

---

# Alternative data sources in targeted marketing: The value of exographics

Received (in revised form): 22nd August, 2005

## Henry Greene

is a consultant for Information Xperts.

## George R. Milne

is Associate Professor of Marketing at the University of Massachusetts, Amherst.

**Abstract** This paper supports the use of exographics in a database marketing context. Exographics, defined as contextual characteristics beyond the scope of individual, household or neighbourhood space, are argued to have an added advantage over data sources traditionally discussed in the literature, such as buyer behaviour, demographics, psychographics, life stage/life event and geodemographics. The value of exographics is demonstrated by constructing regression models to predict response to direct marketing campaigns for a motor insurance company.

**Keywords:** *data issues; direct marketing; geodemographics; measurement performance*

## INTRODUCTION

### Selecting and managing data

Organisations today are warehousing multiple databases, creating data marts to meet the specific needs of functional groups within their enterprises. For the database administrator, it is necessary to determine which data fields should be included in the warehouse. For a marketing department, data items are chosen that contribute to a variety of business objectives: profiling customers, predicting response rates and segmenting customers for specific communication efforts. Data selection in practice<sup>1-3</sup> is more of an art, depending on business strategy, availability, cost, experience and technology.

The trend among database practitioners in the USA over the past 40

years has been to utilise smaller scale geographies. In the 1970s, marketers were only able to capture aggregate data at large scales such as city and ZIP code. Technology has now made it possible to utilise data at very small geographies. In the USA these include census blocks groups (400–500 households), postal carrier routes (200–300 households), census blocks (20–30 households) and postal ZIP+4 (6–8 households). The underlying strategy is to make the data descriptors more relevant and precise.<sup>4</sup> As direct marketers have narrowed their scales of data measurement, however, changes in the market place have also occurred. These changes include geographically larger market places, greater diversity of language, social and political characteristics of potential customers, and larger and more

**George R. Milne**  
Isenberg School of  
Management, University of  
Massachusetts,  
Amherst, MA 01002, USA  
Tel: +1 413 545 5669;  
Fax: 1 413 545 3858;  
e-mail:  
milne@mktg.umass.edu

technically sophisticated distribution channels. In contrast to industry trends, understanding these changes requires larger scales of measurements, focusing on city, region, state or country.

Measurement issues are often overloaded in model development stages. While the literature frequently addresses methodology improvements,<sup>5-7</sup> the quality, relevance, accuracy and completeness of data are often ignored. As underscored by many empirical modellers, no model can succeed when predictor variables are inappropriate, inaccurate or invariant.<sup>8,9</sup>

Given that marketers are looking for variation in data elements to differentiate prospective customers, it seems natural that they would prefer to work at the micro level. All else being equal, small samples of data have more variation than larger samples. But all else is not equal. People do not randomly choose to live within geographical boundaries. Social, political, historical and economic forces influence where people live. Commonalities within geographies exist, and differences between geographies may be greater than differences within geographies when measuring specific attributes of a population or consumer patterns. These data dimensions need to be recognised.

### **New data dimensions**

To go beyond the trend of direct marketers focusing on smaller geographical levels requires consideration of contextual conditions, conditions associated with alternative data sources and measurement scales.<sup>10-12</sup> As this paper argues, there is a class of variables denoted as exographics. These exographics represent data items beyond the spatial being of a person's immediate surroundings. The outer boundary of immediate surroundings is defined as

household neighbourhoods, small, geographically contiguous sets of homes. Examples of exographic variables would include summary statistics at the state, county or city level. This differs from geodemographics, which are typically defined at small geographical levels such as the census block.

In many respects, the motivation for using exographics is supported by observations of behaviours across geographies. For example, cross-country skiing is more popular in Minnesota than in Alabama. Pizza is more popular on the East Coast than in the Midwest. New fashions are first found in cities rather than smaller communities. Yet, there is no framework for database marketers to capture these phenomena. Further, it is not reasonable to assume that prior experience with products and geographical location will uncover useful model predictors.

The purpose of this paper is to develop a data strategy that supports the use of exographics and to demonstrate how these additional data sources may benefit database marketers. The remainder of this paper is organised in four sections. The first section presents an exographic components approach. The second section discusses database marketing applications of exographics. In this section a framework is provided to evaluate exographics and traditional data sources in terms of effectiveness, cost and privacy issues. The third section describes a study using exographics for prospecting and acquiring customers for a motor insurance company. The fourth section discusses implications, limitations and future research considerations.

## **EXOGRAPHIC COMPONENTS APPROACH**

Geographers have recognised the influence of multiple domains on

**Table 1:** Exographic scales and domains

Exographics Scales	Domains	Examples
Region	Nature Culture Built	Hurricanes, climate Spanish settlements Water dam, interstate highways
State	Nature Culture Built	Mountains, lakes Tax laws Bridges
MSA	Nature Culture Built	Weather, snow days, sunny days Catholic festivals Professional sports cost of living
City	Nature Culture Built	Rivers, ports Buddhism Colleges, traffic

MSA, metropolitan statistical area.

consumption.<sup>13–16</sup> This paper modifies the environmental realms of Sach and extends the argument of Webber to define exographic influences, as incorporating two aspects. These are the character of the domains (ie nature, culture and man-built), and the scale (city, county, metropolitan statistical area [MSA],<sup>17</sup> state and regional levels). Table 1 provides examples of exographics by scale and domain.

Exographics are useful for two reasons. First, they provide additional insight into consumer behaviour by examining the interactions among the natural, built and cultural environments. Secondly, they focus marketer attention on the relevant scale of analysis.

When incorporating exographics into models the marketer must decide which variables will serve as useful predictors. In deciding upon specific exographics a three-step process is recommended:

- 1 For each exographic dimension, create a list of characteristics that represent reasonable forces that may affect the consumer response of interest.
- 2 Once the set of variables is constructed, select levels of scale that would be appropriate.
- 3 For each variable selected, test

different levels of scale to determine which is most useful for predicting response.

As an illustration, consider the case of a retailer deciding to stock fashion boots. For the nature domain, one might consider precipitation and average winter temperature; for the cultural domain, one might consider urbanicity and its relationship to the history of boot sales, and the percentage of people owning boots. Finally, for the man-built domain, one might consider the number of retail stores and average distance walked to work. In addition to recognising appropriate forces which may affect consumer consumption, it is necessary to determine the appropriate geographical scale which will maximise explanation. In an effort to determine the potential sales of women's fashion boots, one may choose temperature and precipitation at the MSA level; in addition, fashion boots may be differentiated by an urbanicity measure attributed to city characteristics. In many cases, data may only be available at large scales, thus negating the possibility of analysing the value at smaller-size geographies. It is always possible to aggregate lower-level data items to larger-scale geographies,

although it may not be meaningful. To finalise the scale selection, one can build datum at each scale level and determine which scale maximises the correlation between the exographic and the consumer response of interest.

### **Direct marketing applications of exographics**

Environmental forces have been acknowledged in marketing, specifically in the fields of geodemography and spatial models.<sup>18–24</sup> Applications abound in retail sales assessment and site location.<sup>25</sup> Different contextual conditions have been noted in the marketing literature, which have an impact on consumer behaviour, for example, travel time,<sup>26</sup> retail atmospherics,<sup>27</sup> minority status<sup>28</sup> and population density.<sup>29</sup>

The choice of data used by direct marketers depends in part on the relationship the marketer has with the consumer. The marketing programme that a company selects to address the needs of a customer along the relationship continuum depends on the customer's past behaviour as well as specific information about the customer that can be acquired. The nature of the information, both the breadth and depth, will change over time. Early on in a firm's relationship with a customer there is little information available, so marketers often rely on third-party data sources or more macro level data such as geodemographics. As the relationship with the customer develops, more customer-specific data are collected. Other researchers have described similar chains of relationships. Zeithaml and Bitner<sup>30</sup> describe a sequence of four relationship types in the service industry: financial, social, customisation and structural. This study adds a customer prospect stage to the database relationship process, and addresses how data sources

foster the relationships.

Cost and privacy are two other important issues regarding customer information. When customer demographics, financial history or purchase history are unknown, information may be obtained from third-party vendors such as: Acxion, Claritas, Donnelley Marketing, Equifax, Infobase and InfoUSA. Acquiring personal customer information may be costly (as much as \$0.10 a name for one data item, for one use). Acquiring personal information may be viewed by the consumer as inappropriate and a violation of their privacy. Neither geodemographic nor exographic data are encumbered with these issues. This is because data items from these sources are generated from public records of large aggregations of individuals or forces, such as economic, meteorological and governmental, which manifest on large geographical scales. Geodemographic information is available at much lower rates. In the USA, companies such as ScanUS, Claritas and ESRI typically charge \$1,000 for a set of 20–50 variables, which represents the entire nation, 110,000,000 households. Individual privacy is protected given that the typical size of a census geography (block group) is 300–500 households. Although the Census Bureau releases information at smaller geographies (block level, 20–50 households), if either the data item is regarded as sensitive (eg household income), or the number of households in the geography is quite small, the data item is not made available.

Table 2 demonstrates the importance of different data sources during the customer relationship process, which consists of five stages: prospect, acquire, retain, loyalty, partner. For each of these stages, the importance of eight data sources is indicated. These data sources

**Table 2:** Relationship stage with customers

Data sources for database marketers	Level of importance					Cost	Privacy
	1. Prospect	2. Acquire	3. Retain	4. Loyalty	5. Partner		
Buyer behaviour (RFMV)	na	+	+++	++++	+++++	L	M
Third-party buyer behaviour	+++	++	++	+	+	H	H
Demographics	+	+	++	+++	++++	M	M
Life stage	+	+	+++	++++	+++++	MH	MH
Life events	+	+	++	+++	++++	MH	MH
Psychographics	+	+	++	++	++	MH	MH
Geodemographics	+++++	++++	+++	++	+	L	N
Exographics (physical geography, culture, economics, social conditions, political conditions, transportation)	+++++	+++	+++	++	+	L	N

na, not applicable; N, not an issue; L, low; M, medium; H, high; MH, medium to high — typically dependent on whether information is acquired directly from consumers or from an outside party.

include buyer behaviour, third-party buyer behaviour, demographics, life stage, life events, psychographics, geodemographics and exographics. The importance of the data sources is indicated with either 'na', indicating the data source is not available, or a number of '+' signs, indicating relative importance. This table is constructed from an intuitive understanding of data availability and cost.

Each of the columns in Table 2 indicates a different relationship stage between the customer and the supplier. The first column represents the first stage in the relationship in which a supplier advertises and communicates to customers or households, with which it has had no prior business exchange, in an attempt to sell its merchandise or services. During the initial stage, there is no purchase history; frequently, there is no demographic, life stage, life event or psychographic information known about prospects. Although database marketers can purchase lists of prospects and to some extent select lists from third parties, the value of the data diminishes as a result of cost and shelf life. Third-party data, representing buyer behaviour, may have the highest potential but may also require significant expense or may not be negotiable. Depending on advertising

media, data elements describing individuals may be inappropriate. Direct marketing advertisements, distributed by saturating across cities, zones or ZIP codes, negate the potential precision of individual characteristics. By contrast, geodemographic and exographic data are always available for targeting prospects and therefore are extremely valuable during the first stage of relationship building.

In the second stage, when prospects have made purchases, a variety of customer characteristics become available and are useful for future communications. Customer location, purchase interest and personal characteristics can be gleaned from transaction records. Third-party data can be acquired in a more cost-effective manner (matching known customers with recent addresses) yielding more accurate and reliable information. As individual characteristics become available, the data sources associated with aggregated behaviour will be less informative.

In stages three and four, the consumer has made repeated purchases, and product purchasing patterns begin to emerge. As the customer continues to purchase, inferences can be made regarding future purchases and customer lifetime value; cross-selling opportunities

can be identified, and shopping personality (eg risk aversion, bargain hunter, fussy shopper and indecisive shopper) can be accessed. As this information continues to accumulate, aggregate data sources should continue to diminish in importance.

At the last stage of the relationship (stage 5), the customer participates as a partner. At this stage the customer speaks highly of the products sold by the firm, advertises the company logo and refers acquaintances. Customer characteristics known by the firm should overshadow the potential advantages of aggregated data and exographic data.

In addition to profiling the eight data sources by the five stages of customer relationship, Table 2 shows the relative cost and privacy risk associated with each data source. Exographics, along with geodemographics and internal behavioural data, have the lowest acquisition costs. In contrast to individual third-party data, privacy is not a significant issue when considering exographics or geodemographics. Personal privacy cannot be compromised by examining summary statistics that represent hundreds to millions of households.

Overall, exographics are shown to be a low-cost and noninvasive data source that could be useful to database marketers in the early stages of customer relationships. The next section focuses on the customer acquisition stage of the database marketing relationship.

## A DIRECT MARKETING STUDY

The study investigated the potential contribution of exographic data during the prospecting and acquisition phase of a direct marketing campaign launched by a motor insurance company. It is acknowledged that direct marketing campaigns can be conducted at both the individual/household level and

geographical level. This study was conducted at the geographical level to reflect the approach taken by several marketers, specifically by a motor insurance company, which chooses to market at the postal ZIP/carrier route level in order to derive significant savings from postal costs. By contrast, mailing to small geographical areas (rather than individual households) negates the potential benefit of utilising individual characteristics such as demographics or buyer behaviour information, since all households in a chosen neighbourhood are treated the same (mailed or not mailed). This seemingly inefficient strategy has been tested and evaluated by some direct marketers and has, in some cases, been concluded to be superior to household level targeting. Not only can substantial savings be realised in postal costs, but there is also a saving in the avoidance of household demographic data and associated data processing.

Geodemographics in this study are demographics created at the postal carrier route level based on US Census block group data, and then transformed by a major data supplier to the postal ZIP/carrier route level. There are approximately 500,000 ZIP/carrier routes across the USA, averaging approximately 200–300 household per geography. Postal carrier routes are the actual territories that a postal carrier covers each day. The exographics utilised in this study are data variables extracted from metropolitan statistical areas (there are approximately 336 MSAs in the USA, the smallest comprises approximately 20,000 households). The final data file used in this analysis consisted of 35,000 postal carrier routes, the response variable was continuous, averaged approximately 3 per cent and represents the percentage of households in each postal carrier route that received a direct mail advertisement

and called the insurance company requesting an insurance quote.

### Approach and rationale

First, a data file was created containing a continuous response variable (enquiries) and predictor variables in the form of geodemographics and exographics. The response variable and geodemographics were previously used by the insurance company to target households with response models. In order to demonstrate the additional performance that can be achieved by incorporating exographic data in a response model, it is necessary to keep the geodemographics at the neighbourhood level, a scale in which they are typically used. Aggregating the geodemographic variables to the MSA level would potentially dilute variable effectiveness. The selection of exographics was based on experience and insights offered in the literature. For example, Reinartz and Kumar<sup>31</sup> incorporated cost of living in their catalogue buyer model and Weiss<sup>32</sup> built a case for urbanicity as he profiled different consumption patterns across the USA.

Secondly, correlations were computed between the predictive variables and the response variable to assess the potential value of exographic variables in a regression model. The goal was to test the contribution of exographic variables above and beyond the explanatory power of data dimensions currently being utilised.

Thirdly, regression models were created under two conditions, (1) using only geodemographic variables and (2) using a combination of geodemographics and exographics as predictors. Models were created with a fixed number of variables in an effort to assess the value of adding exographic data. Specifically, pairs of models were built (one

geodemographics only and one with geodemographics and exographics) with the same number of predictors: 5, 10, 15 and 20 variables. In estimating the models, observations were weighted based on number of households, which varied from 50 to 300. A least squares regression model was used given that the response variable was continuous (percentages of responses per neighbourhood).

Fourthly, model performance was evaluated by summarised response performance of models with and without exographics in the form of decile charts and gains tables.<sup>33–38</sup> This approach was taken because it is recognised<sup>39–41</sup> that statistical issues such as multicollinearity, adjusted  $R^2$  and  $F$  tests, although important to consider when measuring model accuracy, are secondary when judging model performance in a direct marketing context.<sup>42</sup> Large data files ( $N > 10,000$ ) with small response rates ( $r < 0.04$ ) tend to exhibit extremely small correlations and  $R^2$  values, which are frequently statistically significant. As expected, during our model testing, all  $F$  values for the regression models were statistically significant.

### Specific details

A data file resulting from direct marketing campaigns by a motor insurance company over a three-month period in 2002, netting 200,000 enquiries, representing 50,000 neighbourhoods (US postal carrier routes), was obtained from Information Xperts, LLC. There are approximately 500,000 neighbourhoods across the country based on ZIP code/postal carrier routes. The universe of interest, however, was limited to those neighbourhoods in which advertising copy was distributed and at least one household made an inquiry. Because the

direct mail vehicle is executed and discounted via 'saturated' mail (90 per cent or more households in the carrier route are mailed), only neighbourhood-level demographics (geodemographics), representing census data, and exographics (built at the US MSA level) were appended to the file. Neighbourhoods on the data file that fell outside the boundaries of an MSA were eliminated because exographic data were not available. These neighbourhoods tend to exist in very remote areas and have very small populations. The final analytical file represented 33,763 neighbourhoods.

Traditionally, direct marketers utilising a saturated mail strategy select geodemographics based on availability and managerial appropriateness to the situation. Predictor variables popular among direct marketers and used in this study consisted of marital status, number of cars per household, average education level of adults in the neighbourhood, home ownership, home value and household income.

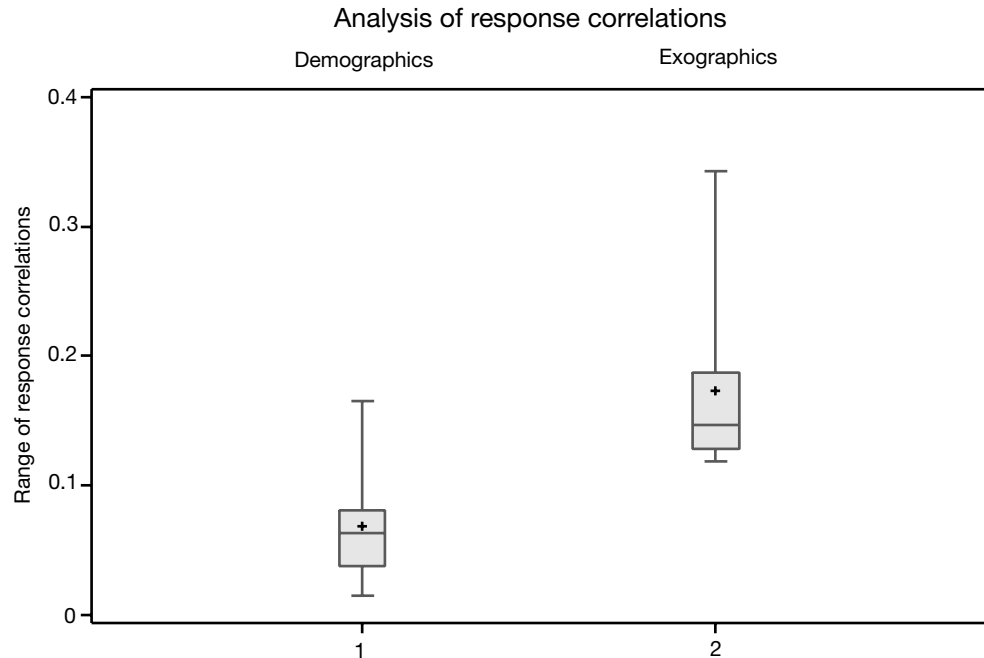
The exographic variables were selected by considering each individual domain and selecting variables from the domain which seemed to be appropriate for predicting motor insurance premium changes. In the domain of nature, weather, precipitation and climate were selected as factors that would affect driving conditions and would have an impact on premium rates. For the social domain urbanicity, car theft and crime rate were selected. For the man-built domain, cost of living, commute time and traffic density were selected. Although it was suggested that different levels of scale be investigated, only one level (MSA) was tested with the understanding that if the value of exographics could be justified with only one scale, utilising multiple scales could only improve the explanation power of

the model. Common transformations (square root, logarithms, squares) were considered to adjust for non-linear relationships. In total, 109 variables, 85 geodemographics and 24 exographics were available for creating the subsequent response models.

To assess the contribution of the potential predictor variables, the predictor variables were correlated with the response variables. As noted in Figure 1, several of the exographic variables exhibit higher correlations with the enquiry response variable as compared with the geodemographic variables.

To evaluate the relative contribution of exographic variables compared with geodemographics, stepwise regression models were run using 5, 10, 15 and 20 variables drawn from large variable pools. In the first set of regression models, the candidate pool of variables was the set of geodemographic variables; in the second set of models, the candidate pool of variables consisted of both geodemographic and exographic variables. In all eight models, all entered variables were significant at the 0.05 level. Although this paper argues that examining adjusted  $R^2$  should not be the primary consideration when evaluating model performance, the adjusted  $R^2$  values for each model pair are presented to demonstrate the superior model fit of models built with exographic variables. The adjusted  $R^2$  values for respective models with five variables were 0.0039 vs 0.0093; models with 10 variables were 0.006 vs 0.013; models with 15 variables were 0.0064 vs 0.0123; models with 20 variables were 0.0069 vs 0.0132. While direct response models typically possess small  $R^2$ , the noticeable increase in  $R^2$  between the two sets of models indicates a potential benefit to incorporating exographic variables.

Using standardised beta weights, the relative contribution of exographics was



**Figure 1** Absolute value of correlations — insurance inquiries with predictor variables

calculated in part by examining the amount of explained variance accounted for by exographic variables. For example, Table 3 shows the variables that entered into the 15-variable regression model using geodemographics and exographics. The standardised betas associated with the exographic variables account for 23.2 per cent of the total beta weights.

Across all models, the relative contribution of the exographic variables was measured by evaluating the relative weights of the standardised exographic betas to the total of all standardised betas. Table 4 shows the relative contribution of the exographic variables; the contributions vary from a low of 23.2 per cent to a high of 67.2 per cent. In addition, the incremental  $F$  statistic was statistically significant for each model, further supporting the contribution of exographics.

Roberts and Berger<sup>43</sup> discuss the merits of gains charts and Pareto curves to evaluate performance rather than

traditional statistics. For the 15-variable models, geodemographics and geodemographics and exographics, response rates by decile (gains chart) were computed and are presented in Figure 2. The charts indicate that the full model, utilising geodemographic and exographic variables, has higher performing segments at the top deciles and lower performing segments at the lower deciles than the geodemographic model. Additionally, the total variation among the decile indices is greater for the full model. Analysts are interested in maximising the separation between the top deciles and the bottom deciles, thus models with greater variation are more desirable.<sup>44</sup> Finally, as measured by economists and other social scientists, a Gini coefficient was calculated to measure deviations from uniformity. A Gini coefficient of 0.231 was calculated for the demographic-only model, and a Gini coefficient of 0.336 was computed for the demographic and exographic

**Table 3:** Standardised betas for 15-variable solution

Variable	Standardised beta
Percentage of households with three or more cars	0.295
Average home price	0.193
Square root of percentage of homes with three or more cars	0.178
Square root of family income	0.053
Square root of average home price	0.179
<b>DEMO</b>	
Square root of percentage of adults who are widows	0.014
Percentage of households with three or more vehicles squared	0.121
Percentage of adults with a college education	0.027
Percentage of adults whose marital status is single	0.037
Percentage of households with three or more cars	0.295
Square root of precipitation of the MSA	0.035
Percentage of households with cars squared	0.029
Square root of climate rankings	0.062
<b>EXO</b>	
Average commute time squared	0.132
Population growth rate of the MSA squared	0.010
Population of the MSA squared	0.062
Total exographic variables	0.33
Total of all variables	1.10
Contribution of exographic variables (%)	23.2

**Table 4:** Contribution of exographics in response models

Number of variables	Number of exographics	Sum of exographic variables	Exographic beta weights as percentage of all beta weights (%)	Comparison of exographic to full model: restricted F-test*
5	3	0.158	67.2	F(3,33759) 186.5
10	4	0.266	33.1	F(4,33758) 180.2
15	6	0.330	23.1	F(6,33756) 203.2
20	7	1.129	25.2	F(7,33455) 214.2

\*Incremental variance explained by exographics. All *F*-tests significant at the 0.01 level.

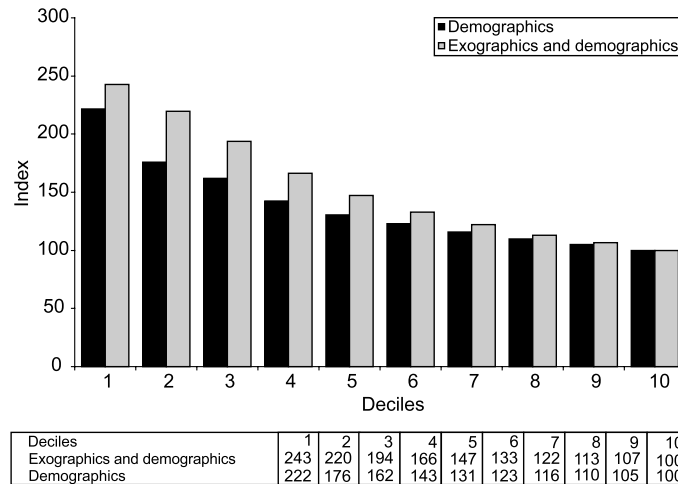
model. In a working paper, Greene and Milne<sup>45</sup> develop an approach to approximate the standard deviation for a Gini of 0.231 to be 0.0497, given assumptions of response rate of 3 per cent and sample size  $N = 35,000$ . Therefore the Gini index for the geodemographic and exographic model falls outside two standard deviations,  $p < 0.02$ . This suggests that for the expanded model, the expected decile responses are further removed from uniformity (predicting greater lift) and the model is more powerful in its predictive power.

Finally, to validate the model, a holdout file (cross-validation test) was scored for the demographic model and

the demographic and exographic model. A similar pattern of Gini coefficients were found (0.210 for demographic and 0.344 for demographic and exographic models), further supporting the claim that exographic data can be used to enhance model performance.

## DISCUSSION

The study described in this paper demonstrates that exographic variables can be useful for enhancing response models and customer profiles with research methodologies currently utilised by database marketers. These data items should be most beneficial during the customer prospecting and acquisition



**Figure 2** Gains chart — Comparison of demographics and demographics + exographics

phases when other data elements (purchase behaviour and individual characteristics) are not available.

Theorising and collecting exographic data are not a trivial task. Simply recognising that proximity to the coastline influences the recreational interests of local inhabitants is not enough. It is still challenging to identify relevant data sources and measurement scales to build a variable that will reflect and predict the characteristic of interest. The work necessary to create appropriate measurements is well worth the effort. Incorporation of exographics can greatly affect the explanatory power of response models used in database marketing.

When considering the use of exographics, the following approach is recommended:

- 1 For each exographic domain, carefully evaluate variables based on theoretical and managerial justification for initial selection.
- 2 Select appropriate level of scale to capture the phenomena.
- 3 Pick the level that maximises correlation with the response variable. The best level of scale will have

enough variation between geographical locations while maintaining enough homogeneity within geographical locations to be cost-effective.

Database marketing managers may find this paper useful in three ways. First, it suggests that investigating data sources on large-scale dimensions has incremental value beyond the data sources usually collected at household or neighbourhood level. Not only are these data sources potentially rich in predictive power, they are less costly, require less storage capacity, and are less expensive and less intrusive. Secondly, although not examined in this study, exographics may prove to be valuable across a larger portion of the customer relationship spectrum, not only during the early stages. Thirdly, marketing collateral, advertising copy and customer communications may also benefit from an understanding of exographic conditions.

### Limitations

This paper has limitations that should be recognised. First, the number of

exographic variables tested was limited by availability. These data items are not ubiquitous among data vendors, nor easily accessed electronically. Secondly, the methods chosen, although representative of the direct marketing industry, are certainly not exhaustive. Exographics may extend to alternative methodologies such as logit, structural equation modelling and factor analysis. Thirdly, the study investigated customers at the prospecting and acquisition phase of the customer relationship life cycle. Although we theorise that exographics would be most meaningful at this stage, it would be worthwhile to investigate their value at additional stages in the cycle. Fourthly, there is the possibility that external factors may have affected geographies differentially; however, it was not possible to control for these factors. Lastly, the data sets chosen reflect one specific industry. Additional data sets from a variety of industries should be investigated to augment external validity.

### **Future research**

This paper focused on customer acquisition and the value of exographic data during the earliest stages of customer relationships. Although it is expected that the value of exographic variables will diminish as behavioural data and personal demographics are acquired, it is believed that there may still be value in the later stages. This is a worthy area of research; data sources to explain consumer behaviour and which are less intrusive on consumer privacy need to be investigated. Exographic data should have value not only in terms of identifying customers with greater purchase propensity but also in the design of advertisements and communications. This is an open area for investigation. The choice of exographic variables in this study reflects theoretical concepts from

geography, sociology and urban development. Other areas that could be examined are history, political science, anthropology and the humanities. Essentially, this paper contends that understanding of consumer behaviour can be increased by examining more fundamental areas of human behaviour.

The intent of this paper is to encourage investigation outside the traditional realms of purchase behaviour, household and neighbourhood characteristics. It is hoped that this paper will encourage researchers to explore additional data sources that will enhance the predictive capabilities of response models and increase model performance in terms of identifying most profitable prospects. Creating a topology to select exographics by product category would be an interesting endeavour. The current suggested method is to use experience, intuition and observations gleaned from the literature.

It is expected that other areas of marketing communication, shelf space allocation, retail site selection, store atmospherics, front-line employee interactions etc, already utilise exographics to some degree, in a less formal setting. For example, an ethnic restaurant developer has to be aware of the population characteristics and traffic of the trade area from which it draws. A local sporting goods store may benefit by stocking products that represent the closest professional team or local university. Marketing collateral, developed by local businesses, may prove to be more responsive by incorporating local, topological, meteorological or natural resources. Beyond traditional database marketing, global database marketers should find many interesting and useful relationships between exographic variables and customer response. It is suspected that exographics will be shown to be useful for

differentiating brands and predicting brand loyalty.

## CONCLUSIONS

The modelling efforts in direct marketing have been focused on micro measures and measuring consumer retention. Yet, as suggested, with geographically larger market places, greater cultural and language diversity and varying sets of activities that people engage in, there are valid reasons to move beyond the trees and consider the forest. Exographics are a source of data that can assist academic scholars and direct marketing managers in their investigations and practice.

The exographic table presented in this paper highlights three domains: nature, culture and built environments, which can be used to identify exographics at various levels of analysis. This exographic approach provides a rationale for the use of this class of variables by making the link between the domains and the impact on consumer behaviour.

This paper suggests that the advantage of exographics is primarily in the prospecting and acquisition stages of the customer relationship. Compared with other sources of variables, exographics can be an effective, low-cost source of variables for targeting prospects, without violating the privacy of consumers. The empirical demonstration of customer acquisition using motor insurance data helps to lend support to the incremental advantage of including exographics in predictive models. Given that external marketing forces play a significant role in explaining and predicting consumer motivation, it would benefit direct marketers to identify and measure relevant environmental conditions.

## References

- 1 Andrew, R. and Curvim, I. (2003) 'A comparison of segment retention criteria for finite mixture logit models', *Journal of Market Research*, Vol. 40, No. 2, pp. 239–243.
- 2 Courtheoux, R. (2004) 'Analysis of list segmentation efficacy and campaign optimization', *Journal of Interactive Marketing*, Vol. 18, No. 1, pp. 70–89.
- 3 Curry, D. (1993) 'The new marketing research systems: How to use strategic database information for better marketing decisions', Wiley, New York, NY.
- 4 Freeman, A. and Poulin, K. (2000) 'Developing a marketing geographic segmentation system using SAS software'. Poster — SUIG (SAS Users Group) Paper 229–28; Available at: [www2.sas.com/proceedings/sugi28/229-28.pdf](http://www2.sas.com/proceedings/sugi28/229-28.pdf).
- 5 Andrew and Curvim (2003) *op. cit.*
- 6 Heidi, H., Leifling, P. and Witlink, D. (2002) 'Semiparametric analysis to estimate the deal effect curve'. *Journal of Market Research*, Vol. 38, No. 2, pp. 197–215.
- 7 Naik, P., Hagerty, M. and Tsai, C. (2000) 'A new dimensional reduction approach for data rich marketing environments: Slice inverse regression', *Journal of Market Research*, Vol. 37, No. 1, pp. 88–101.
- 8 Griffiths, W., Hill, R. and Judge, G. (1993) 'Learning and practicing econometrics', John Wiley & Sons, Inc., New York, NY.
- 9 Solberg, E. and Wong, D. (1992) 'Family time use: Leisure, home, production, market work related travel', *Journal of Human Resources*, Vol. 27, No. 3, pp. 485–510.
- 10 Jackson, R. and Wang, P. (1996) 'Strategic database marketing', NTC Business Books, Chicago, IL.
- 11 Schmid, J. and Weber, A. (1998) 'Desktop database marketing', NTC Business Books, Chicago, IL.
- 12 David Shepard Associates (1999) 'The new direct marketing', 3rd edn, McGraw-Hill, New York, NY.
- 13 Sach, R. (1988) 'The consumer's world: Place as context', *Annals of the Association of American Geographers*, Vol. 78, No. 4, pp. 642–664.
- 14 Zielensky, W. (1973) 'The cultural geography of the United States', Prentice Hall, Englewood Cliffs, NJ.
- 15 Webber, R. (2004) 'Designing geodemographic classifications to meet contemporary business needs', *The International Journal of the Institute of Direct Marketing*, Vol. 5, No. 3, pp. 219–237.
- 16 Steenburgh, T., Ainslie, A. and Engeberson P. (2003) 'Massively categorically variables revealing the information in zipcodes', *Marketing Science*, Vol. 22, No. 1, pp. 40–59.
- 17 Metropolitan statistical area (MSA): a geographic entity defined by the Office of Management and Budget for use by federal statistical agencies, based on the concept of a core area with a large population nucleus, plus adjacent communities having a high degree of economic and social integration with that core.
- 18 Curry (1993) *op. cit.*
- 19 Garreau, J. (1988) 'Edge cities: Life on the new frontier', Doubleday, New York, NY.
- 20 Garreau, J. (1981) 'The nine nations of North America', Houghton Mifflin, Boston, MA.

- 21 Hughes, A. (1996) 'The complete database marketer' 2nd edn, McGraw-Hill, New York, NY.
- 22 Weiss, M. (1988) 'The clustering of America'. Harper and Row, New York, NY.
- 23 Weiss, M. (1994) 'Latitudes and attitudes: An atlas of American tastes, trends, politics and passions', Little Brown and Company, Boston, MA.
- 24 Weiss, M. (2000) 'The clustered world', Little Brown and Company, Boston, MA.
- 25 Rook, D. (1987) 'The buying impulse', *Journal of Consumer Research*, Vol. 14, No. 2, pp. 189-199.
- 26 Hubbard, R. (1978) 'A review of selected factors conditioning consumer travel behavior', *Journal of Consumer Research*, Vol. 5, No. 1, pp. 1-20.
- 27 David Shepard Associates (1999) *op. cit.*
- 28 Grier, S. A. and Deshpande, R. (2001) 'Social dimensions of consumer distinctiveness: The influence of social status on group identity and advertising persuasion', *Journal of Market Research*, Vol. 38, No. 2, pp. 216-224.
- 29 Reinartz, W. and Kumar, V. (2003) 'The impact of customer relationship characteristics on profitable lifetime duration', *Journal of Marketing*, Vol. 67, No. 1, pp. 77-99.
- 30 Zeithaml, V. and Bitner, M. (2000) 'Service marketing', McGraw-Hill, Boston, MA.
- 31 Reinartz and Kumar (2003) *op. cit.*
- 32 Weiss (1988) *op. cit.*
- 33 Courtheoux (2004) *op. cit.*
- 34 Levin, N and Zahavi, J. (1998) 'Continuous predictive modeling: a comparative analysis', *Journal of Interactive Marketing*, Vol. 12, No. 2, pp. 5-22.
- 35 Levin, N and Zahavi, J. (2001) 'Predictive modeling using segmentation', *Journal of Interactive Marketing*, Vol. 15, No. 2, pp. 2-22.
- 36 Ratner, B. (2002) 'Rapid statistical calculations for determining the success of marketing campaigns', *Journal of Targeting, Measurement and Analysis for Marketing*, Vol. 10, No. 4, pp. 385-390.
- 37 Roberts, M. and Berger, P. (1999) 'Direct marketing management', Prentice Hall, Upper Saddle River, NJ.
- 38 David Shepard Associates (1999) *op. cit.*
- 39 Leahy, K. (1992) 'Beyond prediction: The nature and consequence of model meddling in direct response', *Journal of Direct Marketing*, Vol. 6, No. 4, pp. 9-16.
- 40 Malthouse, E. (1999) 'Ridge regression and direct marketing scoring models', *Journal of Interactive Marketing*, Vol. 13, No. 4, pp. 10-23.
- 41 Van Der Scheer, H. (1998) 'Quantitative approaches for profit maximization in direct marketing', Rijksuniversiteit Groningen, Groningen, The Netherlands.
- 42 Rook (1987) *op. cit.*
- 43 Roberts and Berger (1999) *op. cit.*
- 44 Ratner (2002) *op. cit.*
- 45 Greene, H. and Milne, G. (2005) 'Measuring standard deviations for Gini'. Working Paper, University of Massachusetts, Amherst, MA.

Copyright of *Journal of Targeting, Measurement & Analysis for Marketing* is the property of Palgrave Macmillan Ltd. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.