Competitive Location Assessment – the MCI Approach

Boris Tihi, Nermin Oruc *

Abstract:

In this paper, we have investigated the list of determinants that influence customers to create a short-term preference to a particular store. A survey among 240 households in Sarajevo was conducted in order to reveal their preferences. The results of the MCI model have identified several main stores' characteristics which have a key influence on customers' patronage decisions. The findings also suggest that customers, in making their patronage decisions, and hence making their preferences to a particular store, choose among a limited list of alternatives, which implies that competition among outlets is more space limited than had been expected.

Keywords: Store choice behaviour, retail gravity model, revealed preference

JEL: C31, D12, L81

DOI: 10.2478/v10033-012-0013-7

1. Introduction

The most important decision, which to a great extent determines the success of any retail outlet, is the choice of its location. In order to achieve the best possible capture of customers from the location we choose, our choice should be led by our knowledge about the store choice behaviour of customers in the trade area of our new outlet.

Since Hotelling (1929) and Christaller (1933), economists have tried to derive the optimal number, size and placing of retail outlets by assuming that utility maximising individuals will shop at the closest store and tend to minimise the costs of shopping. Yet every day we observe that customers travel to very distant sites to shop just because of its particular features that differentiate it from other stores, so we may assume that an individual's utility is affected by the characteristics of the store where they shop.

The main aim of this analysis is to identify the main determinants of store choice behaviour that cause these observed deviations from the predictions of simple models, with empirical evidence obtained by research about the store choice behaviour of hypermarket customers in Sarajevo, Bosnia. The survey amongst 240 households is conducted in Sarajevo, one of the fastest growing markets in transition and one of the least concentrated retail markets, which raises its attractiveness for the entrance of new international chains.

The model employed for this purpose is the Retail Gravity Model, augmented by the inclusion of the MCI coefficient. The model incorporates both the attraction influences of stores' attributes and the deterring influence of the distance between store and customer on the customer's store choice behaviour, and the interaction of these two influences to estimate the probability of patronisation of particular stores by customers from particular areas.

Since this analysis of the determinants of hypermarket store choice was conducted for the first time in Sarajevo,

*** Nermin Oruc**
Assistant Professor at IUS, Sarajevo and Affiliate Fellow at CERGE-EI, Prague
E-mail: noruc@ius.edu.ba

Boris Tihi
University of Sarajevo, School of Economic and Business;
E-mail: boris.tihi@efsa.unsa.ba

we assumed that the results of this analysis could be very interesting for decision-makers in those companies.

The remainder of this paper is divided into four sections. In the next section, the evolution of both location theories and store choice models is presented, with particular emphasis on evaluation of modern theoretical developments and empirical work. In section three, we present the methodology, description of data collected through a survey and the main empirical findings of this work. Finally, section four gives conclusions and managerial implications based on the empirical evidence presented in section four, along with some limitations of the study and directions for further research.

2. Literature Review

2.1. Location theory

Competitive location literature is concerned with analysis of the optimal location of stores within either discrete or continuous space.

The first of these theories, developed by Haig (1927; cited in Brown, 2001), is bid rent theory. Assuming easy travelling in any direction, a free property market and perfectly informed and profit-maximising buyers and sellers, this theory predicts that different activities on the demand side of the land property market are aware of advantages which central location offers, and therefore bid for it by means of the rent they are willing to bear. Such competition for inelastic land supply will, in the long run, result in a situation where all central urban sites will be occupied by the most competitive activities and thus put land to "its highest and best use"[1]. This theory does not have much applicative value and has many critics, because of its strong assumptions and ahistoric nature.

The second well-known theory is named after its author, Howard Hotelling. In short, Hotelling's theory assumes uniform distribution of buying power along a segment and "patronising closest store" consumer behaviour. On the basis of these assumptions, he derives the optimal location for two facilities (ice cream vendors along a beach strip). When there is no competition, all customers patronise one facility. But after introduction of another facility, customers patronise the closer one, which introduces a breaking point and splits the segment into two sub-segments, where the segment left of the breaking point patronises the facility placed to the left, and vice versa. In order to capture as much of the rival's customers as possible, the best move for the left facility is to change its location and move to the left beside the facility placed on the right side. Subsequent actions and reactions by rivals will continue up to a Nash equilibrium situation, where all competing facilities will be located at the centre of the area, clustered around the breaking point, and none will any have incentive to change its location.

The third theory, and the one which is of the most interest to us, is spatial interaction theory (Reilly, 1929; cited in Drezner, 1994b). Reilly's "Law of Retail Gravitation", based on the Newtonian law of planetary attraction, takes into consideration not only distance but also the attractiveness of a particular site. This theory also predicts that there will be a break point where the customer will be indifferent between two stores. Initially developed for the determination of flows of customers between two cities, this law was soon applied to different issues, such as international trade, immigration flows, tourism, etc.

Introduction of the store's attractiveness, besides distance, as a determinant of consumer choice, where these two determinants influence such decisions in opposite directions, may be considered the main achievement of this theory.

This theory is further developed by Huff (1964; cited in Drezner, 1994b) and Nakanishi and Cooper (1974), who introduce a probabilistic approach to this rather deterministic theory, offering techniques for empirical analysis of a customer's store choice behaviour in their application to operational issues that retailers face, such as determination of capture area, choice of the best location for new outlets, analysis of the efficiency of current marketing program, and so forth.

Finally, another location theory is central place theory (Christaller, 1933; cited in Brown, 2001). Central place theory is based on strong assumptions such as a uniform distribution of identical, perfectly informed customers served by sellers who sell at f.o.b. prices, subject to free entry and equivalent costs and behaving in a competitive, profit-maximizing manner. Furthermore, they assume uniformly priced and equally easy - in all directions - travelling. Fulfilment of these assumptions predicts 'closest store' patronising behaviour of customers in the case of single-purpose shopping.

[1] Brown, 2001

2.2. Store choice literature

Customer store choice behaviour is concerned with the choice of a particular outlet to shop at, out of all possible alternatives.

We shall classify all store choice models into three broad groups (Colome and Serra, 2003). Within the first group of models are those models that use a descriptive – deterministic approach; the other two groups, revealed preference models and direct utility models, use a stochastic approach.

Descriptive – deterministic models use observations or normative assumptions in order to determine consumer behaviour. The three most important models of this group are analogue modelling, the normative theory approach and Reilly's law of retail gravitation.

Revealed preference models rely upon information revealed by consumers' past behaviour to make predictions about consumers' possible future store choice behaviour. This group of models can further be divided into three sub-groups: spatial interaction models; discrete choice logit models; and dynamic spatial models.

Each of these models relies upon Luce's axiom regarding the probability of patronage as a function of consumers' utility received from patronage. This axiom states that the probability of patronage of any particular store may be expressed as ratio between the utility for the customer in patronising that store and the sum of utilities of patronising all alternatives.

Spatial interaction models are based on Reilly's law of retail gravitation. Huff (1963; cited in MacKenzie, 1989) was the first to develop a probabilistic formula for the determination of consumer patronage behaviour by introducing Luce's axiom to Reilly's original formula. Since utility, according to Reilly, is positively related to size of store and negatively related to distance between customer and store, this means that the probability that a customer will patronise a particular store may be expressed by the following formula:

$$p_{ij} = \frac{U_{ij}}{\sum_j U_{ij}} = \frac{\dfrac{S_j\beta_1}{d_{ij}\beta_2}}{\sum_j \dfrac{S_j\beta_1}{d_{ij}\beta_2}}$$

where:

p_{ij} – probability that consumer from zone i will patronise store j,

U_{ij} – utility for customer i from shopping at store j,

S_j – size of store j,

d_{ij} – distance between costumer i and store j,

β_1 and β_2 - sensitivity parameters that have to be estimated

In order to overcome the "single attribute" problem in explaining a store's attractiveness in Huff's formula, Nakanishi and Cooper (1974) introduced the multiplicative competitive interaction (MCI) coefficient[2] into this formula, which allows for more than one variable in explaining the attractiveness of the store. Their model looks as follows:

$$p_{ij} = \frac{U_{ij}}{\sum_j U_{ij}} = \frac{\dfrac{\prod_k A_{kj}\beta_{k1}}{d_{ij}\beta_2}}{\sum_j \dfrac{\prod_k A_{kj}\beta_{k1}}{d_{ij}\beta_2}}$$

where:

A_{kj} – vector of k attractiveness variables of store j; k = 1, 2, … l

β_{kj} – vector of k coefficients on attractiveness variables of store j; k = 1, 2, … l

Calibration of such a model can be done by OLS, after logarithmic transformation and then dividing each variable by its geometric mean (Nakanishi and Cooper, 1974). Thus, the final version of their model, with k variables for stores' attractiveness, is as follows:

$$\ln \frac{p_{ij}}{\tilde{p}_{ij}} = \beta_o + \beta_1 \sum_k \frac{A_{kj}}{\tilde{A}_{kj}} - \beta_2 \frac{d_{ij}}{\tilde{d}_{ij}} + u_{ij}$$

where:

$\tilde{p}_{ij}$ - geometric mean of variable p_{ij}

$\tilde{A}_{kj}$ – $(\Pi A_{kj})^{1/m}$ – geometric mean of variable A_{kj}, k = 1, 2, … m

d_{ij} – $(\Pi d_{ij})^{1/m}$ – geometric mean of variable d_{ij}

The possibility of estimating this model by OLS further contributed to its attractiveness for application in retail business decision-making processes and for its wider use relative to models based on information integration or conjoint techniques.

Discrete – choice logit models use techniques such as logit, multinomial logit, and nested logit to derive probabilities for customers' patronisation. Huff's model can be considered a particular case of discrete choice

[2] This coefficient is a vector of variables capturing the store's attractiveness.

models known as multinomial models, developed by McFadden (1974).

The problem with both Huff's and McFadden's model is satisfaction of the so-called 'independence of irrelevant alternatives' property. This means that models do not take into account the change of probabilities of patronising existing stores caused by the entrance of a new store. Under this property, introduction of a new outlet will increase the denominator of these models, and therefore decrease the probability of patronising any other alternative in equal proportions, which is not supported by the existing empirical evidence.

Finally, dynamic spatial models are recent developments concerned with the analysis of retail market area evolution and open new directions to analysis of spatial phenomena. The main representative of this group of models is Allaway's Spatial Diffusion Model, based on diffusion theory.

The main shortcoming of revealed preference models is the context–dependence of their approach. For example, if stores' characteristics are very similar amongst all of the analysed stores, it will cause the regression analysis results to show statistical insignificance for that variable. This does not necessarily mean that it is unimportant to customers, but only that they choose other attributes to discriminate among stores, since that particular attribute does not give them scope for discrimination. Therefore, conclusions drawn from the results of these models cannot be simply applied to other trade areas.

The third group of models, and the second to use a stochastic approach, is the group of direct utility assessment models. These models use information integration (Louviere and Gaeth, 1987), conjoint (Popkowski and Timmermans, 2001) or logit techniques (Drezner et al., 1998) to estimate the consumers' utility function. It is made by obtaining data on consumers' estimating market shares for new, innovative retail formats, when there is no possibility to obtain past data. The most important representative of these models is one developed by Ghosh and Craig (1983), based on game theory.

The vast majority of methods presented here are intuitively very simple and have been broadly used for empirical analysis of retail trade attraction. Yet a paper by Clarkson et al. (1996), which provides empirical analysis of the usage of particular location techniques amongst major retailers in the UK, reveals that revealed preference models are predominant.

According to the above discussion, and taking into account the characteristics of Sarajevo trade area and the availability of data, it has been decided that the Retail Gravity Model with the MCI coefficient is the most appropriate model for the purpose of analysing consumer store choice behaviour. The model is based on the revealed-preference approach to the analysis of choice behaviour. The survey will be used to reveal past customers' patronage behaviour.

Hypotheses

The original Reilly's Law of Retail Gravitation states, in brief, that the probability of patronising store *j* by a customer *i* is positively correlated with the size (or attractiveness) of the store and negatively related to the squared distance between customer *i* and store *j*, based upon Newton's formula[3]. The main purpose of the analysis is to identify the main attractiveness attribute that customers use for the evaluation of alternative sites.

H_0 - 1: Customers evaluate stores on the basis of a range of store attributes

Another issue we wished to address in this analysis was customers' driving patterns, and for that reason we have chosen Sarajevo as a particularly adequate trade area for such analysis. An explanation follows. Sarajevo city is oval shaped, placed in the narrow valley of the Miljacka river. If we would like to put two equal sized circles within such an oval, the first group of hypermarkets would be concentrated roughly in the middle of one circle, and other group within the other circle. Within each of these groups of stores, there are differences between them that allow different customers to choose the one they like. If travel patterns would show that customers are mainly shopping within the one half of the city where they live, it would suggest that customers are reluctant to travel beyond the area where they already have a sufficient variety of stores. Hence we may define an additional hypothesis that we would like to test, which may be expressed as follows:

[3] According to the Newton's formula, coefficient on size is set to be 1, and coefficient on distance is set to be 2. Before Nakanishi and Cooper's (1974) solution for transformation of the gravity model into linear form, these values for coefficients were widely defined a priori, in order to avoid cumbersome non-linear regression analysis.

H_0 - 2: In their store choice decision-making process, customers evaluate alternative sites up to a certain "reservation distance".

We planned to test this by defining the breaking point between the two clusters of stores. The break point for the purpose of our analysis may be regarded as the limit up to which customers evaluate alternative stores at all, which is explained through the notion of "reservation distance" (Ghosh and Craig, 1991). The first cluster consists of four stores located in the eastern to central part of the city, and the other cluster consists of four stores located in the western part of the city[4]. Then, we planned to compare the number of customers who shop at one of the stores placed in their 'zone of origin' with the number of customers who shop at stores in the other zone. If the majority of customers (more than 50 %) shop at stores not from their zone, it would suggest a rejection of the hypothesis.

3. Empirical Analysis of Store Choice Behaviour

For the purpose of testing the hypotheses set forth in the previous chapter, and the completion of the task to identify the main variables of customer behaviour, a regression analysis of data collected by interviewing customers in Sarajevo was conducted. The model is based on the revealed-preference approach to the analysis of choice behaviour. The data are cross-section data for the period June – July 2005. The dependent variable in this model is the probability of patronage of store j by customers living in area i. Patronage is defined as the patronage of the "main store", preferred over the other alternatives. According to Rhee and Bell (2002, p.227), "the main store is defined as the store that receives the greatest allocation of consumer expenditures in the associated interval". The allocation of customers to the particular store he or she prefers was done on the basis of customers' answers to questions on the number of shopping trips to all of the stores together and the number of trips to particular stores in a given month. Customers' estimations of the importance they give to some store characteristics in their decision-making process and of their individual characteristics were used as regressors. The regression analysis was then run according to the MCI model in order to estimate the relationship between the regressand and the regressors defined above.

[4] Detailed map of Sarajevo with marked locations of supermarkets is available at request.

3.1. The Methodology

The methodology used for completion of this task is combined from and largely relies upon methodologies used in Colome and Serra (2003), Dennis et al. (2002), Smith and Sanchez (2003) and Yavas (2003).

The first step is identification of key supermarket attributes that play the main role for customers in their store choice decision-making process. A wide range of attributes has been drawn from existing marketing and consumer behaviour literature, along with location theory and store choice empirical analyses. Particularly extensive lists of these attributes may be found in McGoldrick (2002) and Smith and Sanchez (2003). Second, the wide list of attributes has been reduced due to data availability and the financial limitations of the project. A survey plan was then specified in order to collect all data necessary for the analysis, sources of data identified and the questionnaire designed[5]. Finally, estimation of the model has been conducted to obtain estimates of coefficients of variables identified in the previous steps.

3.2. Data Description

Hypermarket attributes. Store choice is a large multi-attribute problem. According to the existing literature, there are dozens of attributes which should be included among the list of relevant ones. According to the limited financial and human resources for conducting the research, we have first identified a list of 32 attributes that may be relevant. We then surveyed several teachers, researchers and marketing managers from Sarajevo, in order to shorten the list further. Thus, we made a final list of ten questions in our questionnaire. A survey among customers living in Sarajevo has been conducted in order to collect responses on the importance they give to these various attributes in making a decision where to shop. Importance is measured on a 1 – 5 scale, according to Hutcheson and Moutinho (1998).

Distance. Values for distance between a consumer's origin of trip and preferred store are obtained from answers to questions 5 and 6, which provide data on the consumer's main origin of shopping trip (whether it is home or office) and their perception of the median driving time between the origin of their trip and preferred store, respectively. Driving time as a measure of distance has been chosen as preferred over the distance measured

[5] The questionnaire is available on request

in meters, since we assumed that driving time will capture the disutility of shopping. This also makes different perceptions of distance by customers who use different modes of transport more comparable. In addition, the distance of the hypermarket from the city's main transport artery has been included as an additional variable which should improve capture of the stores' accessibility.

Competition. The level of competition is measured by Fotheringham's (1988) formula for competition, which states that the level of competition is positively related to the size of competitive stores and negatively related to the squared distance between the observed store and rival stores. This model states that the degree of competition decreases exponentially with an increase in the distance between stores, similar to the basic idea of OLS regression analysis. We have chosen to use driving time in this formula as well, since we used this measure for distance between customers and stores, and it seems more appropriate to use the same measures in both cases.

Advertising. We were not able to obtain the actual level of advertising expenditures for all hypermarkets. On the basis of a priori experience, we know that most hypermarkets conducted their advertising campaigns during the introduction period of the store (several months after opening) and all other advertising is limited to promotional flyers and catalogues. Hence, we believe that we may capture most advertising effects on customer behaviour through collecting data on the number of promotional flyers distributed to customers per month and two dummy variables for whether the hypermarket has any additional TV or outdoor campaign during the months we conducted the research (June and July)[6].

Relative prices. A variable on relative price is defined as a sum of deviations of store prices for the best selling products from the lowest price amongst all stores[7]. The product groups that capture the majority of expenditures of households are identified on the basis of results of the survey on households' expenditures conducted by the Statistics Office of Bosnia and Herzegovina[8]. In order to identify products that can be used for comparison, we have excluded all products that are not present in each of the hypermarkets analysed as well as products subject to promotional activities in any of these hypermarkets at the time of collecting the actual prices[9]. Thus, we identified the final list of twelve products that satisfied all these conditions. This way we identify the main products that are part of the common basket content of the majority of customers on most shopping trips and thus the products that customers use to form an impression about the overall price level of the hypermarket.

Basket size. The rationale for inclusion of the basket size variable lies in the assumption that customers weigh most of the store's attributes according to the value of the products for which they intend to shop there (Bell et al., 1998). The variable for basket size is obtained from answers to the question from a questionnaire, namely the one on the average expenditures per shopping trip.

Consumer demographics. The other three questions on customers' demographics, those on the size of their household, age of person in charge of shopping and household's monthly income, were present in a questionnaire and used for the creation of variables. These three variables have shown high statistical significance in most empirical work. Thus we decided to include it in our model, as well.

Tenant stores. One of the main distinctions between hypermarkets in Sarajevo is the presence or non-presence of tenant stores within the hypermarket's building. Four of the hypermarkets analysed (Mercator, Robot Commerce Hrasno, Robot Commerce Rajlovac and VF Komerc Wisa) have tenant stores, and four others (Velpro, VF Komerc Korea, Interex and Sam Shop) do not have tenant stores. The rationale for including a dummy variable for the presence of tenant stores is the assumption that tenant stores extend the product offer of a site, and thus its attractiveness.

3.3. A Survey

Taking into consideration the limited financial and human resources for the completion of the survey, it was decided to limit the number of respondents to 240 customers from the Sarajevo city area.

[6] Usually, the dummy variable for the stores that were opened in the last 12 months is used along with these two variables in order to capture the effect of opening period intensive advertising and a public relation campaign. Inclusion of this dummy in the model has been planned initially. However, since only Robot Rajlovac falls into this category, and since it is excluded from the list of hypermarkets due to the survey results, this variable has not been included in the model.

[7] Taking the deviations from the lowest price has computational reasons. This way we will have only positive values, which enable us to make a logarithmic transformation.

[8] LSMS – Living Standard Measurement Survey, Agency for Statistics of Bosnia and Herzegovina, 2005

[9] This was the period between 7. and 20. of July, 2005.

In order to achieve the highest possible representativeness of such a limited sample, the Sarajevo area was divided into 30 sub-areas, which constitutes an exhaustive list for questions about the origin of shopping trips. Eight households from each sub-area were then chosen for interviewing at random. Furthermore, interviewers were instructed to obtain cooperation from the member of the household who is in charge of weekly shopping and to ensure that the interviewee is shopping at one of the eight hypermarkets, and not at other types of shops (such as smaller supermarkets or open markets). In case interviewers could not obtain these conditions, they were instructed to choose another household next to the previous one.

We have to take into account the issue raised by Colome and Serra (2003) regarding limitations on the computation of geometric means. In case that any of the sub-areas defined does not have at least one customer shopping at any of the hypermarkets considered, the geometric mean for that particular relation would be equal to zero, which would make any further analysis infeasible. In order to make as many zones as possible, taking into account this computational problem, it was decided to divide Sarajevo into 30 sub-areas and to collect at least 8 questionnaires from each sub-area, which gives 240 questionnaires in total. The rationale for collecting 8 answers on the questionnaire is the trade-off between the purpose of the survey and limited resources. We believed that we should have at least 8 answers from each sub-area, because it gives at least the minimum chances for getting answers that solve the computational problem mentioned above.

Since we expected a priori that some of the sub-areas would not satisfy the abovementioned computational limitation, such areas were combined with the closest sub-area in order to make a new sub-area with this limitation satisfied. Also, after the data were collected, we excluded three hypermarkets (Velpro, Robot Rajlovac and Sam Shop) from the initial list because one did not appear as the preferred store for any customers in our sample, and the other two stores had customers who preferred them mainly in their original sub-areas, so it was impossible to derive areas such that each has customers who prefer each of these hypermarkets[10]. Therefore, we finally created 9 zones that, along with the final list of 5 hypermarkets, gave us 45 grouped observations, based on 204 individual ones[11]. It can be expected that estimates based on these nine groups will be reliable (Gujarati, 2003), since we have at least 19 individual observations in each group, which is sufficiently large for this purpose.

From the descriptive analysis of data collected, we can see first that more women are in charge of shopping, which is in line with our expectations. Out of 240 customers, 158, or 65,83 %, are women. This exceeds the percentage of women in the whole population, which is 51.2 %[12]. Furthermore, the level of preference for the store as a percentage of shopping trips to the preferred store from the total number of shopping trips to any of the hypermarkets, according to the analysis of answers to questions 2 and 3, is 78.8 %. The more important finding is that 102 out of 240 customers from the sample, or 42.5 % of customers, state that they shop at one store only. It is also worth noting that 55 customers go shopping once a month only, which may be surprising. However, since we have not included shopping trips to other store formats, this only suggests a high level of inter-format competition that hypermarkets face from supermarkets, open markets, and other retail formats present in Sarajevo.

The average expenditure per shopping trip of customers within the sample is 87.24 KM. The average monthly budget of households within the sample is 885.53 KM. Answers to question 5 reveal that 156 out of 240 customers use a car as their main mode of transport for shopping. Others answered differently; namely, 62 go shopping mainly by public transport, 18 of them by foot and 4 use some other type of transport. All four respondents who answered "other mode of transport" have specified "taxi" as the alternative mode. Customers have different perceptions about the time they need to get to their stores. For 23 customers it takes less than 5 minutes, and for 46 it takes between 5 and 10 minutes. For 89 it takes 10 to 15 min, and for 82 customers it takes even more than 15 min; average driving time is 12.55 minutes. It is important to mention that customers who live very close to each other and use the same mode of transport have sometimes very different perceptions of the distance to the store. We have also checked the

[10] This does not deteriorate the quality of the analysis conducted, since we excluded the hypermarkets which we concluded that are local or out-of town in nature and do not compete with other in-town hypermarkets directly.

[11] We have excluded 36 observations on customers preferring excluded stores and from the sub-areas 1, 28 and 29. These sub-areas are irrelevant for the analysis of the Sarajevo trade area, since most of the questionnaires from these sub-areas are from customers preferring excluded stores.

[12] Thematic Bulletin 03 - Gender, Agency for Statistics of Bosnia and Herzegovina, January 2005

number of customers living in the sub-area where some of the shops are placed. Out of the 50 customers interviewed in those areas, 32, or 64 %, shopped mainly at the store from the same zone.

All attributes are estimated as the main one, or as of above average importance, for choosing a hypermarket by a significant number of respondents. This supports our decision to include these attributes among those for which we wished to obtain information on the importance customers give to them.

3.4. Regression Analysis

The model we intend to use for the regression analysis states that the probability of patronage depends on the ratio of attractiveness of the patronised store to the sum of attractiveness of all alternative stores. Since we have measured the attractiveness of only the preferred store at the individual level, we have to group the data in order to create zones, which will have at least one customer for each of the alternative stores. In order to have zones that contain at least one customer patronising each of the alternative stores, and thus enable computation of the geometric means, we have rearranged data in the following way. First, we have excluded three stores that do not satisfy the condition of having customers from more than the closest sub-area. Those hypermarkets are Velpro, Robot Rajlovac and Sam Shop. Second, we have excluded three areas, since customers from those areas mainly shop at the stores placed in those sub-areas, which means that the excluded hypermarkets are local in nature and are not direct competitors to the other hypermarkets. We then combined the remaining 27 sub-areas in a manner that yielded the final list of 9 zones for the regression analysis. Answers on all questions at the individual level were averaged over the store within each zone. Hence we have obtained the final list of 45 observations (9 zones x 5 stores). Finally, the values obtained by averaging individual data were divided by the geometric mean over all stores within one zone, according to the suggestion by Nakanishi and Cooper (1974).

The MCI model chosen for the analysis of determinants of store choice, developed by Nakanishi and Cooper (1974), may be expressed as follows:

(see Appendix 1)

After log-centered transformation of the previous equation (Nakanishi and Cooper, 1974), we obtain the following equation, which may be estimated by OLS:

(see Appendix 2)

To estimate the model (using the Microfit 4.0 econometric package), the following variables were defined[13]. First, the list of attractiveness variables has been made. The variables named A1 to A10 are the answers collected to questions about the importance of stores' attributes. The problem of a limited number of final observations, and therefore the wish to save as many degrees of freedom as possible, forced the solution in which we have presented these answers through a binary variable; namely, where answers to the these questions that were 4 and 5 on the scale were given a value of 1, and other answers were given the value 0. The variable PRICE measures the difference in the price level between hypermarkets; the variable SIZE measures the size of the supermarket for groceries in m^2; and the variable SITE captures the total size of the hypermarket's site, including the size of supermarkets, tenant stores, warehouses and offices. The variable TRDIS measures the driving time between the main road and tramline in Sarajevo and the hypermarket, and is used as a proxy for the accessibility of the hypermarket from the main transport connection.

Four variables on demographic data are included in the original model. The first one, EXPEN, measures average expenditures of customer per shopping trip. The variable MONB controls for the household's monthly income. The variable AGE controls for customers' age, the variable HHSIZE is for the household's size, CHILDDV controls for number of children in the household, and GENDER controls for the gender of the customer.

Distance between the customer and the hypermarket obtained from the answers to questions about the driving time spent on shopping is presented by the variable DRIVT. To control for advertising expenditures of hypermarkets, we use two variables, TVDV as a dummy for TV advertising during the previous two months, and FLYER for the number of flyers distributed by hypermarkets every month. The presence of tenant stores is controlled by the TSDV variable, and the level of competition the hypermarket is faced with is measured by Fotheringham's formula and presented by the variable COMPSQ.

[13] The explanation of the creation of all the variables is presented in Appendix 3

Test Statistics	LM Version		F Version	
	Computed value	Critical value for 5 % ls	Computed value	Critical F (1, 40) value for 5 % ls
A:Serial Correlation	0.85 E-3[0.977]	3.84146	0.61 E-3[.980]	4.08
B:Functional Form	0.0080[0.929]	3.84146	0.0057[.940]	4.08
C:Normality	0.9762[0.614]	5.99147	Not applicable	
D:Heteroscedasticity	3.7510[0.053]	3.84146	3.9103[.054	4.08

Table 1: Summary results of diagnostic tests

Dummy variables for store characteristics are included, in order to estimate differences in probability that arise from the particular characteristics of stores[14].

Values for all quantitative variables were transformed into logarithms after dividing them by their geometric means. We have changed the names of those variables by adding LN in front of the original names in order to indicate logarithmic transformation.

Before any regression estimation we decided to calculate the Pearson's correlation matrix of all variables we plan to include in the model, since we expect high collinearity among some of the variables mentioned. This matrix suggests high collinearity among some variables[15].

The easiest method of remedying the problem of collinearity among regressors is to exclude one of the regressors correlated to other. We believe this method is appropriate in our case, since we believe that variables left in the model still capture a sufficient part of the specific characteristic we would like to measure. Both results from the correlation matrix and from the VIF analysis suggest exclusion of the variables LNPRICE, LNSIZE, LNFLYER, LNTRDIS and all store dummies. Therefore, we have decided to exclude them. Exclusion of these variables is additionally justified by the "context dependency" nature of the analysis. It means that this analysis is dependent on the context of the Sarajevo trade area, where some store attributes are not significant because of insufficient variation between stores, and customers use other attributes to discriminate between these stores.

After the exclusion of variables due to the results of correlation indicators, we have estimated the first model, with 23 variables. The model has good diagnostics, but shows the insignificance of certain variables, whose exclusion might improve the model. Using stepwise backward regression we deleted regressors one at a time up to the point at which there was no more improvement in values of adjusted R^2 and the F statistic of joint significance. These variables were excluded in the following order: TSDV; TVDV; A5; A9; LNAGE; A1; GENDER; A3; MOTR; CHILDDV and A6. Thus we obtained the final, parsimonious model. There are more insignificant variables in the model but we have decided not to exclude them, because these coefficients have t-values higher than 1 and the exclusion of such regressors is not always recommended (Gujarati, 2003).

If we compare the signs and statistical significance of the same variables in the first and final models, we can see that all coefficients retain the same sign and improve their statistical significance from the first to the final model. This suggests that the results obtained by this regression analysis may be considered robust.

3.5. Results

The final model we estimated is as follows.

$$LNY = \beta_O + \beta_1 A2 + \beta_2 A4 + \beta_3 A7 + \beta_4 A8 + \beta_5 A10 + \beta_6 LNEXPEN + \beta_7 LNDRIVT + \beta_8 LNSITE + \beta_9 LNMONB + \beta_{10} LNHHSIZE + \beta_{11} COMPSQ + u_{ij}$$

The dependent variable in this case, after grouping of individual data, is not binary, but the proportion of customers from zone i patronising, or better, preferring hypermarket j.

The summary results of diagnostic testing of the residuals from the model are presented in the following table.

As we can see from the table, the computed χ^2 and F values for the LM test for serial correlation are well below their critical values at the 5% level of significance, which suggests that we cannot reject the null hypotheses of no serial correlation among the residuals.

Also, the result of Ramsey's RESET test shows that the model we have estimated is well specified as a linear model. Moreover, the computed χ^2 and F values for the Jarque–Berra test for normality are below their critical

[14] We have not included dummy variables for zones, since it is considered as alternative model to the one we used. With these dummies included in the model, it is not necessary to divide variables by their geometric mean (Cooper and Nakanishi, 1988).

[15] We have chosen the rule of thumb that the correlation coefficient above 0.7 suggests high multicollinearity.

Variable's name	Variable's description	Coefficient	Standard error	T-Ratio [prob.]
CON	Intercept term	-0.5116	0.3918	-1.305 [0.201]
A2	**Dummy for importance of the restaurant and café**	**-1.0288**	**0.3182**	**-3.232 [0.003]**
A4	**Dummy for importance of the assortment**	**0.7179**	**0.3538**	**2.029 [0.051]**
A7	Dummy for importance of lower prices	0.3867	0.3284	1.177 [0.248]
A8	**Dummy for importance of fresh products offer**	**1.0217**	**0.3243**	**3.150 [0.003]**
A10	**Dummy for importance of the queue in front of cashier**	**-1.3143**	**0.3870**	**-3.396 [0.002]**
LNEXPEN	**Variable for household's average expenditures per shopping trip**	**-1.2115**	**0.7057**	**-1.717 [0.095]**
LNDRIVT	**Variable for driving time for shopping trip**	**-0.9390**	**0.4035**	**-2.327 [0.026]**
LNSITE	**Variable for the size of the retail site (in m^2)**	**0.7148**	**0.2535**	**2.819 [0.008]**
LNMONB	**Variable for the household's monthly budget**	**1.1392**	**0.5043**	**2.259 [0.031]**
LNHHSIZE	**Variable for household's size**	**-1.2854**	**0.6242**	**-2.059 [0.047]**
LNCOMPSQ	**Variable for the level of competition each retailer faces**	**-0.4317**	**0.1424**	**-3.031 [0.005]**

Table 2: Estimated coefficients of the final model

values for a 5% level of significance, so we can conclude that our assumption of normal distribution of residuals holds. In turn, this suggests that our results are not distorted by outliers. Finally, our assumption of homoskedastic residuals is supported by the results of the Koenker-Bassett test for heteroskedasticity.

To conclude, we cannot reject the hypotheses that the residuals from our model do not indicate serial correlation, wrong functional form, non-normal distribution or heteroskedasticity of variances. Therefore, the values of t-statistics obtained from testing the hypotheses of statistical insignificance of estimated coefficients may be treated as valid.

The following values of coefficients, standard errors and t-statistics for the final model were obtained by OLS estimation and are presented in the table below.

Estimation of the model identified retail site size, distance between customer and store; customer expenditure per shopping trip, household size and monthly budget, presence of restaurant and café within the store, assortment of products, offer of fresh products, and the effectiveness of cashiers as the main discriminatory attributes of the stores[16]. Although we left the price variable in the model, it has shown its statistical significance, so we decided not to give any comment on it.

Probability of patronage increases with store size, but it decreases with the distance to the store. These observations are in line with the underlying theory (Reilly's law) and the logic of the gravity models.

Probability of patronage increases if stores offer a wider assortment of all products and especially of fresh products, since these stores' characteristics are very important to customers.

If customers believe that a store has a better restaurant, café and effective cashiers, such stores will have a lower probability of being patronised by customers, holding other influences unchanged. It may be explained by the fact that customers have assigned high importance only to some of the store's attributes presented as optional. It raises the possibility that customers who assigned high importance to assortment and lower prices to products at the same time consider the presence of restaurants and the effectiveness of cashiers as unimportant. Regardless, we must not conclude that stores would capture more customers if they made their restaurants worse or cashiers less effective.

[16] The level of significance we have used in this case is 10%.

The probability of patronage is higher for customers with a higher monthly budget, who spend less per shopping trip and live in a smaller household, which is as we expected.

The stronger presence of competing hypermarkets, measured by Foteringham's formula, affects the probability of patronage negatively.

The value of F statistics for testing the joint significance of variables from the model is higher than the critical F value, thus we can reject the hypothesis of joint insignificance of the variables at the 1 % level of significance. Regressors included in the final model explain around 61 % of all the variations in the dependent variable.

4. Conclusions and Implications

The location theories of oligopolistic competition are based on the unrealistic assumption that customers shop at the closest store. The empirical evidence presented in this dissertation does not support this assumption. It suggests that customers care not only about distance to the store, but also about other stores' attributes in making their patronage decision.

The main issue of this dissertation was to determine which store's attributes, other than distance, determine customers' patronage behaviour. The methodology used for this purpose was regression analysis of the Multiplicative Competitive Interaction model, with a range of attributes we assumed to be taken into consideration by customers in their patronisation decision–making process. The results have shown the significant influence of the six store attributes on customers' patronisation behaviour. Those attributes are the quality of the restaurant and café within the store, range of product lines offered in the store, offer of fresh products, effectiveness of the cashier, size of the store, and distance to the store. The significance of these store attributes show that customers, in their store choice behaviour, make a trade off between the distance to store and other store attributes. The regression analysis also identified four significant variables included in the regression as control variables. Those are customer's expenditure per shopping trip, household's size and monthly budget and presence of competitive stores close to the store patronized.

The analysis has also shown that there is a "reservation distance" beyond which customers are reluctant to travel for shopping. Hence, retail gravity models should incorporate this fact, by allowing for a different set of alternatives considered by customers from different zones.

4.1. Managerial Implications

As expected, distance showed significance for customers' patronage decisions, both by the results of the regression and of the descriptive analysis. Each hypermarket succeeded in attracting the majority of customers from the nearest area. Customers evaluate alternative outlets in their patronisation decision-making process up to a certain distance. Presence of competing outlets in the nearest areas around the store strongly influence the store's overall performance. Hence, the decision on the location of the retail outlet still remains the single most important decision facing decision makers in retail businesses in Sarajevo.

Nevertheless, the empirical evidence presented in this paper reveals the influence of certain store attributes that drive customers to shop at stores other than the closest one. Although these results reveal the preferences of an average customer toward an average hypermarket, they still show the existence of a variety of factors influencing customers' behaviour. Decision makers in retail companies should be aware of these factors and processes of store choice behaviour.

The range of products and offer of fresh products within the hypermarket increase the probability of patronisation of such a hypermarket. Hence, the choice of the products mix is very significant decision for retailers, since it should have a positive impact on the success of any hypermarket outlet. However, for further conclusions we should have more data about the list of products customers buy at a particular outlet.

The significance of several customer demographic variables, namely variables on household size, monthly budget and expenditures per shopping trip, along with the high correlation between store attributes considered important suggests that managers in retail companies should be well aware of both consumer characteristics and their preferences in order to create the most effective marketing program for their customers.

The descriptive analysis shows that all store attributes were assigned as the most important by a significant part of the sample, but the regression analysis suggested their insignificance. One of the explanations may be the existence of clearly distinguished market segments attracted by different stores, but regression analysis has made this unclear by giving results for the average store. This may be an interesting direction for further research.

4.2. Limitations and Directions for Further Research

The main limitation of this research emerges from the limited sample size.

There are some additional limitations to be mentioned, which emerge from the nature of the gravity-type models. The model assumes that customers evaluate each store within the trade area, which is not always true. As the results from our analysis show, customers reject stores beyond a certain "reservation distance", and also may reject stores that do not satisfy a certain level of the attribute that plays the main role for the customer. Also, we have to take into account the "context dependency" nature of the model. The insignificance of any store's attribute in the regression analysis does not necessarily mean that the attribute is unimportant to customers, but it is possible that there is insufficient variation in the attributes between stores, and thus customers use other attributes to discriminate between stores in making their patronisation decision. This means that the results of this model for one trade area should not be simply applied to other trade areas.

The results that suggest a possible high level of inter-format competition suggest the need for further analysis of the determinants of the store choice behaviour of customers based on data that would include other retail formats. Particularly important for the gravity-based models would be the analysis of inter-format store choice behaviour that would include e-commerce as the new retail format, where distance is no longer the most important variable.

Also, the dependent variable is based upon revealed patronisation during one month, which may suggest the level of preference customers have to a particular store. Nevertheless, customers change their preferences due to various influences according to the stores' marketing activities, changes in the environment, and so forth. Hence, longitudinal analysis of the store choice behaviour of customers through time may reveal the level of loyalty customers have to particular stores, track their switching behaviour and determine the main factors that influence the degree of and changes in loyalty among customers.

References

Amine, A. and Cadenat, S., 2003. Efficient retailer assortment: a consumer choice evaluation perspective, International Journal of Retail and Distribution Management, Vol. 31, No. 10, pp. 486-497

Antonides, G. and Raaij, van F. W., 1998. Consumer Behaviour: A European Perspective, John Wiley & Sons, Chichester, UK

Baltas, G. and Papastathopoulou, P., 2003. Shopper Characteristics, product, and Store Choice Criteria: a survey in the Greek grocery sector, International Journal of Retail and Distribution Management, Vol. 31, No. 10, pp. 498-507

Bell, D. R.; Ho, T.-C.; Tang, C. S., 1998. Determining Where to Shop: Fixed and Variable Costs of Shopping, Journal of Marketing Research, Vol. 35, August 1998, pp. 352-369

Brown, S., 2001. Retail Location Theory: evolution and evaluation,

Clarke, I.; Hallsworth, A.; Jackson, P.; Kervenoael, R.; Perez-del-Aguila, R.; Kirkup, M., 2004. Retail competition and consumer choice: contextualising the 'food desert' debate, International Journal of Retail and Distribution Management, Vol. 32, No. 2, pp. 89-99

Clarkson, R. M.; Clarke-Hill, C. M; Robinson, T., 1996. UK supermarket location assessment, International Journal of Retail and Distribution Management, Vol. 24, No. 6, pp. 22-33

Colome, R. and Serra, D., 2003. Supermarket Key Attributes and Location Decisions: A Comparative Study between British and Spanish Consumers, Institut D'Estudis Territorials, Working Paper, October 2003

Cooper, L. and Nakanishi, M., 1988. Market Share Analysis: Evaluating Competitive Marketing Effectiveness, Kluwer Academics, Boston, MA

Cox, R. and Brittain, P., 2000. Retail Management, Fourth Edition, Pearson Education Limited, FT-Prentice Hall, Essex, UK

Dennis, C.; Murphy, J.; Marsland, D.; Cockett, T.; Patel, T., 2002. Measuring Image: shopping centre case studies, International Review of Retail, Distribution and Consumer Research, 12:4, October 2002, pp. 355-373

Drezner, T., 1994a. Locating a single new facility among existing, unequally attractive facilities, Journal of Regional Science, Vol. 42, No. 2, pp. 287-299

Drezner, T., 1994b. Optimal Continuous Location of a Retail Facility, Facility Attractiveness and Market Share: An Interactive Model, Journal of Retailing, Vol. 70, No. 1, pp. 49-64

Drezner, Z.; Wesolowsky, G. O.; Drezner, T., 1998. On the Logit Approach to Competitive Facility Location, Journal of Regional Science, Vol. 38, No. 2, pp. 313-327

Drezner, T.; Drezner, Z.; Shiode, S., 2002. A Threshold-Satisfying Competitive Location Model, Journal of Regional Science, Vol. 42, No. 2, pp. 287-299

Fotheringham, A. S., 1988. Consumer Store Choice and Store Set Definition, Marketing Science, Vol. 7, No. 3, Summer 1988, pp. 299-310

Fotheringham, A. S.; Brunsdon, C.; Charlton, M., 2000. Quantitative Geography: Perspectives on Spatial Data Analysis, SAGE Publications, London, UK

Ghosh, A. and Craig, C. S., 1991. FRANSYS: a franchise distribution system location model, Journal of Retailing, Vol. 67, No. 4, Winter 1991, pp. 467-495

Gonzales-Benito, O., 2001. Inter-format spatial competition of Spanish hypermarkets, International Review of Retail, Distribution and Consumer Research, 11:1, January 2001, pp. 63-81

Gonzales-Benito, O., 2002. Geodemographic and socioeconomic characterisation of the retail Attraction of leading hypermarket chains in Spain, International Review of Retail, Distribution and Consumer Research, 12:1, January 2002, pp. 81-103

Gujarati, D. N., 2003. Basic Econometrics, Fourth Edition, McGraw-Hill, New York,

Hutcheson, G. D. and Moutinho, L., 1998. Measuring Preferred Store Satisfaction Using Consumer Choice Criteria as a Mediating Factor, Journal of Marketing Management, Vol. 14, pp. 705-720

Kmenta, J., 1971. Elements of Econometrics, Macmillan, New York, USA

Landsverk, O. B.; Hughes, D.; Fearne, A., 2003. Shopper Loyalty and Store Choice: Insight from a study of Norwegian supermarkets, European Retail Digest, Issue 38, Summer 2003, pp. 1-8

Louviere, J. J. and Gaeth, G. J., 1987. Decomposing the determinants of retail facility choice using the method of hierarchical information integration: a supermarket illustration, Journal of Retailing, Vol. 63, No. 1, pp. 25-48

LSMS – Living Standard Measurement Survey, Agency for Statistics of Bosnia and Herzegovina, 2005

MacKenzie, S. Botum, 1989. Retail Gravity Model, The Appraisal Journal, April 1989, pp. 166-172

McFadden, D., 1974. Conditional Logit Analysis of Qualitative Choice Behaviour, in P. Zarembka (ed.) Frontiers in Econometrics, Academic Press, New York, USA

McGoldrick, P. J., 2002. Retail Marketing, Second Edition, McGraw-Hill, London, UK

Moye, L. N. and Kincade, D. H., 2002. Influence of usage situations and consumer shopping orientations on the importance of retail store environment, International Review of Retail, Distribution and Consumer Research, 12:1, January 2002, pp. 59-79

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

Popkowski Leszczyc, P. T. L. and Timmermans, H. J. P., 2001. Experimental Choice Analysis of Shopping Strategies, Journal of Retailing, Vol. 77, pp. 493-509

Rhee, H. and Bell, D. R., 2002. The inter-store mobility of supermarket shoppers, Journal of Retailing, Vol. 78, pp. 225-237

Smith, D. and Sanchez, S. M., 2003. Assessment of business potential at retail sites: empirical findings from a US supermarket chain, International Review of Retail, Distribution and Consumer Research, 13:1, January 2003, pp. 37-58

Steenkamp, J.-B. E. M., 1991. Segmenting Retail Markets on Store Image using a Consumer-Based Methodology, Journal of Retailing, Vol. 67, No. 3, Fall 1991, pp. 300-320

Thematic Bulletin 03 - Gender, Agency for Statistics of Bosnia and Herzegovina, January 2005

Woodside, A. and Trappey, R. J. III, 1992. Finding out why customers shop at your store and buy your brand: automatic cognitive processing models of primary choice, Journal of Advertising Research, Nov-Dec 1992, pp. 59-78

Yavas, U., 2003. A multi-attribute approach to understanding shopper segments, International Journal of Retail and Distribution Management, Vol. 31, No. 11, pp. 541-548

Appendix 1: Specification of the Original MCI Model

$$\ln p_{ij} = \frac{\beta 0 + \sum_k \ln \beta_1 S_{kj} + \ln \beta_2 d_{ij} + \ln \beta_3 C_j + \sum_k \ln \beta_4 A_{ki} + \ln \beta_5 P_j + \sum_k \ln \beta_6 CD_{ki} + \ln \beta_7 TSDV_j + \ln \beta_9 SDV_j + \varepsilon_{ij}}{\sum_j (\beta 0 + \sum_k \ln \beta_1 S_{kj} + \ln \beta_2 d_{ij} + \ln \beta_3 C_j + \sum_k \ln \beta_4 A_{ki} + \ln \beta_5 P_j + \sum_k \ln \beta_6 CD_{ki} + \ln \beta_7 TSDV_j + \ln \beta_9 SDV_j + \varepsilon_{ij}}$$

where:

$\ln p_{ij}$ – logarithm of probability that consumer from sub-area i will patronise hypermarket at site j
β_0 – intercept term
$\Sigma_k S_{kj}$ – k factor variables on store's attractiveness, k = 1, 2, … 10
d_{ij} – distance between sub-area i and hypermarket j,
i = 1, 2, … 9; j = 1, 2, … 5
C_j – level of competition for hypermarket j
$\Sigma_k A_{kj}$ – 2 variables for the hypermarket's level of advertising,
P_j – variable representing relative prices of hypermarkets,
$\Sigma_k CD_k$ – k variables on costumer's demographics, k = 1, 2, … 7
$TSDV_j$ – dummy variable for presence of tenant stores, 1 – hypermarket has tenant stores, 0 – otherwise,
$\Sigma_j SDV_j$ – j hypermarket dummies, capturing specific store ignorance
u_{ij} – error term

Appendix 2: Specification of the Log-Transformed MCI Model

$$\ln \frac{p_{ij}}{\tilde{p}_{ij}} = \beta_0 + \sum_k \ln \beta_{1k} \frac{S_k}{\tilde{S}_k} + \ln \beta_2 \frac{d_{ij}}{\tilde{d}_{ij}} + \ln \beta 3 \frac{C_j}{\tilde{C}_j} + \sum_k \ln \beta_{4k} \frac{A_{kj}}{\tilde{A}_{kj}} + \ln \beta_5 \frac{P_j}{\tilde{P}_j} + \sum_k \ln \beta_6 \frac{CD_{ki}}{CD_{ki}} +$$
$$\ln \beta_7 TSDV_j + \ln \beta_8 \sum_j SDV_j + u_{ij} \quad (15)$$

where:

$\tilde{p}_{ij}$ – geometric mean of variable p_{ij}

$\sum_k \tilde{S}_{kj}$ – geometric means of k variables of store's attractiveness for m shops evaluated by customers from zone i,

$\tilde{d}_{ij}$ – geometric mean of variable d_{ij} to m shops from customers living in zone i,

$\tilde{C}_j$ – geometric mean of variable C_j,

$\sum_k A_{kj}$ – geometric means of variables of advertising,

$\tilde{P}_j$ – geometric mean of variable for relative price,

$\sum_k \tilde{C}D_{ki}$ – geometric means of k variables for customer's demographics,

$u_{ij} = \ln \left(\frac{\varepsilon_{ij}}{\tilde{\varepsilon}_{ij}} \right)$

Appendix 3: Names and Descriptions of Variables

Variable's name	Variable' s description
LNY	Dependent variable. Values obtained from the answers to question 1 from the questionnaire. At the individual level, value of 1 assigned to the sub-area and store answered as being preferred, 0 otherwise, then proportions of customers preferring a particular store were calculated for each zone at the group level. These values were then divided by their geometric mean. Finally, the values were transformed into the logarithmic form.
CON	Intercept term.
A1 – A10	Variables for the importance of stores' attributes. Values obtained from answers to question 9 from the questionnaire. Dummy variable which, at individual level, takes value 1 if customer answered 4 or 5 on the scale, 0 otherwise. At the group level, it is calculated as the proportions of customers within each zone, giving high importance or not to the attribute.
LNEXPEN	Quantitative variable representing customer's expenditure per shopping trip. Values obtained from the answers to question 10 from the questionnaire. At the group level, average expenditures were calculated for each group of customers preferring particular stores within one zone. These values were then divided by their zone's geometric mean. Finally, the values were transformed into the logarithmic form.
LNMONB	Quantitative variable whose values were obtained from the answers to the question on the household's monthly budget from the questionnaire. At the group level, the average household's budget was calculated for each group of customers preferring a particular store within one zone. These values were then divided by their zone's geometric mean. Finally, the values were transformed into the logarithmic form.
MOTR	Dummy variable. Values obtained from the answers to question 5. At the individual level, takes value 1 if customers go shopping by car, 0 otherwise. At the group level, it is calculated as proportions of customers within each zone going shopping by car or otherwise.

Variable's name	Variable' s description
LNDRIVT	Quantitative variable obtained from answers to the questions 7 and 8. Takes mid values of the range offered (e.g. if customers travels to store in 5 to 10 minutes, value of 7.5 has been assigned). Then, averages were calculated at the group level for each group of customers preferring a particular store within one zone. These values were then divided by their geometric mean. Finally, the values were transformed into the logarithmic form.
LNSIZE	Quantitative variable for the size of the hypermarket's selling space. These values were then divided by their zone's geometric mean. Finally, the values were transformed into the logarithmic form.
LNSITE	Quantitative variable for the size of the site's space (including warehouses, offices, and parking). These values were then divided by their zone's geometric mean. Finally, the values were transformed into the logarithmic form.
TVDV	Dummy variable. Takes value 1 if hypermarket had any additional mass media advertising in recent two months, 0 otherwise.
TSDV	Dummy variable. Takes value 1 if hypermarket has tenant stores, 0 otherwise.
LNPRICE	Quantitative variable for the store's price level. Calculated as the sum of deviations of prices of 12 representative products in the store from the lowest prices of products between these stores.
LNFLYER	Quantitative variable for the number of promotional flyers and catalogues store prints distributed regularly every month.
LNTRDIS	Quantitative variable for the distance of store from the main transport artery. Calculated as the mean driving time from the road to the store.
CHILDDV	Dummy variable for the number of children in the household. Takes value 1 if household has 2 or more children, 0 if less then 2. At group level, it is calculated as proportions of customers having two or more children, or otherwise.
GENDER	Dummy variable taking value of 1 if customer is male, 0 if female. At group level, it is calculated as proportions of customers within each zone being males or females.
LNHHSIZE	Values obtained from the answers to question 1 from the questionnaire. At the individual level, a value of 1 assigned to the sub-area and store answered as being preferred, 0 otherwise, then average size of households preferring particular store were calculated for each zone at the group level. These values were then divided by their geometric mean. Finally, the values were transformed into the logarithmic form.
LNAGE	Values obtained from the answers to question 1 from the questionnaire. At the individual level, a value of 1 was assigned to the sub-area and store answered as being preferred, 0 otherwise, then the average age of customers preferring a particular store were calculated for each zone at the group level. These values were then divided by their geometric mean. Finally, the values were transformed into the logarithmic form.
LNCOMPSQ	Quantitative variable for the influence of competition. Calculated by Fotheringham's formula, by using driving time between stores as a measure for distance.
SDV4 - SDV7	Store dummy variables. Store numbered as 3 (VF Wisa – see Appendix 2) is the benchmark store.