.05	1.00	0.05
Year of establishment (/100)	0.55	0.09	0.21	0.09	0.21	0.09
Competitor characteristics						
Number of competitors (/100)	0.40	0.13	0.44	0.13	0.45	0.13
Consumer characteristics						
Distance to the store (in miles)	21.93	6.14	11.85	6.14	11.23	5.94
Distance to next-nearest store (in miles)	31.63	18.87	25.72	18.87	26.21	19.38
% households with children	0.43	0.12	0.43	0.12	0.43	0.11
% couples without children	0.38	0.09	0.39	0.09	0.39	0.08
% households with high SES	0.08	0.13	0.08	0.13	0.08	0.11
% households with above-average SES	0.32	0.19	0.32	0.19	0.32	0.18
% households with average SES	0.42	0.20	0.42	0.20	0.42	0.18
% households with low SES	0.14	0.13	0.14	0.13	0.14	0.12
% of double-income families	0.21	0.05	0.21	0.05	0.21	0.05
Average number of low-educated	0.32	NA	0.34	NA	0.34	NA
> average number of low-educated	0.47	NA	0.46	NA	0.47	NA
Average number of middle-educated	0.74	NA	0.76	NA	0.78	NA
> average number of middle-educated	0.10	NA	0.10	NA	0.10	NA
Average number of high-educated	0.31	NA	0.31	NA	0.32	NA
> average number of high-educated	0.24	NA	0.24	NA	0.22	NA
Time trend	2.50	1.11	2.54	1.11	2.53	1.11
Number of observations	16,032		11,140		9,884	

3.4.2 Trade Area Sizes and Competition

We test for random store effects in the model that explains the size of store trade areas. The resulting test statistic of 16.04 is highly significant under a $\chi^2(1)$ distribution ($p < 0.001$), which allows us to reject the null hypothesis of $\eta^2 = 0$. We therefore conclude that (random) store effects should appear in the model. In the first column of Table 3.3, we provide parameter estimates for this model. We find that many store characteristics do not have a significant impact on trade area boundaries apart from a store's size, which positively affects the extent of its trade area. This finding is consistent with Reilly's law of retail gravitation, which states that consumers are willing to travel farther to visit larger stores (McGoldrick 1990).

As we show in the last column of Table 3.3, the level of competition in a particular area is positively related to the number of people living there. Attractive locations, with large populations and therefore high potential demand, not only attract the focal firm but also appeal to its competitors. However, location attractiveness depends on more than the number of potential customers; the nature of these consumers may be just as important (Duan and Mela 2009). For example, the results in Table 3.3 show a positive relation between the number of middle-educated inhabiting the store trade area and the number of competitors, which suggests that this group matches the chain's target customer profile.

3.4.3 Drivers of Store Performance: Estimation Results

We used Lagrange multiplier (LM) tests for the panel data regression models to test for potential misspecifications of the proposed models. In particular, we employ the LM test derived by Baltagi and colleagues (2003) to check for spatial error correlation and the presence of random effects in the penetration rate model. The resultant test statistic of 8,909 is highly significant under a $\chi^2(2)$ distribution

($p < 0.001$). Therefore, we reject the null hypothesis that $\lambda = \sigma_{\nu}^{2} = 0$. For the hierarchical random effects model with spatial autocorrelation, we derive the LM test statistic numerically and find that ϕ^2 and λ are significantly different from 0 for both the average expenditures and the average number of visits (all p-values < 0.05). We also test for $\eta^2 = 0$ in the ordinary random intercept models for the percentage of sales from outside the trade area (LM = 4.86; $p < 0.05$) and the percentage of sales to nonmembers (LM = 4.88; $p < 0.05$), which indicates that random store effects should appear in both models (Baltagi, Chang, and Li 1992; Breusch and Pagan 1980).

In Table 3.3, we present the parameter estimates for the models estimated at the store level, explaining the trade area perimeter, the percentage of sales from outside the trade area, the percentage of sales to nonmembers, and the number of competitors.

In Table 3.4, we provide the parameter estimates for the models that explain loyalty card penetration rates, members' average number of visits, and members' expenditures per visit. Consistent with Van Heerde and Bijmolt's (2005) results, we find that the explanatory power for the average expenditures model is the lowest. Moreover, the effects of predictor variables differ considerably across criterion variables, which supports our decision to adopt a decomposition framework. For example, the distance to the store negatively affects loyalty card penetration rate and visit frequency, whereas the average expenditures per visit depend positively on travel distance.

Store Characteristics

The parameter estimates for the share of space reserved for women's and children's clothes suggest a gender effect for all sales components (Table 3.4). If more of the assortment consists of clothes for women and children, the loyalty card penetration rate increases. This positive relation may be caused by an increase in the number of households adopting the chain's loyalty program, because they expect to derive more economic benefits from the program if the chain offers a larger assortment (Leenheer et al. 2007). The number of visits also increases if the share of space reserved for women's and children's clothes is larger. Evidence that women visit stores more frequently than men emerges from the meta-analysis provided by Pan and Zinkhan (2006). However, if men visit the store, their expenditures are generally higher than those in the women's and children's assortments.

The results in Table 3.3 show that distance to the nearest store has a positive effect on the percentage of sales outside the store's trade area. Cannibalization among different stores thus is present (Kalnins 2004): sales to members living outside the trade area of each store decrease if two stores of the same chain are located close to each other.

Competitor Characteristics

The number of competitors has positive effects on the loyalty card penetration rate but no significant impact on visit frequency or expenditures. Consumers living in zip codes close to agglomerations of clothing stores, including stores in this particular chain, are more likely to become members of the loyalty program.

The percentage of sales to nonmembers is positively affected by the number of competitors. Because nonmembers are more likely to live far from the store (Allaway, Berkowitz, and D'Souza 2003; Kivetz and Simonson 2003), this positive relationship may be caused by the effect of retail agglomeration. That is, consumers are willing to drive long distances if they can reduce the risk of product unavailability and search and compare among a large set of stores. The spatial concentration of competitors makes a shopping trip more attractive, because it facilitates visits to multiple shops with different assortments during the same trip (González-Benito and González-Benito 2005).

Consumer Characteristics

Loyalty card penetration rates are lower among members living farther from the store (Table 3.4), which is consistent with findings by Allaway et al. (2003) and Kivetz and Simonson (2003). Furthermore, members living closer to the store visit it more frequently than do members living farther away, in line with literature on spatial interaction models, which assumes that the probability of consumers patronizing a store is inversely related to distance to the store (e.g., Huff 1964). Average expenditures appear to increase with distance to the store; that is, members living farther away buy in larger quantities, perhaps because they are more likely to travel by car (Bhatnagar and Ratchford 2004).

Loyalty card penetration rates are higher for households with children than for couples and single-person households. This outcome is consistent with the results of Leenheer and colleagues (2007), who find that consumers compare the expected benefits and costs when deciding to participate in customer loyalty programs. In this view, larger households are more likely to benefit from such programs because of their higher demand levels, which will positively affect their adoption decision. On average, households with children visit the store more frequently than do singles, consistent with the results of Roy (1994), who notes that larger households buy more often than smaller households because of their higher demand levels.

The results also indicate that cannibalization between different stores may exist (Kalnins 2004), because we find a positive effect of distance to the next-nearest store on the average number of visits. That is, consumers living within the trade area of a particular store but close to another store of the same chain may visit the other store. Average expenditures per visit also are positively affected by the distance to the next-nearest store. Yet both cannibalization effects are small compared with the influence of the other factors.

Table 3.3: Parameter estimates of models estimated at the store level

Explanatory Variable	Dependent Variable							
	Trade Area Perimeter		% Unscanned Sales		% Outside Trade Area		Number of Competitors	
	Coeff	**t-value**	**Coeff**	**t-value**	**Coeff**	**t-value**	**Coeff**	**t-value**
Constant	1.292	2.87***	-2.800	-0.83	3.723	0.82	-1.176	-0.45
Store characteristics								
Size (in 10,000 m^2)	16.106	2.77***	-2.083	-0.25	-8.448	-0.79		
% female assortment			-3.445	-1.63	-2.685	-1.01		
% children's assortment			-6.114	-2.52**	-3.539	-1.15		
Proportion of the year the store is open (in months)	0.012	0.13	-0.515	-2.38**	-0.123	-0.39		
Year of establishment (/100)	-0.242	-0.42	0.269	0.34	1.472	1.44		
Distance to the nearest store (in miles)	0.001	0.26	0.008	1.30	0.024	2.86***	-0.009	-2.20**
Competitor characteristics								
Number of competitors (/100)	0.923	1.72	2.304	3.29***	1.017	1.12		
Consumer characteristics								
Population size (/100,000)	-0.175	-1.05	-0.308	-1.29	-0.208	-0.67	0.452	7.59***
% households with children			3.979	1.34	-1.681	-0.42	1.283	0.74
% couples without children			6.924	1.81	-6.100	-1.18	5.388	2.39**
% households with high SES			7.227	1.11	2.164	0.25	-9.531	-3.16***
% households with above-average SES			-3.478	-1.19	0.587	0.14	1.994	0.70
% households with average SES			-3.773	-1.25	-0.013	-0.00	1.940	0.62
% households with low SES			7.889	2.66***	6.002	1.48	2.078	0.70
% of double-income families			-1.959	-0.64	1.156	0.28	-0.188	-0.05
Average number of low-educated			-0.001	-0.00	-2.071	-1.55	-0.668	-0.65
> average number of low-educated			-0.103	-1.26	-2.053	-1.73	0.510	0.58
Average number of middle-educated			-2.927	-3.99***	-2.571	-2.57**	2.116	3.22***
> average number of middle-educated			-3.636	-4.74***	-2.579	-2.45**	1.068	1.10
Average number of high-educated			0.138	0.25	-0.254	-0.34	0.767	1.68
> average number of high-educated			-0.810	-1.56	-0.040	-0.06	0.863	1.32
Time trend	0.014	2.84***	-0.025	-0.39	-0.032	-0.39	0.100	2.58**
R^2	0.38		0.48		0.15		0.58	
Number of observations	102		102		102		102	

Notes: SES = socioeconomic status; *** $p < 0.01$ ** $p < 0.05$

Table 3.4: Parameter estimates of models estimated at the zip code level

Explanatory Variable	**Dependent Variable**					
	Loyalty Card Penetration		**Visits**		**Expenditures**	
	Coeff	**t-value**	**Coeff**	**t-value**	**Coeff**	**t-value**
Constant	-6.130	-15.85***	-0.803	-2.43**	4.890	5.06***
Store characteristics						
Size (in 10,000 m^2)	2.236	2.40**	2.964	1.86	-0.252	-0.11
% female assortment	3.938	6.04***	0.806	1.95	-1.167	-2.01**
% children's assortment	5.890	9.39***	1.705	3.91***	-1.402	-2.29**
Proportion of the year the store is open (in months)	0.907	12.45***	0.451	4.94***	-0.144	-1.37
Year of establishment (/100)	-1.797	-7.08***	-0.071	-0.45	-0.457	-2.05**
Competitor characteristics						
Number of competitors (/100)	1.732	9.52***	-0.171	-1.45	0.075	0.45
Consumer characteristics						
Distance to the store (in miles)	-0.103	-60.14***	-0.022	-40.83***	0.003	3.00***
Distance to next-nearest store (in miles)	0.002	1.80	0.003	9.28***	0.003	5.17***
% households with children	0.553	4.52***	0.102	5.86***	-0.013	-0.21
% couples without children	0.295	2.32**	0.018	0.89	-0.025	-0.33
% households with high SES	0.326	3.55***	0.003	0.18	0.153	2.34**
% households with above-average SES	0.183	2.45**	0.020	1.34	0.094	1.68
% households with average SES	0.061	0.87	-0.014	-0.97	0.131	2.31**
% households with low SES	-0.169	-1.92	-0.008	-0.46	-0.050	-0.74
% of double-income families	1.205	6.91***	-0.086	-2.86***	0.074	0.68
Average number of low-educated	0.038	1.32	-0.010	-2.21**	0.025	1.63
> average number of low-educated	0.109	3.18***	-0.009	-1.54	0.054	2.90***
Average number of middle-educated	0.085	3.86***	0.021	4.59***	0.044	2.90***
> average number of middle-educated	0.066	2.06**	0.036	5.26***	0.038	1.64
Average number of high-educated	-0.011	-0.56	-0.008	-2.22**	-0.010	-0.85
> average number of high-educated	-0.130	-4.55***	-0.026	-4.63***	-0.031	-1.76
Time trend	0.105	24.39***	-0.012	-3.41***	-0.005	-1.10
Spatial autocorrelation coeff (λ)	0.163	9.04***	0.655	81.35***	0.082	5.82***
R^2	0.52		0.53		0.04	
Number of observations	16,032		11,104		9,884	

Notes SES = socioeconomic status; *** p < 0.01 ** p < 0.05

The results in Table 3.3 show that most consumer characteristics of the trade area do not have a significant impact on the percentage of unscanned sales or the percentage of sales outside the trade area. One exception to this trend is the proportion of zip codes belonging to the store's trade area with an average or higher-than-average number of middle-educated, for which all four effects are significant (Table 3.3). These findings indicate that if the chain's target customers (i.e., middle-educated) inhabit a large part of the trade area, sales to nonmembers and members living outside the store's trade area will be lower.

Spatial Dependence

The three spatial autoregressive coefficients of the spatial-error random-effects models are all positive and significantly different from 0 (Table 3.4), which indicates that the error terms are spatially correlated. This means, zip codes close to one another tend to have similar values for each sales component due to unobserved similarities in consumer characteristics, preferences, and behavior. The degree of spatial autocorrelation for the number of visits is much greater than that for the penetration rate and the average expenditures per visit.

3.4.4 Store Performance Evaluation

We now know the effect of each variable on each sales component, but not the total effect on sales. To evaluate the predictive power of our decomposition framework, we calculate the predicted total sales and compare the obtained values with the actual sales figures for each store in the holdout sample (all stores in 2006). For this purpose, we follow the store site evaluation process presented in Figure 3.1. We first determine the geographical boundaries of store trade areas (which assigns those zip codes that reside closer to a store than the predicted value to that store) and the number of competitors, using the coefficients reported in the first and the last column of Table 3.3. Because these two variables are mutually dependent, we solve this problem iteratively. We next use the coefficients reported in Table 3.4 to predict the three sales components (i.e., loyalty card penetration rate, visit frequency, and average expenditures per visit) for each four-digit zip code in the store's trade area. Similarly, we use the coefficients reported in Table 3.3 to derive predictions of the percentage of sales to members living outside the store's trade area and the percentage of sales to nonmembers. We obtain the predicted values for each criterion variable by adding the average BLUP correction and Kelejian and Prucha (2007) terms based on Equations (3.11)–(3.13) to Equations (3.4), (3.6), and (3.8) and applying the correct transformations. When we obtain predictions for all these components, we can calculate total sales using Equations (3.1)–(3.3). As we show in Figure 3.4 (left panel), the model predicts total sales well. The correlation coefficient between observed and predicted sales equals 0.63. This correlation coefficient decreases to 0.60 when the trade area is reduced to zip codes responsible for 75% of total sales and to 0.51 when the trade area is enlarged to zip codes responsible for 95% of total sales. In sum, we conclude that the focus on the primary and secondary zones of a store's trade area produces the best results from a forecasting point of view.

For comparison, we estimated a benchmark model that ignores spatial autocorrelation, which can be obtained by substituting $\lambda = 0$ in Equations (3.5) and (3.7). The correlation coefficient decreases to 0.32 when spatial dependence between the error terms is not taken into account (right panel).[1] As expected, on the basis of econometric theory (Anselin 1988), the parameter estimates do not change dramatically when we ignore spatial autocorrelation, but the standard errors and tests based on them are substantially influenced. We find that all but one of the parameter estimates have the same sign in the proposed model as in the benchmark model without spatial effects. Hence, accounting for spatial autocorrelation leads to better results from a substantive as well as a methodological point of view. An

[1] The tables with the corresponding estimation results are available on request.

important feature of spatial models is that they borrow information from neighboring zip codes to predict each sales component at a particular location. Predictive performance of the proposed model is therefore considerably better than a model that ignores spatial dependence between zip codes in close proximity. These findings underline the usefulness of including spatial error correlation for store location and evaluation decisions.

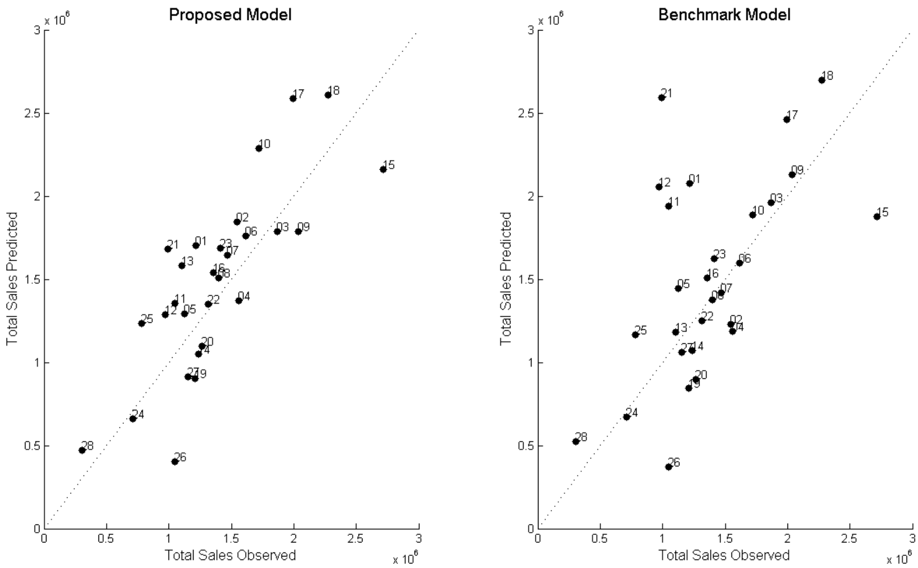


Figure 3.4: Predictive validity of (a) the proposed model and (b) the benchmark model. The data displayed in the scatterplots are observed and predicted values for total sales per store in 2006.

To illustrate how the retailer might use these results for store location evaluations, we use the estimated coefficients from Tables 3.3 and 3.4 to predict each sales component for two new stores (represented in Figure 3.4 by the numbers 27 and 28) and employ the obtained values to calculate their (potential) sales. These stores, opened in 2006, are similar to the existing stores. From Figure 3.4, we note that the decomposition framework predicts total sales for these stores quite well. Although the realized sales figures differ somewhat from the predicted figures, the model predictions are on the same order of magnitude as the observed values. The actual sales level of store 28 (€306,187), for example, is among the lowest of all stores; the same holds true for the predicted sales (€481,117). Similar conclusions apply to store 27 (€1,153,110), which belongs to a larger group of stores with average sales levels and is classified accordingly by the predicted sales figures (€929,543). Actually, predictions for all sales components, including the percentages of sales to nonmembers and members living outside the trade area, are close to the observed values. These results indicate that our modeling approach offers retailers an extremely useful tool for store location evaluation. Retailers that consider various candidate locations may want to use our proposed model to obtain estimates of potential sales for each site, which they then can employ to evaluate their entry decisions better.

By comparing the predicted and observed sales figures for each four-digit zip code, the retailer also can identify areas in which it can improve store performance. Our model predicts three sales components (i.e., loyalty card penetration rate, visit frequency, and average expenditures per visit) for each zip code in the store's trade area. Using the predicted values for each sales component, we calculate total sales to loyalty card holders (hereafter, total scanned sales) for each zip code and compare these values to the realized sales figures. We depict the results in Figure 3.5 by plotting the predicted values for each sales component and the total scanned sales for the trade area of store 27 (Figure 3.5a-d). Because the retailer also may want to know whether the store is currently over- or underperforming in certain areas, we plot the difference between the observed and predicted values for each sales component in Figure 3.5e-h. The colors of the zip codes represent predicted values for each sales component (Figure 3.5a-d) and the differences between observed and predicted values (Figure 3.5e-h). We represent the store itself by a black dot.

From Figure 3.5a, we see that, in general, the loyalty card penetration rate relates negatively to distance to the store, apart from the northeastern part of the trade area in which a large city is located and where the number of loyalty card holders is substantially lower than in other zip codes at similar distances to the store. Visit frequency also decreases with distance to the store (Figure 3.5b), meaning that loyalty card holders living closer to the store visit it more often than do those living farther away. However, average expenditures per visit increase with distance to the store (Figure 3.5c), likely because members who live farther away buy in larger quantities and are more likely to travel by car. This finding holds true for the largest part of the trade area but, again, not for the northeastern part of the market, for which we predict relatively low expenditures per visit. From Figure 3.5d, we note that if any relationship exists between total scanned sales and travel distance, it tends to be negative, which indicates that the negative relationships between the loyalty card penetration rate and visit frequency and distance dominate the positive relationship between average expenditures and travel distance.

To determine if there is room for improvement at certain locations, we plot the difference between the observed and predicted values in Figure 3.5e–h. For a retailer, it is useful to know whether scanned sales in certain areas are lower than predicted and whether this difference is due to the number of loyalty card holders, the number of visits, or the expenditures per visit. From Figure 3.5e, we determine that loyalty card penetration is lower than predicted in a large part of the trade area; therefore, the retailer should try to enhance the number of loyalty card holders by, for example, mailing a door-to-door flyer that informs consumers about the advantages of the store and its loyalty program. The number of visits falls short of the predicted values mainly in the outer parts of the trade area (Figure 3.5f). To increase the number of times existing loyalty card holders visit the store, the retailer may want to reward existing customers according to the number of times they visit the store. The spatial pattern for the average expenditures per visit indicates, as we show in Figure 3.5g, that zip codes differ substantially in the extent to which they over- or underperform. In general, the difference between the observed and predicted values tends to increase in zip codes located farther away from the store, but in these areas, the variation in prediction errors is higher as well. In Figure 3.5h, we indicate the differences in observed and predicted values for total scanned sales. Scanned sales are lower than predicted mainly in some zip codes to the west of the store and in the outer parts of the trade area. The outlet manager could investigate the local situation further by, for example, conducting a customer survey. Combined with the model results, which help explain the causes for the (negative) differences in sales, the manager could use information from the survey to develop marketing strategies specifically for certain store locations or even certain zip codes.

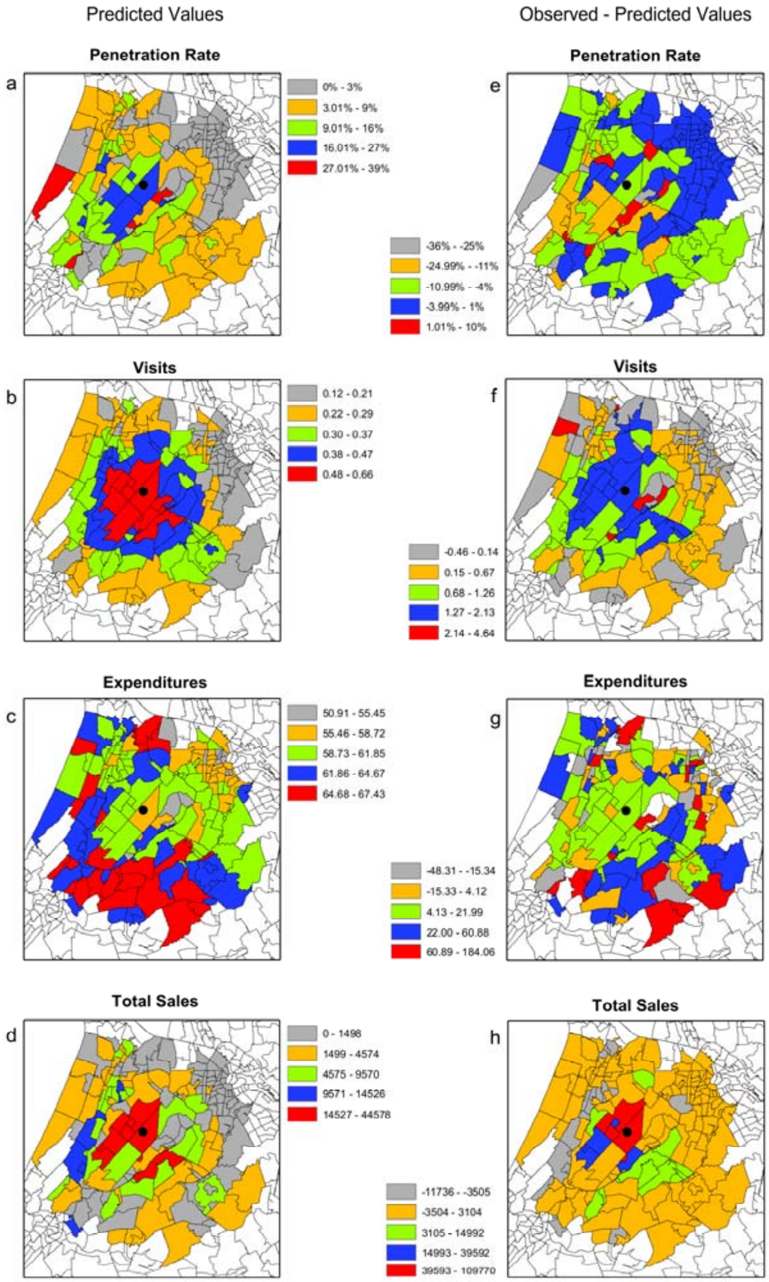


Figure 3.5: (a-d) Predicted sales components for each zip code in the store's trade area. (e-h) Differences between observed and predicted sales components for each zip code in the store's trade area.

3.4.5 Scenario Analysis

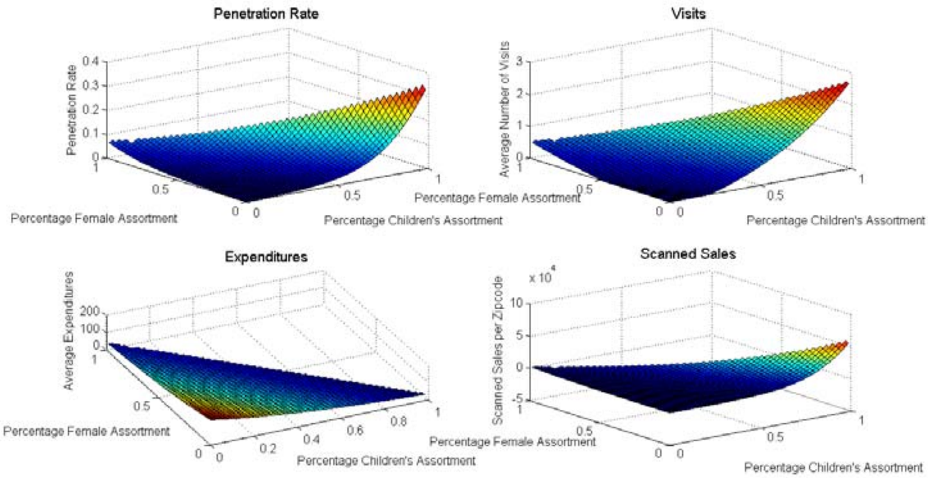


Figure 3.6: Response patterns of different sales components to a change in the relative size of the children's and women's department.

The proposed modelling approach also can answer other "what-if" questions, beyond store site evaluations. To determine the results if, for example, the relative sizes of the children's and women's departments were to change, we use the established parameter estimates, BLUP and Kelejian and Prucha (2007) correction terms to investigate the response patterns of different sales components (Figure 3.6). For this analysis, we change only the variables of interest (i.e., relative sizes of the children's and women's departments) and set all other predictor variables to their average values. Because store size is fixed, we define the relative size of the men's department as 1 minus the sum of the relative sizes of the children's and women's departments.

As we show in Figure 3.6, increasing the size of the children's or the women's department compared with the men's increases the number of loyalty card holders and visit frequency. However, the average expenditures per visit decline. To investigate the changes in scanned sales (per zip code), we determine the sales impact of an increase in the relative size of each department with an average number of households (bottom right graph, Figure 3.6)[2]. The overall effect of an increase in the size of either the children's or the women's department is positive and quite substantial. Therefore, the negative effect of an increase in the size of these departments on average expenditures per store visit is outweighed by its positive effects on the loyalty card penetration rate and the number of visits. We also find that an increase in the size of the children's department has a substantially larger sales impact than does a similar enlargement of the women's department. It therefore seems reasonable to conclude that stores belonging to this particular chain can increase their sales to loyalty card holders by enlarging the children's assortment and by reducing the fraction of floor space devoted to men's clothing.

3.5 Conclusions and Discussion

Store location is crucial to store performance because it determines store attractiveness and thus consumers' shopping decisions and spending patterns. The key objective of this chapter is to provide a general modeling approach to store location evaluation. The proposed model contributes to store location literature in several important ways. First, we use a decomposition framework to split store sales into their constituent parts, which leads to insights that remain unavailable with a model of just sales. Second, we use longitudinal customer data at the zip code level and thereby can explain differences in store performance across the trade area and over time. Third, we account for spatial dependence by specifying and estimating spatial-error models for panel data. We show how to estimate these models using longitudinal data pertaining to stores, purchase behavior, and consumer demographics.

In the empirical study, we apply our decomposition framework to 28 clothing stores of a Dutch retail chain. The customer database, supplemented with survey

[2] The number of households per zip code ranges from 0 to 11,960, with an average of 1,750.

data describing the retail environment of individual stores and commercially available geodemographic information, enables us to estimate spatial-error random-effects models that explain a substantial amount of variance in store sales. We identify several important drivers of store sales, such as travel distance, number of double-income families, and assortment composition. In particular, we find that the effect of predictor variables differs between the loyalty card penetration rate, the average number of visits, and the average expenditures per visit. For example, distance to the store negatively affects the penetration rate and the average number of visits but has a positive relationship with average expenditures per visit. We find empirical evidence of spatial dependence between the observations for each sales component as a result of unobserved similarities in consumer characteristics, preferences, and behavior.

The high predictive performance of our decomposition model underlines its usefulness for store location and evaluation decisions. Retailers who consider a number of candidate locations may want to use our proposed model to obtain estimates of potential sales for each site, which they can use to decide whether to invest in the proposed locations or not. The proposed model also makes it possible to develop criteria to evaluate the performance of existing stores. Furthermore, it can predict the sales impact of future changes within particular markets. Some of these factors entail management decisions (e.g., store location, composition), whereas other factors pertain to exogenous variables.

We acknowledge several limitations of our study that suggest directions for further research. The first limitation exists because we consider only one store chain. Although some of the findings therefore are peculiar to the retailer under consideration, the proposed store location evaluation model can be applied to other retailers or other settings that require evaluations of the location of facilities, such as health clubs, restaurants, banks, or public facilities.

Next, previous research suggests that our empirical findings might change if we choose another operationalization of the spatial weights matrix (Anselin 2002; Leenders 2002; Pace and LeSage 2004). However, in a recent study, Stakhovych and Bijmolt (2009) show that spatial models that use a first-order contiguity weight matrix perform better on average than do those that use other weight matrix specifications, due to their higher probabilities of detecting the true model and the lower mean standard error of the spatial and regression parameters. We are therefore confident in our results and do not consider other specifications.

Notwithstanding these limitations, we believe our modeling approach offers an extremely useful tool for store location evaluations. We also hope this chapter stimulates further research in this area.

Chapter 4

Evaluating Store Location and Assortment Design Based on Spatial Heterogeneity in Sales Potential[1]

4.1 Introduction

Store location and assortment composition are important strategic decisions for many retailing firms. For example, Walmart announced that it would add about 3.4 million m^2 globally in 2011, which will demand capital expenditures of $13–15 billion (Progressive Grocer 2009). The growth of retailers depends largely on their selection of geographical markets and opening of new stores. Retail chains that find the best match between the positioning of their stores and characteristics of the local market are the most likely to succeed (González-Benito, Bustos-Reyes, and Muñoz-Gallego 2007). From a consumer perspective, convenient locations and product assortment drive store choice (Briesch, Chintagunta, and Fox 2009). Therefore, the evaluation of (potential) sites should consider store location and assortment composition together.

The best regions for opening new stores are those that generate the highest demand or sales (Levy and Weitz 2004). The problem, however, is that (potential) sales are not readily observed and do not necessarily match with observed population density (Duan and Mela 2009; Garber et al. 2004). Therefore, retailers use population characteristics associated with local market potential and buying power to infer potential sales from candidate sites (e.g., Kumar and Karande 2000; Putler, Kalyanam, and Hodges 1996). Beyond their impact on store choice (Pan and Zinkhan 2006), location characteristics, such as the geodemographic profile of customers and

[1] This chapter is based on a working paper with the same title co-authored by Tammo H.A. Bijmolt and J. Paul Elhorst.

the presence of competitive stores, affect the relative attractiveness of different product categories and therefore consumer spending patterns (Inman, Shankar, and Ferraro 2004). Retailers that customize their product assortment at the store level according to location characteristics can dramatically increase their profits (e.g., Campo et al. 2000). A logical extension of retail location models therefore is to work to improve store performance by tailoring the assortment composition to local conditions. We therefore develop a methodological model that supports store location and assortment composition for new retail store locations.

This chapter contributes to extant literature in several ways. Our store location model can be used to determine the performance implications of changes in the store's assortment composition for each proposed site. Although other studies have considered location choice in combination with other marketing mix variables before, we combine outlet location and assortment composition for the first time. We also relate department-level store sales to store, competitor, and consumer characteristics and thus provide rich insights into the drivers of department sales. By investigating consumer data at the zip code level, we allow for heterogeneity in consumer characteristics and preferences over space. We further account for unobserved spatial effects in department sales by including spatially autocorrelated error terms.

The remainder of this chapter is organized as follows: We start with a summary of the relevant literature in section 4.2, then introduce our models to explain total chain sales (4.3.1), department sales shares (4.3.2), and the relative sizes of each department (4.3.4). Section 4.4 elaborates on the estimation procedure for the attraction models, followed by a discussion of how these models can help predict each department's sales share and its relative size among the store's total floor space (section 4.5). We apply our modeling framework in an empirical setting (section 4.6), the results of which we discuss in section 4.7. Sections 4.8 and 4.9 suggest two potential applications of the proposed models, namely, the evaluation of new store locations and the impact of changes in the assortment of each store. Finally, we present some conclusions, managerial implications, and directions for research.

4.2 Related Literature

Our work lies at the intersection of several research streams in marketing; we discuss three (see Table 4.1 for an overview). First, we note work on micromarketing, particularly that which shows that outlet location moderates the optimal assortment. We also position our work against spatial econometric models that take spatial dependence among observations into account. Finally, we discuss empirical economics literature, which simultaneously considers outlet location and marketing mix decisions.

4.2.1 Micromarketing and Shelf Space Allocation

Of growing interest to practitioners and academics alike is the possibility of exploiting spatial differences in category appeal by tailoring assortments to local needs. This example of micromarketing is the type of strategy adopted by retailers to tailor their marketing mix elements at the store level instead of following the same policy for every store in the chain (Montgomery 1997). Several factors are responsible for the widespread application of micromarketing, including the desire of retail managers to find new ways to differentiate themselves and lower costs at the same time (Campo et al. 2000; Desmet and Renaudin 1998; Grewal et al. 1999). Furthermore, the adoption of customer loyalty cards and the availability of scanner data have offered retailers greater possibilities for analyzing heterogeneity in consumer preferences and customizing their assortment accordingly.

Table 4.1: Empirical studies on the impact of location factors on store performance

Article	Consumer Response Variable	Marketing Decision(s)	Explanatory Variables		
			Store	Consumer	Competitor
Panel (a): Micromarketing					
Grewal et al. (1999)	Efficiency	Assortment planning	x		
Campo et al. (2000)	Shelf space elasticity	Shelf space allocation	x	x	x
Campo and Gijsbrechts (2004)	Shelf space elasticity	Shelf space allocation	x	x	x
Gijsbrechts, Campo, and Goossens (2003)	Flyer design elasticity	Feature promotions	x	x	x
Hoch et al. (1995)	Price elasticity	Pricing		x	x
Kamakura and Kang (2007)	Price elasticity	Price promotions		x	
Montgomery (1997)	Price elasticity	Pricing		x	x
Mulhern, Williams, and Leone (1998)	Price elasticity	Pricing		x	
Panel (b) Spatial models					
Bronnenberg and Mahajan (2001)	Price promotion elasticity	Price promotions	x		x
Van Dijk et al. (2004)	Shelf space elasticity	Shelf space allocation	x	x	x
Panel (c): Outlet location & Micromarketing					
Chan, Padmanabhan, and Seetharaman (2007)	Consumer utility	Outlet location & pricing	x	x	
Duan and Mela (2009)	Consumer utility	Outlet location & pricing	x		
Thomadsen (2007)	Consumer utility	Outlet location & pricing	x	x	
This study	Shelf space elasticity	Outlet location & assortment composition	x	x	x

Many micromarketing studies have concentrated on explaining differences in consumers' reactions to marketing activities across stores. They typically explain heterogeneity in consumer responses, indicated by differences in price (e.g., Hoch et al. 1995; Kamakura and Kang 2007; Mulhern, Williams, and Leone 1998) or shelf space elasticities (e.g., Desmet and Renaudin 1998), with a second-stage analysis that relates differences in response variables across stores to store and trade area characteristics. Although these studies show that consumers may respond differently to marketing activities across stores, they do not explain why these differences exist or how retailers might exploit them.

Campo and Gijsbrechts (2004) and Campo et al. (2000) use normative models to find the optimal shelf space shares for each product category in a particular outlet, according to local differences in consumer preferences and competition. In particular, Campo et al. (2000) demonstrate that category attractiveness depends on store and trade area characteristics; if more shelf space is allocated to locally appealing categories, chain profits can be improved considerably. Moreover, Campo and Gijsbrechts (2004) show that micro-marketing strategies depend on the store format, so the optimal shelf space allocation should be differentiated across formats.

This stream of literature (e.g., Campo et al. 2000; Chen et al. 1999) also suggests three ways outlet location and micromarketing may drive store sales. First, location factors such as local buying power and the number of inhabitants of the store's trade area can lead to a direct shift in store sales. These direct effects influence all product categories and are fairly well documented in retailing literature (e.g., Levy and Weitz 2004; McGoldrick 1990). Second, location factors can have differential impacts on the local attractiveness of different product categories. In particular, the space assigned to locally attractive categories may prompt a store draw effect that attracts new customers to the store, as well as prompt current customers to spend more. Third, local differences in category appeal determine shopping basket compositions, such that product categories with strong local appeal represent a larger proportion of store sales, as do their complements, whereas substitutes for locally appealing products likely sell less. In addition, category sales may vary from one location to another as a result of differences in category-specific competition.

4.2.2 Spatial Models of Demand

Spatial econometric models apply to a broad range of marketing problems (for overviews, see Bradlow et al. 2005; Bronnenberg 2005). An important advantage of these models is their potential to account for unobserved firm behavior by combining data from multiple markets. For example, Bronnenberg and Mahajan (2001) show that retailers use the current demand levels of three national brands, unobserved to the researcher, to set their advertising and promotion expenditures for these brands in a particular market. If a retailer decides to invest more heavily in markets where the brand (or product category) is already a large share player, we expect a positive correlation between expenditures on marketing variables and (anticipated) sales. Bronnenberg and Mahajan (2001) also effectively capture unobserved retailer behavior by assuming a joint spatial dependence of marketing variables and sales. Using a spatial structure in the error terms based on geographic distances among stores, they estimate more realistic (i.e., smaller) absolute price promotion elasticities than can a model that assumes the predictor variables are truly endogenous. Disentangling simultaneous effects becomes more complicated for variables that change slowly over time, as is typically the case for the amount of shelf space. In these situations, we can better exploit differences in the amount of shelf space attributed to a particular product across stores. Van Dijk et al. (2004) use an approach similar to that of Bronnenberg and Mahajan (2001) to obtain shelf space elasticities for five Dutch brands in the shampoo category. However, instead of using a spatial structure based on geographic proximity, they use similarities between store profiles to obtain reliable shelf space elasticities. They argue that if managers make decisions according to store profiles, it is more informative to use a profile-based similarity measure than geometric distance. Geometric distances are especially inadequate in situations in which stores in close proximity have more dissimilar client groups than stores located farther apart. The results of these studies thus indicate that spatial econometric models can efficiently capture geographical variations in demand and supply side factors, even if these factors are not observed.

4.2.3 Structural Models of Outlet Location

The simultaneous consideration of location and marketing mix decisions has only recently received research attention, such as by Chan, Padmanabhan, and Seetharaman (2007), Duan and Mela (2009), and Thomadsen (2007), who determine equilibrium prices or sales conditioned on outlet location and capacity. Most work in this area belongs to a developing research stream in marketing: empirical economics (Chintagunta et al. 2006). Although we adopt a reduced form approach, we note the relevant empirical economics literature here.

Specifically, Chan, Padmanabhan, and Seetharaman (2007) develop an econometric model of both geographic location and price competition among gasoline stations in Singapore. Their location model exploits the observed spatial distribution of geodemographic variables to infer potential gasoline demand at each local market. The location decision is an optimization problem, in which the Singapore government determines optimal locations from a social welfare perspective. Therefore, the total sum of traveling distances to different store locations across all consumers gets minimized, instead of some firm-specific measure of interest, such as profits or sales.

In a similar vein, Thomadsen (2007) parameterizes spatial demand as a function of observed geographic characteristics, prices, and the travel distances of consumers to stores. He uses demand estimates for each location to identify the most profitable outlet locations for fast-food chains with asymmetric competitive strengths. The results indicate that pricing and location decisions are interrelated, because optimal outlet locations depend on (price) competitive intensity and the size of the market at a particular location. However, firms may adjust their prices in response to the geographic layout of the market. The results indicate that retailers that want to open a new store should evaluate candidate sites carefully according to their potential sales, as well as the intensity of competition.

These studies by Chan, Padmanabhan, and Seetharaman (2007) and Thomadsen (2007) use observed spatial differences in consumer characteristics and consumers' distances to the store to infer spatial demand. Duan and Mela (2009) supplement such observed spatial demand factors with spatially correlated unobserved demand effects. They also augment spatial statistics with a structural model of pricing, which they use to simulate the effect of changes in outlet locations on equilibrium prices and profits.

In summary, this literature stream models consumer choice of outlet locations and shows that variations in demand across locations can be explained by product and outlet characteristics, competition, and the spatial and demographic distribution of consumers within a market. Yet outlet locations are considered endogenous, so they do thus not depend on (the retailer's conjectures about) the location decisions of others. Related literature (Mazzeo 2001; Seim 2006; Zhu and Singh 2009) asserts that entry and location decisions depend on the potential profitability of a particular market and inferences about (future) competitor decisions. Zhu and Singh (2009) find, for example, that discount retailers prefer to locate stores as close as possible to (potentially) large markets, but that the threat of competition may prevent them from doing so. Differentiation, through the adoption of different store formats or assortment, may weaken the impact of such competition.

4.2.4 This Study

We address some important issues not covered by existing literature (Table 4.1). For example, though some traditional micromarketing papers note the impact of locational factors on assortment composition, they consider only the allocation of shelf space for existing stores. As suggested by Campo et al. (2000), the appropriate assortment for new stores also can be determined from their location profile, if known, which is the focus of our study. Furthermore, existing micromarketing literature addresses assortment composition at either the product category (e.g., Campo et al. 2000; Campo and Gijsbrechts 2004) or brand (e.g., Kamakura and Kang 2007; Montgomery 1997) level, ignoring store sales at the department level.

Similar to Duan and Mela (2009), we build on spatial modeling literature by using a model that accommodates unobserved spatial effects in store sales variables. We assume that zip codes in close proximity share unobserved characteristics, which may cause spatially correlated error terms. Failing to account for spatial error autocorrelation when it exists may cause inefficiency (Anselin 1988). Therefore, we adopt models with spatial autocorrelation to account for spatial dependencies in sales components across zip codes and stores. In addition to allowing for unobserved sources of store sales, our model uses a broad range of location characteristics, such as store, competitor, and consumer characteristics observed at the zip code level, to infer department-level sales.

Unlike prior research that considers both outlet location and marketing mix decisions (Chan, Padmanabhan, and Seetharaman 2007; Thomadsen 2007), we investigate assortment composition rather than pricing—the predominant topic thus far. Furthermore, most existing work is built on the premise of Bertrand competition among firms, such that equilibrium prices depend on differences in competitive intensity rather than variations in consumers' price sensitivity across markets. These studies thereby assume that competitors offer nearly the same products, which is obviously not the case for many stores. It is therefore desirable to ascertain whether differences in consumer preferences and competition across products lead to different outcomes. We accordingly model store sales at the department level.

4.3 Model Specification

We define total chain sales as the sum of sales over all departments and zip codes in the area in which the chain operates, equivalent to the product of a department's share of sales and the total amount of sales generated by the chain at that zip code:

$$S_t = \sum_{m=1}^{M}\sum_{j=1}^{J} S_{mjt} = \sum_{m=1}^{M}\sum_{j=1}^{J} DSS_{mjt} \times S_{jt}, \tag{4.1}$$

where j refers to zip codes ($j = 1,\ldots,J$, such that J is the number of zip codes), m denotes the department ($m = 1,\ldots,M$, where M is the number of departments), and t represents a given time period ($t = 1,\ldots,T$, and T is the number of time periods). Furthermore, S_t is chain-level sales to members at time t, S_{mjt} refers to sales of department m in zip code j at time t, DSS_{mjt} is the sales share of department m in zip code j at time t, and S_{jt} represents the chain-level sales in zip code j at time t.

Customers signing up for loyalty programs must provide the retailer with their addresses, so their subsequent purchases are registered by the system. We use this information to obtain detailed insights about the mechanisms driving store sales. Loyalty program members usually are responsible for a large proportion of total sales (Singh, Hansen, and Blattberg 2006; Van Heerde and Bijmolt 2005); we therefore limit our analysis to purchases by these loyalty card customers. By gathering information about the residence of these members, we in turn determine how department-level store sales are distributed geographically.

We develop models for both variables on the right-hand side of Equation 4.1, DSS_{mjt} and S_{jt}, to capture the different ways in which the outlet location may affect store sales. Consider for example the effect of an increase in consumers' buying power in a particular area, which is common to all departments and thus might lead to higher store sales but not necessarily changes in departments' sales shares. This effect is captured by the model that explains the total amount of sales generated in a particular zip code. Other location characteristics, such as the number of children living in a particular area, also might differentially affect the attractiveness and sales shares of individual departments (e.g., children's, women's clothes), accounted for in the model by each department's share of sales at a particular location. With this modeling approach, we also capture the different ways in which assortment composition affects store sales. A store's total assortment drives its attractiveness to certain consumer groups and thus store choice, which generally then leads to higher overall sales (Briesch, Chintagunta, and Fox 2009; Chernev and Hamilton 2009). Moreover, changes in the assortment composition of a store may lead to relatively higher sales shares for categories/departments that constitute a larger proportion of a store's assortment.

Leaving aside the two variables on the right-hand side of Equation 4.1, we recognize that if retailers allocate more floor space to departments that perform well in a particular store, the reverse effect may emerge. A department's (past) sales may determine its (relative) size in the store, as a result of which department sizes should be considered endogenous. We therefore also develop a model to explain the (relative) amounts of floor space attributed to each department. In total, we model three variables: total sales (S_{jt}), departments' sales shares (DSS_{mjt}), and departments' relative floor space sizes (SS_{mit}). In the remainder of this section, we present and discuss these models, which we use to explain the variables.

4.3.1 Overall Sales

To evaluate the impact of outlet location on overall sales, we use a Tobit model. Overall sales per zip code are bounded by zero and skewed to the right. If we analyze the amount of sales generated per zip code, we should also account for nonnegativity and the large number of zero observations. Tobin (1958) developed a model to explain this type of variable, taking the following form:

$$S_{jt}^{*} = \alpha_0 + \sum_{q=1}^{Q} \alpha_q X_{jtq} + \sum_{l=1}^{L} \beta_l Z_{jtl} + \varepsilon_{jt}, \tag{4.2a}$$

and

$$S_{jt} = \begin{cases} S_{jt} & \text{if } S_{jt}^{*} > 0 \\ 0 & \text{if } S_{jt}^{*} \leq 0, \end{cases} \tag{4.2b}$$

where S_{jt}^{*} is a latent variable measuring the amount of sales generated at a particular location, which can be negative, positive, or zero. However, if the (unobserved) sales in a particular zip code, as predicted by Equation 4.2a, are negative, S_{jt}, the observed sales level will be zero, as formalized by Equation 4.2b. The set of explanatory variables includes variables observed at the store (X) and zip code (Z) levels. The store-specific explanatory variables are measured such that they refer to the nearest store.

We extend the regular Tobit model to include spatially autocorrelated error terms, because zip codes in close proximity often share unobservable characteristics (e.g., history, resources, infrastructure), and consumer spending levels in neighboring zip codes cannot be considered fully independent. If the observations $\left(j = 1, \ldots, J\right)$ are stacked in a vector for each cross-section of zip codes at time t, we can account for spatial error autocorrelation by

$$\varepsilon_t = \delta W \varepsilon_t + \varsigma_t, \tag{4.3}$$

where $E\left(\varsigma_t\right) = 0$, $Var\left(\varsigma_t\right) = \sigma^2 I_J$, and W is a row-standardized first-order contiguity weight matrix of size $\left(J \times J\right)$ that describes the spatial arrangement of zip codes.

4.3.2 Department Sales Shares

We use an attraction model to explain a department's sales share at a particular location, based on its relative size compared with the nearest store, other store attributes, and competitor and consumer characteristics observed at the zip code level. Attraction models are useful tools to analyze competitive interactions (Carpenter et al. 1988; Cooper and Nakanishi 1988; Nakanishi and Cooper 1982), because of their logical consistency; that is, market shares sum to unity, and the market shares of individual brands are between 0 and 1. Campo and Gijsbrechts (2004) and Campo et al. (2000) have used the attraction model for purposes similar to ours, but rather than explaining sales shares at the department level, they explain product category sales shares. In our setting, the attraction model takes the following form:

$$DSS_{mjt} = \frac{A_{mjt}}{\sum_{c=1}^{M} A_{cjt}}, \tag{4.4a}$$

where

$$A_{mjt} = \exp\left(\beta_{1m} + \varepsilon_{mjt}\right) SS_{mjt}^{\beta_{2m}} \prod_{k=1}^{K} \exp\left(\gamma_{km} X_{tk}\right) \prod_{n=1}^{N} \exp\left(\lambda_{nm} Z_{jtn}\right), \tag{4.4b}$$

and A_{mjt} is the attraction of department m in zip code j at time t, SS_{mjt} is the fraction of store space devoted to department m in the store closest to zip code j at time t. The set of explanatory variables includes variables observed at the store (X) and at the zip code level (Z). The store-specific explanatory variables (X) will be measured such that they refer to the nearest store.

4.3.3 Department Sizes

As we noted in the beginning of Section 4.3, there is a potential endogeneity problem with the models. In practice, a retailer may decide to allocate more floor space to departments that are selling well in a particular store (Van Dijk et al. 2004;

Van Nierop, Fok, and Franses 2008). In this case, the store manager might use previous department sales (shares) to determine the optimal assortment composition; that is, the (relative) amount of space attributed to a department is a function of its past performance. If this endogeneity is ignored, we would likely overestimate the impact of changes in a department's (relative) size on its share of sales. In technical terms, the explanatory variables SS_{mjt} in Equation 4.4b are not uncorrelated with the error term of this equation. To correct for this point, department sizes should be considered endogenous. We therefore specify a model for assortment composition, in which department sizes are considered a function of store, competitor, and (aggregated) consumer characteristics. We thus again adopt an attraction model specification, because relative department sizes also satisfy the logical consistency requirements of this model type. Hence, we model SS_{mt}^{i} using[1]

$$SS_{mt}^{i} = \frac{Att_{mit}}{\sum_{c=1}^{M} Att_{cit}}, \text{ where} \tag{4.5a}$$

$$Att_{mit} = \exp\left(\rho_m + \upsilon_{mit}\right)\prod_{g=1}^{G}\exp\left(\delta_{gm}X_{itg}\right)\prod_{r=1}^{R}\exp\left(\phi_{rm}\bar{Z}_{itr}\right), \tag{4.5b}$$

in which the $\bar{Z}_{itr}$ variables are the averages for each zip code–level variable for all zip codes for which store i is the nearest store. The set of store-specific explanatory variables now includes a variable measuring the one-period lag of department m's share of sales.

We include explanatory variables that capture a store's profile, which is defined as characteristics of the store, consumers, and competitors. Van Dijk et al. (2004) find that retailers are more likely to allocate similar amounts of shelf space to

[1] Note that the notation for the relative department sizes in Equation 4.5 differs from that in Equation 4.4, because we include a superscript i in Equation 4.5 instead of the subscript j in Equation 4.4. The SS_{mt}^{i} refers to the relative size of department m in a particular store i at time t, whereas SS_{mjt} measures the relative size of department m in the store closest to zip code j at time t.

brands for stores with comparable profiles. We include store profiles as explanatory variables rather than, as Van Dijk and colleagues (2004) do, an operationalization of the spatial weights matrix in which the weights are distances derived from a multidimensional scaling analysis. We believe that our specification offers richer insights, because we can assess the impact of each location variable separately rather, than on just the two dimensions obtained through a principal components analysis.

Following Bronnenberg and Mahajan (2001), we also account for (unobserved) spatial dependencies in relative department sizes across stores. In particular, we assume the unexplained part of department m's relative size in store i to be a function of those of neighboring stores. The rationale behind this assumption is that stores in close proximity should share unobservable characteristics that may lead to similar fractions of floor space devoted to each department.

4.4 Attraction Model Estimation

Estimating the parameters of an attraction model is not straightforward; we must transform the dependent and explanatory variables to obtain a model that is linear in parameters and satisfies logical consistency conditions. Fok, Franses, and Paap (2002) show that this model can be achieved by considering Equations 4.4b and 4.5b as the m^{th} equations in a set of M equations. Because the sales shares of all departments in a particular zip code (and all department shares of total floor space) by definition sum to 1, dependencies across equations exist, so we do not have a full rank system. We linearize the system in Equations 4.4 and 4.5 to an equivalent system of $M-1$ equations by arbitrarily selecting a base brand M^* and taking the ratio between DSS_{mjt} and the share of this brand:

$$\frac{DSS_{mjt}}{DSS_{M^*jt}} = \frac{\exp\left(\beta_{1m} + \varepsilon_{mjt}\right) SS_{mt}^{\beta_{2m}} \prod_{k=1}^{K} \exp\left(\gamma_{km} X_{tk}\right) \prod_{n=1}^{N} \exp\left(\lambda_{nm} Z_{jtn}\right)}{\exp\left(\beta_{1M^*} + \varepsilon_{M^*jt}\right) SS_{M^*t}^{\beta_{2M^*}} \prod_{k=1}^{K} \exp\left(\gamma_{kM^*} X_{tk}\right) \prod_{n=1}^{N} \exp\left(\lambda_{nM^*} Z_{jtn}\right)}. \quad (4.6)$$

If we then take the natural logarithm of both sides, we obtain a system of $M-1$ equations that is linear in the parameters. For notational convenience, we define the log-transforms $ss_{mjt} = \log\left(SS_{mjt}\right)$, log-ratios $\tilde{y}_{mjt} = \log\left(DSS_{mjt} / DSS_{M^*jt}\right)$, and the following differences: $\tilde{\beta}_{1m} = \beta_{1m} - \beta_{1M^*}$, $\tilde{\gamma}_{km} = \gamma_{km} - \gamma_{kM^*}$, $\tilde{\lambda}_{km} = \lambda_{km} - \lambda_{kM^*}$, and $\eta_{mjt} = \varepsilon_{mjt} - \varepsilon_{M^*jt}$. These tactics simplify Equation 4.6 to

$$\tilde{y}_{mjt} = \tilde{\beta}_{1m} + \beta_{2m} ss_{mjt} - \beta_{2M^*} ss_{M^*jt} + \sum_{k=1}^{K} \tilde{\gamma}_{km} X_{jtk} + \sum_{n=1}^{N} \tilde{\lambda}_{nm} Z_{jtn} + \eta_{mjt}, \quad (4.7)$$

for $m = 1, \ldots, M-1$. Consequently, we can only estimate the parameters $\tilde{\beta}_{1m}, \beta_{2m}, \beta_{2M*}, \tilde{\gamma}_{km}$, and $\tilde{\lambda}_{km}$, not (all) the model parameters in Equations 4.4 and 4.5. Yet the identification of these reduced-form parameters is sufficient to calculate elasticities (Cooper and Nakanishi 1988; Fok, Franses, and Paap 2002).[2]

We assume the $M-1$ equations to be correlated, because a zip code with high unobserved variables for the sales percentage of one department probably also has them for other departments. The transformed disturbances $\eta = \left(\eta_1 \ldots \eta_{M-1}\right)'$ therefore follow a normal distribution with mean zero and covariance matrix $\tilde{\Sigma} = L \Sigma L'$. Also, $L = \left(\mathrm{I}_{M-1} \vdots \mathrm{i}_{M-1}\right)$, in which I_{M-1} is a $\left(M-1\right)$-dimensional identity matrix, and i_{M-1} is a $\left(M-1\right)$-dimensional vector; therefore, only $\frac{1}{2} M(M-1)$ parameters of the original covariance matrix Σ of the error terms $\varepsilon = \left(\varepsilon_1 \ldots \varepsilon_M\right)'$ can be identified. Each equation contains a unique set of explanatory variables, so

[2] The coefficient β_{2M^*} is equal across the $\left(M-1\right)$ equations. This restriction is taken into account when the parameters are estimated.

estimating each equation separately by ordinary least squares would lead to inefficient parameter estimates. We use a feasible generalized least squares (GLS) procedure, better known as the seemingly unrelated regressions (SUR) model, to estimate the model parameters. The reduced-form parameters can calculate elasticities (Fok, Franses, and Paap 2002), and the elasticity of department m's relative size in store i on its percentage sales in zip code j equals:

$$\frac{\delta DSS_{mjt}}{\delta SS_{mjt}}\frac{SS_{mjt}}{DSS_{mjt}}=\left(1-DSS_{mjt}\right)\beta_{2m}, \tag{4.8}$$

According to Equation 4.8, elasticity converges to 0 if the department's share of sales goes to 1. If a department's sales share is a increasing function of its relative size in the nearest store, that is, $\beta_{2m}>0$, then the elasticity will go to 0 if SS_{mit} goes to infinity. We can increase a particular department's sales share by enlarging the floor space attributed to that department, but for higher floor space levels, the impact of space on department sales quickly levels off. This relationship, which exhibits decreasing returns to scale, is consistent with prior store space allocation literature (Desmet and Renaudin 1998).

We extend the reduced-form model in Equation 4.7 with spatial error autocorrelation to account for spatial dependencies in (the log of) each department's relative sales shares across zip codes. The spatial error model posits that (the log of) a department's relative sales share in a particular zip code depends on unobservable characteristics, correlated across space, as a result of which the error terms follow a spatial first-order autoregressive process that generates the error terms,

$$\eta_{mt}=\kappa_m W_t \eta_{mt}+\xi_{mt}, \tag{4.9}$$

where η_{mt} is a $\left(J_t\times 1\right)$ vector of spatially correlated error terms for every zip code $\left(j=1,\ldots,J_t\right)$, and ξ_{mt} is a vector of error terms which are not correlated over space. Furthermore, κ_m is the spatial autocorrelation coefficient, which can differ between departments, and W_t denotes a $\left(J_t\times J_t\right)$ non-negative matrix with zeros on the diagonal that describes the spatial arrangement of zip codes. Note that the spatial weights matrix W_t differs for each year, as indicated by the subscript t, because a department's sales shares in a particular zip code are observed only if $S_{jt} > 0$. Therefore, the number of zip codes J_t for which shares must be explained differs for each time period.

4.5 Prediction

An important element of the proposed modeling approach involves predicting the sales impacts of future changes in a store's retail environment and location and assortment changes, as well as new store openings. Prediction from the models is not straightforward though. Fok, Franses, and Paap (2002) show, for example, that it is impossible to calculate the expected values of department sales shares analytically; they can be obtained only through simulations. The market shares averaged over a large number of iterations then provide the basis for calculating expectations.

We therefore randomly draw the $(J \times 1)$ vector ξ_{mt}^{d} a certain number of D times, $d = 1, \ldots, D$, from the estimated covariance matrix $\tilde{\Sigma}$ to obtain η_{mt}^{d} by $\eta_{mt}^{d} = \left(I_{J_t} - \kappa_m W_t\right)^{-1} \xi_{mt}$. We use these disturbances and the parameters of the reduced-form model to predict relative market shares $dss_{mjt} = DSS_{mjt} \big/ DSS_{M^* jt}$ in zip code j for department m:

$$dss_{mjt}^{d} = \exp\left(\tilde{\beta}_{1m} + \eta_{mijt}^{d}\right) SS_{mjt}^{\beta_{2m}} \prod_{k=1}^{K} \exp\left(\tilde{\gamma}_{km} X_{jtk}\right) \prod_{n=1}^{N} \exp\left(\tilde{\lambda}_{nm} Z_{jtn}\right). \qquad (4.10)$$

Because $dss_{M^* ijt}^{d} = 1$, we can compute each department's share in the total amount of sales generated in a particular zip code by using the following equation:

$$DSS_{mjt}^{d} = \frac{dss_{mjt}^{d}}{\sum_{c=1}^{M} dss_{cjt}^{d}}. \tag{4.11}$$

Provided that we use a sufficiently large number of draws (D), we can finally approximate the expected values of the department shares by taking the average of the sales shares over all draws.

If we use the Tobit model in Equation 4.2 to predict overall sales at the zip code level, we also

$$\begin{aligned} E\left(S_{jt} \mid U_{jt}\right) &= P\left(S_{jt} = 0 \mid U_{jt}\right) E\left(S_{jt} \mid S_{jt} = 0, U_{jt}\right) \\ &\quad + P\left(S_{jt} > 0 \mid U_{jt}\right) E\left(S_{jt} \mid S_{jt} > 0, U_{jt}\right), \end{aligned} \tag{4.12}$$

where $U_{jt} = \left[X_{jt}\ Z_{jt}\right]$, $\theta = \left[\alpha_0\ \alpha'\ \beta'\right]'$, α is a $(Q \times 1)$ vector of α_q s, and β is a $(L \times 1)$ vector of β_l s. Because $E\left(S_{jt} \mid S_{jt} = 0, U_{jt}\right)$ equals 0 if S_{jt} is censored from below 0, $P\left(S_{jt} > 0 \mid U_{jt}\right) = \Phi\left(\frac{U_{jt}\theta}{\sigma}\right)$, and $E\left(S_{jt} \mid S_{jt} > 0, U_{jt}\right) = U_{jt}\theta + \sigma \frac{\varphi\left(U_{jt}\theta/\sigma\right)}{\Phi\left(U_{jt}\theta/\sigma\right)}$. In turn, we obtain the following expression for expected sales in zip code j at time t:

$$E\left(S_{jt} \mid U_{jt}\right) = U_{jt}\theta\Phi\left(U_{jt}\theta/\sigma\right) + \sigma\varphi\left(U_{jt}\theta/\sigma\right), \tag{4.13}$$

where $\Phi(z)$ and $\varphi(z)$ denote the cumulative density function (cdf) and probability density function (pdf) of the standard normal distribution, respectively.

4.6 Data

In addition to the store-level variables of the attraction model, we include the number of households living in a particular zip code as an additional covariate in the sales model.

The data set analyzed in this chapter also contains information about 30 stores belonging to a Dutch clothing chain. The chain's positioning is targeted at middle-class families, as reflected in the stores' average price levels and medium-quality assortments for men, women, and children. The chain uses a loyalty program to strengthen its relationship with regular customers. Participants receive a 5 percent cash reward on every purchase, credited to their loyalty cards, which can be spent freely two times a year. Although all stores offer clothes for men, women, and children, the relative amount of floor space devoted to each department differs for each store. Store size ranges from approximately 500 m^2 to 2530 m^2, of which a average of 44 and 21 percent contains women's and children's clothes, respectively.

From the chain's customer database, we collected yearly data on department-level sales for all zip codes in the Netherlands in five successive years (2002–2006); we use the first four years for estimation and the last year for validation. These data are supplemented with survey data on the retail environment in which each store operates and commercially available geodemographic information. We identify the number of competitors for each store using information obtained from a survey among store managers. Other store attributes include store size (in 10,000 m^2), the relative sizes of the various departments, the number of months a store is open in a particular year, and store age (1932 = 0). We do not include other marketing mix variables, because store managers must adhere to the marketing activities dictated by the head office.

We finally use a wide variety of socio-demographic variables observed at the zip code level to evaluate the impact of consumer characteristics on overall and department-level sales. These variables can influence store performance in several ways. First, variables such as the number of households living in a particular area and their socio-economic status determine local buying power and affect overall spending levels. Second, other variables may drive need patterns and differentially affect the sales level of each department. An example would be if many households with children create a higher demand for children's clothes in a particular region. Third, consumer variables may drive store patronage, because channels (or individual stores) differ in their attractiveness to certain consumer groups (Inman, Shankar, and Ferraro 2004). This effect is partly captured through the inclusion of variables that measure the amount of floor space allocated to each department in our model as a means to explain overall sales levels. The assortment composition determines a store's attractiveness to certain consumer groups, which in turn affects its sales level.

Table 4.2: Parameter estimates of attraction model explaining relative department sizes

Explanatory Variable	Dependent Variable			
	Women's Department Size		Men's Department Sales Share	
	Coeff	**t-value**	**Coeff**	**t-value**
Constant	3.854	2.91***	2.165	2.20**
Store characteristics				
Size (in 10,000 m^2)	-7.418	-4.29***	-2.196	-1.77
Lagged sales female assortment				
Lagged sales male assortment	0.966	3.85***	0.963	5.43***
Lagged sales children's assortment	-0.750	-1.48	-0.750	-1.48
Proportion of the year the store is open (in months)	-0.768	-1.81	-0.501	-1.52
Year of establishment (/100)	-0.737	-3.27***	-0.355	-2.23**
Competitor characteristics				
No of competitors female ass. (/100)	-0.287	-1.14	0.570	1.65
No of competitors male ass. (/100)				
Consumer characteristics				
% households with high SES	-2.055	-1.41	2.959	2.68***
% households with low SES	-0.050	-0.03	3.577	2.99***
% of foreigners	2.711	2.74***	1.244	1.56
% couples	-5.245	-2.92***	-6.222	-4.37***
% households with children	-0.937	-0.71	0.171	0.17
D2004	-0.021	-0.50	0.006	0.14
D2005	-0.036	-0.79	-0.022	-0.46
Spatial autocorrelation coeff (λ)	-0.522	0.00	0.054	0.00
R^2	0.71		0.53	
Number of observations	84		84	

Notes: *** p < 0.01 ** p < 0.05

Table 4.3: Parameter estimates of models explaining department sales shares and total sales

Explanatory Variable	**Dependent Variable**					
	Women's Department Sales Share		**Men's Department Sales Share**		**Total Sales**	
	Coeff	**t-value**	**Coeff**	**t-value**	**Coeff**	**t-value**
Constant	1.392	4.19***	1.613	4.80***	-8.199	-1.92
Store characteristics						
m^2 female assortment	0.281	3.15***			15.792	0.41
m^2 male assortment			0.506	6.50***	-19.256	-0.45
m^2 children's assortment	-0.280	-4.12***	-0.280	-4.12***	15.825	0.35
Proportion of the year the store is open (in months)					6.061	1.84
Year of establishment (/100)					1.132	0.42
Competitor characteristics						
No of competitors female ass. (/100)	0.840	2.64***				
No of competitors male ass. (/100)			2.438	4.65***		
Total no of competitors (/100)					7.306	4.01***
Consumer characteristics						
Distance to the store (in miles)	0.018	13.61***	0.013	8.51***	-0.265	-17.66***
Distance to next-nearest store (in miles)						
Number of households (/1,000)					2.991	27.78***
Average household size						
% households with high SES	-0.199	-1.19	-0.360	-2.00**	-1.220	-1.08
% households with low SES	0.883	3.91***	0.671	5.54***	2.622	2.31
% of foreigners	0.311	1.56	1.223	5.68***	-12.365	-7.47***
% couples	1.351	4.50***	1.102	3.37***	8.642	4.71***
% households with children	-2.062	-10.91***	-1.617	-7.92***	3.483	2.59**
D2004	0.057	1.18	0.131	2.49**	-0.337	-0.64
D2005	0.050	1.04	0.291	5.54***	0.437	-0.83
Spatial autocorrelation coeff (λ)	0.194	8.22***	0.212	9.13	0.841	98.98***
R^2	0.08		0.07		0.45	
Number of observations	9,798		9,798		12,024	

Notes: *** $p < 0.01$ ** $p < 0.05$

4.7 Drivers of Department Sales: Estimation Results

4.7.1 Total Sales

The parameter estimates reported in the last column of Table 4.3 indicate that the overall sales level at a particular zip code negatively depends on the distance to the nearest store in the focal chain. This finding is consistent with prior literature on spatial interaction models, which assumes that the probability of visiting a particular store is inversely related to the distance to the store. The theory of the allocation of time advocated by Becker (1965) also predicts that consumers perceive disutility from traveling, due to transportation and opportunity costs, and therefore are more likely to visit stores closer to their residence (Bawa and Ghosh 1999; Bhatnagar and Ratchford 2004).

We find a significant positive effect of the number of households on sales. Regions with large populations constitute potentially large markets, and retail outlets can generate more sales from these regions (Kumar and Karande 2000; Reinartz and Kumar 1999). Zhu and colleagues (2009) go as far as to conclude that population size is the single most important determinant of market structure. They find that in the retail discount industry, markets with no stores have significantly smaller populations than markets with stores. This finding implies that retailers use population size as a proxy for potential sales and employ such predictions to make their entry decisions. Moreover, households with children and couples spend significantly more on clothes than do single-person households at this retailer, likely because larger families have more diverse needs and buy a wider variety of products, which on average produces higher sales for these consumer groups (Bawa and Ghosh 1999).

We also find a positive and significant effect of the number of competitors on the chain’s overall sales level, which means that more rival stores enhance target store performance. This finding contradicts recent results from Zhu and Singh

(2009), who indicate that competition typically exerts a strong negative effect on store performance. Although previous studies show that the effect of competition between stores is lower if the market is large enough to support multiple stores, stores located farther away and/or those with different retail formats (Ailawadi et al. 2009; Gielens et al. 2008; Zhu, Singh, and Manuszak 2009) could produce a positive effect of competition. Nelson's theory of cumulative attraction assumes that stores in close proximity earn more business than those located far apart, because consumers visiting multiple stores have a lower risk of product unavailability (González-Benito and González-Benito 2005). Because shopping for clothes is sometimes regarded as a recreational activity, individual stores also may benefit from the presence of competitors, who offer the promise of comparison shopping (Dholakia 1999).

4.7.2 Department Shares

The second and third columns of Table 4.3 contain the parameter estimates for the models that explain sales shares in the women's and men's departments. We estimated these models using a reduced-form specification in which the children's department was the reference category, so a positive (negative) coefficient for a particular variable means that it affects the sales share of this department more (less) strongly than does the share of the children's department.

The results show that more floor space allocated to a particular department increases the sales share for this particular department. This finding confirms the results of, among others, Campo et al. (2000), Desmet and Renaudin (1998), and Van Nierop, Fok, and Franses (2008). A possible explanation for this effect indicates that products in departments with a large share of the store's total floor space receive more attention from consumers and are more likely to appear in their shopping baskets (Desmet and Renaudin 1998). If more floor space is allocated to a particular department, the retailer also can display more items, giving consumers more products from which to choose so that they are more likely to find what they want and increase sales in this particular department (Hoch et al. 1995). Similarly, more space allows the retailer to hold extra inventory and lower the risk of out-of-stocks, which may have a positive effect on sales (Desmet and Renaudin 1998).

The sales shares of the men's and women's department are positively affected by the number of competitors in each department. If we hold total sales constant, the men's and women's departments thus benefit more from the presence of competitors than does the children's department and obtain a larger share of overall sales. Travel distance also positively affects these sales shares; men's and women's departments achieve higher sales shares when consumers live farther from the store. The combination of these findings suggests that men's and women's departments benefit most from the spatial concentration of apparel stores.

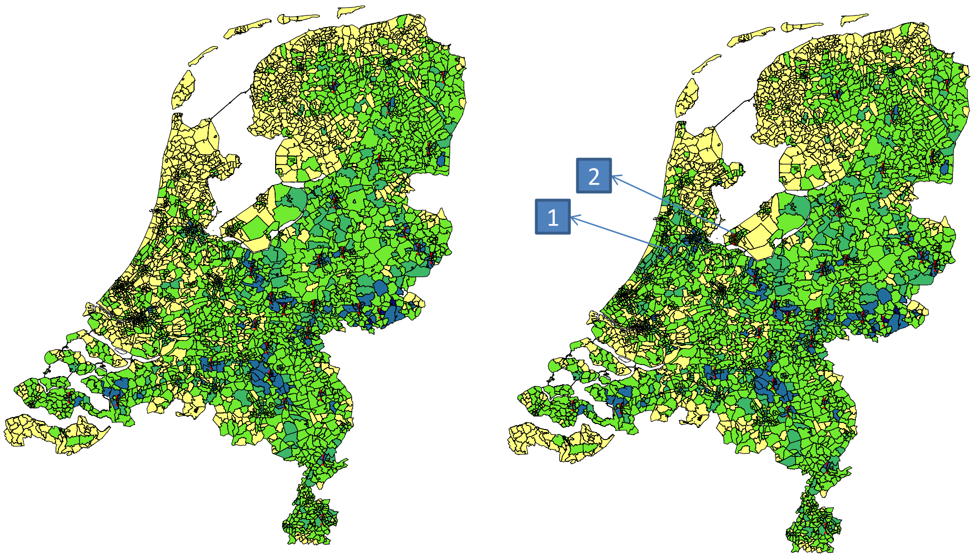


Figure 4.1: Predicted sales levels for each zip code in the Netherlands in the year 2006.
The left panel shows the sales distribution if there are no new stores, whereas the right panel depicts the predictions for total chain sales in a situation with two new stores.

4.7.3 Department Sizes

In Table 4.2, we present the parameter estimates for the models that explain the relative sizes of the men's and women's departments. We include this model to account and control for potential endogeneity in the store space allocation decision, which results when store managers allocate larger (smaller) amounts of store space to better (worse) performing departments. We find evidence of such effects, as indicated by the positive coefficients that measure the effects of (lagged) department sales on the amount of floor space allocated to this department. If departments have high (past) sales levels, more space gets allocated to them. We thus corroborate the findings of Van Dijk et al. (2004) and Van Nierop, Fok, and Franses (2008).

In addition, larger stores appear more likely to have a (larger) children's department. The year of establishment variable indicates a negative relationship with the relative sizes of the men's and women's departments; newer stores have relatively larger children's departments. We also note that though the chain's format generally should appeal to middle-class families, store managers are more likely to enlarge the men's department if the proportion of households with low and high socio-economic status greater large. The focal stores thus could be more attractive to (single) men in these consumer groups.

4.8 Potential Application: Store Location Evaluation

We now know not only which location variables drive the total amount of sales generated in a zip code but also how the performance of each individual department is affected by each variable. To examine whether the proposed model can predict sales correctly for new stores, we use the holdout sample of two newly opened stores in 2006. These two new stores have similar characteristics to the other stores and appeared in the midwestern part of the Netherlands (in Figure 4.1, numbers 1 and 2). Adding these new stores to the data set implies that the variables measuring the characteristics of and travel distances to the nearest store should change for a substantial number of the zip codes located in the western part of the country.

To see how the distribution of a department's sales changes after the two new stores open, we use the coefficients reported in the last column of Table 4.3 to predict the total sales for each four-digit zip code in the Netherlands for a hypothetical situation in which the new stores are not present in 2006 but the values of the variables measuring consumer characteristics represent those observed in 2006. The predicted spatial distribution of sales appears in Figure 4.1a. We also update the set of explanatory variables so that the variables measuring travel distances to and characteristics of the nearest store include the new stores, with the results in Figure 4.1b for the predicted sales distribution. To evaluate and compare the predicted amount of sales with the observed sales figures, we subtract the amount of sales generated in a zip code before from that after the opening of the two new stores. The next step is to determine the geographical extent of the trade area for each store, then sum the sales for all zip codes within the store's trade area. To determine the size of a store's trade area, we use the same trade area perimeters as in Chapter 3, that is, the maximum travel distance to the store for the first zip codes responsible for 85% of total sales. These distances are 13.36 and 14.17 for stores 1 and 2, respectively. If we sum the sales levels for all zip codes that belong to the trade area for each store, we can compare the realized sales figures with the predicted values. As we show in Figure 4.2, the predicted sales figures are very close to the realized values. The predicted sales for store 1 are €1,109,984, very close to the predicted sales (€988,320), and for store 2, actual sales equal €250,406, very close to the predicted value (€154,943). Therefore, the model predictions for total sales approximate the observed values well, so the model is useful for store location evaluations.

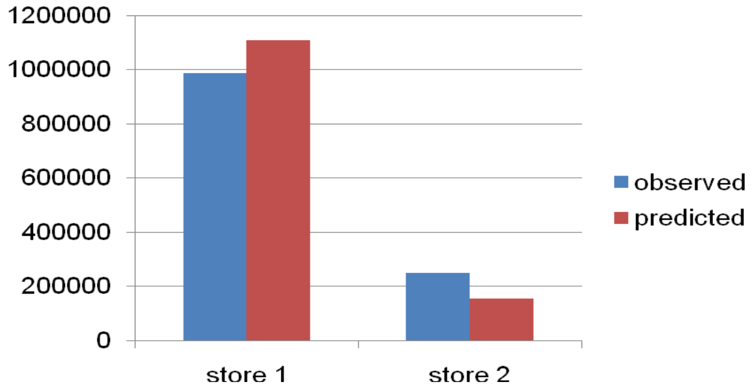


Figure 4.2: Observed and predicted sales figures for two stores opened in 2006.

4.9 Scenario Analysis: Relative Department Size

Our modeling approach also enables us to determine what happens to total sales if we change the sizes of store departments. We use the coefficients in the rightmost column of Table 4.3 to predict the overall sales level for each zip code belonging to the trade area of store 1 but change the variables that measure department sizes, setting all other predictor variables to their average values. To quantify the impact on total store sales, we change the size of each department at increments of 5 m^2. In Figure 4.3, we depict the predicted sales levels for several combinations of individual department sizes, assuming total floor space does not increase. An increase in the sizes of the children's and women's departments enhances total store sales. Specifically, increasing the size of the women's department by 1 m^2 has an effect similar in size (€15,792) to a similar enlargement of the children's department (€15,825). Increasing the size of the men's department negatively affects store sales though. This particular store therefore might increase its potential sales by enlarging the proportion of floor space it devotes to the children's and women's departments, at the expense of the men's department.

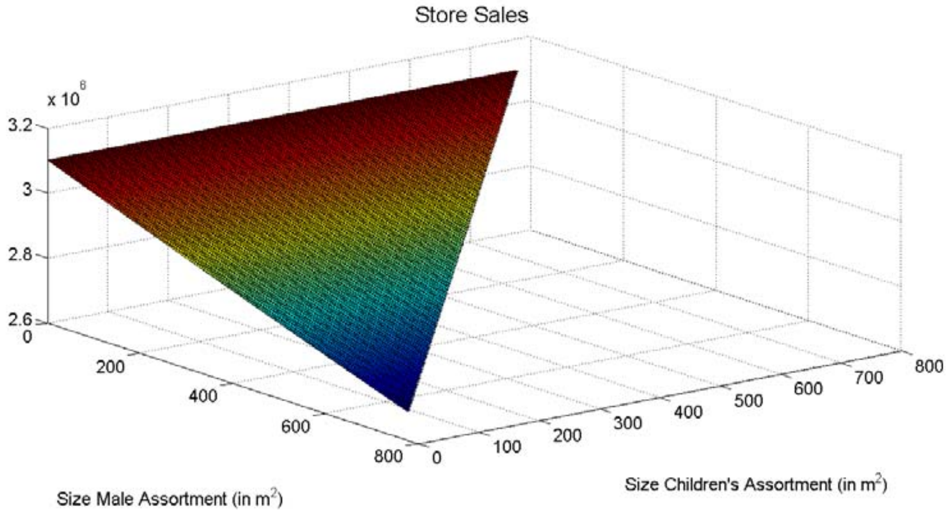


Figure 4.3: Response of total store sales to a change in the size of the children's and men's department.

4.10 Conclusions and Discussion

Travel distance to the store and assortment variety are the two of the most important factors that consumers consider when deciding where to shop (Briesch, Chintagunta, and Fox 2009). Retailers use these elements of the marketing mix to differentiate themselves from competitors, so they determine the structure of the local market and potential profit levels. In this sense, location choice and assortment composition are critical elements of the marketing mix that must be determined in combination. We therefore propose a model for store location evaluations that acknowledges the moderating effect of location characteristics on the optimal assortment composition for each store.

The model we propose in this chapter contributes to prior literature in several ways. We extend micromarketing literature, in that we evaluate the performance implications of changes in new store assortments, a consideration never adopted previously (Campo et al. 2000). We also allow for more heterogeneity in consumer characteristics than currently available models offer, in that we use consumer data observed at the zip code level rather than aggregated socio-demographic profiles for each store (Kumar and Karande 2000). Moreover, we account for spatially correlated error terms that may result from unobserved imitating behavior by consumers (Choi, Hui, and Bell 2010), retailers (Bronnenberg and Mahajan 2001), or other variables that cause spatial dependence in department sales levels across zip codes.

We test the proposed model using data from a Dutch clothing chain that operates 30 stores in various locations, each offering a retail assortment to middle-class families. The empirical study confirms previous findings from Campo et al. (2000): Location variables affect each department's sales shares differently. Travel distance to the store, for example, affects the sales shares of the men's and women's department more strongly than that of the children's department, whereas in areas with more families with children, the sales share of the children's department is greater. Total sales levels are higher in areas closer to the store, where there is intense competition and greater market potential (i.e., number of households). We further find evidence that retailers decide about the amount of floor space devoted to each department, based on each department's past performance. The size of each department in a particular year positively depends on its sales level in the previous year, so department sizes are endogenous, and we would likely overestimate the sales impact of changes in the amount of floor space allocated to each department if we were to ignore this reverse effect.

Not only does this study increase our understanding of the performance implications of tailoring assortments to local store environments, but it also has some limitations that should be addressed in further research. First, we consider only one chain of stores; the findings are therefore peculiar to the positioning of this particular chain and difficult to generalize. Second, we do not really optimize the overall performance of the chain. To do so, we would need data about the (average) unit (gross) margins for each individual store department and an estimate of the costs associated with adding new stock to the assortment of each department. Third, our modeling approach does not take into account the potential endogeneity of market structure (Zhu and Singh 2009), which would imply that we cannot assume the number of competitors, their locations, and their assortments are given. Previous research has shown that retailers note the (anticipated) location choices of their competitors when they choose locations; if we wanted to identify the optimal location and assortment for each store, we should investigate potential (future) reactions of competitors and solve location and assortment decisions for several retailers simultaneously rather than sequentially. Fourth, we consider location and assortment decisions—just two elements of the marketing mix. Finally, we consider a model with limited competition assuming that a department's attractiveness only depends on its own explanatory variables and not on those of other departments. Previous research shows the existence of cross-demand effects at the product category level (Leeflang and Parreño Selva 2010). Future research is needed to investigate whether similar effects can be observed at the level of store departments. Additional research should address local marketing strategies for other elements, such as prices (Hoch et al. 1995) and promotions (Gijsbrechts, Campo, and Goossens 2003).

Despite these limitations, we believe that the proposed model can be very valuable to retailers that want to open new chain stores and tailor their assortments to local conditions. As we have shown in our empirical study, this model effectively predicts potential sales by new store locations and the sales impacts of changes in the assortment composition, which makes it a useful tool to support these decisions.

Chapter 5

Conclusions and Further Research

5.1 Introduction

Retailing literature provides compelling evidence that spatial convenience, that is, the proximity of stores to consumers, is a primary driver of store choice, closely followed by assortment and prices (Briesch, Chintagunta, and Fox 2009; Fox and Sethuraman 2006; González-Benito, Bustos-Reyes, and Muñoz-Gallego 2007). Although this finding underlines the strategic importance of location decisions, it does not imply at all that retailers can just ignore other marketing variables when establishing a store at a particular location. Consumers are more likely to visit a store (format) that matches their consumer profile for example, so store managers can differentiate along dimensions other than space (Inman, Shankar, and Ferraro 2004). The mere presence of competitors in a particular market may force (nearby) stores to adopt different formats, through which they can mitigate the effects of spatial competition. Finally, retail properties themselves come at a cost that also determine local pricing strategies and margins (Fox and Sethuraman 2006). Thus, location is an important marketing variable that should be considered not in isolation but rather as an important facet of the retailer's overall strategy.

The main objective of this dissertation has been to develop new methodologies for store location evaluation and choice. All the proposed methodologies explain store sales at the zip code level and account for substantial heterogeneity in consumer characteristics. The models include the spatial distribution of consumer demographics; previous studies mainly have used aggregate measures for these variables. In this concluding chapter, we summarize the key findings of the research that constitutes this dissertation. We also answer the research questions formulated in Chapter 1 and identify new ones in need of additional scientific inquiry. These avenues for research are discussed further in the final section of this chapter.

5.2 Main Findings

Many retail industries today are dominated by chain outlets; the top 50 supermarket chains in the United States for example operate an average of 378 stores each (Gauri, Pauler, and Trivedi 2009). Location decisions for such large store networks are complicated by the vast number of interactions between individual stores and their powerful performance implications for the whole network. Chapter 2 therefore proposes an optimization model to support location decisions for chain stores, which can determine the optimal number of stores, store locations, and store sizes for a given region simultaneously. Because the model uses the mixed integer linear programming (MILP) paradigm, instances of considerable sizes (i.e., potential store locations) can be solved effectively, which fosters its implementation. The proposed model also presumes that consumers choose among alternative stores on the basis of travel distance to the store and store size—a more realistic assumption than that of the so-called full-capture model that only considers distance. In an empirical application to health clubs in the Rotterdam area, we show that the model replicates the findings of the full-capture model if the store size sensitivity parameter is set to 0. We further demonstrate that the proposed model generates better solutions than a model that first determines optimal store locations and then the store sizes for each location. The difference in profit levels between these two approaches depends (non-monotonically) on the size of the store size sensitivity parameter.

In Chapter 3, we consider the situation in which a retail chain wants to add a new store to its existing network of stores. We have developed a model that supports evaluations of the sales impacts of changes in a store's retail environment, location changes, and new store openings. Rather than modeling total store sales, we split this variable into its constituent parts and develop separate models for loyalty card penetration, average number of visits, and average expenditures per visit. Previous research reveals that the effect of predictor variables may differ across these sales components (Pan and Zinkhan 2006; Van Heerde and Bijmolt 2005), so a model evaluating the effect of each predictor variable on each sales component separately is likely to offer richer insights than a single model for just sales. We find accordingly that the effect of our explanatory variables differs across the loyalty card penetration rate, the average number of visits, and the average amount spent. For example, distance to the store negatively affects the penetration rate and the average number of visits but has a positive relationship with average expenditures per visit. The results further provide evidence of spatial dependence across the observations for each sales component, due to the unobserved similarities of consumers living in close proximity. We evaluate the usefulness of the decomposition framework in an empirical application involving 28 stores of a Dutch clothing retailer. The findings show that the decomposition model achieves high predictive performance and is able to forecast the performance of new stores reasonably well. The model accounting for spatial autocorrelation also performs better than the models ignoring spatial dependence and can identify areas in which store performance can be improved, according to a comparison of the realized sales figures with those predicted by a model for a store's trade area. Another application evaluates the sales impacts of changes in predictor variables such as the relative sizes of individual store departments. The results of these analyses can help retailers establish an optimal assortment for each store location.

In Chapter 4, we extend these approaches by considering a store's overall performance as the sum of the performance of individual store departments. Department-level sales likely vary by space, due to differences in their attractiveness to local consumer groups. The proposed model allows for spatial heterogeneity in consumer preferences by modeling the sales shares of each individual department at the zip code level, as a function of the characteristics of that local environment. The findings indicate that the sales levels of individual store departments are affected differentially by location variables, which offers room for improvements in store performance if the retailers were to adjust each store's assortment to the local environment. Total sales levels are higher in zip codes near focal stores with a potentially large market and in which the number of competitors is high. We also find evidence that store managers allocate more space to relatively stronger departments. Specifically, departments with better past performance (i.e., last year's sales) tend to occupy a larger proportion of a store's floor space, at the expense of other departments. Therefore, (relative) department sizes should be considered endogenously; ignoring this reverse causality likely leads to overestimations of the effects of department size on sales. The model in Chapter 4 also can predict potential sales of new stores, which makes it a useful tool for store location evaluation. We apply this model to evaluate the performance implications of changes in the assortment composition of new stores, an application never done before (Campo et al. 2000).

The combined results of the three studies presented in this dissertation add to store location literature in several important ways. First, by using disaggregated data at the zip code level, our proposed models account for more heterogeneity in consumer characteristics than do currently available models, independent of whether this heterogeneity occurs in (researcher-) observed or unobserved variables. Second, to forecast (future) values for each sales component in a particular zip code, we borrow information from neighboring zip codes. Taking this spatial dependence between zip codes in close proximity into account leads to better predictions, which enhances the usefulness of the proposed models for store location and evaluation decisions. Third, rather than evaluating the impact of drivers of store performance on a store's total sales level, we disentangle the effects of each predictor variable for various components of store sales and thereby offer richer insights than a single model for just sales would. The predictor variables have different effects on each sales component, such as sales of individual store departments. In such a situation, a retailer can enhance the performance of its individual stores by tailoring the assortment of each store to the characteristics of the local markets in which they operate.

From a practical perspective, we demonstrate applications of the proposed models to (1) identify and evaluate new store locations; (2) measure the sales impact of changes in store location, store assortment design, and the retail environment; and (3) assess the (relative) performance of existing stores. As we noted in Chapter 1, we recommend a successive application of the proposed models, in the same order as they appear in this dissertation.

5.3 Limitations and Further Research

5.3.1 Endogeneity of Market Structure

A major limitation of the empirical studies in this dissertation is that we do not always fully correct for the endogeneity of observed market structures, which potentially leads us to assess the effects of some predictor variables on store performance incorrectly. Market structure is a broad concept that entails the number, size, and distribution of buyers and sellers in a market (American Marketing Association 2010). If we contrast the elements of this definition with market structure variables, we would conclude that our featured elements are rather limited. For example, the model in Chapter 3 only considers the number of competitors as an endogenous variable, and competition is treated strictly exogenously in the other chapters. Yet recent studies using observed store location patterns across markets in the U.S. retail industry show that market entry and location decisions depend heavily on the (actual and anticipated) positioning of competitors (Hansen and Singh 2009; Seim 2006; Thomadsen 2007; Zhu, Singh, and Manuszak 2009). Therefore, though retailers prefer to locate stores closer to greater sources of demand, the (anticipated) presence of (direct) competitors may prevent them from doing so. Moreover, rivals in close proximity may weather competition better by differentiating themselves, whether geographically or demographically. If these potential interdependencies among firms' decisions are ignored, sales forecasts might be biased. Consequently, misleading conclusions could result about the effect of competition on store revenues. In such a situation, a retail manager might wrongly decide or not to open a new store in a particular market because he or she underestimates or overestimates the potential detrimental effects of competition on potential performance.

Therefore, further research should include additional information about competitors, including their physical store locations, other store attributes, and marketing mixes. Collecting these data could pose significant challenges, because information about competitors' marketing activities is likely confidential or simply not available, and the number of competitors tends to be large. However, if we succeed in obtaining more information about competitors, we might build on a small but growing literature stream that allows for strategic interactions in firms' decision making.

Seim (2006) and Zhu and Singh (2009) consider a situation in which firms not only decide whether to enter a particular market but also, if they enter, where to locate the new stores. The econometric specification of these models relies on a game-theoretic model that assumes a firm's (latent) profit in a particular location depends on distance from that location to the firm's nearest store and the relative proximity of competitive stores. The model developed by Zhu and Singh (2009) also can be extended to include the (most appropriate) store format for each location, similar to the tailoring of assortment shares to local conditions in Chapter 4. They use a random coefficients model that allows for asymmetric competitive effects due to firm-specific differences (Ailawadi et al. 2009). Therefore, these models meet some of the limitations of the models developed herein, though it should be noted that they suffer their own shortcomings. Possible extensions could include location-specific unobservable variables (such as traffic patterns and the presence of employers, see, e.g, Duan and Mela 2009); dynamics in strategic decision making, such as market entry and exit; or interactions across markets (Bronnenberg and Mahajan 2001).

5.3.2 Market Area Delineations

Another important avenue for research is the delineation of market areas. In this dissertation, we have defined market areas as zip codes that encompass 85 percent of a store's sales. However, because sales depend on the retailer's decision to open a store at a particular location, this trade area definition cannot be treated as an exogenous explanatory variable. We need another way to determine trade area boundaries, one not based on the amount of sales that a location is able to generate. Although the problem of determining a store's trade area has been addressed from many perspectives, it still seems promising to model an individual consumer's store choice decisions with a spatial choice model that allows for heterogeneous consumer preferences. This model can then predict the probability that consumers from a certain location visit a particular store while incorporating (spatial) differences in consumer demographics, store attributes, competition, and travel distances to stores. If modelers were to supplement such information with a market's geographic population density surface, they could infer local demand. Donthu (1991) and Donthu and Rust (1989) use kernel density estimation methods to find a geographic population distribution from a random sample of addresses, but researchers could also use an interpolation method such as spatial kriging (Duan and Mela 2009). If they combine store patronage predictions with the inferred population densities, they might calculate the market shares for each individual store. Furthermore, because a competitor opening a new store in a particular market changes the explanatory variables of the store choice model, it is possible to evaluate the impact of such competition on the market shares of an individual store. One of the main advantages of this approach is that it does not just dismiss location information by aggregation into subareas, such as zip codes.

It is interesting to note that kernel density estimation also has been applied to determine consumer ideal points in perceptual maps (Rust and Donthu 1988). These points represent ideal combinations of product or store attributes from a consumer's perspective and are therefore attractive sites for firms introducing a new product. For retail stores, we might define these sites along several dimensions, such as the travel distances between households and stores and store profiles, which can be defined as factors of consumer demographics and/or utility distances (Van Dijk et al. 2004). If researchers were to contrast the two (perceptual) maps obtained from such an analysis against the actual store locations, they could evaluate the extent of competition along each dimension. An outcome therefore might be that adjacent stores do not compete very much, because they have different store profiles, whereas distant stores compete more intensely because they target the same clientele. This possibility is even greater if stores in close proximity target the same customer groups with different portfolios. In this scenario, the stores might be complementary, such that each individual store benefits from small interstore distances (Arentze, Oppewal, and Timmermans 2005; Dellaert et al. 1998; Popkowski Leszczyc, Sinha, and Sahgal 2004).

In summary, the models in this dissertation might benefit from efforts to make more variables endogenous. However, we concur with Shugan (2004): Models are primarily developed to assist decision makers. In some situations, it might be better not to model particular variables, such as the availability of locations due to zoning regulations. Adding these constraints can make models unnecessarily complicated and therefore detract from their usefulness for real-world marketing practice.

References

Aboolian, R., O. Berman, and D. Krass (2007), "Competitive Facility Location and Design Problem," *European Journal of Operational Research,* 182 40-62.

Achabal, D. D., W. L. Gorr, and V. Mahajan (1982), "MULTILOC: A Multiple Store Location Decision Model," *Journal of Retailing,* 58 5-25.

Ailawadi, K. L. and K. L. Keller (2004), "Understanding Retail Branding: Conceptual Insights and Research Priorities," *Journal of Retailing,* 80 331-42.

Ailawadi, K. L., J. Zhang, A. Krishna, and M. Kruger (2009), "When Wal-Mart Enters: How Incumbent Retailers React and how this Affects their Sales Outcomes," *Journal of Marketing Research,* 47 577-93.

Allaway, A. W., D. Berkowitz, and G. D'Souza (2003), "Spatial Diffusion of a New Loyalty Program through a Retail Market," *Journal of Retailing,* 79 137-51.

American Marketing Association (2010), "Dictionary of Marketing Terms," [available at http://www.marketingpower.com/_layouts/dictionary.aspx].

Anselin, L. (1988), *Spatial Econometrics: Methods and Models.* Springer.

______(2002), "Under the Hood Issues in the Specification and Interpretation of Spatial Regression Models," *Agricultural Economics,* 27 247-67.

Applebaum, W. (1966), "Methods for Determining Store Trade Areas, Market Penetration, and Potential Sales," *Journal of Marketing Research,* 3 127-41.

Arentze, T. A., H. Oppewal, and H. J. P. Timmermans (2005), "A Multipurpose Shopping Trip Model to Assess Retail Agglomeration Effects," *Journal of Marketing Research,* 42 109-15.

Assmus, G., J. U. Farley, and D. R. Lehmann (1984), "How Advertising Affects Sales: Meta-Analysis of Econometric Results," *Journal of Marketing Research,* 21 65-74.

Baltagi, B. H. (2005), *Econometric Analysis of Panel Data.* Chichester: Wiley.

Baltagi, B. H., Y. J. Chang, and Q. Li (1992), "Monte Carlo Results on several New and Existing Tests for the Error Component Model," *Journal of Econometrics,* 54 95-120.

Baltagi, B. H. and D. Li (2004), "Prediction in the Panel Data Model with Spatial Correlation," in *Advances in Spatial Econometrics: Methodology, Tools, and Applications,* L. Anselin, R. J. G. M. Florax, and S. J. Rey, eds. Berlin: Springer, 283-295.

Baltagi, B. H., S. H. Song, and W. Koh (2003), "Testing Panel Data Regression Models with Spatial Error Correlation," *Journal of Econometrics,* 117 123-50.

Basker, E. (2007), "The Causes and Consequences of Wal-Mart's Growth," *The Journal of Economic Perspectives,* 21 177-98.

Bawa, K. and A. Ghosh (1999), "A Model of Household Grocery Shopping Behavior," *Marketing Letters,* 10 149-60.

Becker, G. S. (1965), "A Theory of the Allocation of Time," *The Economic Journal,* 75 493-517.

Benati, S. (1999), "The Maximum Capture Problem with Heterogeneous Customers," *Computers & Operations Research,* 26 1351-67.

Benati, S. and P. Hansen (2002), "The Maximum Capture Problem with Random Utilities: Problem Formulation and Algorithms," *European Journal of Operational Research,* 143 518-30.

Bhatnagar, A. and B. T. Ratchford (2004), "A Model of Retail Format Competition for Non-Durable Goods," *International Journal of Research in Marketing,* 21 39-59.

Bijmolt, T. H. A., H. J. Van Heerde, and R. G. M. Pieters (2005), "New Empirical Generalizations on the Determinants of Price Elasticity," *Journal of Marketing Research,* 42 141-56.

Bowlby, S., M. Breheny, and D. Foot (1984), "Store Location: Problems and Methods 1: Is Locating a Viable Store Becoming More Difficult?" *International Journal of Retail & Distribution Management,* 12 31-3.

Bradlow, E. T., B. Bronnenberg, G. J. Russell, N. Arora, D. R. Bell, S. D. Duvvuri, F. Ter Hofstede, C. Sismeiro, R. Thomadsen, and S. Yang (2005), "Spatial Models in Marketing," *Marketing Letters,* 16 267-78.

Breusch, T. S. and A. R. Pagan (1980), "The Lagrange Multiplier Test and its Applications to Model Specification in Econometrics," *The Review of Economic Studies,* 47 239-53.

Briesch, R. A., P. K. Chintagunta, and E. J. Fox (2009), "How does Assortment Affect Grocery Store Choice?" *Journal of Marketing Research,* 46 176-89.

Bronnenberg, B. J. (2005), "Spatial Models in Marketing Research and Practice," *Applied Stochastic Models in Business and Industry,* 21 335-44.

Bronnenberg, B. J. and V. Mahajan (2001), "Unobserved Retailer Behavior in Multimarket Data: Joint Spatial Dependence in Market Shares and Promotion Variables," *Marketing Science,* 20 284-99.

Bronnenberg, B. J. and C. Sismeiro (2002), "Using Multimarket Data to Predict Brand Performance in Markets for which no Or Poor Data Exist," *Journal of Marketing Research,* 39 1-17.

Bucklin, R. E., S. Siddarth, and J. M. Silva-Risso (2008), "Distribution Intensity and New Car Choice," *Journal of Marketing Research,* 45 473-86.

Buckner, R. W. (1998), *Site Selection: New Advancements in Methods and Technology.* Lebhar-Friedman.

Campo, K. and E. Gijsbrechts (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,* 11 369-83.

Campo, K., E. Gijsbrechts, T. Goossens, and A. Verhetsel (2000), "The Impact of Location Factors on the Attractiveness and Optimal Space Shares of Product Categories," *International Journal of Research in Marketing,* 17 255-79.

Carpenter, G. S., L. G. Cooper, D. M. Hanssens, and D. F. Midgley (1988), "Modeling Asymmetric Competition," *Marketing Science,* 7 393-412.

Chan, T. Y., V. Padmanabhan, and PB Seetharaman (2007), "An Econometric Model of Location and Pricing in the Gasoline Market," *Journal of Marketing Research,* 44 622-35.

Chen, Y., J. D. Hess, R. T. Wilcox, and Z. J. Zhang (1999), "Accounting Profits Versus Marketing Profits: A Relevant Metric for Category Management," *Marketing Science,* 18 208-29.

Chernev, A. and R. Hamilton (2009), "Assortment Size and Option Attractiveness in Consumer Choice among Retailers," *Journal of Marketing Research,* 46 410-20.

Chintagunta, P. K., T. Erdem, P. E. Rossi, and M. Wedel (2006), "Structural Modeling in Marketing: Review and Assessment," *Marketing Science,* 25 604-16.

Choi, J., S. K. Hui, and D. R. Bell (2010), "Spatiotemporal Analysis of Imitation Behavior Across New Buyers at an Online Grocery Retailer," *Journal of Marketing Research,* 47 75-89.

Clarkson, R. M., C. M. Clarke-Hill, and T. Robinson (1996), "UK Supermarket Location Assessment," *International Journal of Retail & Distribution Management,* 24 22-33.

Colomé, R., H. R. Lourenço, and D. Serra (2003), "A New Chance-Constrained Maximum Capture Location Problem," *Annals of Operations Research,* 122 121-39.

Cooper, L. G. and M. Nakanishi (1988), *Market-Share Analysis: Evaluating Competitive Marketing Effectiveness.* Springer.

Craig, C. S., A. Ghosh, and S. McLafferty (1984), "Models of the Retail Location Process: A Review," *Journal of Retailing,* 60 5-36.

Dellaert, B. G. C., T. A. Arentze, M. Bierlaire, A. W. J. Borgers, and H. J. P. Timmermans (1998), "Investigating Consumers' Tendency to Combine Multiple Shopping Purposes and Destinations," *Journal of Marketing Research,* 35 177-88.

Desmet, P. and V. Renaudin (1998), "Estimation of Product Category Sales Responsiveness to Allocated Shelf Space," *International Journal of Research in Marketing,* 15 443-57.

Dholakia, R. R. (1999), "Going Shopping: Key Determinants of Shopping Behaviors and Motivations," *International Journal of Retail & Distribution Management,* 27 154-65.

Donthu, N. (1991), "Comparing Market Areas using Kernel Density Estimation," *Journal of the Academy of Marketing Science,* 19 323-32.

Donthu, N. and R. T. Rust (1989), "Estimating Geographic Customer Densities using Kernel Density Estimation," *Marketing Science,* 8 191-203.

Drèze, X. and S. J. Hoch (1998), "Exploiting the Installed Base using Cross-Merchandising and Category Destination Programs," *International Journal of Research in Marketing,* 15 459-71.

Drezner, T., Z. Drezner, and HA Eiselt (1996), "Consistent and Inconsistent Rules in Competitive Facility Choice," *Journal of the Operational Research Society,* 47 1494-503.

Drezner, T. and H. A. Eiselt (2002), "Consumers in Competitive Location Models," in *Facility Location: Applications and Theory,* Z. Drezner and H. W. Hamacher, eds. Berlin: Springer, 151-178.

Editor Drezner, Z. (1995), *Facility Location: A Survey of Applications and Methods.* Springer Verlag.

Duan, J. A. and C. F. Mela (2009), "The Role of Spatial Demand on Outlet Location and Pricing," *Journal of Marketing Research,* 46 260-78.

Eiselt, H. A. and G. Laporte (2006), "The Maximum Capture Problem in a Weighted Network," *Journal of Regional Science,* 29 433-9.

Elhorst, J. P. (2003), "Specification and Estimation of Spatial Panel Data Models," *International Regional Science Review,* 26 244-68.

Elhorst, J. P. and A. S. Zeilstra (2007), "Labour Force Participation Rates at the Regional and National Levels of the European Union: An Integrated Analysis," *Papers in Regional Science,* 86 525-49.

Farris, P. W., M. E. Parry, and K. L. Ailawadi (1992), "Structural Analysis of Models with Composite Dependent Variables," *Marketing Science,* 11 76-94.

Fernandez, J., B. Pelegrín, F. Plastria, and B. Tóth (2007), "Solving a Huff-Like Competitive Location and Design Model for Profit Maximization in the Plane," *European Journal of Operational Research,* 179 1274-87.

Fok, D., P. H. Franses, and R. Paap (2002), "Econometric Analysis of the Market Share Attraction Model," in P. H. Franses and A. Montgomery, eds. Elsevier Science, 223-256.

Fox, E. J. and R. Sethuraman (2006), "Retail Competition," in *Retailing in the 21st Century: Current and Future Trends,* M. Krafft and M. K. Mantrala, eds. Berlin: Springer, 239-254.

Frees, E. W. (2004), *Longitudinal and Panel Data.* Cambridge: Cambridge University Press.

Garber, T., J. Goldenberg, B. Libai, and E. Muller (2004), "From Density to Destiny: Using Spatial Dimension of Sales Data for Early Prediction of New Product Success," *Marketing Science,* 23 419-28.

Gauri, D. K., J. G. Pauler, and M. Trivedi (2009), "Benchmarking Performance in Retail Chains: An Integrated Approach," *Marketing Science,* 28 502-15.

Gautschi, D. A. (1981), "Specification of Patronage Models for Retail Center Choice," *Journal of Marketing Research,* 18 162-74.

Ghosh, A. and C. S. Craig (1983), "Formulating Retail Location Strategy in a Changing Environment," *Journal of Marketing,* 47 56-68.

Ghosh, A., S. McLafferty, and C. S. Craig (1995), "Multifacility Retail Networks," in *Facility Location: A Survey of Applications and Methods,* Z. Drezner, ed. New York: Springer, 301-330.

Ghosh, A. and S. L. McLafferty (1982), "Locating Stores in Uncertain Environments: A Scenario Planning Approach," *Journal of Retailing,* 58 5-22.

Gielens, K., L. M. Van de Gucht, J. B. E. M. Steenkamp, and M. G. Dekimpe (2008), "Dancing with a Giant: The Effect of Wal-Mart's Entry into the United Kingdom on the Performance of European Retailers," *Journal of Marketing Research,* 45 519-34.

Gijsbrechts, E., K. Campo, and T. Goossens (2003), "The Impact of Store Flyers on Store Traffic and Store Sales: A Geo-Marketing Approach," *Journal of Retailing,* 79 1-16.

Goldberger, A. S. (1962), "Best Linear Unbiased Prediction in the Generalized Linear Regression Model," *Journal of the American Statistical Association,* 57 369-75.

González-Benito, Ó., C. A. Bustos-Reyes, and P. A. Muñoz-Gallego (2007), "Isolating the Geodemographic Characterisation of Retail Format Choice from the Effects of Spatial Convenience," *Marketing Letters,* 18 45-59.

González-Benito, O. and J. González-Benito (2005), "The Role of Geodemographic Segmentation in Retail Location Strategy," *International Journal of Market Research,* 47 295–316.

Gonzalez-Benito, O., P. A. Munoz-Gallego, and P. K. Kopalle (2005), "Asymmetric Competition in Retail Store Formats: Evaluating Inter-and Intra-Format Spatial Effects," *Journal of Retailing,* 81 59-73.

Greene, W. H. (2003), *Econometric Analysis.* London: Prentice Hall.

Grewal, D., M. Levy, A. Mehrotra, and A. Sharma (1999), "Planning Merchandising Decisions to Account for Regional and Product Assortment Differences," *Journal of Retailing,* 75 405-24.

Haans, H. and E. Gijsbrechts (2010), "Sales Drops from Closing Shops: Assessing the Impact of Store Outlet Closures on Retail Chain Revenue," *Journal of Marketing Research,* 47 1025-40.

Hansen, K. and V. Singh (2009), "Market Structure Across Retail Formats," *Marketing Science,* 28 656-73.

Hernandez, T. and D. Bennison (2000), "The Art and Science of Retail Location Decisions," *International Journal of Retail & Distribution Management,* 28 357-67.

Hernandez, T., D. Bennison, and S. Cornelius (1998), "The Organisational Context of Retail Locational Planning," *GeoJournal,* 45 299-308.

Hoch, S. J., B. D. Kim, A. L. Montgomery, and P. E. Rossi (1995), "Determinants of Store-Level Price Elasticity," *Journal of Marketing Research,* 32 17-29.

Huff, D. L. (1964), "Defining and Estimating a Trading Area," *The Journal of Marketing,* 28 34-8.

Huff, D. L. and R. T. Rust (1984), "Measuring the Congruence of Market Areas," *The Journal of Marketing,* 48 68-74.

IHRSA (2006), "IHRSA Global Report: State of the Health Club Industry 2006," 1-72.

Inman, J. J., V. Shankar, and R. Ferraro (2004), "The Roles of Channel-Category Associations and Geodemographics in Channel Patronage," *Journal of Marketing,* 68 51-71.

Iyer, G. and P. B. Seetharaman (2008), "Too Close to be Similar: Product and Price Competition in Retail Gasoline Markets," *Quantitative Marketing and Economics,* 6 205-34.

Jones, K. G. and J. W. Simmons (1987), *Location, Location, Location.* Toronto: Methuen.

Kalnins, A. (2004), "An Empirical Analysis of Territorial Encroachment within Franchised and Company-Owned Branded Chains," *Marketing Science,* 23 476-89.

Kamakura, W. A. and W. Kang (2007), "Chain-Wide and Store-Level Analysis for Cross-Category Management," *Journal of Retailing,* 83 159-70.

Kelejian, H. H. and I. R. Prucha (2007), "The Relative Efficiencies of various Predictors in Spatial Econometric Models Containing Spatial Lags," *Regional Science and Urban Economics,* 37 363-74.

Kivetz, R. and I. Simonson (2003), "The Idiosyncratic Fit Heuristic: Effort Advantage as a Determinant of Consumer Response to Loyalty Programs," *Journal of Marketing Research,* 40 454-67.

Kumar, V. and K. Karande (2000), "The Effect of Retail Store Environment on Retailer Performance," *Journal of Business Research,* 49 167-81.

Lafontaine, F. and K. L. Shaw (1999), "The Dynamics of Franchise Contracting: Evidence from Panel Data," *Journal of Political Economy,* 107 1041-80.

Leeflang, P. S. H. and J. Parreño Selva (2010), "Cross-Category Demand Effects of Price Promotions." Working paper, University of Groningen.

Leenders, R. T. A. J. (2002), "Modeling Social Influence through Network Autocorrelation: Constructing the Weight Matrix," *Social Networks,* 24 21-47.

Leenheer, J. and T. H. A. Bijmolt (2008), "Which Retailers Adopt a Loyalty Program? an Empirical Study," *Journal of Retailing and Consumer Services,* 15 429-42.

Leenheer, J., H. J. Van Heerde, T. H. A. Bijmolt, and A. Smidts (2007), "Do Loyalty Programs really Enhance Behavioral Loyalty? an Empirical

Analysis Accounting for Self-Selecting Members," *International Journal of Research in Marketing,* 24 31-47.

Levy, M. and B. A. Weitz (2004), *Retailing Management.* Boston: McGraw-Hill/Irwin.

Magnus, J. R. (1982), "Multivariate Error Components Analysis of Linear and Nonlinear Regression Models by Maximum Likelihood," *Journal of Econometrics,* 19 239-85.

Mazzeo, M. (2001), "Product Choice and Oligopoly Market Structure," *RAND Journal of Economics,* 33 221-42.

McGoldrick, P. J. (1990), *Retail Marketing.* McGraw-Hill London.

Mendes, A. B. and I. H. Themido (2004), "Multi-Outlet Retail Site Location Assessment," *International Transactions in Operational Research,* 11 1-18.

Miller, D. M. (1984), "Reducing Transformation Bias in Curve Fitting," *American Statistician,* 38 124-6.

Montgomery, A. L. (1997), "Creating Micro-Marketing Pricing Strategies using Supermarket Scanner Data," *Marketing Science,* 16 315-37.

Mulhern, F. J. (1997), "Retail Marketing: From Distribution to Integration," *International Journal of Research in Marketing,* 14 103-24.

Mulhern, F. J., J. D. Williams, and R. P. Leone (1998), "Variability of Brand Price Elasticities Across Retail Stores: Ethnic, Income, and Brand Determinants," *Journal of Retailing,* 74 427-46.

Nakanishi, M. and L. G. Cooper (1982), "Simplified Estimation Procedures for MCI Models," *Marketing Science,* 1 314-22.

O'Kelly, M. E. and H. J. Miller (1989), "A Synthesis of some Market Area Delimitation Models," *Growth and Change,* 20 14-33.

Pace, R. K. and J. P. LeSage (2004), "Chebyshev Approximation of Log-Determinants of Spatial Weight Matrices," *Computational Statistics & Data Analysis,* 45 179-96.

Pal, D. and J. Sarkar (2002), "Spatial Competition among Multi-Store Firms," *International Journal of Industrial Organization,* 20 163-90.

Pan, Y. and G. M. Zinkhan (2006), "Determinants of Retail Patronage: A Meta-Analytical Perspective," *Journal of Retailing,* 82 229-43.

Pauwels, K., P. S. H. Leeflang, M. L. Teerling, and K. R. E. Huizingh (2011), "Does Online Information Drive Offline Revenues? Only for Specific Products and Consumer Segments!" *Journal of Retailing,* Forthcoming .

Plastria, F. (2001), "Static Competitive Facility Location: An Overview of Optimisation Approaches," *European Journal of Operational Research,* 129 461-70.

Plastria, F. and E. Carrizosa (2004), "Optimal Location and Design of a Competitive Facility," *Mathematical Programming,* 100 247-65.

Popkowski Leszczyc, P. T. L., A. Sinha, and A. Sahgal (2004), "The Effect of Multi-Purpose Shopping on Pricing and Location Strategy for Grocery Stores," *Journal of Retailing,* 80 85-99.

Progressive Grocer (2009), "Walmart Reveals Expansion Plans," [available at http://www.progressivegrocer.com/top-story-walmart_reveals_expansion_plans-26090.html].

Putler, D. S., K. Kalyanam, and J. S. Hodges (1996), "A Bayesian Approach for Estimating Target Market Potential with Limited Geodemographic Information," *Journal of Marketing Research,* 33 134-49.

Redondo, J. L., J. Fernández, I. García, and P. M. Ortigosa (2009), "A Robust and Efficient Algorithm for Planar Competitive Location Problems," *Annals of Operations Research,* 167 87-105.

Reinartz, W. J. and V. Kumar (1999), "Store-, Market-, and Consumer-Characteristics: The Drivers of Store Performance," *Marketing Letters,* 10 5-23.

ReVelle, C. (2006), "The Maximum Capture Or "Sphere of Influence" Location Problem: Hotelling Revisited on a Network," *Journal of Regional Science,* 26 343-58.

Roy, A. (1994), "Correlates of Mall Visit Frequency," *Journal of Retailing,* 70 139-61.

Rust, R. T. and N. Donthu (1988), "A Programming and Positioning Strategy for Cable Television Networks," *Journal of Advertising,* 17 6-13.

Santos-Peñate, D. R., R. Suárez-Vega, and P. Dorta-González (2007), "The leader–follower Location Model," *Networks and Spatial Economics,* 7 45-61.

Seim, K. (2006), "An Empirical Model of Firm Entry with Endogenous Product-Type Choices," *RAND Journal of Economics,* 37 619-40.

Serra, D. and R. Colomé (2001), "Consumer Choice and Optimal Locations Models: Formulations and Heuristics," *Papers in Regional Science,* 80 439-64.

Serra, D. and C. Revelle (1995), "Competitive Location in Discrete Space," in *Facility Location: A Survey of Applications and Methods,* Z. Drezner, ed. New York: Springer, 367-386.

Shang, J., T. P. Yildirim, P. Tadikamalla, V. Mittal, and L. H. Brown (2009), "Distribution Network Redesign for Marketing Competitiveness," *Journal of Marketing,* 73 146-63.

Shugan, S. M. (2004), "Editorial: Endogeneity in Marketing Decision Models," *Marketing Science,* 23 1-3.

Singh, V. P., K. T. Hansen, and R. C. Blattberg (2006), "Market Entry and Consumer Behavior: An Investigation of a Wal-Mart Supercenter," *Marketing Science,* 25 457-76.

Stakhovych, S. and T. H. A. Bijmolt (2009), "Specification of Spatial Models: A Simulation Study on Weights Matrices," *Papers in Regional Science,* 88 389-408.

Stanley, T. J. and M. A. Sewall (1976), "Image Inputs to a Probabilistic Model: Predicting Retail Potential," *Journal of Marketing,* 40 48-53.

Steenburgh, T. J., A. Ainslie, and P. H. Engebretson (2003), "Massively Categorical Variables: Revealing the Information in Zip Codes," *Marketing Science,* 22 40-57.

Suárez-Vega, R., D. R. Santos-Peñate, and P. Dorta-González (2004), "Discretization and Resolution of the (r|Xp)-Medianoid Problem Involving Quality Criteria," *Top,* 12 111-33.

Suárez-Vega, R., D. R. Santos-Peñate, and P. Dorta-González (2004), "Competitive Multifacility Location on Networks: The (r|Xp)-Medianoid Problem," *Journal of Regional Science,* 44 569-88.

Ter Hofstede, F., M. Wedel, and J. B. E. M. Steenkamp (2002), "Identifying Spatial Segments in International Markets," *Marketing Science,* 21 160-77.

Thomadsen, R. (2007), "Product Positioning and Competition: The Role of Location in the Fast Food Industry," *Marketing Science,* 26 792.

Tobin, J. (1958), "Estimation of Relationships for Limited Dependent Variables," *Econometrica,* 26 24-36.

Tóth, B., J. Fernández, B. Pelegrín, and F. Plastria (2009), "Sequential Versus Simultaneous Approach in the Location and Design of Two New Facilities using Planar Huff-Like Models," *Computers & Operations Research,* 36 1393-405.

Van Dijk, A., H. J. Van Heerde, P. S. H. Leeflang, and D. R. Wittink (2004), "Similarity-Based Spatial Methods to Estimate Shelf Space Elasticities," *Quantitative Marketing and Economics,* 2 257-77.

Van Heerde, H. J. and T. H. A. Bijmolt (2005), "Decomposing the Promotional Revenue Bump for Loyalty Program Members Versus Nonmembers," *Journal of Marketing Research,* 42 443-57.

Van Heerde, H. J., P. S. H. Leeflang, and D. R. Wittink (2004), "Decomposing the Sales Promotion Bump with Store Data," *Marketing Science,* 23 317-34.

Van Nierop, E., D. Fok, and P. H. Franses (2008), "Interaction between Shelf Layout and Marketing Effectiveness and its Impact on Optimizing Shelf Arrangements," *Marketing Science,* 27 1065-82.

Verbeek, M. (2000), *A Guide to Modern Econometrics.* Chichester: Wiley.

Verhoef, P. C. and P. S. H. Leeflang (2009), "Understanding the Marketing Department's Influence within the Firm," *Journal of Marketing,* 73 14-37.

Wal-Mart Stores Inc. (2009), "Walmart Updates Growth Plans: Walmart U.S. Remodeling Success Leads to Accelerated Program," [available at http://investors.walmartstores.com/phoenix.zhtml?c=112761&p=irol-newsArticle&ID=1345359&highlight=].

Yang, S. and G. M. Allenby (2003), "Modeling Interdependent Consumer Preferences," *Journal of Marketing Research,* 40 282-94.

Zhang, L. and G. Rushton (2008), "Optimizing the Size and Locations of Facilities in Competitive Multi-Site Service Systems," *Computers & Operations Research,* 35 327-38.

Zhang, S. (2001), "On a Profit Maximizing Location Model," *Annals of Operations Research,* 103 251-60.

Zhu, T. and V. Singh (2009), "Spatial Competition with Endogenous Location Choices: An Application to Discount Retailing," *Quantitative Marketing and Economics,* 7 1-35.

Zhu, T., V. Singh, and M. D. Manuszak (2009), "Market Structure and Competition in the Retail Discount Industry," *Journal of Marketing Research,* 46 453-66.

Appendix: Model Specification and Estimation

A.1. The Spatial Error Random Effects Hierarchical Model

Assume that observations are sorted by store and, for each store, by time and then by zip code. Let Y_i (which can be $ln\left(NV_i\right)$ or $ln\left(EXP_i\right)$), X_i, and Z_i denote the observations, and let ε_i indicate the disturbance terms that are stacked for a particular store. The length of these vectors or matrices is store specific, because the number of zip codes within each store's trade area differs, and that number might change over time. Consequently, each vector or matrix consists of $\sum_{t=1}^{T_i} J_{it}$ observations. The full model can be written as:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_I \end{bmatrix} = \begin{bmatrix} \iota_1 \\ \iota_2 \\ \vdots \\ \iota_I \end{bmatrix} \gamma^{NV} + \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_I \end{bmatrix} \alpha + \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_I \end{bmatrix} \beta + \begin{bmatrix} \iota_1 & 0 & . & 0 \\ 0 & \iota_2 & . & 0 \\ . & . & . & . \\ 0 & 0 & . & \iota_I \end{bmatrix} \begin{bmatrix} \nu_1^{NV} \\ \nu_2^{NV} \\ \vdots \\ \nu_I^{NV} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_I \end{bmatrix}. \qquad \text{(A.1)}$$

Because $Var\left(\nu_i^{NV}\right) = \phi_{NV}^2$ and the covariance matrix Φ of the composite disturbance term $diag\left(\iota_1,\ldots,\iota_I\right) \times \nu_i^{NV} + \varepsilon$ is block diagonal, the i^{th} block diagonal is given by:

$$\Phi_i = \phi_{NV}^2 \iota_i \iota_i^{'} + \sigma_{NV}^2 \begin{bmatrix} \Omega_{iJ_{i1}} & 0 & . & 0 \\ 0 & \Omega_{iJ_{i2}} & . & 0 \\ . & . & . & . \\ 0 & 0 & . & \Omega_{iJ_{iT_i}} \end{bmatrix} = \phi_{NV}^2 \iota_i \iota_i^{'} + \sigma_{NV}^2 \Delta_i, \qquad \text{(A.2)}$$

where $\Omega_{iJ_{it}}$ is the $\left(J_{it} \times J_{it}\right)$ covariance matrix for each cross-section of zip codes that belong to the trade area of store i at time t. This matrix takes the form $\Omega_{it} = \left[\left(I_{J_{it}} - \lambda^{NV} W_{it}\right)^{'} \left(I_{J_{it}} - \lambda^{NV} W_{it}\right)\right]^{-1}$.

The inverse of Φ_i is $\Phi_i^{-1} = \dfrac{\Delta_i^{-1}}{\sigma_{NV}^2} - \dfrac{\Delta_i^{-1}\iota_i}{\sigma_{NV}^2}\left[\dfrac{1}{\phi_{NV}^2} + \dfrac{\iota_i^{'}\Delta_i^{-1}\iota_i}{\sigma_{NV}^2}\right]^{-1} \dfrac{\iota_i^{'}\Delta_i^{-1}}{\sigma_{NV}^2}$ (Frees 2004).

If the total number of observations is denoted $nobs = \sum_{i=1}^{I}\sum_{t=1}^{T_i} J_{it}$, and $\varphi^2 = \left(\phi_{NV}^2 / \sigma_{NV}^2\right)$, the log-likelihood function can be written as:

$$\begin{aligned} LogL = \quad & -\frac{nobs}{2}\log\left(2\pi\sigma_{NV}^2\right) - \frac{1}{2}\sum_{i=1}^{I}\log\left|\varphi^2 \iota_i \iota_i^{'} + \Delta_i\right| \\ & -\frac{1}{2\sigma_{NV}^2}\sum_{i=1}^{I}\tilde{e}_i^{'}\left(\varphi^2 \iota_i \iota_i^{'} + \Delta_i\right)^{-1}\tilde{e}_i, \end{aligned} \qquad \text{(A.3)}$$

where $\tilde{e}_i = Y_i - \left[\iota_i \; X_i Z_i\right]\left[\gamma_{NV} \; \alpha^{'} \beta^{'}\right]^{'}$, γ_{NV} is a scalar, α is a $\left(K \times 1\right)$ vector of α_k^{NV} s, and β is a $\left(N \times 1\right)$ vector of β_n^{NV} s.

According to Elhorst and Zeilstra (2007), the maximum likelihood estimators of the response parameters γ, α, and β (provided that X and Z do not include a lagged dependent or any endogenous explanatory variables) are equal to the generalized least squares (GLS) estimator:

$$\begin{bmatrix} \gamma \\ \alpha \\ \beta \end{bmatrix}_{GLS} = \left[\sum_{i=1}^{I} \begin{bmatrix} \iota_i & X_i & Z_i \end{bmatrix}' \Phi_i^{-1} \begin{bmatrix} \iota_i & X_i & Z_i \end{bmatrix} \right]^{-1}$$

$$\times \left[\sum_{i=1}^{I} \begin{bmatrix} \iota_i & X_i & Z_i \end{bmatrix}' \Phi_i^{-1} Y_i \right]$$

$$= \left[\sum_{i=1}^{I} \sum_{t=1}^{T_i} S_{it}^{*'} S_{it}^{*} - \sum_{i=1}^{I} \sum_{t=1}^{T_i} S_{it}^{*'} \iota_{it}^{*} \left[\frac{1}{\varphi^2} + \sum_{i=1}^{I} \sum_{t=1}^{T_i} \iota_{it}^{*'} \iota_{it}^{*} \right]^{-1} \iota_{it}^{*'} S_{it}^{*} \right]^{-1} \tag{A.4}$$

$$\times \left[\sum_{i=1}^{I} \sum_{t=1}^{T_i} S_{it}^{*'} Y_{it}^{*} - \sum_{i=1}^{I} \sum_{t=1}^{T_i} S_{it}^{*'} \iota_{it}^{*} \left[\frac{1}{\varphi^2} + \sum_{i=1}^{I} \sum_{t=1}^{T_i} \iota_{it}^{*'} \iota_{it}^{*} \right]^{-1} \iota_{it}^{*'} Y_{it}^{*} \right],$$

where $S_{it}^{*} = \left[\iota_{it}^{*} \; X_{it}^{*} Z_{it}^{*} \right]$, and the superscript * denotes the transformation $S_{it}^{*} = \left(I_{J_{it}} - \lambda^{NV} W_{it} \right) S_{it}$, applied to the variables ι_{it}, X_{it}, and Z_{it}. In addition, $\sigma_{NV}^2 = 1/_{nobs} \sum_{i=1}^{I} \tilde{e}_i' \left(\varphi^2 \iota_i \iota_i' + \Delta_i \right)^{-1} \tilde{e}_i$.

In contrast, there is no closed-form solution for λ^{NV} and φ^2. Therefore, we develop an iterative, two-step estimation procedure, in which the two sets of parameters are estimated alternately until convergence; γ^{NV}, α, β, and σ_{NV}^2, given λ^{NV} and φ^2, can be estimated using the GLS estimator in Equation A.4, whereas λ^{NV} and φ^2, given γ^{NV}, α, β, and σ_{NV}^2, can be estimated by maximizing the log-likelihood in Equation A.3. A Matlab routine of this estimation procedure can be downloaded from the Web site http://www.regroningen.nl/elhorst/.

A.2. The Spatial Error Random Effects Model

Assume that the observations are sorted by time and then, for each time period, by zip code. Let Y_t $(= logit(PR_t))$ and $U_t = [\iota_t X_t \; Z_t]$ denote observations stacked within a particular time period (Y_t and U_t consist of J observations). Further assume that $\theta = \left[\gamma^{PR}\ \alpha^{'} \beta^{'}\right]^{'}$, where α is a $\left(K \times 1\right)$ vector of α_k^{PR} s, and β is a $\left(N \times 1\right)$ vector of β_n^{PR} s. The model then can be written as:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{bmatrix} = \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_T \end{bmatrix} \theta + \delta, \;\; \delta = \left(\iota_T \otimes I_J\right)\nu + \left(I_T \otimes B^{-1}\right)\varepsilon, \tag{A.5}$$

where ι_T is a $\left(T \times 1\right)$ vector of unit elements, $\nu = \left(\nu_1^{PR}, \ldots, \nu_J^{PR}\right)$, ε is a $\left(TJ \times 1\right)$ vector of disturbance terms ε_{jt}^{PR}, and $B = \left(I_J - \lambda^{PR} W\right)$.

Anselin (1988) and Baltagi (2005) show that the log-likelihood function of this model is:

$$\begin{aligned} LogL = \;& -\frac{JT}{2}\log\left(2\pi\sigma_{PR}^2\right) - \frac{1}{2}\log\left|T\kappa^2 I_J + \left(B^{'}B\right)^{-1}\right| \\ & +\left(T-1\right)\log\left|B\right| - \frac{1}{2\kappa^2} e^{'}\left(\frac{1}{T}\iota_T\iota_T^{'} \otimes \left(T\kappa^2 I_J + \left(B^{'}B\right)^{-1}\right)\right)^{-1} e \\ & -\frac{1}{2\sigma_{PR}^2} e^{'}\left(I_T - \frac{1}{T}\iota_T\iota_T^{'}\right) \otimes \left(B^{'}B\right)e, \end{aligned} \tag{A.6}$$

where $e = \left(e_1, \ldots, e_T\right)^{'}$, $e_t = Y_t - U_t\theta$, and $\kappa^2 = \left(\sigma_\nu^2 / \sigma_{PR}^2\right)$.

Elhorst (2003) also shows that if the determinants of the matrices $T\kappa^2 I_J + (B^{'}B)^{-1}$ and B are expressed as a function of the characteristic roots of W, denoted by ω_j ($j = 1, \ldots, J$), the log-likelihood function can be rewritten as:

$$\begin{aligned} LogL = & -\frac{JT}{2}\log\left(2\pi\sigma_{PR}^2\right) - \frac{1}{2}\sum_{j=1}^{J}\log\left(1 + T\kappa^2\left(1 - \lambda^{PR}\omega_j\right)^2\right) \\ & +T\sum_{j=1}^{J}\log\left(1 - \lambda^{PR}\omega_j\right) - \frac{1}{2\sigma_{PR}^2}\sum_{t=1}^{T}\tilde{e}_t^{'}\tilde{e}_t, \end{aligned} \tag{A.7}$$

where $\tilde{e}_t = Y_t^* - U_t^* \theta$,

$$Y_t^* = P\overline{Y} + B\left(Y_t - \overline{Y}\right) = BY_t + \left(P - B\right)\overline{Y} = \left(I_J - \lambda^{PR} W\right)Y_t$$
$$+\left(P - \left(I_J - \lambda^{PR} W\right)\right)\overline{Y},$$

$U_t^* = \left(I_J - \lambda^{PR} W\right)U_t + \left(P - \left(I_J - \lambda^{PR} W\right)\right)\overline{U}$, and

P is such that $P^{'}P = \left(T\kappa^2 I_J + \left(B^{'}B\right)^{-1}\right)^{-1}$.

Consequently, the GLS estimator of the response parameters θ and σ_{PR}^2 can be computed as:

$$\theta = \left(u^{*'}u^*\right)^{-1}\left(u^{*'}y^*\right) \text{ and } \sigma_{PR}^2 = \frac{\sum_{t=1}^{T} \tilde{e}_t^{'} \tilde{e}_t}{JT}, \tag{A.8}$$

where $u^* = \begin{bmatrix} U_1^* \\ \vdots \\ U_T^* \end{bmatrix}$, and $y^* = \begin{bmatrix} Y_1^* \\ \vdots \\ Y_T^* \end{bmatrix}$.

Just as in the previous case, there is no closed-form solution for λ^{PR} and κ^2, so we need an iterative two-step estimation procedure in which the two sets of parameters get estimated alternately until convergence occurs. A Matlab routine of this estimation procedure can be downloaded from the Web site http://www.regroningen.nl/elhorst/.

Samenvatting (Summary in Dutch)

Beslissingen over toekomstige vestigingslocaties zijn van cruciaal belang voor retailers en andere ondernemers, omdat een juiste locatie essentieel is voor het trekken van klanten. Recent onderzoek wijst uit dat de reisafstand tot de winkel de belangrijkste drijfveer is voor de winkelkeuze van klanten, gevolgd door assortiment en prijzen (Briesch, Chintagunta en Fox 2009). De vestigingslocatie bepaalt dus in belangrijke mate het verzorgingsgebied van een winkel en daarmee het aantal mogelijke klanten. Daarnaast is de keuze van winkellocaties een strategische beslissing die moeilijk is terug te draaien en dus een langdurig effect heeft op de ondernemingsprestaties. Het adagium dat de drie belangrijkste factoren in retailing ‘locatie, locatie en locatie’ zijn, geldt daarom nog steeds.

Terwijl winkeliers in het verleden op basis van ervaring en kennis van de betreffende markt op ‘gevoel’ een oordeel vormden over mogelijk nieuwe vestigingsplaatsen, heeft de toegenomen complexiteit en dynamiek van de retailomgeving geresulteerd in een behoefte aan meer systematische en geavanceerde methoden voor het vinden van geschikte vestigingslocaties. Deze (veelal) kwantitatieve methoden zijn in staat het omzetpotentieel van potentiële vestigingslocaties te schatten en daarmee de bijdrage van nieuwe winkels aan de ondernemingsprestaties. Zij komen daarmee tegemoet aan de aanbevelingen van Verhoef en Leeflang (2009) die stellen dat de marketingafdeling in de toekomst duidelijker moet maken wat marketingactiviteiten opleveren. In dit proefschrift stellen wij drie nieuwe locatiemodellen voor die inspelen op een aantal recente ontwikkelingen, onder andere door gebruik te maken van gegevens op individueel klantniveau en door rekening te houden met (mogelijke) ruimtelijke afhankelijkheden tussen naburige observaties. De voorgestelde methoden bieden daarmee verbeterde mogelijkheden voor de evaluatie van winkellocaties en het beantwoorden van de vraag waar nieuwe winkels moeten worden geopend.

De centrale onderzoeksvraag van deze dissertatie luidt dan ook:

Hoe beïnvloeden de locatie en het ontwerp van een winkel de winkelprestaties en met behulp van welke methoden kunnen inzichten hierover worden gebruikt ter ondersteuning van beslissingen over vestigingslocaties en winkelontwerpen?

Deze centrale vraag is uitgesplitst in een aantal deelvragen dat in de afzonderlijke hoofdstukken van deze dissertatie worden behandeld:

1. *Wat is het optimale aantal winkels voor een bepaalde markt?*
2. *Wat zijn de beste winkelkenmerken (assortimentssamenstelling, winkelgrootte) voor elke winkellocatie?*
3. *Welke factoren bepalen de prestaties van een winkel?*
4. *Hebben deze factoren een unieke invloed op verschillende componenten van de winkelomzet?*
5. *Wat is de invloed van veranderingen in de winkelomgeving op de prestaties van een winkel?*
6. *Hoe presteren bestaande winkels in verhouding tot het omzetpotentieel van hun huidige locaties?*

Deze deelvragen worden aan de hand van drie empirische studies in de hoofdstukken 2, 3 en 4 van deze dissertatie geanalyseerd.

Resultaten en Implicaties voor de Marketingwetenschap

De eerste studie (hoofdstuk 2) ondersteunt beslissingen over vestigingslocaties voor winkels die deel uitmaken van een groter geheel, zoals een winkelketen of een franchiseformule. Locatiebeslissingen voor dergelijke winkelnetwerken zijn vrij gecompliceerd omdat beslissingen over de locatie van iedere afzonderlijke winkel consequenties kunnen hebben voor het netwerk als geheel, vanwege mogelijke interacties tussen winkels (bijvoorbeeld door kannibalisatie of door schaalvoordelen). Dit betekent dat de optimale locaties en winkelgroottes voor alle winkels in samenhang moeten worden bepaald om de meest ideale ruimtelijke configuratie van winkels te vinden. Het in hoofdstuk 2 voorgestelde optimalisatiemodel gebaseerd op gemengd geheeltallig lineair programmeren (Mixed Integer Linear Programming) kan behulpzaam zijn in dergelijke situaties, omdat het voor een bepaalde markt niet alleen het optimale aantal winkels bepaalt, maar ook de beste locaties en winkelgroottes van deze winkels. Dit hoofdstuk heeft dus betrekking op onderzoeksvragen 1 en 2 die beantwoord worden in een empirische studie naar het openen van nieuwe health clubs in Rotterdam en omgeving.

Het voorgestelde model veronderstelt dat consumenten bij hun keuze voor een bepaalde winkel een afweging maken tussen de reisafstand tot de winkel en de winkelgrootte. Omdat onderzoek heeft aangetoond dat meer factoren dan alleen reisafstand een rol spelen bij de winkelkeuze (zie bijvoorbeeld Pan en Zinkhan 2006 en Kumar en Karande 2000), zijn de assumpties van het voorgestelde model realistischer dan die van het klassieke full-capture model dat ervan uitgaat dat alleen de reisafstand bepaalt waar consumenten winkelen (Serra en ReVelle 1995). De empirische resultaten laten zien dat het voorgestelde model het full-capture model kan repliceren als de parameter die de gevoeligheid van consumenten voor winkelgrootte meet op nul wordt gezet. Daarnaast leidt toepassing van het voorgestelde model, dat de beste winkellocaties en bijbehorende winkelgroottes gelijktijdig vaststelt, tot betere resultaten dan een alternatief model dat eerst de optimale vestigingsplaatsen bepaalt en pas daarna de juiste winkelgroottes voor elke locatie. Het verschil in ondernemingsprestaties hangt niet-monotonisch af van de omvang van de parameter die de gevoeligheid van consumenten voor de winkelgrootte meet. Al met al leidt toepassing van het voorgestelde model tot betere resultaten dan vergelijkbare modellen. Deze eigenschap van het model, in combinatie met de flexibiliteit van het geheeltallig programmeren, zorgt er voor dat het model goed toepasbaar is voor de ondersteuning van locatiebeslissingen in de marketingpraktijk.

De tweede studie (hoofdstuk 3) is van toepassing op een retailer die één (of enkele) nieuwe winkel(s) wil toevoegen aan zijn huidige winkelnetwerk. Het voorgestelde model stelt de retailer in staat om: (1) een raming te maken van het omzetpotentieel van nieuwe winkellocaties, (2) het effect van veranderingen in de winkelomgeving op de winkelomzet te voorspellen en (3) de prestaties van huidige winkels te beoordelen. In plaats van een model dat de totale omzet verklaart, is het voorgestelde model in staat om de invloed van verschillende factoren op de penetratiegraad van de klantenkaart, het gemiddeld aantal winkelbezoeken en de gemiddelde bestedingen per bezoek te voorspellen. Eerder onderzoek (Pan en Zinkhan 2006, Van Heerde en Bijmolt 2005) heeft aangetoond dat de componenten van de winkelomzet verschillend worden beïnvloed door kenmerken van klanten, de winkel en concurrentie. Deze studie gaat dus in op onderzoeksvragen 2 en 3, namelijk wat de optimale winkelkenmerken zijn voor een bepaalde locatie en welke (andere) factoren de winkelprestaties beïnvloeden. De empirische studie van dit hoofdstuk heeft betrekking op 28 winkels van een Nederlandse keten van kledingzaken die het bestaande winkelnetwerk wil uitbreiden.

De resultaten van deze studie tonen aan dat onze verklarende variabelen inderdaad een verschillende invloed hebben op de genoemde componenten van de winkelomzet. Reisafstand tot de winkel heeft bijvoorbeeld een negatieve invloed op de penetratiegraad van de klantenkaart en het gemiddeld aantal winkelbezoeken, terwijl de gemiddelde bestedingen per bezoek stijgen naarmate de afstand tot de winkel groter wordt. Daarnaast tonen de resultaten het bestaan van ruimtelijke afhankelijkheden aan tussen naburige observaties voor componenten van de winkelomzet. Een vergelijking van het voorgestelde model dat rekening houdt met deze ruimtelijke afhankelijkheden en een benchmark model wat dit niet doet, laat bovendien zien dat het negeren van deze afhankelijkheden leidt tot minder goede voorspellingen van de winkelprestaties. Het voorgestelde model is goed in staat om de omzet van bestaande en nieuwe winkels te voorspellen, hetgeen het een geschikt hulpmiddel maakt voor zowel het evalueren als het selecteren van winkellocaties. Het model kan bovendien worden gebruikt om gebieden te identificeren waar de winkelprestaties achterblijven bij de voorspellingen van het model en waar dus ruimte is voor verbetering.

De derde studie is een aanvulling op de twee vorige studies, omdat het de totale winkelomzet beschouwt als de som van de omzet van alle winkelafdelingen. Het is aannemelijk dat het omzetpotentieel van winkelafdelingen een grote ruimtelijke variatie kent, omdat dit afhangt van de aantrekkelijkheid van de desbetreffende afdeling voor lokale consumentengroepen. Het voorgestelde model houdt rekening met deze ruimtelijke heterogeniteit in klantvoorkeuren door het omzetaandeel van de verschillende afdelingen per postcode te modelleren als functie van kenmerken van de retailomgeving. Omdat het tevens voor de hand ligt dat retailers meer vloeroppervlak reserveren voor relatief goed presterende winkelafdelingen, verklaart een ander model de relatieve afdelinggroottes. Weer een ander model verklaart de totale omzet in een bepaalde postcode op basis van winkel- en marktkenmerken. Deze studie richt zich daarmee op onderzoeksvragen 2, 3 en 4, namelijk het vaststellen welke factoren van invloed zijn op de ondernemingsprestaties, de afstemming van de samenstelling van het assortiment op de winkelomgeving en het evalueren van potentiële winkellocaties.

De empirische studie toont aan dat het omzetaandeel van winkelafdelingen afhankelijk is van factoren uit de winkelomgeving en dat deze factoren invloed hebben op de relatieve omzet van elke afdeling. Dit biedt retailers de mogelijkheid om de samenstelling van het winkelassortiment af te stemmen op de retailomgeving, zodat de prestaties van individuele winkels kunnen worden verbeterd. De resultaten laten bovendien zien dat de totale omzet in een postcode groter is naarmate er meer huishoudens wonen en de potentiële afzetmarkt dus groter is en naarmate er meer concurrenten actief zijn in de winkelomgeving. Deze laatste relatie wijst er wellicht op dat de winkels van deze modeketen profiteren van agglomeratie-effecten die consumenten in staat stellen om het aanbod van verschillende retailers te vergelijken, waardoor de omzet van elke afzonderlijke winkel kan toenemen. We vinden bovendien bewijs voor de veronderstelling dat deze retailer meer vloeroppervlak toekent aan afdelingen die relatief goed presteren. Dit betekent dat we de relatieve afdelinggroottes moeten beschouwen als endogene variabelen en dat we het effect van deze variabelen op de omzetaandelen van elke afdeling overschatten als we geen rekening houden met dit verband in tegengestelde richting. Tot slot laten we zien dat de resultaten van deze studie nuttig zijn voor het voorspellen van de omzet van nieuwe winkellocaties en voor het evalueren van veranderingen in de assortimentssamenstelling van elke winkel.

Als we de resultaten van deze drie studies samennemen, dan kunnen we stellen dat deze dissertatie een belangrijke bijdrage levert aan de bestaande literatuur over locatiebeslissingen. Ten eerste houden de voorgestelde modellen beter rekening met de ruimtelijke heterogeniteit in klantkenmerken door gebruik te maken van gegevens hierover op postcodeniveau, ongeacht of deze variabelen (door de onderzoeker) geobserveerd worden of niet. Ten tweede houden de in deze dissertatie voorgestelde modellen rekening met mogelijke ruimtelijke afhankelijkheden tussen observaties van winkelprestaties in naburige postcodes. De resultaten tonen aan dat het negeren van deze ruimtelijke afhankelijkheden leidt tot minder goede voorspellingen van de winkelomzet, waardoor een succesvolle toepassing van het model ter ondersteuning van locatiebeslissingen wordt bemoeilijkt. Een ander voordeel van de voorgestelde modellen is dat zij het effect van de verklarende variabelen op afzonderlijke componenten van de winkelomzet vaststellen. Een dergelijke aanpak levert betere inzichten op dan een model dat alleen de totale omzet verklaart. De resultaten in hoofdstuk 4 tonen bijvoorbeeld aan dat variabelen uit de winkelomgeving de relatieve omzet van winkelafdelingen verschillend beïnvloeden, waardoor retailers hun assortiment beter kunnen afstemmen op de lokale markt. De voorgestelde modellen zeer geschikt om te worden toegepast in de marketingpraktijk. In de volgende paragraaf beschrijven we welke toepassingsmogelijkheden de modellen hebben en welke gegevens hiervoor nodig zijn.

Implicaties voor de Marketingpraktijk

De in deze dissertatie ontwikkelde modellen hebben naast een wetenschappelijke ook een grote praktische relevantie. De modellen stellen retailers in staat om het omzetpotentieel van nieuwe vestigingslocaties te voorspellen en zij kunnen daardoor behulpzaam zijn bij het beantwoorden van de vraag of en waar (een) nieuwe winkel(s) moet(en) worden geopend. Op basis van de omzetschattingen en een inschatting van de bijbehorende kosten kan het management vervolgens beslissen of in een bepaalde locatie geïnvesteerd moet worden. De voorgestelde methoden zijn dus hulpmiddelen bij het zoeken naar de juiste winkellocaties, maar zullen niet direct het doorslaggevende antwoord geven.

De modellen kunnen daarnaast worden gebruikt om:

1. De huidige omzet van bestaande vestigingen te verklaren;
2. De toekomstige omzet van bestaande vestigingen te voorspellen;
3. De rol van verschillende beïnvloedende factoren te bepalen;

4. De invloed van veranderingen in klant-, winkel- en omgevingskenmerken op de winkel-prestaties te bepalen.

Hierbij moet worden opgemerkt dat de in deze dissertatie voorgestelde modellen in eerste instantie zijn ontwikkeld voor toepassing in een retailcontext. Toch kunnen deze modellen ook worden ingezet voor locatiebeslissingen in andere branches zoals restaurants, banken en publieke instellingen zoals bibliotheken. Om de voorgestelde modellen in deze verschillende situaties toe te passen door de samenstelling van de set van verklarende variabelen aan te passen aan de specifieke omstandigheden van de betreffende bedrijven. De omzet van een supermarkt zal bijvoorbeeld sterk afhangen van de hoeveelheid (gratis) parkeerplaatsen in de directe omgeving van de winkel, terwijl de prestaties van een restaurant vooral worden beïnvloed door de aantrekkelijkheid van de binnenstad. Hoewel de modellen dus breed inzetbaar zijn, is een belangrijke voorwaarde dat de besluitvormer beschikt over de benodigde data. Om de besproken methoden toe te passen, dienen minimaal de volgende gegevens verzameld te worden:

1. Aankoopgegevens per winkel en op individueel klantniveau;
2. Geodemografische gegevens op postcodeniveau;
3. De totale verkopen per winkel.

Daarnaast is het van belang om gegevens te verzamelen over de directe omgeving van de verschillende winkels en over de winkels zelf. Omdat deze gegevens tegenwoordig in veel branches routinematig worden verzameld, kunnen de modellen relatief eenvoudig worden toegepast door bedrijven.

In hoofdstuk 1 stellen wij dat de in deze dissertatie voorgestelde modellen complementair aan elkaar zijn en dat het combineren van de resultaten van afzonderlijke modellen tot betere locatiebeslissingen kan leiden. Dit voorstel houdt in dat het locatie-allocatiemodel uit hoofdstuk 2 wordt toegepast om het optimale aantal winkels en hun vestigingsplaatsen voor een bepaalde regio vast te stellen, waarna de regressiemodellen uit de hoofdstukken 3 en 4 afzonderlijke winkellocaties evalueren. Het model uit hoofdstuk 4 kan vervolgens behulpzaam zijn bij het bepalen van het optimale assortiment voor elke locatie. Wij spreken de wens uit dat de in deze dissertatie voorgestelde modellen waardevolle instrumenten blijken voor retailers op zoek naar de beste 'locatie, locatie en locatie'.