The Impact of Local Demographics on Retail Centre Health in England and Wales

Karlo Lugomer[1] and Guy Lansley[2]

Department of Geography, University College London

January 08, 2016

Summary

Spatial Interaction Models are a widely accepted means of linking retail centres to their local customer catchments. However, without real consumer data to validate the findings, such models remain as mere estimations based on geographic population and store data. Using a spatial interaction model which accounts for both the residential and local workplace populations, the following study seeks to evaluate the extent of which demographic characteristics of estimated retail catchments are an effective predictor of retail health for retail centres in England and Wales. Overall, this case study explores the value of demographics from estimated centre catchments to retail planners.

KEYWORDS: retail health, geodemographics, catchment areas, spatial interaction model, census

1. Introduction

The recession and the changing nature of the UK retail market have led to the decline of many well-known retailers in recent years. Yet this decline has not been experienced evenly across the country. Certain retail centres have been more resilient, either due to their favourable local consumer characteristics or due to their composition of retailers.

The characteristics of the local population are known to greatly influence store performance, and consequently understanding the population is a crucial part of retail location planning (Birkin et. al., 2004). However, no work fully explores the relationship between the entire spectrum of census variables from retail centre catchment areas and the conditions of the corresponding retail environments. These conditions can be conceptualised in the form of retail health (or vitality), which can be generally described as a measure of the overall presence of favourable physical characteristics in retail environments (such as number of stores, presence of anchor stores, low vacancy rates, etc..).

The aim of this paper is to determine and quantify the impact of local demographics on the retail health, using a spatial interaction model to estimate catchment areas for retail centres in England and Wales.

2. Methodology

Data on 1,206 retail centres from England and Wales was provided by the Local Data Company (LDC). They also provided health indices or location scores whose methodology was compiled by Morgan Stanley Research (2014). The indices are more comprehensive, and pool a large number of variables including vacancy rates, the openings and closures of minor anchor retailers, presence of cinemas and other leisure units and department stores.

While retail performance is often measured by the turnover of individual stores (Campo et. al., 2000), such data is usually not accessible across an entire spectrum of stores in a centre. Therefore, for this

[1] karlo.lugomer.14@ucl.ac.uk
[2] g.lansley@ucl.ac.uk

paper we are considering retail centre health, a proxy indicator built from several variables on the local retail environment, notably including vacancy rates.

In order to investigate the relationship between catchment area demographics and retail centre health, we first developed a spatial interaction model to define catchments, and we subsequently used regression models to explore the relationships. The residential population data was obtained from 2011 Census data at the Lower Super Output Area (LSOA) level. In addition, we also considered the workplace population using Workplace Zones (WZs) also recorded from the 2011 Census, as local workers may also patronise proximal retail centres.

2.1. Spatial interaction modeling

A spatial interaction model (SIM) was implemented to estimate the catchment areas of 1206 retail, generated separately for residents (from LSOAs) and workforce (from WZs).

Catchments were created by adapting a Huff's probabilistic model (Equation 1) derived by Dolega et. al. (2016). The model attempts to identify the probability (P_{ij}) that residents from LSOA i or worker from WZ i will shop at each retail centre j. These could be used to determine the population likely to visit each centre.

$$P_{ij} = \frac{A_j^{\alpha_k} d_{ij}^{-\beta_{sj}}}{\sum_{j=1}^{n} A_j^{\alpha_k} d_{ij}^{-\beta_{sj}}} \qquad \textbf{(1)}$$

A_j is the measure of attractiveness of the retail centre j;
d_{ij} is the distance from i to j;
α_k is the attractiveness parameter dependent on the distance threshold of 500 metres;
β_{sj} is the distance decay parameter dependent on the hierarchical class to which retail centre j belongs: five clusters of retail centres based on their attractiveness scores were found by k-means clustering algorithm and were assigned different distance decay parameters accordingly

The attractiveness of retail centres measure was defined by Dolega et. al. (2016) as a sum of several indicators as described below:

$$\mathrm{A_j} = \left(S_j - V_j\right) + RM_j + L_j + An_j \qquad \textbf{(2)}$$

Where:

S_j is the total number of retail centre units;
V_j is the total number of vacant units;
RM_j is the index of retail diversity;
L_j is the proportion of leisure units;
An_j is the proportion of anchor retailers.

The retail diversity index (RM_j), compiled by Oxford Institute of Retail Management (2013), measures the presence of up to 273 unique categories of retail business at each centre.

2.2. Regression models

Following the creation of retail catchments, the compositions of residents and workers estimated to be probable patrons of each retail centre were calculated. This data could be used to establish the relationship between catchment demographics and retail health of retail centres. Pearson's correlation coefficients were calculated to determine the potentially most relevant residential and workplace variables for the regression procedure. The best subsets regression was then used in order to find the multiple regression model which contains the optimum combination of variables at explaining the variance of the Health Index across the retail centres. Finally, a geographically weighted regression (GWR) model was also tested to account for spatial variation of the coefficient weights. A GWR is an exploratory technique mainly intended to indicate the extent of which locally weighted regression coefficients move away from their global values across a spatial dataset (Fortheringham *et al*, 2003).

3. Results

The residential customer catchments generated for retail centres in England and Wales are shown in Figure 1. In the map LSOAs are only allocated to one catchment area. However, the subsequent analysis allows catchments to overlap based on the probability scores from the spatial interaction model. The created catchments reflect the complexity of retail geography, which is especially true in the case of Greater London.

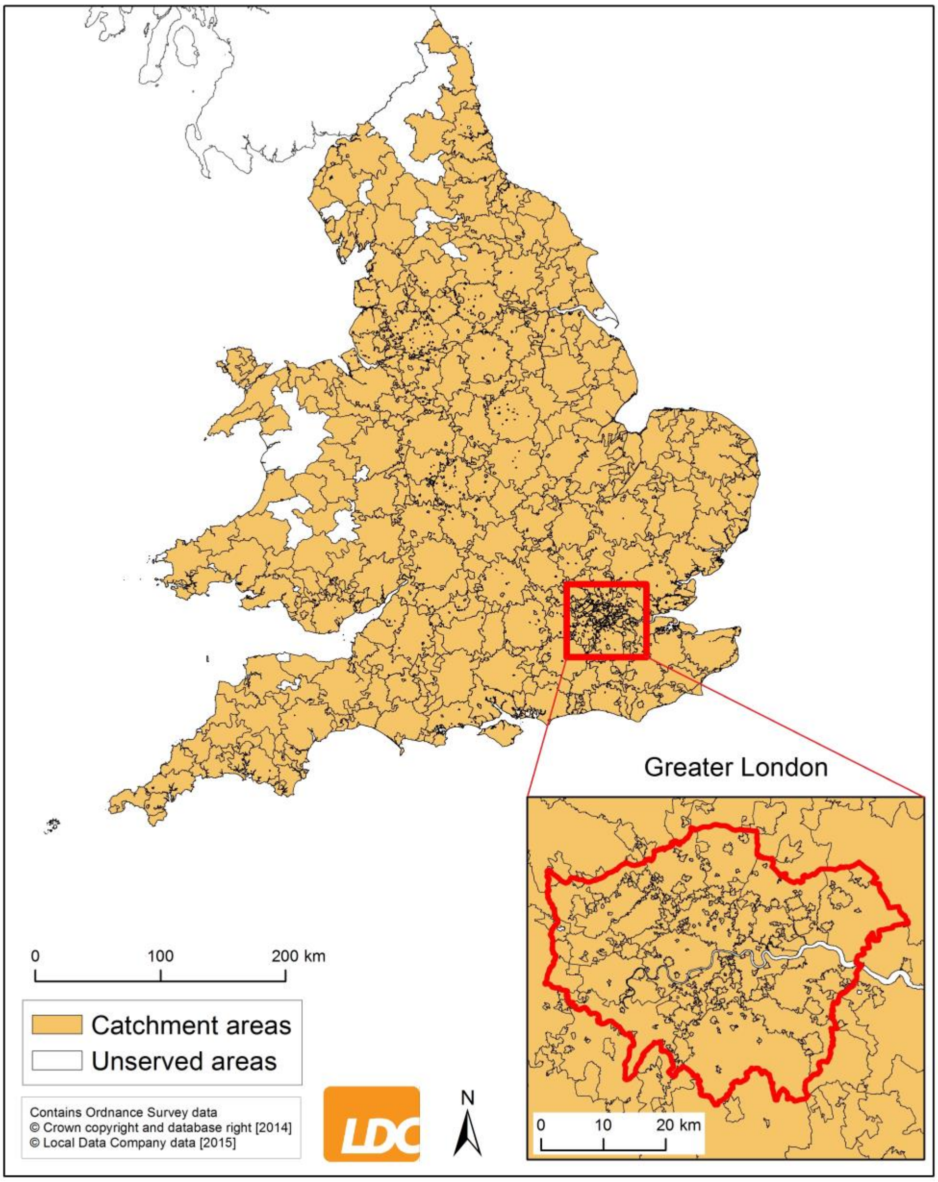

Figure 1: Residential catchment areas of retail centres in England and Wales

3.1. Identification and assessment of the demographics – retail health relationship

Pearson's correlation coefficients for twenty highest correlated demographic variables with the Morgan Stanley Health Index are laid out in the Table 1. The Health Index proved to have significant relationship ($\alpha = 0.05$) with Census variables, with nine of the variables sharing a correlation coefficient greater than 0.4. Our model found that both residential and workplace population statistics were similarly strong predictors of retail health. In general, socioeconomic and educational variables prevail as the best predictors.

Table 1: Predictors with highest Pearson correlations with Health Index. (RES = residential variable; WZ = workplace zone variable)

Rank	Predictor	r
1	% RES Semi-routine workers	-0.451
2	% WZ Qualification level 4	0.447
3	% RES Routine workers	-0.431
4	% RES Lower supervisory and technical occupations	-0.422
5	% RES Qualification level 4	0.422
6	% WZ No qualifications	-0.413
7	% WZ AB social grade	0.413
8	% WZ Semi-routine workers	-0.411
9	% WZ Qualification level 1, 2 or Apprenticeship	-0.402
10	% WZ Lower managerial and professional occupations	0.395
11	% RES AB social grade	0.393
12	% RES manufacturing workers	-0.389
13	% WZ DE social grade	-0.379
14	% RES Information/finance/real estate/science/professional workers	0.378
15	% RES Higher managerial and higher professional workers	0.377
16	% WZ Higher managerial and higher professional workers	0.377
17	% RES C2 social grade	-0.372
18	% RES Very good or good health	0.369
19	% RES Cohabiting with dependent children	-0.367
20	% WZ Business, media and public service professionals	0.366

data source: Geolytix, 2013; Nomis, 2013, 2014; LDC, 2015
Note: All of the correlations shown in the table are significant at 0.05 level.

The variables were subsequently entered into a regression procedure which checked them for their appropriateness in a multivariate model. The results of the final three-variable multiple regression model are shown in Table 2. The model explains 23.04% of the variance of Health Index and makes use of a mixture of NS-SEC, industry of employment and educational level variables.

The R^2 values are rather high given the fact that the demographic characteristics of SIM estimated catchment areas are not the only factors influencing the vitality of retail environments. However, the model also identifies that demographics alone are not an accurate means estimating retail health. Instead the health of a high street, and also retail performance, is a determinant of an intricate range of influencers which will vary across British retail centres.

Table 2: Model statistics for three-predictor model

Coefficients statistics				
Term	Coefficient value	Standardised error	P-Value	VIF
Constant	-1.5464	2.9136	0.596	
RES Semi-routine workers	-0.5393	0.1292	0.000	3.33
WZ Qualifications level 4	0.2494	0.0460	0.000	2.65
RES Manufacturing workers	-0.2414	0.0776	0.002	2.04

Model Summary			
R^2	Adjusted R^2	Joint Wald statistic	Koenker (BP) statistic
0.2324	0.2304	0.0000	0.0000

3.2. GWR model

Finally, geographically weighted regression (GWR) was explored to observe if accounting for local variances would create a more explanatory model. In the example we used the percent semi-routine workers as the predictor variable. The residuals for GWR model were not spatially autocorrelated, indicated by insignificant Moran's I (I = 0.001, p = 0.367) (Anselin, 1995). The improvement of the GWR over a bivariate linear regression model was marginal, with R squared value becoming just 0.88 percentage points higher (R^2 = 21,18%). However, spatial variation of goodness of fit shows that the local R^2 value varies between 0.109 and 0.240 across the retail centres, a spectrum of variance which a simple linear regression neglects.

4. Conclusions

There is a significant linear relationship between local demographics characteristics and the health of retail centres when considering spatial interaction model estimated customer catchments. The most important census predictors of retail health can be drawn from socio-economic variables and their associated proxies. Whilst, age, life stage, ethnicity and other groups of variables were found to be of negligible importance in our models. This research also found that both the workplace and residential statistics were similarly useful predictors of retail health. This was presumably due to the high correspondence between travel to work areas and retail catchments.

However, it is also important to note that a high proportion of variance in retail health across England and Wales cannot be predicted from inferred catchment data when using linear health indices. Retail health is an intricate concept that cannot be fully encapsulated by a universal index, nor be predicted by the composition of local demographics alone. Moreover, the full implications of retail centre health on individual outlets are not fully understood. Therefore, there is scope for further study into the relationship between retail health and the revenue of stores, and how the association varies across high streets and between different types of retailers too. Whilst it was possible to identify a linear correspondence between retail centre health and local demographic characteristics, it has not yet been possible to model the full complexities of retail geography due to data availability.

5. Acknowledgements

This work is funded by the UK ESRC Consumer Data Research Centre (CDRC) grant reference ES/L011840/1. We would like to thank Matthew Hopkinson and Ronald Nyakairu from the Local Data Company for granting us access to their retail centre data and providing advice throughout the research procedure. Gratitude also goes to Les Dolega of University of Liverpool for his useful advice on spatial interaction modelling.

6. Biography

Karlo Lugomer is a PhD student in Geography at UCL. His research interests incorporate GIS applications and quantitative methods in human geography, with particular focus on the geodemographics and population geography, tourism, urban and rural geography. He previously studied Geography BSc at University of Zagreb and Geospatial Analysis MSc at UCL.

Guy Lansley is a Research Associate at the Consumer Data Research Centre, UCL, an ESRC Data Investment. His previous research at UCL has included exploring the temporal geo-demographics derived from social media data, and identifying socio-spatial patterns in car model ownership in conjunction with the Department for Transport. His current work entails exploring population data derived from large consumer datasets.

References

Anselin, L. (1995). Local Indicators of Spatial Association-LISA. *Geographical Analysis*. 27; 93-115.

Birkin, M., Clarke, G., Clarke, M., Culf, R. (2004). Using spatial models to solve difficult retail location problems. In: Stillwell J, Clarke G (eds) *Applied GIS and spatial analysis*. Wiley, Chichester. 35–54.

Campo, K., Gijsbrechts, E., Goossens, T., Verhetsel, A. (2000). The impact of location factors on the attractiveness and optimal space shares of product categories. *Journal of Research in Marketing*, 17; 255–279.

Dolega, L., Pavlis, M., Singleton, A. (2016). Estimating attractiveness, hierarchy and catchment area extents for a national set of retail centre agglomerations. *Journal of Retailing and Consumer Services*, 28; 78–90.

Fotheringham, A.S., Brunsdon, C. and Charlton, M. (2003) Geographically weighted regression: the analysis of spatially varying relationships. Wiley, Chichester.

Geolytix (2013). Open Census Pack. Available at: http://geolytix.co.uk/geodata/ [2 August 2015].

Local Data Company (2015). http://www.localdatacompany.com/ [10 April 2015].

Morgan Stanley Research (2014). UK General Retail 'Dying High Streets' – An updated exposure guide.

Nomis (2013). Census 2011 Table Links - Quick Statistics. Available at: https://www.nomisweb.co.uk/census/2011/quick_statistics. [2 August 2015].

Nomis (2014). Quick Statistics - Census 2011 Table Links: Approximated Social Grade and Distance Traveled to Work. Available at: https://www.nomisweb.co.uk/census/2011/quick_statistics [2 August 2015].

Oxford Institute of Retail Management, 2013. Diversity and the UKs High Streets. Available at: http://oxford-institute.sbsblogs.co.uk/2013/07/19/diversity-and-the-uks-high-streets/ [1 August 2015].