

GEOMARKETING MODELS IN SUPERMARKET LOCATION STRATEGIES

Amparo BAVIERA-PUIG¹, Juan BUITRAGO-VERA²,
Carmen ESCRIBA-PEREZ³

¹*Department of Economics and Social Sciences, Universitat Politècnica de València,
Camino de Vera s/n 46022 Valencia, Spain*

E-mails: ¹ambapui@upvnet.upv.es (corresponding author);

²jmbuitrago@esp.upv.es; ³carespe@upv.es

Received 03 May 2015; accepted 23 October 2015

Abstract. Choosing where to open a new outlet is a critical decision for retail firms. Building on the multiplicative competitive interaction model from retail location theory, this paper develops a geomarketing model that can be used to devise supermarket location strategies. First, attributes that explain a supermarket's pull on consumers were determined. These attributes included objective (taken from databases and empirical observation) and subjective (based on managerial judgements) variables relating to the supermarket and its trade area. Then, geographic information system tools were used to analyse real data at a highly detailed level (road section). From a geomarketing viewpoint, the model shows that sociodemographic characteristics of the supermarket's trade area affect firms' location strategies. The paper also discusses improvements for calibrating and validating this model. Adding the spatial organization of supermarkets to the model yields a different consumer behaviour pattern. This geomarketing model can help managers to design supermarket location strategies according to shop features, competitors and environment, whilst estimating supermarket sales.

Keywords: geomarketing, retail location theory, geographic information systems (GIS), multiplicative competitive interaction model (MCI), business planning, trade area, neural networks, spatial organization.

JEL Classification: L81, M31, R30.

Introduction

Opening a new retail outlet is expensive. Therefore, a firm that opens a new outlet is exposing itself to financial risk (Alarcón 2011). Location also affects retail consumer experience, which influences consumer loyalty (Kim, Choi 2013). Furthermore, if a shop fails because of a poorly chosen location, this failure may damage the parent firm's image. Location analysis is therefore vital for retail firms (Roig-Tierno *et al.* 2013). Although choosing a location for a retail outlet has always been difficult, the state of the current environment has made decision-making even more challenging. Increasingly, firms are facing greater competition, which reduces margins.

To decide where to locate a new outlet, a retail firm needs a strategy (Li, Liu 2012). Cheng *et al.* (2007) argue that geographic information systems (GIS), which rely on detailed digital maps and databases, will become essential tools for retail businesses to develop decision systems and select locations for new outlets. According to Clarke (1998), the use of GIS in retail location theory has become more widespread because of the importance of geodemographic characteristics. The study of geodemographic characteristics has led to great advances in consumer segmentation (Vigneau *et al.* 2014). Geodemographic segmentation discriminates between possible retail outlet sites depending on the profile of consumers living within the sites' trade area (Murad 2003). GIS tools make geodemographic segmentation possible.

This paper develops a sales forecast model that can contribute to retailers' strategies for selecting sites for new outlets. To produce this model, we followed three steps. First, we selected a location theory model to form the focus of this research. Second, we enriched this model by combining other factors and including new variables: sociodemographic characteristics (i.e., geomarketing variables) and subjective variables (i.e., managerial judgement on location, competition and environment). Third, we made methodological improvements to the model. We first develop the theoretical framework for this research. Then, we describe the model. Third, we set forth and analyse results. Finally, we present conclusions and key findings.

1. Theoretical framework

1.1. Retail location theory: MCI model

Retail location theory comprises four main theoretical areas: central place theory, spatial interaction theory, land value theory and the principle of minimum differentiation. The models derived from spatial interaction theory are the most widely used by firms when deciding where to locate new outlets (Merino, Ramirez-Nafarrate 2015; De Beule *et al.* 2014). The multiplicative competitive interaction (MCI) model, which conceptualizes consumers' spatial behaviour, is part of spatial interaction theory. Nakanishi and Cooper (1974) developed this model to generalize the model proposed by Huff (1964). The MCI model is as follows:

$$P_{ij} = \frac{\left(\prod_{k=1}^q A_{kj}^{\alpha_k} \right) D_{ij}^{\beta}}{\sum_{j=1}^n \left[\left(\prod_{k=1}^q A_{kj}^{\alpha_k} \right) D_{ij}^{\beta} \right]}, \quad (1)$$

where: P_{ij} – probability that a consumer living at i chooses retailer j ; A_{kj} – measure of variable k that describes the pull on consumers of retailer j ; α_k – sensitivity parameter with respect to variable k ; q – total number of variables k considered in the measure of pull on consumers; D_{ij} – distance between consumer location i and retailer j ; β – sensitivity parameter with respect to distance; n – number of retail firms considered by the consumer living at i .

By applying a logarithmic transformation and using geometric means ($\tilde{P}_i, \tilde{A}_k, \tilde{D}_i$), we transformed the mathematical equation of the model into the following regression equation (Nakanishi, Cooper 1982):

$$\log\left(\frac{P_{ij}}{\tilde{P}_i}\right) = \sum_{k=1}^q \alpha_k \log\left(\frac{A_{kj}}{\tilde{A}_k}\right) + \beta \log\left(\frac{D_{ij}}{\tilde{D}_i}\right). \quad (2)$$

Following this transformation, however, the model can be used with ratio variables only (Cooper, Nakanishi 1983). Similarly, binary variables must be transformed (Mahajan *et al.* 1978). For the MCI model, the property of independence of irrelevant alternatives holds. This property means that if a new choice (i.e., a new outlet) becomes available to consumers, this choice will compete on a level playing field with existing outlets. This result derives from Luce's (1959) definition of an outlet's utility, which states that an outlet's utility is independent of competitive context. Consequently, a company's trade policy exerts a uniform influence on all competitors. This property can create problems if outlets are not perceived as a uniform choice set, but the zeta-squared transformation helps to overcome this limitation (González-Benito *et al.* 2001).

Cliquet (1990) defines two types of the MCI model: objective and subjective. Objective models contain explanatory variables that are measured objectively. Conversely, variables in subjective MCI models are measured subjectively. Our use of the subjective model owes to two main reasons. First, perceptions of a retail outlet's attributes play an essential role in the selection process. Second, perceptions evolve more quickly than actual characteristics. Nevertheless, both types of variables – objective and subjective – can be incorporated into the same model.

1.2. Geomarketing

Latour and Le Floc'h (2001) define geomarketing as a system – comprising data, IT data-handling programs, statistical methods and graphical representations – designed to produce useful information for decision-making through tools that combine digital cartography, graphs and tables. Geomarketing offers a way of carefully and methodically analysing the location of target consumers to achieve greater profitability. Geomarketing works because local market potential and purchasing power depend on demographic characteristics within a shop's trade area (Grewal *et al.* 1999). According to Johnson (1989), geodemographic characteristics (i.e., geodemographics) classify people according to the type of neighbourhood where they live rather than according to conventional socioeconomic criteria such as income or social class. Sleight (1995) defines geodemographics as the analysis of demographic information (e.g., from censuses or large surveys) by geographical unit. Both concepts assume that people tend to congregate with others who are similar in terms of certain factors that may determine consumption, such as social status, household composition and ethnicity. We thus define geomarketing as the use of GIS to analyse data and make retail decisions, with the aim of meeting consumer needs and wants whilst making a profit.

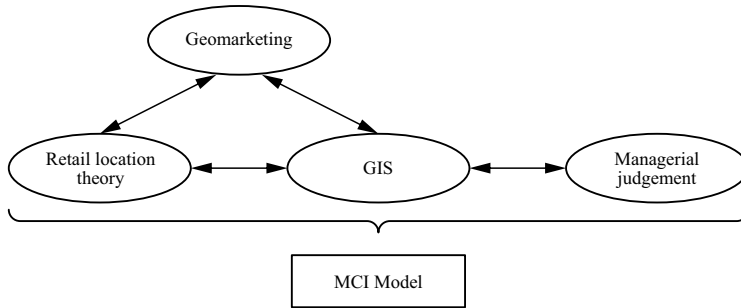


Fig. 1. Theoretical framework
 Source: Compiled by the authors.

1.3. Managerial judgement

Until now, exponents of the MCI model have used subjective variables from the consumer viewpoint (González-Benito *et al.* 2001; Cliquet 1995) but not from the firm perspective (i.e., managerial judgement). Blattberg and Hoch (1990) showed that firms enjoy more advantages if they combine simple statistical models with general intuition, using the two methods to complement rather than compete with one another. Experts are flexible and are capable of diagnosis and prediction, whereas models are more rigid and only predict. Conversely, models never tire nor suffer the influence of political, social or ideological pressures. As regards retail firm location, Mendes and Themido (2004) highlight the large subjective component inherent in these complex decision-making processes.

Figure 1 reflects the theoretical framework.

2. Data and methodology

2.1. Data

Research took place in the city of Castellón de la Plana in Spain. According to the latest data from the Spanish National Statistics Institute, Castellón de la Plana had a population of 180,185 as of 1 January 2013 – 0.01% less than in 2012. We considered 19 outlets belonging to three retail firms in the food sector: eight shops belonged to supermarket A, nine to supermarket B and two to supermarket C. These three supermarket chains are the biggest chains in Castellón de la Plana in terms of number of outlets. We omitted all supermarkets with a surface area of less than 600 square metres. By considering a single shop format (i.e., supermarkets), we ensured that we studied choice sets with a uniform competitive structure. We could therefore assume with some confidence the independence of irrelevant alternatives in the MCI model.

2.2. Configuring the subjective MCI model

The proposed model incorporates the following elements:

1. Objective variables that describe the outlet and its trade area. Distance from consumer to supermarket, measured in metres, is an objective variable.

2. Subjective variables, which correspond to how experts evaluate the outlet and its trade area.

The subscript i corresponds to the road section that falls within the retail outlet's trade area. Road sections lie within census blocks. They constitute the smallest census unit. The total number of road sections considered was 9 899. The model in this research was defined thus:

$$\pi_{ij} = \frac{\left(\prod_s ST_{sj}^{\alpha_s}\right)\left(\prod_t L_{ij}^{\alpha_t}\right)\left(\prod_h E_{hj}^{\alpha_h}\right)D_{ij}^\beta}{\sum_j \left[\left(\prod_s ST_{sj}^{\alpha_s}\right)\left(\prod_t L_{ij}^{\alpha_t}\right)\left(\prod_h E_{hj}^{\alpha_h}\right)D_{ij}^\beta\right]}, \quad (3)$$

where: π_{ij} – probability that a consumer in road section i chooses supermarket j (dependent variable); ST_{sj} – value of objective feature s of supermarket j ; L_{ij} – value of objective feature t of the trade area of supermarket j ; E_{hj} – managerial judgement h regarding features of supermarket j and of its trade area; D_{ij} – distance between road section i and supermarket j ; $\alpha_s, \alpha_p, \alpha_h, \beta$ – sensitivity parameters.

2.3. Variables

Objective and subjective variables alike were subdivided into two groups: features of the outlet and characteristics of its trade area. Data sources varied (see Table 1).

For objective variables of the supermarket (ST_{sj}), we used two sources. First, objective variables drew on data from the Nielsen database of supermarkets and hypermarkets in Spain. Nielsen compiles this database annually. Variables in this group were *sales floor area*, *number of checkouts*, *years operative* and *parking*. This group of variables also included two binary variables: *number of sections* and *refurbishment*. *Number of sections* measured how many departments the supermarket had. We considered 16 departments: fresh fruit and vegetables, fresh meat, packaged meat, deli counter, packaged deli products, bakery, fishmonger, frozen food, ready meals, dairy, savoury foods, sweet foods, drinks, pharmacy, health and beauty, and household (Dussart 1998). *Refurbishment* referred to whether the supermarket had undergone a refurbishment since its opening. We obtained this information by visiting supermarkets.

For objective variables regarding area of influence (L_{ij}), we used sociodemographic variables, which the Spanish National Statistics Institute produces for each road section in all Spanish cities. According to Campo *et al.* (2000), these variables are captured by four factors deriving from principal component analysis of 16 original variables (see Appendix for results of this analysis). These factors can be labelled *family* (Factor 1), *young single* (Factor 2), *separated* (Factor 3) and *unemployed* (Factor 4). We performed this analysis because of problems of correlation between the 16 initial variables. These four factors represented 89% of the original information supplied by all variables under study.

The variable *distance* (D_{ij}) was defined as the distance from the centroid of each road section to the supermarket's centroid. A centroid is the centre of mass of an object with

uniform density. For road sections with length L , the centroid was the mid-point of the line segment. GIS yielded this information. We calculated distances to the supermarket only for road sections that fell within the supermarket’s trade area. Since the MCI model is an attraction model of proximity, we considered only road sections closest to the supermarket. The trade area was a circle with a radius of 400 metres. We chose this distance because scholars have found that more than 80% of a supermarket’s consumers come from within this radius (Baviera-Puig et al. 2013; Applebaum 1966). We analysed 9 899 road sections to calibrate the model.

The subjective variables (judgements of experts, E_{hj}) came from a survey completed by shop managers. Managers evaluated their own supermarkets, those of competitors and the trade area, on a scale from 0 to 10. As per the literature (Smith, Sanchez 2003; Durvasula et al. 1992), we included the following variables in the questionnaire: *visibility, ease of access for pedestrians, ease of access by car, pedestrian volume, brand recognition, growth potential in the trade area, growth rate of competitors’ market share, aggressiveness of competitors’ strategies, and whether the supermarket has a star department.* We defined a star department as a department where the supermarket excels.

Table 1. Independent variables and data sources

Type	Category	Variables	Source
Objective variables	Supermarket features	Sales floor area Number of checkouts Years operative Parking	Nielsen
		Number of sections Refurbishment	Visit to supermarkets
	Characteristics of the trade area	Family Young single Separated Unemployed	Spanish National Statistics Institute
		Distance	Distance from each road section to the supermarket (within trade area)
Subjective variables	Supermarket features	Visibility Ease of access for pedestrians Ease of access by car	Survey to shop managers
		Star department	Survey to shop managers and visit to supermarkets
	Characteristics of the trade area	Pedestrian volume Brand recognition Growth potential in the trade area Growth of competitors’ market share Aggressiveness of competitors’ strategies	Survey to shop managers

Weisbrod *et al.* (1984) highlight the financial costs involved in obtaining information necessary to calibrate the MCI model for the dependent variable π_{ij} , referring specifically to the number of questionnaires needed. Drezner, T. and Drezner, Z. (2002) showed that the information available in secondary sources, with respect to consumers' purchasing power and volume of retail sales, may be used in gravitational models as both sources yield similar results. We therefore opted to use data provided by consumer loyalty schemes as our source of secondary information. Using GIS, we positioned supermarkets and consumers on a map, a process called geocodification. This yielded total sales by road section for each supermarket in 2013.

2.4. Treatment of subjective variables: zeta-squared transformation

Subjective variables tend to be interval rather than ratio variables, and often have arbitrary origins and disparate units of measurement. We therefore had to treat subjective variables before including them in the model. To resolve this problem, Cooper and Nakanishi (1983) developed the zeta-squared transformation, which exploits the physical concept of moments of inertia of a distribution of mass:

$$\zeta_{hij} = \begin{cases} (1 + z_{hij}^2)^{1/2} & \text{if } z_{hij} \geq 0 \\ (1 + z_{hij}^2)^{-1/2} & \text{if } z_{hij} \leq 0 \end{cases} \quad (4)$$

where: ζ_{hij} – transformed variable; z_{hij} – standard values of x_{hij} .

2.5. Treatment of binary variables

Binary variables also required treatment because the model does not admit null values. Mahajan and Jain (1977) evaluated three ways of achieving this goal: a) exponential transformation; b) transformation using a Likert-type scale; and c) transformation applied by Nakanishi *et al.* (1974). Mahajan and Jain (1977) conclude that these three methods are equally valid. We used transformation a): $X_{kj} = \exp(1) = e$, if supermarket j had feature k ; $X_{kj} = \exp(0) = 1$, if supermarket j did not have feature k .

2.6. Calibration and validation

After performing a logarithmic transformation of Equation (3), we compared the regressions yielded by stepwise ordinary least squares (OLS), generalized least squares (GLS) and a feed-forward artificial neural network (multilayer perceptron). To validate this model, we divided our sample into two groups: the training group (80% of the total sample) and the test group (20% of the sample). We used the coefficient of correlation (Ghosh, McLafferty 1987), coefficient of determination (R^2) (González-Benito *et al.* 2001), mean absolute percentage error (MAPE) (Klapper, Herwartz 2000) and root mean square error (RMSE) (Peña, Romo 1997) as goodness-of-fit indicators.

Before completing our analysis, we made one final comparison. Colome (2002) highlights one of the MCI model's deficiencies. Namely, the parameter β (sensitivity parameter with respect to distance) depends heavily on spatial organization of retail outlets because consumers consider both absolute and relative distances between outlets. The

parameter β is therefore spatially non-stationary because it is greater for origins where the ratio of distances between the furthest and closest outlets is higher. To resolve this problem, Ghosh (1984) advocates assuming a linear relationship between parameter β_i and a new parameter R_i , where R_i is defined as the quotient between the maximum and minimum distances from consumer at origin i , such that:

$$\beta_i = \omega + \theta R_i. \tag{5}$$

Upon incorporating this new variable in the model, the impact of distance breaks down into two categories: ω , which measures the impact of distance on choice; and θ , which measures the impact owing specifically to the consumer’s origin. As R_i varies depending on the consumer’s origin, θ measures the specific effect associated with each point of origin i . Substituting this relation into the original model (henceforth, simple subjective MCI model), the resulting model (henceforth, extended subjective MCI model or non-stationary subjective MCI model) becomes:

$$\log\left(\frac{P_{ij}}{\tilde{P}_i}\right) = \sum_{k=1}^q \alpha_k \log\left(\frac{A_{kj}}{\tilde{A}_k}\right) + \omega \log\left(\frac{D_{ij}}{\tilde{D}_i}\right) + \theta \log\left(\frac{D_{ij}}{\tilde{D}_i}\right) R_i. \tag{6}$$

3. Results and discussion

3.1. Methodological aspects

When we applied the simple subjective MCI model, the neural-network approach yielded better results for R^2 and RMSE in both training and test groups. In contrast, results for the correlation coefficient and MAPE imply that the best calibration technique was GLS (Table 2). González-Benito *et al.* (2001) report an R^2 of 13%, whereas González-Benito *et al.* (2000) report 41%. In this research, however, we observed R^2 of 88%, using calibration by neural networks, and 84%, using OLS. Results of the extended subjective MCI model were similar to results of the simple subjective MCI model (see Tables 2 and 3). Table 3 displays the differences between the two models in brackets. The same pattern of results recurred for different calibration methods. In Tables 2 and 3, best results by coefficient are underlined.

Table 2. Results for simple subjective MCI model

	Coef. correlation	R^2	MAPE	RMSE
OLS train	0.92	0.84	0.89	0.70
OLS test	0.92	0.84	0.63	0.70
GLS train	<u>0.96</u>	0.86	<u>0.87</u>	0.64
GLS test	<u>0.96</u>	0.86	<u>0.61</u>	0.64
NN train	0.94	<u>0.88</u>	0.87	<u>0.59</u>
NN test	0.94	<u>0.89</u>	0.62	<u>0.58</u>

Table 3. Results for extended subjective MCI model

	Coef. Correlation	R ²	MAPE	RMSE
OLS train	0.94 (0.02)	0.84 (0.00)	0.90 (0.01)	0.70 (0.00)
OLS test	0.94 (0.03)	0.84 (0.00)	0.63 (0.00)	0.69 (−0.01)
GLS train	<u>0.96 (0.00)</u>	0.87 (0.00)	<u>0.85 (−0.01)</u>	0.63 (−0.01)
GLS test	<u>0.96 (0.00)</u>	0.87 (0.01)	<u>0.59 (−0.02)</u>	0.62 (−0.02)
NN train	0.92 (−0.02)	<u>0.88 (0.00)</u>	0.88 (0.01)	<u>0.60 (0.01)</u>
NN test	0.92 (−0.02)	<u>0.89 (0.00)</u>	0.61 (−0.01)	<u>0.58 (0.01)</u>

Using GLS, numerous authors have specified MCI models that are reportedly simpler than OLS models and yield better results (Klapper, Herwartz 2000; Nakanishi, Cooper 1982). Some report better results with OLS (Ghosh *et al.* 1984). In our case, neural networks and GLS yielded better calibration results than OLS did. Neural networks capture non-linear relationships, which linear regressions are incapable of modelling. Furthermore, unlike linear regression techniques, neural networks do not require data to meet any criteria (i.e., normality, collinearity, homoscedasticity and linearity). Neural networks also quantify the importance of variables in the model and are preferable to linear regression techniques when using large data sets, as in the current research.

3.2. Variables that influence supermarket choice

In light of the advantages of neural networks versus GLS, we analysed results using neural networks for the simple and extended subjective MCI models (see Table 4). In the simple subjective MCI model, *distance* was the independent variables with greatest effect (100%) on the dependent variable. This finding coincides with those of Huff (1964), who defined distance, along with sales floor area, as a choice factor when consumers are deciding where to shop. The second most important variable was *ease of access for pedestrians* (32%). This finding is coherent with the results of Thrall (2002) and Gautschi (1981), who express a preference for measuring ease of access instead of transport cost because access depends on time required as well as cost.

Characteristics of trade area (sociodemographics) *unemployed* (30%) and *separated* (29%) were the next most important factors. This finding implies that consumer profile in a supermarket’s trade area is important when deciding where to purchase. The influence of competitors in consumers’ supermarket choice is reflected by the importance of the variables *growth of competitors’ market share* (27%), *brand recognition* (21%) and *aggressiveness of competitors’ strategies* (20%). These findings are consistent with those of Li and Liu (2012), who improved sales forecast models by considering competitors.

Finally, the most influential supermarket features in consumers’ supermarket choice were *years operative* (23%) and *number of checkouts* (20%). The importance of *number of checkouts* coincides with the findings of Kumar and Karande’s (2000) study into the effects of a retail outlet’s environment on its sales. This variable may relate to

sales floor area – greater area means more checkouts – or service efficiency – having more checkouts means faster payment. *Years operative* may have emerged as important because the longer the supermarket has been open, the greater locals’ awareness of its existence. This finding is coherent with Durvasula *et al.*’s (1992) proposal in their STORELOC model.

Table 4. Importance of independent variables

Variables	Simple subjective MCI model		Extended subjective MCI model	
	Importance	Normalized importance	Importance	Normalized importance
Sales floor area	0.03	14%	0.04	19%
Number of checkouts	0.04	20%	0.05	23%
Years operative	0.05	23%	0.06	28%
Parking	0.02	11%	0.02	10%
Number of sections	0.01	6%	0.02	12%
Refurbishment	0.02	9%	0.03	14%
Family	0.03	15%	0.03	13%
Young single	0.04	18%	0.04	20%
Separated	0.06	29%	0.04	19%
Unemployed	0.07	30%	0.05	24%
Distance	0.22	100%	0.20	100%
R	---	---	0.12	58%
Visibility	0.01	6%	0.03	13%
Ease of access for pedestrians	0.07	32%	0.03	15%
Ease of access by car	0.05	22%	0.04	20%
Star department	0.06	25%	0.03	17%
Pedestrian volume	0.04	17%	0.02	11%
Brand recognition	0.05	21%	0.03	15%
Growth potential in the trade area	0.03	12%	0.05	25%
Growth of competitors’ market share	0.06	27%	0.02	10%
Aggressiveness of competitors’ strategies	0.04	20%	0.05	25%

Note: The importance of an independent variable refers to how much the value forecast by the network model changes for different values of the independent variable. For a given variable, the normalized importance equates to the importance of that variable divided by the maximum importance among variables. It is expressed as a percentage.

In the extended subjective – or spatially non-stationary – MCI model, *distance* was again observed to have the greatest effect (100%) on the dependent variable. The quotient of maximum over minimum distance (i.e., R) equalled 58%. These results imply that the spatial organization of supermarkets and the relative distances between them (not only absolute distances) do indeed influence consumers' choice of supermarket (Ghosh 1984). The most influential variables in consumers' choice of supermarket were *years operative* (28%), *aggressiveness of competitors' strategies* (25%), *growth potential in the trade area* (25%) and *number of checkouts* (23%). The next most important variables corresponded to objective trade-area variables: *unemployed* (24%) and *young single* (20%).

When we considered the entire supermarket network (i.e., when applying an extended subjective MCI model), the variables with the greatest effect on the probability of choosing a supermarket were spatial organization of supermarkets, and supermarket features and competition. Possibly, when consumers in cities evaluate their buying options, they place a higher value on the distinctive features of each supermarket. Conversely, when we excluded supermarket spatial organization from the consumer decision process (i.e., when considering a simple subjective MCI model), the most influential variables in consumers' choice of supermarket were *distance*, *ease of access* and characteristics related to a supermarket's trade area (i.e., sociodemographic characteristics). The different sociodemographic variables used in each case (*separated* in the simple subjective MCI model and *young single* in the extended subjective MCI model) may have led to different behaviour in each sociodemographic group.

3.3. Managerial implications

This research increases our knowledge of consumer purchase behaviour and of key factors in supermarkets' pull on consumers. The MCI model can also forecast supermarket sales. The model predicts the probability that a consumer in any road section will purchase in a particular supermarket. This probability can then be multiplied by Castellón de la Plana's annual food spend per inhabitant to obtain total sales in that road section. The sum of spending in all road sections then yields total supermarket sales. For instance, total sales of one supermarket were 915 026.53 Euros. Using the neural networks calibration technique, the simple subjective model yielded an estimate of 853 590.83 Euros. The sales forecast by the model therefore has an error of –6.71% in this instance.

When opening a new outlet, a supermarket chain can input the expected features of the new supermarket and forecast its total sales. This calculation also yields total sales for all other supermarkets. Supermarket strategists can thus observe how market share varies between establishments. Similarly, supermarket strategists can explore several potential sales scenarios by adjusting the features of the new supermarket (e.g., *sales floor area*, *number of checkouts*, *parking*), and can select the most attractive scenario. Finally, this model contributes to defining a supermarket's product range, regardless of whether the supermarket is a new opening or an existing outlet. If any trade-area variable (i.e., a sociodemographic characteristic) stands out for its importance when defining the model, the supermarket management can adapt its product range to that particular sociodemographic group if it lies within the supermarket's trade area.

Conclusions

This paper develops a spatial interaction model that determines purchase probability. The model combines diverse yet related theoretical fields: retail location theory, geomarketing, GIS and managerial judgement. Main findings fall into two groups: conceptual and methodological.

From a geomarketing viewpoint, the model shows that sociodemographic characteristics of the supermarket's trade area affect firms' location strategies. Using subjective variables, we included managerial judgement in the model, thereby improving the model's goodness of fit. The most common subjective variables were *ease of access* and variables relating to competitors. The advantage of using subjective variables, which can be measured and included in the model, is that they relate to perceptions of attributes. The objective variables with the greatest effect on the dependent variables were *years operative* and *number of checkouts*. This rigorous approach reveals some keys to success in retail strategy and identifies significant variables.

GIS tools allowed us to analyse highly detailed data (9 899 road sections – the smallest census unit). Neural networks yielded the best results for the calibration of the MCI model. Using the multilayer perceptron procedure, the model yielded results that are comparable with those obtained using logistic regression or discriminant analysis, which suggests that data do not relate in ways these models are unable to capture. Adding the spatial organization of supermarkets to the model yielded a different consumer behaviour pattern. In other words, consumers evaluated different attributes, and therefore behaved differently depending on whether they considered the spatial organization of supermarkets (relative distances) or did not consider this organization (absolute differences). Therefore, managers should consider a greater range of variables.

With this model, managers can design supermarket location strategies as it estimates total supermarket sales. Unfortunately, subjective variables and those which take zero or negative values must be transformed before being included in the model. This limitation delays model specification. Similarly, the model should be recalibrated for each location to reflect the local competition. Future research could include other pull variables such as price, shop atmosphere, friendliness of staff, promotions and so forth. Scholars could also improve the methodology for greater ease of calibration and validation of the model. To consider spatial variance in the data, local regression models could be tested instead of global ones. For example, Suárez-Vega *et al.* (2015) report that the geographically weighted regression model performs better than global ones when using the Huff model. The same type of regression could be tested when using the MCI model. Neural networks could also be studied to incorporate local parameters.

Acknowledgements

The authors would like to thank the Consum-Universitat Politècnica de València Chair (Cátedra) for collaborating in this study.

Disclosure statement

The authors of this article state that we don't have any financial, professional, personal interest or benefit arising from the direct applications of our research.

References

- Alarcón, S. 2011. The trade credit in the Spanish agro-food industry, *Mediterranean Journal of Economics, Agriculture and Environment (New Medit)* 10: 51–57.
- Applebaum, W. 1966. Methods for determining store trade areas, market penetration, and potential sales, *Journal of Marketing Research* 3: 127–141. <http://dx.doi.org/10.2307/3150201>
- Baviera-Puig, A.; Roig-Tierno, N.; Buitrago-Vera, J.; Mas-Verdu, F. 2013. Comparing trade areas of technology centres using geographical information systems, *The Service Industries Journal* 33(7–8): 789–801. <http://dx.doi.org/10.1080/02642069.2013.740467>
- Blattberg, R. C.; Hoch, S. J. 1990. Database models and managerial intuition: 50% model + 50% manager, *Management Science* 36(8): 887–899. <http://dx.doi.org/10.1287/mnsc.36.8.887>
- Campo, K.; Gijbrecchts, E.; Goossens, T.; Verhetsel, A. 2000. The impact of location factors on the attractiveness and optimal space shares of product categories, *International Journal of Research in Marketing* 17: 255–279. [http://dx.doi.org/10.1016/S0167-8116\(00\)00026-4](http://dx.doi.org/10.1016/S0167-8116(00)00026-4)
- Cheng, E. W. L.; Li, H.; Yu, L. 2007. A GIS approach to shopping mall location selection, *Building and Environment* 42: 884–892. <http://dx.doi.org/10.1016/j.buildenv.2005.10.010>
- Clarke, G. 1998. Changing methods of location planning for retail companies, *GeoJournal* 45: 289–298. <http://dx.doi.org/10.1023/A:1006995106736>
- Cliquet, G. 1990. La mise en oeuvre du modèle interactif de concurrence spatiale (MICS) subjectif [The implementation of the subjective interactive model of spatial competition (MCIS)], *Recherche et Applications en Marketing* 5(1): 3–18. <http://dx.doi.org/10.1177/076737019000500101>
- Cliquet, G. 1995. Theory and methodology. Implementing a subjective MCI model: an application to the furniture market, *European Journal of Operational Research* 84: 279–291. [http://dx.doi.org/10.1016/0377-2217\(93\)E0313-M](http://dx.doi.org/10.1016/0377-2217(93)E0313-M)
- Colome, R. 2002. *Consumer choice in competitive location models*: PhD thesis. Universidad Pompeu Fabra, Barcelona.
- Cooper, L. G.; Nakanishi, M. 1983. Standardizing variables in multiplicative choice models, *Journal of Consumer Research* 10: 96–108. <http://dx.doi.org/10.1086/208948>
- De Beule, M.; Van den Poel, D.; Van de Weghe, N. 2014. An extended Huff-model for robustly benchmarking and predicting retail network performance, *Applied Geography* 46: 80–89. <http://dx.doi.org/10.1016/j.apgeog.2013.09.026>
- Drezner, T.; Drezner, Z. 2002. Validating the gravity-based competitive location model using inferred attractiveness, *Annals of Operations Research* 111: 227–237. <http://dx.doi.org/10.1023/A:1020910021280>
- Durvasula, S.; Sharma, S.; Andrews, J. C. 1992. STORELOC: a retail store location model based on managerial judgments, *Journal of Retailing* 68(4): 420–444.

- Dussart, C. 1998. Category management: Strengths, limits and developments, *European Management Journal* 16(1): 50–62. [http://dx.doi.org/10.1016/S0263-2373\(97\)00073-X](http://dx.doi.org/10.1016/S0263-2373(97)00073-X)
- Gautschi, D. A. 1981. Specification of patronage models for retail center choice, *Journal of Marketing Research* XVIII(2): 162–174. <http://dx.doi.org/10.2307/3150951>
- Ghosh, A. 1984. Parameter nonstationarity in retail choice models, *Journal of Business Research* 12(4): 425–436. [http://dx.doi.org/10.1016/0148-2963\(84\)90023-7](http://dx.doi.org/10.1016/0148-2963(84)90023-7)
- Ghosh, A.; McLafferty, S. L. 1987. *Location strategies for retail and service firms*. Lexington: Lexington Books.
- Ghosh, A.; Neslin, S.; Shoemaker, R. 1984. A comparison of market share models and estimation procedures, *Journal of Marketing Research* 21(2): 202–210. <http://dx.doi.org/10.2307/3151702>
- González-Benito, O.; Grotorex, M.; Muñoz-Gallego, P. A. 2000. Assessment of potential retail segmentation variables. An approach based on a subjective MCI resource allocation model, *Journal of Retailing and Consumer Services* 7: 171–179. [http://dx.doi.org/10.1016/S0969-6989\(99\)00026-0](http://dx.doi.org/10.1016/S0969-6989(99)00026-0)
- González-Benito, O.; Muñoz, P. A.; Matías, A. 2001. Determinantes estratégicos de la selección de establecimiento minorista: aplicación de un modelo MCI subjetivo de asignación de recursos [Strategic determinants of the choice of retailer: application of a subjective MCI model of resource allocation], *Revista Europea de Dirección y Economía de la Empresa* 10(1): 177–194.
- Grewal, D.; Levy, M.; Mehrotra, A.; Sharma, A. 1999. Planning merchandising decisions to account for regional and product assortment differences, *Journal of Retailing* 75(3): 405–424. [http://dx.doi.org/10.1016/S0022-4359\(99\)00015-9](http://dx.doi.org/10.1016/S0022-4359(99)00015-9)
- Huff, D. L. 1964. Defining and estimating a trading area, *Journal of Marketing* 28(3): 34–38. <http://dx.doi.org/10.2307/1249154>
- Johnson, M. 1989. The application of geodemographics to retailing-meeting the needs of the catchment, *Journal of the Market Research Society* 31(1): 7–36.
- Kim, H.; Choi, B. 2013. The influence of customer experience quality on customers' behavioral intentions, *Services Marketing Quarterly* 34(4): 322–338. <http://dx.doi.org/10.1080/15332969.2013.827068>
- Klapper, D.; Herwartz, H. 2000. Forecasting market share using predicted values of competitive behavior: further empirical results, *International Journal of Forecasting* 16: 399–421. [http://dx.doi.org/10.1016/S0169-2070\(00\)00052-2](http://dx.doi.org/10.1016/S0169-2070(00)00052-2)
- Kumar, V.; Karande, K. 2000. The effect of retail store environment on retailer performance, *Journal of Business Research* 49(2): 167–181. [http://dx.doi.org/10.1016/S0148-2963\(99\)00005-3](http://dx.doi.org/10.1016/S0148-2963(99)00005-3)
- Latour, P.; Le Floc'h, J. 2001. *Géomarketing: principes, méthodes et applications* [Principles, methods and applications]. Paris: Éditions d'Organisation.
- Li, Y.; Liu, L. 2012. Assessing the impact of retail location on store performance: a comparison of Wal-Mart and Kmart stores in Cincinnati, *Applied Geography* 32(2): 591–600. <http://dx.doi.org/10.1016/j.apgeog.2011.07.006>
- Luce, R. 1959. *Individual choice behavior*. New York: John Wiley & Sons.
- Mahajan, V.; Jain, A. K. 1977. An examination of operational problems with multiplicative competitive interaction models, in B. A. Greenberg, D. N. Bellinger (Eds.). *Contemporary marketing thought*. Chicago, IL: American Marketing Association.
- Mahajan, V.; Jain, A. K.; Ratchford, B. T. 1978. Use of Binary attributes in the multiplicative competitive interactive choice model, *Journal of Consumer Research* 5: 210–215. <http://dx.doi.org/10.1086/208733>

- Mendes, A. B.; Themido, I. H. 2004. Multi-outlet retail site location assessment, *International Transactions in Operational Research* 11: 1–18.
<http://dx.doi.org/10.1111/j.1475-3995.2004.00436.x>
- Merino, M.; Ramirez-Nafarrate, A. 2015. Estimation of retail sales under competitive location in Mexico, *Journal of Business Research* 69(2): 445–451.
<http://dx.doi.org/10.1016/j.jbusres.2015.06.050>
- Murad, A. A. 2003. Creating a GIS application for retail centers in Jeddah city, *International Journal of Applied Earth Observation and Geoinformation* 4: 329–338.
[http://dx.doi.org/10.1016/S0303-2434\(03\)00020-5](http://dx.doi.org/10.1016/S0303-2434(03)00020-5)
- Nakanishi, M.; Cooper, L. G. 1974. Parameter estimation for a multiplicative competitive interaction model-least squares approach, *Journal of Marketing Research* 11: 303–311.
<http://dx.doi.org/10.2307/3151146>
- Nakanishi, M.; Cooper, L. G. 1982. Simplified estimation procedures for MCI models, *Marketing Science* 1(3): 314–322. <http://dx.doi.org/10.1287/mksc.1.3.314>
- Nakanishi, M.; Cooper, L. G.; Kassarian, H. H. 1974. Voting for a political candidate under conditions of minimal information, *Journal of Consumer Research* 1: 36–43.
<http://dx.doi.org/10.1086/208589>
- Peña, D.; Romo, J. 1997. *Introducción a la estadística para las ciencias sociales* [Introduction to statistics for the social sciences]. Madrid: McGraw-Hill.
- Roig-Tierno, N.; Baviera-Puig, A.; Buitrago-Vera, J.; Mas-Verdú, F. 2013. The retail site location decision process using GIS and the analytical hierarchy process, *Applied Geography* 40: 191–198. <http://dx.doi.org/10.1016/j.apgeog.2013.03.005>
- Sleight, P. 1995. Neighbourhood watch: geodemographic and lifestyle data in the UK GIS marketplace, *Mapping Awareness* 9(6): 18–21.
- Smith, L. D.; 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): 37–58. <http://dx.doi.org/10.1080/0959396032000051684>
- Suárez-Vega, R.; Gutiérrez-Acuña, J. L.; Rodríguez-Díaz, M. 2015. Locating a supermarket using a locally calibrated Huff model, *International Journal of Geographical Information Science* 29(2): 217–233. <http://dx.doi.org/10.1080/13658816.2014.958154>
- Thrall, G. I. 2002. *Business geography and new real estate market analysis*. New York: Oxford University Press.
- Vigneau, E.; Charles, M.; Chen, M. 2014. External preference segmentation with additional information on consumers: a case study on apples, *Food Quality and Preference* 32: 83–92.
<http://dx.doi.org/10.1016/j.foodqual.2013.05.007>
- Weisbrod, G. E.; Parcells, R. J.; Kern, C. 1984. A disaggregate model for predicting shopping area market attraction, *Journal of Retailing* 60(1): 65–83.

APPENDIX

Sixteen original variables described the distribution of several sociodemographic variables in the population of each store's trade area. The second column of Table A.1 describes each variable. To reduce the number of variables, we used principal component analysis. Four factors explained 89% of total variation. Factor loadings, obtained after a varimax rotation with Kaiser, appear in columns 3 to 6 of Table A.1.

Table A.1. Principal component analysis of sociodemographic variables

Variables	Description	Factor 1	Factor 2	Factor 3	Factor 4
1	% Spanish citizens	0.39	<u>0.80</u>	-0.32	-0.23
2	% of population under 15 years	<u>0.82</u>	0.49	0.05	-0.24
3	% of population 15–29 years	0.20	<u>0.88</u>	0.04	0.22
4	% of population 30–49 years	<u>0.90</u>	0.33	0.08	-0.03
5	% of population 50–64 years	<u>-0.91</u>	-0.05	-0.19	-0.10
6	% men	0.60	0.16	<u>0.61</u>	-0.15
7	% single	-0.05	<u>0.89</u>	0.01	-0.39
8	% married	<u>0.93</u>	-0.09	-0.06	0.16
9	% widowed	<u>-0.74</u>	<u>-0.61</u>	0.10	0.15
10	% separated	-0.14	-0.08	<u>0.88</u>	0.07
11	% students	-0.53	0.19	<u>-0.76</u>	-0.17
12	% employed	<u>0.84</u>	0.49	0.02	-0.02
13	% unemployed	0.26	0.01	0.08	<u>0.86</u>
14	% retired	-0.64	<u>-0.72</u>	0.12	0.20
15	% house workers	<u>0.81</u>	0.29	0.22	0.20
16	% other occupations	-0.27	-0.46	0.05	<u>0.70</u>
Cumulative % of variance explained		40.53	65.85	78.01	88.84

Note: Values of variables most strongly related to the factor are underlined.

Since factors may take positive or negative values but the model only accepts strictly positive values, we transformed the factors as follows:

- First, proportional transformation between 0 and 1:

$$t = (x - x_{\min}) / (x_{\max} - x_{\min}). \quad (\text{A.1})$$

- Second, exponential transformation:

$$x = e^t. \quad (\text{A.2})$$

Amparo BAVIERA-PUIG (PhD Universitat Politècnica de València, 2011) is a senior lecturer at the Universitat Politècnica de València (Spain) where she teaches courses in marketing, communication and CSR. Her primary research interests are in the area of geographical analysis applied to marketing and innovation. Her secondary research interests include the application of multicriteria decision models in marketing and communication. She has published several papers and book chapters on these topics.

Juan BUITRAGO-VERA, PhD, is a senior lecturer at the Universitat Politècnica de València (Spain). In the field of Marketing and Retailing he is author of diverse papers in indexed journals and conferences. He is also author of various chapters in books. He has ongoing research interests in the geographical analysis and models design applied to retailers.

Carmen ESCRIBA-PEREZ, PhD, is a senior lecturer in the Faculty of Business Administration and Management at the Universitat Politècnica de València, Spain. She has authored several journal articles and book chapters on agrifood marketing. She teaches marketing and trade management at the Universitat Politècnica de València.