

The selection of food shopping locations by UK suburban households: A multivariate analysis

A.R. Yates and K.A. Yeomans

Graduate School of Business Administration, University of the Witwatersrand, Johannesburg

In this article a behavioural model of retail location is discussed which differs from established models in two respects. Firstly, in the behavioural model the demand for retail outlets is emphasized rather than the supply of retail outlets. Secondly, psychological and sociological constructs are used to help explain store selection behaviour. The behavioural model differs from established location theory in that an attempt is made to study the act of shopping in relation to other human activities. The basic postulate of the model is that people are constrained in their shopping behaviour. People can then be rated on a scale of constraints, which can be used to predict store selection behaviour.

S. Afr. J. Bus. Mgmt. 1985, 16: 171 – 180

In hierdie artikel word 'n gedragsmodel van plasing van kleinhandel bespreek wat op twee wyses van erkende modelle verskil. In stede van op die voorsiening van kleinhandelsinsette te konsentreer, word met hierdie model op die aanvraag daarvan gefokus. In die model word verder ook van sielkundige en sosiologiese konstruksie gebruik gemaak om winkelseleksie te verklaar. Daar word ook gepoog om die aksie van aankoop te bestudeer in verband met ander menslike interaksies. 'n Basiese aanname by die gebruik van die model is dat mense beperk word in hulle aankoopgedrag. Mense kan dan op 'n skaal van koopbeperking geplaas word, waarmee winkelseleksie dan weer vooruitgeskat kan word.

S.-Afr. Tydskr. Veek. 1985, 16: 171 – 180

Introduction

The theory of retail location has been built on only a few fragile, and sometimes non-related, foundation stones. Its present status is the result of fragmented contributions from various interested groups of people who have different working backgrounds. Such groups have included retailers, planners, geographers, consumers, mathematicians, marketers, economists, and consumer behaviourists. Consequently, there appears to be no commonly accepted basis for an analytical breakdown of the subject which would give it both a comprehensive and a meaningful coverage. Different writers have approached the subject in many varied ways. Some break it down into its theoretical and empirical contributions, some into an analysis of trade areas and shopping patterns, and others into various spatial levels of enquiry. What has tended to happen is that theories, findings, and models have sometimes been wrongly interpreted and applied to situations to which they have no relevance.

We shall concentrate on spatial levels of enquiry in order to develop a workable model of retail location which could be applied to location problems arising within large urban areas. A model applicable to these situations was chosen, firstly, because most retailing activity is currently occurring in and around large urban areas, and secondly, because there is a notable absence of theory which may be applied at this level of analysis.

Existing retail location theory is dominated by models that attempt to assess the attraction potential of existing shopping centres. 'Central Place Theory', first proposed by Christaller (1933) and later by Losch (1940), and the 'Gravity Models' of Reilly (1929) and Converse (1949) have been subject to many refinements in the post-war decades. However, all such supply approaches consider the existing pattern of shopping centres and attempt to assess how retail trade is distributed between them.

It is assumed that the existing centres represent an ideal selection as far as the consumer is concerned. Such an assumption is probably unsound because in practice the consumer must choose an existing centre which best suits his/her conditions and requirements.

What is more, there is currently little or no retail development at the regional level so that much of the successfully applicable regional retail location theory becomes redundant. All the structural activity in retailing in the foreseeable future is likely to be centred in and around large urban areas because in such circumstances both the population size and the distances involved are large enough to enable a

A.R. Yates and K.A. Yeomans*

Graduate School of Business Administration,
University of the Witwatersrand, P.O. Box 31170,
Braamfontein, 2017 Republic of South Africa

*To whom correspondence should be addressed

variety of alternative retailing facilities to develop.

It is interesting to note that at this level of analysis food products become the most significant commodity for attracting people to a particular shopping centre. This fact further exacerbates the redundancy of existing theories because their applications are more suited to shopping/durable goods.

What is needed, is a model of retail location capable of determining the food shopping centres which people will select from a variety of centres in and around a large urban area. The model should utilize constructs developed within the theory of consumer behaviour to focus attention on the demand for, rather than the supply of, retail outlets. Thus, the model developed here adopts a demand approach which focuses attention on households in relationship to existing shopping centres and suggests the important characteristics of household members that influence the selection of retailing centres.

We are concerned with establishing relationships between certain types of shoppers, grouped according to the degree of constraints imposed upon them, and the usage patterns of various types of shopping centres.

In many situations a consumer is confronted not only with the factors that directly influence and give rise to desired behaviour, but he/she is also confronted with factors that act as constraints on behaviour. Behaviour possibilities become a function of desired behaviour adjusted by these constraints. Social class, income, age, family size, and position in the family life cycle simultaneously induce and constrain the behaviour of individuals. However, it is assumed that retail location selection decisions are influenced more by factors that constrain behaviour than by factors that induce behaviour. Based on this assumption a behavioural model of retail location will be constructed which utilizes several demographic variables and incorporates their influence in terms of how they constrain the behaviour of an individual. Constrained behaviour is the single concept that will be used as a basis for segmentation.

A theoretical model of retail location based on constrained behaviour

Time constraints

Shopping for food is only one of a whole range of activities in which people are involved during their weekly routine. It is therefore likely that shopping behaviour patterns will be influenced by some or all of these other activities. This influence will restrict shopping behaviour by imposing time constraints on the shopper. The total time available for shopping during a week will be determined by the total time spent on other activities during the same period. More important, this total available shopping time will comprise several time elements spread over the days of the week. The precise periods when these time elements occur will be determined by certain occasions in the day when the shopper must be engaged in some activity other than shopping. Therefore, for each shopper the complex structure of his/her activities establishes a variety of alternative time elements which may be devoted to the shopping act. In practice the 'work status' of the shopper is likely to be a major source of time constraints.

In addition to the shopper's personal circumstances it is likely that the primary activities of certain other people will impose time constraints on his/her behaviour. Not all people are 'food shoppers', but all are consumers of food. Generally food is purchased on behalf of a household unit by one or more of its members. In this way all household members will

play an important role in establishing the variety of time elements during which the shopping agent may be engaged in food shopping. For example, the employment status of the agent's spouse and the child status of the household are factors which will impose time constraints.

Mobility constraints

An important factor in determining the selection of a retail outlet is the ability of the shopping agent to travel the distance between his/her home base and the shopping centre. This will depend on the speed at which the shopping agent is able to travel and on the time (s)he has available on any one particular occasion. Speed may be interpreted in terms of a measure of mobility (ability to cover distance). Shopping agents will differ in their mobility levels. Furthermore, it is possible for any shopping agent's mobility level to vary considerably throughout the week. Theoretically it would be possible to calculate the distance that a particular shopping agent could travel from his/her home base in each of his/her available time elements. However, the situation is complicated when shopping is combined with other activities which occur outside the home. If an agent combines shopping with other activities the selection of shopping centres will be directly influenced, or even determined, by the physical locations of these activities. It is probably the case that the most important situations would be when the agent combines shopping with work and child-related activities. These are also important conditions for determining time constraints, therefore they are an integral part of the model.

Storage and attitude constraints

For any particular shopping agent the distance boundaries determined by the mobility and time constraints will usually encompass a selection of shopping locations.

Within the model, the factors that will help an agent to make the final selection may be discussed under the general headings of attitude and storage constraints.

Shopping agents will, by their own experience and by peer group influences, develop attitudes towards certain stores and especially towards major retailers. At the level of analysis under investigation, most shopping centres will be dominated by one or two major retailers. If an agent has a favourable attitude towards particular stores then (s)he is likely to seek out the locations where these stores are sited. In practice, the preferred store will be visited if the combined effects of the size of time element and the mobility level during that element provide a sufficient distance boundary. Conversely, unfavourable attitudes toward stores will dissuade agents from visiting certain locations. In this case attitude conditions will constrain behaviour by reducing the range of obtainable shopping opportunities further. It is suggested that a more general attitude construct acting as a constraint to behaviour would be a simple measure of like or dislike of food shopping. Those agents who dislike shopping would possibly be inclined to shop relatively less frequently. This would constrain their behaviour by reducing the number of time elements which they would be prepared to devote to shopping. Also, less frequent shoppers may be persuaded to shop at those locations that would allow most household purchases to be made on one visit. But agents who like shopping would tend to shop more often and seek out a greater variety of shopping opportunities. Such agents would be relatively unconstrained.

Household food storage facilities will act as constraints on behaviour in a similar way to attitude constraints. An

important storage facility which is likely to cause different food shopping behaviour patterns is the possession of a food freezer. Those agents who have inadequate food freezer storage facilities would be forced to shop more frequently than agents with good food freezer storage facilities. In addition, the possession of a food freezer may encourage agents to visit those stores that cater for freezer owners, so that this would in itself determine the location that is chosen.

For each shopping agent attitude and storage constraints would together establish shopping frequency patterns, store preferences, and shopping method preferences (supermarket or personal service). These conditions would have to be balanced against the possible distance boundaries during each time element in order to arrive at a final shopping centre location decision.

The complete model

The representation of the model in Figure 1 is a simple form which may be divided into two stages. The first stage determines distance boundaries around the consumer's home base for every available time element. The maximum distance boundary separates possible shopping opportunities from unobtainable ones. At this first stage only time and mobility constraints have been taken into consideration and in certain cases, when the maximum boundary encompasses only one shopping centre, this is sufficient to determine the selected shopping location. However, in most cases, when the maximum boundary encompasses two or more shopping opportunities, it is necessary to take account of attitude and storage constraints in order to determine which centres are chosen. The selection process that this involves may be designated as the second stage.

In the simple form, the four constraints have been presented

as unrelated variables. In reality it is more likely that the constraints are closely related. For example, an agent's attitudes towards shopping may well stem from his/her storage, mobility, and/or time conditions. Favourable attitudes may well be the result of good storage facilities or good transport availability. However, at this juncture it is not important to establish precisely how these variables are related but only to appreciate that an interrelationship does exist.

Finally, it is necessary to include the demographic segmentation variables in the model as a reminder that the single concept of constrained behaviour, as represented by the four constraint characteristics, is an interpretation of the combined effects of those demographic variables.

The constraints scale

If one or more qualitative characteristics represented each of the four constraints (time, mobility, attitude, storage), then groups of agents could be formed from all the possible combinations of the attributes of each of those characteristics. It would then be the case, according to the theory, that agents in the same group would display similar shopping behaviour patterns. If these groups of agents were ranked according to the extent that they are constrained in their behaviour, a scale would result ranging from highly constrained groups of agents down to unconstrained groups of agents. Such a scale would be referred to as a 'constraints scale'. From Table 1 it can be seen that mobility, work and child status have been described by threefold divisions, while attitude and food storage status have two states each. It is assumed that mobility constraints are most important in determining shopping patterns followed by work, child, attitudinal, and food storage constraints. Thus 108 groups of agents are identified, ranging from the most to the least constrained.

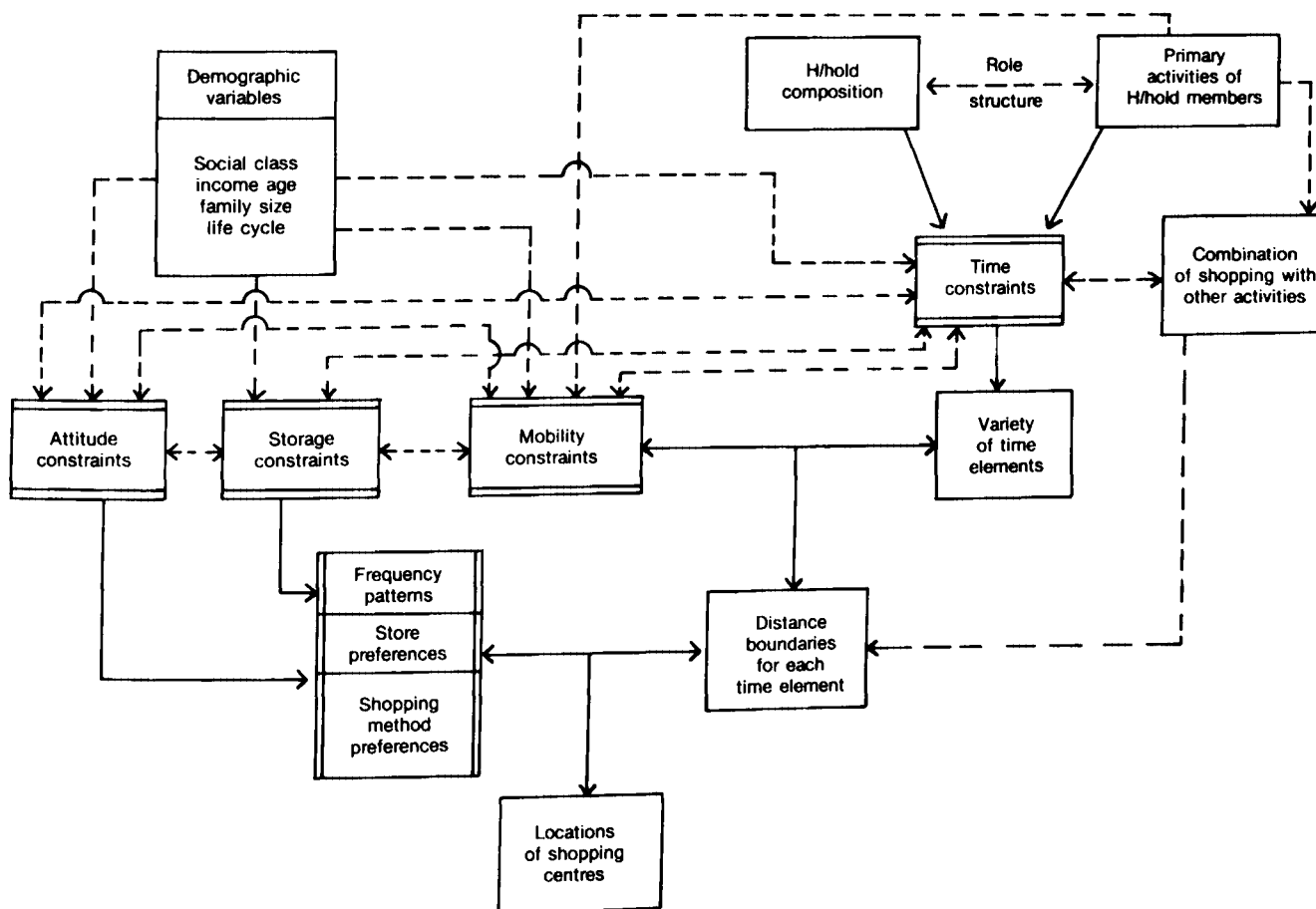


Figure 1 Model of retail location based on constrained behaviour

Table 1 The constraints scale

Use of car	Adults at work	Children	Like/hate shopping	With/without freezer
Most constrained				
1 no car ^a	both ^d	{ under full-time school age }	hate ^g	without
2 no car	both		hate	with
3 no car	both	under age	like	without
4 no car	both	under age	like	with
5 no car	both	at school ^h	hate	without
6 no car	both	at school	hate	with
7 no car	both	at school	like	without
8 no car	both	at school	like	with
9 no car	both	none ⁱ	hate	without
10 no car	both	none	hate	with
11 no car	both	none	like	without
12 no car	both	none	like	with
13 - 24 no car	one work, one not ^e			
25 - 36 no car	none ^f			
37 - 72 restricted use ^b				
Most unconstrained				
73 - 108 unrestricted use ^c				

^aAgent does not have access to a motor car.

^bHousehold possesses a car but it is not always available for the agent's use during shop opening hours. In this case the agent may drive him/herself or rely on another household member to drive him/her.

^cThe agent holds a driving licence and has a car available for his/her personal use at all times during the day.

^dBoth adults in the household worked (no distinction made between full time and part time).

^eOne adult in the household worked outside the home while the other did not.

^fThe adults in the household were unemployed or retired.

^gThe household contained at least one child under full time school age.

^hThe household contained at least one child attending full time school.

ⁱThere were no children in the household.

Hypothesized behaviour

It is now possible to construct hypotheses about the relationships between an agent's shopping behaviour and his/her position on the scale.

- (1) Agents from more unconstrained groups are likely to travel greater distances on home-based trips to purchase food items than agents from more constrained groups.
- (2) There will be a relationship between the extent of constraints imposed on shopping agents and the relative importance (in terms of proportion of money spent) of various types of shopping facilities. In this way, the more constrained agents will place a greater importance on shopping at local stores, whereas the more unconstrained agents would favour suburban centres.
- (3) Highly constrained agents would be more likely to combine shopping with other activities (the other activities are the primary reasons for an agent being out of the home).
- (4) As the constraints imposed on an agent are increased, so the number of food shopping trips that an agent undertakes is likely to increase, the agent purchasing fewer items on each occasion.
- (5) It is likely that for highly constrained agents the stores they use are not the stores they want to use, but unconstrained agents are more able to shop at the stores they prefer.
- (6) Shopping agents from the same constraint group will shop for food items at similar times during the week. The more constrained an agent the more likely that the main food

shopping trip is conducted on Saturdays and/or during evening periods, but the less constrained agents will tend to favour weekday shopping.

- (7) The more constrained an agent the more the constraints will influence the selection of retail outlets directly. Therefore, highly constrained agents will rate constraint characteristics as more important determinants of store selection than any attraction criterion of retail outlets, the reverse will hold true for the more unconstrained agents.

Data collection

In order to test these hypotheses, data were obtained by personal interview from a quota sample of 352 households in a Birmingham suburban area, during the five-week period from mid-June to mid-July 1977.

Information was collected on the characteristics used to represent the four constraints and on the age and social class of shopping agents so that the influences of these additional independent variables on the dependent and constraint variables could be assessed. The dependent variable reflected the extent to which the shopping agent used certain types of shopping centres (measured by frequency of visits and by the proportion of money spent at a particular centre), the extent to which they combined shopping with other activities, the difference in average distance travelled to shopping centres from their home base, the level of satisfaction with the stores which they used regularly and the period in the week during which they conducted their main shopping. Because it was hypothesized that shopping agents at various points on the

constraint scale would patronize different shopping centres to varying extents, the interview area had to be carefully selected to allow residents to exercise a full range of choice. An area bounded by the Kings Heath, Springfield, Hall Green and Billesley districts of Birmingham was deemed appropriate in so far that it was serviced by a wide variety of shopping opportunities which were readily accessible to its residents. There were several small collections of local independent stores in the area; it contained and was surrounded by numerous suburban centres; and it was within easy reach of the city centre. Finally, the interview area was serviced by several bus routes to the city centre, was on the inter-suburban outer-circle route, and included a variety of housing types which it was anticipated would yield a range of social classes of respondents.

Quota controls were based on what were deemed the two most important constraint characteristics, namely mobility and work status. In Table 2 the distribution of respondents within the nine major categories is shown.

Table 2 Ninefold constraint scale

Group		Number of responses
No car	all work	44
No car	one work	40
No car	no work	61
Restricted car	all work	53
Restricted car	one work	45
Restricted car	no work	13
Unrestricted car	all work	44
Unrestricted car	one work	41
Unrestricted car	no work	11
Total responses		352

The design of the questionnaire involved two major preliminary stages. The first of these was concerned with the construction of attitude scales. One scale was required to measure an agent's like or dislike of food shopping so that the attitude constraint could be assessed. Another scale was required to measure an agent's level of satisfaction with the stores (s)he regularly used so that the relationship between the level of satisfaction and the level on the constraints scale could be investigated.

The Likert scaling technique was adopted for both of these attitude measures and involved the assistance of 75 housewives. The results of the internal consistency checks on the attitude statements used were not decisive. This indicated a lack of internal consistency possibly resulting from the fact that certain statements reflected attitude dimensions different from that under investigation. Nevertheless, the scales were used because it was felt that they would still be capable of distinguishing between relatively large differences in attitudes even though marginal differences would be undetected. The criterion used for eliminating statements from the like/dislike scale was a correlation coefficient value of less than 0,4 (significant at the 0,1% level) for the 'satisfaction' scale the corresponding figure was 0,37 (just significant at the 0,1% level).

The second major preliminary state in designing the final questionnaire was to undertake a series of initial pilot surveys. Several questionnaires were drawn up, tested, and refined before an acceptable version was obtained. A final pilot survey

was conducted within the actual survey area by interviewing between 30 and 40 respondents.

Preliminary tests of the simple constraints scale

The simple nine-group constraints scale, based on only the mobility and work status variables, was cross-tabulated against each of the dependent variables in turn. However, this exercise did not uncover any satisfactory relationships. When the constraints scale was broken down into its component variables and the dependent variables compared with each part separately, certain patterns emerged. However, on inspection some of these patterns did not agree with the constrained behaviour hypotheses.

Nevertheless, these preliminary tests of the constraints scale were encouraging to the extent that the variables used to construct the scale individually displayed substantive relationships with the dependent variables. But, as no consistent relationships were indicated, doubt was placed on the simplifying assumptions of relative importance of the constraint variables which were used to form the simple constraints scale.

Consequently analyses were undertaken firstly to see whether other potential independent variables were better predictors of the shopping behaviour dependent variables and secondly to examine the nature and extent of relationships between these and the constraint scale variables.

The results of these investigations indicated that some of the independent variables not used in the construction of the nine-group constraints scale were important determinants of variation in the dependent variables. Because of their uneven distributions throughout the nine-group scale they were a source of error when assessing the predictive capabilities of the scale; as such they could be referred to as possible confounding variables. Also there were suggestions that all the independent variables were directly related to one or more of the dependent variables. In fact, insufficient cell sizes in cross-tabulation and breakdown tables of more than two levels resulted in great difficulty in establishing which of the independent variables were the important determinants of behaviour. This problem was aggravated further by the complex interrelationships which existed between the independent variables. It was, therefore, necessary to adopt a new approach capable of assessing the ability of the constraint variables to predict behaviour.

Principal component analysis

Analysis had shown that the range of independent variables available from the study were highly interrelated. By using principal component analysis it was possible to uncover any underlying patterns in these relationships. Such patterns could be used to rearrange the information contained in the independent variables so as to create a smaller set of new variables, called components or factors, which would replace the original variables. Those original variables which were highly related would be merged to form one new variable. Such a transformation of the original variables into a set of new variables (factors) should reveal important relationships which had proved difficult to uncover among the variables in their original form.

In principal component analysis each new variable (factor) is a linear combination of all the original variables. The first factor is the single best summary of linear relationships exhibited in the data; it accounts for more of the total variance of the original variables than any other linear equation could.

The second factor is the second best linear combination

of the original variables, accounting for more of the remaining variance than any other linear equation and is orthogonal to the first. Subsequent factors are defined similarly until all the variance in the data is exhausted.

For the actual principal component analysis the complete set of independent variables was used (Appendix 1). Variables were required to be at least ordinal in their level of measurement. If an ordering of the attributes of each variable was interpreted in terms of levels of constrained behaviour then the level of measurement condition was satisfied.

The first step in this analysis was to calculate product moment correlation coefficients between all the independent variables. Initially there are as many factors as there are original variables (16 in this case). But many of the factors which are constructed may be deemed insignificant. The significance of a factor is judged by the amount of the total variance in the data which is accounted for by the particular factor; this amount is given by the eigenvalue of the factor. Because all the original variables are standardized, the variance of each is equal to unity. Therefore when the eigenvalue of a factor is less than unity, the factor is considered insignificant because the amount of total variance accounted for by such

a factor is less than an amount accounted for by any one of the original variables.

The respective eigenvalues for the 16 factors that were constructed from the data, are shown in Table 3.

Examination of the eigenvalues reveals that only the first six factors are significant, the seventh having a value less than unity. These six factors accounted for 77,2% of the total variance in the data and, as such, provided a good alternative and compact representation of the original 16 variables.

The final step in the principal component analysis was to rotate the six factors to a terminal solution. This solution simplified the factor structure by finding the position where a factor loaded more heavily on a few variables rather than loading more evenly on all variables, thus making each factor easier to interpret.

Table 4 contains simple correlation coefficients between the six factors in the terminal solution and each of the original variables.

By taking one row of the table at a time, it was possible to see which factor had the best correlation with each original variable. In this way it was discovered that certain variables were grouping together under certain factors. Where groupings occurred, it meant that the factor would provide an adequate explanation of the variance in those original variables in the group. Therefore Factor 1 was able to replace the following variables: the size of household, the presence of children under full-time school age, the presence of children at school, and the presence of children at work. Factor 1 was interpreted as a 'size/family structure' variable which emphasized the presence of children. Factor 2 was able to replace the following variables: the shopping agent's job status and the work status of the adults. This factor was interpreted as a 'work status' variable. Factor 3 was capable of replacing the involvement of other members of the household in shopping and the presence of others in the household (others being persons other than a spouse or children). This factor was difficult to interpret and name. Factor 4 was able to replace the two attitude scales and was itself interpreted as an attitude construct. Factor 5 replaced the age of the respondent, the presence of an agent's spouse, and the job status of the spouse. It was interpreted as an 'age/family structure' variable. Factor 6 was interpreted as an 'affluence' variable by replacing freezer ownership, social class and mobility level.

Table 3 Principal components analysis

Factor	Eigenvalue	Per cent of variance	Per cent cumulative
1	4,685	29,3	29,3
2	2,249	14,1	43,3
3	1,712	10,7	54,0
4	1,394	8,7	62,7
5	1,266	7,9	70,7
6	1,051	6,6	77,2
7	0,800	5,0	82,2
8	0,638	4,0	86,2
9	0,577	3,6	89,8
10	0,387	2,4	92,2
11	0,354	2,2	94,5
12	0,263	1,6	96,1
13	0,226	1,4	97,5
14	0,180	1,1	98,6
15	0,145	0,9	99,5
16	0,073	0,5	100,0

Table 4 Correlation with principal components

Independent variables	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
Size of household	-0,840	0,120	0,377	0,161	0,341	-0,032
Spouse in household	-0,301	0,155	0,582	0,087	0,711	-0,262
Others in household	-0,292	0,059	0,854	0,051	0,273	-0,222
Children under full-time school age	-0,887	0,048	0,084	0,135	0,299	-0,084
Children at school	-0,915	0,187	0,145	0,184	0,269	-0,087
Children working	-0,825	0,198	0,220	-0,031	-0,102	-0,085
Involvement of others	-0,089	0,082	0,851	0,145	0,037	-0,042
Freezer owner	-0,151	0,158	0,116	0,098	-0,009	-0,789
Age	0,449	-0,285	0,015	-0,455	-0,578	0,100
Social class	-0,170	0,075	0,015	-0,143	-0,469	0,605
Score like	0,131	-0,296	-0,115	-0,896	-0,182	0,165
Score satis.	0,066	-0,108	-0,075	-0,912	-0,054	0,094
Spouse job status	-0,359	0,327	0,265	0,155	0,854	-0,264
Shopping agent job status	-0,014	0,934	0,017	0,167	0,003	-0,118
Work status of adults	-0,253	0,911	0,122	0,281	0,362	-0,222
Mobility of shopping agent	0,102	-0,211	-0,163	-0,149	-0,279	0,769

Scores for each factor were obtained by summing the results of the multiplication between the values of the original variables, in standardized form, and their respective regression weights. Only those variables that were highly correlated with a particular factor were used to compute the scores for the factor.

To give some meaning to a high or a low score for each new factor variable it was necessary to relate the coding values of the original variables and the new variables (obtained from the correlation coefficients). This exercise revealed the following meanings (shown in Table 5).

Table 5 Factor definitions

Factor 1 Size family structure variable		
High score	relatively unconstrained	small household, no children under full-time school age, at school, or at work
Low score	relatively constrained	large household, with children under full-time school age, and/or at school, and/or at work
Factor 2 Work status variables		
High score	relatively constrained	shopping agent having a full-time job, both adult members working
Low score	relatively unconstrained	shopping agent no job, no adults working
Factor 3 Involvement variable		
High score	relatively constrained	presence of others, more involvement of others
Low score	relatively unconstrained	no other members, no involvement of others
Factor 4 Attitude variable		
High score	relatively constrained	relative dislike and dissatisfaction
Low score	relatively unconstrained	relative liking and satisfaction
Factor 5 Age family structure variable		
High score	relatively constrained	spouse has full-time job, spouse is present, younger shopping agents
Low score	relatively unconstrained	spouse has no job, no spouse, older shopping agents
Factor 6 Affluence variable		
High score	relatively constrained	no car, lower social class, no freezer
Low score	relatively unconstrained	unrestricted car use, higher social class, large freezer

Cluster analysis

Using the factor scores as input data it was decided to attempt the construction of a shopping behaviour prediction classification employing clustering.

Cluster analysis would focus attention on the individual shopping agents to see if they could be divided into groups (clusters) each containing agents who were relatively similar to one another on the basis of the information in the set of independent variables. The basic principle of this technique is to create groups, in this case groups of agents, so that within-group variance is minimized and between-group variance is maximized. However, many sets of groups (clusters) may be established according to this basic principle and if only one set is required then its selection will depend upon a certain amount of subjective reasoning.

The computer program that was used (Golder & Yeomans, 1973) took as its input the scores from the six variables created

by the principal component analysis. The output contained more than 20 different cluster arrangements. Only one set of groups was required and it was selected in two stages. First, a Beale test (Beale, 1969) was applied to all the cluster arrangements. This test was designed to indicate, by use of an *F* test, when a set of *n* groups produced a significantly better clustering arrangement (better in terms of greater between-group variance *F* test) than a set of *n* - 1 groups. But this test alone was not capable of selecting the best set of groups, it only suggested a range of possibilities. The final decision on which set of clusters to adopt was based on a complete breakdown of all the original variables by each cluster group arrangement. This enables the most meaningful set of groups to be selected, which was the set containing five groups.

Each of the five cluster groups contained shopping agents possessing a mixture of attributes of any one independent variable, but each group was represented by a predominance of certain attributes. Therefore, members of a group could be stereotyped.

The first cluster group may be portrayed by households containing just two adult members, aged between 46 and 60 years, both working, coming from a 'skilled' social class, not possessing a freezer, the shopping agent having restricted motor car usage and indicating a relative dislike of food shopping.

The second cluster group is represented by households with three or four persons, comprising two adults within the age range 26 - 45, one working and one not working, and children at school and/or under full-time school age, coming from the intermediate' and/or professional social class, possessing a freezer, the shopping agent having unrestricted use of a motor car and indicating a relative dislike of food shopping.

The third cluster group shows a predominance of households containing just two adult members, aged over 60 years, neither working, coming from a 'skilled' social class, not possessing a freezer or a motor car, and the agent indicating a relative liking of food shopping.

The fourth cluster group is depicted by large households (four or five members) comprising two adults within the age range 26 - 45, one working and one not working, and children at school and/or under full-time school age, coming from the 'skilled' social class, not possessing a freezer or a motor car, and the agent indicating a relative dislike of food shopping.

The fifth cluster group may be portrayed by households containing one adult over 60 years old, not working, coming from the 'skilled' social class, not possessing a freezer or a motor car, and indicating a relative liking of food shopping.

The value of this clustering exercise was to be judged by the ability of the cluster groups to distinguish between certain behaviour patterns. To this end, mean values of all the dependent variables were calculated for each cluster group. The results were very encouraging and are reproduced in Table 6.

By inspecting the values of the proportion of money spent at various centres and the number of visits to various centres, considerable variations between the cluster groups were revealed. To confirm this statistically an *F* test was undertaken on each dependent variable to see if there was a significant collective difference between the mean values of the five cluster groups. All tests were significant at a 1% level except for the 'number of visits to suburban centres' variable. In addition, the five cluster groups were cross-tabulated against the period when the main shopping trip was conducted, this also showed wide variation (see Table 7).

Several observations may be made about the variety of

Table 6 Means of five clusters of households

Cluster group	Proportion of total weekly food expenditure at various centres			Proportion of money spent on combined trips per week %	Average weekly food expenditure per household £	Number of visits per week					(N) Size of cluster groups
	City centre %	Suburban centres %	Local stores %			To City centre	To suburban centres	To local stores	On all homebased trips	On all combined trips	
Cluster 1	5,4	54,2	8,4	32,0	16,99	0,125	1,18	0,659	1,967	2,22	82
Cluster 2	2,9	70,0	13,2	13,8	18,58	0,95	1,46	1,079	2,630	0,737	76
Cluster 3	18,5	54,8	24,0	2,7	12,43	0,567	1,53	1,825	3,922	0,078	77
Cluster 4	9,8	56,6	17,0	16,7	21,61	0,247	1,51	2,444	4,256	1,607	89
Cluster 5	14,7	43,3	33,1	8,9	6,92	0,321	1,02	1,643	2,982	0,393	28
<i>F</i> test of a difference between means	6,132	3,83	7,112	12,36	34,66	7,982	1,77	8,64	15,20	19,36	

not significant

Table 7 Period of shopping

Cluster group	Period of main trip, percentages			
	Daytime in week	Evening in week	Saturday	All periods
1	45	20	35	100
2	68	16	16	100
3	84	0	16	100
4	70	8	22	100
5	79	0	21	100

behaviour patterns between the cluster groups.

Cluster group 1 (two adults working, aged 46–60 years, skilled, no freezer, restricted use of car, dislike of shopping).

It was very important for agents in this group to combine shopping with other activities for on average one-third of their total food expenditure was undertaken in such a way (largest proportion of all groups). Combined shopping was conducted very frequently but on each occasion relatively fewer items were purchased. On home-based trips, neither the city centre nor local centres were particularly important for this group. The majority of shopping was conducted at suburban centres. Main shopping trips on Saturdays and during the evenings were particularly important.

Cluster group 2 (two adults, one working one not, aged 26–45 years, one or two children at school or under full-time school age, intermediate or professional, with freezer, unrestricted use of car, dislike of shopping).

In total, agents in this group visited the shops on fewer occasions than agents from any other group. On home-based trips, food shopping was concentrated at suburban centres and was very infrequent; indicating bulk purchasing. The city centre and local stores were visited almost as often as suburban centres but comparatively little money was spent in either centre.

Cluster group 3 (two adults not working, aged over 60, skilled, no freezer, no car, enjoy shopping).

Agents in this group used all types of shopping centres but concentrated their expenditure at the suburban centres. Visits to local stores were very frequent but the proportion of money spent there was relatively small. Combined shopping was of little or no importance to these agents and collectively they displayed the greatest dependence on weekday shopping.

Cluster group 4 (two adults, one working one not, 26–45 years of age, two or three children at school and/or under full-time school age, skilled, no freezer, no car, dislike

shopping).

Agents in this group visited the shops at least 50% more often than any other agents, averaging six trips per week. This group also produced the highest ratio of visits to expenditure for local store usage (agents visited local stores very frequently but the money spent per visit was the smallest proportion of all agents). As for all other agents, suburban centres took the largest proportion of the agent's weekly food expenditure. Finally, combined shopping was a comparatively important method of food shopping for this group.

Cluster group 5 (one adult, not working, aged over 60, skilled, no freezer, no car, enjoys shopping).

For these agents the local stores were extremely important, both in terms of the number of visits and the proportion of money spent. This local store usage severely challenged the supremacy of the suburban centre usage. Most main shopping was undertaken on weekdays, some on Saturdays, but never during the evenings.

Discriminant analysis

This technique was used purely for its classification ability. With each and every shopping agent allocated to one of the cluster groups, a set of classification functions were derived which would permit the allocation of new cases with unknown memberships to a cluster. The functions were weighted linear combinations of the original independent variables constructed in such a way as to maximize the separation of the five groups.

Initially all 16 the original variables were included as discriminating variables. Also, the known membership of all the cases (352) was used to construct the five discriminating functions. As a check of the adequacy of these functions the original cases were reallocated to see how many were correctly classified. This was achieved by calculating five values for each case, one from each discriminant function. If the value obtained from the first discriminant function was greater than the values obtained from the other four functions, then the case was assigned to cluster group 1. If the value obtained from the second discriminant function was greater than the values obtained from the other four functions, then the case was assigned to cluster group 2, etc. The results of this check of adequacy were: 96,2% of the cases allocated to cluster group 1 were correctly assigned; 95,7% were correctly assigned to cluster group 2; 92,8% were correctly assigned to cluster group 3; 83,9% were correctly assigned to cluster group 4; and 96,6% were correctly assigned to cluster group 5.

It was felt that inclusion of all 16 independent variables as discriminating variables was unnecessary and that adequate

discrimination could have been obtained from a smaller set of variables.

Thus a stepwise inclusion algorithm was employed. The stepwise procedure begins by selecting the single best discriminating variable. A second variable is selected as the variable best able to improve the discrimination with the first variable, and so on. The first ten variables were selected on the basis of the sudden drop in the *F* value (level of significance for inclusion) from the tenth to the eleventh variable. The ten best variables were, in order of discriminating capability: the number of persons in the household, the presence of a spouse in the household, the presence of members other than a spouse or children in the household, the presence of children under full-time school age in the household, the freezer status, the age of the shopping agent, the social class of the household, the spouse's job status, the shopping agent's job status, and the work status of the adult members of the household.

The known memberships of all 352 cases were used to compute the five discriminant functions. The check of adequacy in this case produced almost as high a successful prediction rate as when all 16 the independent variables were used. Eighty-eight per cent were correctly assigned to cluster group 1; 90% to cluster group 2; 92,3% to cluster group 3; 83,9% to cluster group 4; and 100% to cluster group 5.

In order to test the stability of the discrimination it was decided to run an independent check on the procedure for allocating individuals to groups.

The check involved taking a random sample of shopping agents from the original sample of 352. Half of the agents were selected in this way. Using the known group memberships of these 176 agents, five discriminant functions were computed from all 16 the original variables. Group memberships of the remaining half of the original sample were then predicted using these functions. 81,8% of the remaining individuals were correctly assigned to cluster group 1; 76,9% to cluster group 2; 70,6% to cluster group 3; 84,6% to cluster group 4; and 100% to cluster group 5; an impressive success rate.

Again a smaller set of discriminating variables were used to create discriminant functions: presence of a spouse in the household, shopping agent's job status, presence of children at school, spouse job status, the age of the shopping agent, freezer ownership, and the presence of others in the household. These seven independent variables were again selected on the basis of the drastic change in the *F* value. In this situation the prediction rate was reduced very slightly: 82,4% were correctly assigned to cluster group 1; 68,4% to cluster group 2; 76,1% to cluster group 3; 75,0% to cluster group 4; and 100% to cluster group 5.

With the known differences in behaviour patterns of the members of the five separate cluster groups it would theoretically be possible to predict the behaviour of any new shopping agent. All that would be required is for the agent to be assigned to one of the cluster groups by applying a set of discriminant functions. The most appropriate functions for this purpose would be those derived from the known memberships of all the cases in the original sample which contain the maximum amount of collected information. Information would be required on only 10 independent variables because the discriminating ability of the complete set of 16 variables showed only a marginal improvement over the best 10.

Conclusions

The objective of this research was to establish a workable

model of retail location which could be applied by both retailers and planners to problems concerned with the site selection of retail outlets within a large urban area. It was envisaged that the ideal shopping facilities for a particular household could be determined by utilizing a 'constraints scale'. If a shopping agent's position on the constraints scale can be established, then his/her shopping behaviour could have been predicted by referring to the known behaviour patterns of persons who were similarly constrained. In the UK, residential areas tend to contain households possessing similar constraint characteristics. For example, there are estates containing predominantly old age pensioners or predominantly young married couples with children, and so on. Therefore it would have been a relatively simple process to aggregate the behaviour patterns of shopping agents living in a particular area so that total demand for local, suburban, and/or city centre shopping could have been estimated. However, the constraints scale proved to be an inadequate analytical tool for households within the survey area. The variables that were interpreted as constraint characteristics, although individually displaying significant relationships with some dependent variables, were incapable of being combined under the unifying concept of constrained behaviour.

The theoretical model perhaps overemphasized the importance of factors that constrain behaviour by underestimating the influence of factors that induce behaviour, thereby rendering the constraint scale an inadequate interpretation of the theoretical model. For the majority of shopping agents living within the survey area it is very likely that if distance boundaries were constructed for each agent, most boundaries would encompass a wide variety of shopping centres. This would arise because the survey area was provided with many types of shopping facilities, easily accessible to all of the resident shopping agents. Even a one-hour time element would allow a comparatively constrained agent, one restricted to the bus service, to choose between many local facilities, several suburban centres, and even possibly the city centre itself. In such a situation inducement factors become very important determinants of retail outlet selection. The theoretical model incorporated such factors into the attitude constraint by suggesting that store preferences and shopping method preferences would induce agents to visit certain centres. This would constrain their behaviour by restricting the possible number of centres which an agent could visit. However, the importance of these factors was overlooked when constructing the constraint scale. Attitude constraints were represented only by a measure of an agent's like or dislike of food shopping. On reflection it is ironic that the survey area was chosen so that agents would be seen to exercise a free choice between a wide variety of centres but this fact in part possibly led to the downfall of the proposed constraint scale. Nevertheless, where retail facilities are more sparsely positioned, as in residential areas located just outside city boundaries, a simple additive form of the constraints scale may prove to be more suitable. This is an area for future research.

While investigating the properties of the constraint scale, relationships between retail selection behaviour and certain groups of shopping agents have been suggested. These groups were not formed according to constraint considerations but were formed on the basis of similarities existing in a set of independent variables measured for each shopping agent. The cluster analysis procedure discussed above produced five distinct groups of agents. The groups displayed significantly different expenditure and shopping frequency patterns for city

centre, suburban centre, and local store usage. The known behaviour patterns that were established by the sample data could be used to predict the behaviour of other shopping agents living in analogous situations. The profiles of the members of each of the five groups could be used as a guide to assign any new agent to a group. This approach constitutes a complex model of retail selection behaviour that could be applied only to location problems within the suburbs of Birmingham. For application to other cities a new set of discriminant functions would be required because certain peculiarities existing in Birmingham may not prevail in other cities.

On balance, the research has contributed in many ways to retail location theory. The theoretical model, with a little more emphasis on inducement factors, could provide a framework within which future studies of retail location may be based. Also, the novel way that cluster analysis and discriminant analysis were combined to establish an alternative approach to the constraints scale could be repeated in other studies of retail location. Indeed, this approach could even be used to develop new bases for segmenting markets in studies other than retail location.

Finally, it would be advisable to retest an additive scale of constraints at a more relevant level of analysis for it may still prove to be an easy to use analytical tool for planners and developers.

References

- Beale, E. 1969. Euclidean Cluster Analysis, contributed paper to 37th Session of the International Statistical Institute.
- Converse, P. 1949. New Laws of Retail Gravitation. *J. Mark.*, October.
- Cristaller, W. 1933. *Die Zentralen Orte in Süddeutschland*.
- Golder, P. & Yeomans, K.A. 1973. The Use of Cluster Analysis for Stratification, *J. Royal Stat. Soc.*, Series C. vol.22, No.2.
- Losch, A. 1940. *Die raumliche Ordnung der Wirtschaft*.
- Reilly, W. 1929. *Methods of Study of Retail Relationships*. Austin, University of Texas Press.

Appendix 1 List of independent variables and coding scheme

Variable	Coding scheme
Size of household	continuous variable
Presence of a spouse in the household	(2) yes; (1) no; (0) not applicable (n.a.)
Presence of person(s) other than spouse or children	(2) yes; (1) no; (0) n.a.
Presence of children under full-time school age	(2) yes; (1) no; (0) n.a.
Presence of children at school	(2) yes; (1) no; (0) n.a.
Presence of children working	(2) yes; (1) no; (0) n.a.
Involvement of others in shopping	(5) others; (4) spouse, son and daughter; (3) son and daughter; (2) spouse only; (1) none; (0) n.a.
Freezer ownership	(2) large freezer; (1) small freezer; (0) no freezer
Age of shopping agent	(4) over 60; (3) 60-46; (2) 45-36; (1) 35-26; (0) under 25
Social class	(1) professional; (2) intermediate; (3) skilled; (4) part-skilled; (5) unskilled
Score on like/dislike shopping	continuous: high score = relative liking for shopping
Score on satisfaction scale ^a	continuous: high score = relative satisfaction
Spouse job status	(3) full-time job; (2) part-time job; (1) no job; (0) n.a.
Shopping agent's job status	(2) full-time job; (1) part-time job; (0) no job
Work status of adults of household	(3) all adults work; (2) one adult works, one does not; (1) neither work
Mobility of shopping agent	(3) no car; (2) restricted car; (1) unrestricted car

^aBecause of the high correlation with the score on the 'like' scale, it was decided to use the 'satisfaction' scale as an independent variable to support the 'like' scale.