Advances in methods to support store location and design decisions

Hunneman, Auke

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.

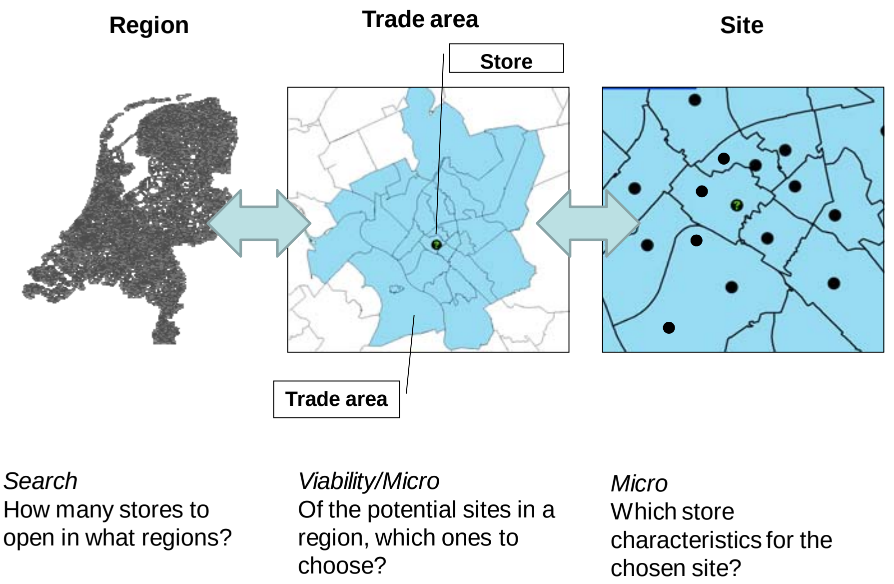

Figure 1.1: The store location decision process.
Adapted from Levy and Weitz (2004); McGoldrick (1990); Mendes and Themido (2004).

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.