METHODOLOGICAL PROBLEMS IN THE CALIBRATION OF SALES FORECASTING MODELS

Michael F. Goodchild
Department of Geography
University of Western Ontario
London, Ontario N6A 5C2

1. INTRODUCTION

Within the general field of Marketing there are several areas where geographers have been very successful over the past few years at applying their particular bag of tools. The combination of computer cartography and geographical data processing, modelling of spatial behaviour and interaction, and location-allocation gives the applied geographer an impressive array of methods which is in many cases far beyond the sophistication of the client. The last ten years have seen a dramatic increase in the number of private sector geographers offering such tools and in the willingness of the private sector to make use of them.

The feasibility of a particular retail location depends directly on sales, and so the ability to forecast sales accurately is crucial to any retailer. This is as true for a company in the process of closing or remodelling stores as it is for one in the process of acquiring new sites. In the case of shopping goods many of the major determinants of sales, such as pricing and advertising, are outside the geographer's normal area of expertise, and others, such as the delimitation of trade area and the prediction of multiple-purpose trips, are notoriously difficult. Sales in the convenience area on the other hand tend to be determined by more readily delimited trade areas and by measurable physical characteristics of each store.

The greatest success in convenience store sales forecasting has undoubtedly been in gasoline operations, for two reasons. First, in North America gasoline is probably the most ubiquitous good of all, and its associated spatial behaviour probably the simplest to model. Second, statistics are readily available on sales for all locations irrespective of ownership, which vastly increases available sample size.

Next in order of suitability are convenience food stores. Spatial behaviour is more complex, but customers can be successfully divided into those making single-purpose trips from home and those making simple stops. Multi-purpose trips involving convenience stores as major destinations are not common, and chain loyalty is low. Unfortunately convenience store chains do not normally have access to competitors' sales figures, so sample sizes tend to be low also.

Other areas with potential include banks, but here behaviour is complicated by a high degree of chain loyalty and even branch loyalty. Higher order retail functions are complicated by product diversity, and forecasting is made difficult by low sample size.

This paper reports on experience with the modelling of convenience food store sales for a major Canadian chain, and describes some of the methodological difficulties which arise in the study. Many of these are general to the whole area of sales forecasting, and indicate the care with which much of the work being done in this area should be interpreted.

2. SAMPLE SIZE

Thousands of factors control the aggregate weekly sales of a convenience food store. These include minute details of the exterior and interior layout of the store, and the surrounding market area and traffic conditions. Many of these are outside the domain of a sales forecasting model: a company seeking to forecast sales for a potential store in a new subdivision is unlikely to consider or even be able to predict the interior layout or the goods offered, for example. Only the physical characteristics of the structure and the permanent features of the site and neighbourhood are likely to be relevant. We have of course already limited the predictive power of any forecasting model by deliberately ignoring in this way a large number of important factors.

The remaining set of factors is still very large. For example sales are influenced by adjacent stores, particularly since many convenience food stores are located in small convenience shopping centres. Dry cleaning is a common neighbouring store type. There are perhaps 50 major classes of stores which are common neighbours of convenience food stores, each with a different impact on sales, and each therefore representing a potential factor. In this study we grouped factors into nine broad categories, as follows:

(1) Neighbourhood: age, presence of industry, institutions and shopping centres
(2) Site: market gravitation, gasoline operation, adjacent tenants
(3) Access: number of curb cuts, median obstructions, stop signs
(4) Parking: number of spaces, front or side, fire corridor
(5) Traffic: counts on both streets, type of traffic

(6) Visibility: facias, poles, obstructions

(7) Store: inside and outside appearance

(8) Population: number of households of various types in trade area

(9) Competition: location and nature of competing stores

The factors within each group are examples: in all, 82 items were collected for each store.

The problem with sample size is that if 82 factors influence sales, it is fundamentally impossible to separate and measure the contribution of each factor without information on sales at at least 83 locations. This assumes of course that no two stores operate in identical circumstances as defined by the 82 factors. However to be of much general use, and not overly dependent on the particular sample of cases chosen, it is necessary to calibrate a model of m variables on many more than m + 1 cases.

Three additional problems arise at this point. First, the convenience food store market is highly regional, to the extent that stores in Vancouver, for example, operate in a very different context than those of the same chain in Toronto. Competition tends to operate at the regional scale: Vancouver might have two major competing chains and Toronto only one. Advertising is also regional, and the chain's image may be affected regionally by historical factors. In turn, there is little reason to suppose that a sales forecasting model developed in the Toronto market will operate successfully in the Vancouver market, or vice versa. In this study mean sales by city varied between a high of 126% of the general mean in Calgary and a low of 66% in Saskatoon, yet these cities were somewhat similar in general economic conditions during the study period.

Second, past sales figures are available only for the chain itself. This is a severe limitation on sample size. For example a company seeking to calibrate a sales model in the Toronto market for gasoline outlets would have a potential sample of about 1500 sites with known sales figures, whereas a very similar model of convenience food stores would have a potential sample of no more than 100.

Finally, the problem is complicated further by the number of ways in which factors can combine. An additive model assumes that a given factor contributes the same amount to sales independently of the other characteristics of the store. For example a school nearby might always contribute $1000 per week to sales, irrespective of the other conditions at the store. Mathematically, an additive model is written:

$$y = a_0 + \sum_{i=1}^{m} a_i x_i$$

where y is the sales forecast

a_o is a constant, the base if all factors are zero

x_i is the volume of the ith factor

m is the number of factors influencing sales

a_i is the response of sales to a unit increase in the value of the ith factor.

It is also possible for factors to combine multiplicatively, in which case the same factor always contributes the same percentage change in sales. The appropriate type of model would be of the form:

$$y = b_o \prod_{i=1}^{m} x_i^{b_i}$$

It is also possible that some factors would have additive effects and some multiplicative. Other types of response are also conceivable. Sales may be affected if both of two different factors are present, but not if either is present by itself, or if neither is present. Any pair of factors may combine in this way. In effect, uncertainty about the form that the model can take adds additional degrees of freedom, and can only be resolved by using additional samples. In summary, there are major difficulties in finding sufficient numbers of convenience food stores or higher order functions to calibrate elaborate sales forecasting models of the type commonly applied to gasoline outlets.

In this study we were able to collect data for 177 stores in 8 markets. This would limit the number of factors which could be successfully incorporated in a nationwide model to perhaps 50. The largest city sample was Calgary at 32, which would allow a useful model of no more than 10 variables. However it is possible that markets in different cities are sufficiently similar to allow aggregation. An analysis of the data for the Ontario cities of Windsor, London, Toronto and Ottawa showed strong similarity in all respects between the markets in Windsor, London and Ottawa and these were combined to give a single sample of 52. In this paper discussion will focus on three models: Calgary ($n = 32$), Windsor/London/Ottawa ($n = 52$) and the national model ($n = 177$).

3. SALES DATA

Sales figures are highly unstable through time, for a number of reasons. Even under entirely uniform conditions there will be short-run variations in sales because the arrival of consumers is essentially a random process, as is their selection of purchases. Superimposed on this are longer term variations through the daily, weekly and seasonal cycles. Every store has a different pattern of variation, and the factors controlling weekday sales may be entirely different from those affecting weekend sales. Finally the lifetime of a convenience food store is relatively short, yet consumer behaviour patterns adjust only slowly to major changes. It takes up to a year for the sales of a store to stabilize following major change. As a result, at any given time a substantial proportion of a sample may be still under the influence of past openings or renovations, or anticipated future closure. The criterion used in this study was that a case was acceptable only if it had been in operation for a full year previous to the study period. The sales figures used were averages of weekly sales over a complete year.

4. SUBJECTIVE VARIABLES

There are perhaps three common purposes in developing sales forecasting models. The client may wish to be able to predict the effects of changes in the system, as reflected in changes in the variables included in the model. For example a gasoline operator may wish to be able to predict the effect on his sales of a new street median proposed by the city, or a new layout of pumps on his forecourt. Second, the client may wish to use the model to compare actual to predicted sales for each of his locations, in order to assess the performance of managers, identify problem stores etc. Finally, the intention may be to use the model in the real estate department to assess the feasibility of new locations, particularly potential leases in new developments. It was clear from the start that the last was the relevant alternative in this study.

At this point the differences between an academic exercise is model-building and the solution of a client's problem become clear. The model will have to be applied in the field by real estate department personnel who already have their own, albeit subjective and intuitive, methods of appraising sites and forecasting sales. The model will be acceptable in this context only if it appears to provide an improvement on existing methods, without at the same time creating the impression that these methods, which are highly personal and individual, are in any way incorrect or inadequate.

With this background, we can now ask what use if any should be made of variables which represent subjective assessments. For example, one of the 82 variables collected by the client was an assessment of "market gravitation". Each store was appraised either as "toward" or "away from", which represented a judgment about the relative positions of the store and its market area. As such, it may contain a wealth of experience gained over many years about the factors influencing store sales, and may by itself prove to be more useful for prediction than many more objective measures.

Another example would be the count of households in the trade area. In this study this was again made by members of the real estate department, who are accustomed to making such judgments daily. The academic inclination would be either to make an expensive survey of consumers in each market, or to apply some rule, such as a maximum driving time, which we know to be overly simplistic. The subjective judgment includes an assessment of competitive locations and again represents many years of accumulated experience. In essence we are trading off the costs of extensive data collection with the disadvantages associated with using subjective data in an objective model. In another sense we are replacing the objective reality of physical, measurable variables with the reality of the world as seen through the eyes of an experienced appraiser. A psychologist would of course find the latter reality more interesting than the former, but the inclination of an applied geographer is probably the reverse.

Perhaps the strongest argument against using subjective variables is the problem of consistency between different individuals. This was never considered a problem by the client who appeared convinced that any one of their staff would have delimited each trade area in the system in the same way.

5. SURROGATE VARIABLES

Many pairs of the variables used in the study are strong predictors of each other, which raises a number of issues. For example a one-way street (variables T11, T12) would not have a median obstruction (variables A6, A7), to the extent that one characteristic is a perfect predictor of the other. To a lesser extent the existence of a stop light in front of the store (A10) is an indication that the traffic count on the side street (T5) is likely to be high. In this sense many of the variables are to a greater or lesser extent surrogates of each other. The subjective variables are frequently surrogates for other objective variables, and surrogates are a useful way of avoiding excessive and expensive data collection. Finally, the presence of strong surrogate relationships means that despite the complexity of the marketplace and the large number of factors influencing consumer behaviour, it may still be possible to predict store sales accurately from a relatively small number of variables.

On the other hand the presence of strong correlations can make cause very difficult to ascribe. Suppose that two variables are predictors of each other, but that only one causes a response in sales. For example if a store is free-standing it also tends to have ample parking. The attraction to the consumer may be the parking, to the extent that an expansion of parking may produce extra sales. From the data, however, it is impossible to determine whether parking or being free-standing causes sales: both show a statistical correlation. The client who was led to demolish surrounding stores or move his own to a free-standing location would be making a serious mistake. Note that it is equally impossible in this situation for the consumer to devise an experiment to determine the reasons for his own behaviour.

6. NUMBER OF PREDICTORS

There are several issues involved in deciding on the number of predictors to be used in the final forecasting model. First, every additional variable adds to the reliability of the prediction, and the sales of n sample stores can be predicted perfectly with no more than n - 1 variables. On the other hand more predictors mean a more cumbersome model, one less likely to be acceptable to the real estate department in the field, and with greater data collection costs associated with its use. A small number of carefully chosen variables may successfully capture a large amount of the observed variation between the stores in the client's chain. Third, although high predictive power can be achieved with a complex model, the result may lack robustness and generality when used to predict the sales of new stores outside the original sample.

The strategy adopted in the study was to add variables sequentially, making extensive searches over variables and forms of combination at each step, so as to maximize the predictive power achieved. This necessitated a rather laborious repetition of the basic multiple regression routine. To deal with nominal variables (but not binary ones), at each step an analysis of variance was made of residuals from the previous step. If the nominal variable produced stronger explanation than any other variable, it was recoded into one or more binary variables and included at the next step. Searches were also made for interaction effects by using two-way analysis of variance, but no strong ones were found.

7. MODELS

The Calgary model uses six variables and is based on 32 cases. In order of decreasing predictive power they are as follows:

- gasoline operation on site (binary)
- major shopping area in market (binary)
- side street traffic transient (binary)
- number of curb cuts on side street
- standing alone (binary)
- number of lanes on main street

The second and third variables are both subjective judgments to some extent, the third being an appraisal of the type of traffic as predominantly residential, transient or industrial. The coefficient of the sixth variable is negative: a larger number of lanes predicts lower sales. None of the trade area variables appears in the model. This may indicate that the subjective assessments of trade area are of no value, or it may indicate the nature of consumer behaviour in the Calgary market.

The Ontario model is based on 52 cases and includes the following variables:

number of high density housing units in market area
number of low density housing units in market area
number of parking spaces at side of building
donut store adjacent (binary)
number of lanes on main street
number of lanes on side street
near or far corner (binary)
hospital in market area (binary)

There is a sharp contrast here with the Calgary model in the presence of two of the three housing unit variables. In general the medium density variable was not useful: this may have been due to difficulties of definition, or to the relatively small numbers of units in this category, or to the wide variation in socioeconomic characteristics of people in units which may range from subsidized row houses to luxury condominiums. The far corner proves to have an advantage over the near corner, as one might expect. Parking spaces at the side can be taken as a surrogate for the store standing alone and for high visibility.

In forecasting sales for the entire sample it was decided to remove the variation between cities by reducing observed sales figures to percentages of the mean sales for each city. Each variable then predicts a percentage rather than a dollar change in sales, relative to the city base. The model used the following nine variables, again in decreasing order of predictive power:

number of low density units
number of high density units
number of curb cuts on side street
gross leasable area of store
school in market area (binary)
near or far corner (binary)
number of medium density units
store stands alone (binary)
major shopping area in market (binary)

The presence of gross leasable area presumably reflects a nation-wide variation in this parameter which was not observed in either the Calgary or Windsor/London/Ottawa samples.

8. APPLICATION

As noted above, it was clear that the client of this study planned to use the results in assessing the feasibility of potential new locations. Previous studies had been made using much smaller samples and with a much more limited set of variables, and the company had been using one of these, a three-variable model incorporating the main street traffic count, market area household count, and subjective judgments about visibility and competition, for some time. The new model was planned as a replacement, and was constrained by the need to be acceptable to the staff.

However it became abundantly clear that there would be difficulties in using the new models in this mode. Consider for example the use of the number of curb cuts on the side street in the national model. This proves to be a highly successful predictor of sales, but it is quite clear that its effect on sales is no more than partially causal. There is no implication that an operator who carved another cut in his side street curb would enjoy the predicted increment to sales. Instead the variable acts as a surrogate for a number of causal variables, including side street traffic density, all of which would have to increase to cause the predicted change.

The difficulty arises when the person using the model has some control on the design of the store, and fails to distinguish between causality and the predictive power of surrogate variables. To be used successfully, it is essential that the model function as a black box, and that its components have no influence on the design of stores, since the model assumes that the observed correlations between variables will continue to be valid for potential sites. Discussions of this problem with the client led to two suggestions: that the model could be resident in a calculator or computer program so that the user would not be aware of its components, and that the model be known only to the head office and used by them to evaluate proposals from the field.

Precisely the same problem arises in using models of this type to forecast the effects of changes to existing stores, where each component of the model is interpreted as having a causal influence on the predicted sales.

9. CONCLUSIONS

Sales forecasting models have been applied with great success to the management of retail gasoline operations, and there is currently a move to apply the same methods to convenience food stores, banks and other low-order functions. The influence of spatial behaviour on sales of convenience goods implies that geographers have a great deal to contribute in this area. However a number of methodological problems suggest that much of this work must be interpreted very carefully, and that models of this type must be used with caution and in clearly defined ways. In an ideal world, geographical techniques of spatial behaviour analysis could be used to great advantage. In reality, academic ideals must be compromised because of the limited resources of the client and the need to make the model acceptable to the client's organization.

Academics who try to apply their ideas and models to the real world often think of the process in two stages: a first stage of scientific, objective modelling in which the main concern is to replicate reality as closely as possible, and a second in which the results are unfortunately modified or rejected entirely by constraints often described as "politics". The approach taken in this paper has been to allow these "political" concerns to affect the modelling directly, as for example in the use of subjective variables. To carry this a little further, it seems that applied research may often have to compromise the normal standards of statistical research in order to produce work which is both acceptable and useful to a client.