Supermarket Key Attributes and Location Decisions: A Comparative Study between British and Spanish consumers

Rosa Colomé [¶]

Daniel Serra [y]

[∂] Dept. of Economics and Business, Universitat Pompeu Fabra, Trias Fargas, 25-27, Barcelona 08005, Spain. Tel: 34-3-5422696, Fax: 34-3-5421746. Mail: rosa.colome@ econ.upf.es

[ψ] Corresponding author: Dept. of Economics and Business, Universitat Pompeu Fabra, Trias Fargas, 25-27, Barcelona 08005, Spain. Tel: 34-3-5421666, Fax: 34-3-5421746. Mail: daniel.serra@econ.upf.es

Supermarket Key Attributes and Location Decisions: A Comparative Study between British and Spanish Consumers

Abstract

The Maximum Capture problem (MAXCAP) is a decision model that addresses the issue of location in a competitive environment. This paper presents a new approach to determine which store's attributes (other than distance) should be included in the new Market Capture Models and how they ought to be reflected using the Multiplicative Competitive Interaction model. The methodology involves the design and development of a survey; and the application of factor analysis and ordinary least squares. The methodology has been applied to the supermarket sector in two different scenarios: Milton Keynes (Great Britain) and Barcelona (Spain).

1. INTRODUCTION

Major food retailing companies in the UK are increasingly moving away from building large superstores and are investing in convenience food stores and middle sized supermarkets (Hunt, 1997). Moreover, 73.2% of the consumers consider supermarkets as the retail channel providing the best overall experience for food shopping (Orgel, 1997).

The trend of the middle sized supermarkets holds true for most European countries with supermarket as the main shopping destination in most of Europe, except in France, Portugal and Greece (The European, April 6, 1998). Specifically, the Spanish trends in food retailing companies reveal an the waxing fortunes of supermarkets and the wane of corner-shops and convenience shops (Pau and Navasmés, 1998).

Given this new trend, one would expect retailing companies to put their hopes for growth in supermarkets. Research findings reveal that the main reasons for choosing this format of food retailing are price (35.2 %), location (19.7 %), quality (18.8%) and variety (13.1 %) (Orgel, 1997). While price, quality and variety can be changed to deal with competitors' policies, the same cannot be said of location which, to all intents and purposes, represents a fixed one-time investment of a unique, unchangeable nature.

In this environment, supermarket's location could be the most important determinant of the supermarket's success or failure. A well-known aphorism states, "the most important attributes of stores are location, location and location". It is therefore not surprising that considerable study has been devoted to this point. One approach to this field is that enshrined by Competitive Location literature in discrete space.

Competitive Location literature in discrete space addresses the issue of optimally locating firms that compete for clients in space. A competitive location model is such that there is more than one firm competing in the spatial market and with interaction between them. The location decision of a firm will affect not only its market share, but also its competitors market shares (Serra and ReVelle, 1996). Freisz, et.al. (Freisz, et.al., 1988) pointed out that one of the three competitive network facility location models that were "likely to serve as foundations for future models" is ReVelle's Maximum Capture Problem (MAXCAP) (ReVelle, 1986). Traditionally, the discrete location modeling literature has been successfully applied to locate public sector services, where the main aim is to optimize some measure of service quality in terms of access (e.g., maximizing service coverage or minimizing average distance to the service). Actually, new models are appearing within a private sector context, where there is competition among providers of the service. The models employed focus on solving problems like hierarchical services and scenarios with different demand and/ or competitor locations. To date, this literature has assumed that consumers shop at the closest store supplying a specific product or service. However, one needs to ask whether this assumption reflects consumer behavior. It seems more realistic to admit that consumers do not merely consider distance when making-choosing retail shops.

Store-Choice literature studies the key variables that influence a consumer when deciding where shop as well as the interaction between these variables. Literature on the subject reveals that distance is not the only variable consumers take into account when deciding where to make their purchases.

This last statement sheds light on the next direction of research in MAXCAP problems, trying to include Store-Choice theories in its models. It stands to reason that any retail location model should take into account the processes underlying consumers' choice of store. The paper follows up this new direction of research whose broader aim is to provide a new version of the MAXCAP model, which could be applied to the retail sector. This broader research work has defined three main stages on the way to achieving this objective:

- First, an analysis of how best to include distance in the new version of MAXCAP model.
- Second, analyze which store attributes (other than distance) should be included in the new version of the MAXCAP model and how these could be incorporated.
- Third, a solution employing the new version of the MAXCAP model and its application to a real case.

Since the first stage was analyzed by Colomé and Serra (Colomé and Serra, 1998), this paper tackles the second stage and its application to the supermarket sector.

In essence, the MAXCAP problem seeks the location of a fixed number of stores for firm entering in a spatial market where competitors' shops are already doing business. Since consumers in an area are captured by a given shop if there is no closer shop, the objective of the entering firm is to maximize its market capture. The MAXCAP model uses the traditional view of all or nothing capture relative to the distance criteria. This assumption emerges in the definition of the parameter ρ_{ij}. Specifically, the parameter ρ_{ij} is defined as a binary variable that takes value 1 (i.e. all consumer's zone *i* will shop at shop *j*) if the shop *j* is the closest one to the consumer's zone *i*. The underlying assumption is thus that consumers will automatically shop at the closest store.

As we have said, this assumption does not reflect the real behavior of consumer's choice. Hence the interest in incorporating Store-Choice theories to define the ρ_{ij} parameter in MAXCAP models. In this paper, ρ_{ij} has been defined by using the revealed preference approach of the Store-Choice Behavior theories. Specifically, the Multiplicative[1] Competitive Interaction[2] (MCI) model (Nakanishi and Cooper, 1974) was used. This model determines ρ_{ij} using information revealed by past consumer's behavior in order to understand the dynamics of retail competition and how consumers choose among alternative shopping opportunities.

The paper is organized as follows. Section 2 reviews the literature. Section 3 presents the new methodology, which involves a survey and the application of several analyses to each scenario (presented in Section 4). Section 5 presents the new MAXCAP model for the supermarket sector. The conclusions are set out in Section 6.

2. LITERATURE REVIEW

The choice of a store's location is considered to be the single most important decision a retail organization makes since it is a critical factor in the enterprise's success or failure. Given the importance of this issue, several lines of study have addressed the question of store's location. The relevant ones for this paper are the Competitive Location Literature and the Store-Choice one.

2.1. COMPETITIVE LOCATION LITERATURE

Competitive Location Literature is one line of study within the retail store field which addresses the issue of optimally locating firms that compete for clients in space. Hotelling pioneered this field (Hotelling, 1929) and assumed that consumers would shop at the nearest store. Different models based on this assumption of consumer behavior have been developed

[1] Note that this model becomes additive after the log-transformation is undertaken (see section 3.4.).

[2] The Competitve Interaction condition arises from the fact that in this model individuals select among alternatives probabilistically, in relation to the utilities offered by each choice alternative.

since then. Friesz, et.al. (Friesz, et.al., 1988) pointed out that there are three competitive network facility location models that were "likely to serve as foundations for future models". These ones are the ones of Lederer (Lederer, 1986), Tobin and Friesz (Tobin and Friesz, 1986) and ReVelle (ReVelle, 1986).

The key one for this paper was the one developed by ReVelle (ReVelle, 1986). ReVelle and his followers constructed a group of models that examined competition among retail stores in a discrete spatial market. The basic model was the Maximum Capture Problem (MAXCAP) (ReVelle, 1986). In essence, the MAXCAP problem seeks the location of a fixed number of stores (p stores) for an entering firm in a spatial market where there are other shops from other firms already competing for clients[3]. The spatial market is represented by a network. Each node of the network represents a local market with a fixed demand, which is given. The location of the shops is limited to the nodes of the network. Competition is based on distance: a market is "captured" by a given shop if there is no other shop closer to it. The objective of the entering firm is to maximize its market capture[4].

This model has been adapted to different situations. The first modification introduced shops that are hierarchical in nature and where there is competition at each level of the hierarchy (Serra, et. al., 1992). A second extension took into account the possible reaction from competitors to the entering firm (Serra and ReVelle, 1994). Finally, another modification of the MAXCAP problem introduced scenarios with different demands and / or competitor locations (Serra et.al. 1996). A good review of these models can be found in Serra and

---

[3] Without loss of generality, it is assumed that there is only one competing firm operating in the market (ReVelle, 1986).

[4] This objective, given the assumptions on the characteristics of the retail stores, is almost equivalent to maximising profits (Hansen, et.al., 1987).

ReVelle (Serra and ReVelle, 1996) and a real application of it in Serra and Marianov (Serra and Marianov, 1999).

The p-median formulation[5] of the basic MAXCAP model states that:

(1) $$\text{MAX } Z = \sum_{i \in I} \sum_{j \in J} a_i \; \rho_{ij} \; x_{ij}$$

Subject to

(2) $$\sum_{j \subset J} x_{ij} = 1, \; \forall i \in I$$

(3) $$x_{ij} \le x_{jj}, \; \forall i \in I, \; \forall j \in J$$

(4) $$\sum_{j \subset J} x_{jj} = p$$

$$x_{ij} = \{0,1\} \;\; x_{jj} = \{0,1\} \; \forall i \in I, \; \forall j \in J$$

Where the parameters are:

$i, I =$ Index and set of consumers' zones.

$j, J =$ Index and set of potential locations for shops.

$J^B (\in J) =$ The set of actual locations of the existing shops.

$d_{ij} =$ The network distances between consumers' zone *i* and a shop in *j*.

[5] The p-median like approach is the one that took into account the fact that the demand depends on the distance to the shop (Serra and ReVelle, 1996).

d_{ib_i} = The network distances from node *i* to the closest competitor shop b_i.

a_i = Demand at consumers' zone *i*.

And the variables are defined as follows:

$x_{ij} = 1$, if consumers' zone *i* is assigned to node *j*; 0, otherwise.

x_{jj} = 1, if a shop of firm's A is opened at node *j*; 0, otherwise.

The constraint set basically that: constraint set (2) forces each demand node *i* to assign to only one facility. But for a demand node *i* to be assigned to a facility at *j*, there has to be a facility open at *j*; this is achieved by constraint set (3). Finally, constraint (4) sets the number of outlets to be opened by firm A.

The objective function defines the total capture that firm A can achieve with the siting of its p servers.

In this model, the parameter ρ_{ij} is assumed to be:

$$\rho_{ij} = 1 \text{, if } d_{ij} < d_{ib_i} \text{; } 0 \text{, otherwise.}$$

The application of Store-Choice theories in Competitive Location models is an attempt to define the ρ_{ij} parameter in a way, which is not just based on proximity

2.2. STORE-CHOICE LITERATURE

Store-Choice literature tries to understand the consumer store-choice process. This literature studies the key variables, which a customer takes into account when shopping at a particular shop, and how these variables interact. This literature usually assumes that the consumer not only cares about which shop is the closest but also considers other variables in making his decision to patronize a particular establishment. The development of the consumer store-choice literature has been extensive.

Store-Choice models may be classified into three groups (Craig, et.al., 1984).

The first group includes models that rely on some normative assumption regarding consumer travel behavior. The simplest model is the nearest-centre hypothesis; i.e., consumers patronize the nearest outlet that provides the required good or service. This hypothesis has not found much empirical support, except in areas where shopping opportunities are few and transportation is difficult.

The empirical evidence suggested that consumers trade off the cost of travel with the attractiveness of alternative shopping opportunities. The first one to recognize this was Reilly in its Reilly's "law of retail gravitation" (1929) based on Newton's Law of Gravitation[6] (1686). Reilly's law states that "the probability that a consumer patronizes a shop is proportional to its attractiveness and inversely proportional to a power of distance to it" (Reilly, 1929). Reilly was the precursor of the "gravity" type of spatial choice models. As this early stage, these models were non-calibrated in the sense that the parameters of the models have a priori assigned value. The best representatives of this group are the models of Reilly (Reilly, 1929) and Converse (Converse, 1949).

These non-calibrated gravity models have some limitations (Diez de Castro, 1997):

- They can only be applied to big stores like hypermarkets and shopping centers.
- They can only be applied when the consumer buys non-usual goods.

- They have a restrictive assumption that forces consumer’s zones to be assigned to only one shop.

The second group includes models that use the revealed preference approach to calibrate the “gravity” type of spatial choice models. These ones use information revealed by past behavior to understand the dynamics of retail competition and how consumers choose among alternative shopping opportunities.

Huff (Huff, 1964) was the first one to use the revealed preference approach to study retail store choice. The Huff probability formulation uses distance (or travel time) from consumer’s zones to retail centers and the size of retail centers as inputs to find the probability of consumers shopping at a given retail outlet. He was also the first one to introduce the Luce axiom of discrete choice[7] in the gravity model. Using this axiom, consumers may visit more than one store and the probability of visiting a particular store is equal to the ratio of the utility of that store to the sum of utilities of all stores considered by the consumers.

The main critique to Huff model is its over-simplification since it only considers two variables (distance and size) to describe consumer store-choice behavior.

Nakanishi and Cooper (1974) extended Huff’s model by including a set of store attractiveness attributes (rather than just one attribute employed in Huff’s model). Attributes such as consumer opinion of store image, store appearance, and service level can be used, as

---

[6] Newton’s Law of Gravitation studies the force between planets and stars in the universe. This law states that the force between two bodies is proportional to the product of the masses of the bodies and inversely proportional to the square of the distance between them.

[7] Luce axiom applied to this case assumes that customers choose the optimal location option as a function of the utility of this option with respect to the level of utility of the other options.

well as objectives measures as travel distance and physical distance (Vandell & Carter, 1993). This more general statement was known as the Multiplicative[8] Competitive Interaction[9] (MCI model).

Revealed preference methods overcome the problems of normative methods because consumers are not assigned exclusively to one shop, and the models can be applied to cases where consumers shopping habits are independent of store size. Despite these improvements, these models also have their drawbacks[10] (Craig, et.al., 1984):

- They assume consumer utility function to be compensatory. But in reality consumers reject stores beyond a certain distance. Consumers may also reject stores unless they possess minimum levels of other attributes.

- Context dependence; i.e., the estimated parameters reflect the characteristics of existing stores in the area. For example, the parameters associated with characteristics on which the existing stores do not differ much would be low. This does not, however, imply that such characteristics are unimportant to consumers but rather, that because of their similarity across stores, other variables are used to discriminate among them.

- The distance decay parameter (β) is highly dependent on the characteristics of the spatial structure. The implication is that in assessing the importance of location on store utilities, individuals consider not only the distance to that stores but also the relative distances to

---

[8] Note that this model becomes additive after the log-transformation is undertaken (see section 3.4.).

[9] The Competitve Interaction condition comes from the fact that in this model individuals select among alternatives probabilistically, in relation to the utilities offered by each choice alternative.

[10] Some of these problems can be alternatively seen as a reflection of the "reality". For this reason, we present here the theoretical limitations of these models, but at the concluding section, these limitations are checked to the case analysed in this thesis.

other stores in the area. The result is that consumers residing in different areas might differentially weight the impact of distance on store choice.

Finally, the third group includes the models that use direct utility. These models overcome the problem of context dependence, estimating consumer utility functions from simulated choice data using information integration, conjoint or logit techniques. Instead of observing past choices, these methods use consumer evaluations of hypothetical store descriptions to calibrate the utility function. The best representative model of this group is the one developed by Ghosh and Craig (Ghosh and Craig, 1983) based on game theory.

Given that the aim of the thesis is the incorporation of one store-choice model in the MAXCAP model, one of the previous store-choice models needs to be chosen. The criterion used in making this choice is how well the resulting model can be applied to the real world.

A recent paper (Clarkon, et.al., 1996) analyzed which location models are used by UK grocery retailers. The research shows that the procedure used by major grocery retailers operating within the UK do not rely on one approach but employ a combination of several. These different approaches were used in a sequence to maximize the overall effectiveness. Firms initially use checklist analysis to reduce the cost and time required to assess a large number of potential site locations before using the analogue approach, regression or a gravity model. Finally, the financial analysis decides which location is the most suitable for the new supermarket.

As can be seen, theoretical models are applied to the real world as part of a wider analysis. The Clarkon's study also shows the fact that the most highly-developed models like MCI and Multiple Store Location (Achabal, et.al., 1982) are usually applied in a retailing context by US firms, but not by UK firms. The reason is that grocery retailers operating within the

UK believed that the consumer spatial structure of shopping opportunities in the UK differs to the one found in the US.

The conclusions of the Clarkon paper show that firms prefer the revealed preference approach to model consumer store-choice behavior. This approach is preferred to normative models since it more faithfully reflects real consumer behavior whilst the direct utility approach is simpler since it uses surveys and linear regressions instead of conjoint, logit techniques or game theory.

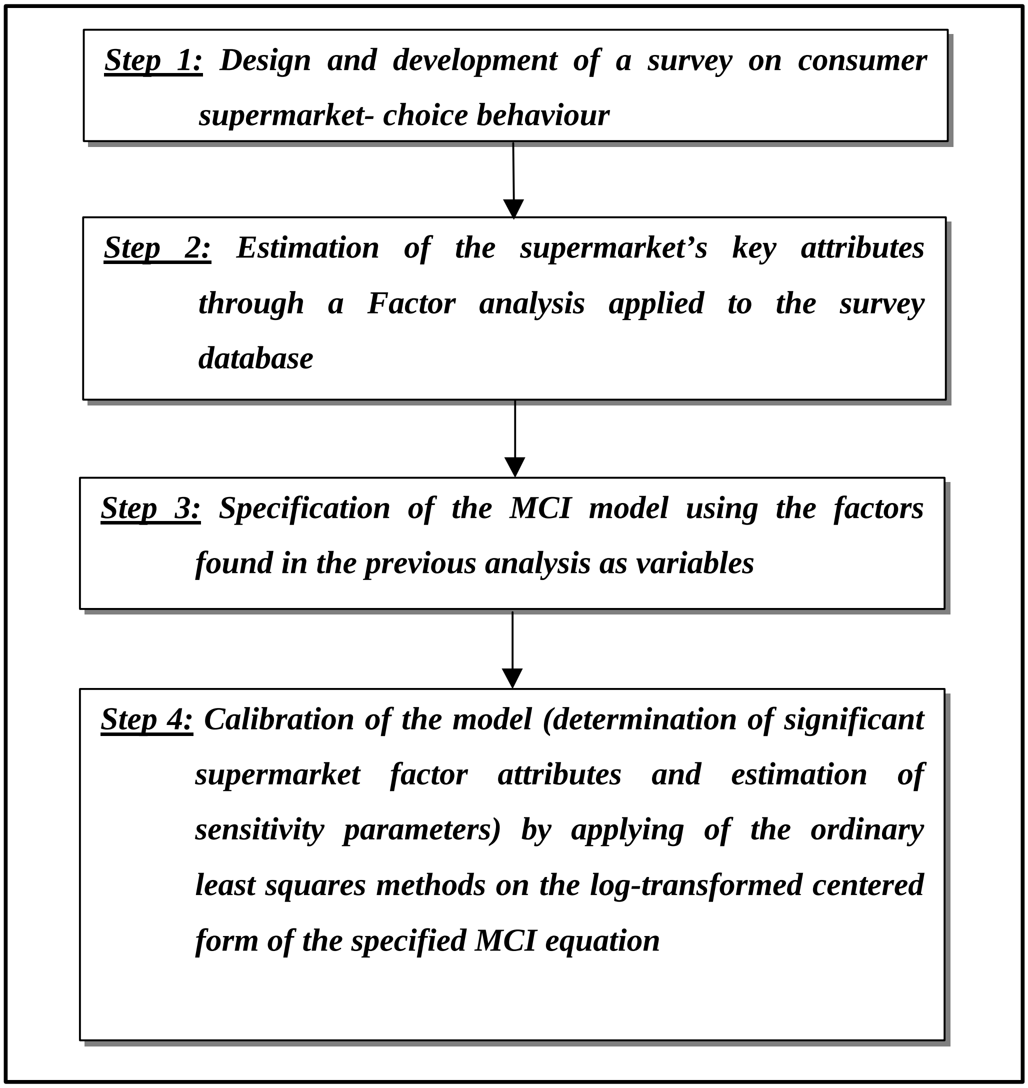
In the revealed preference approach, the most popular model is the MCI model (Craig, et.al., 1984). One of the practical problems of this model is that to date all the calibration had reflected the consumer spatial structure of shopping opportunities of the US market. The problem is overcome in this paper because the surveys were conducted in the UK and Spain. This means that calibration of the MCI in this case reflects British and Spanish Consumer Spatial structure.

3. METHODOLOGY

The main objective of this thesis is the presentation of a new methodology for determining which store attributes (other than distance) should be included in a new version of the MAXCAP model applicable to the retail sector as well as how these parameters ought to be reflected. The parameter r_{ij} included in the MAXCAP model will be determined using the Multiplicative Competitive Interaction model. Specifically, the estimation of parameter r_{ij} will be performed for two scenarios: Milton Keynes (in Great Britain) and Barcelona (in Spain).

The methodology presented and used in this paper is shown in Figure 1.

Figure 1. Methodology



3.1. FIRST STEP

The first step of the paper is the design and development of a survey of consumer supermarket-choice behavior. This is required since MCI is a revealed preference model. In other words, the model uses information revealed by past consumer behavior to calibrate its parameters.

First of all, a questionnaire was design o be used in a personal interview survey. The main structure of the questionnaire included three main parts.

First of all, general questions on shopping behavior were done. Questions 1&2 determine the issue of multi-supermarket shopping; i.e., if consumers went to one or more supermarkets to do their shopping. Question 3 was an open-ended question on the reasons for choosing one supermarket to do the “shopping”. And finally, in question 4, consumers were asked to rank the main supermarket’s attributes. These attributes were extracted from a paper (Burn, 1992) that reviewed the definition of store attributes by different authors.

The second and most important part of the questionnaire includes specific questions on supermarket’s attributes. This general section of specific questions on supermarket’s attributes was structured in blocks representing the main supermarket attributes groups. These blocks were the ones defined by London & Della (London & Della, 1998): *Location, Convenience, Customer Service, Merchandise and Prices.*

Consumers were asked to make scalar judgements in an interval on the importance of various supermarket attributes (except location) when choosing where to do their “shopping”. The specific attributes in each block are the ones defined in Della (London & Della, 1988) and McGoldrick (McGoldrick, 1990). The attributes were measured in accordance with the procedures set out in the Marketing Scales handbook (Gordon, et.al., 1993).

The aim of location was to glean information requires for determining the variables ρ_{ij} and d_{ij}[11] These ones need the determination of the origin and destination of the trip. The destination in this case is clear because it is the supermarket where customers had just done their shopping. But the point of origin is more difficult to ascertain. Most researchers assume that people always travel from home when they go shopping. But nowadays, given demographic changes (e.g., working women), the trip origin may be either home or the workplace[12].

Finally, the third block included some demographic questions.

The survey for this thesis was conducted in Spain and Great Britain. The only differences between both samples were the supermarkets involved. The type of survey, the questionnaire and the sample design were the same. This was so because the aim was to analyze the differences between Spanish and British consumer store-choice behavior.

The target population in Great Britain was British supermarket shoppers. The sampling frame was shoppers at two supermarkets in the Food Centre of the Central Milton Keynes Shopping Centre. The two supermarkets located in this area are *Sainsbury* and *Waitrose*. The target population in Spain was the Spanish supermarket shoppers. The sampling frame was shoppers at two supermarkets in the centre of Barcelona. These are *Bon Preu and Caprabo*.

---

[11] Traditionally, distance has been considered one of the basic reasons for patronising one supermarket. In this paper, distance was computed both in terms of physical distance and travel time distance from home and workplace.

[12] Note that we have assumed that there is only two possible origin for the trip. The reason is that these two are the most important ones and the adding of more options could complicate the analysis.

In principle, the financial constraints of this study determined a sample of 200 consumers in each country. However operational problems in the British survey resulted in a sample of 99 consumers. Thus, the Spanish sample size gives a level of accuracy (confidence level) of ± 7.1 % (for all variables), while the British sample yields a level of accuracy of ± 10 % (for all variables).

The sample procedure selected in this case is a simple random sampling one. Additionally in this case, we split the sample size into different hours and days. The reason was that we wanted to avoid a sample biased toward only one type of supermarket customer (e.g. weekly and weekend shoppers). We therefore decided to conduct 60% of the interviews on Wednesdays (all day) as a guide to weekly shopping habits and 40% on Fridays (afternoon & night) to give a picture of weekend consumers.

After conducting the fieldwork, the Spanish sample followed the previous *a priori* distribution. However operational problems with the British survey prevented this *a priori* distribution being followed. It also proved impossible to follow the a priori daily distribution, although it was possible to split the British distribution by supermarket patronized (59 Sainsbury consumers and 40 Waitrose consumers).

3.2. SECOND STEP

When consumers choose one supermarket to shop, they have to evaluate a large number of attributes. In the questionnaire of this thesis, consumers were asked to evaluate the relative importance of a large number of supermarket's attributes. At this stage, store-choice behavior can be seen as a large multi-attribute problem. But, we need a more parsimonious description of the data to assess a general store-choice behavior. How can we do it?

A theoretical approach for handling multi-attribute judgement problems with a large number of attributes is the Hierarchical Information Integration approach (Louviere, 1984). This approach is based on the assumption that it is a reasonable strategy for consumers to organize individual decision attributes into clusters or sets. Consumers then evaluate and aggregate some property of each of the sets to reach an overall judgement. Moreover, this approach suggests that one could use factor analysis to determine the sets of attributes, and then use these sets as the basis for the hierarchical task.

As the supermarket choice behavior can be seen as a large multiattribute problem (Louviere & Gaeth, 1987), we can use the assumptions of the previous theoretical approach. Using them, the attributes evaluated in the surveys can be categorized into specific factors using Factors analysis.

Moreover, it can be pointed out that a recent research (Hutcheson and Moutinho, 1998) have used factor analysis and regression analysis to estimate the relative importance of each of the factors selecting supermarkets and the way in which they interact to determine the level of customer satisfaction.

3.3. THIRD STEP

After finding the key supermarket factor attributes, the next step is the specification of the MCI model. This specification involves the substitution of the A_{kij} variables of the MCI model, by the factors found in the previous factor analysis and two key variables related to distance[13] (physical distance[14] and travel time distance[15]).

---

[13] The ordinary least square theory states that the omission of relevant variables in a regression analysis could lead to biased estimators (i.e., a biased estimator is one where the estimated value is different from the true one). Then, in this case, the simplest distance variables have been included in the MCI specification to achieve

The MCI version used in this thesis is the original version of Nakanishi and Cooper (Nakanishi and Cooper, 1974) which formulation states that:

(5) $$r_{ij} = \left(\prod_{k=1}^{s} A_{kij}^{\ b_k} \right) \Big/ \sum_{j=1}^{m} \left(\prod_{k=1}^{s} A_{kij}^{\ b_k} \right)$$

Where, at this stage,

r_{ij} = The probability that consumers at location *i* will shop at shop *j*. (i.e., The proportion of capture that a shop in *j* will achieve by consumers' zone *i*)

A_{kij} = The *k*-th attribute describing shop *j* attracting consumers from site *i*; in this case:

- The attributes' factors found by Factor analysis
  - And two distance variables (physical distance and travel time distance from consumers' zone *i* to shop *j*).

i = Index of consumers' zone; $i = 1,..., n.$

j = Index of shops; $j = 1,...,m.$

---

unbiased estimators (although the thesis' aim is the determination of the store's attributes excluding distance variables).

[14] Physical distance is computed as the Manhattan rectilinear distance (because the scenario is a city) from the exact address of the origin to the supermarket in the Spanish survey. Due to some operational problems in the British survey, the physical distance in the British case has been computed with the answer to the second question of the location block: *How far is the store from your home / your workplace?*

[15] Travel time distance has been computed, in both cases, with the answers to the third question of the location block: *How long does it takes to get to the store from your home / your workplace?*

b_k = Parameters still not estimated, which reflect the sensitivity of consumers to the shop characteristics on the probability to shop at a particular shop.

An assumption of the original Nakanishi and Cooper MCI model formulation restricts the estimation of the attribute's effect (b_k) to a single parameter reflecting aggregate market response to all shops alternatives. The use of such market wide parameters allows one to assess how each variable affects patronage but does not permit analysis of these influences for an individual shop (Black, et.al., 1985).

Given this assumption, the Nakanishi and Cooper estimation is not useful in most real cases. The reason is that a firm employing the MCI model usually wants to estimate its individual sensitivity parameters. This is a different case to the one studied in this paper. Here, the variables and the sensitive parameters, which reflect aggregate market response to all shop alternatives, have been estimated. Following the same approach, Jain and Mahajan (Jain and Mahajan, 1979) estimated the original Nakanishi and Cooper MCI model for the food-retailing sector of a large US north-eastern metropolitan area.

3.4. FOURTH STEP

After specification of the MCI model, it only remains to calibrate the model to each specific scenario. The calibration involves two things:

- The identification of the significant attributes in each case (i.e., which attributes are significant to explain the supermarket choice in each scenario).
- The estimation of the sensitivity parameters (b_k) of consumers to the relevant supermarket factor-attributes (i.e., which level of importance is given to each significant attribute).

Nakanishi and Cooper (Nakanishi and Cooper, 1974) showed that the MCI equation could be calibrated by the ordinary least square method on the log-transformed centered form of the equation. They also demonstrated that these estimations could be unbiased and efficient when sampling errors were negligible and specification errors were uncorrelated.

In practical terms, firstly, the original MCI equation[16] (equation (6)) is transformed into its log-transformed-centered form (equation (7)). And then, the ordinary least square method is applied to equation (7) to obtain the parameters' estimators.

(6) $$r_{ij} = \left(\prod_{k=1}^{s} A_{kij}^{\ b_k} \right) V_{ij}^{*} \Big/ \sum_{j=1}^{m} \left(\prod_{k=1}^{s} A_{kij}^{\ b_k} \right) V_{ij}^{*}$$

(7) $$\ln\left(\frac{p_{ij}}{\hat{p}_i} \right) = \sum_{k=1}^{s} \ln\left(\frac{A_{kij}}{\hat{A}_{ki}} \right) + \ln\left(\frac{V_{ij}^{*}}{\hat{V}_i^{*}} \right)$$

Where,

$\hat{p}_i = \left(\prod_{j=1}^{m} \mathrm{p}_{\mathrm{ij}} \right)^{1/m}$ = Geometric mean of the probabilities of consumers at zone *i* shopping at *m*-shops.

$\hat{A}_{ki} = \left(\prod_{j=1}^{m} A_{kij} \right)^{1/m}$ = Geometric mean of *k*-th attributes of *m* shops evaluated by consumers at zone *i*.

[16] Note that a disturbance term has to be included when the parameters of the model were estimated.

$$\hat{V}_i = \left(\prod_{j=1}^{m} V_{ij}^* \right)^{1/m}$$ = Geometric mean of the specification error terms of *m* retail facilities.

Although this estimation seems operationally simple, there is a computational problem for the analyst: if consumers from any zone i ($i = 1...n$) do not shop at a shop j ($j = 1... m$), the resulting p_{ij} and the geometric mean, $\hat{p}_i$, for the consumers' zone will be equal to zero. In such an event, the transformation of the ratio $p_{ij} / \hat{p}_i$ will not be possible for parameter estimation (Jain and Mahajan, 1979). The practical solution is the creation of consumers' zone; each of this consumers' zone has to have consumers patronizing all supermarket alternatives. For example, if the scenario has two supermarkets, each consumers' zone has to have consumers that shop in supermarket 1 and consumers that shop in supermarket 2.

In this study, this computational problem has one practical consequence: each database has individual consumers as cases. This implies that before applying ordinary least squares on the log-transformed centered form of the MCI equation, consumers' zones need to be created. Specifically, the consumers' zones have been created in such a way that consumers of both supermarkets[17] belong to it[18].

4. ANALYSES OF DATA

4.1. PRELIMINARY ANALYSES: GENERAL SHOPPING BEHAVIOR

Before the detailed questions about the supermarket's attributes, several questions related to general shopping behavior were done. In the second general question, consumers were asked

[17] Note that the scenarios analysed in this thesis have two supermarkets.

to rank the main supermarket attributes. In the Spanish survey, there were 8 dimensions to be ranked (from very important 1 to not important 8); while in the British survey, there were 9 dimensions to be ranked (from very important 1 to not important 9)[19]. To summarize these rankings, the mean and standard deviation of each attribute were computed (Table 1).

---

[18] Note that the evaluation of each supermarket attribute for each consumer zone has been computed as the average evaluation of the customers of this specific supermarket in this specific demand node.

[19] The British questionnaire includes the dimension of financial service (defined as the services offered by supermarkets that had a bank). While, the Spanish survey does not include it because Spanish supermarkets do not yet offer this type of service.

Table 1. Ranking of supermarket attributes[20]

IMAGE DIMENSION	*Spanish sample*		*British sample*	
	Mean	*Standard Deviation*	*Mean*	*Standard Deviation*
Convenience	1.62	1.17	2.98	1.98
Quality products	3.68	1.65	2.71	1.59
Range products	4.07	1.47	3.10	1.49
Price products	4.26	2.02	3.46	1.94
Staff	4.28	1.70	5.41	1.41
Hours of opening	4.80	2.24	5.03	1.73
Customer Service	6.15	1.71	5.80	1.80
Customer Account	7.16	1.35	8.12	1.11
Financial Service (only UK)	-	-	8.40	0.85

[20] A c^2 test to the frequencies of these variables shows that all of them are significant.

From the previous Table 1, we can pointed out that, in the Spanish sample, convenience (1.62) was the most important characteristic for customers whilst financial services (7.16) was the least one. The other range of characteristics fell between these extremes in the following order: quality of products (3.68), range of products (4.07), price of products (4.26), staff (4.28), hours of opening (4.30) and customer service (6.15). The standard deviation scores also provided some useful information on the pattern of responses. Relatively low deviation scores were observed for items such as convenience and financial services, whereas higher deviations were observed for items such as price products and hours. This result was to be expected, since the mean perceived importance of items is likely to be dependent, at least to some extent, on how the perceived importance of items differentiates the sample. For example, convenience was importance to most, if not all, respondents and was rated similarly by everyone. In contrast to this, some items appeal more to specific subgroups of the sample and therefore attract different ratings of importance, which increase the standard deviation measure. An example of this is hours of opening, which is not likely to be an important consideration for all respondents.

In the British sample, quality products (2.71) proved the most important characteristic whilst financial service (8.40) was the least one. The other characteristics fell between these two extremes in the following order: convenience (2.98), range of products (3.10), price of products (3.46), hours of opening (5.03), staff (5.41), customer service (5.80) and customer accounts (8.12). In this case, low deviation scores were obtained for items such as financial service and customer accounts; whereas higher deviations were observed for items such as convenience and price of products.

4.2. DETERMINATION OF KEY SUPERMARKET ATTRIBUTES

4.2.1. SPANISH CASE

Factor analysis[21] was applied to the Spanish survey. Eight factors were identified. These factors represented 68 percent of the variance of the 21 variables[22]. This percentage was acceptable given that the criterion of satisfactory percentage of variance explained in social science is 60 % (Hair, et.al., 1998).

The interpretation of the rotated factor matrix was supported by the fact that the minimum significance level for the factor loading in a sample size of 200 is 0.4 (using table 3.2., page 112, Hair, et.al., 1998). In other words, in a sample of size 200, the variables with factor loadings greater than 0.4 are considered significant.

The label and the significant factor loading variables (i.e., the variables with a factor loading greater than 0.4) of each factor are the ones shown in Table 2.

---

[21] In this case, factors were extracted with component analysis and using Varimax rotation.

[22] Note that 5 variables were extracted in the reespecification because their communalities were less than 0.5.

Table 2. Factors for Spanish survey

Variable	Characteristic	Factor loading
Factor 1: Accessibility by modes of transport		
Parksp	It is easy to park at the store	0.862
Publictsp	Easy access by Public Transport	0.715
Dpetrolsp	Petrol discounts	0.846
Dparksp	Parking discounts	0.815
Factor 2: Checkout and shopping assistance service		
Fchecksp	Fast checkout	0.780
Echecksp	Express checkout counters	0.703
Sassistsp	Shopping assitances are courteous and knowledgeable	0.666
Factor 3: Store design and physical facilities		
Crowdsp	No crowded store	0.572
Emovesp	It is easy to move around the store	0.811
Fprodsp	It is easy to find products (readable labels)	0.777
Factor 4: Club card facilities		
Clubcsp	Supermarket Club Card	0.790
Creditsp	The store lets you buy on credit	0.789
Pbrandsp	Store has products of all well known brands and own label ones	0.421
Factor 5: Quality and range of the merchandise		

Prangesp	Store has all basic products and a variety of special items	0.553
Pfreshsp	Store has fresh products	0.713
Pqualsp	Store has high quality products	0.765
Factor 6: Low price policy image		
Offersp	The store does a lot of “promotional offers”	0.892
Advertsp	The store does a lot of advertising of sales	0.869
Factor 7: Wider opening hours		
Omiddaysp	The store is open at noon	0.874
Olatesp	The store is open until late at night	0.864
Factor 8: Location		
Locatedsp	It is well located	0.777

The final step of the Factor analysis was the selection of the surrogate variables[23] of each factor. These surrogate variables were the representatives of the factors found and the ones used in the next regression analysis. In the Spanish case, for example, the first factor of "accessibility by modes of transport" was represented by the variable *parksp*[24] (i.e., "It is easy to park at the store") because was the variable with the higher factor loading. All Spanish surrogate variables were the ones presented in Table 3.

[23] As our objective was the identification of appropriate variables for a subsequent application of the regression technique, a form of data reduction was applied. Given that the aim of this thesis was the practical use of the model (i.e., its replication) to locate supermarkets, the data reduction technique chose in this case was the surrogate variables. Surrogate form of data reduction examines the factor matrix and selects the variables with the highest factor loading on each factor to act as a surrogate variable that is representative of that factor (Hair, et.al., 1998).

[24] Note that it is easy to use a single surrogate variable instead of a linear combination of variables (i.e., Factor Scores).

Table 3. Surrogate variables of the Spanish survey

FACTOR	SURROGATE VARIABLE	DESCRIPTION OF THE FACTOR
Factor 1	*Parksp*	Accessibility by modes of transport
Factor 2	*Fchecksp*	Checkout and shopping assistance service
Factor 3	*Emovesp*	Store design and physical facilities
Factor 4	*Clubcsp*	Club card facilities
Factor 5	*Pqualsp*	Quality and range of merchandise
Factor 6	*Offersp*	Low price policy image
Factor 7	*Omiddaysp*	Wideness of opening hours
Factor 8	*Locatedsp*	Location

The previous surrogate variables that represented the factor-attributes found were the key supermarket's attributes (A_{kij}) that would be included in the Spanish MCI model. As we have explained, additionally, the physical and travel time distance[25] were also introduced in the specification of the MCI model. Using the Spanish surrogate variables found and the distance variables, the specified MCI model in the Spanish scenario is the following one:

(8) $$p_{ij} = \frac{Parksp_{ij}^{b_1} * Fchecksp_{ij}^{b_2} * Emovesp_{ij}^{b_3} * Clubcsp_{ij}^{b_4} * Pqualsp_{ij}^{b_5}}{\sum_{j=1}^{m}(Parksp_{ij}^{b_1} * Fchecksp_{ij}^{b_2} * Emovesp_{ij}^{b_3} * Clubcsp_{ij}^{b_4} * Pqualsp_{ij}^{b_5}}$$

$$\frac{* Offersp_{ij}^{b_6} * Omiddaysp_{ij}^{b_7} * Locatedsp_{ij}^{b_8} * Dhousesp_{ij}^{b_9} * Timehsp_{ij}^{b_{10}}}{* Offersp_{ij}^{b_6} * Omiddaysp_{ij}^{b_7} * Locatedsp_{ij}^{b_8} * Dhousesp_{ij}^{b_9} * Timehsp_{ij}^{b_{10}})}$$

4.2.2. BRITISH CASE

Factor analysis was applied to the British survey. Eight factors were identified. These factors represented 77 percent of the variance of the 19 variables[26]. This percentage was acceptable given the criterion of satisfactory percentage of variance explained in social science is 60 % (Hair, et.al., 1998).

The interpretation of the rotated factor matrix was supported by the fact that the minimum significance level for the factor loading in a sample size of 99 (≈ 100) is 0.55 (using table 3.2., page 112, Hair, et.al., 1998). In other words, in a sample of size near 100, the variables with factor loadings greater than 0.55 are considered significant.

---

[25] Physical distance and travel time distance from consumers' zone *i* to supermarket *j* in the Spanish scenario are represented by *dhousesp* and *timehsp* variables, respectively.

[26] Note that 8 variables were extracted in the reespecification.

The label and the significant factor loading's variables (i.e., the variables with a factor loading greater than 0.55) of each factor are the ones shown in Table 4.

Table 4. Factors for British survey

Variable	Characteristic	Factor loading
Factor 1: Low price policy image		
Lowpuk	Store always has sufficient stock	0.623
Offeruk	Store has fresh products	0.918
Advertuk	Store has high quality products	0.924
Factor 2: Store design and physical facilities		
Crowduk	No crowded store	0.751
Emoveuk	It is easy to move around the store	0.882
Fproduk	It is easy to find products (readable labels)	0.729
Factor 3: Quality and range of merchandise		
Pstockuk	Store always has sufficient stock	0.692
Pfreshuk	Store has fresh products	0.864
Pqualuk	Store has high quality products	0.832
Factor 4: Checkout and shopping assistance service		
Fcheckuk	Fast checkout	0.793
Echeckuk	Express checkout counters	0.841
Sassistuk	Shopping assistance are courteous and knowledgeable	0.661
Factor 5: Facilities for non-car customers		
Parkuk	It is easy to park at the store	-0.733
Publictuk	Easy access by Public transport	0.839

Homeduk	Home delivery	0.704
Factor 6: Wider opening hours		
Osundayuk	The store is open on Sunday	0.862
Olateuk	The store is open until late at night	0.859
Factor 7: Location		
Locateduk	It is well located	0.826
Factor 8: Facilities for car customers		
Dpetroluk	Petrol discounts	0.877

From the previous table, it can be pointed out that the fact that, in this case, two factors were created to represent the importance of modes of transport. Factor 5 represents the non-car customers' facilities, while factor 8 represents car customers' facilities. The polarization of the British society between the car users and non-car users were shown by these two factors; specifically, by factor 5. The reason is that factor 5 included non-car users' variables ("Easy access by public transport" and "home delivery") with positive factors loading and, more important, a car user variable ("it is easy to park at the store") with negative factor loading. In other words, non-car users gave importance to non-car facilities and, at the same time, they did not give any importance to car facilities.

The final step of the Factor analysis was the selection of the surrogate variables of each factor. In the British case, for example, the sixth factor of "wider opening hours" was represented by the variable *osundayuk* (i.e., "the store is open on Sunday") because was the variable with the higher factor loading. All British surrogate variables are presented in Table 5.

Table 5. Surrogate variables of the British survey

FACTOR	SURROGATE VARIABLE	DESCRIPTION OF THE FACTOR
Factor 1	*Advertuk*	Low price policy image
Factor 2	*Emoveuk*	Store design and physical facilities
Factor 3	*Pfreshuk*	Quality and range of merchandise
Factor 4	*Echeckuk*	Checkout and shopping assistance service
Factor 5	*Publictuk*	Facilities for non-car customers
Factor 6	*Osundayuk*	Wideness opening hours
Factor 7	*Locateduk*	Location
Factor 8	*Dpetroluk*	Facilities for car customers

The previous surrogate variables that represent the factor-attributes found were the key supermarket's attributes (A_{kij}) that would be included in the British MCI model. As we have explained, additionally, the physical and travel time distance[27] were also introduced in the specification of the MCI model. Using the surrogate variables found and the distance variables, the specified MCI model in the British scenario is the following one:

$$(9) \qquad p_{ij} = \frac{Advertuk_{ij}^{b_1} * Emoveuk_{ij}^{b_2} * Pfreshuk_{ij}^{b_3} * Echeckuk_{ij}^{b_4} * Publictuk_{ij}^{b_5}}{\sum_{j=1}^{m} (Advertuk_{ij}^{b_1} * Emoveuk_{ij}^{b_2} * Pfreshuk_{ij}^{b_3} * Echeckuk_{ij}^{b_4} * Publictuk_{ij}^{b_5}}$$

$$\frac{* Osundayuk_{ij}^{b_6} * Locateduk_{ij}^{b_7} * Dpetroluk_{ij}^{b_8} * Dhouseuk_{ij}^{b_9} * Timehuk_{ij}^{b_{10}}}{* Osundayuk_{ij}^{b_6} * Locateduk_{ij}^{b_7} * Dpetroluk_{ij}^{b_8} * Dhouseuk_{ij}^{b_9} * Timehuk_{ij}^{b_{10}})}$$

4.3. CALIBRATION OF THE MCI MODEL TO ESTIMATE p_{ij} IN EACH SCENARIO

The calibration of the model identifies, firstly, which of the relevant supermarket's attributes identified by consumers (in the factor analysis) are discriminatory supermarket choice. The calibration, also, estimates the consumers' sensitivity parameters to the significant (i.e., discriminatory) supermarket attributes.

4.3.1. SPANISH CASE

Firstly, the consumers' zone was created from individual consumer responses, using two assumptions:

First, the variable *"timehsp"* coded as interval was transformed to a numeric variable.

---

[27] Physical distance and travel time distance from consumers' *i* to supermarket *j* in the British scenario are represented by *dhouseuk* and *timehuk* variables, respectively.

- Second, consumers that went shopping exclusively from their workplace were excluded from the analysis. Given that only 11.5 % came exclusively from home, 177 consumers forming the initial sample were used to create the consumer zones.

  The reason of this exclusion is the purpose of the MCI model. Its main application is its replication in different zones to predict the market share capture of each supermarket in each zone. The model is estimated with a representative sample, and after this, it is extrapolated to the whole population by means of a census. Usually, this population census reflects the population that lives in these specific zones but not the people working there.

In the Spanish case, 15 zones were created. The next step was the computation of the new A_{kij} and p_{ij} for the consumer zone using the individual A_{ki^*j} and the number of consumers in each zone[28].

The last computational transformation before the ordinary least squares (OLS) estimation was the log-centered transformation of the MCI equation. In this case, this transformation was:

$$(10) \quad \ln\left(\frac{p_{ij}}{\hat{p}_i}\right) = b_1 \ln\left(\frac{\text{locatedsp}_{ij}}{\widehat{\text{locatedsp}}_i}\right) + b_2 \ln\left(\frac{\text{parksp}_{ij}}{\widehat{\text{parksp}}_i}\right) + b_3 \ln\left(\frac{\text{emovesp}_{ij}}{\widehat{\text{emovesp}}_i}\right) + b_4 \ln\left(\frac{\text{omiddaysp}_{ij}}{\widehat{\text{omiddaysp}}_i}\right)$$

$$+ b_5 \ln\left(\frac{\text{clubcsp}_{ij}}{\widehat{\text{clubcsp}}_i}\right) + b_6 \ln\left(\frac{\text{fchecksp}_{ij}}{\widehat{\text{fchecksp}}_i}\right) + b_7 \ln\left(\frac{\text{pqualsp}_{ij}}{\widehat{\text{pqualsp}}_i}\right) + b_8 \ln\left(\frac{\text{offersp}_{ij}}{\widehat{\text{offersp}}_i}\right)$$

$$+ b_9 \ln\left(\frac{\text{dhousesp}_{ij}}{\widehat{\text{dhousesp}}_i}\right) + b_{10} \ln\left(\frac{\text{timehsp}_{ij}}{\widehat{\text{timehsp}}_i}\right) + \ln\left(\frac{V^*_{ij}}{\hat{V}_i}\right)$$

Finally, the ordinary least squares were applied to the log-centered transformation form of the MCI[29]. The regression estimation for the Spanish survey states that:

$$(11) \qquad \ln\left(\frac{p_{ij}}{\hat{p}_i}\right) = -2.989\ln\left(\frac{\text{dhousesp}_{\text{ij}}}{\widehat{\text{dhousesp}}_{\text{i}}}\right) + 0.858\ln\left(\frac{\text{parksp}_{\text{ij}}}{\widehat{\text{parksp}}_{\text{i}}}\right) + 1.645\ln\left(\frac{\text{offersp}_{\text{ij}}}{\widehat{\text{offersp}}_{\text{i}}}\right)$$

The previous equation is the log-centered transformed form of the estimated Spanish MCI model. Using the parameters estimated in equation (11), the original MCI model for the Spanish scenario states that:

$$(12) \qquad p_{ij} = \frac{dhousesp_{ij}^{-2.989} * parksp_{ij}^{0.858} * offersp_{ij}^{1.645}}{\sum_{j=1}^{m}\left[dhousesp_{ij}^{-2.989} * parksp_{ij}^{0.858} * offersp_{ij}^{1.645}\right]}$$

where,

p_{ij} = The probability that a consumer at zone *i* will shop at shop *j*.

$dhousesp_{ij}$ = Physical distance from demand node *i* to the supermarket *j*.

$parksp_{ij}$ = Valuation by zone *i*'s consumers to "the accessibility by modes of transport" to the supermarket *j* (on a 5-point scale).

$offersp_{ij}$ = Valuation by zone *i*'s consumers to "the low price policy image" of supermarket *j* (on a 7-point scale).

---

[28] Note that the A_{ki*j} used are the eight ones identified by the Factor analysis plus the physical and travel time distance.

[29] The OLS procedure was applied using stepwise estimation. After the estimation, the statistical significance determines a R-square of 0.881 and an adjusted R-square of 0.868. Moreover, the t-tests of all three variables, except the constant, prove that all coefficients were significantly different from zero for a significant of 95%. Finally, an analysis of the residuals confirmed that the previous estimations were correct.

Summing up, the calibration of the Spanish MCI model have identified:

- *The discriminatory attributes to the Spanish scenario*

Equation (16) shows that the probability of patronizing the two Spanish supermarkets depends on three variables: "the physical distance from consumer's zone to the supermarket" (i.e., variable *dhousesp*), "the accessibility by modes of transport to the supermarkets" (i.e., variable *parksp*) and "the low price policy image" (i.e., variable *offersp*). In other words, the choice between both Spanish supermarkets depends only on these three attributes, because both supermarkets were very similar in the other relevant attributes.

- *The consumers' sensitivity parameters to the discriminatory supermarket attributes*

In this case, the estimated parameters were –2.989 for the variables *dhousesp*, 0.858 for the variable *parksp* and 1.645 for the variable *offersp*. A positive sign of the sensitivity parameters indicates that a supermarket with higher levels of that attribute would have a higher probability of being patronized; while, a negative sign indicated that a supermarket with a higher level of that attribute would have a lower probability of being patronized. In this case, the supermarket with higher valuations of "accessibility by modes of transport" or "low price policy image" would achieve a higher capture of consumers (i.e., a higher probability (p_{ij})); while the further supermarket from consumers' zone would have a lower probability of being patronized.

Moreover, the absolute values of these sensitive parameters indicate the relative level of importance give to each of the attributes; a higher value of the sensitive parameter indicates that a little change of that attribute in one supermarket would have a higher impact on the probability of being patronized. In this case, the Spanish consumers were more sensitive to physical distance; than to “low price policy image” and “accessibility by modes of transport”, respectively.

4.3.2. BRITISH CASE

Firstly, the consumers’ zone was created from individual consumer responses, using two assumptions:

- First, the variables *timehuk* and *dhouseuk* coded as interval were transformed to numeric variables.

- Second, consumers that went shopping exclusively from their workplace were excluded from the analysis[30].

  Given the operational problems in the British survey, not all the interviewees did their usual “shopping” in the supermarket patronized in the survey. As the thesis’ aim was the analysis of the consumers’ supermarket choice in its usual “shopping”, we needed to exclude the cases that did not comply with this condition[31]. Finally, a sample of 62 consumers was determined after the exclusion of the cases that did not comply with any of the previous conditions.

---

[30] The justification to do this can be found in section 4.3.1.

[31] Note that, in the Spanish survey, all the interviewees were usual customers of the Spanish supermarkets. The reason was that, in this case, the interviewer confirmed that the interviewee did their usual shopping in that supermarket before begin the interview.

In the British case, 6 zones were created[32]. The next step was the computation of the new A_{kij} and p_{ij} for the consumers zone using the individual A_{ki*j} and the number of consumers in each zone[33].

The last computational transformation before the ordinary least squares (OLS) estimation was the log-centered transformation of the MCI equation. In this case, this transformation was:

$$(13) \qquad \ln\left(\frac{p_{ij}}{\hat{p}_i}\right) = b_1 \ln\left(\frac{\text{Advertuk}_{\text{ij}}}{\widehat{\text{Advertuk}}_{\text{i}}}\right) + b_2 \ln\left(\frac{\text{emoveuk}_{\text{ij}}}{\widehat{\text{emoveuk}}_{\text{i}}}\right) + b_3 \ln\left(\frac{\text{Pfreshuk}_{\text{ij}}}{\widehat{\text{Pfreshuk}}_{\text{i}}}\right) + b_4 \ln\left(\frac{\text{echeckuk}_{\text{ij}}}{\widehat{\text{echeckuk}}_{\text{i}}}\right)$$

$$+ b_5 \ln\left(\frac{\text{Publictuk}_{\text{ij}}}{\widehat{\text{Publictuk}}_{\text{i}}}\right) + b_6 \ln\left(\frac{\text{Osundayuk}_{\text{ij}}}{\widehat{\text{Osundayuk}}_{\text{i}}}\right) + b_7 \ln\left(\frac{\text{Locateduk}_{\text{ij}}}{\widehat{\text{Locateduk}}_{\text{i}}}\right) + b_8 \ln\left(\frac{\text{dpetroluk}_{\text{ij}}}{\widehat{\text{dpetroluk}}_{\text{i}}}\right)$$

$$+ b_9 \ln\left(\frac{\text{dhouseuk}_{\text{ij}}}{\widehat{\text{dhouseuk}}_{\text{i}}}\right) + b_{10} \ln\left(\frac{\text{timehuk}_{\text{ij}}}{\widehat{\text{timehuk}}_{\text{i}}}\right) + \ln\left(\frac{V_{ij}^*}{\hat{V}_i}\right)$$

As can be expected, all values of the variable "Dhouseuk" were zeros because the consumers' zones were created using the codes (i.e., the intervals) described by this variable. Then, this variable was excluded from equation (13) to be able to apply ordinary least squares efficiently (i.e., to find unbiased and efficient estimators).

---

[32] The operational problems of the British sample did not allow knowing the exact address of the interviewees. Then, the zones were created using the zones described by the variable *dhouseuk* (i.e., "physical distance from consumer home to the supermarket"). The six zones created correspond to the six intervals defined in that variable (i.e., zone 1 includes consumers living within a radius of less than 2 kilometres round the two side-by-side supermarkets of the Food Centre).

[33] Note that the A_{ki*j} used are the eight ones identified by the Factor analysis plus the physical and travel time distance.

Finally, the ordinary least squares were applied to the log-centered transformation form of the MCI[34]. The regression estimation for the Spanish survey states that:

(14) $$\ln\left(\frac{p_{ij}}{\hat{p}_i}\right) = 2.163\ln\left(\frac{\text{Advertuk}_{ij}}{\hat{\text{Advertuk}}_{i}}\right) + 1.650\ln\left(\frac{\text{Pfreshuk}_{ij}}{\hat{\text{Pfreshuk}}_{i}}\right)$$

The previous equation is the log-centered transformed form of the estimated British MCI model. Using the parameters estimated in equation (14), the original MCI model for the British scenario states that:

(15) $$p_{ij} = \frac{advertuk_{ij}^{2.163} * pfreshuk_{ij}^{1.650}}{\sum_{j=1}^{m}\left[advertuk_{ij}^{2.163} * pfreshuk_{ij}^{1.650}\right]}$$

where,

p_{ij} = The probability that a consumer at zone *i* will shop at shop *j*.

$Advertuk_{ij}$ = Valuation by zone *i*'s consumers to "low price policy image" of supermarket *j* (on a 7-point scale).

$Pfreshuk_{ij}$ = Valuation by zone *i*'s consumers to the "quality and range of merchandise" of supermarket *j* (on a 5-point scale).

Summing up, the calibration of the British MCI model have identified:

- *The discriminatory attributes to the British scenario*

---

[34] The OLS procedure was applied using stepwise estimation. After the estimation, the statistical significance determines a R-square of 0.864 and an adjusted R-square of 0.691. Moreover, the t-tests of both variables, except the constant, prove that all coefficients were significantly different from zero for a significant of 95%. Finally, an analysis of the residuals confirmed that the previous estimations were correct.

Equation (15) shows that the probability of patronizing the two British supermarkets depends on two variables: "the low price policy image" (i.e., variable *advertuk*) and "quality and range of merchandise" (i.e., variable *pfreshuk*). In other words, the choice between both British supermarkets depends only on these two attributes, because both supermarkets were very similar in the other relevant attributes. For example, in this case, distance (i.e., travel time distance) was not significant to explain the supermarket choice because these two supermarkets are located side by side in the Food Centre of the Central Milton Keynes Shopping Centre.

- *The consumers' sensitivity parameters to the discriminatory supermarket attributes*

In this case, these parameters were 2.163 for the variable *advertuk* and 1.650 for the variable *pfreshuk.* Here, the supermarket with higher valuation of "the low price policy image" or "quality and range of merchandise" would achieve a higher capture of consumers (i.e., a higher probability (p_{ij})). Moreover, the absolute values of the sensitive parameters indicate that the British consumers were more sensitive to "low price policy image" than to "Quality and range of merchandise".

5. A MAXIMUM CAPTURE MODEL FOR THE SUPERMARKET SECTOR

The result of this paper is the presentation of a new version of the maximum capture model for the supermarket sector which takes account of revealed consumer store-choice behavior.

The maximum capture model (MAXCAP) presented in this case selects the location of supermarkets for a food retailing company entering a market in which it wishes to maximizes it share of a market where competing supermarkets are already operating. The formulation of this Maximum Capture Model states[35]:

(16) $$\text{MAX } Z = \sum_{i\in I}\sum_{j\in J} a_i \ \rho_{ij} \ x_{ij}$$

Subject to

(17) $$\sum_{j\subset J} x_{ij} = p+q \ , \ \forall i \in I$$

(18) $$x_{ij} \leq x_{jj} \ , \ \forall i \in I \ , \ \forall j \in J$$

(19) $$\sum_{j\subset J} x_{jj} = p$$

$$x_{ij} = \{0,1\} \ x_{jj} = \{0,1\} \ \forall i \in I \ , \forall j \in J$$

This formulation is similar to the one in the P-median problem (the one presented in epigraph 2.1.), except in two things:

- We have reformulate constraint (2): $\sum_{j\subset J} x_{ij} = 1 \quad \forall i \in I$, the one that forces each consumers' zone *i* to assign to only one shop. Instead, we use constraint set (17) which states that every consumer zone makes $p + q$ assignments to the p new and q existing supermarket shops.

- The parameter ρ_{ij} is defined using the results found in the previous calibration of the Multiplicative Competitive Interaction model. Using this consumer store-choice model to define ρ_{ij}, the new version of MAXCAP model takes into account how consumers choose among alternative shopping opportunities.

---

[35] The notation is the same to the one used in section 2.1.

The calibration of the parameters of the ρ_{ij} was performed separately for each country's database (as explained in Section 3.4.). Next, we present the two ρ_{ij} (Spanish and British) values for use in the new MAXCAP model. The use of each will depend on the country where the model is applied.

The Spanish $\boldsymbol{r_{ij}}$ resulted from the previous analysis states that:

(20)
$$p_{ij} = \frac{dhousesp_{ij}^{-2.989} * parksp_{ij}^{0.858} * offersp_{ij}^{1.645}}{\sum_{j=1}^{m}\left[dhousesp_{ij}^{-2.989} * parksp_{ij}^{0.858} * offersp_{ij}^{1.645}\right]}$$

where,

p_{ij} = The probability that a consumer at zone *i* will shop at shop *j*.

$dhousesp_{ij}$ = Physical distance form demand node *i* to the supermarket *j*.

$parksp_{ij}$ = Valuation by zone *i*'s consumers to the accessibility by modes of transport to the supermarket *j* (on a 5-point scale).

$offersp_{ij}$ = Valuation by zone *i*'s consumers to the low price policy image of supermarket *j* (on a 7-point scale).

The British $\boldsymbol{r_{ij}}$ resulted from the previous analysis states that:

(21)
$$p_{ij} = \frac{advertuk_{ij}^{2.163} * pfreshuk_{ij}^{1.650}}{\sum_{j=1}^{m}\left[advertuk_{ij}^{2.163} * pfreshuk_{ij}^{1.650}\right]}$$

where,

p_{ij} = The probability that a consumer at zone *i* will shop at shop *j*.

Advertuk$_{ij}$ = Valuation by zone *i*'s consumers to "low price policy image" of supermarket *j* (on a 7-point scale).

Pfreshuk$_{ij}$ = Valuation by zone *i*'s consumers to the "quality and range of merchandise" of supermarket *j* (on a 5-point scale).

6. CONCLUSIONS

6.1. A SIMPLE APPLICATION OF THE NEW METHODOLOGY

The aim of this paper is the presentation of a new methodology for determining which supermarket attributes should be included in the new version of the MAXCAP model applicable to the retail sector as well as how these parameters ought to be reflected. The methodology presented in this thesis is the second stage of a broader methodology that derives the optimal location of new supermarkets in a market where other supermarkets were already operating.

In this section, we are going to present briefly the flow of a simple application of this broader methodology and how the analyses presented in this paper were included.

We can consider a scenario represented by a network. This network represents a small town in Great Britain and each node represent a consumer zone (i.e., neighborhood zones). In this little town, there are several supermarkets located. A new supermarket chain wants to locate a store in that little town. The entering Supermarket Company decides to apply the MAXCAP methodology to find the optimal location of the new store. To do this, the company applied the following stages:

First Stage: Development of a survey of supermarket - choice behavior

The first stage would be the development of a survey in this British town. In this survey, consumers would be asked to make judgements on the importance of various supermarket's attributes when choosing where to do their shopping.

1. To simplify the analysis, we could use the list of general attributes found in this paper for the British case. In this case, the attributes evaluated in the survey do not need to be categorized into factors using Factor analysis because general attributes have been used from the beginning.

Second Stage: Calibration of the MCI model in this British scenario to determine the parameter p_{ij}

The calibration of the MCI involves:

- The computation of the new A_{kij} and p_{ij} for the consumers' zones using the individual A_{kij} and the number of consumers in each consumers' zones.
- The transformation of the MCI equation in its log-centered transformed form
- Finally, the application of the ordinary least square method to the previous log-centered transformed form of the MCI equation.

The calibration of the MCI model would give the estimated MCI for this small town scenario. Specifically, the calibration would identify which attributes are discriminatory in the choice between the supermarkets in that British town. Moreover, the calibration would also estimate the level of importance (i.e., sensitivity parameters) given by consumers to each of the previous discriminatory attributes.

Third Stage: Resolution of the MAXCAP model

Using the p_{ij} found in the previous stage, we can solve the new MAXCAP model[36].

The resolution of the MAXCAP model would give the optimal location for the new supermarket. Moreover, we could assume different levels of the significant key attributes for the new supermarket and find, in each case, the optimal location.

6.2. LIMITATIONS OF THE ANALYSIS

The main limitations of the analysis are the ones identified for revealed preference methods (Craig, et.al., 1984) and specifically for the MCI model used in this thesis. We shall now discuss these theoretical problems and their applicability to this case.

- This model assumes consumer utility function to be compensatory. But really consumers reject stores beyond a certain distance and may also reject stores unless they meet threshold levels of other attributes.

  This problem does not apply here because the supermarket alternatives in both scenarios have a minimum level of all key attributes. Additionally, the supermarkets in both cases were closely enough to be alternative choices for all the consumers in the sample.

- The model is context dependent; i.e. the estimated parameters associated with characteristics on which the existing stores do not differ much would be low. This does not, however, imply that such characteristics are unimportant to consumers but rather that other variables are used to discriminate among them. This limitation applies to both scenarios. In the Spanish case, the ranking of supermarket attributes (identified in Section 4.1.1.) was convenience (location and access by transport mode), quality products, range of products and price products. Despite this ranking, the key

---

[36] Note that this stage is not explained in this paper, but it is the next step of the broader research.

discriminating variables between both Spanish supermarkets were convenience (distance and accessibility by transport mode) and price policy. This means that both supermarkets are very similar in terms of product quality and range. In a similar way, distance was not significant to explain the British supermarket choice because the two British supermarkets were located side by side in the Food Centre of the Central Milton Keynes Shopping Centre.

- The distance decay parameter (β_d) is highly dependent on the characteristics of the spatial structure.

  This limitation is also applicable to this study. Although both surveys were designed to be as similar as possible, it was not possible to overcome the issue of different spatial structure in both countries.

  - The Spanish scenario is the centre of Barcelona. Barcelona is a traditional Mediterranean city where supermarkets and grocery shops are located throughput the city.

  - The British scenario is the centre of Milton Keynes. Milton Keynes is a big residential area. Basically, its roundabouts and American style road network were designed to ensure that any part of the city would be within 15 minutes drive time. In terms of supermarkets, the city has a big shopping centre in the middle of the city (called the Central Milton Keynes Shopping centre) and several small malls on the city outskirts. The two supermarkets used in the British survey are located side by side in the Central Shopping Centre.

Finally, there is a statistical limitation of the analysis. This is due to:

- The sample size: the Spanish survey had a sample size of 200 questionnaires, which gave a level of accuracy (confidence level) of ± 7.1%. The British survey had a sample size of 99 questionnaires, which gave a level of accuracy (confidence level) of ± 10%.

- The Spanish sample was distributed *a priori* as a function of the day of the week and the supermarket involved. The distribution chosen tries to avoid bias in choosing only one type of supermarket shopper (i.e., weekly or weekend one's). The British survey posed a problem in this respect. Operational difficulties meant the British survey could not be split as the basis of this *a priori* distribution. Likewise, we were able to establish the daily distribution of the sample afterwards. The British sample may therefore be biased toward one type of supermarket consumer.

6.3. NEW DIRECTION

The new direction of this research is obviously the analysis of the third stage of the broader research work which aim is the presentation of a new modified version of the MAXCAP model applicable to the retail sector. This last stage involves resolving the new modified version of the MAXCAP model and its application to a real case.

Firstly, model resolution will involve new metaheuristics[37] techniques such as: Heuristic Concentration (Rosing and ReVelle, 1996, 1997) or Greedy Randomised Adaptive Search (GRASP) (Feo and Resende, 1989). The resolution of the model involves a computational procedure in simulated networks to check the optimality of the resolution method.

---

[37] A metaheuristic us a process which applies a subordinate heuristic at each step which has to be designed for each particular problem. Although there is no guarantee of optimality of these methodologies; Metaheuristics have proved highly successful in obtaining high quality solutions to many real world complex problems.

Once an optimal resolution has been identified, the model can be applied to real cases. A real case where a new food retailing company wishes to enter a market with a fixed number of shops to maximize its market share given that another food retailing company is already operating with a determined number of shops. The computation of the new optimal locations for the entering firm will be the result of the resolution of this new MAXCAP model that takes into account the consumers' behavior in choosing supermarkets.

REFERENCES

Achabal, D., Gurr, W.L. and Mahajan, V. **MULTILOC: A Multiple Store Location Decision Model.** *Journal of Retailing*, v58, Summer 1982, p5-25.

Agustina, L. **¿Alcanzará Caprabo a Pryca y Continente?** *Actualidad Económica*, 17th to 23th of May, 1999.

Black, W.C., Osthind, L.E. and Westbrook, R.A. **Spatial Demand Models in an Intrabrand Context.** *Journal of Marketing*, v49, Summer 1985, p106-113.

Burn, D. J. **Image Transference and Retail Site Selection.** *International Journal of Retail & Distribution Management*, Sept-Oct, v20 n5, 1992, p38-46.

Clarkon, R.M., Clarke-Hill, C.M. and Robinson, J. **UK Supermarket Location Assessment.** *International Journal of Retail & Distribution Management*, v24 n6, 1996, p22-33.

Colomé, R and Serra, D. **Consumer Choice in Competitive Location Models: Formulations and Heuristics.** *Economics Working Paper 290, May 1998, Pompeu Fabra University, Barcelona, Spain.*

Converse, P.D. **New Law of Retail Gravitation.** *Journal of Marketing*, v14, 1949, p379-384.

Craig, C.S., Ghosh, A. and McLafferty, S. **Models of the Retail Location Process: A Review.** *Journal of Retailing,* v60 n1, Spring 1984, p5-36.

Davies, J.B. and Flemmer, M. **Consumer Behavior in the European Union.** *In the book: The internationalisation of Retailing edited by Gary Akehmist and Nicholas Alexander,* Frank Cass London, 1995, p177-190.

Diez de Castro, E. *Distribución Comercial.* McGRaw Hill, second edition, 1997.

Feo, T and Resende, M (1989). **A probabilistic heuristic for a computational difficult set covering problem.** *Operations Research Letters*, n8, 1989, p67-71.

Friez, T., Miller, T. and Tobin, R. **Competitive Network Facility location Models: A Survey.** *Papers of the Regional Science Association*, v65, 1988, p47-57.

Ghosh, A. and Craig, C.S. **Formulating Retail Location Strategy in a Changing Environment.** *Journal of Marketing*, v47, 1983, p56-68.

Gordon, C., Bruner, I and Hensel, P.J. *Marketing Scales Handbook. A Compilation of Multi-Item Measures.* American Marketing Association, Chicago, Illinois, USA, 1993.

Hair, J., Anderson, R., Tatham, R. and Black, W. *Multivariate Data Analysis.* FifthEdition, Prentice Hall International, Inc., 1998.

Hansen, P., Labbé, M., Peeters, D. and Thisse, J.F. **Facility Location Analysis.** *Fundamentals of Pure and Applied Economics*, v22, 1987, p1-70.

Hogan, B.A. **What drivers shoppers' decisions supermarket News.** *Supermarket News,* May 6, 1996, v46, n19, p50-52.

Hotelling, H. **Stability in Competition.** *Economic Journal*, v39, 1929, p41-57.

Huff, D. **Defining and Estimating a Trading Area.** *Journal of Marketing,* v28, 1964, p34-38.

Hunt, J. **The Big Squeeze (Food Retailing in the United Kingdom).** *Grocer*, v219 n7308, 1997, p38-44.

Hutcheson, G.D. and Moutinho, L. **Measuting Preferred Store Satisfaction Using Consumer Choice Criteria as a Mediating Factor.** *Journal of Marketing Management*, v14, 1998, p705-720.

Jain, A. K. and Mahajan, V. **Evaluating the Competitive Environment in Retailing using Multiplicative Competitive Interactive Model.** *Research in Marketing*, v2, 1979, p217-235.

Kinnear, T.C. and Taylor, J.R. *Marketing Research: An applied Approach.* McGraw-Hill Series in Marketing, 1987.

Lederer, P.J. **Duopoly Competition in Networks.** *Annals of Operations Research*, v6, 1986, p99-109.

Lehmann, D. R., Gupta, S. and Steckel, J.H. *Marketing Research.* Addison-Wesley, 1998.

London, D and Della, A.J. **Chapter 20: Purchasing Processes.** *Consumer Behavior: Concepts and Applications.* McGraw-Hill Series in Marketing, 1988.

Louviere, J.J. **Hierarchical Information Integration: A New Method for the Design and Analysis of Complex Multiattribute Judgement Problems.** *Advances in Consumer Research,* Provo Utah: Association for Consumer Research, 1984, p148-155.

Louviere, J. J. and Gaeth, G. J. **Decomposing the Determinants of Retail Facility Choice Using the Method of Hierarchical Information Integration: A Supermarket Illustration.** *Journal of Retailing,* v63 n1, spring 1987, p25-48

McGoldrick, P.J. *Retail Marketing.* McGraw-Hill Book Company, 1990.

Nakanishi, M and Cooper, L.G. **Parameter Estimation for a Multiplicative Competitive Interaction Model – Least Squares Approach.** *Journal of Marketing Research,* vol. XI, August 1974, -303-311.

Orgel, D. **Consuming Issues: Responses to a SN Consumer Survey show Preferences about Supermarket Shopping and Reveal Merchandising Opportunities.** *Supermarket News,* August 25, v47 n34, 1997, p1-4.

Pau, J. and Navasmes, R. *Manual de Logística Integral*. Diaz de Santos, 1998.

Reilly, W.J. *The Law of Retail Gravitation.* New York, Knickerbocker Press, 1929.

ReVelle, C. **The Maximum Capture or "Sphere of Influence" Location problem: Hotelling revisited on a Network.** *Journal of Regional Science,* v26, n2, 1986, p343-358.

Rosing, K.E. and ReVelle. **Heuristic Concentration: A Metaheuristic for Combinatorial Location Problems.** *Studies in Locational Analysis.* n9, 1996, p109-112.

Rosing, K.E. and ReVelle. **Heuristic Concentration: Two Stage Solution Construction.** *European Journal of Operations Research*, 1997, p75-86.

Serra, D., Marianov, V. and ReVelle, C. **The Hierarchical Maximum Capture Problem.** *European Journal of Operational Research*, v62 n3, 1992.

Serra, D. and ReVelle, C. **Market Capture by two Competitors: The Preemptive Location Problem.** *Journal of Regional Science*, v34 n4, 1994, p549-561.

Serra, D, Ratick, S. and ReVelle, C. **The Maximum Capture problem with Uncertainty.** *Environment and Planning,* v62, 1996, p49-59.

Serra, D. and ReVelle, C. **Competitive Location in Networks.** *Facility Location: A Survey of Application and Methods.* Z. Drezner (ed), Springer, 1996.

Serra, D and Marianov, V. **The P-Median Problem in a Changing Network: The Case of Barcelona.** *Location Science,* v4(1), 1999.

The European editorial, **Hyper Growth in Food Sales (Small Retail Traders lose out to Supermarkets in Europe),** *The European,* n411, April 6, 1998, p61.

Tobin, R and Friesz, T. **Spatial Competition Facility Location Models: Definition, Formulations and Solution Approach.** *Annals of Operations Research,* v6, 1986, p49-74.

Vandell, K.D. and Carter, C.C. **Retail Store Location and Market Analysis: A Review of the Research.** *Journal of Real Estate Literature,* n1, 1993, p13-45.