Modelling Meso-Level Marketing Phenomena Using Geographically Weighted Regression

Hongwei Lu, Mark Uncles, Lihua Zhao, Gary Gregory, University of New South Wales

Abstract

Marketers have an interest in targeting consumers in particular geographical areas. The effectiveness with which they can do this depends on an understanding of meso-level marketing phenomena. However, conventional multiple regression models can only address the issue at a global level (e.g., within a trading area), whereas it is desirable to examine spatial variation at a much more local level (e.g., within smaller geographical areas, such as city blocks and Census Collection Districts, CCDs). Therefore, Geographically Weighted Regression (GWR) from quantitative geography is introduced to solve the problem. Specifically, taking the sales of a European car brand as a case study, GWR is used to model the relationship between area brand performance (i.e., car sales in each CCD) and related marketing phenomena (e.g., retail outlet accessibility, competition intensity and brand ownership penetration). This analysis of meso-level marketing phenomena has implications for marketing strategy.

Introduction

From a marketing perspective, it is important to know how marketing phenomena vary over space and to be able to draw strategic inferences (e.g., competitive intensity tends to vary and this has implications for any strategy designed to exploit patterns of competitive intensity). A strategy is likely to be more effective if it is based on accurate spatial modelling. Currently, this is achieved using multiple regression models (MR), which help marketers examine meso-level marketing phenomena. However, to increase accuracy it is necessary to examine spatial variation at the level of small geographical areas, such as city blocks and Census Collection Districts (CCDs). For this purpose, Geographically Weighted Regression (GWR) is introduced into marketing research.

First, several influential spatial theories are reviewed – these provide a theoretical basis for understanding meso-level marketing phenomena and the inferred marketing strategies. Second, GWR is explained. Third, measurement issues and models used in this paper are illustrated, followed by a case study and brief discussion of new car sales in part of the Sydney metropolitan area.

Spatial Theories and Marketing Strategies

Two key theories have been used to study spatial aspects of store/product performance. First, those theories based around the gravity model, including the gravity model itself (Huff, 1964) and extensions (i.e., spatial interaction and spatial choice models), which suggest store/product performance within a trading area is a function of the store's attractiveness and the store's accessibility for local customers. Thus, to improve store/product performance, the retailer must consider ways to improve his attractiveness and accessibility. Second, by considering the influence of competitors, Hotelling (1929) and subsequent analysts (e.g., Brown, 1989), suggest that while a limited market must be shared by competitors, there might be positive effects on store/product performance arising from retail outlet agglomeration. These additional benefits derive from the larger numbers of consumers who are attracted by the breadth of merchandise and the convenience of product comparisons that can be made when stores agglomerate.

Based on these two theories, franchise distribution models have been developed (e.g., the location-allocation models presented by Ghosh and Craig, 1991 and Kaufmann and Rangan, 1990). In such models the goal is to maximise franchise performance by opening more franchisees/dealers (which improves accessibility and network coverage), while at the same time minimising losses from having more franchises/dealers in the system (which intensifies competition). The approach described here can be seen as a successor to these earlier studies. Essentially, a spatial dependence (distance-decay) effect is assumed to exist, whereby franchisees/dealers close to a neighbourhood will have greater impact on each other than those that are far apart. Similarly, consumers close to one another will have more of an effect on each other than those far apart. Conventional MR models do not accurately capture these effects and therefore attention is turned to GWR as an alternative.

It is noted that the approach contrasts with geo-demographic profiling, where study areas are classified and clustered into different groups. Such classifications help marketers to identify the complex geographies of product consumption, consequently helping them to understand local population behaviours, attitudes, and deprivations that are shaped by socio-economic landscapes (Harris *et al.,* 2005).

Geographically Weighted Regression

In quantitative geography, GWR is a relatively simple and practical technique used to estimate local parameters based on the use of a traditional regression framework in combination with kernel probability density estimation (Brunsdon *et al.*, 1999; Silverman, 1986; Fotheringham and Brunsdon, 1999; Fotheringham *et al*., 2002). Kernel probability density estimation is a non-parametric smoothing technique to estimate a local value based on values in adjacent points (distance-decay effect).

The objective of using GWR is to estimate a linear model (within each CCD) that relates the dependent variable to its determinants after taking into account spatial correlation among observations in neighbouring locations. GWR takes advantage of distance decay in the data (or spatial dependency in geographical terminology). Spatial dependence implies that data available in locations near the focal location are particularly informative about the relationship between the independent and the dependent variables in the focal location. When calculating estimates for a focal location, GWR gives more weight to data from closer locations than to data from more distant locations. It is assumed that the relative weight of the contributing locations decays at an empirically determined rate as their distance from the focal location increases. Statistically, the weighting matrix contains weights for all locations that are used in computation of the regression equation for the focal location (Fotheringham *et al.,* 2002). Using this geographic weighting matrix, we can obtain the weighted least squares estimates for any location (i.e., small area). Thus, during the modelling process, the data are in fact weighted twice, the local version of global model is weighted by least square estimation, then it is geographically weighted by GWR to allow for the local spatial effects to be measured (Fotheringham *et al.*, 2001).

Thus, as an exploration tool, GWR tell us: (a) if there is significant spatial variation with the studied independent variables (marketing phenomena in our case, such as accessibility, level of agglomeration, ownership density, etc.), and (b) what is the spatial variation pattern with each independent variable across the study areas. The significance of this is only beginning to be recognised by marketing researchers – e.g., Mittal *et al.* (2004) use the technique to identify and estimate non-stationary customer satisfaction, which cannot be directly measured across a spatially continuous market.

Methodology

New car sales of a European brand (Brand X) are examined as a case study. The study area is comprises a part of Sydney (shown in Figures 1-6). The unit of analysis is the CCD, which is widely used in meso-level analysis. In total, 4708 CCDs are studied. Based on the above review, we consider four independent variables: dealership accessibility, competition intensity, local brand ownership, and recent sales. Brand X sales in 2001 are taken as the dependent variable.

To calculate dealership accessibility, the number of available dealerships within a 10km travelling distance for Brand X is counted first. Then this figure is divided by the area of each CCD (square km) to obtain a measure of dealership accessibility. To measure the competition intensity of the local market (i.e., the average level of competition faced by each brand), the competition intensity index CI_i in each small area i is measured as:

$$CI_i = 1 - \frac{\sqrt{\frac{1}{N}\sum_{j=1}^{N}\left[(MS_{ij} - \overline{MS}_i)^2\right]}}{\overline{MS}_i}$$

i - the studied small area; j- the studied competitor;
MS_{ij} - the market share of Brand j in small area i;
$\overline{MS}_i$ - the normative (average) market share in small area i.

The idea of this index is to represent the ratio of the market share standard deviation of the competing brands to their average market share. This ratio is subtracted from one to derive the final measure of competition intensity. A higher value indicates more serious competition. Brand ownership is measured as the number of current Brand X owners in each CCD – this measures long-term sales effects. Recent sales are measured as sales in the last year – this highlights short-term sales effects.

Based on MR, at the global level, brand sales can be modelled as Equation (1):

Equation (1) $$BrandSales_i = a_0 + a_1 \bullet AC_i + a_2 \bullet CI_i + a_3 \bullet BO_i + a_4 \bullet RS_i$$

i - The studied small area (CCD); AC_i - Brand X dealership accessibility in CCD i;

CI_i - The competition intensity in CCD i; BO_i - Brand X ownership in CCD i;

RS_i - Recent sales of Brand X in CCD i.

In the global model (Equation (1)), the coefficients for each variable are constant across the studied CCDs. By contrast, with the local level GWR model (Equation (2)), the coefficients for each variable vary across the studied CCDs. While localized relationships are specified across CCDs, the global linear relationship is replaced by a non-linear relationship in a global view. Thus, the spatial drift from 'average' global relationships is modelled directly.

Equation (2) $$BrandSales_i = a_0 + a_1(i) \bullet AC_i + a_2(i) \bullet CI_i + a_3(i) \bullet BO_i + a_4(i) \bullet RS_i$$

Case Study: Brand X Sales in Sydney

Conventional MR is conducted to model global level relationships. Results show that the global model is significant with an adjusted R^2 of 0.52. The coefficient of each variable is significant except the intercept. Specifically, the global relationship is modelled as (Equation (3)):

Equation (3) $$BrandSales_i = -0.009 + 0.104 \bullet AC_i - 0.156 \bullet CI_i + 0.039 \bullet BO_i + 0.234 \bullet RS_i$$

In this global model, brand accessibility has a positive influence, which is consistent with the gravity model. Competition intensity has a negative effect. Also, comparatively, recent sales (short-term sales effects) have a stronger influence on sales than brand ownership in an area over time (long-term sales effects). Based on this global model, it is possible to compare the relative significance of each relationship, and perhaps infer some strategic responses (e.g., in terms of opening new dealerships).

However, do these rules apply in all the small areas studied? This question cannot be answered using the global regression and we must turn instead to GWR to model local relationships. Results from GWR show the adjusted R^2 is significantly improved to 0.84, with significant coefficients for each variable except recent sales.

The localized coefficients of each studied independent variable across CCDs are mapped in Figures 1 to 6. Firstly, the effectiveness of brand dealership accessibility varies from place to place (Figure 1). City centres (e.g., Sydney, North Sydney and Parramatta) have a higher positive influence in terms of dealership accessibility. In contrast, a negative influence can be found in areas that might be characterised as “traditional residential” (e.g., Ku-ring-gai) and as “non-English/European speaking areas” (e.g., Auburn, Ashfield, etc.). Comparing Figures 1 and 2, a high positive effect with very low brand dealership accessibility appears in Baulkham Hills. This suggests a new dealership is needed in this area.

Secondly, spatial variation is found in relation to competitive intensity (Figure 3). Specifically, a negative effect occurs in most areas, but a positive effect can be found in the Ku-ring-gai to North Sydney corridor. This might be because local residents have been “educated” in the types of car on offer from high previous and current levels of market competition. Thus, additional marketing effort by Brand X in this area (e.g., a bigger investment in direct marketing) might be more profitable than investing in a newer market (where it would take time and considerable expense to build awareness).

Thirdly, in most study areas there are positive long-term sales effects, especially in northern areas which might be seen as “traditional”, where the descendents of early European migrants dominate, often occupying larger land blocks (Figure 5). Comparatively, targeting the North-East would be more efficient than targeting the North-West, although currently brand ownership penetrations are almost equal. This is because higher levels of brand ownership penetration effectiveness are found in the North-East, which may be directly related to demographic features of the local resident. In contrast, short-term sales effects (i.e., recent sales) are not significant.

Discussion

Spatial variation – in terms of brand dealership accessibility, competition intensity and sales effects – is found across the study areas. Specifically, brand dealership accessibility, competition intensity, and recent sales (a short-term effect) tend to have a higher positive effect within city centre areas, whereas brand ownership in an area (a long-term effect) is more significant in traditional residential areas. These results are evident from using GWR, showing the potential value of this technique in marketing research. Such work can be extended by investigating other factors that might have an influence on sales in the context of car dealerships (e.g., a deeper appreciation of the influence of consumer demographics) and for spatially dispersed outlets more generally (e.g., in grocery retailing, real estate agencies, and retail banking).

Figure 1. Brand Dealership Accessibility Effectiveness

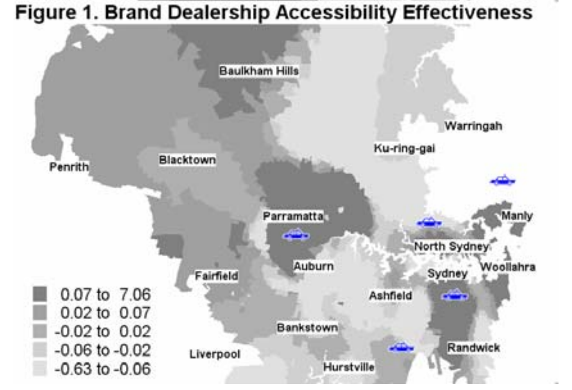


Figure 2. Brand Dealership Accessibility

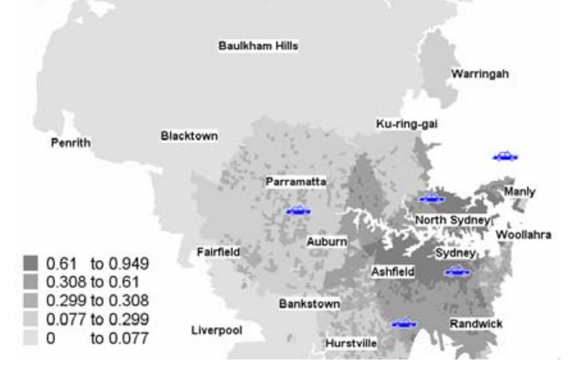


Figure 3. Competition Intensity Effectiveness

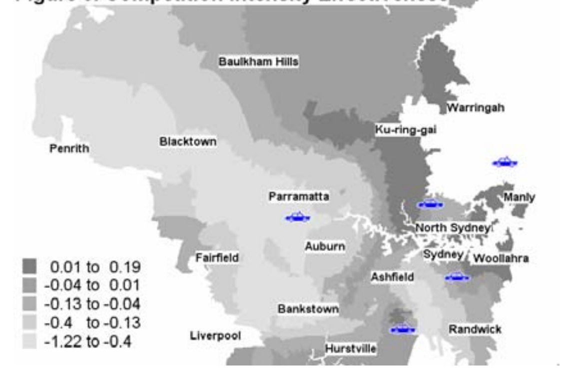


Figure 4. Competition Intensity

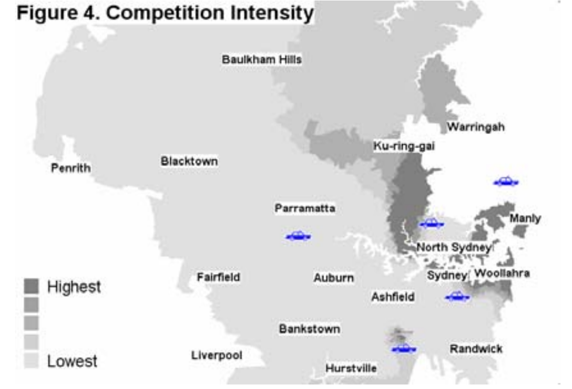


Figure 5. Brand Ownership Penetration Effectiveness

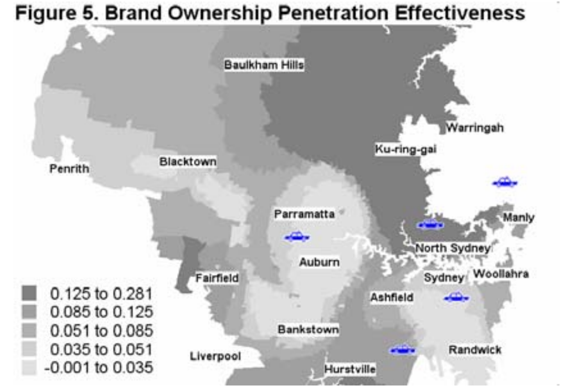
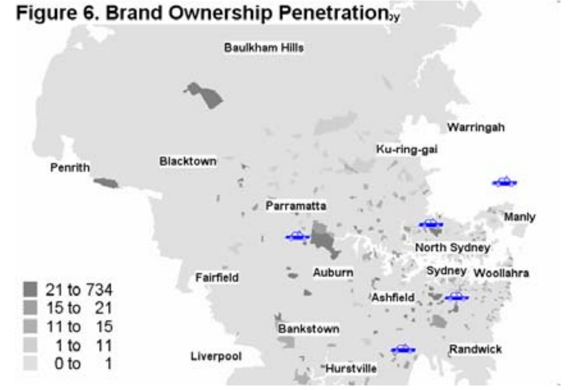


Figure 6. Brand Ownership Penetration



* The little car in each map represents the dealership location of Brand X.

References

Brown, S., 1989. Retail location theory: The legacy of Harold Hotelling. Journal of Retailing 65 (4), 450-470.

Brunsdon, C., Fotheringham, S., Charlton, M., 1999. Some notes on parametric significance tests for geographically weighed regression. Journal of Regional science 39 (3), 497-524.

Fotheringham, S. A., Brunsdon, C., Charlton, M., 2001. Spatial variations in school performance: a local analysis using Geographical Weighted Regression. Geographical & Environmental Modelling 5 (1), 43-66.

Fotheringham, S. A., Brunsdon, C., Charlton, M., 2002. Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. John Wiley & Sons, Chichester, England.

Fotheringham, S. A., Brunsdon, C., 1999. Local forms of spatial analysis. Geographical Analysis 31, 340-358.

Ghosh, A., Craig, C.S., 1991. FRANSYS: A franchise distribution system location model. Journal of Retailing 67 (4), 466-495.

Harris, R., Sleight, P., Webber, R., 2005. Geodemographics, GIS and Neighbourhood Targeting, John Wiley & Sons, Chichester, England.

Hotelling, H., 1929. Stability in competition. Economic Journal 39, 41-57.

Huff, D. I., 1964, Defining and estimating a trading area. Journal of Marketing, 28 (July), 34-38.

Kaufmann, P. J., Rangan, V. K., 1990. A model for managing system conflict during franchise expansion. Journal of Retailing 66, 155-73.

Mittal, V., Kamakura, W. A., Govind, R., 2004. Geographic patterns in customer service and satisfaction: An empirical investigation. Journal of Marketing 68 (3), 48-62.

Silverman, B. W., 1986. Density Estimation for Statistics and Data Analysis, Chapman and Hall, London.