How does competition impact grocery click and collect performance?

Alec Davies*[1], Les Dolega†[1], Dani Arribas-Bel‡[1] and Alex Singleton§[1]

[1]Department of Geography and Planning, University of Liverpool

April 2, 2017

Summary

Grocery click and collect is a relatively new service being offered in the supermarket industry. Little is known of how competition impacts the service. Typically catchments are estimated using drive distances which fail to consider regional disparities. Using the empirically tested gravitational estimation method of Huff (1963) catchments were created for Sainsbury's click and collect stores showing application of the methodology of Dolega et al. (2016), assuming continuous geographic space across the UK (Dennis et al., 2002; Birkin et al., 2010). Catchments are then used to analyse store performance measured by demand per day investigated using regression.

KEYWORDS: Grocery; click and collect; huff; gravity modelling; spatial competition.

1 Introduction

A retail gravity theory led Huff (1963) methodology, positioned in the context of national exploration, used a supermarket case study and open source software to generate store catchments for Sainsbury's grocery click and collect sites. Grocery click and collect is an online service provided by supermarkets enabling customers to order products online and collect in store (Sainsbury's, 2016). The service launched in 2015 (Sainsbury's, 2015). Grocery industry online retail sales accounted for 10.4% total sales (Roby, 2014) with 26% of shoppers using the service and 6% exclusively as their only shopping method (Henry, 2015), showing the service has a valid share of a major industry. Resultant catchments will be used in the decision making process for the next 5 years and 100 click and collect site openings for Sainsbury's grocery click and collect operation, having strong commercial value due to the scarcity of similar study, and the immediate implementation potential.

Retail competition is a complex field where market leaders have advantage from incumbency benefits of size, via proprietary technology, preferential access to suppliers, pre-emption of favourable

*a.e.davies@liverpool.ac.uk
†l.dolega@liverpool.ac.uk
‡d.arribas-bel@liverpool.ac.uk
§alex.singleton@liverpool.ac.uk

geographic location, brand identity and cumulative experience (Porter, 2008). Retailers typically cluster with competitors, termed 'co-opetition', which is performance beneficial; infrastructure is shared whereas wallet share is competed for (Teller and Reutterer, 2008). Competition is also internal highlighting complexity of the retail environment. There is a large body of academic and commercial research focused on store sales estimation, new store location prediction and exploration of retail catchments, which are typically limited to a local or regional extent (Dolega et al., 2016).

Supermarkets often use drive times derived catchments, coupled with intuition based on industry experience and common sense, although the rise of low cost computing and datafication has led to the adoption of modelling in decision making (Hernandez and Bennison, 2000). Site attractiveness, a key determinant of trips to store per annum (Dennis et al., 2002), is integrated in the laws of retail gravitation, initially designed by Reilly (1931), which aid market strategic decision and understanding consumer behaviour (Teller and Reutterer, 2008). Grocery specific and competition study fail to consider multiple store networks (De Beule et al., 2014), showing a gap for national level modelling. Customers shop in continuous geographic space (Dennis et al., 2002; Birkin et al., 2010), and national level boundary free modelling brings spatial consistency in model calibration (Dolega et al., 2016).

2 Method

The study follows a multi-disciplinary approach combining 'geographic data science' and retail study, set in the 'geo-informatics' paradigm (Koutsopoulos, 2011), taking influence from the emerging big data paradigm (Kitchin, 2014), using open source software. New forms of data generated through the societal change of infrastructure and urbanization mean that the data landscape is more advanced and boasts greater detail (Arribas-Bel, 2014). 'Big data', 'e-commerce' (Laney, 2001), and computing power advances have occurred. The data used includes secure demand data and store descriptives from Sainsbury's, LSOA polygons from UKDS (2016), GeoLytix (2016) store descriptives, and geodemograpics from the CDRC (2016). The secure data was provided via a non-disclosure agreement and was multiplied by a constant and random error applied of -1 to 1% to protect the data.

Gravity modelling uses Newtonian laws of physics (Joseph and Kuby, 2011) to model spatial interaction for analysis of retail outlets (Griffith, 1982). Huff (1963) modelling delineates retail catchments as consumers shop based on attractiveness not closest distance, focusing on customer origin data to explain patronage decisions (Joseph and Kuby, 2011), considered the expected reward and cost (McGoldrick and Andre, 1997). Catchments, the areal extent from which patrons of a store will be found (Dolega et al., 2016), were generated. Huff (1963) calculates the probability Pij that a consumer located at i would chose to shop at store j using equation 1.

$$P_{ij} = \frac{A_J^{\alpha} D_{ij}^{-\beta}}{\sum_{j=1}^{n} A_j^{\alpha} D_{ij}^{-\beta}} \tag{1}$$

where: Aj is a measure of attractiveness of store j; Dij is the distance from i to j; α is an attractiveness parameter estimate for empirical observations; β is the distance decay parameter estimated for empirical observations; n is the total number of stores including j.

Huff (1963) allows simultaneous estimation of probabilities of multiple stores (Joseph and Kuby, 2011) following a market penetration approach, where the trade area is considered a probability surface that represents customer patronage, allowing the creation regions of patronage probability or catchments (Dramowicz, 2016). Historically Huff (1963) was limited by processing power, although this is overcome by cheap, powerful computing and open source software (Dramowicz, 2016). Point in polygon analysis of competitors was then performed using the catchments to generate competition counts. The LSOA boundaries meant census data and geodemographics could easily be aggregated to the catchments, leading to further insights of the catchment characteristics. Regression was used to explore performance, explaining the dependent variable as a linear function of explanatory variables (Gelman and Hill, 2006).

3 Empirical Strategy

3.1 Huff Catchments

Huff catchments were generated using the shortest distance function using road networks in the 'huff-tools' R package (Pavlis et al., 2014). Attractiveness is typically measured using store size, although further descriptives lead to better explanation (Dolega et al., 2016), and thus further store characteristics where included. The model had an α of 0.5km to account for walking distance. A distance constraint of 15km excluded outliers. The huff probability was set at above .25. Stores open less that 100 days were removed as outliers. Demand could then be plotted using fisher jenks classification for the catchments (shown Figure 1).

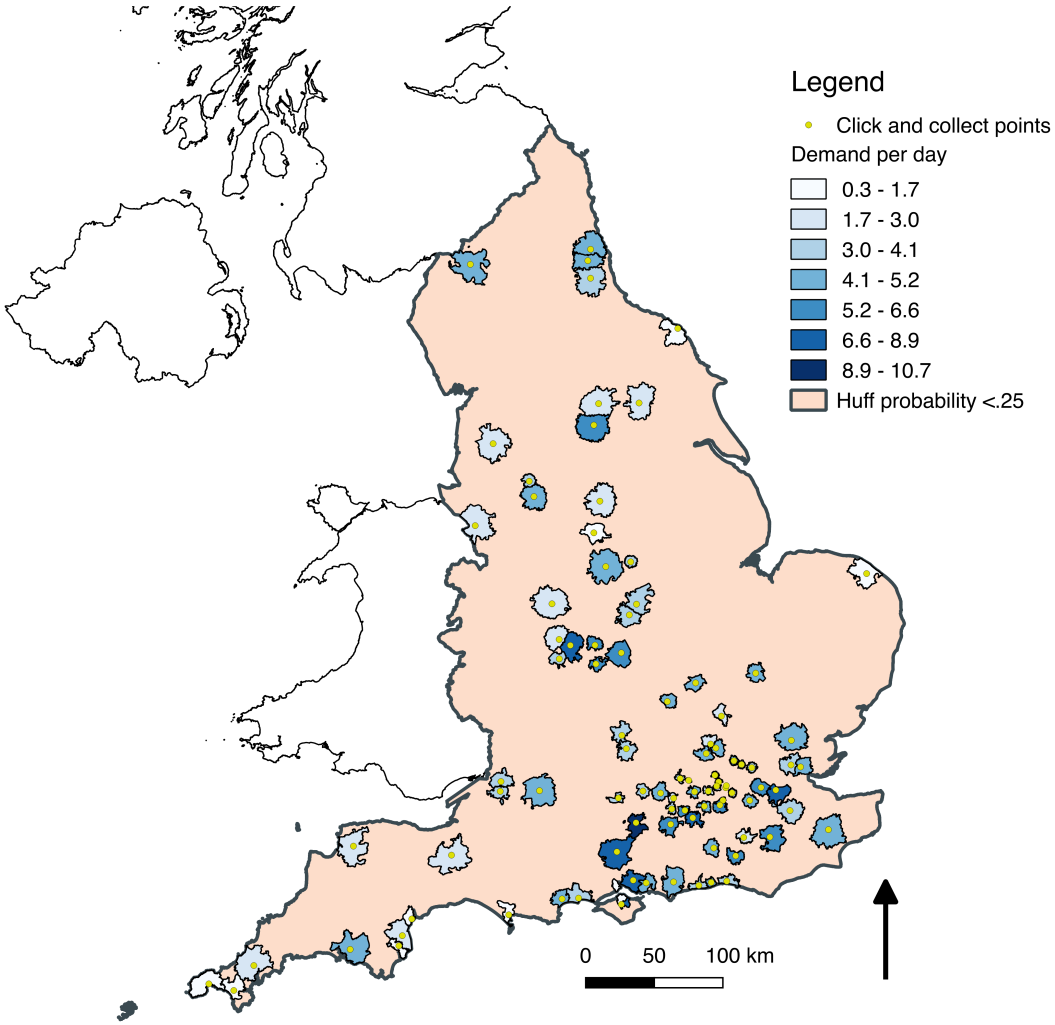


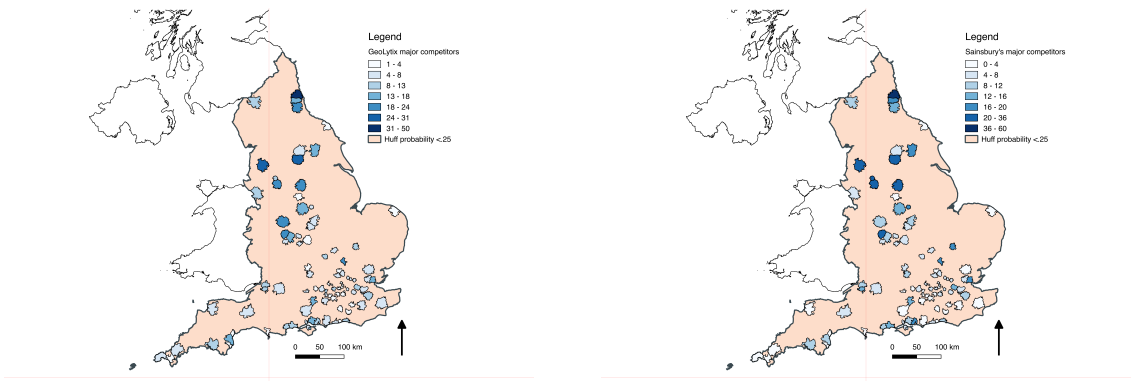
Figure 1: Huff catchments and store demand

3.2 Competition measures

Data for competitors was derived from Sainsbury's competition dataset and checked againstGeoLytix (2016). Counts were subset to include only 'major competitors' stores over 15069ft^2 (counts are shown Table 1 and Figure 2). The data shows catchments and competitor counts larger in rural areas, compared with more urban stores. The differences between the two datasets are very minimal.

Table 1: GeoLytix (2016) and Sainsbury's competitor breakdown

	Asda	Marks and Spencer	Morrison's	Sainsbury's	Tesco	Waitrose	Total
GeoLytix	442	79	453	511	711	208	2404
Sainsbury's	333	238	259	552	856	59	2297



(a) GeoLytix

(b) Sainsbury's

Figure 2: Competition counts

3.3 Performance measures

Store descriptives were used to account for internal competition in terms of performance. These include sales area and trade intensity which both were measured 1-5 (5 as the largest), trading hours (weekly), GOnline - whether the products are picked from that store and Canopy a dummy variable for the best performing collection format.

3.4 Geodemographic measures

Census variables, and geodemographics were used to further explain demand as they are considered to influence loyalty (McGoldrick and Andre, 1997). Census variables used were 'Number of cars', 'Managers', and 'British and Irish population population'. Other data used were the Internet User Classification, workzone population, train stations and Rural-Urban classification.

3.5 Regression

The first model looks at competition, the second store descriptives and the third geodemographics. The final model used demand per day explained by the parameters of competition, performance and geodemographics, shown in equation 2.

$$Demand/day_i = \alpha + \beta_1 Competition_i + \beta_2 Performance_i + \beta_2 Geodemographic_i \epsilon_i \tag{2}$$

Regression results are shown in Table 2.

4 Results

Table 2 shows the regression models. In model 1 a unit increase in competition has a 0.024% increase in demand per day, although this is insignificant meaning there is no confidence as to whether the result is positive(Gelman and Hill, 2006). In model 2 all of the store characteristics are significant, showing that with internal competition, an increase in these increases demand per day and thus performance, although this could be biased by the attractiveness in huff creation. Model 3 shows all cars, managers, British and Irish population, and urban LSOA count as significant. Surprisingly all cars has a negative coefficient, although the R-squared is very low. The final model includes most of the variables (excluding two to remove multicolinearity), where only store descriptives are significant. The R-squared accounts for 58.7% of the variance. Further investigation would explore more variables.

Table 2: Regression models

	Dependent variable:			
	Demand per day			
	(1)	(2)	(3)	(4)
Competition	0.024			0.013
	(0.021)			(0.019)
Sales Area		0.496^{**}		0.484^{**}
		(0.214)		(0.220)
Trade Intensity		0.323^{**}		0.295^{**}
		(0.125)		(0.137)
Trading Hours (Weekly)		0.068^{***}		0.066^{***}
		(0.020)		(0.023)
GOnline		-0.890^{**}		-0.940^{**}
		(0.373)		(0.402)
Canopy		2.294^{***}		2.270^{***}
		(0.387)		(0.416)
All cars			-0.398^{***}	-0.039
			(0.124)	(0.061)
British and Irish population			0.053^{**}	
			(0.023)	
Managers			0.186^{***}	0.038
			(0.062)	(0.052)
Workzone population density			-0.0001	-0.0002
			(0.0003)	(0.0002)
Train station count			-0.001	-0.036
			(0.052)	(0.033)
ERural (IUC)			-0.496	-0.068
			(0.480)	(0.325)
Urban LSOA Count			0.005^{*}	
			(0.003)	
Constant	3.718^{***}	-3.194^{***}	13.854^{***}	-1.396
	(0.272)	(1.001)	(3.900)	(2.948)
Observations	95	95	95	95
R^2	0.013	0.570	0.237	0.587
Adjusted R^2	0.002	0.546	0.176	0.532
Residual Std. Error	1.854 (df = 93)	1.250 (df = 89)	1.685 (df = 87)	1.270 (df = 83)

Note: *p<0.1; **p<0.05; ***p<0.01

5 Conclusion

Huff (1963) modelling gave new insights into grocery click and collect for Sainsbury's, benefiting from large data set analysis (Laney, 2001). The catchments are easily applied to the real world to accurately generate catchment estimations, replacing linear buffers. Competition was considered throughout, with internal competition via the Huff (1963) catchment estimation and then in the regression models, but also external with the inclusion of catchment competitor counts. Ideally the data would include all stores including competitors that offer click and collect, but there is no dataset available - using such a dataset would lead to much more accurate catchments and a greater idea of competition, but for an explanatory study there are valuable insights. Regression showed that increase competitor numbers in a catchment leads to increased demand, although not significant. Store characteristics show size impact demand, which allows further investigation. Insights from the geodemographics show impact on demand, but loyalty needs definition (O'Malley, 1998) to be able to further explore this. There are also many unmeasurable aspects in every day life (Roby, 2014), so the data insights only goes so far. Traditional shopping will never cease (**?**Roby, 2014; Hand et al., 2009) although insights into a new grocery channels are particularly relevant for the companies, as little is known currently. The study ultimately finds competition to impact demand, with performance and geodemographic aiding understanding.

6 Acknowledgements

Appreciation goes to the NWDTC for funding my study via a 1+3 studentship. Appreciation also goes to the CDRC for there sponsored dissertation projects and Sainsbury's for sponsoring the study and providing the data - particular thanks to Guy Lansley and Matthew Pratt. Thanks also goes to the GDSL at the University of Liverpool for providing an excellent working environment.

7 Biography

Alec Davies - Postgraduate student at the University of Liverpool on a 1+3 studentship from the NWDTC in PhD year 1, based in the Geographic Data Science Lab. Research interests are in geographic data science particularly geodemographics and health.

Dr Les Dolega - Lecturer in Geography and Planning at the University of Liverpool. Research interests are economic/ retail geography and GIS.

Dr Dani Arribas-Bel - Lecturer in Geographic Data Science at the University of Liverpool. Research interests include urban economics and regional science, spatial analysis and spatial econometrics, open source scientific computing.

Prof Alex Singleton - Professor of Geographic Information Systems at the University of Liverpool. Research is concerned with how the complexities of individual behaviours manifest spatially and can be represented and understood through a framework of geodemographic data science.

References

Arribas-Bel, D. (2014). Accidental, open and everywhere: Emerging data sources for the understanding of cities. *Applied Geography*, 49:45–53.

Birkin, M., Clarke, G., and Clarke, M. (2010). Refining and operationalizing entropy-maximizing models for business applications. *Geographical Analysis*, 42(4):422–445.

CDRC (2016). Data services. `www.cdrc.ac.uk/data-services/`.

De Beule, M., Van den Poel, D., and Van de Weghe, N. (2014). An extended huff-model for robustly benchmarking and predicting retail network performance. *Applied Geography*, 46:80–89.

Dennis, C., Marsland, D., and Cockett, T. (2002). Central place practice: shopping centre attractiveness measures, hinterland boundaries and the uk retail hierarchy. *Journal of Retailing and Consumer Services*, 9(4):185–199.

Dolega, L., Pavlis, M., and 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.

Dramowicz, E. (2016). Retail trade area analysis using the huff model. `www.directionsmag.com/entry/retail-trade-area-analysis-using-the-huff-model/123411`.

Gelman, A. and Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

GeoLytix (2016). Geolytix. `geolytix.co.uk`.

Griffith, D. A. (1982). A generalized huff model. *Geographical Analysis*, 14(2):135–144.

Hand, C., Dall'Olmo Riley, F., Harris, P., Singh, J., and Rettie, R. (2009). Online grocery shopping: the influence of situational factors. *European journal of Marketing*, 43(9/10):1205–1219.

Henry, V. (2015). Pushing online shopping growth further. `http://www.igd.com/Research/Shopper-Insight/Pushing-online-shopping-growth-further/`.

Hernandez, T. and Bennison, D. (2000). The art and science of retail location decisions. *International Journal of Retail & Distribution Management*, 28(8):357–367.

Huff, D. (1963). A probabilistic analysis of shopping center trade areas. *Land Economics*, 39(1):743–768.

Joseph, L. and Kuby, M. (2011). Gravity modeling and its impacts on location analysis. In *Foundations of location analysis*, pages 423–443. Springer.

Kitchin, R. (2014). Big data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1):2053951714528481.

Koutsopoulos, K. (2011). Changing paradigms in geography. *European Journal of Geography*, 1(1):54–75.

Laney, D. (2001). 3d data management: Controlling data volume, velocity, and variety. `blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf`.

McGoldrick, P. J. and Andre, E. (1997). Consumer misbehaviour: promiscuity or loyalty in grocery shopping. *Journal of Retailing and Consumer Services*, 4(2):73–81.

O'Malley, L. (1998). Can loyalty schemes really build loyalty? *Marketing Intelligence & Planning*, 16(1):47–55.

Pavlis, M., Dolega, L., and Singleton, A. (2014). The huff-tools package.

Porter, M. E. (2008). *On competition*. Harvard Business Press.

Reilly, W. J. (1931). *The law of retail gravitation*. WJ Reilly.

Roby, H. (2014). 'click and collect'; how will this affect e-tailing?

Sainsbury's (2015). Sainsbury's launches grocery click and collect in stores. `www.j-sainsbury.co.uk/media/latest-stories/2015/0303-sainsburys-launches-grocery-click-and-collect-service-in-stores/`.

Sainsbury's (2016). Click and collect groceries. `www.sainsburys.co.uk/shop/gb/groceries/get-ideas/click-and-collect/click-and-collect`.

Teller, C. and Reutterer, T. (2008). The evolving concept of retail attractiveness: what makes retail agglomerations attractive when customers shop at them? *Journal of Retailing and Consumer Services*, 15(3):127–143.

UKDS (2016). Boundary. `census.edina.ac.uk/`.