



UNIVERSITY OF LEEDS

This is a repository copy of *Developing and applying a disaggregated retail location model with extended retail demand estimations*.

White Rose Research Online URL for this paper:  
<http://eprints.whiterose.ac.uk/83956/>

Version: Accepted Version

---

**Article:**

Newing, A, Clarke, GP and Clarke, M (2015) Developing and applying a disaggregated retail location model with extended retail demand estimations. *Geographical Analysis*, 47 (3). pp. 219-239. ISSN 0016-7363

<https://doi.org/10.1111/gean.12052>

---

**Reuse**

Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

# **Developing and applying a disaggregated retail location model with extended retail demand estimations**

Final accepted version

Dr Andy Newing\*<sup>1</sup> Prof Graham P. Clarke<sup>1</sup> Prof Martin Clarke<sup>1</sup>

\* Corresponding author [a.newing@soton.ac.uk](mailto:a.newing@soton.ac.uk) Tel: 02380 598756

<sup>1</sup> School of Geography, University of Leeds, Leeds, LS2 9JT, UK.

## **Acknowledgement**

This research is supported by an ESRC CASE Award (2010–2013) as part of RIBEN. Data have been provided by a collaborating retailer who guided the initial selection of study stores.

## **Abstract**

The Spatial Interaction Model (SIM) is an important tool for retail location analysis and store revenue estimation, particularly within the grocery sector. However, there are few examples of SIM development within the literature that capture the complexities of consumer behaviour or discuss model developments and extensions necessary to produce models which can predict store revenues to a high degree of accuracy. This paper reports a new disaggregated model with more sophisticated demand terms which reflect different types of retail consumer (by income or social class), with different shopping behaviours in terms of brand choice. We also incorporate seasonal fluctuations in demand driven by tourism, a

major source of non-residential demand, allowing us to calibrate revenue predictions against seasonal sales fluctuations experienced at individual stores. We demonstrate that such disaggregated models need empirical data for calibration purposes, without which model extensions are likely to remain theoretical only. Using data provided by a major grocery retailer, we demonstrate that statistically, spatially and in terms of revenue estimation, models can be shown to produce extremely good forecasts and predictions concerning store patronage and store revenues, including much more realistic behaviour regarding store selection. We also show that it is possible to add a tourist demand layer which can make considerable forecasting improvements relative to models built only with residential demand.

## **1. Introduction**

The spatial interaction model (SIM) has a long and distinguished history in the fields of geography and regional science and has been widely used in studies of retail location analysis. The SIM (often referred to as the ‘gravity model’ within the retail industry) has become an important tool for revenue estimation within the grocery sector in particular (Birkin et al., 2010b; Reynolds and Wood, 2010). Although many theoretical extensions to the models have been made over time, there are few examples in the literature that discuss the types of extensions that are necessary to produce models which can be proved to work in a commercial environment – i.e. that not only predict revenues to a high degree of accuracy (within 5% of actual sales) but also capture the complexities of different types of consumer behaviour. This lack of case study material partly reflects the limited work academics publish on work undertaken with particular clients (sometimes due to confidentiality) and the lack of work published by organisations themselves which may use these techniques taken from the academic world, but are likely to customize them to make them more operational.

In section 2 we argue that more research has been undertaken within spatial interaction modelling on measuring the characteristics of retail destinations and their attractions to

different types of customer. There has been relatively less research on incorporating different types of retail demand, especially in relation to modelling brand choice of different household income groups (within the interaction model framework) and in relation to the inclusion of non-residential demand which, in certain regions, can be as important as residential demand (see for example Newing et al., 2013b). One of the main reasons for this has been the lack of commercial data sets (on a large scale) to calibrate such disaggregated models. However, for this paper, we have been given access to a major UK grocery retailer's internal data sets – namely data on store revenues and data from the company loyalty card, plus data drawn from a large commercial consumer survey produced by Acxiom Ltd. (a major UK consumer survey and data analytics company) and additional supply side data from GMAP Ltd., a UK retail consultancy. The provision of this store and consumer data means that this is one of very few examples within the academic literature of an applied SIM that has been developed, calibrated and validated with reference to empirical data supplied by a major retailer and consultancy. The data provided is for Cornwall, a region in the south west of England which is experiencing expansion of retail provision to meet the needs of both local residents and tourist visitors, the latter representing one of Cornwall's major industries (VisitCornwall, 2010). Aside from examples mentioned by Birkin et al. (2010a) (which were generally carried out on a consultancy basis), the authors know of very few examples, reported within the literature, of spatial analytical retail studies that have drawn upon extensive commercial data and been reported comprehensively within the literature. Nevertheless, recent industry-academia collaborations in the UK, such as those facilitated through the RIBEN network<sup>1</sup> are beginning to generate research outputs drawing on commercial collaborations.

In this paper we report on the development and calibration of a SIM, disaggregated on both the demand and supply side, which has been developed using our own demand estimates and

---

<sup>1</sup> <http://www.riben.org.uk/>

calibrated with reference to empirical commercial data. Our final disaggregated model allows us to add more sophisticated demand terms in the model which reflect different types of retail consumer with very different shopping behaviours in terms of brand choice. We are also able to incorporate seasonal fluctuations in demand driven by tourism, a major source of non-residential demand in this region, allowing us to calibrate monthly revenue predictions against seasonal sales fluctuations experienced at individual stores. The richness of the client data also allows a greater ability to calibrate the models accurately and prove the concept of ‘goodness of fit for purpose’ or ‘goodness of forecast’ (Birkin et al., 2010a).

In section 2 we briefly review the literature on the retail SIM and look at the major ways the model has been disaggregated within applied modelling. Section 3 describes the model disaggregation in detail, starting with disaggregation by household type and brand. Then, we disaggregate the demand side by introducing tourist demand in section 4. Section 5 deals with calibration overall, using client data, and finally section 6 evaluates the ability of the model to produce accurate revenue predictions and replicate observed consumer behaviour at a store level. We discuss broader implications for location based modelling in section 7.

## **2 Spatial interaction modelling for retail location analysis**

Spatial interaction models (SIMs) have become a fundamental tool for retail location analysts and are used to forecast flows of consumer expenditure from an origin, usually the consumer’s home, to one of many accessible competing stores. Inherent in the design of the model is the concept that expenditure flows and subsequent store revenue are driven by store attractiveness and constrained by distance, with consumers exhibiting a greater likelihood to shop at stores that are geographically proximate. Accessibility is usually a function of the relative ‘cost’ in terms of distance or travel time ( $C_{ij}$ ), calibrated using a distance decay parameter ( $\beta$ ) which reflects the willingness or ability of consumers to travel to stores in the

modelled region. Store attractiveness is commonly identified using variables such as store size, brand name and range of products stocked (Birkin et al., 2002; Birkin and Heppenstall, 2011). The production-constrained model (Wilson, 1971; Wilson, 2010), is the most commonly used in grocery retail applications, where expenditure estimated in origin zones is given as fixed and is distributed among the competing retail destinations.. A basic form of the production-constrained model used to forecast the expenditure flow ( $S_{ij}$ ) between zone ( $i$ ) and store ( $j$ ) is shown in equation 1.

$$S_{ij} = A_i O_i W_j \exp^{-\beta C_{ij}} \quad (1)$$

Where:  $S_{ij}$  represents the interaction or expenditure flow between zone  $i$  and store  $j$ ;

$A_i$  is a competition factor which ensures that all demand is allocated, it is calculated as:

$$A_i = \frac{1}{\sum_j W_j \exp^{-\beta C_{ij}}} \quad (2)$$

$O_i$  represents the demand or expenditure available in residential zone  $i$ ;

$W_j$  accounts for the attractiveness of store  $j$ ;

$\exp^{-\beta C_{ij}}$  is the distance deterrence term, incorporating  $\beta$ , the distance decay parameter, and  $C_{ij}$ , the distance or travel time between zone  $i$  and store  $j$ .

(Source: Adapted from Birkin and Clarke, 1991; Birkin et al., 2002; Wilson, 1971; Wilson, 2010).

Based on their 2010 survey of location planning departments, Reynolds and Wood (2010) suggest that around two thirds of retail location planning teams (across all sectors) make use of SIM for location planning. Survey respondents identified that such models had become a

flexible and increasingly accurate tool for revenue estimation, adding complexity and sophistication to location analysis (compared to analogue approaches), accounting for expenditure flows over space that result from consumers decision making processes. Birkin et al. (2010a) assert that one reason why these models may have become so popular in an industry context is because the clear return on investment achieved through using these models can be quantified. Birkin et al. (2010b) cite one example, based on a major DIY retailer in the UK, whereby an investment in spatial modelling reduced the margin of error in their new store revenue forecasts from 30% to 10%, giving the company confidence to invest in 25 new stores over a 5 year period, generating profits of around £40m. Investment in this form of modelling can thus be used to achieve robust predictions of store revenue at the pre-investment stage, allowing investment decisions to be made with confidence.

On the supply side, factors such as overall floorspace drive store attractiveness with larger stores more appealing to consumers. In reality, other site specific factors may make a smaller store relatively more attractive than its size would suggest. Birkin et al. (2004) suggest that for grocery retailers, factors including price, product range, opening hours and the availability of parking all have an important influence on consumer's perception of store attractiveness. Consequently, there has been a wealth of research investigating alternative formulations of the attractiveness term (e.g Oppewal et al., 1997; Pacione, 1974; Spencer, 1978). Work by Fotheringham in the 1980s (1983; 1984; 1986) in particular, highlighted that the attractiveness term within the basic spatial interaction model failed to account for the spatial distribution of individual stores in relation to one another, though Krider and Putler (2013) note that supermarkets generally do not exhibit a tendency to cluster, since, from a consumer perspective, there is little to be gained from comparison shopping for groceries. Consequently, Birkin et al. (2010a) suggest that SIMs require a careful choice to be made as

to whether stores are considered as individual stand-alone sites or clustered together into centres.

Thus, whilst considerations of supply-side disaggregations are common place in the literature, demand-side disaggregations are less common. There are some notable exceptions. A number of authors have tried to capture the elasticity of demand which is noticeable in the patronage of cinemas, restaurants, fast food outlets and even ATMs (Birkin et al., 2010a; Ottensmann, 1997; Pooler, 1994). Such models tend to include an accessibility function within the demand term so that demand effectively increases substantially when residents are in close proximity to the retail destination. There have also been a number of studies which have explored demand variations by social class or income group. Thus, for example, expenditure is estimated to be much higher in areas of higher income. This type of work is more common when the chosen methodology is discrete choice or random utility models rather than SIMs (Solgaard and Hansen, 2003; see also Fotheringham and Trew (1993) for a good summary). There is also a set of papers on brand choice within such discrete choice models (Wrigley and Dunn, 1984) but these are not typically based on the choices made by consumers within SIMs (which need to include not only the choice of the brand itself given social class etc., but also the choices given a competing set of retail locations of the different brands in question which typically exist in a particular region).

At the very least, different income groups may have a different propensity to travel further to the store of choice based on the cost and accessibility of transport. Disaggregation may thus be as straightforward as applying different  $\beta$  values for different groups of consumers to account for the fact that a single  $\beta$  value is unlikely to be able to represent all the different complex consumer flows that exist. For example, in an application of a SIM to estimate the impacts of the new Silverburn regional shopping centre near Glasgow, Scotland, Khawaldah et al. (2012) applied different  $\beta$  values for consumers in each postal area, recognising that



those residents in geographically remote postal areas were less likely to be over-sensitive to the impact of distance due to the inevitable longer journeys involved in accessing principal shopping centres.

Thus whilst it is realistic to assume that, based on factors such as age, socio-economic status or income, consumers will exhibit more individualised behaviours with respect to store choice, the availability and relative accessibility of different brands and store options is an important consideration. Certain groups of consumers may have a higher propensity to travel further to the store of choice and as such retail brand is often an important driver of consumer behaviour. Literature suggests that consumers will exhibit brand preferences based on perceptions of store quality, service and price (Clarke et al., 2006; Clarke et al., 2012; Jackson et al., 2006; Kirkup et al., 2004). For example, UK consumers tend to perceive that Sainsbury's brand has a more upmarket position than Tesco, ASDA and Morrisons, with Clarke et al. (2012) noting that consumers from more affluent areas were considerably less satisfied if they had a Tesco nearby, rather than a Sainsbury's. As a consequence, evidence suggests that consumers who shop at Sainsbury's exhibit a tendency to have travelled past an alternative store closer to their home in order to shop with their brand of choice (Intel, 2012) but their propensity to do so will often depend upon factors such as income and car ownership (Kirkup et al., 2004).

There is also a paucity of work on estimating demand from non-residential sources. Based on a comprehensive study of consumer habits, Jackson et al. (2006) note that consumer decisions about when and where to shop are increasingly embedded within complex lives and carried out around responsibilities such as childcare and work. As such, residential grocery demand may often originate from workplaces or leisure destinations, particularly where residents commute into major settlements on a regular basis for these purposes. There has been some theoretical work on multi-purpose trip making models (i.e. Arentze et al., 2005;

Mulligan, 1983; O'Kelly, 1981) (Mulligan, 1987) (McLafferty and Ghosh, 1986) (Borgers and Timmermans, 1986) and implications for store choice and location based decision making, but these are hard to calibrate for entire populations given the lack of data on such complex consumer behaviours (although future agent-based models may be able to offer new insights). Thus, they have rarely been considered within applied modelling.

There are few systematic studies of how retail demand can be disaggregated between residential and non-residential origins, yet Birkin et al. (2010a), Birkin et al. (2004) and Birkin and Foulger (1992) note the importance of work-based demand in certain high street or city centre locations. In particular, there is little research in the literature which considers other forms of non-residential demand, notably demand from tourism which in certain regions can be sizeable. In coastal regions of the UK for example, the store-level demand uplift for groceries driven by tourism can be as high as 200% in key months of the tourist season (Newing et al., 2013b). This form of demand is unique in that it is highly concentrated spatially, exhibiting clear clusters around major resorts and destinations, whilst also giving rise to a highly seasonal pattern of fluctuation driven by institutional factors such as school holidays alongside short-term fluctuations owing to the weather and local events (Newing et al., 2013a). Thus models which do not include tourist demand may seriously under-predict the revenue estimates which are such an important output from the SIMs.

The rest of the paper aims to address these two gaps in the literature in relation to model disaggregation from a demand perspective, considering both brand attractiveness by demand type and incorporation of seasonal non-residential demand driven by tourism. To do this we develop a SIM that is disaggregated on both the supply and demand side. The model which is outlined in section 3 draws on estimates of small-area grocery demand (Newing et al., 2013a) and incorporates supply side data from GMAP Ltd. and a major UK retailer. The model is able to estimate flows of consumer expenditure between all Census Output Areas (OAs) that

make up the county of Cornwall (UK) and major food stores serving those residential neighbourhoods, workplaces and tourist accommodation sites. An OA represents the lowest level of a series of hierarchical zones used for the aggregation and dissemination of census and administrative data in the UK. OAs are built around residential addresses (representing an average of 124 households (Vickers and Rees, 2006)) and are an important spatial scale for local-level analysis and decision making, commonly used for store-level demand estimation and market share analysis by retailers e.g. see (Dugmore, 2013).

Section three outlines how key parameters and constraints within a SIM can be disaggregated by household and store type, allowing the model to handle some of the more complex and individualised behaviours of different groups of consumers, and to take account of key socio-economic characteristics that drive expenditure and store choice. In section 4 we consider disaggregation by demand type including demand originating from visitors.

### **3. Model extensions 1: disaggregating by brand and person type**

It is recognised that the characteristics of demand, the attractiveness of the retail destination and the propensity to travel to the retail destination of choice, will vary according to the income, age, ethnicity or other socio-economic characteristics of the consumer, and may also vary depending on the type of product in question. This section seeks to account for these factors within our model and makes extensive use of consumer survey data, explored below after a brief conceptual outline of the disaggregate model.

The model takes the same form as the classic production-constrained SIM, yet the balancing factor ( $A_i$ ) demand ( $O_i$ ) supply ( $W_j$ ) and distance deterrence ( $exp^{-\beta C_{ij}}$ ) terms have been modified to incorporate behaviours by different household or visitor types ( $k$ ). The model also accounts for the relative attractiveness of different store brands ( $n$ ) to different consumer types, operationalised through the introduction of a power function ( $\alpha^{kn}$ ) incorporated

within the attractiveness term in order to apply a measure of relative brand attractiveness. The inclusion of these additional terms allows both supply and demand to be disaggregated independently, yet the links between them maintained through the recurrence of household type or visitor ( $k$ ) on both the demand and supply side.

The new model can be written as:

$$S_{ij}^{kn} = A_i^k O_i^k W_j^{\alpha^{kn}} \exp(-\beta^k c_{ij}) \quad (3)$$

Where:  $S_{ij}^{kn}$  represents the predicted expenditure flow between zone  $i$  and store  $j$  (of store brand  $n$ ), by household of type  $k$ .

$A_i^k$  is a balancing factor which takes account of competition and ensures that all demand from zone  $i$  by household of type  $k$  is allocated to stores within the modelled region. The balancing factor thus ensures that:

$$\sum_j S_{ij}^{kn} = O_i^k \quad (4)$$

It is calculated as:

$$A_i^k = \frac{1}{\sum_j W_j^{\alpha^{kn}} \exp(-\beta^k c_{ij})} \quad (5)$$

$O_i^k$  is a measure of the demand or expenditure available in demand zone  $i$  by household/visitor of type  $k$ .

$W_j$  reflects the overall attractiveness of store  $j$ , whilst  $\alpha^{kn}$  represents the additional or perceived relative attractiveness of store  $j$  for household type  $k$  and by store brand  $n$ .

$C_{ij}$  is the distance (although in this application, travel time is used) between zone  $i$  and store  $j$ , and incorporates the distance deterrence/decay parameter  $exp^{-\beta^k}$  for household of type  $k$ .

Household type is used here to segment consumers by geodemographic status or income based on inferred household characteristics as outlined below. When considering visitor demand (section 4),  $k$  refers instead to ‘visitor party type’, driven by the nature of the tourist visit (day or overnight) and the type of accommodation used, considered a key indicator of grocery expenditure habits. This disaggregation by both household type and retailer brand affords tremendous potential for the model to incorporate flows between different types of consumer and different retailers, as outlined in the following sub-sections.

### **3.1 Categorising households by geodemographic status**

Households have first been categorised using the Office for National Statistics’ ‘Output Area classification’ (OAC) geodemographic system (Vickers and Rees 2006). The classification is based on 2001 census data and classifies all 175,434 Output Areas (OAs) in England and Wales into a hierarchy of 7 Supergroups and 21 groups based on 41 census variables (Vickers and Rees, 2006). The variables used for the classification reflect the socio-economic nature of the households that make up each OA and include demographic, housing and employment characteristics. It is the only area based geodemographic classification accredited as a national statistic and is based entirely on census data due to the unrivalled geographic coverage and robustness of household level socio-economic data collected by the census (Vickers and Rees, 2006).

The OAC is used extensively for targeting central and local government resources and is also widely used by the private sector for commercial decision making and targeted marketing (Allen, 2008). The ‘Living Costs and Food Survey’ (LCF) (ONS, 2010) is also reported by

OAC group and forms an important link to surveyed household expenditure data used as part of demand estimation. A 2011 OAC was under development at the time of data analysis and preparation of this paper, but not yet available for use. Consequently, the 2001 OAC remains an up to date and respected geodemographic classification scheme, used extensively for the reporting of large scale governmental surveys.

The OAC classification is itself used to determine relative brand attractiveness as explored below. It is also used to control the importance placed by different household types on distance or travel time, using inferred income, as a key driver of interaction patterns, via the distance deterrence parameter ( $\beta$ ). Within our model,  $\beta$  is a key parameter used to control flows by determining the importance of distance/travel 'cost' in household decision making behaviour. Within the model calibration (section 5), different  $\beta$  values are used to simulate the ability or willingness of consumers, from different household types, to travel further to their store of choice (irrespective of brand). For example, less mobile groups may be more likely to shop at their closest store even if it does not represent the brand that is most attractive to their age, income or socio economic group. By contrast, low income groups with access to transport may be more willing to travel further in order to access discount stores even if more proximate options exist, whereas higher income households who are more likely to own cars can be given the 'freedom' to travel further to their store of choice.

Unfortunately the UK census does not collect or report information on household income in order to operationalise the distance deterrence parameter for use in the model. Nevertheless, the reporting of key governmental surveys such as the LCF (ONS, 2011) allows reported household income to be linked to household geodemographic status via the OAC. The LCF is a national sample survey of around 5,000 households per annum (ONS, 2011). Surveyed households complete a diary of expenditure for a two week period, with results weighted to account for the characteristics of all households and reported alongside key household socio-

demographic characteristics. As such our study OAs have been categorised into three income groups (high, mid and low) based on their household income (relative to other households in the study area). It is this categorisation of residential demand by OAC group and subsequently by income that reflects the disaggregate potential of this model, allowing model parameters to be set independently for households within each OAC or income group, discussed below.

### **3.2 Incorporating brand preferences**

In addition to their mobility and the relative proximity of available stores, each consumer or household will make decisions about where to shop based on their perceptions of each brand's offering in terms of value, service etc. To incorporate this form of decision making within the disaggregate model, consumer data from Acxiom's research opinion poll (ROP) (2009 and 2010) is used following extensive analysis by Thompson et al. (2010; 2012). The ROP is a detailed annual consumer lifestyle survey of approximately 750,000 households in Great Britain, reported at the household level, conducted by Acxiom Ltd. (Thomas 2012). Acxiom are a UK based private sector market research company and their ROP represents a rich and valuable dataset rarely available for academic investigations, containing detailed information on household spending habits, store choice and socio-demographic information, alongside their geographic location (Thompson et al., 2012). Within the ROP, households report which major retailer they use for their main weekly grocery shop and this dataset presents a good indication of actual brand choice by respondents, which in turn can be linked to their geodemographic status.

Thompson et al. (2012) used Acxiom's ROP in combination with the OAC to identify the consumer base for each major UK grocery retailer. They report location quotients for each retailer-OAC supergroup combination ( $n = 7$  supergroups and 10 retailers), dividing that

retailers' observed customer breakdown (by OAC group) by the underlying distribution of population across the OAC groups in their study region. As such, their location quotients identify whether a particular OAC group is over or under represented in a retailers' customer profile. They note, for example, that Waitrose, M&S and to some extent Sainsbury's all generate greater patronage from the affluent 'city living' supergroup than would be expected based on the prevalence and spatial distribution of those households alone. The same is true of ASDA in the 'blue collar communities' supergroup, Co-Op in the 'countryside' supergroup and Sainsbury's in the more affluent 'prospering suburbs' supergroup.

We have used the location quotients produced by Thompson et al. (2012) to set brand attractiveness, incorporated as part of the attractiveness term, for each retailer and each residential household type in our SIM. The location quotients have been rescaled around the value of 1 maintaining the relative difference between them, since alpha operates as a power function on store attractiveness (floorspace) in the model. As such, store floorspace is raised to a power, depending on the individual combination of household type and store brand, thus recognising that a unit of floorspace of a Waitrose store is more attractive than a unit of floorspace of ASDA to certain household types. The rescaled location quotients by OAC supergroup, used to operationalise the alpha term within the model are shown in the matrix within Table 1.



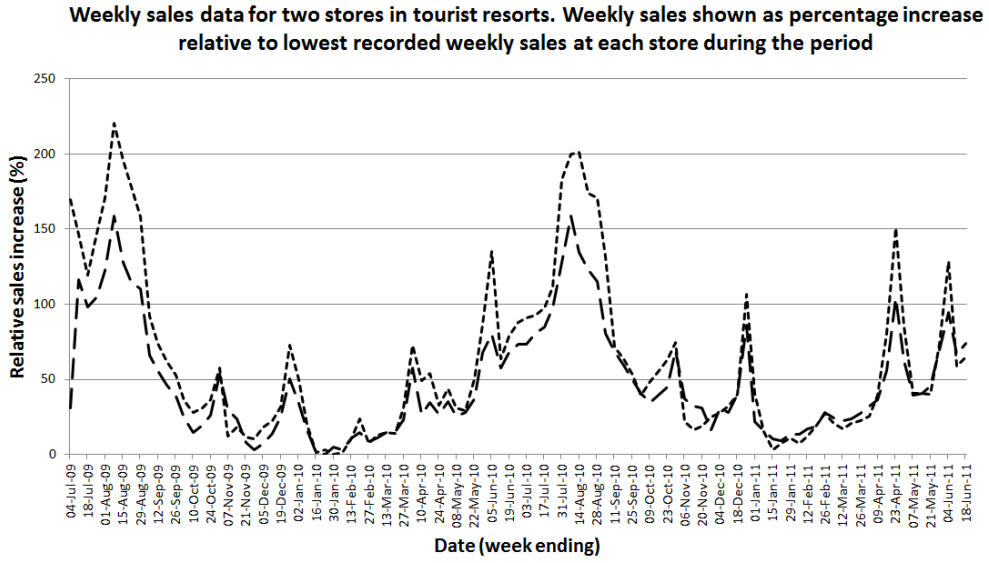
**Table 1 - Brand location quotients used to set alpha values**

<b>Brand (Retailer)</b>	<b>OAC Supergroup</b>							
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	
	<b>Blue Collar</b>	<b>City Living</b>	<b>Countryside</b>	<b>Prospering</b>	<b>Suburbs</b>	<b>Constrained by Circumstances</b>	<b>Typical Traits</b>	<b>Multicultural</b>
<b>Aldi</b>	0.9980	0.9970	1.0051	0.9987	1.0025	1.0005	0.9952	
<b>ASDA</b>	1.0076	0.9912	0.9904	0.9970	1.0023	0.9992	1.0013	
<b>Co-Op</b>	1.0020	0.9990	1.0157	0.9922	1.0008	1.0000	0.9894	
<b>Lidl</b>	1.0015	0.9995	1.0066	0.9962	0.9957	0.9997	1.0091	
<b>M&amp;S</b>	0.9891	1.0381	0.9967	1.0066	0.9952	1.0051	1.0003	
<b>Morrisons</b>	1.0005	0.9942	0.9997	0.9987	1.0020	1.0005	0.9990	
<b>Sainsbury's</b>	0.9904	1.0121	1.0013	1.0088	0.9942	1.0028	0.9997	
<b>Tesco</b>	0.9992	0.9987	1.0071	1.0010	0.9965	0.9990	0.9985	
<b>Waitrose</b>	0.9811	1.1000	1.0061	1.0124	0.9843	1.0023	1.0068	
<b>Iceland</b>	0.9997	0.9982	1.0058	0.9975	0.9991	1.0001	1.0021	

Alpha is intended to control the relative attractiveness of different brands to different household types, based on their inferred geodemographic status. The values used for alpha have been ‘set’ based on the analysis carried out by Thompson et al. (2012) and are ‘static’ within the model and are not actively calibrated since we do not have access to the actual Axiom data itself. The impact of alpha on household grocery shopping trip making behaviour (by income group) is shown in section 6.

#### **4. Model extensions 2: disaggregating by type of demand**

Another important disaggregation needed on the demand-side in many UK regions is demand which is tourist or visitor led. As noted above, demand for tourism can increase sales by 200% in the summer months (Newing et al., 2013b). Figure 1 shows the seasonal variation in sales for two grocery stores in Cornwall, with both stores in seaside locations. Analysis of loyalty card data for these stores for the corresponding period identifies that a considerable portion of this sales uplift is driven by visitor demand – with the proportion of loyalty card spend originating from customers with a home address outside the store catchment reaching over 60% during August at one of these stores. This is not unexpected given the nature of these resorts as major tourist destinations which are heavily geared towards highly seasonal self-catering trips, which can be expected to generate considerable retail spend, especially on food and drink sourced locally (BH&HPA, 2012; Dudding and Ryan, 2000; Mottiar, 2006; Timothy, 2005). Additionally, expenditure associated with visitors also originates from households hosting visiting friends and relatives (VFR) or staying in a second or holiday home (Quinn, 2010), with evidence from local econometric modelling in Cornwall (South West Tourism, 2010), coupled with surveys of hosts (ETC, 2002), identifying over £10m of additional grocery expenditure associated with these visitors in Cornwall during 2008 (the most recent year for which data is available).



**Figure1: Seasonal variations in sales across two grocery stores in Cornwall**

Recall that residential demand has been segmented by household type ( $O_i^k$ ). This allows the small area available residential expenditure to be built up from a household level based on geodemographic and socio-economic characteristics (and surveyed expenditure). Residential demand has been calculated as:

$$O_i^{kt} = e^{kt} n_i^{kt} \quad (6)$$

Where:

$O_i^{kt}$  is a measure of the total available expenditure available in zone  $i$  by householdtype  $k$  during seasonal time period  $t$ .

$e^{kt}$  is a measure of the average weekly groceries expenditure for household type  $k$  during time period  $t$ , taken from the living costs and food survey.

$n_i^{kt}$  reflects the number of households of type  $k$  in zone  $i$  during time period  $t$ .

Visitor demand, which is seasonal in nature, is added as a separate series of layers representing the additional (average weekly) visitor induced demand during 12 monthly periods (a 52 week ‘average’ visitor demand is also used). Visitor demand has been calculated as:

$$V_i^{kt} = e^{kt} n_i^{kt} \quad (7)$$

Where:

$V_i^{kt}$  is a measure of the total available expenditure available in zone  $i$  by visitor of type  $k$  during seasonal time period  $t$ .

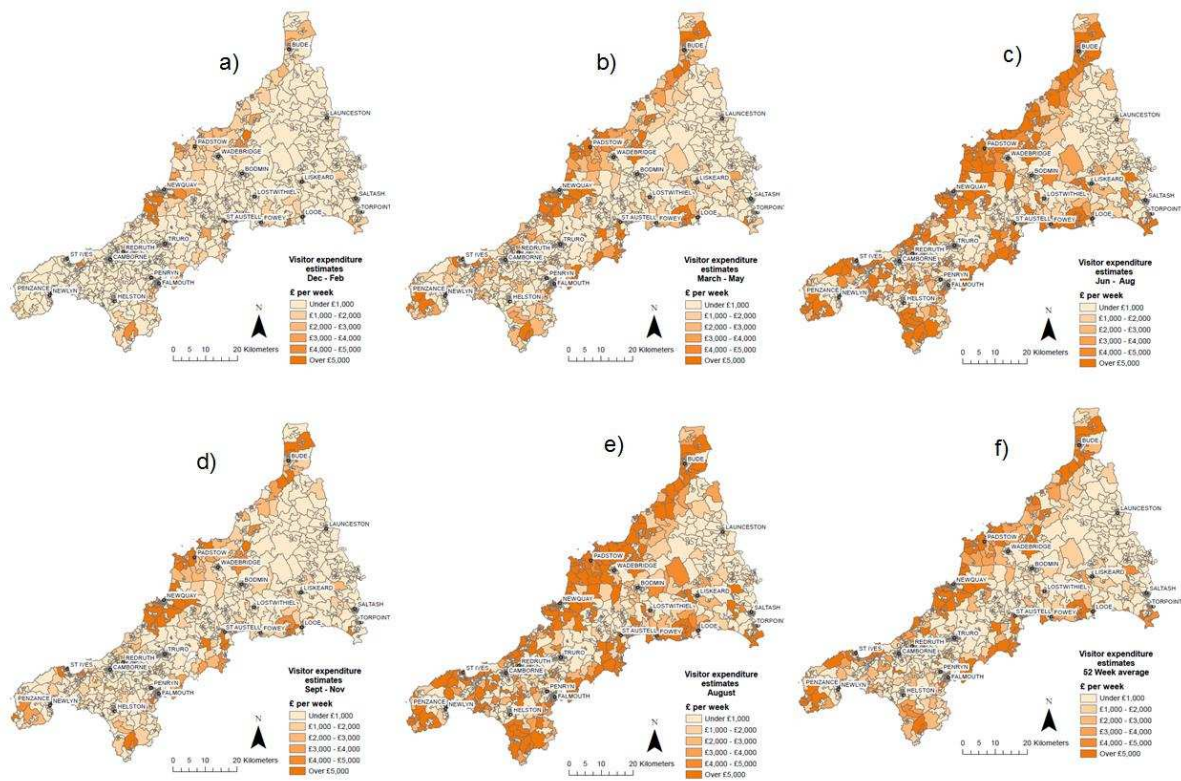
$e^{kt}$  is a measure of the average weekly groceries expenditure for visitor type  $k$  during time period  $t$ , drawn from a variety of survey sources (outlined fully in Newing et al., 2013a) and informed by loyalty card analysis.

$n_i^{kt}$  reflects the number of visitors of type  $k$  in zone  $i$  during time period  $t$ .

These layers incorporate spending by visitor parties (typically a family group, equivalent to a ‘household’) of different types ( $k$ ), which influence their likely grocery spend. We consider visitors using all forms of overnight accommodation, including visitors using rented self-catering accommodation, camping and caravanning, staying in a second home or with friends and relatives (see Newing et al., 2013a). Additional spending by those hosting visitors, along with spending by day visitors visiting local attractions, resorts and beaches is also included.

Using the approach outlined fully in Newing et al. (2013a) we produced a series of demand layers to estimate small-area seasonal and spatial demand originating from tourists, shown in Figure 2. These seasonal demand layers clearly show the spatial concentration of tourist demand around key coastal resorts such as St Ives, Newquay, Bude and Padstow and the

importance of the summer season in driving tourist demand uplift. We have estimated visitor demand, where possible, from the ‘bottom-up’, using the supply of individual ‘units’ of commercial accommodation as a building block to which surveyed expenditure and occupancy rates are applied.



**Figure 2 - Seasonal visitor demand estimates (average weekly spend)**

a) Winter (Dec-Feb), b) Spring (March – May) c) Summer (June – Aug) d) Autumn (Sept - Nov), e) August (peak school summer holidays) and f) 52 week Average

This approach is built on the premise that the spatial distribution of visitor spending is predominantly driven by the spatial distribution of the visitor accommodation stock (see Newing et al. 2013b). Given that no comprehensive or complete database of visitor accommodation exists within the UK (e.g. see Johns and Lynch, 2007), these estimates are based on considerable validation and updating of fragmented local databases held by tourist organisations in South West England. Occupancy rates for commercially operated

accommodation are routinely collected and reported (see White, 2010) and have been used to determine seasonal patterns of accommodation utilisation. No nationally representative survey of visitor spend on groceries exists in the UK (although key headline surveys such as the United Kingdom Tourism Survey (UKTS) contain broader spending categories), however surveys by key trade organisations such as the British Holiday and Home Parks Association (BH&HPA, 2012) provide an excellent indication of grocery spend associated with visitors and have been used, in conjunction with loyalty card analysis (reported fully in Newing et al. 2014) to apply seasonal visitor grocery expenditure rates to the occupied accommodation stock.

In addition to grocery spend associated with visitors using commercial accommodation, additional expenditure associated with visitors using a second or 'holiday' home unit, staying with friends and relatives or visiting local resorts on a day trip basis have been incorporated. The estimates utilise outputs from the 'Cambridge Local Impact Model' (Cambridge Model), a key econometric modelling tool employed by the tourist sector, providing headline estimates of trip volumes and value (DCLG, 2006). These have been disaggregated seasonally and spatially across the study area, in conjunction with other regional and local survey data, in order to estimate seasonal grocery expenditure associated with these visitors at the OA level. Since little is known about this form of demand, no established methodology or data sources exist and the approach used results from an extensive literature review, search for and exploration of potential data sources. These estimates benefit from access to the only comprehensive source of data about commercial accommodation within Cornwall and considerable input and validation from the authors, and offer tremendous scope to model seasonal grocery expenditure fluctuations driven by tourism. Section 5 now considers model calibration, incorporating the demand side model enhancements outlined throughout this section.

## 5. Model calibration

As mentioned in section 1, a unique aspect of this paper is the use of commercial data, supplied by a collaborating retailer, to calibrate the model based on known consumer flow data and actual observed store revenues which can be used to assign values to model calibration parameter beta such that the model is able to reproduce observed consumer behaviour, and thus estimate store revenue, to an acceptable level of accuracy (within the grocery sector an accuracy threshold of +/- 5% of observed revenue would be expected). If observed consumer behaviour can be consistently replicated by the calibrated model, the model can be used in a predictive capacity within the retail sector, for example to consider the impact of new store openings.

Although the objective is to predict store revenue, in practice calibration involves setting model parameters in order to optimize conditions that are thought to be representative of flow patterns. Birkin et al. (2010a) identify that SIM calibration is traditionally undertaken by comparing observed and predicted average trip distance (ATD). Batty and Mackie (1972) assert that this is the most appropriate calibration statistic to use for a SIM which employs an exponential distance function. The premise is simple: if the model can replicate observed consumer trip making characteristics then it is likely to estimate the spatial patterns of trade (or store catchment area) effectively. Assuming that demand estimates are reasonable, and that the model has an appropriate representation of store attractiveness, actual expenditure flows to stores, and thus individual store revenue should then represent reality as closely as possible. The calibration routine reported here thus seeks to minimise the difference between observed and predicted ATD and to demonstrate, via selected goodness-of-fit (GOF) statistics ( $R^2$ , SRMSE), that the subsequent modelled flows can replicate observed flows, and predict store revenue, to an acceptable level of accuracy.

Equation 8 outlines the calculation used to minimise the difference between observed and predicted ATD.

$$ATD = \frac{ATD^{Pred}}{ATD^{Obs}} \quad (8)$$

Where:

$$ATD^{Pred} = \frac{\sum_{ij} S_{ij} C_{ij}}{\sum_{ij} S_{ij}} \quad (9)$$

$$ATD^{Obs} = \frac{\sum_{ij} \hat{S}_{ij} C_{ij}}{\sum_{ij} \hat{S}_{ij}} \quad (10)$$

and  $S_{ij}$  represents predicted flows, and  $\hat{S}_{ij}$  represents observed flows.

Effective calibration is dependent upon the availability of sufficient observed customer flow data. Obtaining observed flow data can be tricky and inevitably involves generalising from a small sample of customers. Observed flow data is based on the individual transaction level records derived from the retailer's loyalty card database for four study stores during the 2010 trading year. These transactions have been aggregated to the OA level and used to calculate observed ATD. Rather than straight line distance, our model employs a travel time matrix in order to reflect the car-borne nature of trade in this predominantly rural area. The road travel times used here were provided by the client and extracted from MapInfo Drivetime (version 7.1) software using the 'Street Pro' (2011 edition) road network. The quickest off-peak route (rather than the shortest) was applied. The drive time software itself is a powerful tool for calculating drive times, taking account of routing restrictions such as roads with limited access/exit restrictions, long-term roadworks and traffic signals. Since the model operates using road travel time in place of distance, ATD can in fact be thought of as the average trip



'cost', and reflects the average road travel time (in minutes) between the centroid of the OA containing the loyalty card holders registered home address, and the OA containing the store itself.

In order to calibrate the model, which was built by the authors, a calibration routine was developed utilising an iterative procedure, whereby a series of incremental beta values were cycled through by the model, using increasingly narrow ranges and smaller incremental values, in order to identify values that most closely replicated the observed flows, with a view to minimising the difference between  $ATD^{Pred} / ATD^{Obs}$ . Recall that consumers have been segmented into three income groups, allowing different beta values applied for each income group in order to replicate different trip-making behaviours of these households. The application of beta values, driven by income group, is again based on analysis of consumer grocery shopping habits and interaction patterns carried out by Thompson et al. (2012). They identify consumer interactions between their home address and their stated grocery store. Using road travel time at the postal sector level, they identified average travel distance for consumers within three income categories, and use this to apply appropriate values within modelling framework in order to capture the propensity (through either choice or need) for higher income consumers to travel further than lower income consumers. For the analysis within this paper, the iterative procedure maintains the relative difference between the beta values applied to high, mid and low income households based on Thompson et al's (2012) findings, accounting for differences in interaction behaviour between different income groups.

**Table 2 - Observed and predicted ATD (travel time) for Cornish study stores - based on 52 week average flows.**

ATD	Road Travel Time (Minutes) – OA Level		
	$ATD^{Pred}$	$ATD^{Obs}$	$\frac{ATD^{Pred}}{ATD^{Obs}}$
Store 1	9.91	8.84	1.12
Store 2	10.70	10.27	1.04
Store 3	12.16	11.70	1.04
Store 4	25.80	27.34	0.94
<b>Average</b>	<b>14.64</b>	<b>14.54</b>	<b>1.04</b>

Table 2 shows observed and predicted ATD, based on road travel time, for four study stores in Cornwall, based on 52 week average flows. No observed flow data is held for visitor demand (since the local origin zone for tourist visitor loyalty card holders is unknown) and so the comparison of ATD is based solely on residential demand. In order to generate the largest possible dataset of observed flows, 52-week average weekly observed flows are used for calibration, without any seasonal disaggregation. Table 2 identifies a close correspondence between predicted and observed ATD, with a trade-off between the slight over-estimation at ‘store 1’ and under-estimation at ‘store 4’, which, due to its size and location on the principal road network, is able to draw consumers from a wider trade area. The ability of the model to predict ATD such that it closely resembles observed ATD across four diverse stores suggests that the model parameters set are appropriate.

Having optimised consumer flows using ATD, measures of GOF (see Fotheringham and Knudsen, 1987; Knudsen and Fotheringham, 1986; and Openshaw, 1975 for more detail) have subsequently been used to validate and test the degree of statistical fit between the observed and predicted flows. GOF statistics provide an overall assessment of model performance, validating its ability to reproduce the known flow volumes supplied by our retail partner, measuring systematic differences between observed and predicted values (Batty and Mackie, 1972). Knudsen and Fotheringham (1986) note that this assessment of the model's ability to replicate an observed set of data is an important component of model building. We made use of two GOF statistics:  $R^2$  (or the coefficient of determination) which is commonly used to assess SIM performance, and SRMSE (standardised root mean square error) which is observed to be very sensitive to any differences between the observed and predicted flow matrix (Harland, 2008). These are both considered to be some of the 'better performing' and more commonly used GOF statistics (Fotheringham and O'Kelly, 1989) and whilst space does not permit a full discussion of their calculation and relative strengths and merits, an overall SRMSE of 0.05 (where 0 denotes exact fit between observed and predicted) and  $R^2$  of 0.88 (where 1 denotes exact fit) suggests that the model is performing very well with respect to the observed consumer flows at the four stores of interest.

It is important to recall that attempts have not been made to calibrate the model through variation within the values used for the alpha term (Table 1), since this study does not have access to any form of reliable surveyed data for consumer brand preference in Cornwall. Any attempt to fit the alpha values to the Cornwall flow data (which is limited to one retailer and four stores) would represent too much of an attempt to fit the model to the observed data, which Birkin et al. (2010a) term 'over-paramatization'. It would be all-to-easy to artificially alter the alpha values such that the model exactly replicated the observed flows for the study stores, but with absolutely no concern for actual consumer behaviour with regard to

preference for other brands not covered by our loyalty card data. Notwithstanding this point, the impact of incorporating the matrix of alpha values shown in Table 1 can be assessed with further reference to ATD.

Table 3 illustrates the impact of the alpha term on ATD (road travel time is used) for both low and high income households. Table 3 clearly demonstrates that the incorporation of alpha values (from Table 1) improves the ability of the model to replicate the type of spatial consumer behaviour anticipated, relative to  $\alpha = 1$ , which effectively disables the alpha term within the model. Following the introduction of alpha as a model parameter we would expect higher end retailers, such as M&S, Waitrose and Sainsbury's to be more attractive to higher income households and less attractive to low income households, whilst discount retailers (such as Lidl, Aldi, Iceland and, to an extent, ASDA) to be relatively more appealing to lower income households. Considering low income consumers, the use of alpha values (that vary by consumer income and brand type) increase these consumers' average travel time to an ASDA store by over 9 minutes (compared to  $\alpha = 1$ ), suggesting that the model can now account for the fact that these consumers are willing to travel further to reach ASDA stores, which become relatively more attractive, by-passing stores that are geographically proximate in order to do so. Similarly, high income consumers exhibit increasing willingness to experience longer average travel times (increasing by around 50%) to shop at M&S, and considerably reduced average journey times for visits to ASDA.

**Table 3 - Impact of alpha parameter on ATD (travel time in minutes) for low and high income consumer groups**

Retailer	Low income consumers		High income consumers	
	$\alpha = 1$	$\alpha$ varies by k and n	$\alpha = 1$	$\alpha$ varies by k and n
Aldi	6.80	6.69	6.90	5.17
ASDA	21.83	30.86	25.21	15.83
Lidl	11.39	11.61	9.65	7.00
M&S	4.88	4.02	3.73	6.83
Morrisons	20.65	24.97	16.46	18.40
Sainsbury's	23.03	15.91	19.79	26.62
Tesco	29.89	25.50	22.64	29.31

Tables 2 and 3, alongside the GOF statistics presented above, suggest that the model can replicate observed ATD very well, accounting for expected behavioural characteristics associated with household income and brand attractiveness. Nevertheless, the real value of the model is its ability to predict store revenue with accuracy, such that it can be used in a predictive capacity. Birkin et al. (2010a) even suggest a move away from traditional concepts of goodness-of-fit statistics to a more 'applied' approach to model validation, considering whether the models are able to accurately replicate customer flows and store revenue, effectively termed goodness-of-forecast and considered in section 6.

## **6. Model's ability to estimate revenue (goodness-of-forecast)**

Since the model is intended for use in an applied, predictive capacity, the ability to generate accurate revenue predictions at the store level is crucial. Revenue estimation is considered in terms of the four stores used for calibration, and an additional ‘test store’ (store 5), that has not been part of the calibration process (and for which limited data are available).

The revenue data used here has been supplied by the client and considers store level revenue, derived from food and drink sales, on a week-by-week basis. Store revenue within the model can be estimated by summing all flows terminating at a given store. Table 4 shows the ratio of observed to predicted store revenue for the four study stores derived using the disaggregate SIM. A value of 1.0 demonstrates exact correspondence between observed and predicted store revenue, a value above 1 demonstrates that the model has over-predicted revenue, whilst a value of less than 1 demonstrates an under-prediction. Table 4 shows the excellent fit between the observed and predicted revenues across the four study stores in Cornwall (used for calibration) and an additional control store (store 5) operated by the same retailer for which revenue data (but no consumer flow data) were provided. This out of town store in a Cornish tourist resort was thus used as test of model performance and 52 week average store revenue was estimated to within 4% of observed values.

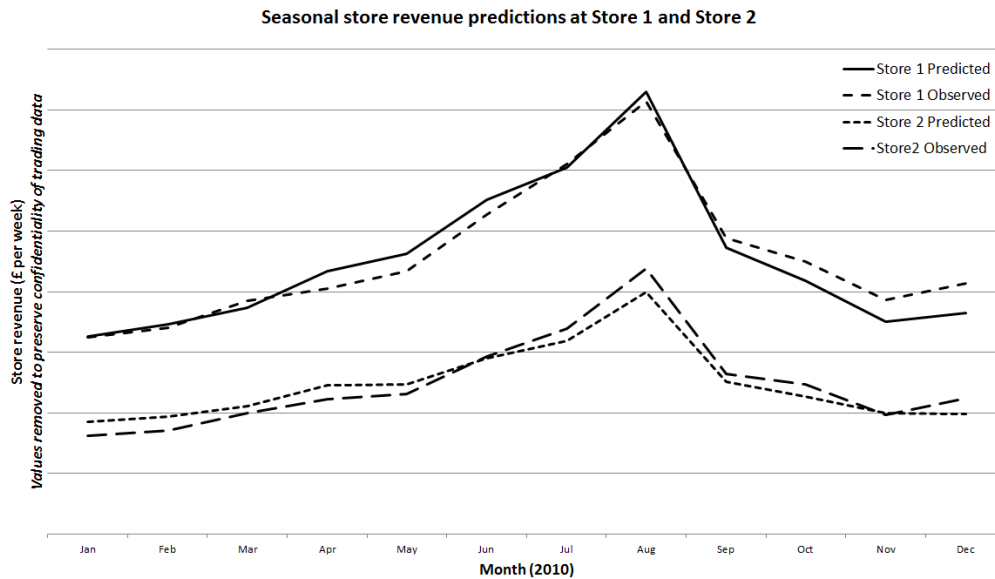
**Table 4: Observed v predicted model fits in Cornwall**

52 Week average – 2010 trading year	Status	Ratio of observed to predicted store revenue
Store 1	Calibration store	0.99
Store 2	Calibration store	1.00
Store 3	Calibration store	0.97
Store 4	Calibration store	0.98
Store 5	Control store from collaborating retailer	0.96

Whilst it is recognised that we must be cautious in using only one control store in order to assess model performance, the difficulties in obtaining data of this nature from commercial organisations should not be underestimated. The control store is located within a different part of Cornwall, and unique in comparison to the four study stores in terms of its size, facilities and catchment. The models clear ability to estimate revenue at this store, which has not been part of the calibration process, is a very encouraging sign of model performance.

It is also through revenue estimation that the impact of incorporating visitor demand can be evaluated, since seasonal variations are reflected in the store’s weekly revenue data. Since flow data is not available for visitors, it is impossible to incorporate visitor demand in model calibration based on observed and predicted flows, and reference to recorded store revenue and seasonal sales fluctuations is the only way to assess the impact of the inclusion of visitor demand. Retailers traditionally think of store revenue on a weekly basis and as such our seasonal demand estimates consider average weekly demand on a month-by-month basis.

Observed average weekly store revenue can thus be compared to predicted average weekly store revenue for each seasonal time period (Figure 3).



**Figure 3 - Observed versus predicted store revenue at Store 1 and Store 2**

Figure 3 shows the excellent fit again between observed and predicted revenues at two highly seasonal stores, both located in major coastal resorts in Cornwall. Although actual values have been removed in order to preserve confidentiality, the ratio of monthly observed to predicted revenue at both stores is consistently within 15% (and in many cases to within 5%), demonstrating our confidence in the model performance. Comparison to revenue estimations (not shown) based solely on residential demand, without the inclusion of seasonally induced visitor demand, demonstrates considerable improvement in the robustness of revenue estimation, particularly during the peak summer tourist season, when use of residential demand alone was seen to under predict revenue at some stores by almost 50% (Newing et al., 2013b) .

It is the ability of the model to predict expenditure flows and subsequent store revenue for other stores and operators that represents the crucial test of model accuracy. Birkin et al.



(2010a) note that “undertaking predictive experiments is the only realistic way to prove that models work”. Typically, these predictive experiments involve testing the predictive capacity of the model against additional stores for which data is held, but which have not formed part of the model development or calibration. Ideally these should be competitor stores in order to demonstrate that the model assumptions and parameters hold true across all competing retailers. Average weekly revenue predictions (52 week average) were also obtained for three stores operated by a competing retailer in Cornwall. These revenue predictions were derived from our collaborating retailers’ own assessment of competitor performance and from a comprehensive independently carried out ‘Cornwall Retail Study’ (CRS) (GVA Grimley, 2010). Whilst it is impossible to verify the accuracy of the CRS or collaborating retailer revenue predictions, the close correspondence between both organisations independent estimates is encouraging. Modelled revenue predictions were within 5% of the independently predicted revenue at these stores (taking the average of our collaborating retailers own assessment and the CRS estimate). Since this retailer traditionally attracts a very different type of consumer to our collaborating retailer, the very close correspondence between modelled revenue and independent revenue predictions suggests that the incorporation of the brand attractiveness via the alpha parameter has generated a model with robust predictive capacity.

Writing in 2010, Birkin et al. (2010a) asserted that there remains a lack of papers within the academic modelling literature that consider issues encountered when seeking to apply spatial location-based models in commercial contexts (where the needs of clients and the limitations inherent in their data need to be taken into account). This paper clearly represents one such application and Ince and Jackson (2012) assert that it is increasingly important for retailers to exploit the potential of academic research in order to best prepare themselves for continued challenges and opportunities in this sector. By engaging in the research reported within this

paper, our commercial partner has benefitted directly from an established modelling framework that has been applied to support new store development within Cornwall. The SIM and associated demand estimates can be used to predict consumer flows, revenue and associated market share for proposed stores in tourist resorts and reflects an industry-wide interest in understanding store-level demand. There are thus clear benefits available to commercial partners through collaboration with academic researchers. Our collaborating retailer, and ultimately this retail sector, is able to develop similar disaggregated demand estimates, utilising their own understanding of their brand positioning, in order to develop and enhance store level revenue estimation, as discussed fully in section 7.

## **7. Conclusions**

The spatial interaction model has been a widely applied tool in retail location analysis. A number of the largest UK retailers are known to have developed and calibrated models to a high level of accuracy. Some of this disaggregation has been explored in the literature to date. However, we believe that there has been more work on disaggregating supply-side factors than there has been on developing more effective ways of handling complexities on the demand-side. In this paper we have sought to give greater consideration to brand choice and store location by geodemographic status within spatial interaction modelling and also to non-residential demand, particularly in areas experiencing high levels of tourism. We have also sought to demonstrate that such disaggregated models need better data for calibration purposes. Without such data, model extensions are likely to remain theoretical only. With such data, in this case provided by collaboration with a major UK grocery firm, the models can be shown to produce extremely good forecasts and predictions concerning store patronage and store revenues.

In their review and experience of applied spatial interaction modelling, Birkin et al. (2010a, p442) note that “models must be seen to work in the most obvious sense – they must reproduce known trip patterns and store revenues”, if they are to be taken seriously by retailers. We hope we have demonstrated that statistically, spatially and in terms of revenue estimates, the new disaggregate model presented here, with its extensions in relation to demand, is able to replicate known flows to a very high level of accuracy. First, we have been able to include much more realistic behaviour regarding store selection – thus the attractiveness of every individual outlet is measured not just by size and brand, but also by person type, with higher income customers drawn more to higher-end grocery retailers such as Sainsbury’s, Waitrose and Marks and Spencer’s in the UK. Second, we have shown how it is possible to add a tourist demand layer which can make considerable improvements to models built only with residential demand included.

The end product is that when considering 52 week average flows, the model can predict revenue to within 5% at five stores for which revenue information is held (and in a number of cases within 2-3%). That said, more research would still be useful on understanding the remaining small error margins. The stores in coastal resorts are inevitably far harder to model, not just because of seasonal demand fluctuations, but also due to the location of these stores offering car parking and other facilities in close proximity to the beaches, town centre and nearby attractions. Thus, there is probably another element of store attractiveness which could be added in relation to the micro geographies of certain locations. However, perhaps model fit ratios of 95% will be acceptable to all given inevitable noise around consumer behaviour modelling, and the need to be able to apply these models across entire store networks.

## **Bibliography**

- Allen, P. 2008. Comparing clarifications: Rural and Urban with Area. *Regional Trends*, 40(1), pp.21-30
- Arentze, T., Oppewal, H. and Timmermans, H. 2005. A multipurpose shopping trip model to assess retail agglomeration effects. *Journal of Marketing Research*, 42(1), pp.109-115
- Batty, M. and Mackie, S. 1972. The calibration of gravity, entropy and related models of spatial interaction *Environment and Planning A*, 4, pp.205-233
- BH&HPA. 2012. Economic contribution: holiday and touring parks across the UK, The British Holiday and Home Parks Association, Gloucester.
- Birkin, M. and Clarke, G. 1991. Spatial interaction in geography. *Geography Review*, 4(5), pp.16-21
- Birkin, M., Clarke, G. and Clarke, M. 2002. *Retail Geography and Intelligent Network Planning*. Wiley, Chichester.
- Birkin, M., Clarke, G. and Clarke, M. 2010a. Refining and operationalising entropy-maximising models for business applications *Geographical Analysis*, 42(4), pp.422-445
- Birkin, M., Clarke, G., Clarke, M. and Wilson, A. 2010b The achievements and future potential of applied quantitative geography: a case study. Unpublished working paper: School of Geography, University of Leeds. Copy available from martin.c.clarke@btinternet.com.
- Birkin, M., Clarke, G., Clarke, M. and Culf, R. 2004. Using spatial models to solve difficult retail location problems. In: J. STILLWELL and G. CLARKE, eds. *Applied GIS and Spatial Analysis*. Wiley, Chichester, pp.35-53.
- Birkin, M. and Foulger, F. 1992. Sales performance and sales forecasting using spatial interaction modelling: the WH Smith approach, School of Geography: University of Leeds, Leeds.
- Birkin, M. and Heppenstall, A. 2011. Extending Spatial Interaction Models with Agents for Understanding Relationships in a Dynamic Retail Market. *Urban Studies Research*, 2011, pp.1-12
- BH&HPA. 2012. Economic contribution: holiday and touring parks across the UK, The British Holiday and Home Parks Association, Gloucester.

- Borgers, A. and Timmermans, H. 1986. A Model of Pedestrian Route Choice and Demand for Retail Facilities within Inner-City Shopping Areas. *Geographical Analysis*, 18(2), pp.115-128
- Clarke, I., Hallsworth, A., Jackson, P., de Kervenoael, R. and Perez del Aguila, R. 2006. Retail restructuring and consumer choice 1. Long-term local changes in consumer behaviour: Portsmouth, 1980 - 2002. *Environment and Planning A*, 38, pp.25-46
- Clarke, I., Kirkup, M. and Oppewal, H. 2012. Consumer satisfaction with local retail diversity in the UK: effects of supermarket access, brand variety, and social deprivation. *Environment and Planning A*, 44(8), pp.1896-1911
- Cornwall Single Issue Panel. 2004. The impact of tourism on Health Services in Cornwall, Cornwall Single Issue Panel (now part of Cornwall Council), Truro.
- DCLG. 2006. Good Practice Guide on Planning for Tourism, Department for Communities and Local Governemtn, London.
- Dudding, V. and Ryan, C. 2000. The impacts of tourism on a rural retail sector: a New Zealand case study. *Tourism Economics*, 6, pp.301-319
- Dugmore, K. 2013. The census and future provision of population statistics in England and Wales - Public consultation September 2013. Response by the Demographics User Group (DUG) Demographics User Group, London.
- ETC. 2002. Visits to Friends and Relatives, English Tourism Council, London.
- Fotheringham, A. S. 1983. Some theoretical aspects of destination choice and their relevance to production-constrained gravity models. *Environment and Planning A*, 15, pp.1121-1132
- Fotheringham, A. S. 1984. Spatial Flows and Spatial Patterns. *Environment and Planning A*, 16, pp.529-543
- Fotheringham, A. S. 1986. Modelling hierarchical destination choice. *Environment and Planning A*, 18(1), pp.401-418
- Fotheringham, A. S. and Knudsen, D. C. 1987. Goodness-of-fit Statistics. Concepts and Techniques in Modern Geography. Geo Books, Norwich.
- Fotheringham, A. S. and O'Kelly, M. 1989. Spatial Interaction Models: Formulations and Applications. Kluwer Academic Publishers, Dordrecht.
- Fotheringham, A. S., & Trew, R. 1993. Chain image and store-choice modeling: the effects of income and race. *Environment and Planning A*, 25(2), pp.179-196.
- GVA Grimley. 2010. Cornwall Retail Study, GVA Grimley, Bristol.

- Harland, K. 2008. Journey to Learn: Geographical Mobility and Education Provision. Doctor of Philosophy Thesis, University of Leeds.
- Ince, M. and Jackson, B. 2012. Retail revolution. *Society Now* (Autumn 2012).
- Jackson, P., Perez del Aguila, R., Clarke, I., Hallsworth, A., de Kerwenael, R. and Kirkip, M. 2006. Retail restructuring and consumer choice 2. Understanding consumer choice at the household level. *Environment and Planning A*, 38, pp.47-67
- Johns, N. and Lynch, P. 2007. The self-catering accommodation market: A review of electronic and other sources. *Hospitality Management*, 26, pp.293-309
- Khawaldah, H., Birkin, M. and Clarke, G. 2012. A review of two alternative retail impact assessment techniques: the case of Silverburn in Scotland. *Town Planning Review*, 83(2), pp.233-260
- Kirkup, M., De Kerwenael, R., Hallsworth, A., Clarke, I., Jackson, P. and Perez del Aguila, R. 2004. Inequalities in retail choice: exploring consumer experiences in suburban neighbourhoods. *International Journal of Retail & Distribution Management*, 32(11/12), pp.511-523
- Knudsen, D. C. and Fotheringham, A. S. 1986. Matrix comparison, goodness-of-fit and spatial interaction modelling. *International Regional Science Review*, 10(2), pp.127-147
- Krider, R. E. and Putler, D. S. 2013. Which Birds of a Feather Flock Together? Clustering and Avoidance Patterns of Similar Retail Outlets. *Geographical Analysis*, 45(2), pp.123-149
- McLafferty, S. and Ghosh, A. 1986. Multipurpose Shopping and the Location of Retail Firms. *Geographical Analysis*, 18(3), pp.215 - 226
- Mintel. 2012. The Retailing of Food and Drink [online]. [Accessed 17th August 2013 ]. Available from: <http://academic.mintel.com/>.
- Mottiar, Z. 2006. Holiday Home Owners, a Route to Sustainable Tourism Development? An Economic Analysis of Tourist Expenditure Data. *Journal of Sustainable Tourism*, 14(6), pp.582-599
- Mulligan, G. 1983. Consumer Demand and Multipurpose Shopping Behavior. *Geographical Analysis*, 15(1), pp.76-81
- Mulligan, G. 1987. Consumer Travel Behavior: Extensions of a Multipurpose Shopping Model. *Geographical Analysis*, 19(4), pp.364 - 375

- Newing, A., Clarke, G. and Clarke, M. 2013a. Visitor expenditure estimation for grocery store location planning: A case study of Cornwall. *International Review of Retail, Distribution and Consumer Research*, pp.221-244
- Newing, A., Clarke, G. and Clarke, M. 2013b. Identifying seasonal variations in store-level visitor grocery demand. *International Journal of Retail & Distribution Management*, 41(6), pp.477-492
- O'Kelly, M. 1981. A Model of the Demand for Retail Facilities, Incorporating Multistop, Multipurpose Trips. *Geographical Analysis*, 13(2), pp.134-147
- ONS. 2011. *Family Spending: A report on the 2010 Living Costs and Food Survey*, Office for National Statistics, Newport.
- Openshaw, S. 1975. Some theoretical and applied aspects of spatial interaction shopping models. *Concepts and Techniques in Modern Geography*. Geo Abstracts, Norwich.
- Oppewal, H., Timmermans, H. J. P. and Louviere, J. J. 1997. Modelling the effects of shopping centre size and store variety on consumer choice behaviour *Environment and Planning A*, 29(6), pp.1073-1090
- Ottensmann, J. 1997. Partially constrained gravity models for predicting spatial interactions with elastic demand *Environment and Planning A*, 29, pp.975-988
- Pacione, M. 1974. Measures of the attraction factor; a possible alternative. *Area*, 6, pp.279-282
- Pooler, J. 1994. A Family of Relaxed Spatial Interaction Models. *The Professional Geographer* 46(2), pp.210-217
- Quinn, D. 2010. *Holiday Home, Sweet Home: A Phenomenological Approach to Second Home Living in Ireland*. Doctor of Philosophy Thesis, University of Wales Institute, Cardiff.
- Reynolds, J. and Wood, S. 2010. Location decision making in retail firms: evolution and challenge. *International Journal of Retail & Distribution Management*, 38(11/12), pp.828-845
- Solgaard, H. S., & Hansen, T. 2003. A hierarchical Bayes model of choice between supermarket formats. *Journal of Retailing and Consumer Services*, 10(3), pp.169-180
- South West Tourism. 2005. *TOWARDS 2015: Shaping Tomorrow's Tourism (Strategy Document)* South West Tourism, Exeter.
- South West Tourism. 2010. *Value of Tourism 2008: Cornwall*, South West Tourism Research Unit, Exeter.

- Spencer, A. H. 1978. Deriving measures of attractiveness for shopping centres. *Regional Studies*, 12, pp.713-726
- Thomas, M. 2012. Exploring the potential of microdata from a large commercial survey for the analysis of demographic and lifestyle characteristics of internal migration in Great Britain. Working paper 12/03, School of Geography, University of Leeds, Leeds.
- Thompson, C., Clarke, G., Clarke, M. and Stillwell, J. 2012. Modelling the future opportunities for deep discount food retailing in the UK. *The International Review of Retail, Distribution and Consumer Research*, 22(2), pp.143-170
- Thompson, C., Stillwell, J. and Clarke, M. 2010. The changing grocery market in Yorkshire and Humber. *The Yorkshire and Humber Regional Review*, Winter 2010, pp.12-14
- Timothy, D. 2005. *Shopping Tourism, Retailing and Leisure Aspects of Tourism*. Channel View Publications, Clevedon.
- Vickers, D. and Rees, P. 2006. Introducing the Area Classification of Output Areas. *Population Trends* 125, pp.15-29
- VisitCornwall. 2010. *Tourism in Cornwall: 1992 to 2008*, Visit Cornwall, Truro.
- White, S. 2010. *Measuring Tourism Locally: Guidance Note Five: Measuring the Supply Side of Tourism* Office for National Statistics, Newport.
- Wilson, A. G. 1971. A family of spatial interaction models, and associated developments. *Environment and Planning* 3, pp.1-32
- Wilson, A. G. 2010. Entropy in urban and regional modelling: retrospect and prospect. *Geographical Analysis*, 42(4), pp.364-394
- Wrigley, N. and Dunn, R. 1984. Stochastic panel-data models of urban shopping behaviour: 3. The interaction of store choice and brand choice. *Environment and Planning A*, 16, pp.1221 - 1236