84
10 | 04

THE RELATIVE POWER OF GEODEMOGRAPHICS VIS A VIS PERSON AND HOUSEHOLD LEVEL DEMOGRAPHIC VARIABLES AS DISCRIMINATORS OF CONSUMER BEHAVIOUR

Richard Webber

The relative power of geodemographics vis a vis person and household level demographic variables as discriminators of consumer behaviour

Richard Webber

Visiting Professor, Centre for Advanced Spatial Analysis, UCL

Abstract

Geodemographics is a field of study which involves the classification of consumers according to the type of neighbourhood in which they live. As a method of segmenting consumers it has long been of value to direct marketers who, being often unable to identify the age, marital status or occupational status of people in mailing lists, found it a useful means of applying selectivity to their mail shots. By analysing the behavioural characteristics of consumers in different types of neighbourhoods they found they could improve business performance by targeting promotional activities to names and addresses falling within specific types of postcode. From direct marketing the application of geodemographics spread to the targeting of door to door distribution and customer communications and to the retail industry where it was found to be useful input into the process of deciding where to site new outlets. Government is increasingly using such methods to improve the targeting of its own communications to tailor local service delivery to the particular needs of local communities.

During the 25 years since geodemographics was first introduced few users have had a clear understanding of precisely neighbourhood differences come about. Are differences in consumption patterns at neighbourhoods level simply the predictable result of differences in the age, household composition, educational status or occupational profile of their residents? Or do additional, incremental neighbourhood effects operate? When deciding neighbourhoods to live in do people select ones whose values and consumer preferences are broadly similar to their own? Or is it only after they have moved that their behaviours change, as they become subject, consciously or not, to the prevailing ethos of the new community in which they find themselves?

To set these alternative explanations this study analyses a random set of consumer behaviours covered by the Target Group Index, one of a number of market research surveys whose respondents have been coded by the type of neighbourhood in which they live; it uses a statistic to measure the extent to which the Mosaic geodemographic system is effective in discriminating on these behaviours; it then measures the relative effectiveness of other frequently used household and person level demographics in predicting of these behaviours; finally it compares the predictive efficiency of different discriminators.

The conclusion that can be drawn from the exercise is that, across these behaviours as whole, the type of neighbourhood in which a consumer lives is a significantly more predictive piece of information that any person or household level discriminator (such as age or social grade). By implication therefore it is almost certain that significant neighbourhood effects must operate for many of the behaviours tested. However the relative discriminatory power of geodemographics and person and household level discriminators varies considerably from behaviour to behaviour. Even when taking measures of status which one might have expected to be highly correlated, such as social grade, terminal education age or household income, there are considerable differences in their relative predictiveness across most consumer behaviours.

Background

The idea of classifying small areas throughout Great Britain into a limited number of 'residential neighbourhood' types originated in 1977 when the Office of Population Censuses and Survey (OPCS) commissioned the Centre for Environmental Studies (CES) to create a typology of wards and parishes based on the 'Small Area Statistics' tables of the 1971 census published by OPCS[1]. This was followed in 1978 by the commissioning of a separate classification system for census enumeration districts[2]. These two projects originated as a result of a studies previously undertaken by CES, one on behalf of the Department of Environment which sought to differentiate neighbourhoods in according to differences in their manifestations of social deprivation and hence in terms of their appropriateness for different policy interventions[3] the others classifications commissioned by OPCS, one of Local Authority Districts[4], the other of Parliamentary Constituencies[5].

In 1979 Ken Baker, Chief Statistician of the Target Group Index (TGI), an annual market research survey of some 40,000 respondents undertaken by the British Market Research Bureau (BMRB), obtained from CES the list of residential neighbourhood categories for each GB ward. Baker then appended the relevant classification code to each TGI survey respondent. Although the original purpose of this exercise was to investigate the representativeness of the TGI sample, Baker thought it might be interesting to test whether type of neighbourhood was an effective predictor of the consumption of a number of products. These tabulations consistently demonstrated that products which it had been assumed by advertisers should be broadly aimed at 'ABC1' audiences , such as the Sunday Observer or the Daily Telegraph, were much more commonly purchased in certain types of up market neighbourhood than in others.

The extent to which the ward typology could differentiate neighbourhoods according to consumption was demonstrated in a paper presented by Baker and colleagues from BMRB at the 1979 annual conference of the Market Research Society[6]. This paper proved seminal in introducing the concept of geodemographics to marketers and media specialists, many of whom relied on the TGI for analysing purchasers of products and media by personal and household demographics such as age, income, social class or terminal education age.

In 1980 the information services company CACI, which was the first organisation to establish an agency agreement with OPCS allowing it to redistribute small area statistics under licence, took possession of the 'CRN' classification as it was then called, CES having by that time been disbanded by the incoming Conservative administration. CACI enhanced the value of geodemographics as a targeting tool first by attributing the CRN typology at the finer level of census enumeration district and then by building a correspondence table between census enumeration district and postcode geographies, postcodes at that time being introduced by the Post Office. As a result of these developments marketers, and in particular direct marketers, were now able to direct their targeting at a much finer level of resolution, the country now being divided into 120,000 rather than 16,000 pieces. The Target Group Index obliged by re-coding its respondents according to the characteristics of the enumeration district rather than the ward in which their postcodes fell.

Mass and Direct Communications

Prior to 1980 owners of most large brands would typically invest their promotional budget in mass media channels, such as newspapers and commercial television, to reach existing and potential new customers. The Target Group Index played a very important role in the selection of these media. By questioning survey respondents on their media consumption as well as on the products and brands they purchased, the TGI provided brand managers and their advertising agencies with the means of qualifying different media according to how well they delivered access to the target markets they wanted to reach.

Nevertheless many marketers would also rely on the TGI database to explore the quite fine distinctions between the markets for their own brands and those of competitor brands. It was in this brand profiling work that conventional demographics played a crucial role. When agencies first needed to explore the subtle nuances of the appeal of different brands within a product category such as biscuits or breakfast cereals, most would commission analyses of competitor brands by measures such as age, sex, household composition, income, social grade and educational attainment. Many still do.

Many professional advertisers therefore regarded geodemographics as a poor man's demographics. It was perceived as being of limited use in 'above the line' advertising where for the selection of mainstream broadcast media one could rely on the TGI for direct tabulations between readership of particular titles and each of the well established demographics such as age and social grade. By contrast it was thought to be of more use in 'below the line' media applications, such as direct mail and leaflet distribution which, without some form of geodemographic classification, were otherwise quite untargettable. However, although applying geodemographics was better than no selectivity at all, many marketeers not unreasonably supposed that these media channels would have been far more targettable if one had been able to select/reject prospects on the basis of individual demographic variables, such as age and social class, rather than on the basis on the 'average' age or social class of the entire census enumeration district in which they lived.

Such aspirations proved less fanciful than one might have imagined at the time with the introduction to the UK in 1984 of what came to be known as 'lifestyle' surveys. Pioneered by US marketing information businesses such as CMT and NDL, these surveys differed from conventional market research surveys in that there was no implication of respondent anonymity or of sample representativeness. Indeed by virtue of intelligently designed incentives significant proportions of the population could be persuaded to fill in extensive questionnaires covering a wide variety of the products and brands respondents purchased as well as their personal demographics. These questionnaires, once captured and stored, created a very valuable source of qualified prospects. If Saga wanted to mail tens of thousands of people on their fiftieth birthday, then they could rent these names from CMT and NDL using person level demographics rather than geodemographics as selection criteria.

Those who were closest to geodemographics however began to recognise that whilst neighbourhoods with above average proportions of old people would tend to have higher than average propensities to purchase products, such as P & O cruises or the Daily Express, generally favoured by older people, the level of this difference, the degree to which such areas exceeded the national average was, in most instances, much greater than could have been expected simply on the basis of the proportions of old people they contained. Similar patterns were evident in neighbourhoods with above average proportions of very wealthy people, on indeed of very poor people.

The frequency with which such pattern were demonstrated suggested a rather different view of neighbourhoods : that neighbourhoods were more than the statistical average of their individual members and that (contrary to the quoted belief of the Prime Minister of the time) the social context in which consumers found themselves did have a significant effect on their consumption patterns as well as their attitudes, values and indeed voting intentions. These 'neighbourhood effects', as they were subsequently to be described by geographers, became of increasing interest in branches of the social sciences other than business studies and marketing which used 'multi-level modelling' techniques to quantify their extent[7].

Much of the discussion of these neighbourhood effects was framed within the broader discussion of the concept of 'social capital'[8] and, within government, of the need to facilitate the building (or rebuilding) of community networks as a prerequisite for addressing many of the manifestations of inner city deprivation[9]. Thus the concept of what constituted a neighbourhood, how it was defined in terms of physical as well as social boundaries, became an increasingly important focus for academic research.

With time the notion that residents or consumers could be described as sharing a common neighbourhood, with a set boundary distinguishing it from other neighbourhoods, gave way to a more realistic recognition that residents are simultaneously located within a network of neighbourhoods of different sizes and scales. Residents may engage at one level by participating in a management committee of a block of flats whose freehold they share; local planning issues may cause them to participate in amenity society covering their block of flats and other houses in their street; their street and adjacent streets may

belong to a common 'neighbourhood watch' area; the parents they meet at the school gates may draw them into a social network which is even more geographically extensive in scale; wider still may be the area from which the church they attend draws its congregation or the golf club they play at draws its membership; the catchment area of the nearest retail park or the territory covered by the closest branch of some professional or arts association will be even larger. Each of these different scales delivers social contact with people who are more or less similar to themselves and are likely to have some effect, however marginal, on their interpretation of the world and as a result on their tastes and values.

At the very finest level of geography the architectural style of the houses in the street will attract people with some tastes and values more than others. Whether the houses are built to a modern or traditional style, whether retaining authentic features or whether utilitarian in character, these factors contribute to a process of self selection of residents with common attitudes as well as at similar stages in the family lifecycle. At the coarser level, the option as to whether a family chooses to buy in Kew or Richmond rather than in Esher or Weybridge shares similarities to decisions involving the selection between between brands of cars or of clothes. The complex set of aspirations or values associated with suburbs can easily be viewed as similar to the brand values which are associated with other products and which attract and repel accordingly.

The relevance to geodemographics of these concepts of 'neighbourhoods' as brands is that they can operate at a geographic scale which is much more extensive than that which has traditionally driven the construction of geodemographic classifications. Most geodemographic classifications, in the UK at least, are based on statistical information gathered at the full postcode level, which averages 17 addresses, as well as from the 120,000 or more enumeration districts (now output areas) for which census small area statistics are now published.

Thus we arrive at a paradox. Viewed as a 'poor man's demographics', it would appear that the way to enhance the discriminatory power of geodemographics is by building classifications at ever finer levels of granularity. If, on the other hand, geodemographics is seen as reflecting the social influences on individuals of the communities in which they live, discrimination may be improved, counter-intuitively, by building classifications which use statistics for quite coarse broad scale geographical areas as well as for finely detailed postcodes.

These considerations led Experian to undertake a series of 'multi level' modelling experiments during recent builds of the 2003 Australian Mosaic classification[10], of the 2003 UK Mosaic classification[11] and of the 2004 Netherlands Mosaic classification. These experiments were designed to test the benefits of incorporating into the set of data used to build these classifications data pertaining to the coarser area around each statistical unit as well as pertaining directly to it. The objective was to see whether, by incorporating data for multiple levels, it would be possible to improve the meaningfulness of the categories to the typical users as well as the ability of the classification to discriminate on consumer behaviour.

For example, when allocating postcodes to their best fit clusters, the 2003 UK Mosaic used information on the social class and employment structure of the 1km, 2km and 5km radii around each postcode as well as statistics for the postcodes themselves and for the census areas they fall in. This process results in new residential neighbourhood types such as 'Ex Industrial Legacy' which it would have been impossible to have identified using data at the postcode and census output area level only. Likewise the appearance of a 'Rural Non Farm' cluster in Netherlands Mosaic arises from the use in the classification of statistics on the proportions of the workforce employed in agriculture at the 4-digit and 5-digit postcode level as well as the finer 6-digit postcode level.

In the Netherlands it was eventually decided that as much as 25% of the weight of input variables used in the classification should apply to variables used at a coarsest level of geography, 75% of the weight of input variables applying to statistics pertaining to the finest level of geography for which the data were available.

Whilst the inclusion of statistics designed to address these neighbourhood effects contribute significantly to the process of maximising the discrimination that geodemographic classifications achieve it is important to recognise a second important characteristics of geodemographic classifications which distinguish them from person and household level demographics – that is that they are is based on more than one dimension, 'multivariate' in character. The multivariate form of the classifications allows them to incorporate additional dimensions to those incorporated in standard demographics. Thus for predicting whether or not a person owns a lawnmower whether that person lives in a flat or a house with a garden is likely to an important discriminator. For certain types of travel product ethnicity may well be as important or indeed more important than age or income. How far away a person lives from major towns and how well or badly they are therefore served by public transport is likely to have a significant effect on the share of household income devoted to motoring. 'Military Quarters' is further example of a neighbourhood whose rather unusual consumption profile is the result of factors outside the conventional set of demographic discriminators.

Analyses

For the purpose of the analysis of the behaviours covered by the TGI we have calculated the discriminatory power of ten discriminators, all but neighbourhood operating at a person or household level. We have applied these discriminators to a set of 54 behaviours. The survey database we have used in the TGI. The respondent file which we have used for the analysis is the 24,750 respondents during the period April 2002 to March 2003.

In order to avoid potential accusations of bias in our selection of behaviours, we invited two leading industry experts, both independent of any commercial interest in the outcome, to nominate a set of response codes from the survey for use in the analysis. These experts were Ken Baker, formerly Chief Statistician of the Target Group Index, who has extensive knowledge not just of the TGI questionnaire but of which questions in

the TGI are commonly used in client analyses, and Peter Sleight, author of 'Targeting Customers'[12] who is generally recognised as the leading independent consultant in the area of geodemographics and lifestyles.

These two individuals were invited to nominate a wide variety of behaviours, covering channels of communication and media as well as product categories and brands. They were invited to nominate behaviours from the widest possible range of market sectors (eg grocery products, holidays) but with the proviso that they should not select response codes where the aggregate number of respondents was likely to be too small to support analysis by all 61 of the Mosaic categories. They were also asked to select questions which are asked of all adults and not to include behaviours which were asked of one sex only. The reason for this is that sex was to be used as one of the discriminators and that therefore it would be impossible to measure the discriminatory power of the discriminator if responses were from one sex only. They were also asked to avoid selecting behaviours which are known to have a particularly strong regional bias, such as the consumption of porridge or whisky.

The primary purpose of the analysis was to examine the relative discriminatory power of Mosaic (as a neighbourhood classification) as compared with 'conventional' person and household level demographics. However a secondary intention was to examine what we believed to be the increasingly de-alignment of indicators of status, namely income, occupational status, terminal education age and house tenure. Our hypothesis was that in an increasingly post industrial (and post modern) society, there is an increasing fragmentation within what were previously described as middle classes in terms of the source of their social status. At the same time we supposed that the degree to which status accrues to individuals on the basis of the display of their tastes (for which we may take education to some degree as a proxy), the nature of the work they do (assuming any one other than in their own work organisation can understand their job title) and the degree to which they conspicuously consume does vary between social contexts and between residential locations. If, as is quite clearly the case with the current Mosaic classification, we find that many types of neighbourhood rank significantly differently on each of these dimensions and display quite different consumption patterns, then it likely that the relationship between consumption and these different indicators of status will have become increasingly de-aligned.

Next it was necessary to define a statistic whereby we could make reliable comparisons between different demographics in terms of their ability to discriminate in different behaviours. The statistic which we have used for this purpose is one which we refer to as 'Total Weighted Deviation' and has proved reliable as the measure used to optimise performance during the process of building the Mosaic classifications.

Suppose that we wish to measure the degree to which the demographic variable 'Income' is an efficient discriminator of readership of The Guardian. The Total Weighted Deviation statistic for income as a discriminator of readership of The Guardian is calculated as follows.

For each class interval by which the income of TGI respondents is recorded we calculate the level of readership of The Guardian.

We express this level as a percentage of the average level of Guardian readership across all respondents such that an index of '100' indicates an average level of Guardian readership, an index of '120' a level 20% above the survey respondent average.

We then calculate the absolute difference between this percentage level and 100%. This is the deviation of that income class from the mean. In the example above this deviation would be one of '20' points

We then multiply (ie weight) this deviation by the proportion of all TGI respondents falling within that class.

Finally we calculate this weighted deviation across all other classes of income in relation to readership of The Guardian and sum the result.

The maximum possibly efficiency of the discriminator would be manifested in the event that each class on the discriminator would contain either only positive or only negative respondents on the behavioural question, in this case if each income groups contained only respondents who did read The Guardian or who did not read The Guardian. In such a limiting case the TWD score would be 100. By contrast if Guardian readers and non Guardian readers were distributed evenly across all income classes the TWD score would be 0. The higher the TWD score therefore the higher the level of discrimination afforded by the demographic in relation to that particular behaviour.

	Total positive	Average twd
Total number of respondents	24750	
SAILING/YACHTING- PLAYED OR TAKE PART IN REGULARLY	120	34.1
ELECTRIC DISHWASHER- OWNED NEW OR SECONDHAND	6886	33.0
BINGO- PLAYED OR TAKE PART IN REGULARLY	907	31.7
DAILY TELEGRAPH- LOOK AT ALMOST ALWAYS/QUITE OFTEN	1711	30.9
AIR SUNDAY TIMES	1537	30.8
THE GUARDIAN- LOOK AT ALMOST ALWAYS/QUITE OFTEN	1040	30.7
OWN HOME COMPUTER	14215	30.5
PRE-MIXED SPIRITS OR ALCOHOLIC CARBONATES- DRUNK ONCE A WEEK OR MORE	537	30.5
ATTITUDES- TO DO MY SHOPPING BY INTERNET MAKES LIFE EASIER	2247	29.7
OWN PATIO DOORS	5498	29.4
ANY USE-HAND ROLL TOBACCO	2813	29.0
HAVE A STORE CARD	7887	28.9
DRAUGHT LAGER- DRUNK 8 PINTS OR MORE IN LAST WEEK	1043	28.6
BOTTLED TABLE WINE- DRUNK 5 BOTTLES OR MORE IN	3102	28.6

LAST WEEK		
CONSERVATORY/SUNLONGE- OWNED NEW OR SECONDHAND	3110	28.1
EVER VISIT LICENSED CLUBS	5792	27.0
DIGITAL CAMERA- OWNED NEW OR SECONDHAND	3655	26.8
CIGARETTES- SMOKE 20 OR MORE PER DAY	3082	24.6
THEATRE- ONCE A MONTH OR MORE OFTEN	424	24.4
FRESH GROUND COFFEE- ONCE A WEEK OR MORE OFTEN	4539	21.9
ATTITUDES- WOMAN'S PLACE IS IN THE HOME	947	20.8
HOLIDAY IN NORTH AMERICA- LAST HOLIDAY	934	20.6
HOLIDAY IN REST OF WORLD (EXCL. N.A.)- LAST HOLIDAY	666	20.4
HOLIDAY IN EUROPE- LAST HOLIDAY	6438	20.2
DRAUGHT BEER/MILD OR STOUT- DRUNK 8 PINTS OR MORE IN LAST WEEK	845	20.0
MOST OFTEN USERS RADOX BATH CUBES	648	19.8
TV SETS IN HOUSEHOLD- THREE OR MORE	10508	19.7
BUYING BULBS (GARDENING)	9466	19.4
OWN A DOG	3198	19.3
OWN FOUNTAIN PEN	5805	19.2
FRESH SOUP- ONCE A WEEK OR MORE OFTEN	2566	18.0
MOST OFTEN USERS CINZANO BIANCO	438	17.9
DAY TRIP TO EUROPE IN LAST 12 MONTHS	1837	17.7
KEEP-FIT EQUIPMENT- OWNED NEW OR SECONDHAND	3907	17.2
BROWN/OTHER SAUCES- ONCE A WEEK OR MORE OFTEN	8825	17.0
SPECIALLY CHOOSE TO WATCH RICHARD & JUDY	2440	16.5
SPECIALLY CHOOSE TO WATCH RWHO WANTS TO BE A MILLIONAIRE	7735	16.1
BOTTLED MINERAL WATER- DRINK ONCE PER WEEK OR MORE OFTEN	6799	15.3
MOST OFTEN USERS SAVLON	774	15.3
MOST OFTEN USERS DAIRYLEA TRIANGLES	3696	15.2
2 OR MORE HOLIDAYS (EXCLUDING SHORT BREAKS)	4994	15.1
SPECIALLY CHOOSE TO WATCH TONIGHT WITH TREVOR MCDONALD	4586	14.7
REGULAR MAJOR SHOPPING ASDA	5625	14.5
TINNED SOUP- ONCE A WEEK OR MORE OFTEN	5045	14.3
MOST OFTEN USE DULUX PAINT	6613	13.7
MOST OFTEN USERS MCCAIN OVEN CHIPS	3278	12.9
BOUGHT FROM MAIL ORDER CATALOGUE IN LAST 12 MONTHS	10941	12.9
ANY- TRYING TO SLIM	8771	12.4
HOLIDAY IN BRITAIN- LAST HOLIDAY	7087	11.0
ANY USE WHITE RUM	2592	10.7
ANY USE MOUTHWASHES AND GARGLES	8829	8.6
HEAVY/MEDIUM USERS BUTTER	8660	8.5
EVER BUY TOFFEES & CARAMELS	9096	6.1

Table one: Average TWD values of studied behaviours, all discriminators.

The list of behaviours selected by Baker and Sleight is shown in table one. They are organised in descending order according to how effective the selected discriminators were on average at discriminating them, ie according to their average TWD scores. For instance of the 54 behaviours 'sail regularly' is the one most closely aligned with particular age groups, income groups, types of neighbourhood and so on. By contrast the market for toffees and caramels is the one for which conventional demographics (including geodemographics) appears to be the least effective. As would be expected niche and up market behaviours tend to be positioned to the top of the list, mass market and down market behaviours being largely concentrated at the bottom. It is also evident, as one would expect when doing a test of this sort, that the levels of TWD achieved tend to decline the higher the proportion of positive respondents. A behaviour which is undertaken by 50% of the adult population is one which is much less likely to lend itself to effective segmentation, whichever demographics one has available, than a behaviour which is undertaken by 5% of the population. This is confirmed by the negative correlation (-0.343) between the proportion of TGI who respond affirmatively to a survey question and the average TWD value of the demographic characteristics used as discriminators.

Most powerful discriminator

Behaviour

Age

PRE-MIXED SPIRITS OR ALCOHOLIC CARBONATES- DRUNK ONCE A WEEK OR MORE

ATTITUDES- TO DO MY SHOPPING BY INTERNET MAKES LIFE EASIER

EVER VISIT LICENSED CLUBS

TV SETS IN HOUSEHOLD- THREE OR MORE

BUYING BULBS (GARDENING)

OWN A DOG

SPECIALLY CHOOSE TO WATCH RWHO WANTS TO BE A MILLIONAIRE

SPECIALLY CHOOSE TO WATCH TONIGHT WITH TREVOR MCDONALD

ANY USE WHITE RUM

ANY USE MOUTHWASHES AND GARGLES

HEAVY/MEDIUM USERS BUTTER

Marital Status

SAILING/YACHTING- PLAYED OR TAKE PART IN REGULARLY

BINGO- PLAYED OR TAKE PART IN REGULARLY

DAILY TELEGRAPH- LOOK AT ALMOST ALWAYS/QUITE OFTEN

HOLIDAY IN BRITAIN- LAST HOLIDAY

Gender

DRAUGHT LAGER- DRUNK 8 PINTS OR MORE IN LAST WEEK

DRAUGHT BEER/MILD OR STOUT- DRUNK 8 PINTS OR MORE IN LAST WEEK

MOST OFTEN USERS CINZANO BIANCO

SPECIALLY CHOOSE TO WATCH RICHARD & JUDY

MOST OFTEN USERS SAVLON

MOST OFTEN USERS MCCAIN OVEN CHIPS

BOUGHT FROM MAIL ORDER CATALOGUE IN LAST 12 MONTHS

ANY- TRYING TO SLIM

Education

AIR SUNDAY TIMES

THE GUARDIAN- LOOK AT ALMOST ALWAYS/QUITE OFTEN

BROWN/OTHER SAUCES- ONCE A WEEK OR MORE OFTEN

Income

ELECTRIC DISHWASHER- OWNED NEW OR SECONDHAND

OWN HOME COMPUTER

HAVE A STORE CARD

DIGITAL CAMERA- OWNED NEW OR SECONDHAND

ATTITUDES- WOMAN'S PLACE IS IN THE HOME

HOLIDAY IN EUROPE- LAST HOLIDAY

FRESH SOUP- ONCE A WEEK OR MORE OFTEN

KEEP-FIT EQUIPMENT- OWNED NEW OR SECONDHAND

BOTTLED MINERAL WATER- DRINK ONCE PER WEEK OR MORE OFTEN

Social Grade

BOTTLED TABLE WINE- DRUNK 5 BOTTLES OR MORE IN LAST WEEK

FRESH GROUND COFFEE- ONCE A WEEK OR MORE OFTEN

HOLIDAY IN NORTH AMERICA- LAST HOLIDAY

OWN FOUNTAIN PEN

Tenure

ANY USE-HAND ROLL TOBACCO

CONSERVATORY/SUNLONGE- OWNED NEW OR SECONDHAND

Mosaic Type

OWN PATIO DOORS

CIGARETTES- SMOKE 20 OR MORE PER DAY

THEATRE- ONCE A MONTH OR MORE OFTEN

HOLIDAY IN REST OF WORLD (EXCL. N.A.)- LAST HOLIDAY

MOST OFTEN USERS RADOX BATH CUBES

DAY TRIP TO EUROPE IN LAST 12 MONTHS

MOST OFTEN USERS DAIRYLEA TRIANGLES

2 OR MORE HOLIDAYS (EXCLUDING SHORT BREAKS)

REGULAR MAJOR SHOPPING ASDA

TINNED SOUP- ONCE A WEEK OR MORE OFTEN

MOST OFTEN USE DULUX PAINT

EVER BUY TOFFEES & CARAMELS

Table two : behaviours grouped according to the most effective discriminator

Table two shows the behaviours which are most powerfully predicted by each of the discriminators used in the study. This table might be of more interest if the study had included a larger number of behaviours. However it is evident even from this limited list that behaviours which are most likely to driven by age and gender tend to relate to the consumption of alcohol to television programmes watched. As might be expected behaviours for which income is the best predictor include most of those involving the ownership of high ticket consumer durables. The few behaviours for which terminal education age is the most powerful predictor include newspaper readership, a behaviour with a low cost but from which satisfaction is gained as a result of education. The behaviours for which Mosaic type was the most powerful discriminator include a number of holiday questions and quite a high proportion of the fmcg questions included in the study.

Total Weighted Distribution					
	average	maximum	minimum	standard deviation	times highest
Discriminator					
Mosaic Type	26.09	50.83	8.87	11.09	12
Social Grade	22.59	49.01	4.05	12.53	4
Income	22.26	45.69	0.00	11.39	9
Age	22.23	69.25	6.21	14.11	11
Mosaic Group	21.88	43.44	5.34	10.82	0
Education	21.64	53.67	4.36	11.92	3
Tenure	17.69	47.12	0.91	12.30	2
Gender	17.15	71.95	0.59	18.04	8
Marital status	15.31	56.72	2.85	10.95	4

Table three: discriminators – consistency of performance

In table three the discriminators have been ranked according to their average TWD on the 54 behaviours. On this basis neighbourhood (Mosaic type) proves a more discriminating measure than any of the person or household level demographics. Measures of status, social grade, income, education and age proved to be more useful discriminators than tenure, gender and marital status.

However from the 'standard deviation' column which measures variability in the TWD scores of each discriminator against different behaviours it is evident that gender, and to a lesser extent age, differ from the other discriminators in performing very well in a limited number of behaviours but very badly on many others.
.
For example out of the 54 behaviours there were eleven where age was the more effective predictor and a further eight where gender provided the best discriminator. Whereas Mosaic type was the best predictor on average and was the most predictive discriminator on most occasions (twelve) its TWD values clearly lie within a narrower spread than those of any other discriminator other than marital status.

It is also interesting to note that although social grade performs marginally better than income across the board, income is more likely to be the best discriminator in instances where status is the principal correlate with behaviour.

We have mentioned earlier the hypothesis that a likely characteristic of a post industrial economy is a progressive de-alignment of different measures of status such as neighbourhood, occupation, income and terminal education. Table four seeks to explore this hypothesis by measuring the extent to which the behaviours for which different discriminator achieve a high TWD happen to overlap. The table is created by correlating the TWD scores of each discriminator on the different behaviours. Where two discriminators tend to achieve a similar level of discrimination across individual behaviours values in the corresponding cells will the high.

	Mosaic Group	Mosaic Type	Gender	Age	Marital status	Income	Social Grade	Education	Tenure
Mosaic Group	1.00	0.96	-0.19	0.10	0.24	0.56	0.86	0.62	0.85
Mosaic Type	0.96	1.00	-0.08	0.12	0.23	0.57	0.85	0.67	0.77
Gender	-0.19	-0.08	1.00	0.03	-0.06	-0.21	-0.24	-0.05	-0.33
Age	0.10	0.12	0.03	1.00	0.63	0.21	0.07	0.31	0.10
Marital status	0.24	0.23	-0.06	0.63	1.00	0.31	0.11	0.16	0.36
Income	0.56	0.57	-0.21	0.21	0.31	1.00	0.75	0.61	0.56
Social Grade	0.86	0.85	-0.24	0.07	0.11	0.75	1.00	0.82	0.71
Education	0.62	0.67	-0.05	0.31	0.16	0.61	0.82	1.00	0.35
Tenure	0.85	0.77	-0.33	0.10	0.36	0.56	0.71	0.35	1.00

Table four : correlation matrix showing the correlation between the TWD of each discriminator across individual behaviours and the TWD of other discriminators

From this table it is evident that behaviours for which Mosaic type is a good discriminator tend to be the same behaviours for which social grade, tenure, terminal education age and income are also powerful discriminators. By contrast the behaviours for which Mosaic is a powerful discriminator tend to ones for which marital status, age or gender are relatively poor discriminators and vice versa. Neighbourhoods, it would seem, are more homogeneous in respect of status than they are in respect of life stage.

We can also see from this table that as a discriminator of behaviour social grade is better aligned with neighbourhood, with terminal education age and with tenure than is income, which is surprisingly poorly aligned with other discriminators. Tenure, although it is fairly well aligned with neighbourhood and social grade, is poorly aligned with income and very poorly aligned indeed with education.

Products which are bought principally by people (or more often households) with high income, we might suppose, are bought by these people primarily because they have a lot of money to spend, not because they are indicative of a high level of taste or sophistication or indeed because they confer a high level of prestige.

Results support the contention that if we are able to infer the type of neighbourhood in which a prospect or customer lives, then a larger incremental gain can be achieved if we can also establish at the same time the age and gender of that prospect or that customer than would be the case if we were to establish their occupation, income or terminal education age. With the correlation matrix indicating that the behaviours predicted well by age are wholly different from those that are predicted by gender, it would seem that type of neighbourhood, age and gender represent three items of information which are 'orthogonal', ie complementary to each other in that they operate in three quite independent domains. Given that both gender and age can be inferred from a person's first name with a fair degree of reliability (especially when also using public information such as years at their current address on the electoral roll and the presence and name (if present) of a partner) then it would seem that most behaviours could be predicted for any consumer from their name and address with a fairly high degree of success relying solely on external data and without recourse to other demographic measures for which survey or questionnaire data would otherwise have to be obtained.

Evaluation

Perhaps the most likely criticism of the methodology is that the different discriminators used do not have equal numbers of classes. Thus the discriminator gender has two classes only whilst the Mosaic type has 61. Comparison between the TWD of discriminators with different numbers of classes might be seen in some way to be unfair.

There is some validity to this objection in so far as clearly the larger the number of classes into which a discriminator is organised the less statistically reliable the 'index' values will be, particularly where classes with small sizes are concerned, as for example with 'Military Quarters' or where the discriminator is applied against a behaviour which is undertaken by only a small minority of respondents.

The TWD statistics addresses this problem only in part by incorporating the size of each class into its calculation. Thus a statistically unreliably high or low index in a small neighbourhood type such as Military Quarters will have a disproportionately low effect, albeit some effect, on the computation of the overall TWD score.

To an extent we tried to address this problem also by requesting that the behaviours selected for analysis were ones which have a relatively high penetration across the sample file. By this means we hoped to reduce the impact of wobbly index values based on small sample sizes.

Another method of addressing the problem is to restrict the analysis of behaviours with smaller sample size to the eleven Mosaic Groups into which the more detailed 61 Mosaic Types are organised. Table three provides useful evidence of the extent to which this reduction in the number of classes affects average TWD across the set of 54 behaviours used in the study. Maybe a fairer evaluation would be one which reported on the TWD

of Mosaic Types for behaviours with a larger sample of positive respondents and for the TWD of Mosaic Groups for behaviours with a smaller sample of positive respondents.

The other possible objection, that a classification system with more classes is likely to be inherently more predictive than one with fewer classes, is clearly true. However the counter argument is that one of the limitations of gender is that it only has two classes. Age and Income clearly could be disaggregated into more classes than the set used in the TGI. However where a classification is one-dimensional rather than multivariate there is a natural tendency for a given increase in the number of classes to result in incrementally less improvement in discrimination than would be the case for classification systems which are multi-dimensional in character.

Clearly the detail into which a classification system is divided may affect the workload involved in accessing that classification. Thus a more detailed breakdown of social grade or of tenure would involve a marginal additional to fieldwork time and cost. The use of a finer classification of age and income would also involve additional fieldwork cost if the respondent is required to specify which class he or she belongs to. By contrast the level of detail by which respondents neighbourhoods are classified has no effect on cost because the classification is generated electronically from a postcode to neighbourhood type correspondence table not from the interview process.

Conclusion

From the evidence of this limited set of behaviours, there is support for the contention that geodemographics is more than 'a poor man's demographics' and that variations in consumption between neighbourhoods is greater than can be explained simply by the averaging out of the behaviours of their constituent age and income profiles. Clearly there are sociological reasons why a consumer's behaviour is affected by the behaviour of people who live in close proximity and with whom he or she comes in contact on a regular basis. There does seem evidence that these neighbourhood effects may in part depend on influences which cover a wider geographical spread than that of the immediate street or postcode. Based on comparisons with an earlier study (for which the results unfortunately have been lost) we believe that the superiority of discriminatory power of neighbourhood over occupational and income classifications is probably of recent origin. From which we believe that the relative effects of neighbourhood will increase vis a vis the effects of age, social class and income over time, notwithstanding the 'loss of distance' that some social commentators have forecast as being a consequence of the electronic age.

[1] Webber, R. (1977) 'The National Classification of Residential Neighbourhoods : An introduction to the Classification of Wards and Parishes' PRAG Technical Paper No 23, Centre for Environmental Studies

[2] Webber, R (1979) 'Census enumeration districts: a socio-economic classification', PRAG Technical Paper, Centre for Environmental Studies

[3] Webber, R. (1975) 'Liverpool Social Area Study, 1971 data', PRAG Technical Paper No 14, Centre for Environmental Studies

[4] Webber, R. and Craig, J. (1978) 'Socio economic classifications of Local Authority Areas', OPCS Studies on Medical and Population Subjects, no 35.

[5] Webber, R. (1978) 'Parliamentary constituencies : a socio economic classification', OPCS occasional paper 13

[6] Baker, K., Bermingham, J. and McDonald, C. (1979) 'The utility to market research of the classification of residential neighbourhoods', paper presented at the Market Research Society Conference, Brighton, March, 1979

[7] Jones, K., (1991) 'Muti-level models for geographic research', Concepts and techniques in modern geography, 54. Norwich : Environmental Publications

[8] Bourdieu,P. (1985) 'The forms of capital', in Richardson,J. Ed. *Handbook of Theory and Research for the Sociology of Education*, 241-58. New York: Greenwood

[9] Keans, A., (2004) 'Social Capital, regeneration and urban policy', CNR paper 15, School of Policy Studies, University of Bristol

[10] Pacific Micromarketing (2003) 'Multimedia Guide to Australian Mosaic' (CD)

[11] Experian International Ltd (2004) 'Multimedia Guide to UK Mosaic' (CD)

[12] Sleight, P., (2004) 'Targeting Customers, Third Edition, How to use geodemographic and lifestyle data in your business', NTC Publication, Oxford