Evaluating alternative methods for forecasting convenience grocery store sales

Nicholas Andrew Hood

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds
School of Geography

September 2016

The candidate confirms that the work submitted is his/her own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

The segmentation of the convenience grocery market in Yorkshire and the Humber reported in chapter 6 of this thesis is in print at the time of submission and is detailed below:

> Hood, N., Clarke, G.P and Clarke, M (2015) Segmenting the growing UK convenience store market for retail location planning. *The International Review of Retail, Distribution and Consumer Research*. **26**(2), pp. 113-136.

This paper also includes the result of the analysis in chapter 5 of this thesis and includes the final table of analysis quantifying the growth of the convenience grocery market for the four largest grocery retailers in GB from 2003-2012. Moreover, the paper includes the spatial battle for the convenience market reported in chapter 5 of this thesis containing the market shares of each of the prominent convenience retailers' at the postal area level in GB.

The research within the paper detailed above was carried out by the first named author. The manuscript was prepared by the first author and input from the co-authors was advisory and editorial.

Acknowledgements

Abstract

Convenience grocery stores have become more commonplace in grocery retailing in Great Britain since the 1990s, with a substantial increase in the proportion of stores operated by the largest grocery retailers in Great Britain that can be defined as convenience grocery stores. Geographically, the convenience networks operated by the largest retailers are more spatially concentrated than their overall grocery networks bringing them into direct competition with retailers more traditionally associated with convenience retailing in some, but not all areas of Great Britain.

As convenience stores have grown, so too has interest in site location research in finding techniques to best predict their success. This thesis is carried out with the support of Sainsbury's and GMAP Ltd and specifically considers location based decision making for convenience grocery stores in Great Britain. Grocery retailers and their location planning teams employ models that are adept at predicting supermarket revenue. However, they find it more difficult to consistently estimate revenue to new or existing convenience store locations. From the outset of this research it was hypothesised that different locations in which convenience grocery stores are found may, in theory, require a different optimal methodology for forecasting revenue accurately. This thesis first offers a segmentation of the convenience market into 7 statistically distinct location types to begin to address this problem.

Using the 7 location types as a framework, three methodologies for forecasting grocery sales are tested for their suitability for predicting convenience grocery sales in the different locations in which convenience grocery stores are found. These are: GIS buffer and overlay modelling, regression modelling and spatial interaction modelling. The different methods were found to have mixed success in predicting convenience store revenue. The regression model was found to be the most effective model on average whilst the spatial interaction model was found to be the best model for generating very good revenue forecasts. Contrary to popular belief, the GIS buffer and overlay model was outperformed by the regression model and spatial interaction model in the majority of locations in which convenience grocery stores are found. Overall, the modelling frameworks presented in this thesis provide a plausible kitbag of techniques which can be applied in different convenience location circumstances.

Table of Contents

List of Tables

List of Figures

List of abbreviations

ACS	Association of Convenience Stores
ATD	Average Trip Distance
DoE	Department of the Environment
DEFRA	Department for Environment, Food and Rural Affairds
EDM	Early Day Motion
ESRC	Economic and Social Research Council
FES	Family Expenditure Survey
FOE	Friends of the Earth
GB	Great Britain
GIS	Geographical Information Systems
GMAP	Geographical Modelling and Planning
GOF	Goodness of Fit
GOR	Government Office Region
IGD	Institute for Grocery Distribution
LCF	Living Costs and Food Survey
LFS	Labour Force Survey
LSOA	Lower Layer Super Output Area
M&S	Marks and Spencer
NFS	National Food Survey
NFWI	National Federation of Women's Institutes
NSPD	National Statistics Postcode Directory
NUTS	Nomenclature of Territorial Units for Statistics
MAUP	Modifiable Areal Unit Problem
MSOA	Middle Layer Super Output Area
PLC	Public Limited Company
OA	Output Area
OAC	Output Area Classification
OFT	Office of Fair Trading
OLS	Ordinary Least Squares
ONS	Office for National Statistics
PA	Postal Area
PPG	Planning Policy Guidance Note
RIBEN	Retail Industry Engagement Network
RPM	Resale Price Maintenance
SOA	Super Output Areas
SIM	Spatial Interaction Model
SRMSE	Standardised Root Mean Square Error
UK	United Kingdom
WPZ	Workplace Zone

Chapter 1 - Introduction

The research presented in this thesis is the result of an ESRC Retail Industry Business Engagement Network (RIBEN) studentship. The RIBEN project aimed to encourage collaboration between universities and the retail industry by working on issues of relevance to both university research and the retail sector. The research reported in this thesis was undertaken as part of a partnership between University of Leeds and two retail industry partners; GMAP and Sainsbury's. The Geographical Modelling and Planning (GMAP) team are part of Callcredit information Group and are experts in market analysis, retail network planning and scenario modelling in Asia Pacific, Europe and the Americas. Sainsbury's are a multinational grocery retailer based in Great Britain and have the second largest share of the grocery market with 16.1% in August 2016 (Kantar World Panel, 2016).

Location Planning Problem

Sainsbury's operate an in-house 'Location Network and Planning team' specialising in analytics in the grocery industry. Location planning is an integral part of the grocery industry and the largest retailers operate dedicated teams to inform the decision making of retailers. Reynolds and Wood (2010) highlight a number of functions served by these location planning teams which include:

- Site screening and selection
- Competitor analysis
- Catchment area identification
- Monitoring store or branch performance
- Analyses of trade cannibalisation by own or competitor behaviour
- Setting store and regional sales targets
- Customer profiling
- Network review and planning
- Market share mapping and analysis
- Store portfolio segmentation and planning
- Acquisition and merger planning
- Merchandising mix analysis
- Targeting direct mail
- Promotional and media analysis
- Logistics planning

One of the major roles of these teams is to evaluate potential new store location sites by estimating the revenue that could be expected if a retailer were to open a store in a given location (Birkin et al. 2014). This informs a strategically integral function within these large grocery businesses as it aims to ensure that only sites with potential to make net profits are opened by the retailer. In order to perform many of their functions, location planning teams must have robust and reliable models that capture both supply and demand conditions of the grocery market in Great Britain.

A specific modelling team responsible for developing and maintain forecasting tools sits within Sainsbury's Location and Network Planning team. It is with this team that the author has maintained correspondence throughout this project. This team identified a particular problem regularly encountered by the retailer in which they find it more difficult to accurately predict sales to smaller convenience grocery stores in comparison with their large supermarket and hypermarket revenue estimates. Sales to larger supermarkets operated by the major retailers in Great Britain are often driven by demand primarily originating from residential consumers and goods are generally purchased as part of relatively predictable trips, either directly to the grocery store or as part of easily identifiable linked trip behaviour: i.e. trips to an out-of-town retail centre comprising stores from heterogeneous sectors of the retail market (e.g. grocery, fashion, electricals). The volume and characteristics of residential grocery demand are well understood and surveys disaggregating spending by types of consumer are readily available.

Research from the grocery market experts IGD anticipates that the UK grocery market will be worth £179.1bn by the end of 2016 and convenience retailing will account for approximately £67.2bn (37.5% of the total grocery market in the UK) having grown extensively in recent years (IGD, 2016). The convenience market is expected to continue growing and present a large opportunity for growing the grocery businesses of the major grocery multiples. Sainsbury's is one such retailer that has exploited this channel in order to maintain and grow their share of the grocery market in GB. However, in order to maintain and extend their success in this market, the retailer must be able to accurately evaluate potential and existing store sites in terms of the revenue that can be anticipated.

Sainsbury's rely on an in-house spatial interaction model (SIM) with its origins in Newtonian gravity modelling when forecasting supermarket revenue and this model effectively captures the supply and demand considerations required to accurately

predict revenue to large grocery stores in GB. However, they find that this model is less effective in estimating convenience store revenue and thus they rely more heavily on other methods to forecast existing stores and potential new locations for this type of store. Convenience stores present a different set of challenges to supermarkets when looking at predicting the revenue that this type of store will achieve. In this thesis, the conventional definition of the convenience market is followed, namely grocery stores of 3000 sq. ft. or smaller (Kirby 1986; Baron et al. 2001). This is important as stores below this size have been exempt from the Sunday Trading Laws which have restricted larger grocery stores to six hours of trading on a Sunday in England and Wales.

Convenience stores have presented a challenge to major grocery retailers in GB in terms of estimating revenue for a number of reasons. These stores can be located in very different types of places – rural villages, city centre train stations, suburban town centres. Each of these different types of location could, in theory, require a different optimal methodology for sales forecasting. Moreover, the places in which these stores are located are often situated in catchments containing a more complex set of grocery destinations than the larger out-of-town supermarket format. For example, neighbourhood convenience stores are likely to be competing for business with major retailers, smaller retail operators such as symbol group retailers or Co-operative group stores and independent retailers. This creates a greater challenge when developing supply side layers in a modelling framework designed to forecast convenience store sales. The overall aim of this research is to develop a model (or series of models) that more effectively forecast revenue to convenience stores in the different locational contexts in which they are found.

A further issue in the forecasting of convenience store revenue is the more complex (or at least greater) range of interactions between consumers and stores that is present in this market. Retailers describe the reason for a person visiting a store as the customer mission and they use information on this to inform decision making such as the location of certain goods within a grocery store. Examples of customer missions upon entering a convenience store are numerous, diverse, and include; weekly one stop shopping trips in more isolated rural neighbourhoods, extra trips to supplement larger grocery shops (e.g. to purchase bread or milk), passing trade such as consumers buying goods at train station stores whilst commuting, work-based lunch shopping, and service station purchases as part of long car journeys. The varying customer missions in relation to store visits creates a modelling challenge in quantifying and placing

parameters on human behaviour in any model and accounting for available competing destinations that a consumer may choose from. The different missions identified above produces customer trips that vary in a number of their characteristics including distance travelled, goods purchased and total spend. Moreover, for the different customer missions, there will be varying contributors to the likelihood of a potential customer being in a store catchment at any given time and accounting for potential reasons for these simultaneously for different potential destinations can be very difficult.

One further issue of note is the process by which major grocery retailers, and in particular Sainsbury's and Tesco, came to operate a large portfolio of convenience stores in GB. Many of the stores that they now operate were acquired very quickly in bulk mergers or acquisitions of smaller retail chains in the early to mid-2000s. These acquisitions were often made on the basis of the overall property value of the store networks purchased and individual forecasting of store revenues for each potential site were less rigorous than equivalent forecasting of potential sites for larger supermarkets. As a result, less is known about the accuracy of the forecasting that took place at this time and how effective the retailer has been in forecasting convenience grocery store revenue.

Moreover, the retailer are understandably cautious in terms of releasing information on their sales forecasting process in a highly competitive market. From an academic perspective this can make it difficult to judge the quality of academic location analysis work against the quality of forecasts devised by the retailers themselves. However, Sainsbury's provided revenue data for 95 convenience grocery stores in Yorkshire and the Humber and more latterly provided revenue data for a further 31 convenience grocery stores in the North West of England for the purposes of validation. It is these datasets that have formed and shaped the direction of this research project. In order to contribute to the established knowledge of the convenience grocery market in GB, and to attempt to boost the accuracy of forecasting of convenience grocery stores, a number of aims and objectives for the research project were formulated.

Aims and objectives

This research strives for an improvement in established convenience store site location evaluation in GB that can be used to more accurately forecast convenience store revenue in the varying location types in which convenience grocery stores are found. The overall aims of this research are as follows:

1. To review the existing academic and industry literature on the convenience grocery market in Great Britain in relation to; the growth of major retailers in the convenience grocery market, the growing demand for convenience groceries and the attempts at forecasting revenue of both convenience grocery stores and grocery stores more generally both in academic and in the retail industry.

2. To quantify the extent to which major retailers have committed to the convenience grocery market and assess the geographical extent to which they play a role in convenience grocery retailing in Great Britain.

3. To develop and test a series of predictive models for forecasting convenience grocery store revenue in the varying location types in which this type of grocery store is found.

In order to achieve these three broad aims, and in particular the third and main aim of the research reported in this thesis, the three aims were broken down into smaller objectives by which the main aims would be realised. The objectives of this research as to:

- Review the literature on the conditions by which major grocery retailers came to be active in the convenience grocery market in GB (Chapter 2).

- Assess the demand for convenience grocery retailing in GB (Chapter 2).

- Review existing revenue forecasting methodologies used in grocery retailing (and convenience grocery retailing in particular) both in the academic literature and in applied store location analysis within retail organisations (Chapter 3).

- Quantify the growth of the four largest grocery retailers' convenience store networks both nationally and regionally in GB and situate this network in the wider grocery operations of each retailer (Chapter 5).

- Ascertain the geographical extent of the convenience grocery network of each of the four largest grocery retailers in GB as a proportion of the total convenience grocery retailing taking place in each postal area in GB and compare this to the overall grocery network of each of the four largest grocery retailers (Chapter 5).

- Using cluster analysis, disaggregate convenience grocery store locations in Yorkshire and the Humber into statistically distinct location types in order to explore different approaches to predicting store revenue (Chapter 6).

- Develop and test three distinct methodologies for predicting Sainsbury's convenience grocery store revenues in Yorkshire and the Humber. (Chapters 7, 8, and 9).
- Compare the effectiveness of each of the three modelling approaches by forecasting accuracy in each of the location types identified in chapter 6. (Chapter 10).
- Validate the effectiveness of each model by testing the capacity of each model to replicate findings in an additional study region. (Chapters 7, 8 and 9).

Thesis structure, scope and contribution

In order to meet the aims and objectives of the research, this thesis is structured as follows. Chapter 2 reviews the development of the supply side of grocery retailing in Great Britain attempting to ascertain the conditions by which convenience grocery retailing came to form a large part of the store network of two major retailers in particular, Tesco and Sainsbury's. In doing so, the chapter reviews the changes seen in planning policy, grocery market changes and the diversification of the strategy of major retailers in response to a number of factors. Chapter 2 also contains a review of the evidence supporting a demand for convenience in the grocery shopping process alongside a review of increasing societal changes which have contributed to the success of major retailers' convenience grocery stores and resulted in them becoming a major source of grocery shopping destination for many consumers in Great Britain.

Chapter 3 reviews the academic and industry literature pertaining to the methods used to predict the revenue of grocery stores, concentrating on specific attempts to forecast convenience grocery store revenue. The chapter introduces the three methodologies for estimating store revenue adopted in this thesis, focusing particularly on the use of spatial interaction modelling and linear regression modelling in revenue forecasting but also introducing the GIS buffer and overlay technique (making up the three approaches used in this research project). In doing so, it justifies the use of the three methodologies used in this research whilst giving a brief justification for those methods that were not used in this work.

Chapter 4 describes the data, study area, and geography used to achieve the aims of this research, setting up the main analysis chapters which begin in chapter 5 of this thesis. Chapter 5 looks at the changes seen in the grocery network of each of the four largest grocery retailers in GB at the outset of this study. These major grocery multiples are; Tesco, ASDA, Sainsbury's, and Morrisons. The chapter quantifies the growth of

the convenience grocery stores and convenience grocery floorspace for each of these retailers over a ten year period beginning in 2003. The changes to the convenience grocery floorspace operated by each of the grocery retailers is situated in the wider grocery operations of each retailer in an attempt to quantify the relative importance of convenience grocery retailing to each of the four largest grocery retailers in Great Britain. The final analysis presented in Chapter 5 disaggregates the overall supply of major retailer convenience grocery operations by using a regional geography in GB at the end of the ten year period for which data was provided.

Chapter 5 explores the geographical extent of the convenience grocery network of the grocery retailers that are most heavily involved in the convenience grocery market in GB. This is achieved by assessing the share of the convenience grocery market commanded by each retailer expressed as the proportion of total convenience grocery floorspace operated by each retailer in each postal area in GB. This provides an extra level of geographic disaggregation of the convenience grocery network, identifying and analysing the types of location around the country (parts of GB) that each retailer has chosen to site its convenience operations in.

Whilst chapter 5 assesses the macro geographic locations in which different retailers have chosen to locate their convenience stores, chapter 6 looks more closely at the types of micro location that convenience grocery retailing takes place in Great Britain. The analysis in this chapter is a k-means cluster analysis which segments the convenience grocery network in Yorkshire and the Humber into 7 distinct location types in which convenience grocery retailing takes place. This analysis is designed facilitate the evaluation of different methods for forecasting convenience grocery store based on the theory that each of these different types of location could, in theory, require a different optimal methodology for sales forecasting. The output of this research results in each store being assigned a cluster type identifying the types of location that each retailer has chosen to locate in and allowing for a comparison of the location strategy of major retailers and other smaller retail chains.

Chapter 6 sets up the analysis reported in the final 4 chapters of this thesis and the location types identified form the basis for assessing the effectiveness of the different methodological approaches to forecasting convenience grocery store revenue in chapters 8 to 11. These chapters analyse the effectiveness of the three methodologically distinct approaches to forecasting convenience grocery store revenue in Yorkshire and the Humber, the main study region of this thesis. Chapter 7 reports on the development and results of a GIS buffer and overlay approach to forecasting

convenience grocery store revenue, chapter 8 reports on the development and results of a spatial interaction model to forecast convenience grocery store revenue. Chapter 9 reports on the development and results of a linear regression approach to forecasting convenience grocery store revenue. The reporting of the results of each model include two indicators of model performance. These are: 1) global model performance in predicting sales of all stores, and 2) model performance in predicting sales of stores by the distinct cluster types identified in the segmentation of the convenience grocery market in chapter 6.

Chapter 10 compares and summarises the effectiveness of each of the three modelling frameworks in forecasting convenience grocery sales both across all location types and in each specific location context as identified in the cluster analysis in chapter 6.. Finally, in conclusion, suggests a future agenda for advancing convenience store revenue forecasting both in academic research and in the grocery industry whilst identifying the limitations of the analysis conducted as part of this thesis.

Contribution and outputs of this thesis

As noted above, this thesis reports on a collaboration between GMAP, Sainsbury's and the University of Leeds as part of an ESRC funded Retail Industry Business Engagement Network studentship. An important aspect of this relationship is the commercially sensitive data that is made available as part of the research project. As part of this project, revenue data for 95 convenience grocery stores in Yorkshire and the Humber and 31 convenience grocery stores in the North West of England were made available. This is an unusual volume of sales data to be made available to researchers in the UK and provides a unique opportunity to empirically test theory developed in academic work on grocery store revenue forecasting.

Birkin et al. (2014) highlight the lack of papers in the academic literature on attempts to apply spatial location models in commercial contexts. Whilst the work of Andy Newing and others at University of Leeds (Newing, 2013; Newing et al. 2013a; Newing at al. 2013b; Newing et al. 2014; Newing et al. 2015) contributed to this body of literature, with specific reference to accounting for seasonal demand for groceries driven by tourism, a dearth of academic literature remains. This thesis and associated papers aims to fill part of this gap in the specific context of convenience grocery retailing by using methods more traditionally associated with supermarket retailing. The limited literature on convenience grocery location planning has identified the difficult in applying methods more associated with supermarket forecasting to convenience store

revenue forecasting. This thesis empirically judges the extent to which these misgivings are true and is thus a major contribution made by this research project.

The results of this thesis are in the process of being disseminated in three ways. Firstly, through peer reviewed academic papers. Analysis based on the segmentation of the convenience grocery market in Yorkshire and the Humber reported in chapter 6 of this thesis is in print at the time of submission and is detailed below:

> Hood, N., Clarke, G.P. and Clarke, M (2015) Segmenting the growing UK convenience store market for retail location planning. *The International Review of Retail, Distribution and Consumer Research*. **26**(2), pp. 113-136.

This paper also includes some of the results of the analysis in chapter 5 of this thesis and includes the analysis quantifying the growth of the convenience grocery market for the four largest grocery retailers in GB from 2003-2012. Moreover, the paper includes the spatial battle for the convenience market reported in chapter 5 of this thesis containing the market shares of each of the prominent convenience retailers at the postal area level in GB.

On the back of this thesis, two further papers are planned to report on the combined results of the three approaches to modelling convenience grocery sales presented in chapter 7 to 10. The first of these papers will focus on developing and calibrating of the models and the second of these papers will focus on the application of these models in the grocery sector, looking at how they could be combined into the suite of models already adopted by major grocery companies with extensive location planning experience.

The second method of dissemination of the results of this thesis has been through the presenting of the findings at a number of international conferences. These conferences have covered the strands of research involved in this thesis and have included presenting at conferences centred on the topics of spatial modelling, GIS, retailing and big data. The final method of dissemination of the results of this thesis has been the inclusion of many of the findings in the teaching materials administered at the University of Leeds. The work has become part of the teaching syllabus and reading lists on three modules at the University of Leeds covering undergraduate and taught postgraduate geography courses. These modules are:

- GEOG 2025: Service Analysis and Planning
- GEOG 3010: Advanced Retail Planning
- GEOG 5881M Applied GIS and Retail Modelling

This has allowed major trends in the grocery industry and advances in store location planning research to be taught to students and is contributing to the continued relationship between the University of Leeds and the retail analytics industry which many students join following the completion of their studies. This research looks to enable the students to leave the University of Leeds with the skill set required to succeed in the retail analytics industry and is therefore another major contribution of this research.

To begin, chapter 2 reviews the supply and demand trends and changes that have resulted in a number of major grocery retailers becoming key players in the convenience grocery industry in GB.

Chapter 2 - A review of the literature on the rise of branded convenience grocery retailing major and the demand for convenience groceries in the population of GB

Chapter 1 introduced this research project and set out the aims and objectives of this thesis. The research presented in this thesis is based on the entry and subsequent rise in prominence of major grocery multiples into the convenience grocery market. The first aim of this research is to review the existing academic and industry literature on the convenience grocery market in Great Britain in relation to; the growth of major retailers into the convenience grocery market, the growing demand for convenience groceries, and the attempts at forecasting revenue of both convenience grocery stores and grocery stores more generally both in academic and in the retail industry. This chapter achieves the first two points in this aim, to review the growth of major retailers into the convenience grocery market and explore the demand for convenience grocery retailing in GB.

In doing so, the chapter reviews the academic literature and wider body of evidence of the changes seen in the grocery landscape in Great Britain between the 1960s and present day, reviewing the conditions by which small-format grocery retailing emerged as a strategy of major retailers attempting to consolidate and expand their share of the grocery market in the UK. The chapter is split into four sections. Section 2.1 identifies the types of grocery retailing taking place in GB focusing on the types of stores that are available and the retailers that are present in the market. Section 2.2 explores the supply side changes seen in British grocery retailing from 1964 to 2016, drawing particularly on changes seen in planning legislation and section 2.3 focuses on the strategies made by retailers to grow their market share in the GB grocery industry. Changes in planning legislation are explored as a major driver of supply side shifts in the grocery market. Thereafter, section 2.4 explores the demand for convenience grocery retailing in Great Britain, particularly looking at demographic changes as an enabler for major retailer expansion in to the small store grocery industry.

Firstly, the supply side section explores a number of key events and market shifts in the GB grocery industry between 1964 and 2016. These include:

- The repealing of Resale Price Maintenance in 1964 allowing price to be used as a tool of comparative advantage, an opportunity greatly benefiting the major retailers.

- The Golden Age of grocery retailing for the major retailers in which a select group of retailers amassed a significant proportion of market share and year on year growth was achieved through the opening of large number of supermarkets.
- The end of the Golden Age for the major grocery retailers, focusing on shifts in planning policy that occurred during this time and looking at the effect this had on retailer operations
- The responses adopted by a major retailers in response to market changes, particularly focusing on growth into the convenience grocery market.

2.1 The grocery market in Great Britain

It is important to first set out the foundations for investigating the convenience grocery market in GB by identifying the types of grocery retailing that take place and the sorts of retailers that are involved in the convenience grocery market.

2.1.1 Grocery Retailing Formats

Grocery retailing in GB takes place in a variety of channel formats. These refer to stores of different sizes alongside non hard store retailing. The IGD identify six different formats through which consumers engage with the grocery market in GB (IGD, 2016). These are:

- **Hypermarkets:** Large format stores that sell a full range of grocery items and a substantial non-food range. Sales areas are typically 60,000 sq. ft.+
- **Supermarkets:** Defined as food-focused stores with sales areas of between 3,000 and 60,000 sq. ft.
- **Convenience stores:** Stores with a sales area of less than 3,000 sq. ft., which are open for long hours and sell products from at least seven grocery categories. Includes standalone forecourts with convenience stores.
- **Discounters:** Includes food discounters Aldi, Lidl and Netto and the grocery sales of the high street discounters such as Poundland and B&M.
- **Other retailers:** Includes stores with a sales area of less than 3,000 sq. ft., typically newsagents, off-licences, some forecourts and food specialists, such as butchers and bakeries. This channel also includes the grocery sales of predominantly non-food retailers such as department stores.
- **Online:** Internet orders placed at grocers and online food specialists for home delivery and customer collection.

Convenience stores operated by branded areas have become a key feature of the grocery market supply in GB and is the focus of this thesis. Moreover, the growth of major retailers into this market has put pressure on smaller retailers, falling into the other retailers category in the definition of IGD (2016). However, many of these stores, particularly those on petrol forecourts and other food specialists below 3000 sq. ft. are often referred to as convenience stores in this thesis as they are seen to compete in the same market. As of 2016, convenience stores had become a large part of the grocery market and were worth 37.5% of the total grocery market in the UK (IGD, 2016).

2.1.2 Retailers in the grocery market in GB

There are a number of retailers operating the grocery formats described in section 2.1.1 of this chapter. These can be broadly grouped into categories based on the size of the retailer, how they run their operations and the types of stores that they have traditionally operated in the grocery market in GB. Retailers broadly fall into the following groups and are referred to as such throughout this thesis:

- **Major Grocery Multiples**: These are large grocery firms operating extensive store networks that are often spread across formats including convenience stores, supermarkets, hypermarkets and online retailing. The main retailers falling into this category in GB are Tesco, ASDA, Sainsbury's, Morrisons, Marks and Spencer and Waitrose. Much of the analysis in this thesis refers to the four largest grocery retailers in GB. These are; Tesco, ASDA, Sainsbury's and Morrisons. They were the four largest grocery retailers in GB at the commencement of the research reported in this thesis (Kantar Worldpanel, 2013).

- **Co-operative Group Retailers**: A consortium of 22 different societies across the whole of GB. Although each has its own name, this thesis often considers them as a whole. Historically, the Co-op has made the greatest commitment to growth through small-format convenience store retailing. However, co-operative group retailers also operate larger grocery stores throughout GB.

- **Discounters**: As detailed in section 2.1.1, discounters include food discounters Aldi, Lidl and Netto and the grocery sales of the high street discounters such as Poundland and B&M. The datasets of grocery retailers used in this thesis focus on the food discounters Aldi, Lidl and Netto and are thus the retailers implied

when discounters are discussed in this thesis. These retailers operate small grocery stores although the majority of them are greater than 3,000 sq. ft. and therefore do not comply with Sunday Trading Laws although they are seen to compete with major retailers convenience stores in a number of locations.

- **Symbol Group Retailers:** These are independent retailers that are members of larger umbrella retail organisations known as symbol groups (IGD, 2012). They have become major players in the convenience grocery market in GB. Musgrave group, Premier, Costcutter and Spar have all made significant inroads into small-format grocery retailing and continue to operate many stores in a number of areas across GB. Stores operated by symbol group retailers are generally below 3,000 sq. ft. and often rely on newspapers, tobacco products and alcohol to generate revenue. Retailers aligning with symbol groups are required to buy a proportion of their goods from the symbol group retailer in return for a range of benefits (IGD, 2012):
  - ➢ A branded shop fascia
  - ➢ Advantageous buying terms
  - ➢ Access to own brand ranges
  - ➢ IT and logistical support
  - ➢ Marketing and promotional programmes
  - ➢ Professional guidance and advice

- **Small and independent retailers:** These are sole grocery traders or small groups of retailers operated independently. These retailers are referred to throughout this thesis and incorporated into analysis where possible. However, data on these retailers is sparse, particularly in respect of independent single store retailers.

2.2 The changing grocery market in GB: 1964 - 2016

The present form of British grocery retailing began to emerge in the 1960s following a post war revolution bringing about self-service grocery retailing (Poole et al. 2002). This led to the emergence of a market increasingly dominated by a few major retailers operating major store networks (Guy, 1998). This has continued to the present day in which a select few retailers dominate the retail grocery market in the UK. The dynamics of the retail market have been in constant evolution from the early 1960s to the present day and the major retailers have adopted various strategies along the way to respond

to a number of changes that have occurred. This chapter discusses these changes under a series of headings, building up to towards the entrance of large multiple grocery retailers into the convenience market.

2.2.1 Resale Price Maintenance

In 1964 Resale Price Maintenance (RPM) on food products was abolished. RPM had been a regulatory red tape controlling the market artificially, in which retail companies were powerless to alter pricing and therefore the ability to gain competitive advantage through passing on cheaper costs to consumers was difficult and often of a short term nature. Abolition through the 1964 Act allowed price to be used as a competitive tool, initiating the growth of many early retail grocery multiples, the precursor of the current market dominance of a select group of large retailers (Harris and Ogbonna, 2001). Many commentators identify this as the origin of genuine competition in the retail sector in the UK and the beginning of free market trading (e.g. Gabor, 1977; Burns et al. 1983).

This led to a period in UK grocery retailing coined as the 'price wars' in which price of goods for the consumer is argued to have become the most important factor as the majority of retailers now competed directly on price (Harris and Ogbanna, 2001). This period of 'price wars' altered the retail landscape and changed the traditional retail hierarchy, placing traditional smaller retailers - often operating local convenience shopping monopolies - under threat (Poole et al. 2002). By this stage the balance of power had also started to shift to the larger retailers. Pommering (1979) identified the shift from 'manufacturers as kings' during the 1950s (due to RPM and food shortages), 'consumers as kings' in 1960s (driven by a decreasing shortage of food and an increase in competition) to 'trade is king' in the 1970s in which the major multiples began to grow and become increasingly powerful. This period began the exertion of dominance by the major retailers that would later evolve into the major retailers also diversifying into the convenience grocery market.

2.2.2 The 'Golden Age'

Price competition, alongside a rising post-war affluence and increasing disposable income led to the population transforming its food purchasing habits. Set alongside increases in income, the rise of car ownership and an increased proportion of consumers owning a fridge-freezer allowed many consumers (and an increasing proportion of the population) to do weekly shops in a single trip (Guy, 1997).The multiple grocery retailers, having initially taken advantage of the abolition of RPM, were

able to best respond to changes in consumer demands brought on by continued post-war rise in prosperity. People began demanding increases in product range and variety, including a growth in demand for international products which the large grocery multiple were best able to satisfy resulting in them growing in favour among consumers (Fernie, 1997).

The major retailers were very well placed to adapt to changing needs of consumers and did so primarily through the development of large store formats allowing for increasing economies of scale. This coincided with (and was used to take advantage of) the increasing suburbanisation of the population as the retailers chose to locate in off centre, out-of-town locations, previously unchartered territory in grocery retailing. The first out-of-town store to appear in British grocery retailing opened in West Bridgford in Nottingham in 1964 (Whysall, 2005). During this so-called 'golden era' of superstore retailing the major retailers adopted a strategy of 'spatial switching of capital', as smaller town and city centre stores were closed and large 'cathedrals of consumption' appeared regularly on the edges of UK towns and cities (Wrigley 1987, 1994). This spatial move out of town was in part driven by consumer changes but was also propelled by market forces; smaller central stores were less profitable than larger stores and run on tighter margins making them less appealing to the large multiple retailers. Tesco, for example, although reducing its number of UK stores from 552 in 1980 to 374 by 1989, increased its floorspace in the same period from 6.2 million sq. ft. to 8.5 million sq. ft. (Wrigley, 1991).

Thus, the major grocery multiples initially gained competitive advantage over smaller retailers and independents through price during the 'price wars'. However, throughout the period of growth it was recognised that location was increasingly of prime importance if firms were to continue to grow and increase their dominance in the grocery market. The need to predict revenues prior to decisions on location being made and stores being built brought store location planning to the forefront of retailer's operations. This was in part the driving force behind the move out of town to larger, more profitable stores and the beginning of an era of 'store wars' in which location became king (Wrigley, 1994). In this period the grocery retailers opened unprecedented numbers of stores and the grocery industry became a battleground for ever increasing store openings described by Howard (1995) as a 'race for space'. The commitment to location planning has continued and store location decisions remain of utmost importance to retailers.

Further structural changes played a significant role in altering the behaviour of retailers throughout this period, particularly in the form of economic shifts and government policy. The 1979-82 recession severely hampered the manufacturing base of the UK economy and the subsequent consumer-led recovery benefited the retailers over the manufacturers (Dawson, 2004). This combined with a favourable regulatory environment driven by the Thatcher Conservative Governments' laissez faire policy in terms of freedom to open stores and weakened regulation of the grocery market, created conditions in which the large multiple retailers thrived, often at the expense of smaller retailers and independent store owners.

The ascendance of major retailers continued to put pressure on the smaller grocery retailers and by the mid-1980s, although small independent outlets were still numerically superior to multiples, their market share had been squeezed by chain retailers who themselves increased national market share from 44% in 1950 to 70% in 1984 (Wrigley, 1992). The increasing domination of major retailers in the grocery industry led to an Office of Fair Trade (OFT) investigation into the grocery market in 1985 addressing concerns over the impact of increasing retailer concentration on competition, and the potential negative effect the large multiples were having on both suppliers and consumers. Despite these concerns, the investigation found no conclusive evidence of unfair advantage being gained by large multiples in the grocery industry and business was allowed to continue as usual (Burt and Sparks, 2003).

The changes seen throughout the 1960s, 1970s and 1980s resulted in unprecedented growth of a small group of five powerful multiples who by the mid-1980s had a combined 43% share of the grocery market (Poole et al. 2002). These retailers were Sainsbury's, Tesco, Asda, the Argyll Group and Dee Corporation (later Gateway). The effect of the early dominance of these retailers can still be felt in the market in the present day in which three of these five multiple grocery retailers -Tesco, Asda and Sainsbury's - are the three largest grocery retailers in the UK commanding 28.6%, 16.6% and 16.5% of the grocery market respectively in June 2015 (Kantar, 2015). The increased domination seen through the 1960s, 70s and 80s came as a result of considerable organic growth by multiples occurring alongside significant acquisitions of smaller retail chains from competitors, the latter occurring again in the growth of the branded convenience grocery market almost 30 years later.

2.2.3 The end of the 'Golden Age'

The early 1990s witnessed the beginning of the end of the so-called 'Golden Age'. A number of factors began to limit the expansion and advancement of the major grocery

multiples through their preferred method of supermarket openings, often out-of-town off-centre developments. Wrigley (1994, 1998) documents 4 key reasons for the slowing down of superstore development in the 1990s:

1. Financial problems
2. Belief in saturation
3. Arrival of the deep discounters
4. New planning policy restrictions

2.2.3.1   Financial problems

Financial problems for the retailers began in the early 1990s and heightened in 1993-94 during which major retailers were engulfed in a property crisis; particularly in the form of property overvaluation, unrecoverable initial investment and depreciation of property and land assets (Wrigley, 1998). The retailers began to acknowledge that they had paid too much for property assets which they had attained during the height of the 'store wars' driven by competition to expand. The limited space available for the most valued locations had significantly inflated the price paid by the multiples for sites in the late 1980s and they were plunged into chaos due to many of the decisions that were made in the Golden Age (Shiret, 1992), which has led the retailers to overvalue their assets (against the price which they could get for alternative land use).

The retailers' response to the overvaluation of assets was to embark on a course of asset depreciation in order to avoid a collapse in market confidence (Poole et al. 2002). The first retailer to initiate this was Argyll group who in 1993, began to gradually depreciate its store values and decrease the rate at which it developed new store locations. Tesco and Sainsbury's followed close behind with their depreciation strategies, the most significant of which was the latter writing down £365m in a single day decreasing the market valuation of the firm by £850m (Wrigley, 1996).

2.2.4 Belief that grocery saturation was imminent

Capital concentration in the UK grocery market by the late 1980s and early 1990s in which five firms controlled over 40% of the market led to a widespread belief that the domestic market was nearing saturation (Duke, 1989). The financial problems experienced by the retailers detailed above were additionally read as signalling market saturation (Poole et al. 2002). Many commentators have alluded to the apparent slowing down of superstore growth in the 1990s as an indicator of saturation approaching. However, with the exception of ASDA who experienced huge debt

problems after the expensive purchase of a number Gateway stores and subsequently drastically reduced supermarket openings – to 0 in 1992/03 - the other retailers continued to open a similar number of stores year-on-year as they had in the late 1980s (Langston et al. 1998).

What was occurring was the slowing of the rate of growth, experienced due to the sheer number of stores that they had opened over the period meaning that to maintain growth rates the retailers would have been required to open an increasing number of stores each year. Further arguments centred around a debate on profitability and an increase on stores being assumed to be placing the profits of existing stores in jeopardy. However, Myers (1993) argued that the 'Golden Age' era had been exaggerated and the growth in floorspace experienced by the multiples had not been as significant as many people argued due to the closure of networks of stores in central locations in lieu of opening large format stores out of town.

Additionally, Langston et al. (1997) argued that the issue of saturation was geographical in context and that in some areas saturation may have been far closer than in others. For example, Surrey, Tayside and Cleveland came out as having a significant provision of grocery stores per head of household and were therefore the most 'saturated', whereas other areas such as Cornwall, Dyfed and Central London remained relatively less 'saturated', and continued to present an avenue for expansion for the major retailers.

Whilst it is evident that saturation had not in fact been reached, it is understandable why many retail commentators felt that the growth levels of the multiples throughout the 1980s and early 1990s could not continue in the long run and it is likely that the retailers had at least an acknowledgement of the increase difficulty (perceived or real) at the time. Whilst they did not abandon their pursuit of large supermarkets as a primary investment mechanism, this may have prompted them to turn their attention to other avenues of growth. This took a number of forms, one being a significant investment in the convenience market by a number of retailers, investigated in more detail later in this chapter and throughout this thesis.

2.2.5 Arrival of the deep discounters

Coinciding with financial problems and a widespread belief that saturation may be imminent, the early 1990s saw the arrival of the deep discount retailers from Germany and Scandinavia into the UK retail market. The three main retailers to move into the market were Aldi, Netto and Lidl. They were drawn to the market by the high profit

margins being experienced by British food retailers (in comparison to mainland European retailing) and recognition of a gap in the market, again partly undermining the saturation thesis (Poole et al. 2002).

These retailers spotted a gap in the market in which they could offer value to consumers through cheaper prices and a spatial gap through which they could locate and serve poorer communities in more deprived towns and cities (Burt and Sparks, 1994). Up to this point, the big three retailers had generally avoided expanding their store networks in low income areas. This brought these new discount retailers into competition with the larger domestic grocery multiples; particularly in traditional blue collar communities. In the early 1990s this disproportionately affected retailers in northern England including Morrisons and Asda, leaving Tesco and Sainsbury's relatively unscathed due to their traditional southern middle-class hinterlands.

Following their initial investment in the UK, the discount market continued to rapidly grow and thrive adding increased pressure to traditional non-discount retailers and becoming less limited to the traditional battleground of northern blue collar communities (Thompson et al. 2012). These new discount retail formats provided greater competition and threat to the hegemony of the largest grocery retailers and are widely believed to have been a driving force in the adoption of different strategies by the multiple food retailers.

2.2.6 Changing planning policy

The history of retail planning policy between 1988 and the present day is significant when considering the conditions through which the convenience store format has emerged as a retail channel used by many retailers. When considering store developments, local planning authorities must take into account national planning policy guidance notes (PPGs) when considering local planning policy. As highlighted earlier, the 1980s is viewed as a time of extensive growth in the large superstore sector of grocery retailing due to laissez faire planning policy and a less regulated trading environment for the major retailers.

This began to change with the introduction of planning policy guidance note 6 (PPG6) in 1988, setting the ball rolling on a series of regulations in retail policy that have subsequently altered the landscape of the UK grocery market (Wrigley, 1994). Recent trends in the UK retail market have seen the slowing of organic growth in the out-of-town superstore sector alongside unprecedented growth in smaller format convenience grocery stores operated by major grocery multiples more traditionally associated with

large supermarket retailing. In this landscape, *"…major retailers continue to turn their attention to smaller stores, less constrained by retail regulation"* (Wood and Browne, 2007, P. 249). This section looks at the recent history of planning policy change resulting in convenience retailing being a more attractive proposition for major grocery retailers.

Initially, PPG6 supported the development of large, out-of-town superstores for reasons of consumer choice and the reduction of traffic in town centres (Pal et al., 2001). Major retailers including Tesco, Sainsbury's, Asda and Morrisons seized this opportunity to build large supermarkets, many in excess of 60,000 sq. ft., to increase their market share in the UK. In subsequent years, amid growing fears over the ability of the town centre to compete against out-of-town retail formats, opposition to major retailer supermarkets gained momentum. This was reflected in a revision of PPG6 in 1993, in which the need to 'protect and promote' town centres was incorporated into planning policy (Department of the Environment, 1993), thus setting into motion the preconditions through which the superstore ceased to be the overwhelmingly preferred format for the major retailers to increase market share.

Furthermore, PPG6 was once again adjusted in 1996 through the introduction of the sequential test, effecting retailers in that if they proposed an out-of-town site for development*, "… the onus will be on the developer to demonstrate that he or she has thoroughly assessed all potential town centre options."* (Department of the Environment, 1996, P. 6). Consequently, off-centre development was now considered only as a last resort in granting planning permission (Wood et al., 2010).

2.2.7 Competition Commission Investigations

In addition to increased pressures from revised local planning regulation, in April 1999, the conduct of large grocery retailers in the UK market was referred to the Competition Commission by the Office of Fair Trading over concerns about their relationships with suppliers and how this may be giving major retailers an unfair advantage in the UK grocery market. The commission's report published in 2000 was *"… basically favourable to supermarkets, (although) it expressed some concerns about their relationship with suppliers"* (Seely 2012, P. 3). Subsequently, a Supermarket Code of Practice was established requiring supermarket retailers that had shares of more than eighty per cent of grocery purchases for resale from their stores (judged as the point at which supermarkets could control relationships with suppliers) to give undertakings, to promise to act in a more responsible manner (Competition Commission, 2000). The

four major grocery retailers meeting this criteria; Tesco, Sainsbury's, Safeway and Asda, did so accordingly. This made it more difficult for major retailers to continue organic growth through large supermarkets, placing restrictions aimed at preventing grocery market domination by a small number of major retailers.

Concerns over breaches to the Supermarket Code of Practice by major retailers in purchasing existing convenience chains between 2002 and 2004 resulted in criticism of both the Code and the behaviour of major retailers by MPs, leading to growing support among individual and smaller retailers for changes to grocery market policy to lessen the dominance of major retailers. Parliamentary early day motion (EDM) 1248 submitted to parliament in the 2003-2004 session stated:

"That this House recognises the importance of diversity and consumer choice in grocery shopping now and in the future, especially the strong tradition of independently run local grocery convenience stores; recognises that on-going acquisitions and consolidation in the grocery market led by the major supermarket groups threatens that choice, diversity and tradition." (EDM 1248, 2004)

Additionally, the Association of Convenience Stores (ACS), Friends of the Earth (FOE), FARM and the National Federation of Women's Institutes (NFWI) urged the Office of Fair Trading (OFT) to have a renewed investigation into the grocery sector (Seely, 2012). They argued staunchly that the Competition Commission's two-market ruling distinguishing 'one-stop' shopping and 'secondary' grocery shopping as distinct separate sectors in UK grocery retailing had given a 'regulatory green light' for major retailers to acquire chains of convenience stores (Wrigley et al., 2009).

'One-stop' shopping had originally been the preserve of the major grocery retailers such as the big four discussed in depth in this chapter. This involves customers making a specific trip to acquire a large amount of groceries, often on a weekly basis. The two-market ruling angered many small and independent retailers and groups because it did not prevent the major retailers from expanding in the 'secondary grocery market', the space in which many smaller retailers were traditionally located and consumers did top-ups on their weekly groceries, bought newspapers, bought cigarettes and picked up daily newspapers. They argued that in terms of supply, *"…the issues arising from the big supermarkets' massive buying power apply to all products they sell, whether they are destined for a superstore or a convenience shop"* (ACS et al., 2004, P. 3).

The ruling meant that the major convenience stores were not considered alongside larger grocery stores when assessing if the market is fair and just in terms of competition. It was as a result of this buying power that could be transferred to smaller stores that the ACS believed the major retailers were harming the UK retail market and forcing many small and independent retailers out of business as they were unable to compete with major retailers with the advantage of comprehensive supply chains and in-house teams of analysts. Despite this opposition, the OFT found no grounds on which to refer the market to the Competition Commission at this time.

Subsequently, the All-Party Parliamentary Small Shops Group published a report in early 2006 declaring that there was widespread belief among small and independent retailers that over the proceeding ten years many small shops were forced to close as a result of a difficult, unbalanced trading environment (All-Party Parliamentary Shops Group, 2006). On the back of this report, the OFT referred the whole grocery market to the Competition Commission. Despite widespread concern, in the final report published in February 2008, the Competition Commission found no cause for concern in the major grocery firm's expansion into the convenience sector as it argued that: *"Whilst we have been sympathetic to those finding themselves under pressure in this market, particularly independent retailers, this does not mean that competition is not working well – it is often the effects of rivalry between retailers which benefit the consumer"* (Competition Commission, 2008, P.2).

Nonetheless, the report proposed the implementation of a new Groceries Supply Code of Practice and the founding of an independent Ombudsman to regulate the Code of Practice. There has not yet been any legislation preventing the growth of major retailers into the convenience sector despite claims from independents and smaller convenience chains that this has been damaging to the UK retail market. It can be seen that through the combination of PPG6 and the Competition Commission's two market ruling policy, potential store locations within town centres and other central locations have become increasingly considered by site location teams of major retailers as an alternative to large out-of-town sites suitable for large supermarkets and hypermarkets. As these locations have more limited space, the choice to diversify into smaller format stores became necessary to major grocery retailers. Whilst planning policy preventing the major retailers growing in the convenience grocery market did not occur, it no doubt left the large multiples wary of further potential investigations in the future.

2.3 Retailer responses to market changes

Facing increasing difficulty in expanding their offer in the UK grocery sector in the way in which they were extensively successful throughout the 1980s and into the 1990s, the major retailers turned to different ways to improve their store offer and to continue to grow their businesses. Wood et al. (2010) documents four such responses which helped continue retail growth in different forms. These were:

1. International growth,
2. Increased growth in non-foods
3. E-commerce
4. Store format changes (including convenience stores)

2.3.1 International growth

In the face of a slowing of domestic growth in supermarkets and hypermarkets, the major retailers were increasingly weighing up potential investments in grocery markets in other countries. This section details the history of international growth of the major UK grocery players from the 1980s to the present day. International markets appeal to retailers for a number of reasons including; the potential for cheap labour, weaker planning restrictions, competition at home, better profit margins and a huge potential for growth in newly established and growing markets (Lamey, 1997). Major UK grocery retailers first ventured into international markets prior to the barriers to growth in the large supermarket sector listed above and then more wholeheartedly committed to extending their borders and operations following the height of the challenges faced in the early 1990s.

Sainsbury's and Marks and Spencer moved into the North American grocery market in the 1980s. Sainsbury's purchased a 21% stake in the Shaw's supermarket chain in North East USA in 1987, transforming the chain from a minor to major regional chain (Wrigley, 2000). The following year Marks and Spencer's acquired the Kings Supermarket chain in the US (Burt et al. 2002). By the time of the property crisis, the major retailers had already begun to test the water of international markets and seen aggressive expansion abroad as a viable option in the face of a difficult climate for growth at home (Field, 1997).

After the onset of the property crisis in GB, Tesco ventured into the French market through the acquisition of the already profitable Catteau chain in 1992, with the intention of using it as a springboard to further growth in mainland Europe (Palmer, 2005). Furthermore, Tesco continued its extension into Europe throughout the 1990s by extending into Hungary, Czech Republic and Slovakia. The Hungarian venture was

the first foray into the convenience grocery market for the retailer as the stores purchased were small format grocery stores, albeit in a different market with its own nuances compared to the convenience grocery market in GB. However, rather than this being a growth strategy of Tesco in its own right, the small grocery stores were intended to be used to gain knowledge of the market in mainland Europe before opening larger supermarkets and hypermarkets. Fast forwarding to the present day, Tesco has advanced further in Europe and currently runs operations in both Poland (launched in 2002) and in Turkey (established in 2003) along with a continued presence in the European markets previously discussed in this chapter.

As the 1990s progressed, further investment by major retailers in the North American market occurred. Sainsbury's acquired 50% of Giant Foods in 1994, a market leading retailer in the Washington DC/Baltimore area, further growing its portfolio outside the UK. Unfortunately for Sainsbury's its US operations would unravel in 1997, resulting in a difficult year for the retailer (Wrigley, 1997). Shaw's encountered fierce competition placing a squeeze on its operations resulting in a sharp drop in profits and in the Washington area Giant suffered greatly from a truck driver strike having a detrimental effect on sales in the area and a 50% profit loss for the retailer (Wrigley, 2000).This came at a time when Sainsbury's had a renewed commitment to preserving and growing its operations in the UK in which convenience retailing would be of crucial importance for the retailer.

The growing Asian grocery market presented an opportunity for expansion for the British retailers that Tesco took on strongly. During the late 1990s and early 2000s the multiple advanced across the continent into the retail markets of Thailand, South Korea, Taiwan and Malaysia in the late 1990s followed by Japan in 2002 and China in 2004 (Butler and Neville, 2013). These ventures proved largely successful for the retailer and they continue to trade in these markets, with the exception of Japan. Shackleton (1998) argues that much of this success came as a result of embedding trade in local markets and supply chains, not operating as a global retailer from afar. Tesco has proved tremendously successful in Thailand and South Korea. In the latter it is the second largest retailer despite taking recent problems due to changes in the laws surrounding Sunday trading hours in a bid to protect smaller stores, a parallel of the environment faced by the retailer domestically in the 1990s.

Following the success of its 'Express' convenience store format in the UK, discussed later in this chapter, Tesco made an ill-fated move into the US small format grocery market through the launch of a chain of typically 10,000 sq. ft. stores branded 'Fresh and Easy' and targeted at western states with high growth rates, opening the first stores in Los Angeles and Phoenix in November 2007 (Lowe and Wrigley, 2010). Amid

zero profits, the retailer pulled out of Fresh and Easy in the USA in 2013 citing costs of £1.2bn for a chain that had never turned a profit amid the first fall in annual profits for the retailer as a whole in almost 20 years. This brought to an end the most significant foray into the convenience market by the retailer in the US and signalled an end to the retailer's operations in North America. British grocery retailers thus had mixed success in their ventures into international markets but by mid-2000s, Tesco operated over 50% of its store network overseas producing $600m of profits and had succeeded in markets in which other huge retail multinationals such as Walmart and Carrefour had failed (Wrigley and Lowe, 1999).

Despite some of these ventures failing, Tesco and Sainsbury's in particular have attempted to grow their offer internationally, partly in response to the issues they were experiencing in growing through their supermarket formats domestically. A number of these ventures were outside the retailers comfort zones in acquiring and opening small format grocery chains, and whilst these were in markets with differing characteristics to the convenience grocery market GB, the retailers no doubt used the knowledge gained when strategizing their convenience grocery operations domestically. This could perhaps be argued to be another advantage gained over smaller retailers and independent grocers in developing their wealth of experience in testbed locations where costs are lower than the UK, notably in Eastern Europe.

2.3.2 Increased growth in non-foods

Feeling increasingly squeezed in the grocery market, the major grocery multiples increasingly saw potential for growth in diversifying into non-food products (Wrigley, 1991). This was acted upon through new commitments to a number of products including petrol, financial services, pharmaceuticals and clothing in the pursuit of greater sales and increased profits, with non-food products often accounting for as much as 40% of space in larger stores (Guy, 1996b).

The first example listed in the preceding paragraph was the petroleum market. Tesco, Sainsbury's, Morrisons and ASDA all have many stores with petrol stations on their parking lots. They are often able to sell petrol at a low cost with small profits being made as the use of the petrol station can attract people to the store and attract non-petrol spending from consumers. This has created an issue for oil companies due to the intense competition over price from hypermarket retailers (Cohen, 1998). Additionally, partnerships allowing convenience stores on existing petrol forecourt stores have proved lucrative for many retailers and will be discussed in more detail later in this chapter, and although this isn't necessarily a change away from non-food, it

does highlight the diversification of store format and changes away from (or alongside) traditional mode of trading for the major grocery multiples.

Many retailers including Sainsbury's, Tesco and Marks & Spencer advanced increasingly into the financial sector in the 1990s (Colgate and Alexander, 2002). The retailers diversified from food into the personal finance industry through partnerships with financial institutions in one of the following two ways; 1) Joint Ventures. 2) Tie-Ups. Joint ventures were brought about through the creation of an independent subsidiary with a share of equity owned by both companies to provide financial services either via telephone or in-store financial desks (Alexander and Pollard, 2000). This was the method through which both Tesco and Sainsbury's chose to pursue growth in the finance industry. Tie-ups, the second type of partnership, are strategic alliances between the grocery retailer and the financial institution to offer in-store space to the financial institutions and were adopted by both ASDA (in partnership with Lloyds) and Morrison's (in partnership with Midland HSBC) and resulted in the customer being serviced mainly by the financial institution whilst the retailers earned commission (Martinelli and Sparks, 2003).

 Food retailers have committed to understanding their customers' behaviours and responding to them quickly through innovation; this has been argued to be beneficial to them when moving into the financial services sector. Following its launch in 1994, Tesco extended its loyalty scheme into financial services with the clubcard plus in 1996 which essentially functioned as a high interest savings account (Pitcher, 1997). The strategic move into finance by many grocery multiples has proved positive and allowed the retailers to deepen relationships with consumers and helped to centralise consumer needs through offering an increased range of goods and services, adding an extra element of convenience to the consumer.

Retail Price Maintenance on over-the-counter drugs was abolished in 2001 bringing the free market to the pharmaceutical industry in which the grocery multiples saw a potentially lucrative opportunity. In a similar way to the price wars created by the abolishment of RPM on groceries in 1964, the major grocery retailers saw this as a major opportunity to diversify their offer to include pharmaceuticals. Pharmaceutical products became increasingly part of the on-shelf offer of the major retailers in the UK and Tesco have gone as far as opening Tesco pharmacies in many locations across the UK (Brondoni et al. 2013). Now a common channel for the purchasing of medicines, this brought increasing customers to grocery stores and once again allowed customers to centralise their purchases of a number of products into the grocery store.

Furthermore, the structure of the UK clothing industry was rapidly altered from the mid-1990s as a result of the entry of the major food retailers into the market. Asda, Tesco and Sainsbury's all established own-branded clothing lines to be sold alongside the traditional grocery goods in supermarkets up and down the UK (Pretious and Love, 2006). The enormous shift in the market is highlighted by the fact that Marks & Spencer was selling less than ASDA's own brand - George at ASDA - across all garment categories by 2006, within ten years of the major grocery retailer's entry into the market (Rider, 2004).

As this thesis is focused on the grocery industry in GB, this section has briefly described the non-grocery responses and strategy by the major grocery multiples. However, it is clear that the retailers have diversified beyond grocery retailing into other retail sectors. This has also had some advantages to their grocery businesses. The centralisation of a multitude of services to consumers gives them a competitive advantage over smaller retailers as it boosts the convenience offered to consumers and brings people back to the brand again and again.

2.3.3 Growth through e-commerce

Technological changes were rapidly occurring at a time in which the large multiples were looking to adapt and diversify from supermarket grocery retailing. Increasing internet usage in households created a new channel for grocery retailing that the major retailers were keen to embrace. Huge growth in internet usage and further improvements in internet technologies have maintained a continuously growing e-commerce market in the UK for a number of products, including groceries. The broadband revolution growing from 200,0000 users in 2002 to over 20 million in 2010 facilitated online shopping through unbridled internet access (without the need for a dial-up connection) and a faster infrastructure for the consumer (Youde, 2010). The grocery retailers attempted to counter the threat of market saturation and difficulty growing in traditional channels domestically through diversification into this emerging channel.

It was often surmised that consumers would be somewhat reluctant to purchase fresh food products without being able to see them (Hackney et al. 2006). However, the online grocery market in the UK began to take off in the mid-1990s and continued growing as the UK supermarkets overcame these obstacles as customer confidence in the supermarket brands increased and the convenience of internet retailing benefitted many consumers (Morganosky and Cude, 2000).

Tesco was the first of the major grocery retailers to enter the e-commerce market through an online wine delivery service in the mid-1990s. Following its entry into the online wine market, Tesco rolled out the first home shopping service in the UK in 1996, growing this channel and rebranding it 'Tesco.com' in 2000 (Thompson, 2013). The retailer committed to becoming the largest online grocery retailer in the world amid its continuing growth in the UK (becoming the first online grocery service to break even) in which it delivered to 90% of the population in 2002 using a model in which goods were picked and delivered by staff at supermarkets. The retailer highlighted its commitment to this distribution channel by developing technological advancements in semi-automated in-store picking for delivery reducing time and labour in the online retail process (Delaney-Klinger et al. 2003). Tesco now operate a number of dark stores from which they distribute goods purchased through their online channel. These stores are not open to the public and purely serve the retailer's online customers. Many of these stores form a ring around London and the retailer identified a desire to build a national network (Wood, 2012).

After Tesco entered the e-commerce channel through their wine delivery service, Sainsbury's followed suit by announcing their own wine delivery service in 1995. Sainsbury's committed to further expanding its e-commerce venture in 1997, experimenting with home delivery from 7 stores which would be extended to 30 stores in 1998. Furthermore, the retailer dedicated itself to the channel, *"…we intend that E-commerce will form an integral part of our offer to customers and we will also be looking at innovative ways of using all aspects of our business to create new value*" (Sainsbury plc, 2000). Sainsbury's became the first grocery retailer in the UK to build dedicated online distribution warehouses from which goods could be delivered to people's homes and used a combination of delivery from both distribution centres and stores to achieve a coverage of 71% of the population by 2002 (Ellis-Chadwick et al., 2007)

Asda were initially sceptical of online food and drink retailing and refused to enter the market of grocery home deliveries and instead began to diversify their offer to respond to changes in the market through increases in its non-food retailing in areas such as furniture and carpets (Owen, 2003). However, in 2000, after witnessing the successes of other retailers in this new distribution channel, the retailer began online deliveries from a limited number of locations serviced by 32 physical stores from which the goods were obtained. Despite a late entry, the retailer began committing to the channel by building dedicated online depots to adopt as their online customer base grew. At first

the retailer conducted its operations through picking of goods from existing stores but continued to advance its online offer into the late 2000s by developing online specific warehouses akin to those operated by Sainsbury's to better serve areas in which its online presence was strongest (Ellis-Chadwick et al., 2007). An even later entry into the online market was Morrisons. The retailer only developed a website (purely to display company information) in 2003 and only fully embraced online retailing in 2013, over ten years after both Tesco and Sainsbury's had successful established a comprehensive online presence.

Despite advances in internet use and uptake of online grocery retailing by the consumer, the online grocery retail market still only accounted for around 5% of total food and drink sales in 2015 (IGD, 2015). However, online has become a valued distribution channel for retailers and they have recognised the need to manage their online operations alongside physical stores to provide the best offer for consumers to stay at the forefront of the industry (Birkin et al. 2002). Furthermore, sales in this grocery channel are expected to grow to over £17bn by 2020 and online grocery retailing is predicted to account for around 8.5% of sales by the end of the decade, continuing its trend as one of the main growth areas in the grocery market in the UK (IGD, 2015).

The establishment and maintenance of online grocery retailing is a sign of the major retailers commitment to multi-channel retailing, allowing consumers to purchase goods from the retailer both from home and in visiting a physical store. Although retailers often begrudge the lower profit margins and logistical difficulties associated with online retailing, the move into the market is essential in guarding against other retailers (new of existing) moving into and securing market share in online groceries.

2.3.4 Store format changes (including investment in the convenience grocery market)

In response to the difficulties in increasing market share through growth in large supermarket retailing synonymous with major retailers, they diversified their offer both in grocery store formats and the type of goods that they traded in. As discussed in this chapter, retailers diversified into new products such as petrol and financial services, invested in new markets internationally such as in mainland Europe and Asia and kept up to date with technological changes through embracing e-commerce retailing domestically. However, the large retailers also began to diversify their offer in the physical store market in GB, making significant investments and adopting new

strategies for growth. This section looks at the change in store formats operated by the major retailers, focusing on growth into the convenience grocery store market.

Through the combination of PPG6 and the Competition Commission's two market ruling policy influences, out-of-town development has been increasingly restricted. The choice to diversify into smaller format stores became attractive to major grocery retailers who have increasingly considered diversifying from out-of-town supermarket developments into other types of locations. These include town/city centre developments, residential neighbourhood stores and train stations. By 2003, 40% of retail development was in town centres (Cheshire et al. 2011). Guy (2011) suggests that Tesco and Sainsbury's in particular opened smaller, convenience store formats to exploit the 'basket' shopping market along with circumnavigating planning policy restrictions.

Following a strategic review in 1997, the Co-op decided first that it would turn its attention to smaller store formats. The Co-operatives *"… inability to compete nationally eventually led to it having to retreat from the large superstore format so favoured by its rivals"* (Hallsworth and Bell, 1998, P. 301). Through this redirection, the Co-op became the first large retailer to become a major player in the convenience grocery market and the introduction of large retailers to this market was initiated.

Tesco announced a joint partnership with ESSO Petroleum in 1997 in which the retailer would operate its new convenience store format 'Tesco Express' out of a number of petrol station forecourts across the UK (Hughes *et al.,* 2003). This proved a success and the retailer forwarded its move into the convenience market by rolling out this format through the opening of convenience stores alongside forecourt petrol trading.

The arrival of the major grocery retailers into the convenience grocery market had begun and a period of change in the grocery market was initiated. Furthermore, Sainsbury's piloted its first convenience store format branded 'Sainsbury's Local' in Hammersmith in 1998 (J Sainsbury PLC, 1998). Another of the four major grocery retailers had entered into the convenience store market through the opening of its first convenience grocery store. Hallsworth and Bell (2003) acknowledge this as the time at which the market leaders began a return to smaller format stores located close to suburban residential areas or in town/city centres using new convenience store formats.

The introduction of the major retailers into the smaller format neighbourhood grocery market in GB *"… exposes smaller neighbourhood retailers to competition along with*

complex, efficient supply chains and a strong tradition in location management" (Wood and Browne, 2007, P. 234). This has created an interesting situation with a complex set of dynamics that will be investigated in this research as major retailers have the advantage of location planning teams whereas independents and smaller retailers do not. Additionally, Marks and Spencer Simply Food opened its first convenience store format in Twickenham in 2001. M&S is a very large retailer in GB although more associated with fashion retailing than grocery retailing. However, many of their larger fashion centred stores do contain supermarkets. They planned to progress their 3,000 sq. ft. offer to complement a portfolio of small to medium-sized in house supermarket stores and continued to grow in this format size, often in rail and motorway service stations (Grocer Online, 2001).

Furthermore, Morrisons also committed to developing their offer in the small food-retail-based convenience store sector in 2011. This was highlighted in a statement by Dalton Philips the Chief Executive of Morrisons in which he announced: *"Convenience is one of the fastest growing sectors of the market and developing our offer in this channel is a key part of our growth strategy."* (Philips, 2011). This commitment from another major retailer was likely to increase the competition in a market already consisting of multiple retailers alongside large retailers. However, the retailer withdrew from the convenience grocery market in 2015 following the sale of its convenience outlets, a transaction discussed in more detail later in this chapter. The most recent larger grocery retailer to enter the convenience market is Waitrose. The retailer opened its first convenience grocery store (branded Little Waitrose) in South Kensington in 2011 before beginning to roll out the format in London and further afield (Whiteaker, 2011). Morrisons later re-entered the convenience market in later 2015 with opening of a 1200 sq. ft. Morrisons Daily convenience store at a Motor Fuel Group petrol station in Crewe (Ruddick, 2015).

Asda have followed the trend of involvement by major grocery retailers in smaller format stores, albeit much later than both Tesco and Sainsbury's. The retailer purchased Netto in 2010 at a cost of £778 million (Finch and Wood, 2010). Despite this commitment to smaller stores, the stores acquired were larger than the convenience stores operated by other major retailers, averaging around 8,000 sq. ft. and therefore cannot be opened for longer hours under Sunday trading laws and are not strictly defined as convenience stores as part of this research. ASDA were slower to uptake both online and small format grocery retailing than both Sainsbury's and Tesco. However, the retailer led the way in terms of committing to an alternative strategy of growth, a comprehensive store expansion programme (Wood and McCarthy, 2013).

Rather than investing in the convenience market, ASDA chose to diversify their offer by expanding existing store sites to increase their operating space in GB. This strategy was known as 'space sweating' as it expanded on existing space rather than developing new space (Wood and McCarthy, 2013). Despite ASDA leading the way on this, other retailers also adopted this strategy to an extent with Tesco expanding some of their supermarket stores to become hypermarkets and rebranding them with their Tesco Extra fascia, the retailers largest store format fascia. This strategy goes hand in hand with the diversity into non-grocery products due to the fact that as much as 50% of the floorspace in ASDA's largest hypermarket stores is now dedicated to higher margin goods such as clothing and electrical goods. Sainsbury's and Tesco also diversified into other non-supermarket formats with the development of city centre stores that were larger than convenience stores. Tesco developed the Tesco Metro fascia to give an individual brand image to this type of store.

The dynamics of the convenience grocery market in the UK has been heavily influenced by mergers and acquisitions by major retailers aiming at growing their market share in this sector. Alongside organic growth in standalone store developments and the conversion of units previously used for other purposes (both retail and non-retail), Tesco and Sainsbury's in particular of looked to grow their convenience operations through acquisitions of (and mergers with) smaller convenience retail chains. This section details the changes in the convenience retail market brought about by mergers and acquisitions involving the major grocery multiples in GB.

In 2002, the Co-operative Group became the largest convenience grocery retailer in the UK following the acquisition of the Alldays brand. This increased the number of stores owned by the Co-operative Group by approximately 600 units to around 2200 stores in total, pushing the retailer to the forefront of the convenience retail market. Furthermore, the Co-operative Group continued its convenience market growth in 2003 through the acquisition of Balfour, a convenience chain with 121 stores in the UK grocery market.

Similarly, Tesco increased its portfolio of convenience grocery store units with the acquisition of a large number of units from T&S Stores in 2003. This acquisition comprised 862 stores previously trading under the One Stop, Dillons and Day and Nite brand names and significantly boosted the number of convenience stores owned by Tesco in the UK. Prior to this, Tesco only operated approximately 130 convenience stores in the UK operating under the Tesco Express fascia, primarily as part of ESSO

petrol forecourts (Wood *et al.,* 2006). The majority of the One Stop stores continue to trade under this original name. Furthermore, the London based convenience store chains Europa, Harts and Cullens stores were purchased by Tesco from their parent company Adminstore in January 2002. These acquisitions catapulted Tesco to the forefront of convenience retailing in Great Britain and was the first significant acquisition of a convenience store chain by a major grocery multiple in GB.

In response to Tesco's growth in the convenience store market, Sainsbury's launched a series of acquisitions in 2004 during which the company purchased a total of 174 units from other convenience retailers. This comprised of fifty-four units from Bells stores in the North-East of England, 114 units from Jacksons stores in Yorkshire and the Midlands and finally six units acquired from JB Beaumont stores in the East Midlands in November 2004. Despite initially retaining the names of these retailers in store fascia branding, Sainsbury's subsequently rebranded the units with the Sainsbury's Local name used on their non-acquired convenience stores. Additionally, the retailer acquired five convenience store units in the South East previously trading under the SL Shaw name in early 2005, further increasing the number of convenience stores owned by Sainsbury's.

At this point, the competition for market share in the convenience sector had intensified amongst the major retailers, coinciding again with questions of fairness being raised with the Competition Commission. Whilst changes in planning policy preventing major retailers growth in to the convenience market in GB did not materialise, it may have been this pressure that contributed to Sainsbury's and in particular Tesco maintaining the original fascias of acquired convenience stores in order to reduce the visibility of their growth in the market. Further acquisitions by major retailers have occurred in subsequent years although not to the same extent as in the period of 2002-2005. In late 2010, Tesco's One Stop brand announced the purchase of the Mills chain of 76 convenience stores operating in the Midlands, South Wales and the North East of England, further increasing Tesco's stock of One Stop convenience store units in England and Wales to a total of 598 convenience stores.

In 2012, the Co-operative Group acquired Somerfield (880 stores) expanding its offer in the small to medium grocery store offer (Finch, 2008). Earlier in the late 1990s the retailer had committed to advancing in the convenience market in light of pressure from larger retailers in the superstore market.. However, the acquisition of Somerfield could be described as a move away from the retailers earlier commitment to the small-store convenience market in the GB. This is evidence of the retailer being squeezed on

both the smallest and largest stores by the largest grocery retailers in the market and evidence that a spatial battle for dominance is taking place across the convenience grocery sector in Great Britain.

The case of the Co-operative Group as discussed in the literature review highlights the pressures placed on smaller retailers. In the late 1990s, the retailer committed to advancing in the convenience market in light of pressure from larger retailers in the superstore market.. However, in 2012, the retailer acquired 880 stores from Somerfield expanding its offer in the small to medium grocery store offer (Finch, 2008). It could be argued that this signalled the retailer moving away from its earlier commitment to the small-store convenience market in the Great Britain. This is evidence of the retailer being squeezed on both the smallest and largest stores by the largest grocery retailers in the market and evidence that a spatial battle for dominance is taking place across the convenience grocery sector in Great Britain. However in 2012, the Co-operative Group acquired 10 London based convenience stores from the Costcutter brand, increasing its share of the South East market. Additionally, the retailer acquired a further 28 stores from the David Sands retail chain based in Fife, Kinross and Perthshire, increasing the presence of Co-op owned convenience stores in Scotland showing that the retailer still had an interest in the convenience grocery market.

Despite being a late entrant into the convenience store market, it was announced in June 2012 that Morrisons were in talks with Costcutter with the intention of looking to purchase a large number of stores belonging to Nisa when a previous contract between Costcutter and Nisa expired (Leyland *et al.*, 2012). This deal did not materialise but it confirmed the growing interest in the convenience market by Morrisons. Moreover, early 2013 acquisitions by Morrisons of stores previously operated by Blockbuster, Jessops and HMV confirm the retailer's intentions to advance their convenience grocery offer by converting these units into small grocery stores (Neville, 2013). This is an interesting case in respect of the dynamics of the overall retail market in the UK. Whilst many retailers are struggling and indeed closing, the convenience grocery market appears buoyant and in this case Morrisons have capitalised on other retailers' demise. Some of these stores opened up as M Local, the Morrisons convenience fascia but many remained unconverted to convenience retailing and were sold on.

Morrisons had continued to grow their convenience operations through both acquisitions and organic growth resulting in them operating 140 stores by 2015. However, in 2015 the retail sold its convenience store network to private investors

effectively starting a new smaller retailer (Sparks, 2016). This was a reverse of the majority of acquisitions seen in the convenience grocery market in GB in which large retailers have generally purchased smaller retail chains. The stores remained very similar, reopening under the name My Local and stocking products acquired through the same supply chain as the symbol group retailer Nisa.

Additional partnerships between petroleum companies and supermarket retailers have also provided sites for convenience retailing in the UK market as detailed earlier in this chapter. In 1994, Tesco undertook its first venture into convenience store retailing through a joint venture with ESSO Petroleum to open a Tesco Express store on each forecourt (Wood *et al.,* 2006). Over a decade later, Tesco continued its ventures in forecourt retailing by acquiring 25 petrol forecourts from Morrisons in September 2005 to be branded Tesco Express convenience stores (Wood and Browne, 2006).

By the end of 1999, another large grocery retailer was forming an extensive portfolio of units at petrol stations. Safeway operated 45 forecourt stores in conjunction with British Petroleum Ameco in the UK (Baron *et al.,* 2001, P. 399). Furthermore, the major advancement by Sainsbury's into the convenience store market was accompanied by a partnership with Shell petrol stations; this alliance resulted in the retailers opening 100 convenience stores on petrol forecourts across the UK (Guardian Press Association, 2003). This partnership subsequently ended and Sainsbury's no longer operates convenience stores on Shell forecourts. The large number of convenience stores operated on forecourts has become a major part of major grocery retailers move into convenience retailing and must therefore be considered when investigating the dynamics of this sector of the UK retail industry.

In 2005, BP and M&S Food announced a joint venture in forecourt retailing. This proved successful and, in 2008, the 100th store operating under this partnership opened (Forecourt Trader, 2008).Waitrose followed other retailers into this market by announcing a partnership with Shell in 2011 to open two trial 1600 sq. ft. stores on petrol forecourts. However, following pressure from campaign group Greenpeace over Shell's drilling in the Arctic, Waitrose decided against further expanding the partnership into more petrol stations (Smithers, 2012).

2.4 A growing demand for convenience groceries

The first aim of this research is to review the existing academic and industry literature on the convenience grocery market in Great Britain in relation to; the growth of major retailers into the convenience grocery market, the growing demand for convenience groceries and the attempts at forecasting revenue of both convenience grocery stores and grocery stores more generally. This section looks to review the literature on the growing demand for convenience groceries in the population of Great Britain.

In responding to changes in local planning legislation and shifts in the grocery market, it could be argued that the major retailers created a demand for branded convenience groceries in GB. However, it is unlikely that they would have chosen to diversify their networks through growth into small format retailing had they not anticipated an existing and potential growth in demand for this type of grocery store. This section looks at a number of aspects of the demand for convenience retailing. This includes demand generated through; the types of consumer prone to convenience shopping behaviours, population change, demographic change, shifts in living arrangements, shifts in working patterns, behaviour-driven grocery shopping trips and demand generated by the retailers themselves. These trends are linked to the types of location in which branded convenience grocery stores are found in GB, ranging from prime city centre pitches to branded convenience stores serving rural villages.

In doing so, section 2.4.1 looks at the types of consumers known to be receptive to convenience grocery retailing. Thereafter, section 2.4.2 reports on the literature surrounding population and demographic changes which have increased the appetite for branded convenience grocery retailing in GB. Next, section 2.4.3 discusses the changes in living and working patterns of the population and the link to a demand for branded convenience grocery stores, whilst section 2.4.4 identifies trip behaviour that has a link to the growth in branded convenience grocery stores. Finally, section 2.4.5 looks at the suggestion that demand for convenience grocery retailing has been created by the branded convenience grocery stores (or retailers) themselves.

2.4.1 Types of consumer

Small format convenience grocery stores can provide the customer with a service through *"…enabling the consumer to increase the number of tasks that can be accomplished during a single visit to the retailer, or reduce the amount of time required to complete the shopping task"* (Morganosky & Cude, 2000, P. 17). Burt et al., (2010)

highlight the importance of consumers in dictating the actions of retailers. They highlight *that "leading retailers now talk of a demand rather than supply channel, which reflects the emphasis on responding to customers and providing innovation"* (Burt et al., 2010, P. 189). When looking at the grocery market in GB, convenience in food is more than just the time spent cooking, but *"… includes the effort required in purchasing, storing, preparing and consuming food."* (Buckley et al., 2007, P. 600). This includes attending a store and purchasing the groceries required.

A number of studies have identified the demand for locational convenience being a prime motivation behind consumer's grocery purchasing behaviours (Morschett et al. 1995; Lindquist, 1975; Barich and Srinivasan, 1993). In a study conducted on Portsmouth residents from 1980-2002, Clarke et al., (2006) found that food shopping behaviour over time has become increasingly dependent on convenience and geographic proximity of retail units. In a study of residents in Argyll, Bute and Edinburgh, McEachern and Warnaby (2006) found that 27% of respondents travelled less than one mile in shopping for food. Furthermore, 87% of these respondents resided in urban areas. It has been argued that major retailers extending 'convenience' through smaller format stores (with longer opening hours and Sunday trading) has reduced the competitive advantage that independent stores have in catering for consumer needs and that *"the erosion of the competitive advantage of convenience may be nearing completion by the start of the 21st century."*(Baron et al., 2001, P. 396).

Buckley et al., (2007) conducted a survey of 1004 respondents in which they asked 80 questions on aspects of food convenience. The research was conducted on respondents from 79 UK locations and administered evenly across the spectrum of ACORN area classifications. Following a factor analysis of participant responses, the study identified four distinct categories in which respondents could be placed: 26% of respondents were categorised as 'food connoisseurs' that were unlikely to select convenience food to make life easier; a further 25% of respondents were termed 'home meal preparers'; the least likely group to seek out convenience food. Of the remaining participants, 16% were categories as 'kitchen evaders' and 33% were termed 'convenience seeking grazers'. The latter two categories include those people that are likely to place a high emphasis on convenience in the shopping process and are least likely to plan trips to buy food. The research found that 49% of consumers are likely to prioritise convenience in purchasing, preparing and consuming food and groceries. This supports the presence of a demand for convenience grocery retailing among consumers.

Jackson et al., (2006) conducted in depth qualitative research into consumers' shopping practices at the household level in Portsmouth and identified that "..*consumer choice is socially embedded within households increasingly complex everyday lives with shopping 'fitted in' around other peoples' responsibilities and commitments (childcare, work, leisure)"* (Jackson et al., 2006, P. 59). The research found convenience to play a key role in shopping decisions; however, convenience often means a variety of different things to different consumers that had to be unpacked. Mendes and Themido (2004) believe that a *"…change in the consumer's behaviour and the fact that consumers are more demanding force the retail groups to invest strongly at stores of smaller dimension."* (Mendes and Themido, 2004, P. 2).

Lal and Rao (1997) found that grocery shoppers can be disaggregated into two distinct categories; 1) Time-constraint shoppers, and 2) Cherry Pickers. They identified time-constraint shoppers as being attracted by convenience in the shopping process. They argue that this can be achieved in one of two methods: by combining grocery shopping through linking groceries with other activities or by purchasing more goods at once in a large shop. Those customers combining multiple activities were found to have a willingness to pay higher prices and use more centrally based stores which often inherently come with increased costs to the consumer due to higher rents for the retailer (Popkowski Leszczyc et al. 2004). Iyer (1998) supports the existence of a large proportion of consumers who have less fixed shopping costs and are often eager to save time in the shopping process.

In a study involving interviews with 560 consumers in Germany, Morschett et al. (2005) found that grocery shoppers could be divided into four distinct types. One category was identified as time-constrained consumers placing increased emphasis on convenience in the shopping process through reducing the need for dedicated grocery shopping trips. In the study these customers preferred nearby store locations reducing the time spent shopping and were willing to sacrifice price in the process.

These studies show that a larger proportion of consumers value convenience in the grocery shopping process and that branded convenience grocery stores can meet this need. This is likely to have contributed to major grocery retailers investing in the convenience grocery market from the mid-1990s and sustained the commitment to this type of retailing to the present day.

There is a body of literature from the 1990s on groups of consumers that are in some way constrained in their shopping behaviour. These studies highlight the general trend

of limited accessibility to stores, primarily due to limited mobility, such as the elderly (Bone,1991; Hare et al. 1999), low income (Stitt et al. 1999), the disabled (Kaufman, 1995) and ethnic minorities (Hill and Somin, 1996). Societal trends such as increased car ownership have contributed to changes in grocery supply, particularly the increase in out-of-town/off centre retail developments at the expense of central stores (Smith and Sparks, 1997). This has often disproportionately affected deprived areas in which the population has limited mobility. These types of consumers benefit from easily accessible neighbourhood convenience stores close to residential areas. These type of stores can be the sole grocery option in rural and semi-rural areas and generate substantial trade. The Co-operative group is a major player in the convenience market in GB and is known to locate in this type of rural area.

Piacentini et al. (1999) investigated the access to grocery provision of deprived consumers in Scotland and found that the majority of these consumers could be categorised as 'economic shoppers'. They found these shoppers to be heavily dependent on local convenience stores due to financial and mobility limitations that they experience. They highlight previous research that has reported convenience shoppers to generally be associated with consumers trying to save time (being willing to spend extra money in doing so) (Williams et al. 1999), but also arguing that disadvantaged consumers often seek convenience for the opposite reason, for lack of choice. The research also emphasises the fact that convenience retailing is often more expensive for the consumer, further disadvantaging the deprived consumer.

2.4.2 Population and demographic change

Population size and location has a significant impact on available expenditure and the ability of an area (or location) to support profitable grocery stores. Belief (or not) in imminent market saturation played a role in retailer operations in the grocery market in GB in the 1980s and 1990s. One counter argument to the belief in saturation was the year-on-year growth of the UK population, boosting available expenditure, amplifying the available grocery demand particularly in areas of high population growth (Birkin et al., 2002). Figure 2.1 identifies the population at each of the last four UK censuses of the population.

Figure 2.1 Population Growth in the UK, 1981-2011. (Source: ONS, 2011)
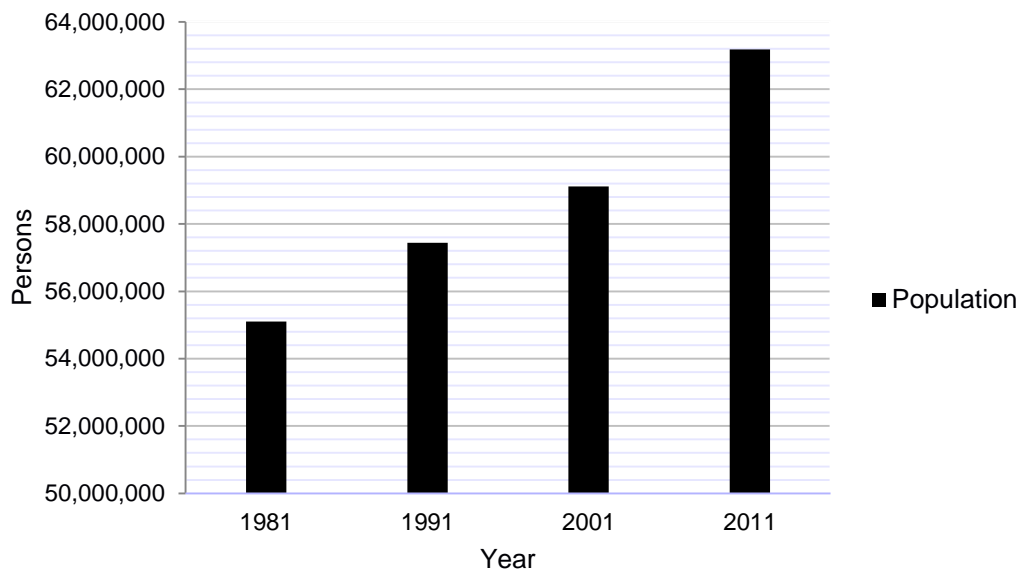
We can see that the population has grown from around 55 million in 1981 to a significantly larger 63 million in 2011 with the rate of growth increasing between 2001 and 2011 when compared to the period between 1991 and 2011 (ONS, 2012). Moreover, it is expected that the population of the UK will continue to rise and to reach 71 million people by 2031 (ONS, 2012). The growth in population has increased the overall available expenditure on groceries and in these areas of growth has increased the viability of additional stores. As convenience stores became a more widely used strategy of major retailers, large and growing population areas have been acknowledged as ripe for convenience grocery stores and the major retailers began opening convenience stores in this type of location.

Demographics can also be seen to play a key role in the shopping behaviour of consumers. The major demographic trend in the developed world in the 21st century is population ageing. The proportion of the population aged over 65 has grown from 13% in 1971 to 19% in 2010 and is predicted to rise to 23% of the population by 2035 (IGD, 2012). The major retailers have become linked to the ageing population trend through the debate around the extent to which large grocery multinationals contribute and have an influence on the quality of life of consumers.

During the golden era of superstore retailing the major retailers adopted a strategy of 'spatial switching of capital', as smaller town and city centre stores were closed and large 'cathedrals of consumption' appeared regularly on the edges of UK towns and cities (Wrigley 1987, 1994). This trend, along with the closing of many independent stores and outlets operated by smaller, less successful retailers had a disproportionate

effect on different groups of consumers in the UK. It has been argued that the elderly, often of limited mobility, that have suffered the most considerably by the closing of neighbourhood grocery stores in favour of larger out-of-town warehouse style supermarkets and hypermarkets (Wilson et al. 2004). Jarvis et al. (1996) highlight the diminishing ability of a person to do the household shopping with time as their age increases. Nearby branded convenience stores are often the only easily accessible means of purchasing fresh grocery produce for elderly consumers and this type of consumer contributes to the demand for branded convenience grocery retailing.

Leighton and Seaman (1997) found that store location was of primary importance to elderly grocery consumers and found that location of out-of-town stores often disadvantaged older customers disproportionately. They hypothesised that local investment in shopping may be the best way to meet elderly shopper's needs in order to provide better value and more accessible groceries. It could be argued that through the major retailer's increased commitment to small-formats, more local shopping channels have presented an avenue to tap into an existing demand, driven by elderly consumers reliant on convenience stores. The Association of Convenience stores found that 27% of over 65s visited their local stores once a day, a significant market available to branded convenience grocery retailers in residential locations.

2.4.3 Changing living and working patterns

Kinsey and Senauer (1996) identify what they describe as a 'sea change' in demographic lifestyles in the USA. The examples of change they highlight are applicable across western democracies and parallel much of the experience of communities in other western cultures. They highlight an increase in ethnic diversity, an ageing population, widening income disparity and an increase of women in the labour force. As a result of these sea changes they identify two distinct types of consumers, those that are economists (or price conscious) consumers and those that are convenience oriented. The rise in number of the latter can be attributed to an increase in dual income households and the growing proportion of professional occupations alongside an increased desired to spend limited free time with family and at home in general (Crossen and Graham, 1996).

The role of single-person households in the shift to a convenience culture has been articulated. De Kervenoael et al., (2006) highlight the increase in single-person households between the 1991 and 2001 Censuses of Population, the latter identifying

30% of households as containing only one person. Research conducted by Hallsworth et al. (2010) has found that, by 2002, 78% of meals consumed in the UK had only one or two people present. They found evidence of an increase in shopping frequencies amongst consumers in the UK retail market. They found that from 1980 to 2002 the percentage of persons food shopping three times a week increased from 9% to 21%. Young professionals who are cash rich but time poor (and may choose to eat ready prepared food to free up time for other activities) are a prime target consumer for convenience grocery retailing. Moreover, these consumers are more likely to live in central towns and cities and close to centre suburban locations located in easy reach of goods and services. The increased cost of goods in many branded convenience stores does not deter these consumers and they are therefore important generators of demand for this type of grocery store.

Gofton (1995) identified the impact of changing household composition on grocery and food retailing. He suggests that an increase in dual-income households in which parents select convenience foods over home cooked meals has resulted in an increase in consumption of take-away and fast foods. This assertion is supported by De Kervenoael et al., (2006) who identify a shift from the immediate post-war period in which families survived on a single social wage (predominantly male) to the emergence of modern, dual-income households in which time is a precious commodity. In this context, it has been acknowledged that there are signs that smaller shops are rising in importance in the daily lives of consumers (Guy, 2009).

In addition to the increased growth of the population in the UK, we have seen changes to the way people are geographically distributed by residence. Between 1981 and 1991, UK cities were losing population whilst in the throes of 'counter-urbanisation'. However, Champion (2014) identifies the growth then witnessed in UK cities in the interim periods (between the next two censuses of the population, in 2001 and 2011). The city-based population of the UK grew by 500,000 between 1991 and 2001 and by a further 2,400,000 persons between 2001 and 2011, leaving a total growth of almost 2.9 million over the 20 year period.

High density urban areas are prime locations for convenience stores for a number of reasons, both in relation to the available demand in central urban areas and the geography of these locations. In demand terms, these include a large available population in a relatively small area with a significant expenditure available to be tapped into, a large proportion of people without cars due to the increased likelihood of

good quality public transport or because they live close to work and therefore potential customers that are likely to seek close, convenient options for grocery shopping and a population often engaged in fast paced, busy lives in which a large value is placed on reduced time in the shopping process.

2.4.4 Behaviour driven convenience grocery shopping trips

Early retail studies (around the time of the emergence of the power of grocery multiples) found a demand for multi-purpose trips often involved in the purchasing of groceries (Hanson, 1980; O'Kelley, 1981). The emergence of shopping centres providing many retail stores across a number of categories offered greater convenience to consumers (Arentze et al. 2005). Some of these shopping centres offered grocery retailing in the form of large supermarkets (enclosed or nearby) allowing customers to purchase multiple types of products in a single trip. The major retailers were keen to open stores in this type of retail pitch as consumers were attracted in large numbers. In this way it could be argued that the major retailers offered convenience in the shopping process well before what is now known as a branded convenience grocery store was operated by this types of retailer.

Messinger and Narasimham (1997) argued that the demand for this type of one stop shopping was a major driving force behind the scaling in operation of major grocery multiples to become dominant in the market. The demand for what are now referred to as convenience grocery stores (those under 3,000 sq. ft. in size in GB) has also been driven by trip choice behaviours. Popkowski Leszczyc et al. (2004) identify a desire for consumers to optimise their time spent shopping. This can come in the form of eradicating a single trip for groceries into smaller visits fitting conveniently around daily movement patterns. Branded convenience grocery stores are often well equipped to satisfy this desire by being located in convenient locations. Timmermans and Louviere (1997) found that locational convenience was more important in grocery retailing than in the purchasing of other goods including fashion and electricals (for Dutch consumers).

The types of location in which convenience grocery stores are found can help the consumer limit the time spent shopping (or perceived time spent shopping) by offering ad-hoc grocery opportunities fitting around the everyday lives and movements of consumers. Retailers have attempted to provide this convenience through opening

convenience stores in a number of locations including: train stations, close to bus stops, near large workplaces and close to universities.

2.4.5 Retailer generated demand

It is possible that branded convenience grocery stores generated demand for themselves by their presence in locations previous without this type of grocery store. Wrigley et al. (2007) conducted interviews with 200 consumers in Hampshire identifying the impact of the conversion of three One Stop grocery stores to become Tesco Express stores with an increased range of fresh produce and the branded fascia of a very well-known major grocery retailer. They found that as a result of the change there was almost a 5% increase in consumers using these localised stores as their main shopping destination. Moreover, they found an almost 15% increase in consumers using the Tesco Express store for their secondary grocery purchases to supplement their main shopping trip. Consumers identified the better chilled food facilities with the increase in fresh fruit, vegetables, meat and fish as the driving force behind increased patronage. This is an example of retailers generating their own demand for convenience grocery stores by providing a better offer to consumers. In this study, there was also a rise in customers using other businesses within the area and a reduction in travelling outside the local area for grocery goods. This is a sign that this type of store may benefit the local area and create demand for branded convenience grocery stores simultaneously.

Wrigley and Dolega (2011) surveyed 267 high street/town centre locations in the UK looking at their performance both prior to and during the economic recession, from 2006 through 2007 to Q4 2008 through 2009. They found that even during the economic crisis, the strong growth identified in small food-retail-based convenience stores identified by the UK Competition Commission in the early-to-mid 2000s had continued, with convenience-retail store units increasing by 5.6% on average between the pre-crisis period (2006-2007) to within crisis period (2008/2009). (Wrigley and Dolega, 2011). This has invariance to the overall findings of the study which found that vacancy rates of units in the town centres/high streets had increased by 2.7% points between the two surveys.

On this evidence, branded convenience grocery retailing appears to have been more durable to economic change on the high street than other types of retailing. Clarke et al. (2012) found that a larger number of stores in neighbourhoods and a greater variety

maximise consumer choice and welfare by positively influencing customers'
satisfaction with their local mix of grocery stores. In this sense consumers can be
receptive to branded convenience grocery retailing to improve the diversity and
vibrancy of local high streets and shopping centres, areas in which this type of store
are often found in GB.

2.5 Summary

Chapter 1 introduced this research and set out the aims and objectives of this thesis.
The research presented in this thesis is based on the entry and subsequent rise in
prominence of major grocery multiples into the convenience grocery market. The first
aim of this research is to gain an understanding of the context by which branded
convenience grocery stores operated by major grocery multiples came to form a
prominent part of the retail landscape in GB. This chapter has reviewed the academic
literature and wider body of evidence of the changes seen in the grocery market in
Great Britain between the 1960s and the present day, identifying the major trends in
grocery retailing, the strategy of major retailers in the market and the growing demand
for branded convenience groceries among consumers in GB.

Section 2.1 looked at the supply side shifts in grocery retailing in GB between the
1960s and 2010s, identifying the players in the market and the types of grocery store
formats operated by different types of retailer. Moreover, section 2.2 looked at the
changes to the grocery market in GB from 1960 to 2016 that have resulted in a market
in which branded convenience grocery retailing is commonplace. The changes began
in the early 1960s when resale price maintenance on groceries was abolished meaning
retailers could now legally compete on price. This benefitted the major grocery
multiples and they grew their operations rapidly. This era of dominance for a select few
grocery retailers was known as the 'golden age' and the major retailers grew their store
networks, particularly through the opening of large off centre supermarkets, during this
time.

This era began to come to an end for a number of reasons including financial
problems, a belief in market saturation, the arrival of the deep discounters and most
importantly in the context of the rise of branded convenience grocery retailing, new
local planning policy restrictions. Changes in the grocery market, particularly in the
form of new local planning legislation, made growth in market share through traditional
large supermarket formats more difficult and major retailers responded in different

ways. One such response (by two retailers in particular, Tesco and Sainsbury's) was extensive growth in the convenience grocery store market. The two retailer's focus on the convenience market has resulted in a situation in which branded convenience grocery stores are now very common in many areas of GB.

In responding to changes in local planning legislation and shifts in the grocery market, it could be argued that the major retailers created a spiralling demand for branded convenience groceries in GB. However, it is unlikely that they would have chosen to diversify their networks through growth into small format retailing had they not anticipated an existing and potential growing of demand for this type of grocery store. A number of drivers of demand for branded convenience grocery retailing were identified in section 2.4. These included demand generated through more consumer becoming prone to convenience shopping behaviours, population change, demographic change, shifts in living arrangements, shifts in working patterns, behaviour driven grocery shopping trips and demand generated by the retailers themselves. The review of literature found that there were many types of consumer that are 'natural' customers of branded convenience grocery retailing and they are able to support the various location types in which branded convenience grocery stores are found.

The third major aim of this thesis is to develop and test a series of predictive models for forecasting convenience grocery store revenue in the varying location types in which this type of grocery store is found. Chapter 3 begins this process by identifying methodologies used by major retailers to forecast grocery store sales, reporting on the various procedures theorised and developed by academics and retail consultancies and commenting on the potential of different methods to forecast convenience grocery store revenue.

Chapter 3
Models and methods of retail site location

Chapter 1 introduced the research reported in this thesis. Three major aims were identified in order to interrogate the convenience grocery market in GB. The first aim was to review the existing academic and industry literature on the convenience grocery market in Great Britain in relation to; the growth of major retailers into the convenience grocery market, the growing demand for convenience groceries, and the attempts at forecasting revenue of both convenience grocery stores and grocery stores more generally both in academic and in the retail industry. The review of literature in chapter 2 reported on the growth of major retailers into the convenience grocery market and the growing demand for convenience groceries among the population of GB. This chapter reports the review of the academic and industry literature on attempts at forecasting revenue of both convenience grocery stores and grocery stores more generally both in academic and in the retail industry.

The third and central aim of this thesis is to develop and test a series of predictive models for forecasting convenience grocery store revenue in the varying location types in which this type of grocery store is found. The review of literature in this chapter reports on the methods identified for forecasting grocery store sales both in academic literature and industry knowledge and sets up the predictive models in chapters 7, 8 and 9. This focuses on those methods that fit two criteria. They must have a theoretically plausible application to forecasting convenience grocery stores and be able to be incorporated into the suite of methods used by the location planning teams of major grocery retailers with little upheaval. Part of this is assessing the extent to which models traditionally applied to the forecasting of large supermarket could be used to forecast revenue of convenience grocery stores.

Location planning teams in major grocery retailers operate at two levels; the strategic level and the operational level (Reynolds and Wood, 2010). At the strategic level they evaluate sites and produce revenue predictions, thus informing investment decisions. Moreover, at the operation level they assist other teams in the business to understand their customer base and monitor store performance, often updating revenue estimates in light of shifts in the market (Newing, 2013). A 2010 survey of location planning teams in major grocery retailers found that the primary role of these teams was to support the

financial business case for new stores and search for potential new sites for a retailer to locate (Reynolds and Wood, 2010). The importance of being able to accurately predict the revenue of a store cannot be understated when it comes to supporting the financial business case for the store. The accuracy of a store forecast is integral to judging the financial viability of a potential store location and comparing potential sites for development.

This chapter reviews the different potential approaches to forecasting convenience grocery stores as follows. Section 3.1 explores the applied use of different models in the retail industry in 2016. Section 3.2 discusses the origins and theory of GIS buffer and overlay method for forecasting store revenue, a technique already associated with convenience revenue estimation. Thereafter section 3.3 reports on the literature surrounding the origins and development of spatial interaction modelling (SIM) for forecasting grocery store revenue, a method more traditionally associated with predicting large supermarket revenue. Thirdly, section 3.4 discusses the theory behind linear regression modelling as a tool for forecasting store revenue. Finally, section 3.5 reviews the methodologies not included in the suite of models reported in this thesis, discussing the reasons for their exclusion and justifying the inclusion of the models incorporated into the revenue predictions in chapter 7 to 11 of this thesis.

3.1 Applied use of location analysis techniques in the retail industry

Location planning teams are now centrally placed within large grocery retailers in the UK with specialised teams of analysts informing decisions using a suite of methods to identify new sites, estimate market shares associated with new and existing stores and forecasting revenues prior to new store network investment (Birkin et al. 2002). A survey conducted in 2010 by Reynolds and Wood found that the primary role of location planning teams within the grocery sector was to support and evaluate the financial case for new store investments (Reynolds and Wood, 2010). It is of primary importance for retailers to synergise the understanding of both supply and demand when evaluating possible new (or existing) sites for their potential (Birkin et al. 2002). Location analysis in the retail industry became increasingly sophisticated following the 'store wars' era detailed earlier in this thesis. Over time, the advancement of technology and increased availability of data have driven shifts in store location analysis to the point where complex models such as the SIMs and regression approaches are widely prevalent and commonly utilised tools informing the location

practice of major retailers. However, in reviewing the history of site location analysis, Clarke (1998) identifies that this was not always the case.

3.1.1 A Short History of site location in major grocery retailers in Great Britain

Although many new techniques were being theorised throughout the 1960s and 1970s the majority of retailers relied on gut-feeling by staff members in selecting locations for stores (Clarke, 1998). Reynolds and Wood (2010) identify this period as a time in which the retail location techniques used nowadays were poorly understood by retailers and were largely ignored by site location teams in favour of a 'gut feeling' approach based on the experience of senior managers. Gut feeling generally relies on the instincts of a senior member of staff who makes a decision following a site visit. This basic technique uses 'rules of thumb' derived from previous observations and trial and error by retail decision makers (Clarke *et al.,* 2000). This is a highly subjective approach relying on the experience of those making decisions and there are many anecdotal stories of stores forecasted incorrectly as a result of gut feeling.

Mendes and Themido (2004) highlight the most methodologically simple form of analogue site location analysis adopted by retailers; this involves creating a profile for a store location based on a checklist of factors deemed as strengths and weaknesses of existing stores of a specific type within a retailer's portfolio of outlets (Mendes and Themido, 2004). These factors are broken down into a number of store characteristics that may be desirable or undesirable depending upon the type of retail outlet proposed (e.g. supermarket, bank). The new site is then considered in a formula including rules and weights and a rating for that potential location is given which can then be used to make a decision and is generally referred to as checklist analysis. These are based on rules of thumb combined with trial and error but have been criticised for over-simplicity and being of a subjective nature. Despite limitations, these simplistic approaches have often been adopted in the development of an individual store or when time and capital are limited (Sulek *et al.,* 1995).

These analogue-based models came as a progression to increase the level of objectivity in site location assessment (Clarkson et al., 1996). Applebaum (1966) was the first to combine site selection experience with empirical data on the drivers of store performance. This method begins by selecting either one or multiple commercial spaces that are potential store locations. These locations are then classified based on a number of likely store attributes and compared on a pre-defined scale based on existing stores, drawing analogies between the proposed site and existing ones. This

can then be used to forecast sales if a store were to be place in one or more of the commercial spaces and used to make a decision on the optimum location for a new store. This method can be performed manually or by using regression modelling. By the mid-1980s, the development of Geographical Information Systems (GIS) had become widely available both academically and within retail organisations, prompting a change in analysis to increasingly complex forms of store forecasting. This advancement of technology was identified as a paradigm shift in which increasingly complex forms of market analysis become possible through the use of GIS (Morrison, 1994). More recently, a time in which computationally complex models can be used in retailing has arrived, greatly attributable to technological sophistication and the increasing availability of large and varied datasets.

3.1.2 Sophistication of approaches to location analysis

The removal of Retail Price Maintenance in 1964 detailed in the literature review in chapter 2 of this thesis and subsequent price wars led to a huge growth in the scale and power of the large grocery multiples and in an increasingly competitive market in which site location became crucially important to the success of stores and by extension retail firms (Ghosh and McLafferty, 1987). Bowlby et al. (1984) suggest four reasons for the increased adoption of systematic location analysis in favour of traditional simplistic methods of selecting suitable sites:

1. Easy site saturation
2. Decreasing usefulness of experience
3. Increasing cost of mistakes
4. Increasing Competition

With increasing expansion of large grocery firms the availability of easy sites in obviously prime locations diminished, increasing the importance of assessing potential non-prime pitches for their feasibility. Additionally, as the grocery market transformed with the increasing presence of the grocery multiples along with geodemographic and consumer shifts, experience gained by location analysts in the past became less useful as the market changed. Bowlby et al. (1984) also highlight the increased cost of investments due to the heightened competition for space leading to the costs of mistakes growing. Location analysts became central in reducing the level of risks in new store builds and acquisitions through accurate forecasting. Finally, a growing number of firms became Public Limited Companies (Plcs) pressuring retail managers to boost profitability increasing the desire of retailers to be located in the most profitable

locations. The effect this had is questionable as this was already happening to a certain extent, although it was likely to have maintained pressure of retail planning teams to find the best possible locations for new stores.

3.1.3 Retail consultancies and academic contribution

Academia has contributed significantly to location analysis models through the work of Reily (1929), Huff (1963), Lakshamanan and Hansen (1965), Applebaum (1965) and Wilson (1971) among others which are discussed in more detail in the next chapter of this thesis. Despite the increasing sophistication of theoretical models in academia, the use of such models was not in vogue in retailers in the 1960s and 1970s. This work has detailed the contributions made by Wilson and his contemporaries to transforming traditional gravity models into sophisticated SIMs that can be applied to support real world retail location decisions.

Wilson's contribution continued and gained an increasingly applied focus through the formation of GMAP Ltd. Founded in the late 1980s by Alan Wilson and Martin Clarke, GMAP was a consultancy section of the School of Geography at the University of Leeds specialising in developing models from the theoretical approaches developed in the 1960s, 70s and beyond into models that work in practice within a variety of retail organisations. Newing (2013, p 21) highlights the contribution of these academics to retail site location analysis:

> *"It was undoubtedly the work carried out by Sir Alan Wilson,*
> *Mark Birkin, Graham Clarke, Martin Clarke and their colleagues*
> *and clients at GMAP in the 1990s that has fully embedded the*
> *link between academic and industry practice in the*
> *development and application of spatial models for location-*
> *based decision making in the retail sector"*

Since GMAP paved the way, several other consultancy firms now work alongside blue chip clients in informing location decisions to attempt to give large retailers an advantage over competitors. Firms such as CACI, Javelin and CallCredit Information Group now operate alongside retailers and aid in location decision making. The development of geodemographic packages such as CACI's ACORN classification and Experian's MOSAIC classification system have advanced retail insight in attempting to gain an understanding of shopping behaviours by profiling customer types (Mitchell and McGoldrick, 1994). These products are widely used by in-house location planning teams within the large grocery multiples and represent a symbiosis between various

actors in the retail sector. Consultancy firms play a wide role in location planning and offer a wide range of services such as Store location assessment, Catchment area analysis, Assessment of store performance, forecasting of store turnover, market share estimation, geodemographic classification, customer targeting and credit scoring (Eyre, 1999).

3.1.4 Retailer in-house operations

Site location teams making use of a complex suite of sophisticated methods are now prevalent in the retail industry. When surveying the use of site location models, Reynolds and Wood (2010) found that grocery retailers are the most likely to have in-house specialised location planning teams and tend to use the most sophisticated techniques. The major grocery multiples discussed in depth in this thesis are at the forefront of location planning in retailing and have specialised in-house location planning teams with a data driven approach to location decision making (Broadley, 2013). Sainsbury's set the ball rolling on dedicated in-house location planning teams in the 1960s when it formed its Site Potential Statistics department. Far from adopting the increasingly complex gravity based models being developed by academics at the time, this team focused largely on simple analogue and checklist approaches comparing potential new sites to existing stores when assessing the potential for trade in new locations (Wright, 2008).

In the 1980s, Tesco became the first major grocery retailer to wholeheartedly commit to using modelling techniques to support location-based decision making. The retailer heavily invested in a project aimed at forecasting store turnover accurately. At the time Tesco was not the retailing giant it is today and much of the modelling focused on traditional simple techniques alongside multiple regression modelling. At this time Sainsbury's significantly improved their infrastructure for location planning teams through the adoption of desktop PCs with GIS capabilities allowing for a computerised spatial forecasting model to be employed to estimate store revenue (Wright, 2008). This allowed for a significant improvement in the applied nature of models to predict real world shopping behaviour as the GIS software enabled the retailer to incorporate drive-time data into their spatial models (Wrigley, 1998). In the 1980s and 1990s Sainsbury's remained sceptical of the implementation of gravity modelling as they believed it to be insensitive to factors important influences on their store performance, notably competitor strength and store access, knowledge gleaned from their checklist and analogous approaches to site location (Newing, 2013).

Tesco were the first major UK grocery multiple to incorporate a gravity model in the early 1980s into their suite of location planning techniques and by the early 1990s the retailer had an in-house system linking census data, branch databases, competitor insights and a digitized road network (O'Malley, 1995). The use of Spatial Interaction Modelling continued to grow substantially at the retailer and it now operates a complex gravity model (Mendes and Themido, 2004) and has grown to become the largest grocery retailer in the UK. It has been widely acknowledged that this commitment to accurate store forecasting has given Tesco the competitive advantage to be able to gain the market position it has today. Sainsbury's joined Tesco in developing a complex in-house Spatial Interaction Model built on gravity principles in the 1990s called the 'Grocery Store Potential Model' built out of the increased knowledge available to the retailer through the use of GIS and other more computationally complex computing hardware and software that became available (Wright, 2008).

Tesco and Sainsbury's along with other major grocery retailers including Marks and Spencer, Co-op, Morrisons and Aldi now employ in house location planning teams running suits of complex models including both gravity and regression approaches to store forecasting. The commitment to multi-channel and multi-format retailing is reflected in these teams as many retailers disaggregate their location planning to specific teams specialising in forecasting supermarket, online and convenience store sales along with teams associated with different product lines, including food, alcohol and non-food items such as home and garden products.

3.3.5.1 Uptake of specific methodologies by retailers

It is useful at this point to assess the extent to which different methods of revenue forecasting are adopted by retailers. Hernandez and Bennison (2000) conducted a survey of 50,000 retail outlets across a number of retailers across a variety of retail sectors in 1998 and found that the various site location techniques varied in their use by retailers. Reynolds and Wood (2010) continued research in this field with a follow up survey looking at the change in adoption of various site location methods between the two time periods. The findings of these studies are displayed in table 3.2.

Table 3.2 - Use of site location methods by retailers, 1998 and 2010

Method	Used At All (1998)	Used At All (2010)
Experience	96%	98%
Checklist	55%	91%
Analogue	39%	83%
Multiple Regression	40%	63%
Gravity	39%	64%

Source: Hernandez and Bennison (2000); Reynolds and Wood (2010)

The studies found that simple techniques were still most frequently adopted by retailers. However, several retailers adopt more complex modelling techniques and the number of retailers adopting these techniques is growing. Reynolds and Wood (2010) found that retailer's recourse to adopt more analytical techniques has continued to catch up with the reliance of companies on the experience of their location planning teams. Although retailers adopt more complex techniques, they do so in combination with traditional techniques and in fact adopt a wide range of methods to solve location-based decisions. Both multiple regression and gravity methodologies continue to be uptaken by retailers and have roughly the same rate of adoption, although gravity modelling is considerably more utilised in grocery retailing than in other retail businesses.

Newing (2013) highlights the example of Sainsbury's adopting more complex methods whilst maintaining an existing reliance on traditional location planning methods. The retailer uses its 'Grocery Store Potential Model' to generate revenue estimations. However, the predictions act as a guide and are adjusted by experienced analysts making use of site visits and the drawing of analogies to reach final forecasts for new sites (Newing, 2013). Wood and Tasker (2008) highlight the continuing importance of the site visit in store location assessment that is often missed in other academic work focusing on quantitative methodologies. Brown (1994) supports the use of the site visit in selecting the prime location to place a store within a selected catchment. It has been highlighted that such methods are far more practical in use as they can be performed by large numbers of people whereas complex models can often only be applied by specialists.

However, if we consider the major grocery retailers, all have site location teams using sophisticated models for site location assessments and the use of these models cannot simply be discredited especially when planning for greenfield developments. Yet, Wood and Tasker (2008) advocate that *"… despite huge advances in the tools and*

assessment techniques of site appraisal, there is no substitute for the field visit and the observations across a range of spatial scales and times of day…" (Wood and Tasker, 2008, P. 152). They additionally suggest that adjustments made to gravity models following site observations can lead to far more accurate forecasts. In 2004, Justin King the chief executive of Sainsbury's Justin King conceded that they had made mistakes in a number of site location decisions and they were forced to close 12 stores (Wood and Browne, 2007) highlighting the importance of retail site location to retailers and the potential affect poor location decisions can have on a major grocery retailers operations.

3.2 GIS buffer and overlay

The retail modelling literature often discusses and suggests a methodology of buffering and overlaying in a Geographic Information System (GIS) as a basic method of forecasting grocery store sales (Birkin et al. 1999). As recently as 2013, representatives from both Sainsbury's and Morrisons gave talks to students at the University of Leeds in which they discussed the continued use of this method to aid in the forecasting of grocery stores at each retailer. They note the continued used of this method for forecasting convenience grocery stores in which they believed that the immediate catchment area around the store is integral to the success of a store location. They argued that simple buffer analysis is more utilised in understanding revenue of convenience stores than the elaborate suite of spatial models that they apply in supermarket revenue forecasting (Gell and Mulchacy, 2013; Brodley, 2013).

A number of steps are required to forecast revenue using this method. Firstly, a travel time (or distance) from which customers would be willing to travel to a store location is decided upon. This may vary based on a number of factors such as the size of the store or the age of the population in an area. Next, this distance is drawn out using a GIS to limit the population available within the given distance or travel time around the store (Birkin et al. 1999). This is the buffer procedure. The overlay procedure is next and is the process of overlaying an expenditure layer pertaining to the available spend on a given product for a given geography in the buffered area. In the example presented in this thesis, this is the weekly household level expenditure on groceries at the census output area level. In its crudest form, the total expenditure within the buffer is allocated to the store that is being predicted. However, the analysis generally

involves the overlaying of additional data. A store layer is generally added identifying all potential grocery destinations available to consumers (Elliot, 1991).

In the presence of additional stores to the store being buffered for revenue prediction, a method of allocating expenditure to each competing grocery destination is required. Beaumont (1991) identifies the fair share method of dividing available revenue within the buffer by each available store. In practice, the fair share method is allocated in one of two ways which are acknowledged by Birkin et al. (1999). Expenditure can either be allocated equally to each store in which the characteristics of each store are not considered or revenue can be allocated based on the characteristics of each store in the buffer. The most common form of this is by taking into account the size of each store and allocating based on the fair share method so that each square foot of floorspace is deemed to attract an equal proportion of available expenditure. Following allocation of available expenditure, each store or potential new store site will then have an estimate of available revenue.

3.2.1 Advantages of GIS buffer and overlay for forecasting store revenue

A major advantage of this method is in its ease of use which comprises a variety of components. Firstly, the method requires little time to implement in a GIS and can be built into a black box model which can be utilised very quickly and requires minimal computational power to maintain and run. The analysis produced from this model is easily repeatable and reproducible in different contexts requiring some knowledge about the market in which revenue forecasts are to be applied to.

Aside from building and running a black box model, it is relatively straight forward to learn and train to be able to have the GIS knowledge required to run this model from scratch meaning large retailers would have little issue in adding this method to their suite of forecasting tools - although the largest grocery retailers in GB already adopt this methodology to at least some extent. With the fundamentals of a series of buffers, an overlay of available demand for a product(s), overlay of available supply and a method of distributing revenue between competing retailers it is relatively straightforward to tailor the model to the type of product/store being forecast. This could include introducing extra data to the process such as identifying consumer preference for certain retailers over others.

3.2.2 Disadvantages of GIS buffer and overlay for forecasting store revenue

The literature surrounding GIS buffer and overlay modelling for forecasting grocery stores has generally focused on the disadvantages of this method over the advantages of using this methodology. Many of these criticisms have focused on how simple the model is until problems are reached which can be difficult or cumbersome to overcome effectively. Birkin et al. (1999) identify the issue of allocation of money within each buffer. The assumption often made within the model means that customers close to the store are deemed to be equally likely to travel to the store as those on the edge of the buffer. Moreover, the buffering process can be problematic if catchment areas overlap leading to issues in allocating expenditure and is compounded if two stores are located very close together and have overlapping catchment areas (Birkin et al. 1999). This issue is more likely in the convenience market in which retailers compete side by side, particularly in attractive neighbourhood catchments.

Benoit and Clarke (1997) and Clarke (1998) discus the GIS buffer and overlay methodology for forecasting grocery stores and both papers are critical of the method for its simplistic representation of a complex system. The method is criticised as it does not allow for the complex set of real interactions between residential areas and retail locations which are distorted by intervening opportunities Clarke (1998). To this end the model places bounds on the distance consumers in certain zones are able to travel to a given store when the barrier may not exist in reality. The opposite problem of this can also be experienced in which consumers in the model are allowed to travel beyond barriers such as an extensive retail centre thus distorting revenue predictions. Spatial interaction modelling (SIM) has been noted as a methodology capable of overcoming some of the issues of GIS buffer and overlay modelling which essentially allow free catchment areas constrained by the retail supply rather than a defined catchment size which can often be arbitrary in nature.

3.3 Spatial Interaction Modelling

The second major method of revenue forecasting used in this research is a Spatial Interaction Model (SIM). SIMs are used to predict the interactions (often referred to as flows) between a combination of origin locations and destination locations and can models flows of a number of variables including money, people and goods. They are a commonly used forecasting tool and have a long history of application in the context of retail systems and consumer behaviour (Birkin et al. 2002). A SIM in a grocery

revenue prediction context is designed to capture the flow of expenditure on groceries between geographical zones (traditionally residential) and a number of competing grocery centres (or stores). The model's theoretical underpinning rests upon two key assumptions; 1) flows between an origin and a destination will be proportional to the relative attraction of competing destinations; and 2) flows will relate to the relative accessibility of competing destinations from the origin (Guy, 2011). Investigating the application of a SIM to the convenience grocery market, it is first necessary to review the work that has already been completed in the field of spatial interaction focusing on the application of SIMs in a retail context.

3.3.1 The Origin of Spatial Interaction Models

Often referred to as gravity models, SIMs have their foundations in Sir Isaac Newton's work in the 17th century on the theory of interaction between physical bodies in space known as the Theory of Universal Gravitation (Roy and Thill, 2004). In this theory, the gravitational force (*F*) between two masses (M_i and M_j) increases proportionally with the product of the two masses, and inversely with the distance between the two masses and is given by:

$$F_{ij} = G \frac{M_i M_j}{d_{ij}^2}$$

(Equation 3.1)

Where *G* is the gravitational constant and d_{ij} is the distance between the two masses M_i and M_j. The model can be expressed to more acutely show the effect of distance so that:

$$F_{ij} = G M_i M_j \left(d_{ij} \right)^{-2}$$

(Equation 3.2)

In the study of human retail interaction, F_{ij} could represent the number of people travelling from a residence zone to a shopping location. The mass of the origin M_i could represent the population of a residential zone and M_j may represent the size of a retail centre (or store) and d_{ij} may represent some measure of physical distance between *i* and *j* (or some measure of the cost of travel between the two which will be addressed later in this chapter) (Dennett, 2010). The inverse square of the distance is the distance decay factor in the Newtonian model which works well for physical systems but has been found to be inadequate in human systems which will also be discussed later in this chapter.

The first attempts at applying gravitational principles to model consumer behaviour in the retail industry are generally attributed to Reilly (1929) who used this context to analyse retail trade areas. This was a form of central place theory attempting to determine the relative attractiveness of two competing retail centres (cities A and B) to customers living between the two centres (in an intermediate city). The research hypothesised that attractiveness (and therefore flows of people between the origin and the two competing destinations) was in direct proportion to the population of the town and in inverse proportion to the distance between the town and each centre. The mathematical expression of Reilly's hypothesis is as follows (Huff, 1963):

$$\frac{B_a}{B_b} = \left(\frac{P_a}{P_b}\right) \left(\frac{D_b}{D_a}\right)^2$$

(Equation 3.3)

Where

B_a = the proportion of the trade from the intermediate city attracted by city A

B_b = the proportion of the trade from the intermediate city attracted by city B

P_a = the population of city A

P_b = the population of city B

D_a = the distance from the intermediate town to city A; and

D_b = the distance from the intermediate town to city B.

In the late 1940s, Converse (1949) made a significant extension to the Reilly (1929) model by introducing the concept of a 'break point' in trade or the point up to which one centre exerts a dominant trading influence and beyond which a different centre becomes dominant (Huff, 1963). Both the original and adapted version of Reilly's 1929 model is theoretically underpinned by Central Place Theory, a theory postulating that consumers make rational decisions and are likely to visit the nearest centre (Dawson, 1980) and the social physics in that larger centres have a greater pull factor and the flow is greater from an origin location with a greater mass. Whilst the adaptation of the model to include a breaking point brought increased sophistication to the technique, it was still unrealistic as in reality trading dominance will decrease gradually with distance, rather than suddenly stop. Furthermore, Eyre (1999) identifies the issue that break points do not allow for overlapping trading areas which would be seen in real world purchasing behaviour.

3.3.2 The probabilistic model

Along with evaluating the positive and negative components of early attempts at retail specific gravity models and testing their effectiveness in different contexts, Huff (1963) made arguably the first significant advancement in applying gravitational concepts to accurately model human behaviour through an increased consideration of the consumer as an actor which overcame some of the problems and criticisms of Reilly's model. Huff's model introduces probabilities by calculating the chance of a consumer from a given origin *i* will travel to centre *j* to fulfil their shopping needs. This is computed mathematically by identifying the attractiveness of centre j as a proportion of the attractiveness of all available retail centres and relative accessibilities of the competing destinations.

This model advanced the distance term (d_{ij}) in the context of retail SIMs by reflecting travel time rather than straight line Euclidean distance as has been used previously, in affect creating a cost component to distance that is measurable by the consumer (Roy and Thill, 2004). The model also incorporated an additional parameter which attempted to empirically assess the effect of the cost (or distance) of travel on different kinds of shopping trips. This was the first work to identify the potential for variable distance decay values reflecting the relative impact of travel cost/time/distance on the ability different consumers to travel different distances for different shopping missions. Furthermore, this model overcame Reilly's issue of heterogeneous trading catchments by allowing trading areas to be graduated by probability contours (Eyre, 1999). Lakshmanan and Hansen (1965) built upon Huff's modelling efforts in the form of a model of Metrotown Centres in Baltimore by advancing a model that looked at the retail sales potential of retail centres. Their model postulated that sales are directly linked to the size of a centre, its location to consumers and the purchasing power of those consumers (Lakshmanan and Hansen, 1965). Furthermore, sales would therefore be inextricably linked with the location and size of other competing centres.

A major criticism of many social physics applications of gravity model theory is the mathematics of the Newtonian equation. The multiplicative nature of the equation means that if an origin and destination mass are doubled, the size of the interaction between them is not doubled but quadrupled (Dennett, 2010). For example:

$$1 \times 10 \times 10 = 100$$
$$1 \times 20 \times 20 = 400$$

In order to counteract this problem, it is possible to introduce constraints into the system based on known characteristics of the system being modelled. This was first comprehensively explored and detailed by the work of Alan Wilson in the 1960s and 1970s who proposed a family of spatial interaction models using different constraints depending on the type of system being modelled or the part of the system being predicted (Batty and Mackie, 1972).

3.3.3 The Family of Spatial Interaction Models

To begin his exploration into gravity modelling for human behaviour applications, Wilson (1971) expressed the general SIM version of the gravity model as:

$$T_{ij} = kO_i D_j f(d_{ij})$$

(Equation 3.4)

In this form of the model T_{ij} replaces the F_{ij} term in the Newtonian gravity model and the M_i and M_j terms are replaced by O_i and D_j which denote information about origin and destination masses. He proposed a family of four models, each incorporating information about a system and applying the information in the form of a constraint to stop unexpected increases in interactions between the O_i and D_j terms. Additionally, the inverse square distance function in the gravity model is substituted for the $f(d_{ij})$ function denoting the cost of travel between an origin and a number of destination zones, the decreasing interaction between two locations the further apart they are in space. The gravitational force term *G* in the Newtonian model is replaced by the constant *k* which acts as a balancing factor ensuring that the sum of all interactions between origins and destinations is equivalent to known information about the total flows in the system.

Wilson's four models cover various combinations of known information about the origins and destinations in the system being modelled and take the following forms:

1) O_i and D_j are unknown – the unconstrained model
2) O_i known, D_j unknown – the production constrained model
3) O_i unknown, D_j known – the attraction constrained model
4) O_i and D_j both known – the production-attraction or doubly constrained model

3.3.3.1   The unconstrained model

The first of Wilson's models is the unconstrained model in which the O_i and D_j terms are unknown and are therefore replaced by proxy terms referred to as an

attractiveness value. Wilson uses the mathematical symbols $W_i{}^1$ when O_i is unknown and $W_j{}^2$ when D_j is unknown (Wilson, 1971). Harland (2008) highlights this model as being more accurately described as the total constrained model as an overall constraint is applied in the form of:

$$\sum_i \sum_j \hat{T}_{ij} = T$$

(Equation 3.5)

Where T is the total number of interactions (overall flows) that is known and must be equalled by the sum of interactions between the O_i and D_j terms. The overall gravity model equation therefore becomes:

$$\hat{T}_{ij} = kW_i{}^1W_j{}^2f(d_{ij})$$

(Equation 3.6)

In which the balancing factor *k* is calculated as:

$$k = \frac{T}{\sum_i \sum_j W_i{}^1W_j{}^2f(d_{ij})}$$

(Equation 3.7)

3.2.3.2   The production constrained model

When the O_i term is known but the D_j is not, the following constraint can be applied:

$$\sum_j \hat{T}_{ij} = O_i$$

(Equation 3.8)

This means that all simulated flows to destinations from an origin *i* must add up to the known origin outflow value O_i. A retail analogy could be if we know the population in residential zone *O* spend £1000 per week on groceries and are limited to two potential grocery stores, the total sum of interactions from residential zone *O* to the two grocery stores must add up to the initial £1000 outflow value. To ensure that the constraint is satisfied the equation becomes:

$$\hat{T}_{ij} = A_iO_iW_j{}^2f(d_{ij})$$

(Equation 3.9)

In which Wilson replaced the constant *k* with the balancing factor A_i calculated as:

$$A_i = \frac{T}{\sum_j W_j{}^2f(d_{ij})}$$

(Equation 3.10)

3.2.3.3   The attraction constrained model

When the D_j term is known but the O_i term is not, the following constraint can be applied:

$$\sum_i \hat{T}_{ij} = D_j$$

(Equation 3.11)

This means that all simulated flows from origins to a given destination *j* must add up to the known destination inflow value D_j although we do not know the actual flows from each origin which sum to give the total inflow D_j. The balancing factor B_j replaces the constant *k:*

$$\hat{T}_{ij} = B_j W_i^1 D_j f\left(d_{ij}\right)$$

(Equation 3.12)

in which B_j is calculated as:

$$B_i = \frac{T}{\sum_i W_i^1 f\left(d_{ij}\right)}$$

(Equation 3.13)

3.2.3.4   The production-attraction (doubly constrained) model

The final model of Wilson's family of SIMs is the doubly constrained production-attraction Spatial Interaction Model in which both O_i and D_j are known. This model requires no W attractiveness terms relating to unknown origin and destination terms. This creates a mathematical problem in which both the A_i and B_j constraints needs to be satisfied and are interdependent in the equation and can be resolved using an iterative process (Senior, 1979) as part of the equation:

$$\hat{T}_{ij} = A_i B_j O_i D_j f\left(d_{ij}\right)$$

(Equation 3.14)

in which the balancing factors A_i and B_j are calculated:

$$A_i = \frac{1}{\sum_j B_j D_j\left(d_{ij}\right)}$$

(Equation 3.15a)

$$B_j = \frac{1}{\sum_i A_i O_i f(d_{ij})}$$

<div align="right">(Equation 3.15b)</div>

These four models have continued to be the major theoretical underpinning of Spatial Interaction Models in an academic context but have been developed further since Wilson first proposed their use.

Further developments in spatial interaction modelling have occurred since the work of Alan Wilson in the 1970s. The work of Fotheringham (1983; 1986) introduced the idea of competing destinations which was devised due to a theoretical criticism of the nature of the standard entropy model. He argued that stores grouped close to each other were often seen as a single destination when a consumer was making decisions on where to travel to for retailing. Moreover, other studies have further disaggregated the demand side of grocery retailing. Examples of these have been disaggregating for different retail sectors and channels (Birkin et al. 2004), for varying socio-economic characteristics of consumers (Khawaldah, 2012), for incorporating demand for discount retailing (Thompson, 2013) and disaggregating demand by residential and visitor expenditure patterns (Newing, 2013)

Operationalising SIMs

A major issue in developing applied SIMs is the requirement of good quality, available data. Birkin et al. (2010) highlight four areas of data that are important when constructing and applying SIMs; demand estimations, supply estimations, calculation of the impedance function and calibration. Demand data is often calculated by extrapolating the sample acquired from surveys, these surveys can vary in scope, availability of the data, accuracy of reporting in the data (particularly if the data is self-reported) and size (Thompson, 2014). A further issue in demand estimation in retail modelling is the problem of determining potential customers; what proportion of the population are potential customers? Is there potential for customers from outside the study area? E.g. Visitor demand in tourist areas (Newing, 2013). When collating destination data in retail models, an up-to-date database of retail centres or stores and their attractiveness is desirable.

 The fast-paced nature of retailing and the often short-notice opening and closing of outlets can be difficult to keep pace with, as highlighted by the rapid growth of the branded convenience market discussed in this thesis. Measuring store attractiveness can prove problematic as drivers of attractiveness may differ by individuals or by groups of individuals. Store (or centre) size and turnover have traditionally been used

as a measure of attractiveness but other factors such as brand, price, customer perception, available parking or access by public transport could be used and may be important to the consumer (Birkin et al. 2010). Calculation of the impedance function, the traditional distance (or cost function) in the traditional Spatial Interaction Model presents another set of choices in the application of SIMs that present a challenge to the modeller. It is possible to use a number of measures including but not limited to; Classic Euclidean straight line distance, distance on a transport network, travel time, or travel cost in monetary terms.

Finally, calibration is a key component of SIM application if improvements in accuracy are desired. Access to good quality real world data to calibrate and test models is often limited as retailers are reluctant to release data due to its value in a competitive market. However, through the emergence of loyalty card data collection retailers now have a wealth of data at their disposal to aid in their model development and application. Thompson (2014) draws attention to the fact that gaining access to one retailers data within a complex market of multiple actors may still lead to bias in the form of inaccurate parameter estimates for other competitors (Birkin et al. 2010; Thompson, 2013). However, it is difficult to acquire store revenue for a single retailer, let alone two or more retailers operating in a given geographical context.

The application of the SIM in chapter 10 attempts to address the challenges and limitations presented in this section which are discussed further before number of potential solutions that have a particular application in the forecasting of convenience grocery store sales are presented and evaluated to hopefully add to the theoretical understanding of applied SIMs in a new context.

3.3.4 Criticisms and limitations of spatial interaction models

It is important to identify and understand the limitations of this method before it is adopted in the context of predicting convenience grocery store sales. As detailed in the previous sections, a number of early criticisms of gravity models have been overcome (to varying degrees) by changes to the terms in the model equation. Senior (1979) criticises the Newtonian gravity model's quadrupling of interactions with the doubling of origin and destination masses. However, we have seen that this problem was dealt with by Wilson through the introduction of constraints into the model to maintain that the overall number of interactions cannot exceed a known total number of interactions. A further major criticism of this type of model is its roots in an analogy to the physics of celestial entities. Early commentators criticised the method as lacking a sound intellectual base in human behaviour and urban theory (Foot, 1981; Harvey, 1969). In

being a 'borrowed' theory, this type of model has been criticised in its lacking of a theoretical behavioural justification and the danger of assuming that statistical relationships equate to causal relationships. However, the entropy maximising work of Alan Wilson and other academics overcame this issue effectively.

The aggregate zonal geography of the model has also been to various degrees criticised and viewed with a degree of mistrust. Out-and-out opponents to zonal models have criticised its inability to capture individual human behaviour or allow individuals in the model to exert agency (Thomas and Huggett, 1980). Senior (1979) highlights the potentially problematic nature of assuming that the behaviour of a single individual can correspond to the placing of behavioural traits on a group of those individuals. Less harshly, a criticism of all research using zones as a base unit is the two broad issues of the Modifiable Areal Unit Problem (MAUP) and ecological fallacy, both of which must be considered when developing a Spatial Interaction Model.

Wood and Browne (2007, P. 241) highlight the traditional gravity model in which *"[the]…probability of a consumer using a given store diminishes at an exponential rate as travel time increases".* For many convenience store locations, this is problematic as it is difficult to accommodate passing trade into this type of model and the flow within the model (potential customers to store) is difficult to conceptualise. This creates difficulty in setting a catchment area from which trade may flow into a store and makes the gravity flow model difficult to apply to convenience stores. Other site location methods may therefore be more appropriate; moreover, Wood and Tasker (2008) suggest gravity modelling is arguably less appropriate in forecasting stores of this size. As detailed earlier, the distance parameter in gravity modelling is problematic when applying it to convenience stores. Whilst drive-times are generally applied in supermarket forecasting, a walking time parameter may be more useful for convenience stores. This requires a far more detailed level of geography that is often unavailable (Wood and Browne, 2007).

3.4 Regression Modelling

Regression modelling is the third method of grocery sales forecasting used in the research in this research and is used to predict the revenue of convenience stores operated by Sainsbury's in Yorkshire and the Humber. The process of predicting one variable from another can be used in retailing to predict sales. Regression is a form of analysis by which we can predict an outcome variable (in this case store sales) from one (simple regression) or several predictor variables (multiple regression). It can be

an incredibly powerful tool as it allows the user to go one step beyond the data that has been collected to produce new data, such as store revenue for additional stores (Field, 2009). As will be discussed later in this chapter, this method is commonly applied in the retail industry in the UK and worldwide. However, little testing of its effectiveness in predicting grocery sales, and in particular convenience grocery store sales, appears in the academic literature.

3.4.1 The simple regression equation

Functionally, regression fits a model which best describes the data in the form of a straight line. This straight line contains two pieces of information: 1) the slope (or gradient) of the line and 2) the intercept, the point at which the line crosses the x axis of the graph. These two terms are shown in the simple regression model in equation 3.18, a model with one predictor variable (Field, 2009):

$$Y_i = (b_0 + b_1 X_1)$$

<div align="right">(Equation 3.18)</div>

Where:

- Y_i is the outcome we want to predict for store i (e.g. store sales)
- b_1 is the slope of the straight line
- b_0 is the intercept or that line, often referred to as the constant
- X_1 is the value of variable X for store i

Let's look at a hypothetical example in which we look to assess the relationship between store size and store sales, a commonly investigated relationship in grocery retailing. Table 3.1a presents data on 11 grocery stores, including their size and sales. When running this simple regression model, the outcomes of the model (Y_i, b_0, b_1) are shown in table 3.1b, a typical regression analysis output.

Table 3.1 a) input variables, and **b)** regression output

a) Input Variables		
Store	**Sales (£)**	**Store Size**
1	1000	50
2	1500	150
3	1500	175
4	4500	200
5	10,000	210
6	75,000	250
7	100,000	375
8	500,000	425
9	200,000	500
10	150,000	100
11	75,000	225

b) Regression Output		
Model		Coefficients
1	(Constant)	- £67697.74
	Store Size	+700.44
a. Dependent Variable: Store Sales		

3.2.2 Using a simple regression model to predict sales

Table 3.1b identifies the constant intercept value b_0 and the coefficient b_1 (= +£700.44) for the store size variable being used as a predictor of store sales. These two values can be used to predict the sales of a new, out of sample store, denoted by the letter Z in equation 3.19.

$$Z_i = (-67967.74) + (£700.44 * Store\ Size\ Value)$$

(Equation 3.19)

In the example above the coefficient for store size is positive, meaning other things equal, a one square foot increase in store size gives an increase in sales of £700.44. Furthermore, the constant (intercept value) is the point at which the straight line crosses the X axis when graphed and the base point from which to calculate the equation. If we consider a hypothetical 225 square foot store, we can plug in the simple regression model to predict sales in which:

$$225\ square\ foot\ store\ sales = (-67967.74) + (£700.44 * 225)$$
$$Sales = £59,631.26$$

(Equation 3.20)

3.4.2 Multiple Regression Modelling

The example detailed in the previous involved a single predictor variable being used to predict the outcome variable (store sales). In reality, the variation in store sales experienced by individual grocery store sales will be the product of the variation in a number of predictor variables. Thus, a methodology allowing the quantification of the relationship between a number of variables and store sales is required. A multiple linear regression framework lends itself well to this aim. The equation maintains the

same structure as the equation shown in equation 3.18, with the addition of additional predictor variable coefficients. The general multiple regression equation is expressed in equation 3.21:

$$Y_i = (b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \ldots b_nX_n)$$

(Equation 3.21)

The change made when moving from the simple one predictor model is the addition of n coefficients for each additional predictor variable introduced into the model. Each b represents an additional coefficient and b_n represents the number of predictor variables in the model. Once again, the model tries to fit the best plane to the data based on the multiple straight line relationships between each predictor variable and the dependent variable (Field, 2009). When defining the relationship between each predictor variable and the dependent variable, the model takes into account the other variables explaining the variance in the dependent variable, thus giving the chance to assess the relationship between each variable and sales, with other things being equal (i.e. accounting for the other predictor variables).

3.4.2.1   Advantages of multiple regression modelling

Multiple regression modelling has a number of positives when applied to forecasting grocery store sales. Firstly, it lessens the subjectivity as results can be expressed in terms of a quantified statistical confidence in the strength and significance of a relationship between a predictor variable and store sales (Wilson, 1984). This advances the degree of sophistication in the model and moves beyond a simple checklist approach or the reliance placed on site visits. Moreover, Rogers (1992) identifies a number of keys strengths of adopting a multiple regression methodology for predicting store sales. As detailed above, Rogers highlights the importance of objectivity in that each model can be purpose built to the demand required or to the type of retailer being analysed giving a level of flexibility to the model and its specifications (Rogers, 1992). However, the selection of which variables enter the model can be very subjective. This type of model can also be applied 'backwards' to evaluate existing stores against the performance of other store across the network. Finally, and significant in the context of this research is the model's ease of use once developed. This is somewhat in contrast to a spatial interaction methodology which requires a significant investment in time in collecting new information and inputting this in to the model with any changes to the grocery market.

3.4.2.2 Limitations of multiple regression modelling

The positive increase in sophistication comes at a cost. Much of this cost comes in the form of an increase in computing requirements when compared to simple checklists and gut feeling (Hernandez and Bennison, 2000). However, this computing requirement is less than that required for a comprehensive spatial interaction model being used to predict sales for the same number of stores. This quantitative sophistication also creates additional data requirements, particular in terms of the volume of data required for the model to run and produce robust coefficients. Rogers (1992) suggests that a database of at least 30 stores is required for a multiple regression model to be applicable. This created issues in the context of this research which are discussed later in this thesis. Moreover, in order to quantify the relationship between a number of predictor variables and store sales, store sales data itself is required which can prove difficult in academic studies, but is more plausible for in house retailer operations (Simkin, 1989). The high number of observations required can create an issue in the collection of data (Mendes and Themido, 2004). However, with the exception of the sales data, all data used as input variables in the regression model presented in this thesis is freely available and can be batch downloaded from various official UK data sources.

The additional sophistication requires a degree of expertise in much the same way as spatial interaction modelling. Rogers (1992) highlights the potential for misuse of this technique through a lack of understanding of the potential and limitations of it, highlighting the importance of a nuanced understanding on the part of the user. This misuse may come as a by-product of the easy to use nature of this modelling approach, resulting in it being applied wrongly or not kept up to date. Coates et al. (2006) raise a criticism of the fact that multiple regression provides a static representation of a dynamic industry. In spatial interaction modelling, it is relatively simple to tweak the model as a result of shifts in consumer behaviour or changes to the supply side of the grocery offer in the study area without rewriting the model. However, in regression modelling, new parameters must be defined which can be time consuming and the model itself may have a limited shelf life.

The static nature of the model is problematic when it comes to utilising the model for future planning. If there is a change in the market, this is difficult for the model to account for this as the coefficients are a static representation of the relationship between a number of variables and store revenue. The model cannot account for the effect of an increase or decrease of competition dynamically whereas a SIM can

account for this easily. Moreover, whilst a regression model can be used to generate total revenue for a store, it does not model flows from individual demand zones to each store. Moutinho et al., (1993) highlight a problem that arises when using regression modelling to forecast sales in retailing. If variables within the model are added or removed, the coefficients will change and the model may therefore be difficult to trust. This is problematic when considering small-store retailing as the factors driving expenditure may differ markedly for stores for different location types. Morphet (1991) suggests a stepwise regression method of adding variables one by one to isolate the effect of each variable.

An additional issue in using a regression model to forecast grocery store revenue lies in defining a catchment area and defining the type of demand that may be originating from the catchment area. Whilst SIMs allow for 'natural' catchment areas to occur in the model by modelling the distance travelled by different types of consumers, regression approaches require the manual definition of a catchment areas from which to derive predictor variables. This can be problematic as convenience stores differ markedly in their customer base and type of area in which they trade. Unlike supermarket one stop trips, convenience retailing often relies on linked trips combining purchasing with other activities. This creates an issue in defining the population falling within the catchment area of a store (Wood and Browne, 2006). Whereas residential population is of prime importance in forecasting sales of a supermarket, daytime populations such as nearby workplace populations may drive trade for a convenience store; this can create issues in forecasting as reliable micro-scale workplace population data is often difficult. Wood and Tasker (2008) adopt a more comprehensive strategy in dealing with the population in a stores catchment area. They take into account both census residential population within the store catchment area and the workplace population within a ¼ mile radius of the proposed convenience store.

A further problem often encountered in using regression to estimate store revenue is multicollinearity. Multicollinearity is an occurrence in which two or more explanatory variables in a regression model are highly correlated. This potentially creates a number of issues. Firstly, it can be difficult to identify the predictor variable which has a relationship with store revenue as the variance it accounts for in the dependent variable may be coincidentally accounted for by a third variable that has no bearing on the dependent variable in reality. Secondly, it can cause the coefficients created by the model to be unstable making the model unreliable as a tool for prediction revenue. In developing a model to predict grocery retailing, experiencing multicollinearity is likely.

For example, city centre locations will have a high workplace populations, retail adjacencies and transport hubs in comparison to non-central store locations and the values for these variables may correlate highly across different store locations as a result.

3.4.3 Applied regression modelling using retail data

Mendes and Themido (2004) highlight the smaller economies of scale available in smaller format grocery stores and emphasise the importance in site location analysis in ensuring the success of this type of store. As major grocery retailers with advanced site location teams direct attention to convenience retailing, maximising the effectiveness of location decisions is of prime importance. Despite this, analysis of store location of convenience stores has been relatively neglected in the academic literature in comparison to supermarkets, often attributed to the habitual nature of weekly expenditure in a grocery 'big shop' that lends itself to statistical modelling (Wood and Brown, 2007). According to Tasker, it is anecdotally believed that superstore forecasting uses 80% office and 20% site visit study whereas the reverse is applicable for convenience stores (Tasker, 2005).

Various studies have looked at the variables that should be included when applying a multiple regression approach for the purpose of retail sales prediction. Simkin (1989) reviewed the factors taken into consideration by UK retailers (multiple product types) and found that variables relating to trading area composition, store accessibility, store characteristics and catchment demographics were the broad categories in which variables were tested for their relationship with store sales data. Moreover, in a review of the academic literature into sales forecasting using this approach, Hoch et al. (1995) found a number of broad categories in which variables could be placed including population characteristics, economic factors (of the target population), competition and store characteristics.

The method by which variables are selected, tested and finalised for the final regression equation is important. The majority of literature advocates for a stepwise regression approach by which a number of variables are selected and tested, with a computational algorithm placing each variable into the model and finding the variable that is associated the greatest variance in the dependent variable (store sales in the literature analysed) before adding each variable in order of their association until the addition of variables does not improve the model (Simkin, 1989; Coates et al. 2006; Rogers, 1992).

74

Previous academic work into convenience store location analysis has been conducted using regression modelling. Morphet (1991) categorised grocery stores into being either free standing or not free standing, the analysis focusing solely on free standing convenience stores, judged as separate from high streets and shopping precincts. The study conducted used multiple regression analysis with the objective of relating turnovers of stores in a chain to the characteristics of the urban area in which they were located, defined as being the trade area of that store. The regression equations formulated included variables related to residential population, share of floorspace, distance to larger centres and the relative attractiveness of a centre.

In terms of the general ability to apply regression modelling to sales forecasting, Coates et al. (2006) found that this type of approach paid dividends when forecasting fashion store sales and found little evidence in the research to suggest that it would not be appropriate when extending to other retail sectors, provided sufficient exploration of important explanatory variables is conducted. Moreover, Rogers (1992) found this type of model to be an effective predictor of sales, despite warning that claims of consistent accuracy at within ±10% of actual sales from consultancy firms is dubious to say the least, and found no reason for it not be considered as a viable model type when evaluating the best method of sales forecasting.

Wood and Tasker (2008) forecasted the opening of a Sainsbury's convenience store in Hayes, Kent. They used a simple multiple regression model giving locational attributes points based on Sainsbury's existing convenience store portfolio. These locational factors were; population, footfall, stopability, footfall drivers (transport nodes, bus stops, etc.) and competition. The store was then given an overall points score, the sum of all attribute scores, equating to a predicted store turnover value. The store was given a high attribute score for footfall drivers due to it being located close to the train station. An evaluation of performance a few months after trading began determined that the store was struggling to attract passing trading acting more like a top-up store, similar to convenience stores in residential neighbourhoods. As a result of this, the stores score for footfall drivers was reduced and the store traded on forecast. The paper encourages analysts *"…to proactively amend in-store modelling when their experience 'on the ground' gives them cause to over-ride quantitative outputs."* (Wood and Tasker, 2008, P. 149).

A recurring theme in the literature is the theory (tested in some cases) that stores of a similar type will be more analogous to each other than those of a different type, and therefore a series of models for different store types may prove the most fruitful when

attempting to accurately and robustly predict store sales. Rogers (1992) advocates a method by which existing store sales are compared with measures of variables which are expected to influence (positively or negatively) store sales which are reasonably analogous to those which will be developed in the future. Reading between the lines, if stores are of a different type (or in a different location) they may warrant drawing upon a different set of store sales and their relationship with a number of potential predictor variables. Furthermore, Rogers (1992) argued that this methodology might be of particular appeal to retailers serving a particular segment of the population (or market) as it could derive statistical relationships between their target customers and sales.

Morphet (1991) deals with the issue of small grocery stores inhabiting various locations by only selecting standalone convenience stores within the study area and adjusting the variables accordingly. In a comprehensive study, it would be possible to disaggregate regression based models, assigning a different set of predictor variables to the regression equation for different stores types (dependent variables). In an applied study, Themido et al. (1998) developed a series of multiple regression models for the Portuguese gasoline market and found that the utility of 6 individual specific models of predicting gasoline sales in different petrol station location types was more robust than using a one size fits all model for predicting sales across all petrol stations (Themido et al. 1998). In theorising a practical research agenda for investigating location analysis in convenience stores, Wood and Browne (2007) suggest assessing different drivers of performance for different classifications of neighbourhood store location and advocate the use of a separate model for each locational type. Dividing the convenience grocery market into locational types and attempting to apply a suite of regression models to predict store sales raises practical concerns due to the limitations of multiple regression modelling highlighted earlier in this chapter. The regression model presented in chapter 9 of this thesis investigates the utility of multiple regression modelling for forecasting convenience grocery sales in general and specifically in each locational type identified in chapter 6.

3.5 Methodologies not adopted in this thesis

Whilst this thesis focuses on three methods of store location analysis, namely GIS buffer and overlay, multiple regression and the spatial interaction model, other researchers and practitioners have (and indeed continue to use) alternative methodologies for site selection and evaluation. This can be due to a number of reasons such as limitations in computational power (meaning simple techniques are

more suitable) to a belief that modelling behaviour in aggregate (through zonal geographies) is problematic due to heterogeneous behaviours in the real world (in the case of agent based methodologies). Thompson (2013) provides an overview of the methodologies alternative to SIMs, the major methodology used in his work. This section draws upon Thompson's findings whilst justifying the use of the methodologies used in this thesis at the expense of those that were excluded from analysis.

Gut feeling, as detailed earlier, was very commonplace in retail organisations until the late 1980s (Simkin, 1989) and relied upon a member of staff (generally senior) using their experience to make a judgement on the sales potential of a new or existing site based on a site visit. Davies (1977) highlight that this has been known to be successful as experienced individuals often have a very good instinctive judgement on the potential of a site. However, it is in this reliance on individual staff members that Clarke (1998) finds criticism in this approach due to it being highly subjective and thus lacking in set rules and guidelines that can be followed when deciding on a site. Moreover, the loss of experienced staff members can prove very problematic if an overreliance on this method has been commonplace.

Thompson (2013) highlights the expense and logistical difficulty incurred by visiting all possible sites, particularly when considering large scale planned expansions of a retailers portfolio of stores. This is particularly relevant to the growth of major grocery retailers into the convenience market which has seen a rapid expansion in the total number of stores operated by retailers, particularly Tesco and Sainsbury's whose location planning teams will be required to maintain analytics on a growing portfolio of stores. However, Thompson (2013) partially counters this argument in suggesting that site visits are still required in conjunction with model estimates once sites potential sites have been chosen using more complex methodologies, although this is significantly lessened through the narrowing down of potential sites to statistically viable locations. Both Clarke (1998) and Eyre (1999) highlight the increasing complexity of UK retailing that emerged throughout the 1980s and 1990s making it more difficult to make simple predictions. This complexity continued throughout the 2000s and continues in the 2010s and it could be argued that convenience retailing falls on the more complex end of the grocery market with stores appearing in different locations from bustling city centres to rural villages alongside stores serving vastly different customer missions from single weekly grocery shops to multi-purpose top-up shopping trips as discussed in the review of the literature on the demand for convenience grocery retailing in chapter 2.

Analogues techniques were also prevalent throughout the 1980s and 1990s and continue to be so today. This methodology relies on drawing comparisons (analogies) between a potential new site and existing sites already operated by a retailer. The characteristics on which comparisons may differ by industry but generally focus on physical, location and trade area conditions (Reynolds and Wood, 2010). Whilst this techniques increased the knowledge of store performance across retailers' networks, Eyre (1999) highlights the wide variation in performance seen by outlets in a retail chain across similar markets which make this technique problematic. Moreover, Thompson (2013) states the improbable task of finding stores with similar characteristics across a retailers network as highlighted by Ghosh and McLafferty (1987). Whilst this is certainly easier for very large retailers, one of the reasons this method remains alive and well in large grocery retailers such as Sainsbury's, this is highly problematic in terms of convenience grocery retailing in which stores are found in a variety of locations serving a variety of purposes for the consumer, making finding comparison stores problematic.

Agent-based methodologies are not widely used in the retail industry and are indeed scarce in the academic literature. It is a 'bottom-up' approach which involves *"…creating artificial agents designed to mimic attributes and behaviours of their real-world counterparts"* (Thompson, 2013, Page 223). One of the major advantages in this type of model is in the overcoming the allocation of homogenous behaviours to groups of people bound up in a zonal geography as found in top-down modelling approaches. In the context of convenience grocery retailing this would allow two different people within the same zone (e.g. output area) to adopt separate propensities to travel to grocery stores rather than the single parameter used in a SIM approach to site location analysis. A major downside of this method is the computational expense incurred when creating enough agents to simulate the market effectively (Axelrod, 1997). Additionally, capturing the mechanism underlying agents' behaviour can prove problematic (Twomey and Cadman, 2002) and acquiring data for validation purposes can be a difficult task for a researcher or practitioner meaning the models can be over complex which can detract from understanding the components at the heart of a complex system (Malleson, 2010).

The adoptions of the SIM and multiple regression methodologies in this thesis are designed to overcome some of the issues raised in this section in the context of the application of store location analysis to convenience grocery retailing. Firstly, both methodologies are designed to overcome the subjectivity that is problematic when

using the gut feeling approach discussed in this chapter. They are both objective approaches that result in the quantifying of set rules in the forms of model parameters in an applied SIM and coefficients in multiple regression modelling.

The adoption of a multiple regression methodology is designed as not to discount the history and precedence of analogue approaches to store sales but to sophisticate the approach and attempt to avoid some of the pitfalls of it. The advantage in using model coefficients of a multiple regression model lie in the quantification of the effect of one unit of each variable on store sales which reduces the need to find stores with a similar volume of a given variable. For example, it enables the quantification of the effect of residential population on sales across all stores which can in turn be used to predict performance of a potential site with a large residential population in its catchment area and also predict performance of a potential site with a small residential population in its catchment area. Additionally, the regression methodology teases out the effect of each variable when accounting for all other variables within the model, thus allowing a more nuanced analysis of the effect of each individual variables on sales than is possible using a purely analogue method. Furthermore, in segmenting the network of convenience stores as seen in chapter 6 and applying the clustering results to the multiple regression model in chapter 9, the fundamentals of the analogue methodology have been somewhat adhered to when using a multiple regression methodology.

In using a SIM methodology over an agent based methodology, this study does not discount some of the valid criticisms of the homogenous behaviour built into any aggregate model. By using as small a zonal geography as is possible within the confines of the research, the applied SIM presented in chapter 8 attempts to overcome some commentators concerns about applying homogenous behaviour to consumers with potentially heterogeneous real world behaviour. Using small-area geography based on characteristics of the population derived from the decadal UK census (i.e output area geography) groups households with similar characteristics into zonal geographies in which behaviours such as grocery spending can be measured and validated. This creates a high degree of accuracy whilst maintaining the ease of validation as afforded by census geography based surveys of the population such as the annual living costs and food survey (LCFS).

3.6 Summary

The third major aim of this thesis is to develop and test a series of predictive models for forecasting convenience grocery store revenue in the varying location types in which

this type of grocery store is found. This chapter has reviewed the available methods for forecasting convenience grocery store revenue. In doing so, section 3.1 discussed the use of site location models in the retail industry, charting the emergence of the suite of methods now used by the major grocery retailers in GB. GIS buffer and overlay, regression modelling and spatial interaction modelling were identified as methods that were; theoretically plausible in terms of an application to the branded convenience grocery market, have a potential to be used by location planning teams to predict convenience grocery store revenue and have academic applications that can be adapted to the forecasting of convenience grocery stores.

The theory and development of these methods were discussed further in the sections 3.2 to 3.4 and prior application of these methods to the grocery market were explored. It was hypothesised that these methods could be used to forecast sales for different locations in which convenience grocery stores are found and anticipated that different locations may have a different optimum methodology for accurately predicting store sales. The data required to develop and test these model's is discussed in the next chapter, a review of the data and study regions used in the analysis of this thesis.

Chapter 4 - Data and Study Regions

Chapter 1 set out a number of aims and objectives of this research. Chapters 2 and 3 achieved the aim of reviewing the existing academic and industry literature on the convenience grocery market in Great Britain in relation to; the growth of major retailers into the convenience grocery market, the growing demand for convenience groceries and the attempts at forecasting revenue of both convenience grocery stores and grocery stores more generally both in academia and in the retail industry. The two remaining aims of this thesis are:

1. To quantify the extent to which major retailers have committed to the convenience grocery market and assess the geographical extent to which they play a role in convenience grocery retailing in Great Britain.
2. To develop and test a series of predictive models for forecasting convenience grocery store revenue in the varying location types in which convenience grocery stores are found.

This chapter introduces the data, study area and scale of analysis used in this research to meet the two aims described above. In doing so, the data used in each piece of analysis are discussed with reference to their method of application, their acquisition and reliability and the limitations of the available data for studying the supply of, and demand for, convenience grocery retailing in GB. The chapter is structured as follows.

Section 4.1 identifies the study areas used for each piece of analysis identifying the scales from the national to the sub-national level. Section 4.2 discusses the geographical scale at which each piece of analysis is conducted. Next, section 4.3 identifies the data provided by GMAP Ltd. and Sainsbury's on the supply side of grocery retailing in GB. This section also details the sales data provided by Sainsbury's for use in the revenue estimations in this thesis. Section 4.4 explores the consumer loyalty card dataset provided by Sainsbury's for use in this research project. Finally, section 4.5 identifies the demand data used in each piece of analysis in this research, looking at census data, survey data, geodemographic classifications and transport infrastructure data.

4.1 Study Areas

Analysis in this work is conducted at a number of scales, from the national level to the small area level. The selected case study area differs depending on the requirements of each piece of analysis within the research. The work in part one of chapter 5, identifying the changing mix of stores by size operated by the major grocery retailers between 2003 and 2012, uses Great Britain (GB) as the unit of analysis. Whilst the UK grocery market is referred to in the summary of other work on this subject, this analysis was conducted for GB as the store location datasets discussed later in this section were less complete for Northern Ireland and thus stores in this area were removed from the data. Furthermore, the grocery market in Northern Ireland is sufficiently different, particularly in terms of the mix of retailers, to feel comfortable in excluding the area to focus solely on GB. Figure 4.1 shows a map of GB and its constituent countries.

Figure 4.1 Countries of Great Britain

Great Britain is divided into 11 former Government Office Regions (GOR) as shown in Figure 4.2. GORs are the highest level of sub-national division in Great Britain and have a population of between 2.5 and 8 million people. The majority of analysis presented in this thesis is based on Yorkshire and the Humber in Northern England, a region estimated to have a population of 5,283,700 in the 2011 UK Census of the Population.

Figure 4.2 Former Government Office Regions of Great Britain

Yorkshire and Humber was chosen as an ideal primary region of study. Sainsbury's, one of the industry partners of this project through which sales data was obtained, have an established convenience store network in this region, comprising stores in a variety of locations. At the outset this research anticipated that a variety of differing modelling challenges would be presented by stores in different types of location (e.g. city centre vs rural pitch). Yorkshire and Humber presented an ideal opportunity to investigate this as the region is diverse and encompasses large urban areas such as Sheffield and Leeds, the 2nd and 3rd largest English districts by population along with the NUTS2 region of North Yorkshire, in which the majority of North York Moors national park lies, making the area a mixture of both urban and rural living. Finally, the author (and supervising academics) has extensive knowledge of the region and its retail geographies. Additionally, the use of the North West as an additional region provides a testbed for validation as it has a similar mix of grocery retailers, population and urban/rurality as Yorkshire and the Humber.

The segmentation of the convenience network was not based on sales data but was conducted for store locations in Yorkshire and the Humber to set up further analysis of store sales data in the region. A database of convenience stores in Yorkshire and the Humber was provided by GMAP Ltd. and validated using Sainsbury's store locations database. Thereafter, Sainsbury's provided sales data for a sample of their convenience stores, a total of 95 stores, in Yorkshire and the Humber allowing for the testing of a number of methods of forecasting store revenue. Both the GIS buffer and overlay model in chapter 7 and the regression model in chapter 9 use the whole of Yorkshire and the Humber as a study region and attempt to accurately predict the store revenue of the whole sample of convenience stores discussed above. The spatial interaction model presented in chapter 8 uses part of Yorkshire and the Humber as its study region. This is due to the availability of nectar card data for calibration and will be discussed in more detail in chapter 8. In order to validate the predictive capacity of the GIS buffer and regression model, Sainsbury's provided store revenue data for a sample of their convenience stores in the North West region. Thus, with the exception of the spatial interaction model, the validation presented in this thesis uses the North West as its study region. The North West is an ideal testbed for validation as it has a similar mix of retailers, population demographics and urban/rurality as Yorkshire and the Humber. In validating the spatial interaction model, the other portion of Yorkshire and the Humber not used for calibration in chapter 8 is used to validate the ability of the model to predict store revenue.

4.2 Geographical units of analysis

The previous section introduced the study areas used for each substantial analysis chapter in this thesis. This section looks at the geographical scale within the study region used for each piece of analysis. It is vitally important to conduct research at appropriate geographical scales to maximise the accuracy and utility of any output produced. Thompson (2013) highlights the hierarchical nature of geography in GB and the fact that they are built on a number of systems including census, electoral, environmental, postal and historical boundaries. When disaggregating into subnational geography, this research uses a mix of postal and census geographies, the former to give an identifiable unit at which to show the market share of a number of prominent convenience retailers and the latter to develop a comprehensive picture of grocery demand at as fine a geographical level as possible at which census data is released in Great Britain.

4.2.1 Great Britain

Section 4.1 introduced the analysis presented in the first part of chapter 5 analysing the change in stores operated by the major grocery retailers in Great Britain by store size from 2003-2012. This is a global analysis attempting to quantify the extent to which convenience stores have become a more prominent feature of the store networks operated by large retailers, in particular Tesco and Sainsbury's in the grocery market in Great Britain. Great Britain is taken as a single geographic unit for this analysis. However, other analysis is conducted using subnational divisions in this thesis.

4.2.3 Former Government Regions

As described in section 4.1, the Yorkshire and the Humber and North West regions form the study areas for the modelling in chapters 7, 8 and 9 of this thesis. However, all 11 former government office regions are used as geographical units in the second part of analysis in chapter 5 identifying the regional growth of branded convenience grocery stores in Great Britain.

4.2.4 Postcode Areas

Market share analysis investigating the geographical extent of each major convenience grocery market retailer's convenience store network is presented in chapter 5 at the postcode area level in Great Britain. This is a national level study in Great Britain but disaggregates the network of each retailer's convenience operations by this smaller geography. There are 121 postcode areas in the UK that are defined by the Royal Mail and used for the purpose of administering the delivery of mail. As Northern Ireland was

removed from the analysis in this thesis, the removal of the Belfast (BT) postcode area left 120 postcode area units in Great Britain. Figure 4.3 is a postal area map in which postal area codes are labelled. The corresponding glossary of postal area codes can be found in the appendix.

Figure 4.3 Postal areas of England and Scotland
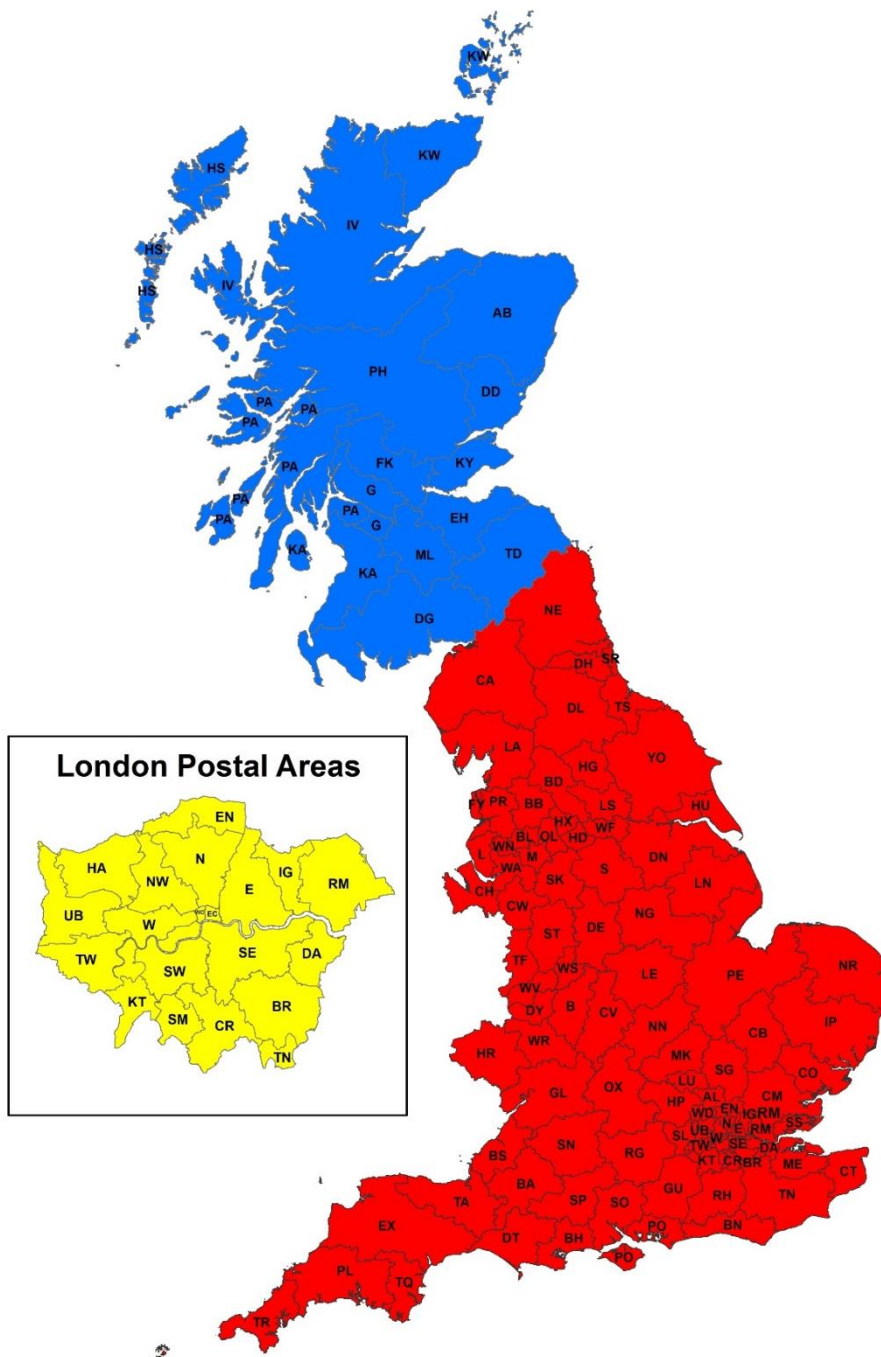
Postal areas are comprised of one or two letter codes identifying a place that is served by a Royal Mail sorting office. For example, the SW postcode synonymous with the Wimbledon Tennis Championships is a mnemonic for South West London and the LS postcode is a mnemonic for the postal area of Leeds. This geography was selected for market share analysis as it provided finer detail than the former government office regions scale whilst allowing for identifiable place names at which persons without extensive knowledge of geographical hierarchies could understand the geographical extent of the store network of each major convenience grocery retailer.

4.2.5 Census Geography

The UK census of the population is undertaken every ten years and collects data on a number of demographic and socio-economic variables. This research uses data from both the 2001 and 2011 censuses of the UK population. The 2011 census is used because it is the most recent UK census and therefore the most up to data and the 2001 census is used in specific pieces of analysis due to limitations on available data for the 2011 census. The major geographical units directly related to census data are the output area (OA) and the super output area (SOA). All census statistical geographies have unique nine letter codes which match digital boundary data allowing for comprehensive GIS analysis of census variables across the UK.

Output Areas (OAs)

First introduced in Scotland in 1981, and the rest of the UK in 2001, output areas (OA) were created specifically for the publication and output of census data and are the smallest geography for which census data is available in the UK. They were generated by aggregating adjacent unit postcodes into contiguous clusters. OAs are designed to be socially homogenous, of similar population size, of approximately regular shape and void or urban/rural where possible (i.e. comprised entirely of rural or urban postcodes). The average population in an OA was 309 in 2011, having risen from 297 in 2001. Census data can be generated for the majority of geographical scales by fitting OAs to higher geographies such as the postal area geography discussed earlier in this chapter.

Census data at OA level is used to derive socio-economic and demographic characteristics of defined catchment areas of convenience stores in Yorkshire and the Humber and the North West of England. This data is used in a number of chapters of this thesis. Firstly, the network segmentation in chapter 6 uses census variables at OA level aggregated up to catchment areas around each convenience store to attempt to

find distinct statistical location types in which convenience grocery retailing is taking place.

Both the GIS buffer and overlay model in chapter 7 and the spatial interaction model in chapter 8 require an accurate residential grocery demand layer that is then allocated to competing grocery store destinations. The GIS buffer and overlay model does this by aggregating available grocery expenditure at OA level in Yorkshire and the Humber to the catchment area of each convenience grocery store operated by Sainsbury's in Yorkshire and the Humber. This is done by combining census data on the number of households in each output area with survey data from the Living Costs and Food Survey (discussed later in this chapter) to generate estimates of available residential grocery expenditure. This process is recreated for the North West region in the model validations.

The regression model in chapter 9 required the selection of predictor variables which would be explored for the extent to which they account for the variation in store revenues experienced by Sainsbury's convenience grocery stores. Data for the socio-economic and demographic variables chosen were collected at OA level and aggregated to various catchment area sizes around each Sainsbury's convenience store in Yorkshire and the Humber and repeated for the North West in the validations in chapter 9.

Super Output Areas (SOAs)

Super output areas are derived by grouping a number of output areas into a larger, contiguous cluster. They give an extra hierarchy of census geography through which area level demographic and socio-economic data can be derived at a larger scale than OAs. Initially released for England and Wales in 2004, they come in two forms; Lower layer super output areas (LSOA) containing between 1000 and 3000 people and middle layer super output areas (MSOA) with a population of between 5000 and 15000 people. Table 4.1 identifies the population and household thresholds that LSOAs and MSOAs meet.

Table 4.1 Population and household thresholds of census geographies (Source: ONS, n.d.)

Geography	Min. Pop	Max. Pop	Min. Households	Max. Households
LSOA	1,000	3,000	400	1,200
MSOA	5,000	15,000	2,000	6,000

This research does not use data at the MSOA level. However, the demand layer in the spatial interaction model in chapter 8 is constructed using LSOA geography as its unit of analysis. Residential expenditure estimates at the OA level are aggregated to the LSOA level for input into the model. This is in part due to computational constraints of the model along with the use of an existing LSOA model by Sainsbury's allowing for comparisons to be drawn. Moreover, if the model was to be extended to encompass the whole country, the computational requirements of an OA level model would be very large.

Workplace Zones (WPZs)

The OA is an ideal unit for analysing residential demographic patterns in the UK as it was developed based on the spatial distribution of homogenous residential populations. However, they are not ideally suited to investigating populations that do not follow the same distribution as residential populations (Berry et al. 2016). In a UK context, one such population is the work based population. This creates two distinct issues when applying residential OAs to the work based population as was attempted as part of the 2001 census of the population. Firstly, the small numbers of workers in certain OAs resulted in a failure to meet statistical disclosure thresholds required for the release of workplace statistics at this fine level of residential geography (Martin, Cockings, and Harfoot, 2013; Mitchell, 2014). Secondly, OAs, mainly consistent of commercial and industrial land uses, often had to be very large in order to contain the threshold for statistical disclosure of the residential population, a minimum of 40 households or 100 people (ONS, n.d.). This loses the fine spatial granularity that is the advantage of the OA when looking at residential population distributions.

The Office for National Statistics addressed these issues and released a new form of census output geography based on census respondents' place of work called 'Workplace Zones' (WPZs) in 2013. They were created using the spatial extent of OAs and aggregated or disaggregated to match this geography perfectly. Some OAs remained the same and are therefore both an OA and a WPZ. They were developed to contain a similar number of workers in each and to include workers employed in the same industry or sector of employment (e.g. Retail) (Mitchell, 2014). Just as the OA geography was designed so that individual residents could not be identified, WPZs were designed to contain a minimum of 200 workers and at least three unit postcodes to maintain individual worker anonymity (Mitchell, 2014). In central locations containing

a number of workers, this creates a far more nuanced picture of population distribution. A comparison of OAs and WPZs in central Leeds, the largest city in Yorkshire and the Humber can be seen in figure 4.4. The two central OAs in figure 4.4 are comparatively large due to their commercial retail and office based land use. Thus, in any application demanding a fine spatial understanding of the location of workers, WPZs perform much better in understanding the location of workplace populations. WPZ geography is used in the analysis in this thesis.

Figure 4.4 Comparison of OAs and WZs in central Leed
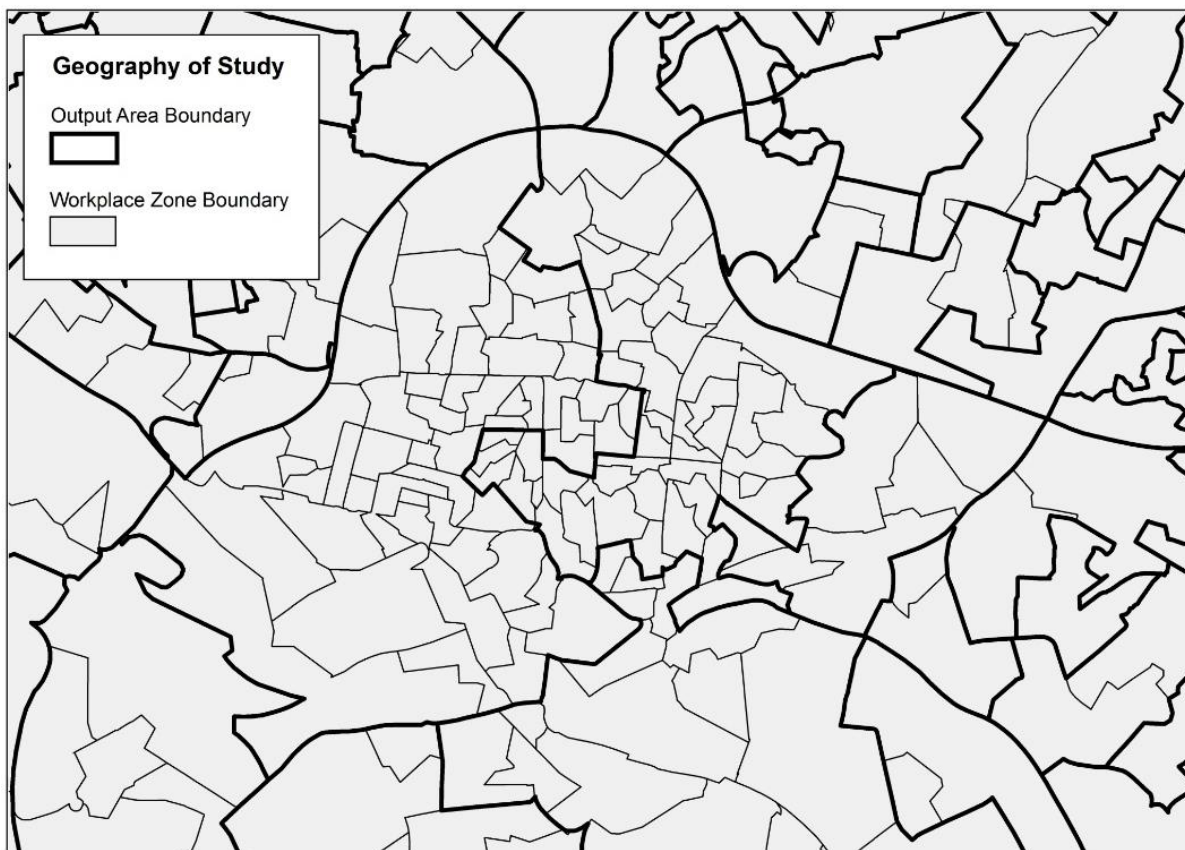
The workplace demand input layer in the GIS buffer and overlay model uses the population of WPZs to derive estimates on work based grocery expenditure aggregated to the catchment area of convenience grocery stores in Yorkshire and the Humber, a procedure repeated for the North West study region for model validation. Moreover, the applied SIM in chapter 8 has two input layers, an LSOA residential demand layer and a WPZ based workplace demand layer. Finally, the linear regression model in chapter 9 uses data at a WPZ level aggregated to buffered catchment areas of convenience grocery stores in Yorkshire and the Humber as free predictors in the model. The statistics used are the number of workers in each WPZ employed in the different NS-

Sec socio-economic occupation classifications along with total WPZ population within the catchment area of each store, a procedure repeated for the North West study region in chapter 10.

4.3 Retailer Data

Store location data was obtained from both GMAP Ltd. (a subsidiary of Callcredit Information Group) and Sainsbury's. GMAP are a leading provider of market intelligence, retail planning and predictive modelling solutions for major retail organisations and purchase location data from market research organisations such as Panorama and Retail Locations Ltd. (Thompson, 2013). Sainsbury's employ an in house team to track competitor's activities and keep a record of Sainsbury's own changes, maintaining an up to date database of grocery retailers in the UK.

These two databases were combined to create an accurate picture of the supply of stores in the grocery market in the UK. Moreover, groundtruthing was conducted by the author visiting various locations in Sheffield and Leeds to see if the retail mix matched the joined GMAP/Sainsbury's database. Reassuringly, the dataset was found to be highly accurate. This work uses the same definition of grocery retailers as Thompson (2013), defining grocery stores as those selling food, non-edible groceries and varying ranges of non-food products. These datasets contain a number of variables relating to each store in the UK. The store data available for Sainsbury's and competitor retailers are listed in table 4.2.

Table 4.2 Combined store location data

Variable	Retailer	
	Sainsbury's	**Competitors**
Fascia	Type of store (convenience or supermarket).	Retailer and type of store where applicable (e.g. Tesco Express, Tesco Metro).
Branch Number	Sainsbury's identifier code for each of their own store.	Sainsbury's identifier number for each competitor store in there database.
Branch Name	Location of each store, identified by a street name and/or town	n/a
Postcode	Store postcodes	Store postcodes
Sales area (Sq. Ft.)	Trading area of store by floorspace in square feet.	Trading area of store by floorspace in square feet.
XY co-ordinates	The XY co-ordinates of all JS store locations.	The XY co-ordinates of all competitor store locations

This dataset included convenience grocery stores operated by all of the major convenience grocery players in GB such as Tesco, Sainsbury's, the Co-operative Group and the symbol groups alongside all grocery stores (small and large) operated by retailers more associated with larger supermarket retailing but less integrated into the convenience grocery market in the UK. These are retailers such as ASDA and Morrisons. Store location data was used in each analysis chapter of this thesis.

The time series data was used to assess the changes in the supply of available grocery stores in GB both nationally and by region in the analysis in chapter 5. This simply used the XY coordinates of each store along with the store size aggregated to GB and each former government region to analyse the change in grocery store availability by size between 2003 and 2012. Moreover, the market share analysis in chapter 5 used the convenience stores within the store database to assess the relative strength of each retailer in the convenience grocery market in each postal area of GB. The location of the convenience grocery stores in the database of stores was used to develop the typology of convenience grocery stores in the Yorkshire and the Humber study region found in chapter 6 of this thesis.

Additionally, store locations were used to build the GIS buffer and overlay model in chapter 7, the applied SIM in chapter 8 and the regression model in chapter 9. Sainsbury's competitor data was used in the GIS buffer and overlay model to divide available workplace and residential grocery expenditure, to build the supply layer in the applied SIM presented in chapter 8 and as a predictor variable in the regression model in chapter 9. The location of Sainsbury' stores combined with known revenue data (discussed later in this chapter) provided by the retailer was used to calibrate each model and assess each model's ability to forecast store revenue. Moreover, location of other Sainsbury's stores in the vicinity of store locations was used as a predictor variable in the regression model in chapter 9 and to divide available grocery expenditure in the GIS buffer and overlay model in chapter 7. The analysis listed in this paragraph was repeated for each model for the North West study region to validate the Yorkshire and Humber analysis. In each model, store size, as detailed in the combined database of stores, was used as a proxy for the attractiveness of each retail store.

4.3.1 Revenue Data

Sainsbury's provided mean weekly sales for 2013 for 95 of their convenience grocery stores in Yorkshire and the Humber. Mean weekly sales was not a variable used in the segmentation of the market in chapter 6 but was used to analyse Sainsbury's store

revenue data in the different location types identified in chapter 6. The main use of revenue data was for the calibrating and testing of the accuracy of the GIS buffer and overlay model in chapter 7, spatial interaction model in chapter 8 and the regression model in chapter 9. In addition to the 95 store revenues provided for stores in Yorkshire and the Humber, Sainsbury's provided an additional mean weekly revenue for 2013 for an additional 31 stores in North West England. This was used to test the effectiveness of the GIS buffer and overlay model and regression model outside the initial Yorkshire and the Humber study area in the model validations.

4.3.2 Retail stores (non-grocery)

In a separate database to the grocery store database, GMAP Ltd. provided data on retail stores in the UK ranging from fashion stores to electronic stores. This data came with two pieces of information, retailer name and store location co-ordinates. This data was geocoded in the same way as the grocery retailer data and used in a number of pieces of analysis presented in this thesis. It was used as an input variable in the segmentation of convenience grocery locations in chapter 6. Total non-grocery stores were aggregated to 1 kilometre catchment areas around each convenience grocery store location and used as an adjacency variably to identify vibrant retail locations in which people may be attracted to the retail mix available and thus is a prime central store location for convenience grocery retailing.

Data on non-grocery retailers was also used as a predictor variable in the regression model in chapter 9 to look at the relationship between the number of available stores in an area and the store revenues of convenience stores. It is anticipated that the availability of other retailer stores may be a significant pull factor in attracting potential customers to a store location and thus a greater density of retailers will result in greater store sales. Moreover, this data was also used in model validation of the regression model.

4.4 Consumer Data

Retailers have increasingly adopted loyalty card schemes as a method of collecting information on customer purchasing habits. This is true both in the grocery industry and in other forms of retailing. One such scheme is the Nectar card operated by Sainsbury's which can be used in all of the retailers' shopping channels; Supermarket, convenience and online. As part of this research project, Sainsbury's provided Nectar card data for Yorkshire and the Humber.

4.4.1 Nectar Card Data

The Nectar card data contains geocoded transaction data for the 95 Sainsbury's convenience stores in Yorkshire and the Humber for which mean weekly revenue was provided by the retailer. This complements the revenue data and is all sales to these stores for a 12 week period in 2013 starting on 1st September. This period avoids any significant holidays and is therefore thought of as typical spending time for each store. The data links customer postcodes from anywhere in the UK based on the address registered when a customer applies for a Nectar card to the Sainsbury's branch in Yorkshire and the Humber in which they purchased groceries. The data includes the number of transactions made in a given time period (in this case 12 weeks), the total value of transactions and the value spent on each type of product. A typical database is shown in table 4.3.

Table 4.3 Example Nectar card entry

Postcode	Branch	Week	Sales Value (£)	Transactions
LS6 4AJ	4206	2013-35	2.50	2
In this example, the customers postcode is LS6 4AJ (a postcode in North West Leeds), the purchases were made in branch 4206 in the 35th week of 2013/14 financial year and the value of sales totalled £2.50 spread over 2 transactions.				

In the context of this research, the primary use of this data was in the applied SIM presented in chapter 8. The nectar card transaction data was used to help calibrate the model by examining average trip distance of consumers in the model with the real world behavioural patterns among consumers at the OA level in Yorkshire and the Humber.

4.5 Demand Data

Data on demand is a necessity when analysing consumer behaviour in relation to convenience grocery stores. This data is used for a number of purposes; to compute variables in the network segmentation in chapter 6, to calculate available expenditure in the GIS buffer and overlay model in chapter 7 and the spatial interaction model in chapter 8 and as explanatory variables in the regression analysis in chapter 9.

4.5.1 Census Data

The principal and most comprehensive source of geodemographic data on the population in the UK is the UK Census of Population. Administered every 10 years, it is a legal requirement for every resident in the UK to fill in the form giving an almost 100% sample of the population, unrivalled in any other survey of the UK population. In reality, a representative sample of the whole population is created in each census as those small numbers of persons not captured in the census are imputed so that a 100% sample is produced (Rees et al. 2002). Geographically, census data is coded to both the OA and LSOA levels discussed earlier in this chapter. The census data in this research comes from a combination of the 2001 and 2011 censuses of the population. This is because it was not possible to synthesise some of the 2011 census data with other data sources, notably the survey data disaggregating household spending by OAC group in the Living Costs and Food Survey.

The 2001 and 2011 censuses of the population was used in this research to derive variables in a number of analysis chapters. The 2011 census was used to acquire a number of the variables adopted in the network segmentation in chapter 6 used to identify a number of distinct location types in which convenience grocery retailing takes place. These variables were residential population, daytime population and social class. The use of these variables is discussed in more detail in chapter 6. Secondly, the 2011 census was used to provide a count of households at the OA level from which an estimate of available household expenditure on groceries was computed for the whole of Yorkshire and the Humber. This provides the base demand layer for the buffer model in chapter 7 and is aggregated to the LSOA level for use in the applied spatial interaction model in chapter 8. Additionally, a number of variables from the census were used as explanatory variables in the regression model in chapter 9. These variables were:

1. Residential population demographics
   - Volume and density
   - Age
   - Mobility
   - Relative deprivation
   - Education
2. Work based population demographics
   - Volume and density

- Socio-economic status

4.5.2 The Output Area Classification (OAC)

The Office for National Statistics (ONS) 2001 Output Area Classification (OAC) groups geographic areas according to key characteristics of the population. These groupings (or clusters) are generated from 2001 census data (Vickers and Rees, 2007). The classification was created using a k-means clustering algorithm, not dissimilar to the segmentation procedure used in the classification of convenience grocery store location types in chapter 6 of this thesis. The OAC involves grouping areas based on 41 variables that can be grouped into the following themes; Population density, age, marital status, ethnic identity, health, employment, industry, occupation, commuting, housing tenure, type of accommodation, car availability, household size, household amenities and households composition (Vickers and Rees, 2007) This results in a comprehensive geodemographic clustering of the whole population of the UK. For geographic analysis, output areas provide a stable unit of analysis at the small area level giving a fine resolution in which to derive area characteristics.

The final product of the OAC is a classification separating OAs demographically to produce 7 'Supergroups', which are broken down into 21 'Groups' and 52 'Subgroups'. Table 4.5 details the 'Supergroups and 'Groups' in the classification. The 'Subgroup' level is at a detail finer than is required in this research but a more detailed description of the clusters can be found in Vickers and Rees (2007).

Table 4.5 Output Area Classification 2001

Supergroup	Supergroup Name	Group	Group Name
1	Blue collar communities	1a	Terraced blue collar
		1b	Younger blue collar
		1c	Older blue collar
2	City Living	2a	Transient communities
		2b	Settled in the city
3	Countryside	3a	Village life
		3b	Agricultural
		3c	Accessible countryside
4	Prospering suburbs	4a	Prospering younger families
		4b	Prospering older families
		4c	Prospering semis

		4d	Thriving suburbs
5	Constrained by Circumstances	5a	Senior communities
		5b	Older workers
		5c	Public housing
6	Typical traits	6a	Settled households
		6b	Least divergent
		6c	Young families in terraced homes
		6d	Aspiring households
7	Multicultural	7a	Asian communities
		7b	Afro-Caribbean communities

The names of the groupings and subgroupings in the OAC nominally infer a sense of the type of areas that they are and the people that reside in them. This geodemographic classification is used in a number of analyses in this research. Furthermore, the output area classification is primarily used as a tool for calculating estimated expenditure on groceries at the OA level which is used as the base demand layer in the GIS buffer and overlay model for predicting store revenue in chapter 7 and the spatial interaction model for predicting store revenue in chapter 8. This is done by combining the number of households in each OA (from the UK Census) with average household grocery spending by OAC group shown in table 4.5 derived from the family expenditure section of the Living Costs and Food Survey (LCFS) discussed in the next section.

At the outset of this research, the 2011 OAC classification was yet to be released. Although the classification has subsequently been released, there is yet to be an updated version of the LCFS in which grocery spending is disaggregated by the 2011 OAC. Moreover, it is important for the demand layer in a model predicting store revenue to match the time period in which that level of revenue was generated by a retailer. Therefore it was important in this case to understand retail demand at the time of the sales data (2013) which was best achieved by using the disaggregation of grocery spend by OAC 2011 available in the Living Costs and Food Survey Family Spending Report (2014) reporting on household spending for 2013.

4.5.3 Household Expenditure: The Living Costs and Food Survey

The ONS Living Costs and Food Survey (LCFS) was formed through the amalgamation of the Family Expenditure Survey (FES) and the National Food Survey (NFS), both of

which have run since the 1950s and chart changes and patterns in spending and food consumption in Great Britain. It is a household level survey in Great Britain and is administered by the Office for National Statistics in conjunction with the Department for Food and Rural Affairs (DEFRA). The LCFS is carried out on a calendar year basis and collects information on household level spending on a variety of goods, including food and drinks. Annually, approximately 12,000 households are randomly selected from the Royal Mail's Postcode Address file and subsequently interviewed and asked to keep a diary recording individual (and total household) consumption expenditure.

The ONS subsequently analyse and report the results via an annual report entitled 'Family Spending'. The report disaggregates consumer consumption patterns by socio-economic and geodemographic characteristics of the population. The Family Spending Survey of 2014 reported on household spending for the year 2013 (the year for which store revenue data was provided as part of this research) which required a matching demand layer for the same time period. Consumer spending on twelve categories are disaggregated by the 2001 ONS Output Area Classification discussed in the section above. The survey provides information on a number of themes related to spending and consumption at the household level including but not restricted to; income by region, expenditure by region, expenditure by income, expenditure by age, expenditure by socio-economic classification and expenditure by the 2001 OAC.

 Expenditure by 2001 OAC was used to develop the demand layer in the GIS buffer and overlay model in chapter 7 and the spatial interaction model in chapter 8. The 2013 Living Costs and Food Survey used the 2001 OAC to disaggregate household spending as the 2011 OAC was not yet released. Hence, the use of some products of both the 2001 census and the 2011 census. Household expenditure by OAC supergroups is divided into the 12 categories shown in table 4.6

Table 4.6 Household expenditure by OAC supergroup, 2013 (ONS, 2014)

OAC 2001 Group	Food & non-alcoholic drinks	Alcoholic drinks, tobacco & narcotics	Clothing & footwear	Housing, fuel & power	Household goods & services	Health	Transport	Communication	Recreation & culture	Education	Restaurants & hotels	Miscellaneous goods & services
Blue collar Communities	**52.30**	**12.60**	18.70	65.70	22.10	4.20	50.90	13.70	51.00	2.10	28.30	26.10
City living	**52.80**	**10.40**	20.50	125.50	40.10	6.70	68.00	15.90	54.30	29.00	56.10	42.90
Countryside	**71.40**	**14.10**	27.80	75.00	47.20	9.50	91.30	15.30	81.20	10.00	47.50	49.20
Prospering suburbs	**67.70**	**11.90**	27.70	60.00	42.70	9.00	94.70	15.10	89.50	6.20	50.10	52.50
Constrained by circumstances	**45.90**	**12.30**	13.90	64.70	22.60	3.40	41.20	11.40	43.00	1.60	23.60	22.60
Typical traits	**55.90**	**11.70**	23.20	73.00	30.90	5.80	72.90	14.80	61.00	9.40	42.10	40.00
Multicultural	**57.30**	**10.10**	21.00	97.80	23.80	3.10	52.40	15.00	43.00	17.00	35.00	32.60
Mean Household Spend	**58.80**	**12.00**	22.60	74.40	33.10	6.20	70.40	14.50	63.90	8.80	40.40	39.10

In studying consumer expenditure available the first two categories of spending are used:

1. Food & non-alcoholic drinks
2. Alcoholic drinks, tobacco & narcotics

These are the categories of spending that are directly associated with convenience grocery retailing. The LCFS further disaggregates spending on these two groups by the 2001 OAC groups, the level below supergroup. It is then possible to multiply the average spending by OAC household in each group by the number of households in each OA to give estimated grocery expenditure for each OA in the study. Average household spending on these products by the 2001 OAC can be seen in table 4.7. Whilst other non-grocery goods are sold in convenience grocery stores (such as domestic cleaning products), this makes up a relatively small proportion of sales, thus the focus on definitive grocery related products.

Table 4.7 Mean household grocery spend on groceries by OAC, 2013 (ONS, 2014)

OAC Code	OAC Description	Mean Household Spend		
		Food & non-alcoholic drinks	Alcohol, tobacco & narcotics	Total Spend
1A	Terraced blue collar	48.90	12.30	61.20
1B	Younger blue collar	50.00	12.20	62.20
1C	Older blue collar	51.40	12.30	63.70
2A	Transient communities	42.50	8.80	51.30
2B	Settled in the city	50.10	12.30	62.40
3A	Village life	54.80	12.60	67.40
3B	Agricultural	57.00	10.50	67.50
3C	Accessible countryside	58.80	13.50	72.30
4A	Prospering younger families	62.00	10.20	72.20
4B	Prospering older families	61.20	11.20	72.40
4C	Prospering semis	57.30	10.20	67.50
4D	Thriving suburbs	61.20	11.50	72.70
5A	Senior communities	33.60	7.00	40.60
5B	Older workers	42.40	11.40	53.80
5C	Public housing	43.20	12.50	55.70
6A	Settled households	55.10	11.20	66.30
6B	Least divergent	51.40	10.20	61.60
6C	Young families in terraced homes	45.60	12.10	57.70
6D	Aspiring households	54.00	10.90	64.90
7A	Asian communities	49.70	9.60	59.30
7B	Afro-Caribbean communities	48.10	9.20	57.30

4.5.4 Building a residential demand layer

The total expenditure on food, non-alcoholic drinks, alcoholic drinks, tobacco and narcotics in each output area can be calculated by combining census data on the number of households and the mean household spend data shown in table 4.7, calculated using the formula:

$$A \times B = C$$

Where A is the combined average household expenditure on food, non-alcoholic drinks, alcoholic drinks, narcotics and tobacco by census output area classification; B is an ONS mid-year estimate of the number of households by output area and C is the total expenditure on the grocery products by output area. This results in the creation of

a demand layer that can be plugged into a GIS and attached to a digital boundary file as shown in figure 4.5, a map of residential grocery demand by OA in West Yorkshire.

Figure 4.5 Total residential grocery expenditure by output area in West Yorkshire, 2013.
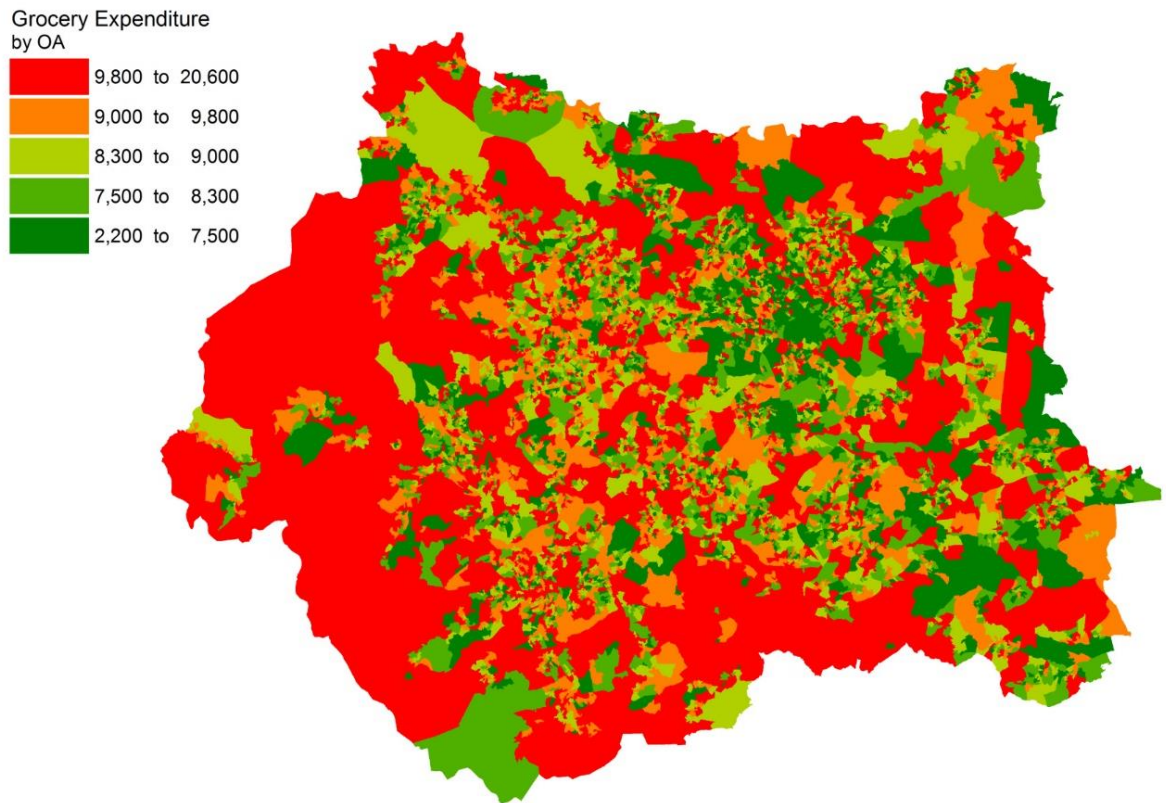
This residential demand layer was used to allocate grocery expendtiure to available grocery stores in both the GIS buffer and overlay model in chapter 7 and the spatial interaction model in chapter 8. Moreover, the process for estimating residential grocery expenditure in Yorkshire and the Humber was repeated for the North West and used in the GIS buffer and overlay and spatial interaction model validations.

4.5.6 Building a work based grocery demand layer

The previous section detailed the development of a residential demand layer at the OA level across the UK. However, this method does not capture work based demand for groceries. It is possible to download work based population statistics for OAs from the 2011 census in the UK; however, this is fraught with issues of large spatial units for small numbers of workers and disclosure issues (Berry et al. 2016; Martin, Cockings and Harfoot, 2013; Martin, 2013). The WPZ geography discussed earlier in this chapter provides a viable alternative to the OA geography when developing a work based demand layer for use in store revenue forecasting in this research.

Berry et al. (2016) looked at the trading characteristics of a Co-op convenience store trading in central London. They found that the residential census output geographies did not accurately capture drivers of retail demand in central locations in which the work based population significant exceeds the residential population. They found that *"(the)… recent provision of specific output geography for the provision of workplace population statistics is a major enhancement which considerably strengthens the potential for incorporation of workplace populations in retail analysis and decision-making"* (Berry et al. 2016, P.392). They suggest a future research agenda in which this type of geography is incorporated into predictive models to introduce an added dimension to improve location based decision making for retailers (Berry et al. 2016).

In this research a WPZ based demand layer for the workplace population in Yorkshire and the Humber was developed and applied to two of the modelling methodologies presented in this chapter. Whilst comprehensive survey data on residential expenditure on goods available in convenience grocery stores is available through surveys such as the Living Costs and Food Survey, such surveys encompassing work based demand for groceries do not exist. In-house research by Sainsbury's has found, on average, £5 per worker per week is a good estimate of workplace grocery demand when forecasting revenue to their UK based store network. They find that a £5 mean expenditure per worker improves the accuracy of their in house gravity model, and has been extensively verified against sales in larger stores. This is a useful starting point from which to build a work based demand layer.

£5 per week per worker was multiplied by the number of workers in each WPZ in Yorkshire and the Humber and the North West of England and combined with digital boundary files in a GIS to develop two work based demand layers for use in the analysis in chapters 7 and 8. In this GIS buffer and overlay analysis in chapter 7, the WPZ demand layer was aggregated to buffered catchment areas of each convenience grocery store operated by Sainsbury's using point in polygon analysis coded in SQL in MapInfo Professional 12.5. This process was repeated for Sainsbury's North West convenience store network in the validation of the GIS buffer and overlay model. This demand layer was also used in the applied SIM as a separate demand layer to the residential demand layer creating a dual origin matrix. This dual origin layer also featured in the validation of the applied SIM.

Figures 4.6a and 4.6b compare available residential demand by OA level and available work based demand by WPZ in central Leeds, the largest city in Yorkshire and the Humber. If the model inputs used purely contain residential demand estimates by OA,

a crude understanding of potential customers in areas with large non-residential populations is a major potential pitfall. However, introducing available expenditure by WPZ gives a better understanding of the spatial distribution of daytime populations which will enhance the likelihood of accurate store revenue predictions in areas with large non-residential populations

Figures 4.6a and 4.6b Comparing OA demand estimates with WPZ demand estimates in central Leeds

a) OA
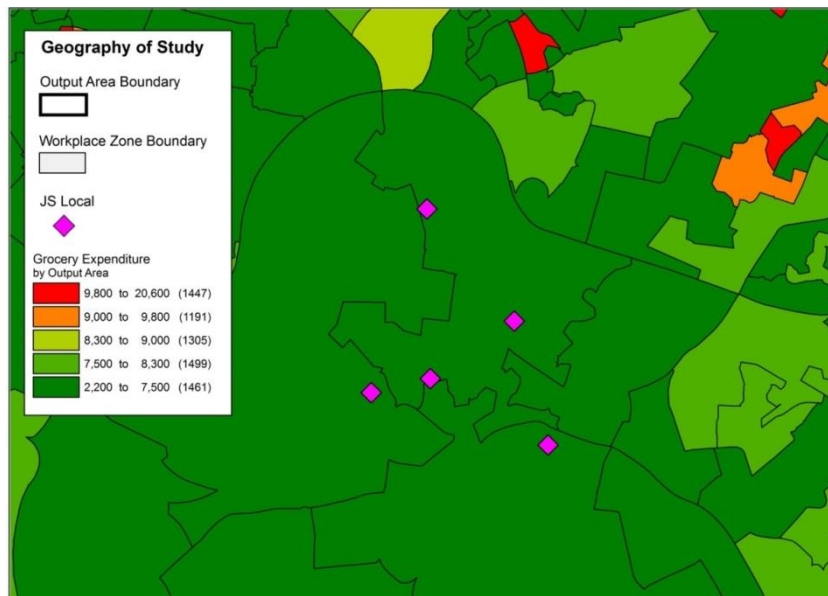
b) WPZ
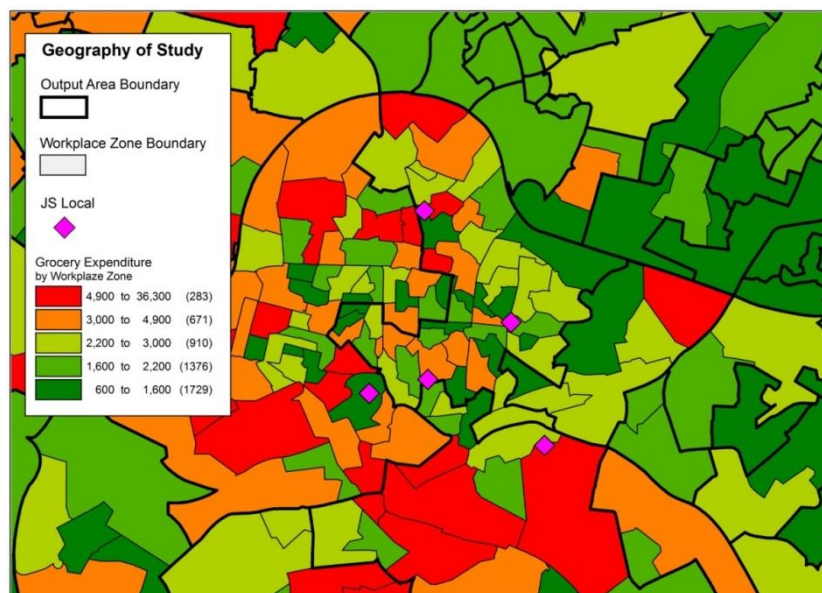
4.5.7 Rail Passenger Data

One further source of data used in this thesis is rail passenger data from the Office for Rail Regulation for 2012/2013. This data contains station usage figures derived from ticket sales and contains entry and exit data for every rail station in the UK which was converted from the postcode of each station to a set of XY co-ordinates for use in the analysis in this thesis. This dataset was used to calculate the total number of train passengers entering or exiting a given catchment area around each convenience grocery store in Yorkshire and the Humber and was used in the network segmentation in chapter 6 and as an explanatory variable in the regression analysis in chapter 9.

4.6 Summary

This chapter has identified the varying study areas, geographies of analysis and datasets used in order to meet the three major aims of this thesis. The major contribution of this chapter to the overall aims of this research is the development of a residential demand layer at the OA layer and a workplace zone (WPZ) demand layer which are used in chapters 7 to 10 in the varying methodologies for predicting store revenue of convenience grocery stores in Yorkshire and the Humber and the North West of England. Later chapters delve deeper in to the collection, manipulation, analysis and presentation of data related to convenience grocery retailing in GB. The proceeding chapters present the main body of analysis conducted as part of this research, starting with the quantification of the importance of different store size formats (including convenience stores) to the overall network of stores operated by the four largest grocery retailers in GB from 2003 to 2012.

Chapter 5
Growth of branded retailers in the convenience grocery market in Great Britain, 2003-2012

Chapter 1 introduced the aims of this thesis. The first aim was to review the existing academic and industry literature on the convenience grocery market in Great Britain in relation to; the growth of major retailers into the convenience grocery market, the growing demand for convenience groceries, and the attempts at forecasting revenue of both convenience grocery stores and grocery stores more generally both in academia and in the retail industry. Chapter 2 identified the conditions by which major retailers came to operate convenience stores which were traditionally the reserve of small and independent neighbourhood grocery retailers. The chapter went on to explore the growing demand for convenience grocery retailing among the population of GB. Moreover chapter 3 reviewed methodologies and existing attempts to date at forecasting convenience grocery store revenue.

The second aim of this thesis is to quantify the extent to which major retailers have committed to the convenience grocery market and assess the geographical extent to which they play a role in convenience grocery retailing in Great Britain. Amid difficulty in growing market share through supermarket growth, Tesco and Sainsbury's were pioneers of this diversification into small store grocery retailing and showed that success in a sector previously the reserve of smaller retailers was possible. Both retailers committed and continue to commit considerable resources to maintain a convenience grocery presence for their brand. This chapter gives an overview of the rise in prominence of small format grocery retailing for the major grocery retailers in Great Britain and explores the spatial growth of the largest grocery retailers playing a role in convenience grocery retailing in Great Britain.

This chapter positions the convenience network of the major grocery retailers in the wider context of convenience grocery retailing across GB. The four largest grocery retailers in GB have the greatest overall market share in the grocery market in GB. However, given that they have all expressed a desire to be more prominent in the convenience market, it is important to assess their presence versus other grocery retailers with a large presence in the convenience grocery market, most notably the co-operative group and a number of prominent symbol group retailers.

The analysis presented in this chapter is reported as follows; Section 5.1 identifies the methodology by which the growing importance of convenience retailing to a number of major grocery retailers was quantified. Secondly, the national growth in convenience store numbers and floorspace for each of the four largest grocery retailers in GB is analysed in section 5.2, grounding this growth in the context of the total supply of grocery stores of all sizes operated by these major retailers. Moreover, convenience store numbers and floorspace by region are investigated in section 5.3 to give a geographical overview of branded convenience grocery retailing and its growth across different parts of Great Britain.

Moreover, Section 5.4 reports on the methodology by which the market share of the prominent convenience grocery retailers in GB in 2012 was measured. Thereafter section 5.5 looks at the contribution of branded convenience grocery retailing to the overall share of the grocery market at the postal area level across GB. Next, section 5.6 analyses the contribution of branded convenience grocery stores operated by a number of prominent convenience grocery retailers to the overall branded grocery market at the postal area level in GB.This includes Tesco and Sainsbury's, the major retailers identified as being significant players in the market along with the co-operative group and four significant symbol group retailers; Musgrave Group, Costcutter, Premier and SPAR.

5.1   National growth

Chapter 4 identified the data used in this thesis. Both GMAP Ltd. and Sainsbury's provided a database of store locations for the years 2003-2012. The two databases were combined to create a picture of the changing branded convenience grocery market in GB. The database contains data on store location (XY coordinates) and store size in square feet of floorspace. From this database, the four largest retailers as of the final year of data (2013) were extracted for analysis to quantify the extent to which the largest grocery retailers in GB have committed to convenience grocery retailing. The first piece of analysis reported in this chapter analyses the ten-year period from 2003 to 2012 looking at the change in store network by store size of the four largest grocery retailers. The retailers analysed are:

1) Tesco
2) ASDA

3) Sainsbury's
4) Morrisons

The review of the literature on the supply side of grocery retailing in chapter 2 identified Tesco and Sainsbury's as being at the forefront of convenience grocery retailing among the major grocery retailers. ASDA were identified as having pursued a different path in adopting a strategy of 'space sweating' in which existing supermarkets were extended to become very large hypermarkets in which non-grocery products played a prominent role (Wood and McCarthy, 2013). Finally, Morrisons were identified as a retailer who had announced a commitment to entering the convenience grocery market but had been slow to commit to this change in store format choice having opened their first convenience grocery stores in West Yorkshire in 2011.

The stores operated by the major retailers listed above were divided into five store format categories based on store size in order to identify the contribution of convenience stores to the total stock of grocery stores operated by the two groups of retailers. The store format categories assigned to stores are as follows:

1) Under 3,000 sq. ft.
2) 3,000 – 10,000 sq. ft.
3) 10,000 – 25,000 sq. ft.
4) 25,000 – 60,000 sq. ft.
5) Over 60,000 sq. ft.

The breakdown of stores into the 5 store categories listed above follows the precedent set by the methodology of the work of Thompson (2013). Under 3000 sq. ft. is the preferred convenience format of major retailers and are covered by Sunday trading laws. 3-10,000 sq. ft. stores are indicative of small grocery stores and are not compliant with Sunday trading laws distinguishing them from convenience grocery stores, 10-25,000 sq. ft. stores are medium sized supermarkets, 25-60,000 sq. ft. supermarkets are large supermarkets and 60,000 sq. ft. stores are hypermarkets, often achieved by extending large supermarkets, a strategy heavily pursued by ASDA and quantified later in this chapter. The breakdown into these categories also looks to quantify the varying strategies of store growth adopted by the four largest grocery retailers in GB.

The analysis reported in this chapter looks at the contribution of stores of each of the five store formats listed above to the overall stock of grocery stores operated by the

four largest retailers. The contribution of the less than 3000 sq. ft. store format (convenience stores) was used to assess the growth of major grocery retailers into the convenience market over the ten year period 2003 to 2012. As previously discussed, stores under 3000 square feet are allowed increased trading hours (especially on Sundays in GB) and are therefore the preferred format of major retailers convenience stores.

5.2 Growth of the branded convenience grocery market in GB

In identifying the contribution of stores of each size format to the overall portfolio of stores operated by each retailer, it is important to look at two measures in order to quantify the importance of convenience stores (and other store formats) to overall retailer operations. These are total numbers of stores and total floorspace. Total numbers of stores allows the research to examine the proportion of stores operated by each retailer that can be defined as convenience stores. However, as convenience stores are comparatively small, it is important to compare this measure of growth against the proportion of total operating floorspace delivered by each format size for each retailer. This is particularly relevant in the grocery industry as revenue per square foot is a very common and useful indicator of performance. The indicators used in this section take the following form:

1. Total number of stores by store size format for each retailer;
2. Total floorspace of stores by store size format for each retailer;
3. Total number of stores by size format for the four largest retailers; and
4. Total floorspace of stores by store size format for the four largest retailers

5.2.1 Tesco

Tesco had the largest share of the grocery market in Great Britain at the end of 2012 and was a truely national retailer, operating stores across the country. In the face of increasing difficulty in growing their store network through large supermarket openings a major strategy of Tesco was to enter the convenience market . As discussed in the literature review in chapter 2, they have grown their convenience offer substantially since first venturing into the petrol forecourt market in 1994. They now operate two convenience store formats, Tesco Express and One Stop convenience stores, the latter not bearing the Tesco logo but operated by them. They have achieved growth both through organic store openings and significant acquisitions. This section looks to quantify the retailers growth in the convenience market and changes to the firm's

operation in other store formats. Figure 5.1 shows the total number of stores by each size format operated by Tesco in Great Britain from 2003 to 2012.

Figure 5.1 Tesco store portfolio by store size in GB, 2003 to 2012. Source: GMAP Ltd.
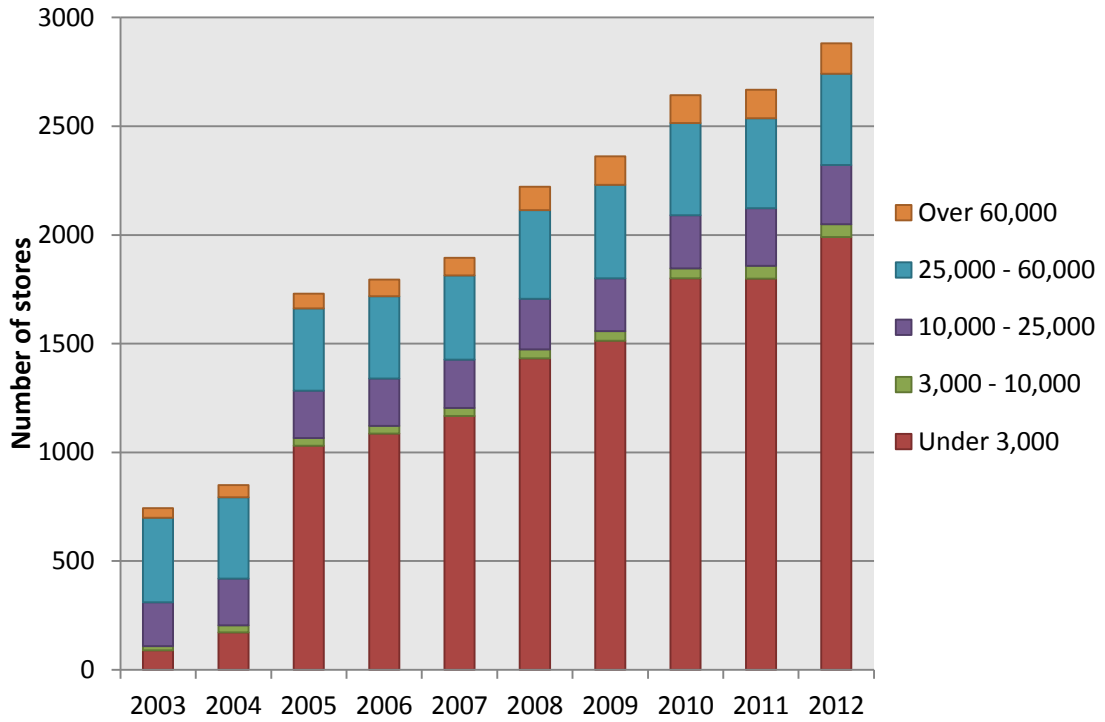
In the ten year period from 2003 to 2012 the total number of stores operated by Tesco grew by over 2000, with the majority of this growth accounted for by convenience grocery stores. Tesco operated close to 2000 convenience stores in 2012, having grown from under 100 in 2003. A significant proportion of this growth in convenience stores (45%) occurred between 2003 and 2005 and was predominantly driven by the retailer acquiring the T&S Store chain. This acquisition occurred in 2003 but appeared in the data in 2005 following the rebranding of stores from the original store fascia to the Tesco Express format.

The majority of the subsequent growth in small format store numbers occurred as a result of organic growth of Tesco Express convenience stores opened by the retailer. Further growth of the brand has also been operationalised through the growth of very large hypermarkets over 60,000 square feet in size - the Tesco Extra store fascia - which when considered alongside the growth in convenience stores confirms the movement of the retailer towards rapid growth in formats other than the traditional large out-of-town supermarket format. However, the retailer has still continued to grow its

offer in the supermarket store format of between 10,000 and 60,000 sq. ft. in size, but at a slower rate than the smaller and larger formats.

It is useful to look at the impact of changes in store numbers in terms of floorspace provision offered by retailers. Figure 5.2 shows the total floorspace offered by Tesco in each store size format in Great Britain from 2003 to 2012:

Figure 5.2 Tesco floorspace by store size in GB, 2003 to 2012. Source: GMAP Ltd.
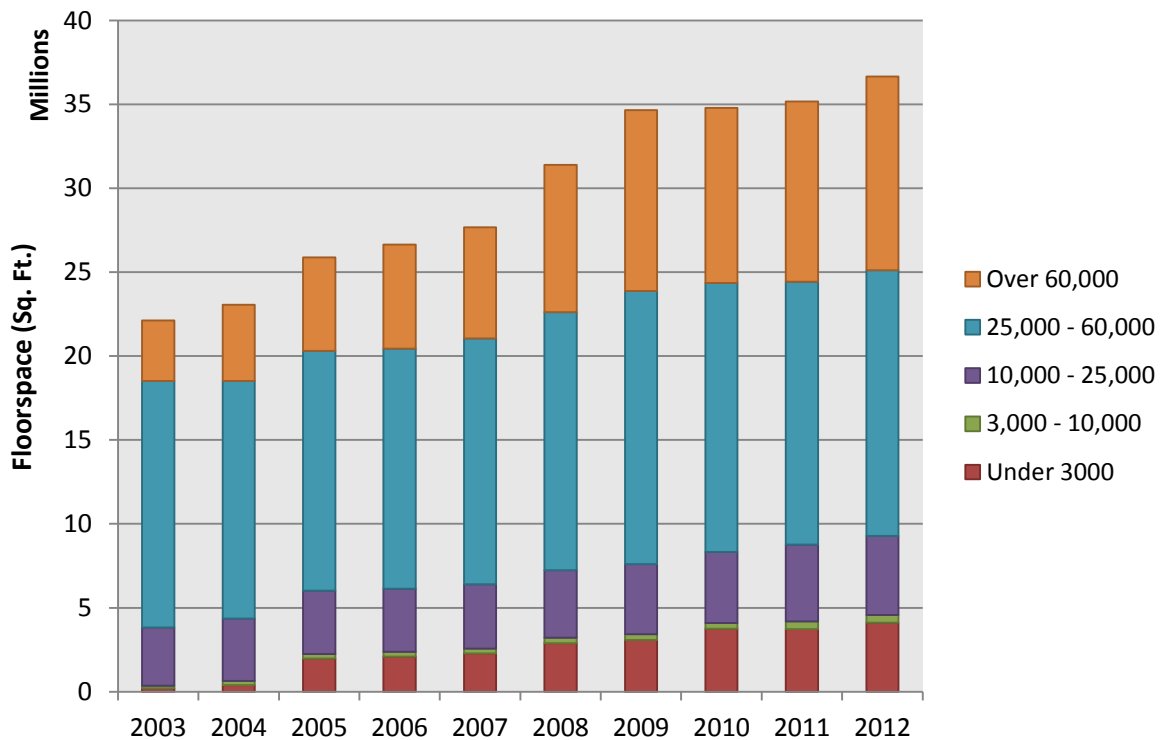
Convenience store floorspace has grown from around 200,000 square feet in 2003 to over 4 million in 2013, increasing the contribution of small format stores from less than 1% to over 12% of total floorspace. In 2003, the 25,000 to 60,000 mid to large supermarket sized format accounted for over 66% of Tesco's floorspace, although by 2012 this had been squeezed to just over 43% of floorspace, no longer the majority of the retailer's presence. The most significant growth in the proportion of floorspace contributed has occurred in hypermarkets over 60,000 square feet in size. The growth in floorspace in this format (from around 3.5 million square feet in 2012 to over 11.5 million in 2012) has increased this formats contribution from 16.3% of square feet to 31.5%.

5.2.2 ASDA

ASDA were the second largest grocery retailer in GB at the end of 2012. The literature review in chapter 2 found little evidence that ASDA had committed to growing their offer in the convenience market in response to the changes in the market. In contrast, the evidence points to a different example of how a major retailer responded to changing market conditions and attempted to grow through a strategy of store expansion rather than extensive growth in the convenience grocery market. Figure 5.3 shows the total number of stores in each store size format operated by ASDA in Great Britain from 2003 to 2012.

Figure 5.3 ASDA store portfolio by store size in GB, 2003 to 2012. Source: GMAP Ltd.
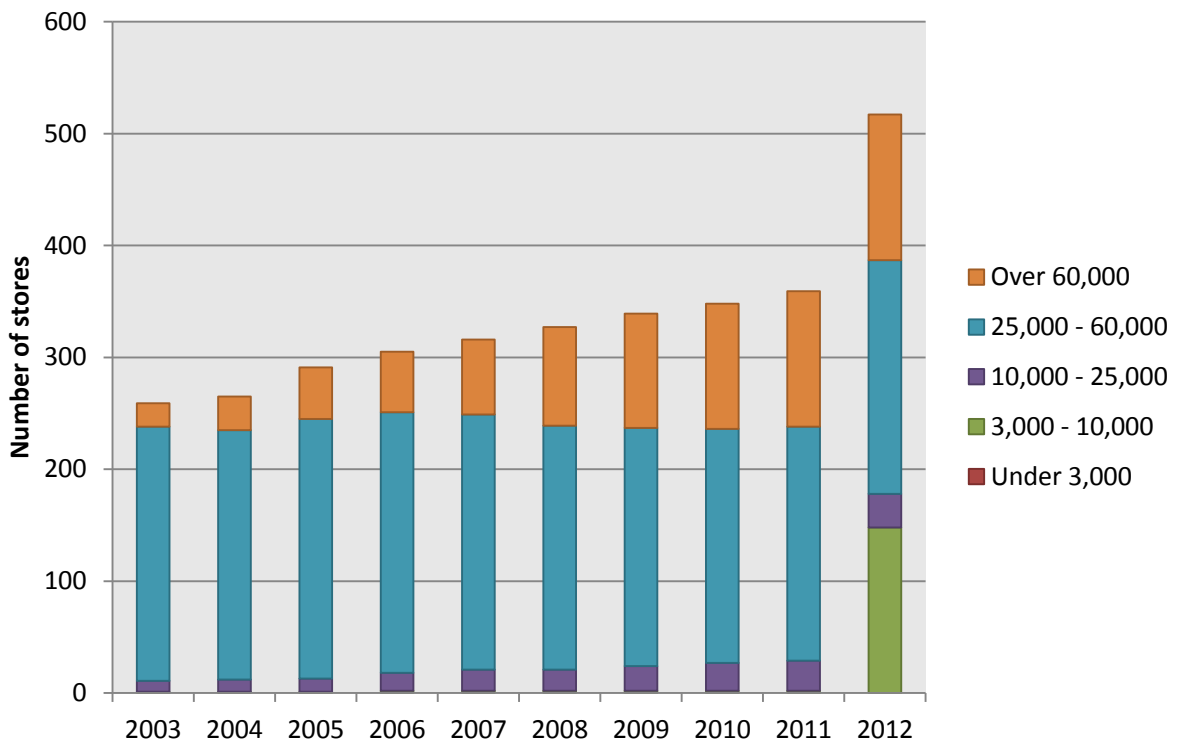
Between 2003 and 2012 ASDA's portfolio in GB grew by over 250 stores, much of this growth accounted for by the retailer's acquisition of Netto in 2011, appearing in the data in 2012. This growth is seen in the 3,000 to 10,000 square foot store format category as the Netto stores acquired by the retailers averaged 8000 sq. ft. in size. The retailer purchased 193 stores in the deal but was required to sell off a number of outlets to comply with Competition Commission guidelines leaving 147 to be opened as new ASDA stores. The remaining growth in store numbers is predominantly accounted for by the growth in very large hypermarkets, partly at the expense of stores in the 25,000 to 60,000 square feet category that were extended to become over 60,000 square feet

hypermarkets. Unlike both Tesco and Sainsbury's, convenience had not been at the forefront of ASDA's operations in the 2000s and the retailer did not express a significant commitment to smaller format grocery retailing until 2010, in which they acquired a large number of stores from Netto, although the majority of these stores were over 3000 sq. ft. in size. (Brooks, 2014). ASDAs store portfolio expressed in terms of floorspace from 2003-2012 is shown in Figure 5.4.

Figure 5.4 ASDA floorspace by store size in GB, 2003 to 2012. Source: GMAP Ltd.
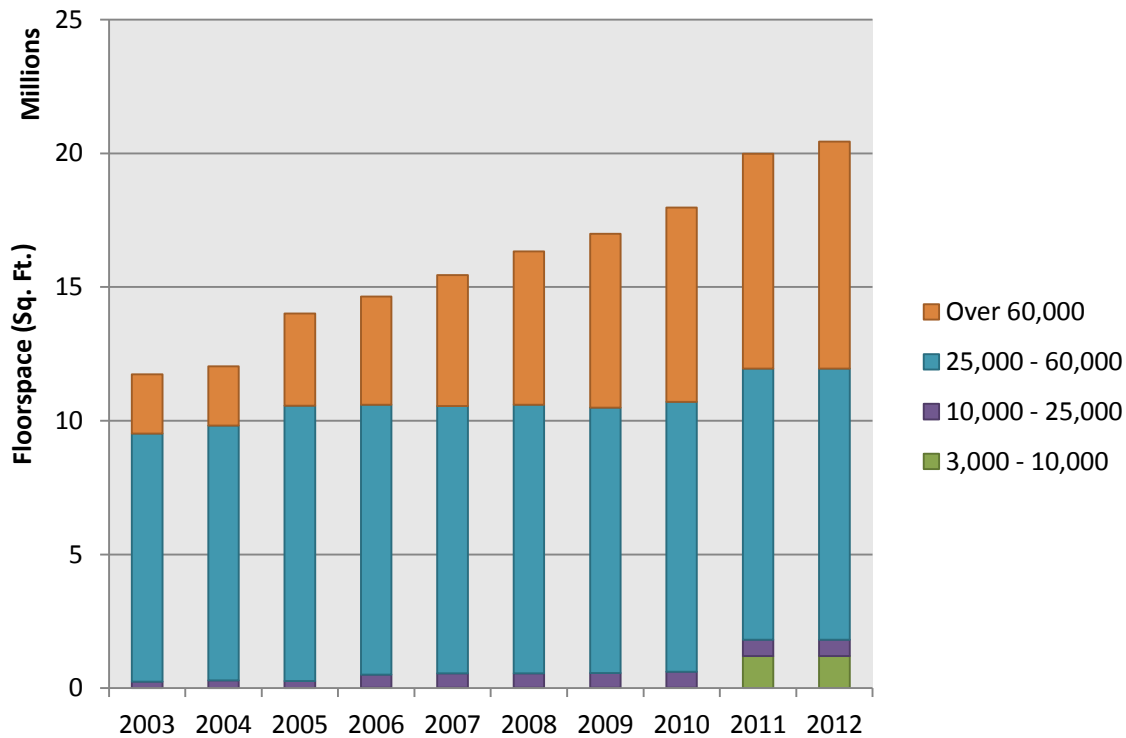
The growth in stores shown in Figure 5.3 corresponds to an increase in floorspace of approximately 9.2 million square feet. The significant growth in number of stores resulting from ASDA's acquisition of Netto stores accounted for 57% of the increase in total stores. However, as the stores were small sized supermarkets, the resulting floorspace growth of around 1.2 million square feet accounts for around 13% of the total floorspace growth.

As Figures 5.3 and 5.4 show, the over 60,000 square feet store format category has grown whereas the 25,000 to 60,000 square feet category has reduced both in store numbers and floorspace. Much of this change is due to ASDA's store extension programme highlighted in the literature review in chapter 2. This extension programme was often conducted through the introduction of mezzanine flooring in existing large warehouse style stores and the reduction of car park sizes resulting in stores with

floorspace exceeding the 60,000 square feet hypermarket threshold (Wood et al. 2006).

5.2.3 Sainsbury's

Sainsbury's had the third largest share of the grocery market in Great Britain at the end of 2012. In the face of increasing difficulty in growing their store network through large supermarket openings, a major strategy of Sainsbury's was to enter the convenience market. As discussed in the literature review in chapter 2, in response to shifting market conditions and the continuing pursuit of market share, Sainsbury's adopted a strategy akin to Tesco in advancing the retailer's offer in the convenience market between 2003 and 2012. Sainsbury's store portfolio by store size can be seen in Figure 5.5.

Figure 5.5 Sainsbury's store portfolio by store size in GB, 2003 to 2012. Source: GMAP Ltd.
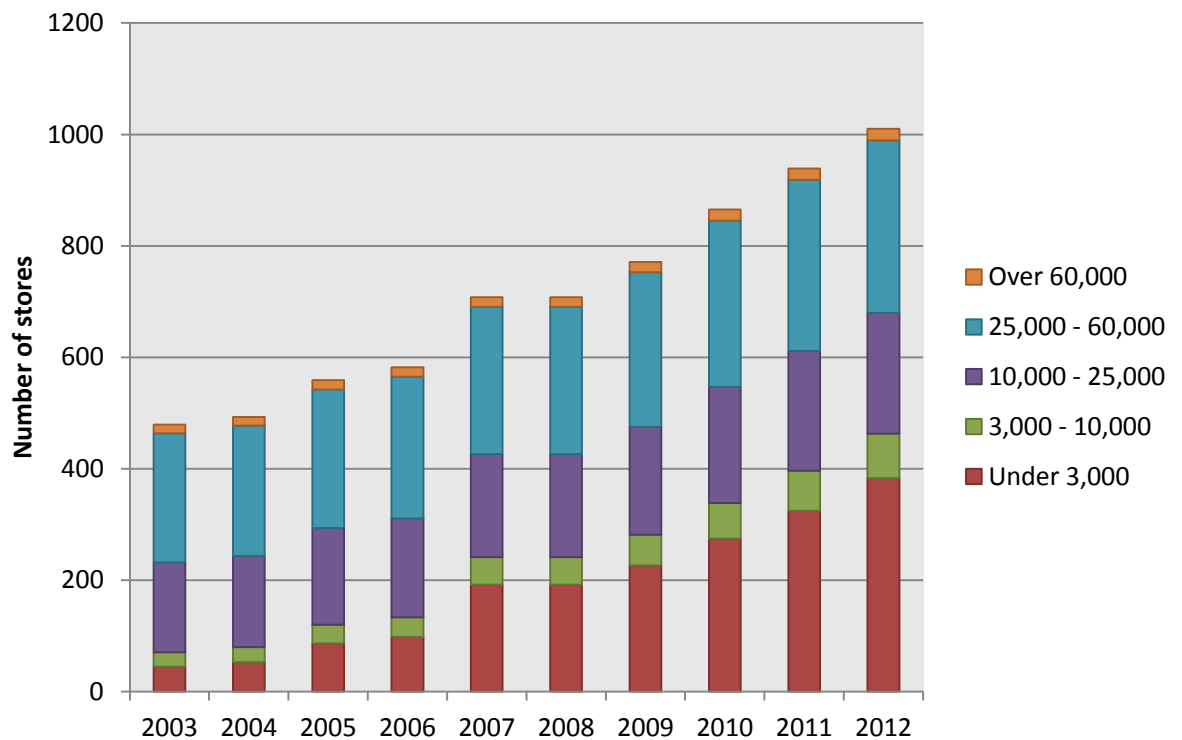
In the ten year period Sainsbury's more than doubled its store portfolio from 479 stores to over 1000 stores. The majority of this growth came from investment in the convenience market which accounted for 64% of the growth in stores, an increase of 339 convenience stores. Over 25% of the growth in convenience stores operated by Sainsbury's occurred between 2006 and 2007 as a result of the rebranding of earlier

acquisitions made by Sainsbury's in 2004, when the retailer acquired 174 convenience stores, including 114 stores previously owned by Jackson's in Yorkshire and the North Midlands (Finch, 2004). Much like Tesco, a significant proportion of the retailer's growth in the convenience market has resulted from acquisitions of smaller convenience store chains.

Whilst the convenience format accounted for the majority of growth, there was growth in Sainsbury's total number of stores in all five store size formats and the retailer has experienced steady year-on-year growth of stores of between 3 and 21% between 2003 and 2012. Unlike Tesco and ASDA, Sainsbury's chose not to advance their offer in the large hypermarket format extensively with this format accounting for only 1% of the total growth of Sainsbury's stores. Figure 5.6 shows the total floorspace offered by Sainsbury's in each store size format in Great Britain from 2003 to 2012.

Figure 5.6 Sainsbury's floorspace by store size in GB, 2003 to 2012. Source: GMAP Ltd.
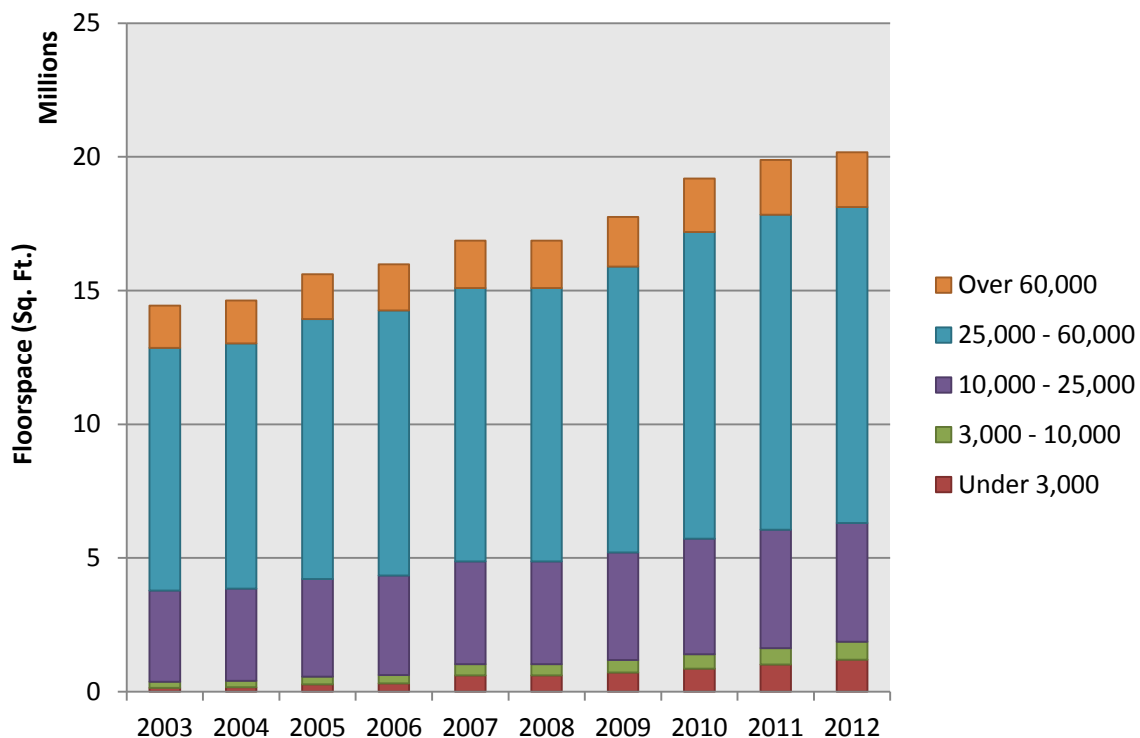
Convenience stores contributed 64% of the growth in Sainsbury's store portfolio from 2003 to 2012 which translates to a growth in floorspace of 18%. In 2003 stores under 3,000 square feet in size made up 1% of Sainsbury's floorspace, but after significant investment in the convenience market, this Figure had grown to 5.9% of floorspace, becoming a substantial part of the retailer's presence in GB. Despite evidence that it

became more difficult for the major grocery firms to advance their offer in supermarkets, stores between 25,000 and 60,000 square feet, the traditional medium to large supermarket format size, accounted for almost half of the retailer's growth in floorspace between 2003 and 2012. Stores in this store size format have, however, become less prominent in their contribution to Sainsbury's total floorspace offer in GB between 2003 and 2012, falling from 62.9% of floorspace to 58.6%, yet still make up the majority of the retailers grocery offer.

5.2.4 Morrisons

Morrisons was the fourth largest grocery retailer in GB in 2012 with an extensive network of stores, particularly in Northern England. The retailer adopted a different strategy to both Tesco and Sainsbury's convenience investment and ASDA's commitment to creating hypermarkets through store expansion. The retailer expanded their offer in mid to large supermarkets of between 10,000 and 60,000 sq. ft. in size. This section will highlight the method of growth of Morrisons between 2003 and 2012. Figure 5.7 shows the total number of stores in each store size format operated by Morrisons in Great Britain from 2003 to 2012.

Figure 5.7 Morrisons store portfolio by store size in GB, 2003 to 2012. Source: GMAP
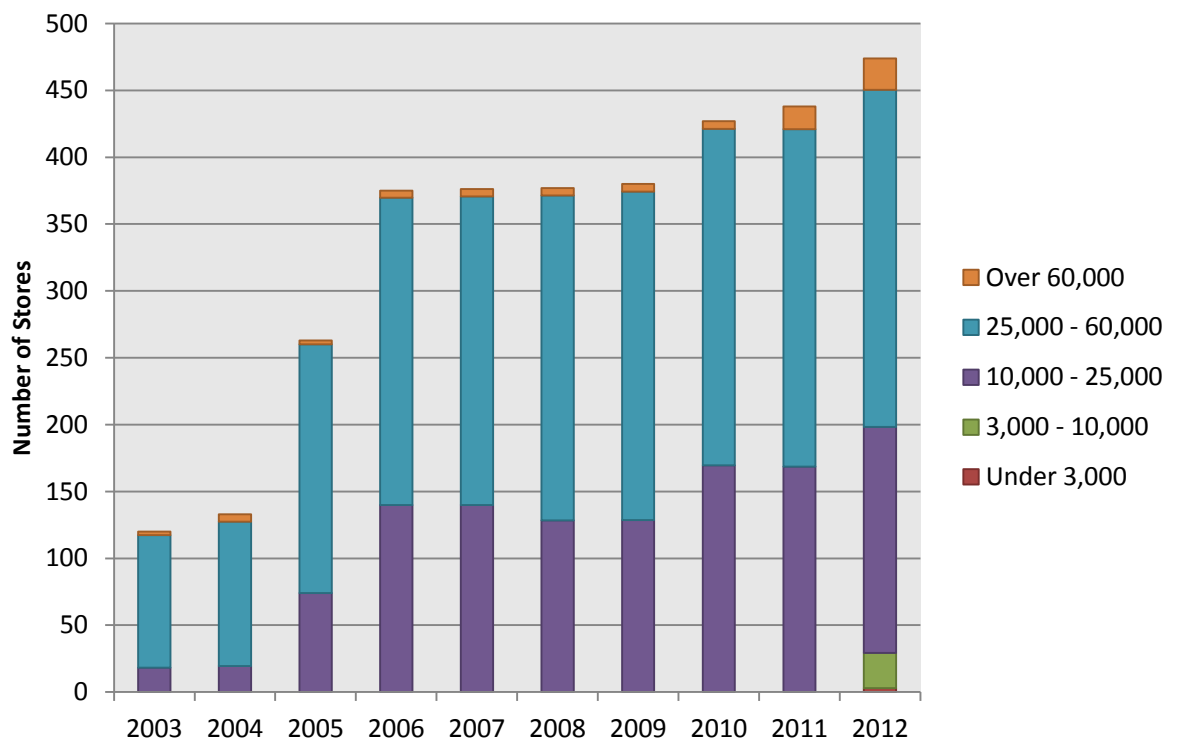
Between 2003 and 2012 the total number of stores operated by Morrisons in GB rose from 120 to more than 470, an increase of almost 400%, the largest percentage increase of any of the major four retailers over the period studied. Convenience retailing played a limited role in the retailer's growth strategy between 2003 and 2012. Whilst Sainsbury's and Tesco began significant investments into the convenience market in the late 1990s and early 2000s, Morrisons were relative latecomers to the party and did not pursue small format grocery retailing until 2011 when it opened its first M Local convenience store in West Yorkshire (Hall, 2011), appearing in GMAP Ltd.'s store database in 2012. However, the retailer later pulled out of the convenience market by selling its small stores to investment firm Greybull Capital in 2015 (Armstrong, 2015)

The majority of the growth of Morrisons in the ten year study period was encapsulated by traditional format stores ranging from 10,000 to 60,000 square feet in size. These stores accounted for an almost equal share of 85% of the growth in stores operated by Morrisons. As Figure 5.7 shows, the retailer advanced its offer in the 3,000 to 10,000 square foot range between 2011 and 2012, some of this was through taking advantage of ASDA's sell off of a portion of its Netto investment. The retailer in fact beat ASDA to the first opening of a newly branded Netto by opening three rebranded stores in May 2011, appearing in the GMAP data in 2012 (Brooks, 2014). Figure 5.8 shows the total floorspace offered by Morrisons in each store size format in Great Britain from 2003 to 2012

Figure 5.8 Morrisons floorspace by store size in GB, 2003 to 2012. Source: GMAP Ltd.
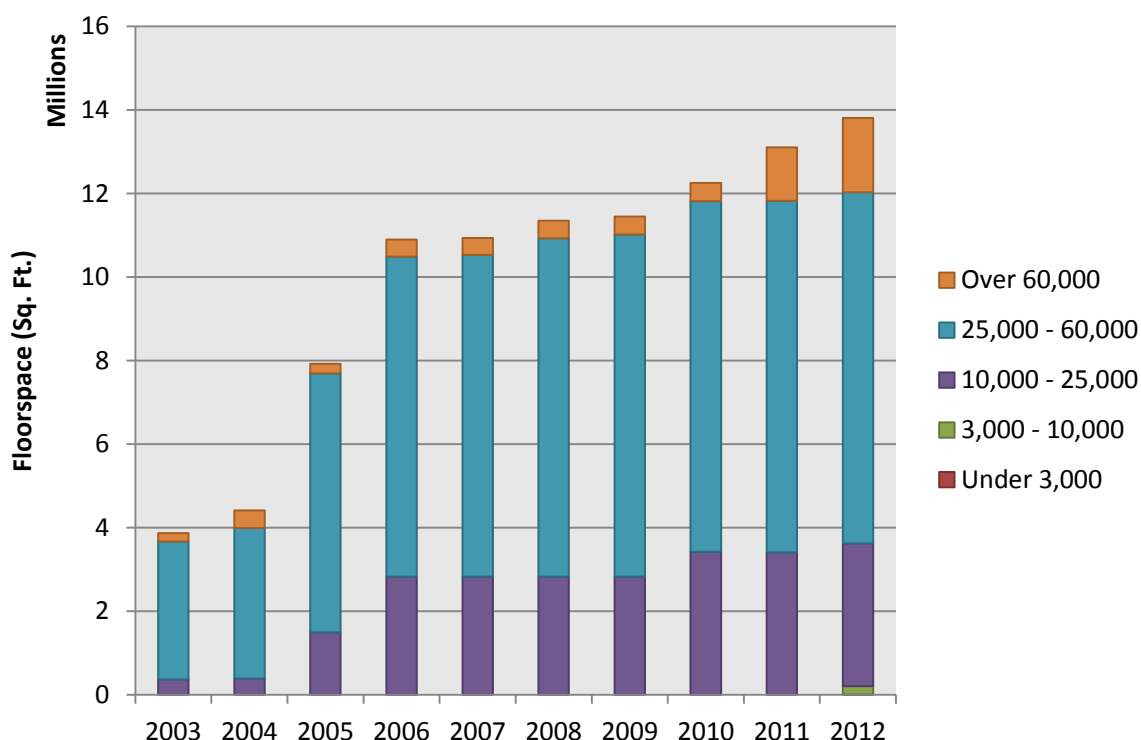
Following its late entry into the convenience grocery market, convenience stores operated by Morrisons accounted for just 0.1% of the retailer's floorspace in 2012. When compared to the 12% and 5% contribution of Tesco and Sainsbury's convenience floorspace respectively, this highlights the extent of Morrisons lag in developing convenience operations. In 2012, traditional supermarket store formats between 25,000 and 60,000 square feet accounted for 64% of Morrisons floorspace, having reduced from 85% in 2003. Whereas the reduction in the contribution of traditional supermarkets for ASDA, Tesco and Sainsbury's was filled by a combination of small format convenience stores and very large hypermarkets, the growth of small supermarkets between 10,000 and 25,000 square feet was the biggest growth area in proportion of floorspace for Morrisons, identifying the fact that the retailer adopted a unique strategy amongst major retailers.

5.2.5 The four major retailers combined convenience network

To bring together the analysis in this section, it is important to consider the operations of the major grocery retailers as a whole. Figure 5.9 identifies the total number of stores in each store size format operated by the four major retailers between 2003 and 2012.

Figure 5.9 The four largest grocery retailers (Tesco, ASDA, Sainsbury's and Morrisons) store portfolios by store format in GB, 2003 to 2012. Source: GMAP Ltd.
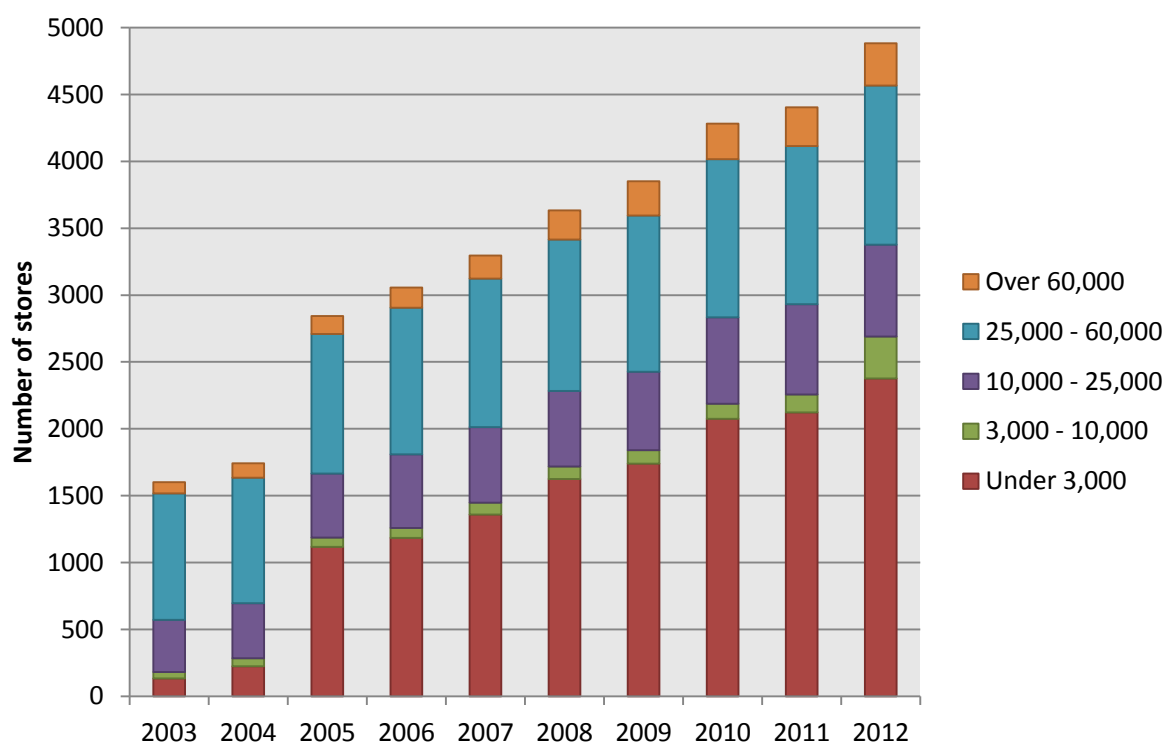
Figure 5.9 shows the changing nature of the four largest grocery retailer's total stores by store size format. Between 2003 and 2012, the total number of stores operated by the four retailers in Great Britain increased by over 3200 stores, from 1601 to more than 4883. Figure 5.10 shows the corresponding growth of floorspace for the major retailers over the ten year period, the trebling of stores operated by the major retailers resulted in a 1.8 times increase in floorspace.

Figure 5.10 The four largest grocery retailers (Tesco, ASDA, Sainsbury's and Morrisons) floorspace by store format in GB, 2003 to 2012. Source: GMAP Ltd.
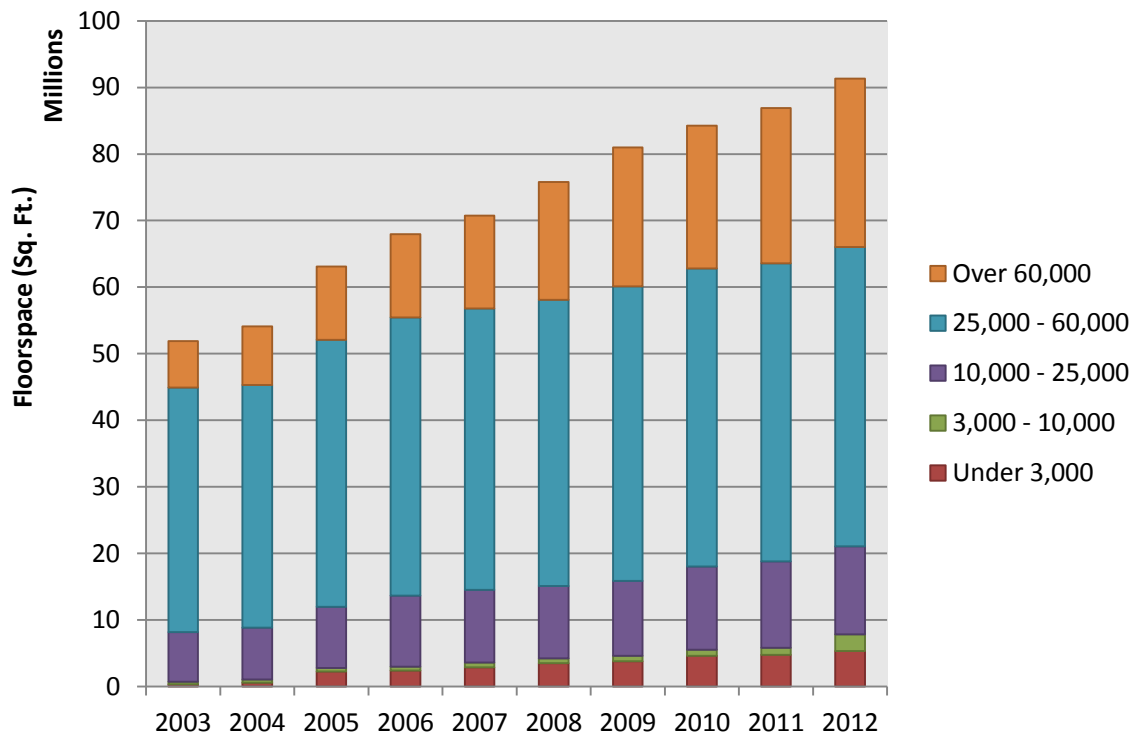
Furthermore, Table 5.1 highlights the contribution of each retailer to the overall growth (or decline) of major retailers in each store size format. Tesco contributed the largest share of additional stores, 65.2%, followed by Sainsbury's, Morrisons and ASDA contributing 16.2%, 10.8% and 7.9% of the growth in stores respectively. The total number of convenience stores operated by the four largest multiples grew from 134 to 2377, an increase of 2244, accounting for over 68% of the increase in total stores. The almost 18 fold increase in the number of major multiple convenience stores corresponded in a growth in convenience floorspace of almost 5 million square feet, accounting for over 12.5% of the total growth in floorspace. The increase in small-format convenience stores is mainly a result of Tesco and Sainsbury's pursuing the convenience market aggressively. The growth in Tesco Express convenience stores

and Tesco's ownership of several hundred One Stop stores were responsible for almost 85% of the growth in convenience stores with Sainsbury's accounting for the majority of the remaining growth.

Table 5.1 Contribution of each retailer to the growth of each store format among the four major retailers (Tesco, ASDA, Sainsbury's and Morrisons), 2003-2012

Store Format	Tesco	ASDA	Sainsbury's	Morrisons	Total
Under 3,000	84.8%	0.0%	15.1%	0.1%	100%
3,000 - 10,000	14.4%	55.3%	20.4	9.9%	100%
10,000 - 25,000	24.0%	6.7%	18.4	50.8%	100%
25,000 - 60,000	12.6%	-7.4%	32.1	62.7%	100%
Over 60,000	41.8%	47.0%	2.2	9.0%	100%
Overall	*65.2%*	*7.9%*	*16.2*	*10.8%*	*100%*

47% of the increase in 60,000+ sq. ft. stores is a direct result of ASDA extending superstores through mezzanine flooring (Wood *et al.,* 2006). The number of hypermarkets operated by the major retailers rose by 232 stores over the ten year period with this format contributing 7% of the increase in total stores, corresponding to a 46.5% increase in total floorspace, the biggest contributor of any of the store formats. Much has been written about the difficulties faced by retailers in continuing to grow in the traditional supermarket formats, however, between 2003 and 2012, stores between 10,000 and 60,000 square feet were responsible for around 16.5% of the growth in major retailer store numbers. However, the majority (62.7%) of the growth in these store formats was driven by the growth of Morrisons, a retailer less restricted by changes in retail legislation as it didn't have the share of the market commanded by Tesco and Sainsbury's during the Competition Commission's investigations.

Both the increase in small and large stores can be seen as a result of changes to planning policy alongside shifts in consumer demand encouragin the large retailers traditionally focused on superstore store formats to diversify into other retail channels. Different responses have yielded different compositions of stores operated by the largest grocery retailers operating in Great Britain. The varying responses hold one common trend, they have all served to advance the major retailers network of stores, widening the scope of the retailers. Tesco and Sainsbury's (and possibly Morrisons moving forward) advancement into the convenience grocery market has brought them

into more direct competition with different types of independent retailers and retail groups who were previously more commonly associated with small format grocery retailing.

5.3 Regional growth of the branded convenience grocery market

The change in convenience stores and total floorspace by former government office region in Great Britain is examined in this section. This allows a sub national look at the growth of convenience stores at a level easily recognisable by both retail practitioners and the general public. Moreover, Sainsbury's disaggregate by region when allocating staff to perform analysis on their existing store portfolio and potential new sites to expand their market share in the grocery market in Great Britain. Between 2003 and 2012, the number of convenience stores operated by the major grocery retailers increased by over 2000 stores.

This section looks at how this increase in stores has been spatially distributed by former Government Office Region (GOR) in Great Britain, also taking into account national growth in Wales and Scotland. This analysis uses the same dataset as the GB analysis already presented in this chapter. The four major retailer's convenience stores have been aggregated together and analysed using point in polygon analysis using SQL in MapInfo Professional 12.5 for each year in the period 2003 to 2012. The provision of convenience grocery retailing is analysed both in terms of store numbers and relative floorspace before the level of provision per capita in each region is assessed. The majority of convenience stores are operated by either Tesco or Sainsbury's as Morrisons did not commit to convenience retailing until 2011 (with three stores operated by the retailer appearing in the data in 2012) and ASDA were yet to open a convenience grocery store by the end of 2012.

5.3.1 Total branded convenience stores operated by the four largest grocery retailers by former government office region in GB, 2003-2012

Figure 5.11 shows the trend in total number of convenience stores operated by the four largest grocery retailers in each former Government Office Region of England along with Scotland and Wales between 2003 and 2012.

Figure 5.11 Major retailer convenience stores by former Government Office Region (GOR) in GB, 2003 to 2012. Source: GMAP Ltd.
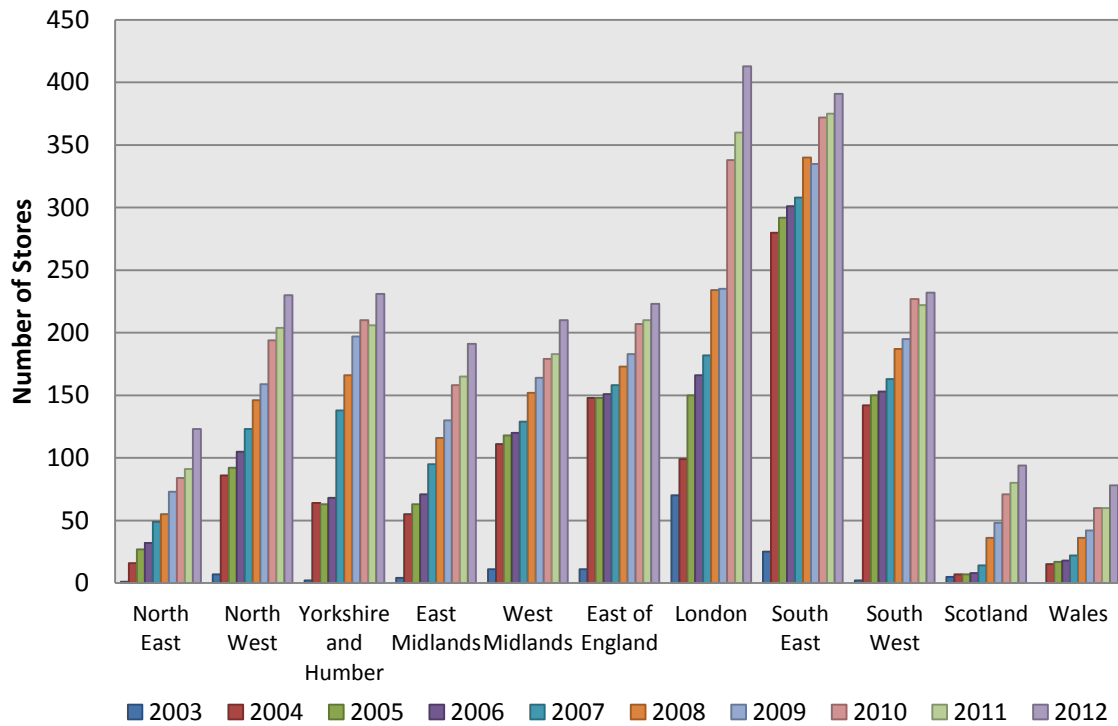
In 2003, London was the region with the most convenience stores (74), many more stores than the nearest competitor, the South East with 29 stores operated by the four largest grocery retailers. By 2012, the average number of convenience stores per region had grown to 220. However, both the growth and total number of stores per region has been geographically variable. Wales experienced the smallest rise in convenience stores in the ten year period, increasing their offer from 0 stores to 77. Conversely, the South East was the region experiencing the greatest growth in the raw number of major retailer convenience stores with a rise of 360 stores. Table 5.2 shows the total growth of stores in each region and the relative contribution of each region to the total growth of convenience stores in Great Britain.

Table 5.2 Growth in major retailer convenience stores by region of GB, 2003 to 2012. Source: Abstracted from data provided by GMAP Ltd

Region	Growth	Proportion of store increases (%)
South East	360	16.0
London	338	15.1
Yorkshire and Humber	226	10.1
South West	226	10.1
North West	220	9.8
East of England	209	9.3
West Midlands	196	8.7
East Midlands	184	8.2
North East	120	5.3
Scotland	88	3.9
Wales	77	3.4
Total	*2244*	*100*

London and the South East are the regions that have experienced the largest growth in convenience stores operated by the major retailers and continue to have the greatest number of convenience stores. London and the South East combined to experience almost one third of the total growth between 2003 and 2012, making up 15.1% and 16.0% of the growth in major multiple retailer convenience stores respectively. Conversely, both Scotland and Wales individually accounted for less of the growth than any of the English regions, collectively contributing less than 10% of the increase in major retailer convenience stores.

Much of the growth experienced across the regions occurred between 2003 and 2004 and appear in the dataset in 2004. The largest incremental growth happened between these dates in the North West, East Midlands, West Midlands, East of England and the South East and comes as a result of Tesco's acquisition of T&S stores. The One Stop chain previously operated by T&S and purchased by Tesco has a wide geographic distribution and accounts for the single largest acquisition of convenience stores by any of the major retailers to date. London has experienced a more steady growth, less effected by large acquisitions as this has been an area of significant organic growth for both Tesco and Sainsbury's. Yorkshire and the Humber has witnessed two large peaks in convenience growth, 2003-2004 as a result of Tesco's T&S acquisition and another

significant acquisition by Sainsbury's of the Bells, Jacksons and JB Beaumont convenience store chains in 2004.

5.3.1 Branded convenience floorspace by region

Figure 5.12 shows the trend in total floorspace operated by the four largest grocery retailers by region. Between 2003 and 2012, convenience store floorspace grew across all regions. In 2012, London and the South East had the greatest regional provision of convenience grocery floorspace operated by the four largest grocery retailers, with 954,000 sq. ft. and 901,000 sq. ft. respectively. Furthermore, by 2012, the North East, Scotland and Wales were the regions with the smallest provision of floorspace operated by the four largest grocery retailers, with 284,000 sq. ft., 217,000 sq. ft. and 179,000 sq. ft. respectively.

Figure 5.12 Major retailer convenience floorspace by former Government Office Region (GOR) in GB, 2003 to 2012. Source: GMAP Ltd.
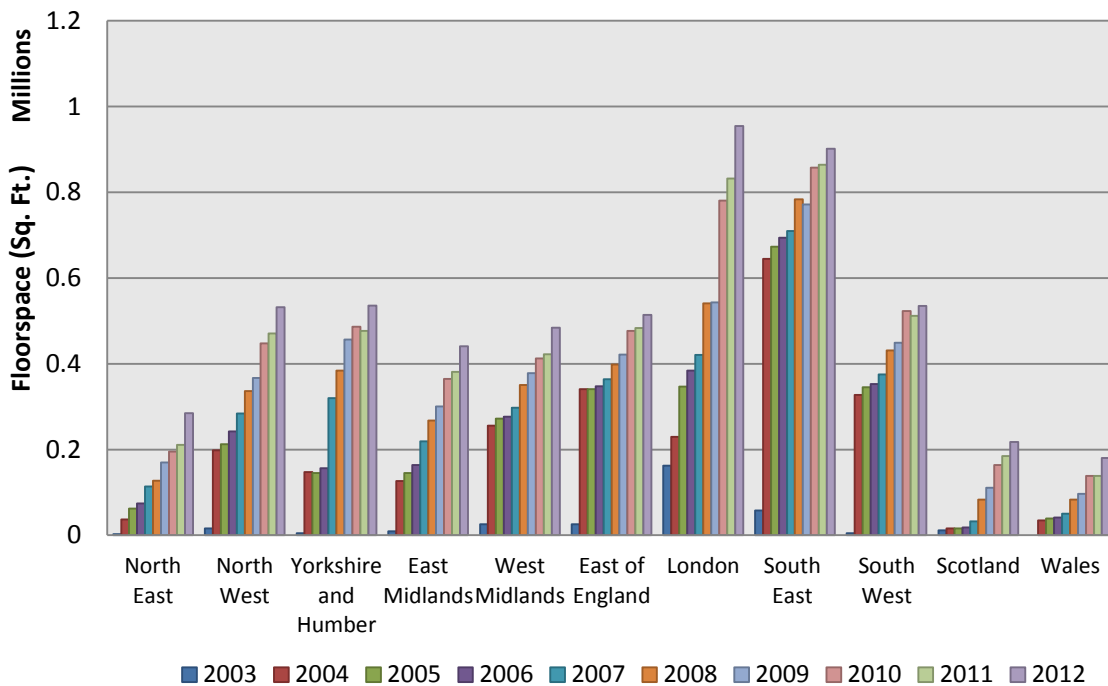
Additionally, Figure 5.13 shows the regional floorspace of convenience stores when controlling for population, giving a per capita provision of convenience grocery retail floorspace operated by major grocery retailers. This presents an alternative picture of provision of convenience store floorspace and appears to be more evenly distributed amongst the regions of England by 2012.

123

Figure 5.13 Major retailer convenience floorspace provision per capita by former Government Office Region (GOR) in GB, 2003 to 2012. Source: GMAP Ltd.
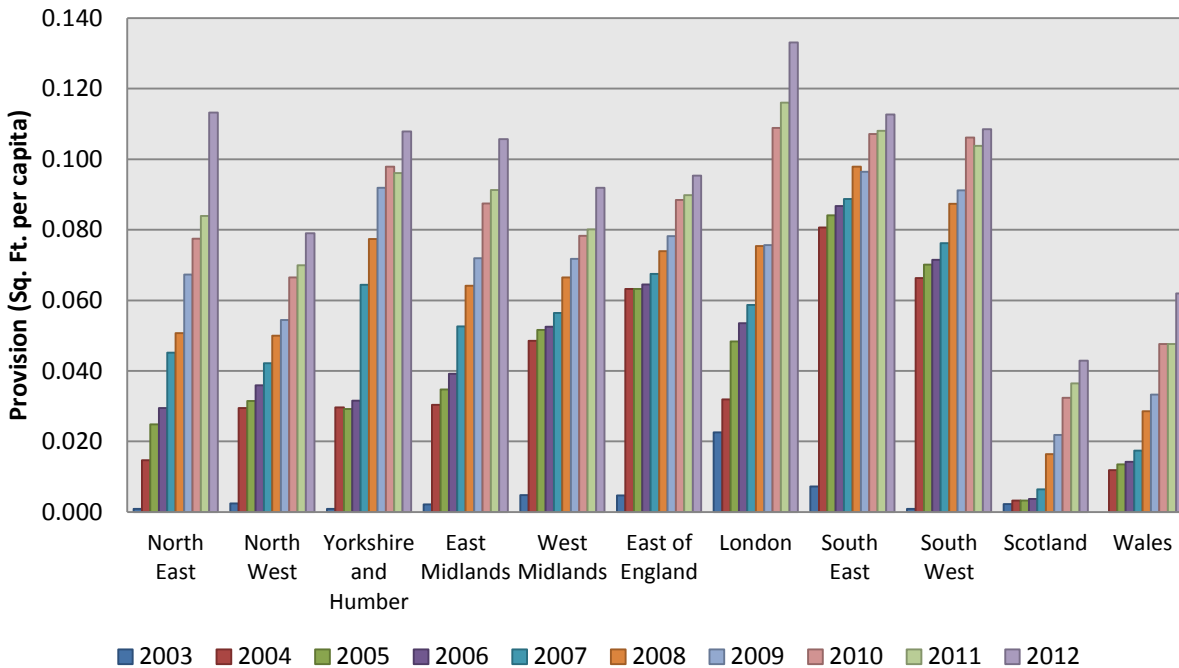
The North East region appears more prominently in this measure of provision; however, both Scotland and Wales continue to have the two lowest floorspace provisions across all regions. In 2012, the North East region had a provision of 0.113 sq. ft. per capita, the highest relative provision of floorspace outside of London (0.133 sq. ft. per capita).

5.4 Grocery Market Share

As discussed in chapter 4, store location data was obtained from both GMAP Ltd. and Sainsbury's for the years 2003-2012. These two databases were combined to create an accurate picture of the supply of stores in the grocery market in the UK. Moreover, groundtruthing was conducted by the author visiting various locations in Sheffield and Leeds to see if the retail mix matched the joined GMAP/Sainsbury's database. This chapter analyses this dataset for the year 2012, looking at the market share by floorspace of a number of prominent convenience grocery retailers in Great Britain at the end of 2012.

This data was used to calculate the share of the grocery market based on total floorspace operated of the prominent major grocery retailers and prominent symbol

groups retailers as a proportion of total convenience grocery market at the postal area level in Great Britain. The floorspace operated by each retailer was aggregated to the postal area level using point in polygon analysis using SQL in MapInfo Professional 12.5. As described in chapter 4, postal areas are comprised of one or two letter codes identifying a place that is served by a Royal Mail sorting office.

Subsequently the total convenience floorspace operated by each retailer as a proportion of total convenience floorspace in each postal area was used to give a market share for each retailer for each postal area. Ideally, market share would be calculated using sales data; however, it is very difficult to acquire sales data for a single retailer let alone all retailers involved in convenience grocery retailing in Great Britain. However, the proportion of floorspace operated by each retailer in each postal area provides a useful overview of the presence of each retailer in convenience grocery retailing at the postal area level. This method of calculating market share has precedence in the academic retail literature and has been used in at least two other similar studies (Langston et al. 1998; Poole et al. 2002)

5.5 The convenience grocery market

Figure 5.14a shows the spatial extent of the convenience grocery market in Great Britain in 2012. Not surprisingly, when plotted as raw numbers, the highest floorspace totals (650,000 sq. ft.) are in key urban areas, notably Glasgow (G), Newcastle Upon Tyne (NE), Sheffield (S), Nottingham (NG), Birmingham (B), Swansea (SA) and Cardiff (CF). If we aggregate the postal areas of London, the city has the greatest total floorspace of any of the major conurbations. Figure 5.14b plots the spatial variations in the market share of the overall convenience market (floorspace 3000 sq. ft.), expressed as a percentage of all floorspace in the grocery sector. When expressed in this way, a very different pattern emerges – high market shares (often with low total floorspace) can be seen in the more rural areas. In the more rural retail landscape in North and West Wales, for example, branded fascia convenience stores feature prominently and account for over 30% of total grocery floorspace in 2012. Similarly, other mainly rural postal areas, such as Aberdeen (AB), Durham (DH) and Plymouth (PL) are also strongholds for branded convenience grocery floorspace, with over 30% market share in each case.

(a)   **(b)**
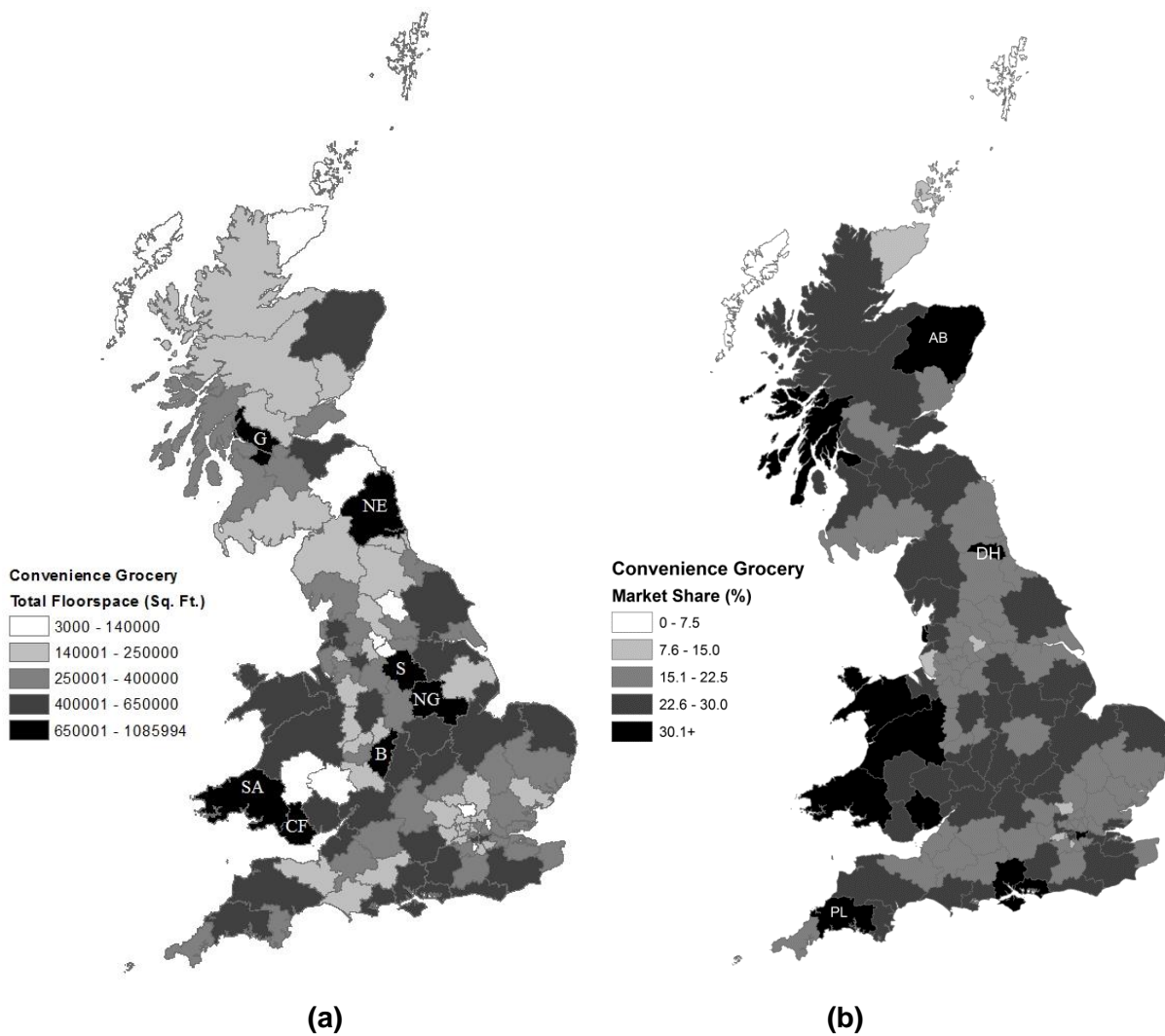
Figure 5.14 Convenience grocery (a) total floorspace (3000 sq. ft.) and (b) market share of total floorspace by postal area in GB, 2012. *This Figure is abstracted using data provided by GMAP Ltd and based on boundary data provided through EDINA UKBORDERS.*

5.5.1 The Co-operative Group

The leading UK major convenience retailer in 2012 was the Co-operative Group, a consortium of 22 different societies across the whole of GB. Although each has its own name we shall look at the combined market share under the banner of the Co-op, shown in figure 5.15 for both the convenience market and the total grocery market. The Co-op has historically made the greatest commitment to growth through small-format retailing. Following a strategic review in 1997, the Co-op opted to turn its attention to small-store grocery retailing due to an inability to compete with larger retailers in the superstore market (Bell and Hallsworth, 2003; Wood et al. 2006). Through this redirection, the Co-op became the first major grocery retailer to commit to the

convenience market. In 2002, the Co-op also acquired the Alldays brand of 600 convenience stores, becoming the largest convenience retailer among the major grocery firms in the UK, with over 2200 convenience stores. However, in 2012, the Co-op acquired 880 stores from Somerfield, expanding its portfolio of small to medium grocery stores (Finch, 2008). It could be argued that this signalled the retailer moving away from its earlier primary commitment to the small-store convenience market which was being increasingly squeezed by the presence of large retailers previously associated with large supermarket retailing.

The Co-op is well represented across GB for both the total grocery market and the convenience grocery market, given that many of Co-op stores are under 3000 sq. ft. The impact of the Somerfield purchase is evident in Figure 5.15b as Somerfield had traditionally been strong in Wales and the west of England. The Co-op comprises a consortium of different companies and the most powerful of these are the Co-operative Group (which merged with the second biggest Co-op 'United Co-op' in 2007), the east of England, the Midlands, Southern and Scotmid. This can be seen in the pattern of high market share seen in Figure 5.15b. As seen in Figure 5.15a, in 2012 the Co-op also had a large share of the convenience market in rural postal areas in northern Scotland. Moreover, the retailer had a large convenience market share in much of northern England, including north and west Yorkshire, Lancashire, north west England including Greater Manchester, and large parts of the south coast.

(a)   **(b)**
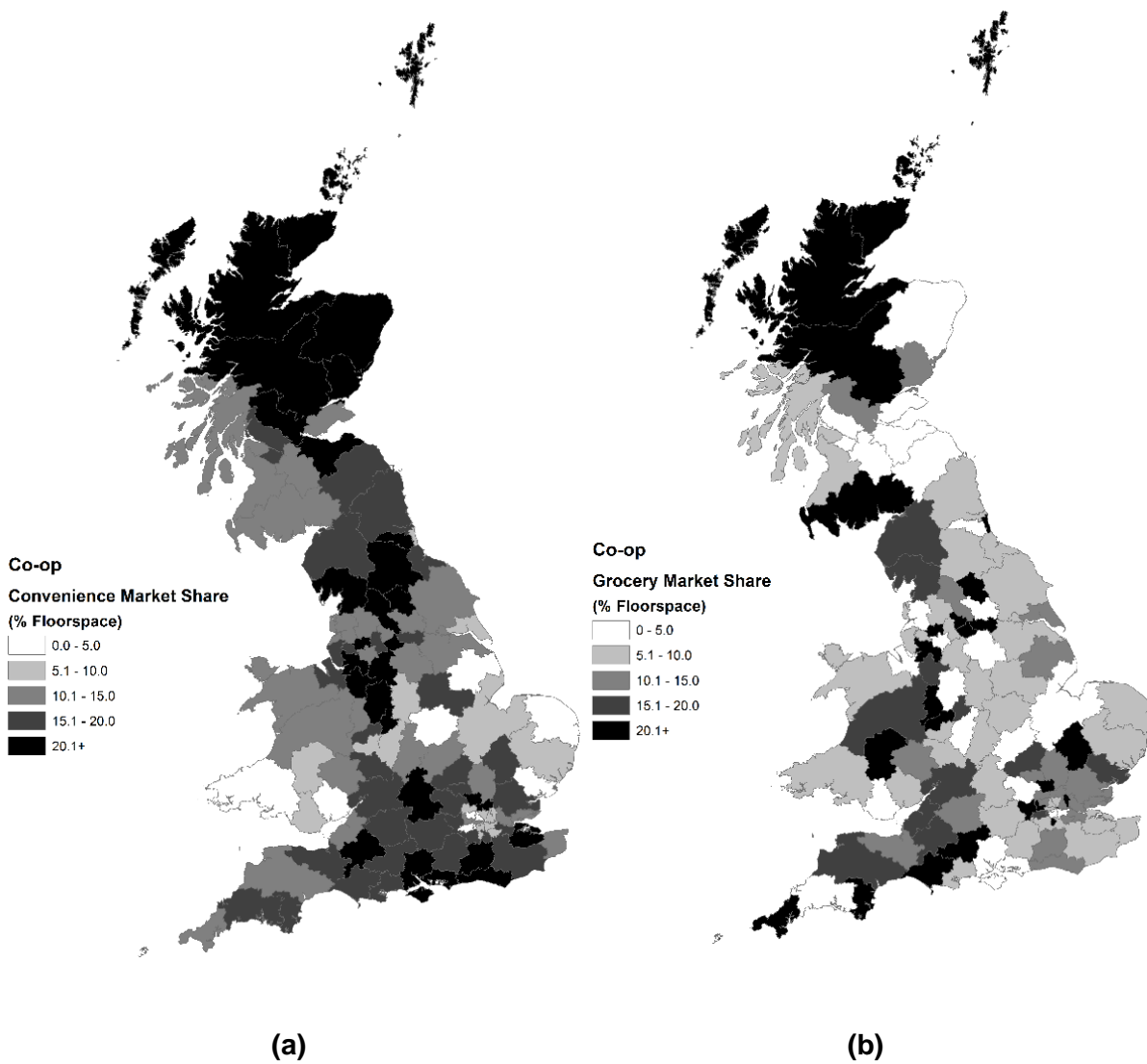
Figure 5.15 (a) Convenience market share and (b) grocery market share of the Co-operative group by postal area in GB, 2012. *This Figure is abstracted using data provided by GMAP Ltd and based on boundary data provided through EDINA UKBORDERS.*

5.5.2 Tesco

As quantified in chapter 5, Tesco and Sainsbury's are at the forefront of convenience retailing among large grocery retailers. In 1994, Tesco undertook its first foray into convenience store retailing through a joint venture with ESSO to open 'Tesco Express', branded convenience stores at petrol forecourts. This proved successful and the retailer continued to pursue convenience retailing through both forecourt and non-forecourt stores. Wood et al. (2006) argue that the competitive landscape of the convenience store sector was transformed in January 2003, when Tesco purchased 862 convenience stores from T&S Stores, boosting the total small-format stores operated by the retailer to around 1000. These stores retained the original One Stop

store branding under which they were previously trading. Additionally, Tesco acquired the London based convenience store chains Europa, Harts and Cullens in 2002 and in late 2010, Tesco's One Stop brand purchased the Mills chain of 76 convenience stores operating in the Midlands, South Wales and the North east of England, increasing Tesco's One Stop chain to 598 convenience stores. Thus by 2012, Tesco had a total of 1,946 convenience stores when combining the One Stop and Tesco fascias. Figure 5.16 shows the market shares of Tesco for both the convenience market and for the total grocery market.

(a)   (b)
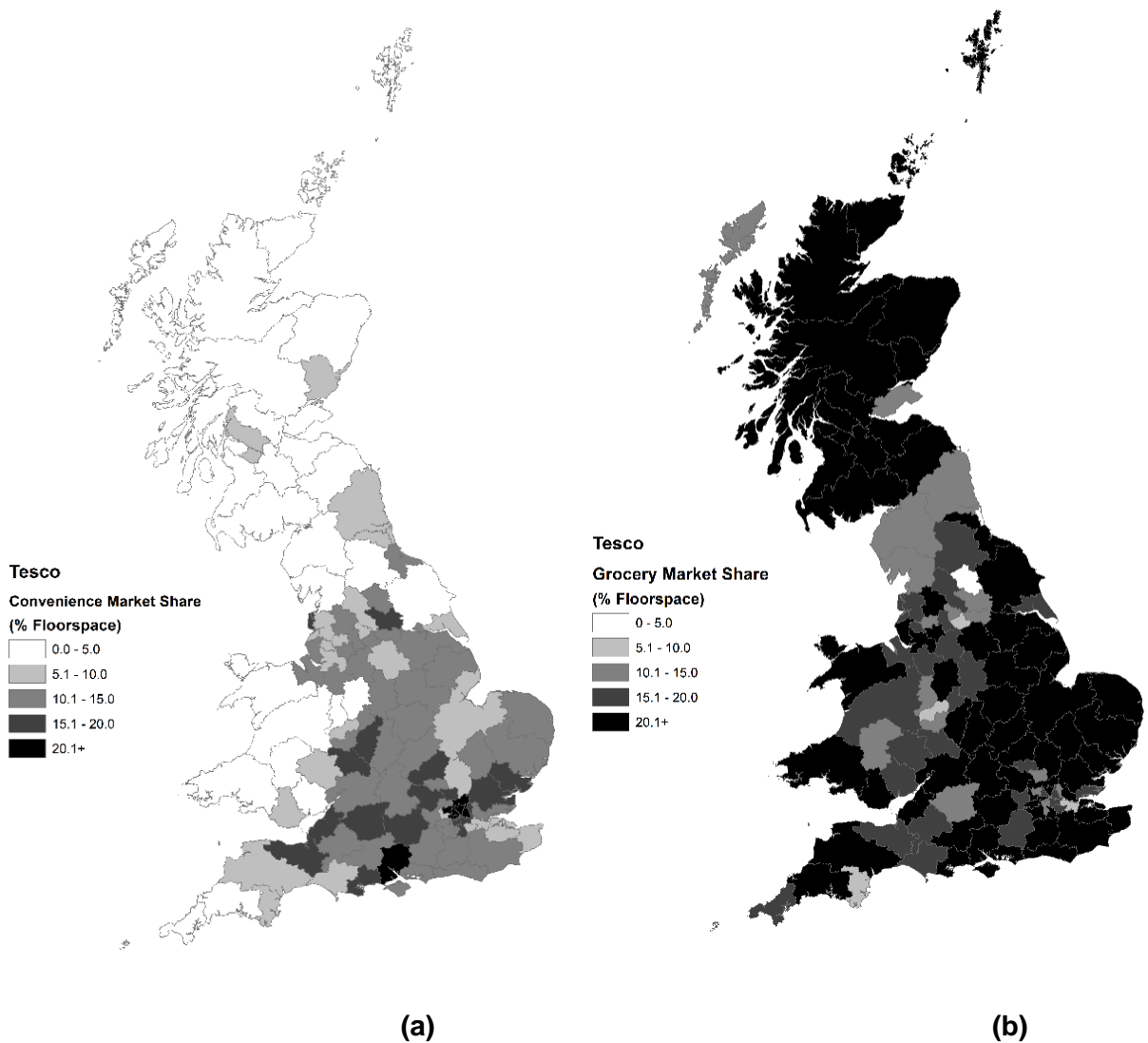
Figure 5.16 (a) Convenience market share and (b) grocery market share of Tesco by postal area in GB, 2012. *This Figure is abstracted using data provided by GMAP Ltd and based on boundary data provided through EDINA UKBORDERS.*

With around 30% of the total grocery market in 2012, Tesco is the most national of all UK grocery retailers in terms of spatial coverage. Tesco's convenience stores are more

clustered in terms of their spatial distribution. As Figure 5.16a shows, Tesco has its largest share of the convenience grocery market in the postal districts in the south of England, particularly around London and the South East. Conversely, Tesco has a relatively low market share in Wales, northern Scotland and north east England.

5.5.3 Sainsbury's

Sainsbury's piloted its first convenience grocery store format branded 'Sainsbury's Local' in Hammersmith in 1998. Following Tesco's acquisitions in 2003, Sainsbury's launched a series of rival acquisitions in 2004, acquiring 54 stores from Bells in the North East of England, 114 stores from Jacksons in Yorkshire and the Midlands, and 6 stores from JB Beaumont in the East Midlands. Figure 5.17 shows Sainsbury's share of the convenience grocery market and the

total grocery market

(a)   **(b)**
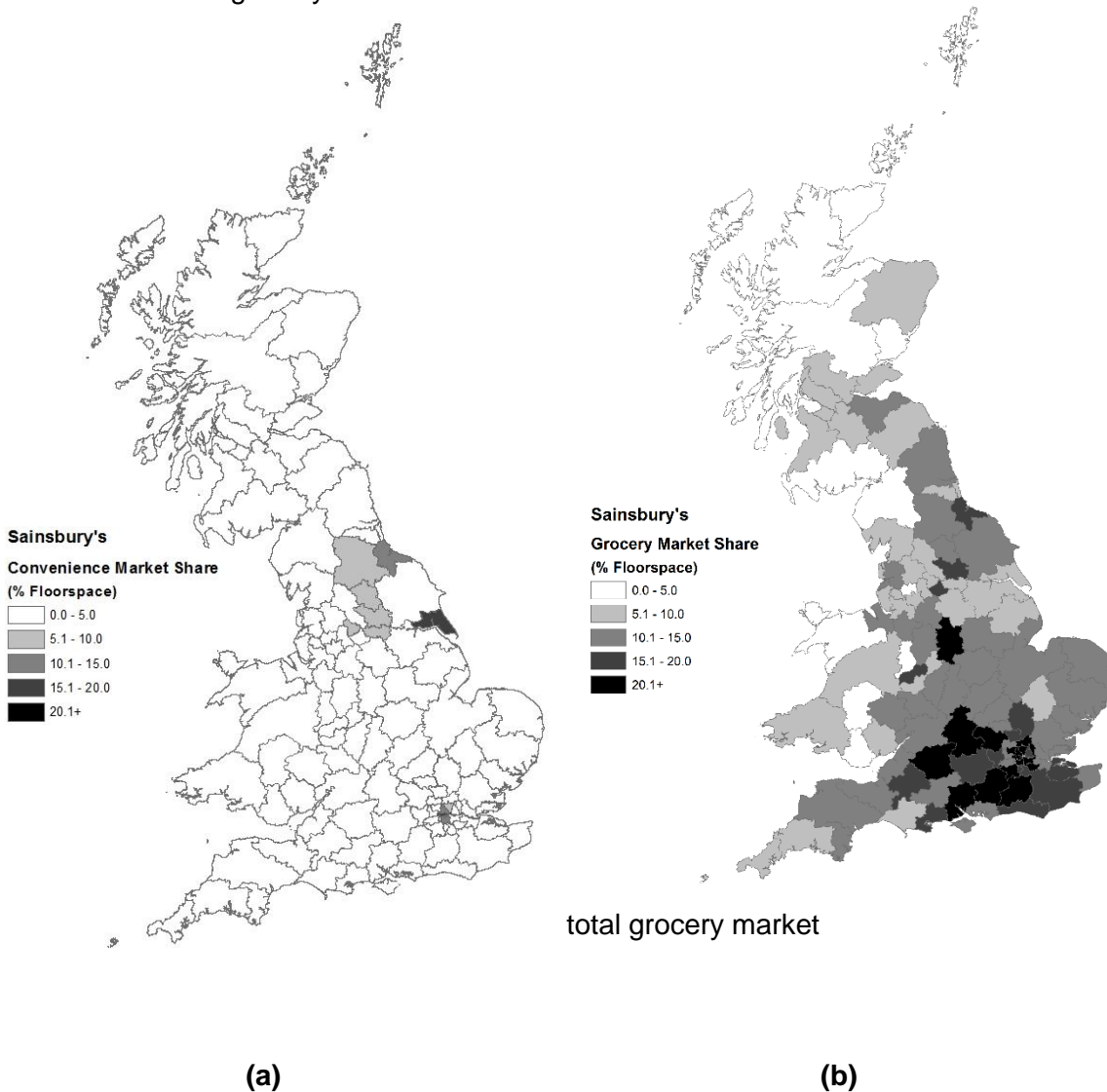
Figure 5.17 (a) Convenience market share and (b) grocery market share of Sainsbury's by postal area in GB, 2012. *This Figure is abstracted using data provided by GMAP Ltd and based on boundary data provided through EDINA UKBORDERS.*

Unlike Tesco, Sainsbury's market share of the total grocery market (Figure 5.17b) is highly clustered in London and the South East. In particular, Sainsbury's has a high market share in postal areas west of London, extending south towards the south coast of England. On the other hand, North Wales, the majority of Scotland and much of the North West and northern England area areas of low market share (under 10%) for the retailer. Although Sainsbury's has many convenience stores it makes little impact on market shares in any area except the postal districts in Yorkshire, the North East and central London, the former the result of the acquisitions described earlier in this section.

5.5.4 Symbol Groups

As detailed in the literature review in chapter 2, many of the symbol group retailers have become major players in the convenience grocery market. Musgrave group, Premier, Costcutter and Spar have all made significant inroads into small-format grocery retailing. Figure 5.18 shows the share of the convenience market of each of these symbol group retailers, providing insight into their different location patterns.

The Musgrave Group, a prominent symbol group retailer (based in Ireland) owns two of the largest symbol group operators, Londis and Budgens. Both these companies have their origins in London and the south of England. Therefore, it is unsurprising that together they have a large market share in much of southern England, particularly in central London and the postal areas south of London and much of the South West of England. Moreover, the retailer has a strong presence in Humberside and East Yorkshire, but a comparatively low market share in East and North Scotland and a relatively small market share in much of Wales.

Costcutter has a comparatively high market share in postal areas in the West Midlands and Yorkshire and Humberside. Costcutter bases its headquarters in Yorkshire, reflecting the firms' strong share of the convenience grocery market in the area. Comparatively, Costcutter has a low share of the convenience market in large parts of the South West and much of north west England. Premier, the largest symbol group retailer, has a large convenience market share across much of the UK. In contrast to other symbol group firms, the retailer has a high market share across the north of

England. Given the strength of Musgrave and Costcutter in London and the south east, Premier has a relatively low market share in and around London.
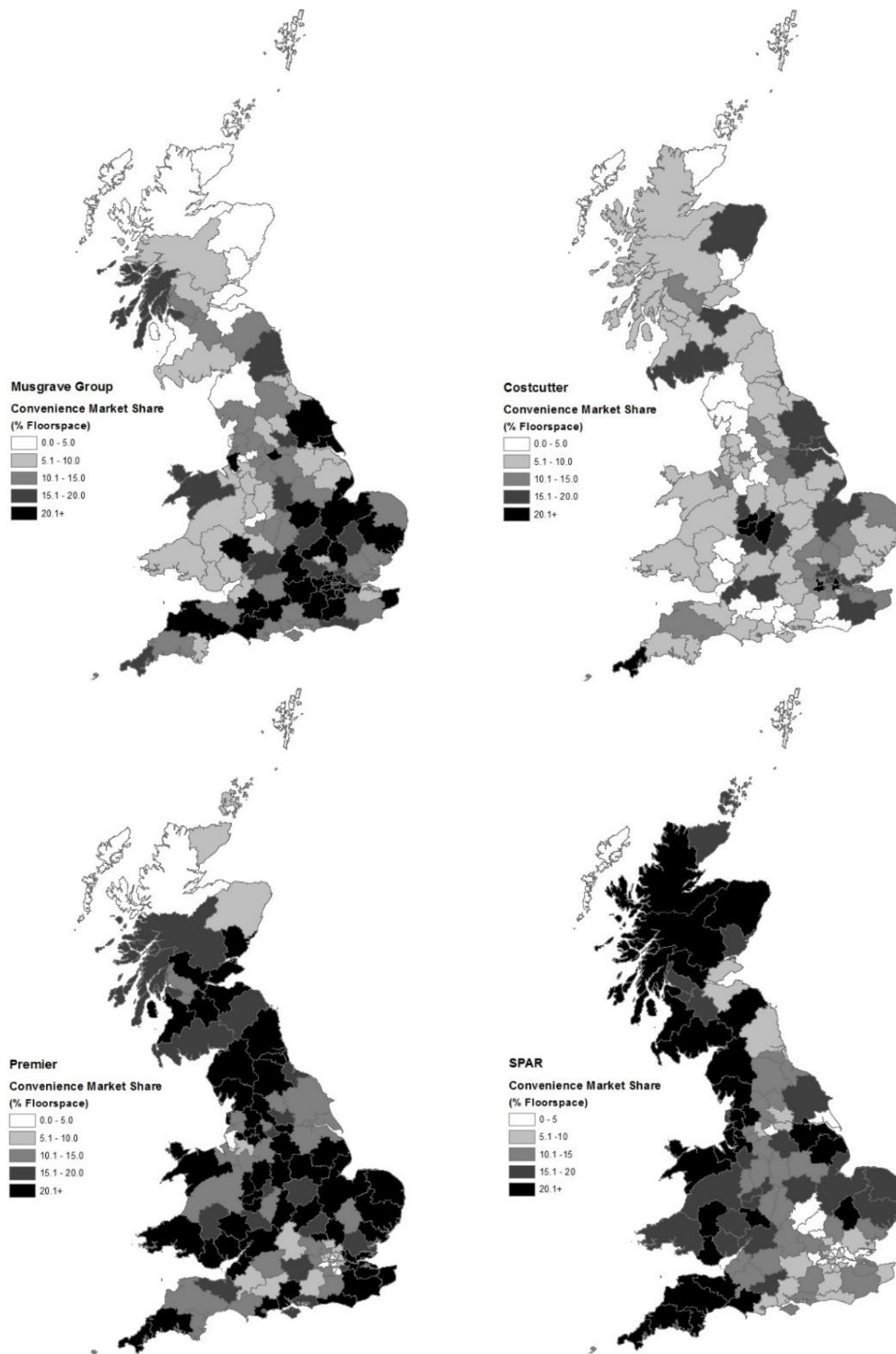
Figure 5.18 Convenience market shares of prominent symbol group retailers in GB, 2012. (a) Musgrave Group. (b) Costcutter. (c) Premier. (d) SPAR.

The final symbol group shown in Figure 5.18 is SPAR, a major international retailer with headquarters in Amsterdam. In 2012, SPAR had a high share of the convenience market in Scotland, North West England, North Wales and many parts of South West England. However, as with Premier, the greater competition in much of Southern England, particularly in London and the South East, means SPAR's market share is lowest in the south-east.

5.6 Chapter Summary

Chapter 1 set out the aims of this thesis, one of which was to quantify the extent to which major retailers have committed to the convenience grocery market and assess the geographical extent to which they play a role in convenience grocery retailing in Great Britain. This chapter has met this aim by quantifying the extent to which convenience retailing has become a part of the operations of the four largest grocery retailers in GB by analysing the change in stores in each size format operated by each of the retailers. The research has found that by 2012, grocery retailing in Great Britain had entered an age of increased convenience. It can be seen that through the combination of PPG6 and the Competition Commissions' two market ruling policy influences, town centre spaces have become increasingly considered by site location teams of major retailers. As these locations have more limited space, the choice to diversify into smaller format stores became attractive to major grocery retailers.

On the supply side, this research has found a marked shift in the choice of format by major grocery retailers between 2003 and 2012. Driven predominantly by Tesco and Sainsbury's full throttle pursuit of convenience, the dynamic of the major four retailers store formats on a national level has shifted towards a greater emphasis on small-format grocery retailing within the remit of the Sunday Trading Act. Between 2003 and 2012, convenience stores as a proportion of total stores increased by 40.1%, from 8.2% to 48.3% of total stores. Whilst Sainsbury's and Tesco diversified their store networks by growing in the convenience market, ASDA operationalised a strategy of 'space sweating' in which they extended existing supermarkets to become hypermarkets, often through the introduction of mezzanine flooring (Wood and McCarthy, 2013).

This chapter addressed the second part of the aim in attempting to assess the geographical extent by which major grocery retailers more traditionally associated with supermarket grocery retailing now play a role in convenience grocery retailing across GB. This chapter quantified the market share of each of the notable convenience

retailers by postal area level geography in GB in 2012, comparing this market share to each retailers total market share across all store formats in GB. The research found that the convenience operations of Tesco and Sainsbury's are more geographically concentrated than their total grocery offer, suggesting that they have specifically targeted certain areas to concentrate their convenience efforts in. This has created a concentrated spatial battle, particularly in large urban areas in which the major grocery retailers are vying for space against the more traditional convenience retailers.

On a regional level, the study found that branded convenience floorspace per capita in 2012 was significantly lower in Scotland and Wales in comparison to most regions of England. This may signal a potential for major retailers to grow their market shares through entering the convenience market outside of England more fervently. This may be due to both Sainsbury's and Tesco having strong traditional connections to the South and Midlands of England and having few operations based in either Scotland or Wales. The regional geography identified in this chapter is somewhat crude and amid evidence that a battle for space in the convenience market has taken place, it would be useful to identify the extent to which retailers are competing for the convenience retail market across GB at a smaller level of geography. This is investigated at the postal area level in GB in chapter 5.

The analysis in this chapter has identified a spatial battle for the convenience grocery market in a number of parts of GB. However, the analysis thus far is yet to take into account the micro locations chosen by retailers for their store networks in GB. The final aim of this thesis is to develop and test a series of predictive models for forecasting convenience grocery store revenue in the varying location types in which this type of grocery store is found. This research has hypothesised that convenience stores can be located in very different types of space – rural villages, city centre train stations, suburban town centres. Each of these different types of location could, in theory, require a different optimal methodology for sales forecasting. The next chapter segments locations in which convenience grocery retailing takes place in Yorkshire and the Humber, identifying the types of statistical locations chosen by each retailer to locate convenience stores and disaggregating the convenience grocery market by location type in preparation for evaluating a series of predictive models for forecasting convenience grocery store revenue.

Sainsbury's commitment to convenience grocery retailing has been mirrored in the company's internal structure. The location department of the retailer is now split into

two sections, supermarket and convenience, a situation mirrored at Tesco. This identifies the importance retailers are placing on making optimum location decisions in the race for small-format success. As a result, the presence of major retailers in convenience retailing *"… exposes smaller neighbourhood retailers to competition along with complex, efficient supply chains and a strong tradition in location management"* (Wood and Browne, 2007, P. 234). Major retailers have the luxury of location planning teams, a significant advantage over both smaller retailers and independent stores. However, other retailers such as Waitrose and Marks and Spencer have continued to invest in the convenience grocery and Morrisons re-entered the convenience market in later 2015 with opening of a 1200 sq. ft. Morrisons Daily convenience store at a Motor Fuel Group petrol station in Crewe (Ruddick, 2015). The persistent presence of major grocery retailers in the convenience market may continue to increase the pressure experienced by small and independent retailers traditionally associated with neighbourhood convenience grocery retailing.

In terms of retail planning policy, it is feasible that the grocery market will once again be referred to the Competition Commission in light of the increasing number of convenience stores operated by the large grocery retailers. As these retailers attempt to increase market share in convenience grocery retailing, a review of the two-market ruling distinguishing between supermarket and convenience retailing may occur. If the two-market ruling were to be changed, major retailer's convenience stores may be considered alongside their larger store formats. This may lead to retailers being forced to sell off part of their portfolio of convenience stores, further altering the dynamics of the convenience grocery market in GB.

Chapter 6
Segmenting the growing convenience grocery market for store location planning

Chapter 1 of this thesis highlighted the aims of this project. The primary aim of this research is to develop and test a series of predictive models for forecasting convenience grocery store revenue. Convenience stores have presented a challenge to major grocery retailers in GB in terms of estimating revenue for a number of reasons. These stores can be located in very different types of location – rural villages, city centre train stations, suburban town centres. Prior to commencing the development and empirical testing of varying methodologies of forecasting store sales it was assumed that different locations may, in theory, require a different optimal methodology for forecasting revenue accurately.

The analysis reported in chapter 5 identified the areas of the country that different convenience retailers had decided to locate their convenience stores in. The analysis considered the macro-geography of location and did not take into account the micro-locations of convenience stores. For example, it did not distinguish precisely between rural and urban areas or between residential areas and business areas. During the golden era for the large grocery multiples in GB, large grocery stores (in excess of 25,000 sq. ft.) became widespread. Locations for these stores were often limited by areas with available space on the outskirts of large conurbations. The catchment areas immediately surrounding large supermarkets are often statistically similar, as the stores were designed to generate trade from a relatively large area surrounding the store and weren't solely reliant on customers on the doorstep. These stores were often within reach of a number of residential areas from which people would be willing to travel further distances for larger stores offering a wider range of products. Moreover, many consumers would have travelled to these stores by car, making one large weekly grocery shop. Traditional models such as applied SIMs were effective in accurately capturing this type of behaviour.

Due to their small size and more limited product ranges in comparison to large supermarkets, convenience stores do not attract trade from as great a distance as larger supermarkets as customers are less likely to make a specific car trip to shop at convenience grocery stores. Convenience grocery stores (particular those operated by the major grocery multiples) are generally of a standard format, with the majority falling

between 2000 and 3000 square feet in size, restricted by the 3000 square feet limit to allow extended Sunday opening hours (Baron et al. 2002). These stores are located in areas with more limited space and in catchments that can vary distinctly. This research hypothesised that the immediate local characteristics of the area surrounding convenience stores will on average have a greater impact on revenue potential of a site than the immediate areas surrounding larger supermarkets.

6.1 Producing a typology of convenience stores

This chapter reports on a segmentation of the convenience grocery market in Yorkshire and the Humber, statistically distinguishing between different store location types. The analysis in this chapter serves two purposes. It accompanies the macro location analysis presented in chapter 5 and explores the types of micro location in which different types of retailers have chosen to locate. Secondly, the segmentation provides a starting point for testing the effectiveness of different modelling methodologies for estimating revenue for different locations in which convenience retailing takes place. The discussion of the varying location types in which convenience grocery stores are located has been largely anecdotal and descriptive in nature thus far in this thesis.

However, it is has been demonstrated that stores with different catchment characteristics will have differing customer profiles which will be borne out in different missions (or purposes) for visiting the store. This will have an impact on the revenue potential of different store locations. This raises a number of important questions:

1. Is it useful to segment the locational types of convenience stores?
2. Do locational types of convenience stores differ significantly?
3. What implications does this segmentation have for predicting convenience grocery store sales

The preceding chapter demonstrated that convenience stores are both geographically widespread and found in differing types of location in Great Britain. It has become apparent that convenience store sales have been more difficult for all the major retailers to predict, certainly with the same accuracy levels normally expected for superstore predictions. Wood and Browne (2006, 2007) discuss the importance attached to analogue techniques, site visits and gut feeling. They also imply that the more sophisticated modelling techniques (cf. Birkin et al 2002.) will not work at such micro spatial scales: 'forecasting convenience stores sees the traditional techniques of

market analysis for large scale food stores become largely redundant' (Wood and Browne 2007, 353). Further work is needed to test whether that statement is true – but, at the very least, this thesis argues that different techniques need to be considered for different locations of convenience stores. This thesis evaluates the use of different methods for different location types in order to assess the extent to which this statement is true.

Guy (1998) discusses one of the major issues in much of the literature surrounding the classification of stores by size, a popular method of classifying retail outlets. In the midst of the major retailers expanding through the opening of large supermarkets, research has often neglected smaller stores or placed them under the category of 'small shop'. Often in the UK that means that all grocery stores between 3,000 and 10,000 square feet in size are termed small supermarkets and those stores falling under 3,000 square feet are termed convenience grocery stores. However, as highlighted earlier in the previous chapter, these stores appear in a variety of locations that have been hypothesised to rely on different drivers of trade in generating sales. Thus it would be useful to offer a classification of convenience stores which splits the sector by store location. In this section we provide a segmentation of the network in the Yorkshire and Humber region of the UK; a classification based on 1185 branded convenience grocery stores. The classification presented in this research adopted a cluster analysis to generate groups (clusters) of similar convenience stores.

6.2 Classifications in geography and consumer research

The classification presented in this research, postulated as a classification of convenience stores, is actually a classification of the social, demographic, economic and environmental characteristics of immediate areas (or catchment areas) surrounding convenience stores. There is a long history to the development and application of area based classification systems in geography and related disciplines which have sought to make sense of areas and their environments. Area based classification in the social sciences is generally focused on placing areas into groups based on socio-economic characteristics.

Vickers et al. (2005) highlight the iconic work of Charles Booth conducted between 1886 and 1903 surveying the life and labour of the people of London. The maps produced from his work classified areas of London by their socio-economic characteristics and infamously characterised the poorer people living in some areas of

the capital as 'vicious, semi-criminal persons'. The classification of the people was attributed to the area in which they lived and was thus an area-based segmentation of London. In a review of data clustering methodologies, Jain (2010) stresses the importance of the purpose of the clustering of data to the method by which it is clustered. Booth's work in London attempted to identify the extent of poverty in London which he believed to be less widespread than reported. The result of Booth's classification in fact revealed that poverty was far more widespread than anyone had previously thought (Simey and Simey, 1960). Dennett (2010) highlights a further classic example of area classification in geography, that of Burgess (1925) classifying areas of Chicago differentiated by the process of urban expansion.

Many of the clustering algorithms applied to retail research have centred around two broad themes; first, identifying homogenous groups of consumers within a larger population and secondly, segmenting the market of stores to gauge some idea of strategy or performance. (Punj and Stewart, 1983). Work in the field of area classification has continued to the present day and now the area based segmentation of locations by socio-demographic and socio-economic characteristics falls under the banner of 'geodemographics'. Dennett (2010) argues that a convincing case can be made for commercial interest being a significant driver of this type of research in recent times. A number of commercial organisations have recognised a lucrative market in defining areas by the characteristics of the population that live within them. CACI's Acorn classification (A Classification of Residential Neighbourhoods), Experian's Mosaic Classification, Axciom's Parsonik classification and Callcredit Information Group's CAMEO classification are examples of private sector geodemographic classifications. These classifications segment consumers (by area) into groups and the classifications and associated insights are frequently sold to other businesses as an aid in decision making processes on customer-facing activities. The classifications are marketed as an aid to the understanding of customer preferences and trends, often linked to their day to day behaviours including shopping habits.

From a retail store location perspective, the marketing of these classifications is often centred on knowing the characteristics of existing or potential store catchment areas. Callcredit Information Group posits this question on their website; *"How do you know you're positioning your retail outlets in the right place"* (Callcredit Information Group, 2015) - Similarly, Experian MOSAIC identifies a benefit of MOSAIC in the ability to … *"Find optimum locations of new stores by understanding catchment profiles?"* (Experian, 2015). Retail businesses (including the major supermarkets discussed in

depth in this research) often purchase classifications such as those discussed above in order to aid in the understanding of their customers and/or potential customers in different localities. It is widely acknowledged that geodemographic classifications have been extremely successful as tools for market analysis, used by a significant number of business service firms (Vickers and Rees, 2011).

Recent years have also seen a renewed government interest in area classification in the UK, often based on data collected in the national census of the population. Rees et al. (2002) provide a comprehensive summary of census-based area classification typologies. Of particular importance to this research is a joint project between the School of Geography, University of Leeds (and latterly University College London) and the Office for National Statistics (ONS). The 2001 census-based ONS Output Area Classification (OAC) is a typology of output areas - the smallest census boundary geography - taking into account over 40 census variables and giving each output area a classification type or label based on the characteristics of the population within it. The classification was the first freely available small scale area level geodemographic classification of the UK and groups each of the UK's 223,060 census output areas into a cluster hierarchy of 7 (supergroups), 21 (groups), and 52 (sub-groups) (Vickers et al. 2005; Vickers, 2006; Vickers and Rees, 2011). This work was continued using the 2011 census of the UK with the development of the 2011 Output Area Classification by Chris Gates at University College London.

Guy (1998) discusses the general ways in which stores have been classified or segmented in retail research. Stores have been categorised by the type of goods sold with food retailers falling into the convenience good bracket. Furthermore, stores have been categorised by store type; in food retailing this has generally resulted in convenience grocery stores being set aside as being a different type of grocery store than larger supermarkets. This was indeed the way in which the Competition Commission's two market ruling distinguished between small and large grocery stores when investigating the dominance of major retailers in the grocery market. Moreover, further general segmentation of grocery stores has often been made on the grounds of ownership. This classification of food stores is evident throughout this research as the operations of retailers has been discussed in the context of ownership, with outlets being operated in one of three broad ways; by major grocery retailers, as part of a symbol group umbrella organisation or as an independent store. More bespoke segmentation centred around the effects of specific characteristics on store performances has been conducted by researchers such as Spiller and Lohse (1997)

looking at online store performance in relation to online presence, Benedict and Wedel (1991) assessing the effect of store image on store sales, Day and Heeler (1971) assessing the contribution of a number of store characteristics to revenue from the sale of a convenience food product and Harrigan (1985) investigating the potential use of cluster analysis in assessing competing retail firms.

6.3 Disaggregating the convenience grocery network

"The goal of data clustering, also known as cluster analysis, is to discover the natural grouping(s) of a set of patterns, points or objects" (Jain, 2010, p. 652).

The natural groupings of store locations in this cluster analysis are investigated in the context of identifying the likely drivers of trade to convenience grocery stores for different locations. This involves identifying both the type of trade that a store may experience, which may be residential, workplace, visitor or passing trade, or any combination of the four and the volume of this potential trade. Cluster analysis is prevalent in any discipline involving the analysis of multivariate data and its ability to compress large volumes of data into more manageable chunks of information with added value is the reason for this research adopting this method.

Guy (1997) highlights two reasons as to why it may be necessary to cluster retail outlets. First, he suggests that a good classification should assist in systematic and well informed discussions in retail research. The introduction to this chapter discussed the different types of location in which convenience stores are found. A statistical clustering of stores into different types would aid (and help in quantifying) the type of locations in which retailers have opted to engage in convenience grocery retailing. Secondly, Guy (1997) argues that a consistent classification system allows for comparing and contrasting empirical findings. The classification system presented in this research has the intention of improving the ability to assess sales performance of stores through enabling the ability to compare store sales performance to that of statistically similar stores. Second, it is used to aid the development of predictive models by understanding which modelling techniques are the most effective at predicting sales in each type of location in which convenience stores are found.

6.4 Methodology: K-means cluster analysis

The convenience stores were segmented or clustered using K-means cluster analysis, a widely used clustering method in the social sciences that has been adopted in many previous research segmenting both consumer types and store networks. The method is an iterative partitioning method (Aldenerfer and Blashfield 1984) involving separating into k classes *"… two-way, two-mode data (that is, N objects each having measurements on P variables) …"* (Steinley 2006, 1). Clustering algorithms can be broadly split into two types; partitional and hierarchical. Hierarchical clustering algorithms can either be agglomerative - starting with each data point in its own cluster and merging most similar clusters - or divisive, top-down clustering starting with all points in a single cluster and subsequently dividing one overall starting cluster into smaller clusters. Partitional algorithms, in which K-means cluster analysis lies, simultaneously partition data producing a complete set of clusters.

6.4.1 K-means algorithm

The k-means clustering algorithm is comparatively simple and works as follows in its SPSS implementation (Everitt et al. 2001).

1) Choose an initial grouping of objects into the desired k clusters; compute the means for the groups over all variables and the sums of squared deviations of objects from group means.
2) Move each object from its own group to each other group and re-compute the sums of squared deviations (the clustering criterion).
3) Choose the change which leads to the greatest improvement in the clustering criterion.
4) Repeat steps 2 and 3 for all objects until no transfer of an object to a new group results in improvement in the clustering criterion.

Distance in this instance denotes the within cluster similarity and is calculated using the formula:

$$Distance\ (or\ similarity) = \sqrt{\left[(CC_X - O_x)^2 + \left(CC_y - O_y\right)^2 + \cdots\right]}$$

(Equation 6.1)

where CC_X and CC_y are the cluster centres of a variable $x\ and\ y$ and $O_x\ or\ O_y$ are the observed values for that variable from a given case. The algorithm attempts to

143

minimise the distance for all observations belonging to a cluster and maximise the distance between cluster centres. The most common distance function applied in K-means cluster analysis (and that which is applied in this research) is Euclidean distance which measures the straight line distance between a point $x(x_1 x_2 \ldots x_n)$ and a point $y(y_1 y_2 \ldots y_n)$ (Vickers, 2006). This involves computing the square root of the sum of squares of the distances between corresponding values, an extension of the Pythagoras theorem (Gordon, 1999).

6.4.2   Number of Clusters (K)

One challenge that presents itself in all K-means cluster analyses is the decision on the number of clusters, which is the responsibility of the researcher (Punj and Stewart, 1983). There are several different rules of thumb that have been suggested. However, these can be contradictory and different solutions can be applicable in different situations. Vickers (2006) argues there is no correct answer to the selection of the number of clusters and the solution should be judged as much on its usefulness as being a 'correct' representation of patterns in the data. Examples of rules of thumb applicable to portioning clustering algorithms include (after Vickers, 2006, 68-69):

1) If you can't choose between two solutions then the larger number of clusters should be selected.
2) Select the solution which has the most suitable number of clusters for purpose.
3) Select the solution which is most homogeneous in terms of the number of objects within each cluster, for example the solution which has the smallest difference between the number of objects in the smallest and largest clusters.

In the convenience grocery classification presented in this chapter, both 6 and 7 cluster solutions were effective in partitioning the data into discrete clusters that could be taken forward to aid in understanding of store sales performance. It was decided that a 7 cluster solution was the most effective and conformed to two of the rules of thumb listed above: namely rules 1 and 3. Furthermore, the clustering solution decided on was tested against its ability to partition the market for the purpose of more accurately predicting convenience store sales, therefore serving its purpose and meeting rule of thumb 2. The clustering process by which this was achieved is discussed in the next section.

6.4.3 The clustering process

The cluster analysis in this chapter used a multiple step process involving;

1) Defining the number of clusters initially.
2) Running the k-means clustering algorithm.
3) Assessing the output for similarity of stores within clusters and the distinction between clusters.
4) Deciding which clusters made logical sense statistically and geographically.
5) Re-running the clustering algorithm for the remainder of stores and repeating through the steps until a distinct set of clusters was reached that made sense in the retail landscape. This method allowed us to identify the distinct characteristics that defined each cluster.

6.4.4 Variables in the clustering algorithm

A number of variables that retailers consider when evaluating a store location were considered for the clustering analysis. The theory behind the classification lies in attempting to identify the major demographic and environmental drivers of sales across different locations in which small grocery stores are present.

6.4.5 Socio-demographic variables

The distribution and characteristics of the population play a role in the likelihood of consumers spending money in convenience stores for different locations. Residential population and social class are the variables most readily fitting into this category of drivers of trade. Other things being equal, a large residential population around a store is likely to drive increased trade. Additionally, daytime population was also included as a variable in the segmentation. This variable is likely to be a key driver of trade for a number of convenience stores in which catchments are dominated by a work based population and customers are using the store due to a different shopping mission than residential based stores. Customers are likely to use these stores to purchase lunch time goods such as sandwiches and small basket top up shops on an evening. These can be seen as alternative customer missions to a purely residential catchment store in which customers are using convenience stores as a weekly shopping destination.

The social class variable was included as retailers are known to identify specific target populations (often based on affluence) when selecting a store location. Moreover, affluent consumers are more likely to spend more. Consumer preferences for specific retailers has been an important aspect of work conducted at the University of Leeds into predicting retail sales and market shares by Thompson (2013) and Newing (2013).

A number of income/deprivation related variables were considered for analysis. However, the proportion of population employed in social class 1 occupations was selected due to its link to money rich, time poor households in which convenience is more likely to be a higher priority in grocery shopping . Other income variables such as the 'Index of multiple deprivation' produced very similar results: hence social class 1 was retained as the main income variable.

6.4.6 Environmental and adjacency variables

The wider environment in a potential or existing store location will have a profound effect on the potential for that store to generate customers (and therefore revenue). The major environmental variables included in this study are proximity to railway stations, expressed in the form of total passengers entering and exiting a store catchment area.

The rail variable contributes to a store locations potential by both facilitating accessibility to the location and providing potential consumers passing through a location. In this respect it can be used as a measure of people present in a store catchment area that may not be captured by either the work based or residential population variables. Additionally, the retail store variable is designed as a measure of the wider attraction of a store location. An increase in other retail businesses in an area may increase the number of shopping tasks a consumer can complete when in a location, making an area more attractive and therefore likely to attract more potential customers to a store location. This variable will pick out stores that are present in large city centres and towns alongside major residential high streets likely to attract customers from a wider radius than standalone stores.

A final set of five variables were selected as they exhibit a lack of multicollinearity, but correlate with known sales data and link strongly with specific customer missions (Birkin et al. 2002). The five variables fit into the assessment of drivers of trade in their conformity with two broad themes; demographics and environmental. The final 5 variables are:

1) Residential population
2) Daytime population
3) Other retail attractions (shopping centres or parades of shops attracting further business),

4) Proximity to railway stations (which provide trade based on a major journey to work mode of transport and also helps to capture additional visitor demand for groceries)

5) Residential population by social class – the proportion of the population employed in higher social class occupations.

6.4.7 Generating data for each store location

Variables were generated by computing the value of each variable for each store in a 1 kilometre buffer around each convenience store, deemed as the store's immediate catchment area of a store in this study. The variables were derived as point data in the case of rail passengers (co-ordinates of each station) and other retail businesses, and derived at the output area level for the residential population, daytime population and social class variables (the smallest area level data for which population data is available in the UK).

In order to derive the data in a one kilometre catchment around each store, MapInfo Professional 12.5 was used to conduct point in polygon analysis (in the case of the rail and retail point data) and polygon in polygon analysis (in the case of the OA level population and social class data). There were no issues in the computation of the point in polygon analysis; however, the polygon in polygon analysis required the overcoming of an aggregation issue associated with overlapping polygons.

The output area polygons contained entirely within the 1 kilometre buffer present no problem in the aggregation process. However, output areas which fall partially within the 1 kilometre buffer around a store present an issue around interpolation. Do we include the whole output area and its data in the aggregation? Do we discount the whole output area from the aggregation? Do we include part of the output area in the aggregation? This research used the proportion of the area of the output area polygon that is within the 1 kilometre buffer. This was done by multiplying this proportion by the total output area value for each variable. For example, if the residential population of an output area is 500 and 20% of the output area falls within the buffer, 100 people from that output area are included in the total buffer aggregation.

6.4.8 Variable standardisation

Standardisation of variables is particularly important when the units used to measure the variables differ (Dennett, 2010). In the cluster analysis presented here, three of the variables are measured in numbers of people, one variable is measured in number of

shop units and the final variable is measured as a percentage of the total population aged 16 to 64. Furthermore, it is also useful to standardise variables when the ranges in which they fall differ (Steinley and Brusco, 2007). The rail variable in this classification differed from a value of 0 to almost 25 million passengers per annum, whereas the social class variable ranged between 0 and 100%, widely differing ranges. For these two reasons, variable standardisation was applied as part of this classification.

There are a number of available methods to standardise data before conducting cluster analysis, of which no general consensus is present in the literature. However, as Dennett (2010) cites, a number of researchers highlight the work of Milligan and Cooper (1998) in which they argue that the most effective way to standardise data is to do so as a function of the range of the data for that variable. This method was adopted by this study and can be expressed as:

$$Z_i = \frac{[X_i - \min(X)]}{\text{Max}(X) - \text{Min}(X)}$$

(Equation 6.2)

where:

Z_i = the standardised value for a variable for area i,

X_i = the value of variable X for area i,

Min (X) = minimum value of variable X for all areas, and

Min (Y) = minimum value of variable X for all areas.

6.5 The Final Clusters

The following section summarises the key findings of the cluster analysis, looking at the characteristics of the catchment areas of each cluster, mapping and analysing the geographic location of stores in each cluster and looking at store location for different retailers across the region. Table 6.1 identifies the distribution of stores among the 7 clusters. Moreover, table 6.2 identifies variable z-scores for each cluster. Examining cluster z-scores allows us to statistically distinguish between clusters, giving an indication of the retail environment in which stores in each distinct cluster operate. Additionally, figure 6.1 gives a visual representation of the cluster means in the form of a radar plot and allows for a comparison of cluster characteristics, whilst table 6.3

identifies cluster membership among the branded convenience retailers in Yorkshire and the Humber.

Table 6.1 Cluster means

Variable	Cluster (number of stores)							Global Average
	A (55)	**B** (161)	**C** (294)	**D** (50)	**E** (277)	**F** (182)	**G** (166)	
Residential Population	14565	17390	10612	8284	5674	6474	2056	8630
Workplace Population	41869	7154	4134	4336	1483	1400	641	4776
Train Passengers	8656036	185169	153209	469985	45848	15753	5968	498730
All Stores	371.4	48	29	60.8	10.0	10.5	6.0	38.3
Social Class 1 (%)	24.3	18.8	19.5	23.4	28.5	13.7	26.8	22.0

6.5.1 Calculation of variable Z-scores

One common method for standardising data is the calculation of z-scores. In this case they allow for an effective comparison of variable means for each cluster and take into account the global means for each variable and how the mean for each cluster relates to this global mean. If a z-score lies below 0, the cluster has a below average mean for that variable when compared to the average of all stores and if a z-score is positive it has an above average mean. Z-scores are measured using standard deviations and the equation to calculate them is as follows (Dennett, 2010, 143):

$$Z_i = \frac{X_i - \hat{X}}{\sigma_x}$$

(Equation 6.3)

where $\hat{X}$ is the global mean of variable X and σ_x is the standard deviation of variable X calculated as:

$$\sigma_x = \sqrt{\frac{(X_i - \hat{X})^2}{N}}$$

(Equation 6.4)

149

Table 6.2 identifies the z-score for each variable for each cluster. Moreover, figure 6.1 presents this in the form of a cluster wheel allowing a visual comparison of the variable scores in each cluster.

Table 6.2 Z-Scores for variables in each cluster

| Variable | Cluster (and number of stores) | | | | | | |
	A (55)	B (161)	C (294)	D (50)	E (277)	F (182)	G (166)
Residential Population	1.1	1.6	0.4	-0.1	-0.6	-0.4	-1.2
Daytime Population	2.5	0.2	0.0	0.0	-0.2	-0.2	-0.3
Train Passengers	2.5	-0.1	-0.1	0.0	-0.1	-0.1	-0.2
All Stores	2.5	0.1	-0.1	0.2	-0.2	-0.2	-0.2
Social Class 1 (%)	0.4	-0.6	-0.5	0.3	1.3	-1.6	0.9

Figure 6.1 Wheel of cluster z-scores
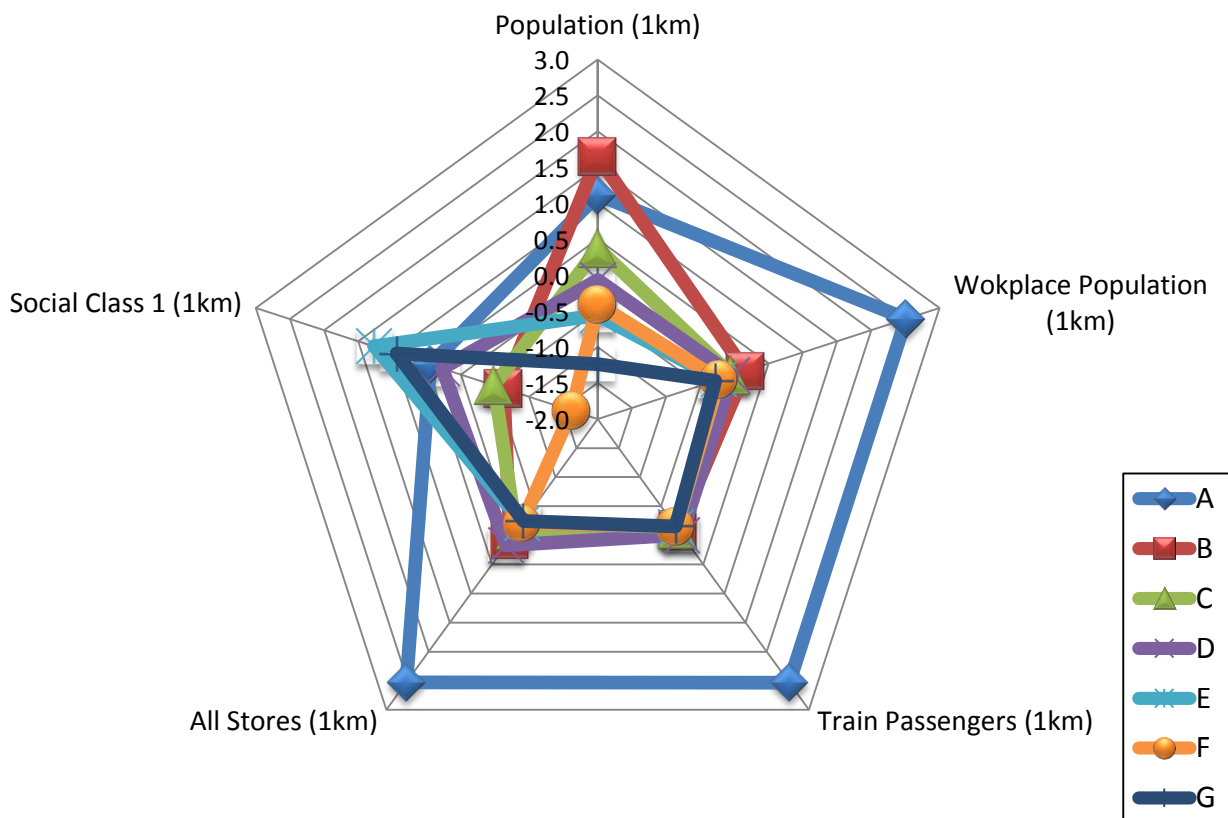
Moreover, the naming of clusters within a socio-demographic or market segmentation is an important aspect of cluster analysis. It allows the user to identify the type of group that is being dealt with and it must represent the characteristics of a given cluster as

accurately as possible. Table 6.3 identifies the chosen names of each location cluster and table 6.4 identifies the number of stores for each convenience retailer falling into each cluster type.

Table 6.3 Naming the clusters

Cluster	Cluster Name	Number of stores
A	Central Urban Cluster	55
B	Large Population Suburban Cluster	161
C	Smaller Population Suburban Cluster	294
D	Satellite Cluster	50
E	Outer Suburban Affluent Cluster	277
F	Outer Suburban Less Affluent Cluster	182
G	Rural Cluster	166

Table 6.4 Number of convenience stores in each cluster by retailer in Yorkshire and the Humber, 2013.

Retailer	Cluster Membership							Total
	A	B	C	D	E	F	G	
Alfred Jones	1	2	3	1	1	0	0	8
Asda Supermarket	0	0	0	0	0	1	0	1
Budgens	1	0	0	0	0	0	3	4
Co-Op	2	25	44	9	76	30	39	225
Costcutter	3	15	33	8	32	23	37	151
Farmfoods	1	1	0	2	0	0	0	4
Heron Frozen Foods	1	0	0	0	0	0	0	1
Iceland	3	0	0	0	0	0	0	3
JS Local	14	22	29	6	12	11	1	95
Londis	1	15	24	5	41	14	27	127
M Local	0	1	4	1	0	0	0	6
M&S Simply Food	2	0	0	0	1	2	0	5
Nisa	3	11	18	0	15	6	5	58
One Stop	3	9	18	1	10	18	6	65
Premier	5	33	67	9	42	51	6	213
Proudfoot	0	0	0	0	0	0	1	1
SPAR	2	14	25	6	34	16	37	134
Tesco Express	13	13	29	2	13	10	4	84
Grand Total	**55**	**161**	**294**	**50**	**277**	**182**	**166**	**1185**

6.6 Geography and characteristics of the final store clusters

This section details the characteristics of each cluster, their location geographically in the Yorkshire and the Humber former government office region and the types of retailers that are more or less prevalent in each of the location types. In doing so, the justification for naming the clusters becomes apparent.

6.6.1 Geography and characteristics of the central urban cluster A

The distinctive features of the central urban cluster are that daytime population, rail footfall and retail activity are very high among stores in this cluster, significantly larger than average. Figure 6.2 identifies the geographical location of the stores in the central urban cluster in Yorkshire and the Humber. When looking at the geography of stores in this cluster, they are located in large urban centres across the region. These include the cities of Leeds, Sheffield, Bradford, York and Hull, along with stores in the centre of a number of the large towns in the region, including Huddersfield, Doncaster and Harrogate. When compared to the average among all retailers, we can see that a greater than average proportion of convenience stores operated by Tesco and Sainsbury's are found in this central urban cluster. Over 15% of stores operated by the two largest retailers engaged in extensive convenience retailing fall into the central urban cluster, in comparison to just 2% of symbol group stores and 1% of Co-operative convenience stores.

Figure 6.2 Locations of stores in the central urban cluster A
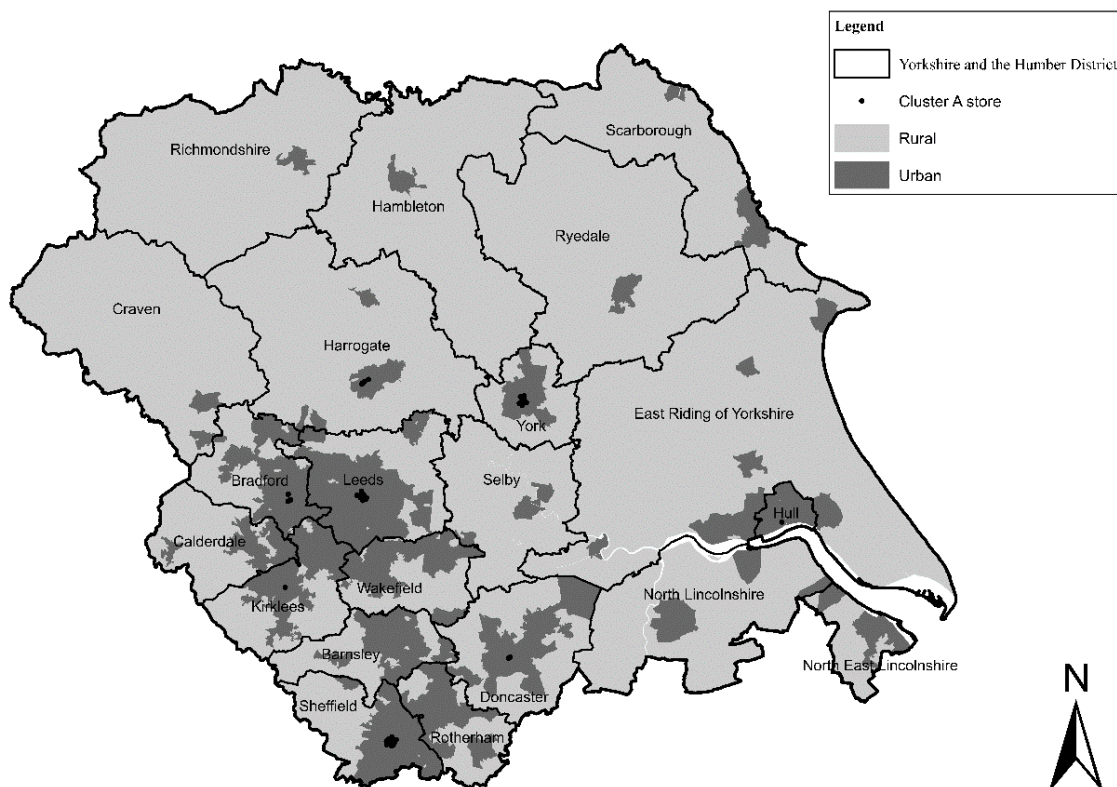
6.6.2 Geography and characteristics of the larger population suburban cluster

On average, the stores in this cluster are distinguished by having catchment areas with significantly above average residential populations and an above average daytime population and other retail outlets in the 1km catchment of these stores. In addition, a smaller proportion of residents employed in social class 1 live in the catchments of these stores. Figure 6.3 shows a map of this cluster showing the geographic location of stores in this cluster. These stores are located on the outskirts and suburbs of the major towns and cities inhabited by stores in the central urban location type. The divide is less stark when looking at company stores in large population suburban cluster: over 20% of stores operated by the major retailers fall into this category compared with around 15% of symbol group stores and just over 10% of Co-operatives.

Figure 6.3 Locations of stores in larger population suburban cluster B
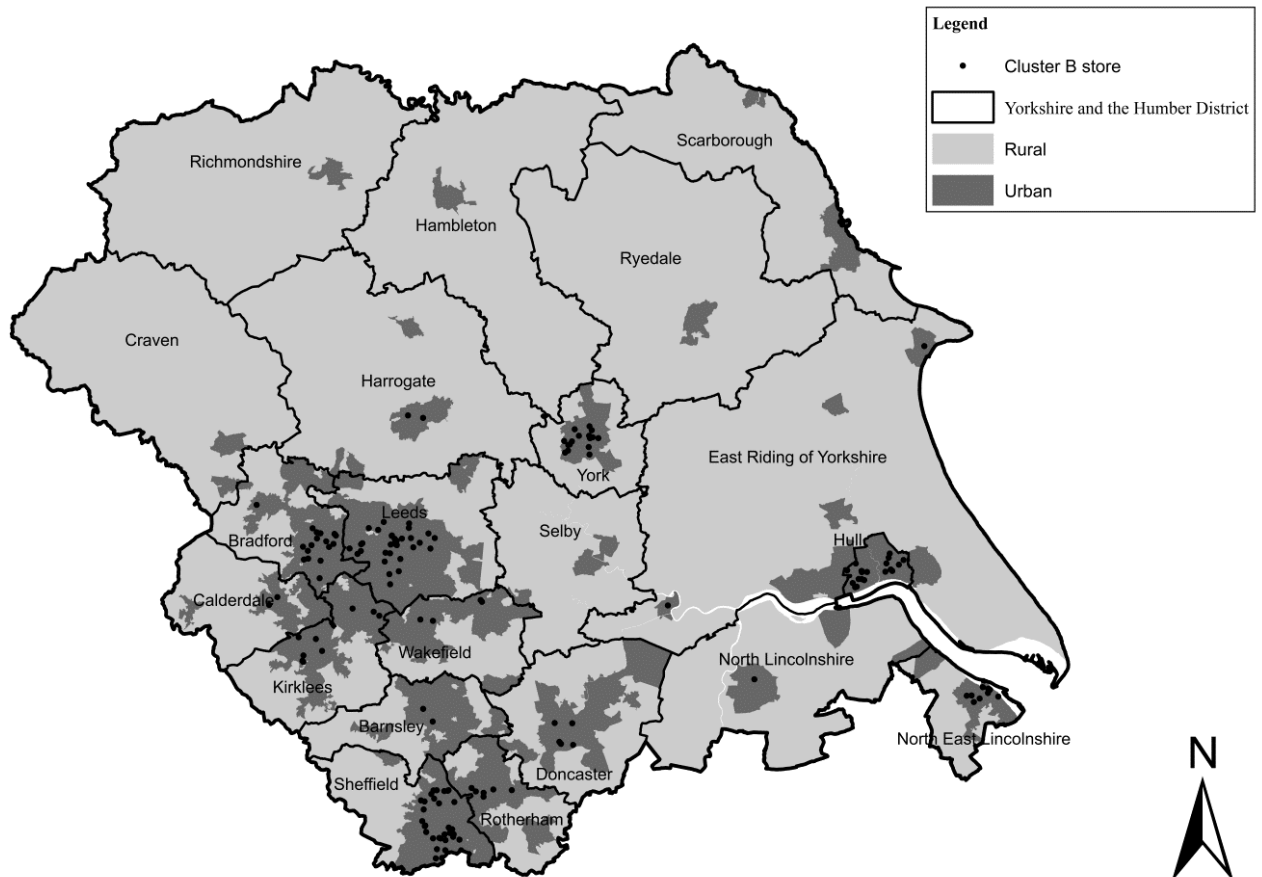
6.6.3 Geography and characteristics of the smaller population suburban cluster

These stores are located on the outskirts and suburbs of the major towns and cities inhabited by stores in the central urban cluster. Furthermore, some of these stores are located in the centre and large suburbs of other important, but smaller towns in the area such as Wakefield, Barnsley, Rotherham, Scunthorpe and Grimsby. Stores in this cluster have a smaller residential population and less other retail stores in the area than in the large population suburban cluster. This cluster primarily contains residential areas, with stores having an above average residential population but a below average daytime population, other retail stores, affluent residents and rail passenger volumes. This locational type is common among the major convenience grocery players: over 30% of stores operated by the major retailers fall into this category compared with around 20% of symbol group stores and just over 20% of Co-operatives. Figure 6.4 maps the stores in this cluster.

Figure 6.4 Locations of stores in the smaller population suburban cluster C
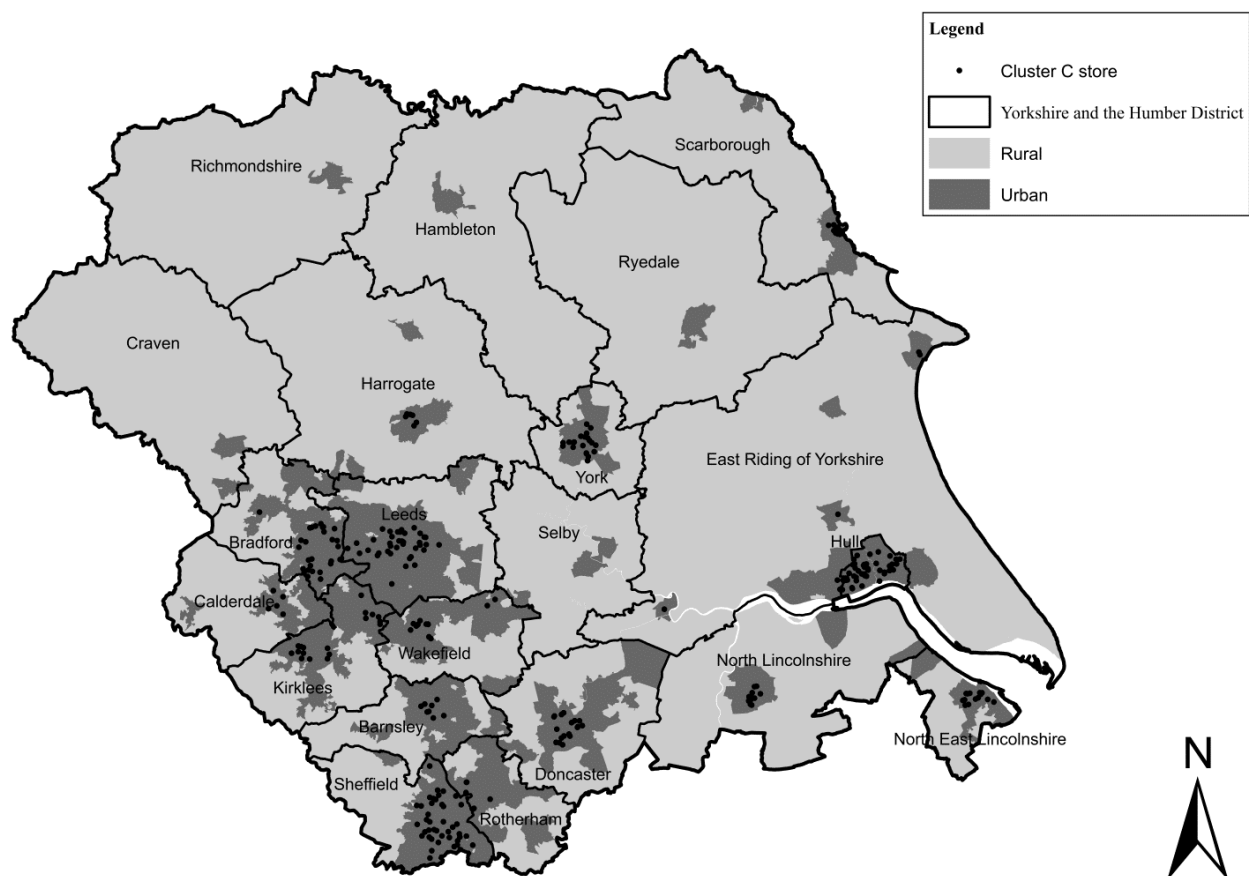
6.6.4 Geography and characteristics of satellite cluster

This cluster is distinguished by having, on average, the second largest average rail footfall among clusters, significantly higher than all clusters other than the central urban cluster. On average, stores in this cluster have a catchment area with a slightly below average residential and daytime population, an above average level of retail stores and proportion of residents in social class 1 occupations. Thus, they are more prevalent close to railway stations and in the suburbs of smaller towns. Figure 6.5 identifies these stores in small market towns such as Market Weighton in East Riding of Yorkshire, Northallerton in Hambleton district, Knaresborough in Harrogate District and Malton in Ryedale district. These stores are also found in towns close to the region's larger cities in places such as Castleford and Pontefract in Wakefield district, Garforth and Guiseley in Leeds district and Bingley and Keighley in Bradford district. As table 6.2 highlighted, the satellite cluster has the lowest number of stores of any cluster. Moreover, Sainsbury's stands out as the retailer with the largest proportion of stores falling in this category.

Figure 6.5 Locations of stores in the satellite cluster D.
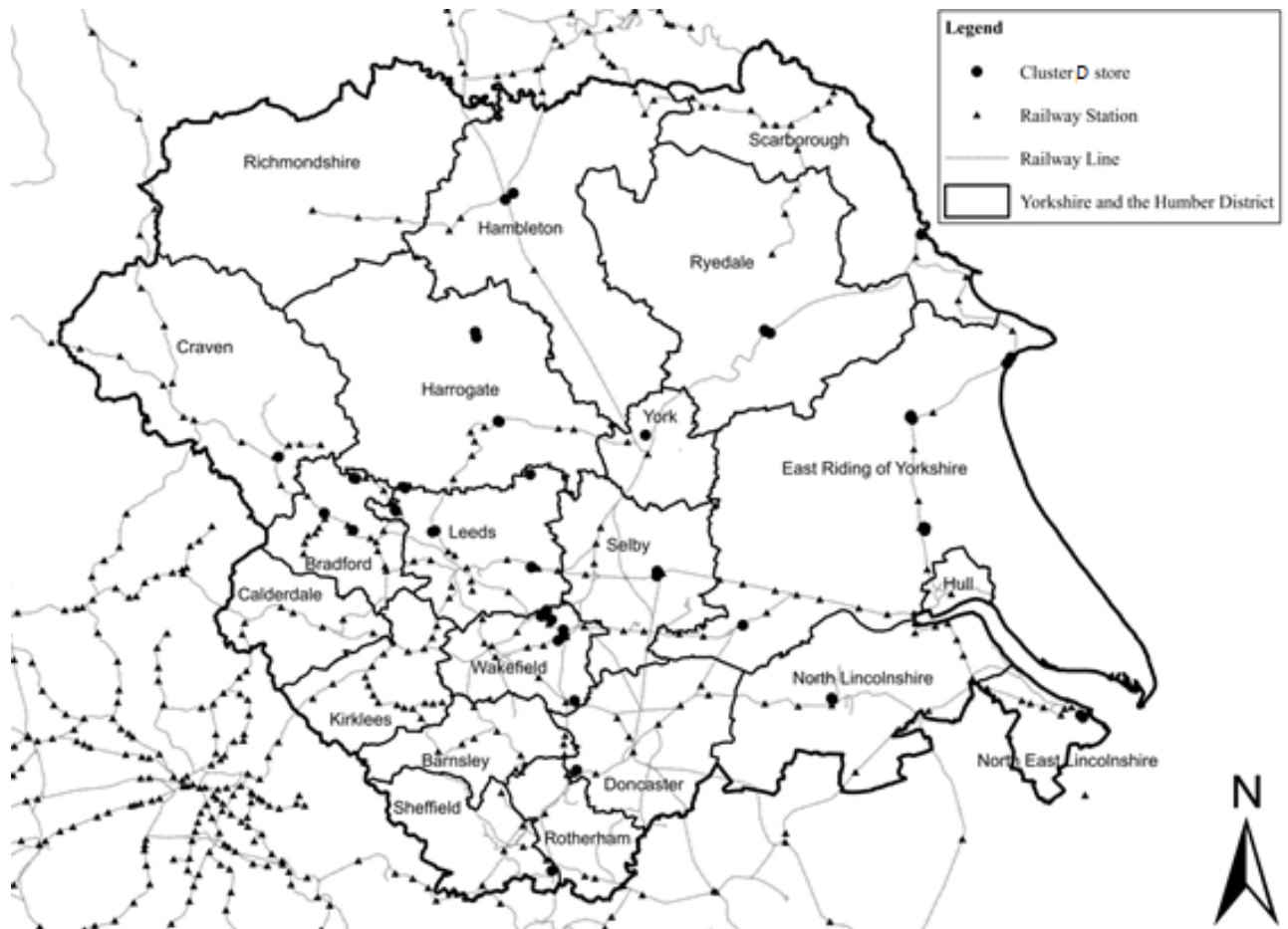
6.6.5 Geography and characteristics of the outskirts affluent cluster and the outskirts less affluent cluster

Clusters E and F are similar in many respects. Both clusters have a significantly below average population, a below average daytime population and lower number of neighbouring stores. However, outskirts affluent cluster E is distinguished by its significantly above average proportion of the population employed in social class 1 occupations, 28.5% of the population on average. This distinguishes it from outskirts less affluent cluster F, in which stores have on average 13.7% of the population in this type of occupation, less than half of outskirts affluent cluster E.

Figure 6.6 maps the location of stores in the outskirts affluent cluster and figure 6.7 maps stores in the outskirts less affluent cluster. The stores in the outskirts affluent cluster tend to be located, unsurprisingly, in the affluent outskirts of the larger towns and cities and in relatively affluent larger villages and small towns across the region. The outskirts affluent cluster locations are strongly favoured by the Co-operative group. Additionally, symbol group stores are significantly more likely to be located in less affluent cluster locations than major retailer stores. However, major retailer stores are still more likely, although less comprehensively so, to be located in the more affluent locations.

Stores in the outskirts less affluent cluster tend to be located in the less affluent outskirts of the larger towns and cities and in relatively less affluent larger villages and small towns across the region including much of the corridor running from South Leeds down the north and south east of Barnsley through into the northern outskirts of Rotherham (areas which struggled with major job losses in mining and manufacturing in the 1980s and 1990s). Additionally, these stores can be found in areas such as the outskirts of Doncaster and in south and south east Sheffield. When looking at membership of stores in the outskirts less affluent cluster F, symbol groups are the most likely of the groups of retailers to locate in the less affluent locations in Yorkshire and the Humber, although symbol group retailers still operate a considerably fewer number of stores in this cluster when compared to the more affluent outskirts locations.

Figure 6.6 Locations of stores in the outskirts affluent cluster E.
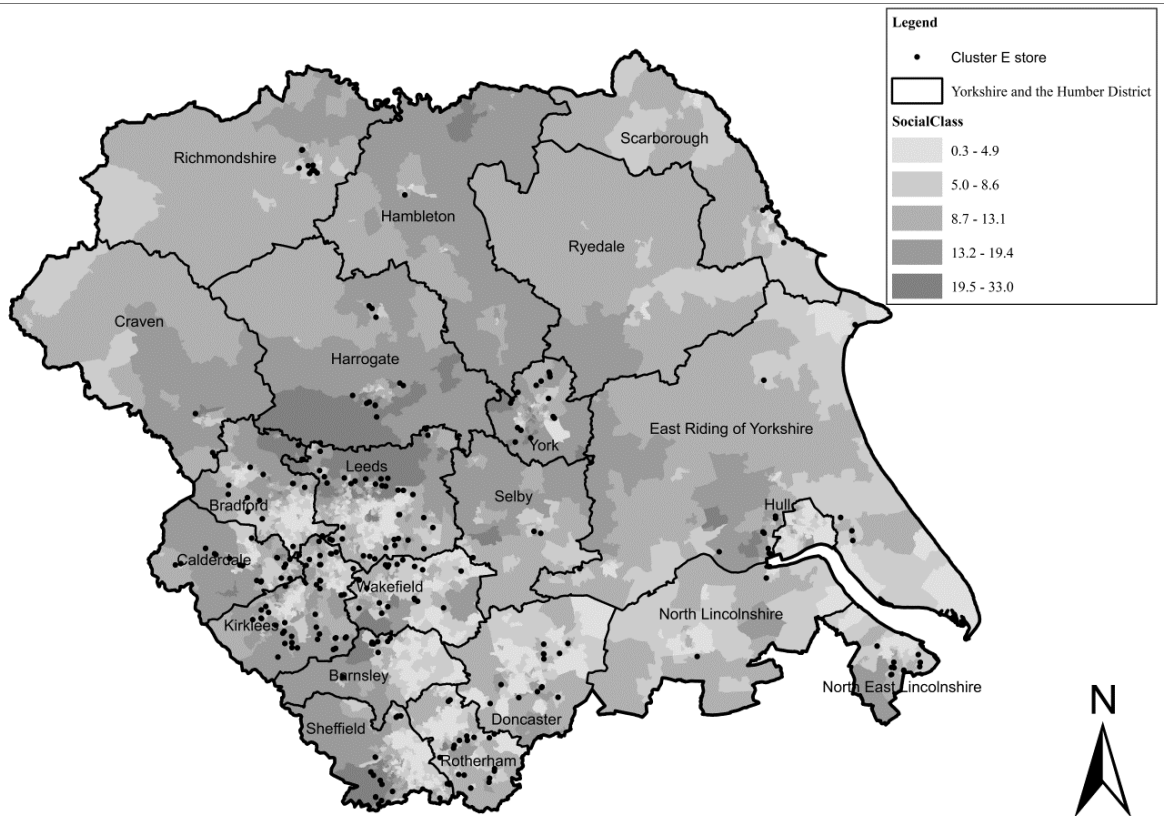
Figure 6.7 Locations of stores in the outskirts less affluent cluster F.
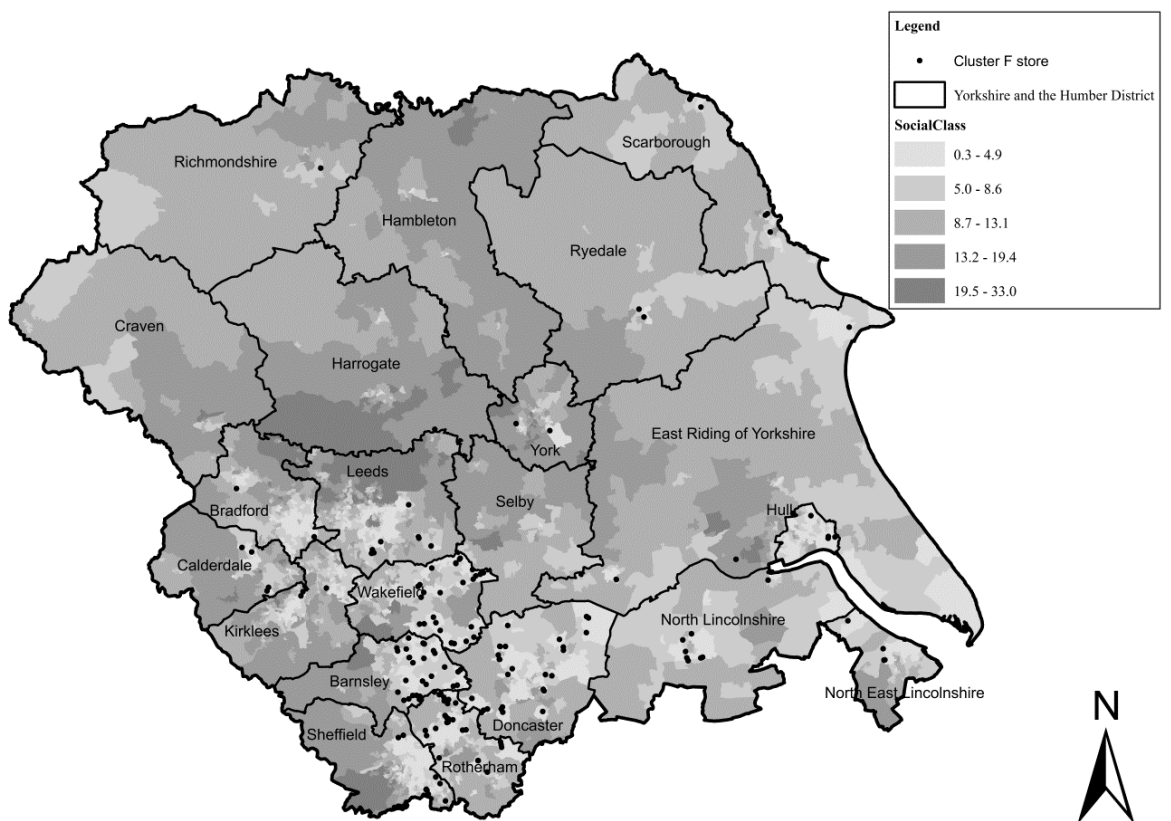
6.6.6 Geography and characteristics of the rural cluster

This cluster produces a distinct group of stores in outlying rural areas. The rural urban classification defines areas as rural if they fall outside of settlements with more than a 10,000 resident population. The 1km catchment area of all stores in the rural cluster meet these criteria. On average, 27% of residents in the catchment surrounding stores in the rural cluster are employed in social class 1 occupations, 5% higher than the average among the catchment areas of all convenience stores in the study. These stores are shown in Figure 6.7. They are prevalent in much of the more rural districts of Yorkshire and the Humber such as Richmondshire, Hambleton, Ryedale, Craven, East Riding of Yorkshire and North Lincolnshire and major retailer stores are considerable less likely to be located here. Fewer than 3% of Sainsbury's Local and Tesco Express stores are located in the rural cluster in comparison to over 17% of stores operated by symbol group retailers and the Co-operative group.

Figure 6.8 Locations of stores in the rural cluster G.
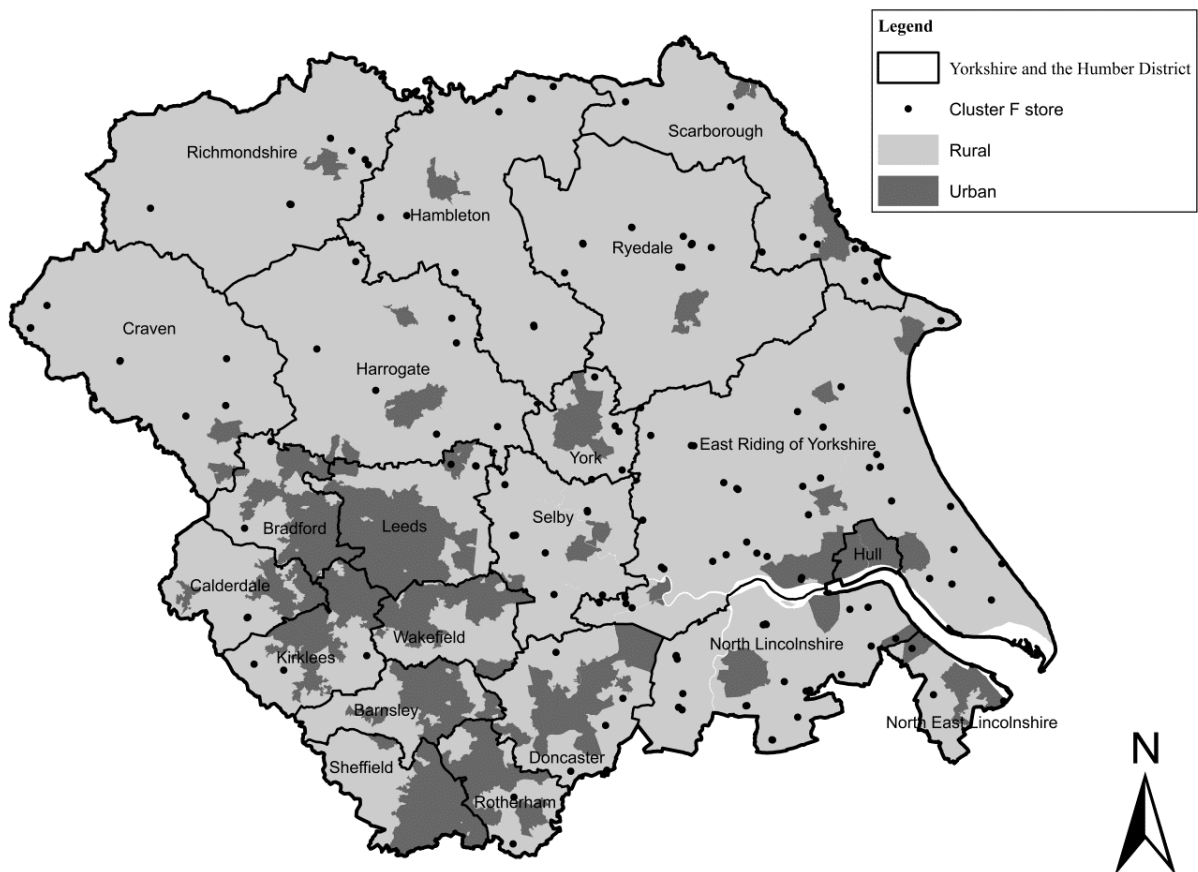
6.7 Conclusions

Prior to this chapter, this thesis reviewed the existing academic and industry literature on the convenience grocery market in Great Britain in relation to; the growth of major retailers into the convenience grocery market, the changing and growing demand for convenience groceries, and the attempts at forecasting revenue of both convenience grocery stores and grocery stores more generally both in academic and in the retail industry. Thereafter, chapter 5 quantified the extent to which the major retailers have committed to the convenience grocery market and assessed the geographical extent to which they play a role in convenience grocery retailing in Great Britain.

The analysis reported in this chapter forms part of the analysis required to meet the final aim of this thesis which is to develop and test a series of predictive models for forecasting convenience grocery store revenue in the varying location types in which this type of grocery store is found. From the outset of this research it was hypothesised that different locations in which convenience grocery stores are found in GB may, in theory, require a different optimal methodology for forecasting revenue accurately. This chapter has reported on a k-means cluster analysis which has segmented the convenience grocery market in Yorkshire and the Humber in to 7 statistically distinct location types based on residential population, daytime population, transport, affluence and retail vibrancy characteristics of locations in which convenience grocery stores are found. The classification of convenience stores presented here is important for future store location analysis as it will be used as a framework to test different methodologies for predicting convenience store sales.

The analysis found that certain locations were more favoured by certain types of retailers than others. Sainsbury's and Tesco prefer central locations more than other types of retailer with 14.7% of Sainsbury's Locals and 15.5% of Tesco Express stores falling in the central urban cluster in comparison with an average of 4.6% of stores across all retailers. There was also a greater than average presence of the major grocery retailers among locations defined as being central suburban areas. Conversely, Co-operative group convenience retail stores are more likely than average to be found in outskirts locations, particularly in outskirts affluent locations and rural locations. This makes sense and is in line with the retailers generally higher pricing

than other large grocery retailers. Moreover, this may also be as a result of the retailer being located in rural areas in which they command spatial monopolies. The major symbol group retailers identified in the market share analysis in chapter 5 are more likely than major retailers to favour smaller population suburban locations, are less likely to be placed in large satellite towns connected to large urban centres by rail links and are considerably more likely to be located in outskirts and rural locations.

In summary, not all convenience stores serve the same markets, in the same types of location, at the same time of day. Hence, if we are to offer retailers new insights into convenience store sales forecasting, it is important to think about forecasting revenues in alternative methodologies for alternative locations. If we can understand the different drivers of trade for different locations (through the cluster analysis here) we can begin to test alternative methodologies for producing sales forecasts by cluster, a significant advance on current operations. The forecasting of convenience grocery sales in Yorkshire and the Humber is reported in the next three chapters, starting with the development and testing of a GIS buffer and overlay model in chapter 7.

Chapter 7
Using a GIS buffer and overlay method to forecast convenience grocery store sales in Yorkshire and the Humber

The first aim of this thesis was to review the existing academic and industry literature on the convenience grocery market in Great Britain in relation to; the growth of major retailers into the convenience grocery market, the demand for convenience groceries, and the attempts at forecasting revenue of both convenience grocery stores and grocery stores more generally both in academic and in the retail industry. Chapter 2 and 3 achieved this aim with the latter identifying various methodological approaches to forecasting grocery revenue of grocery stores, identifying GIS buffer and overlay, regression and spatial interaction modelling as the three approaches to forecasting convenience store sales used in this research.

The second aim of this thesis was to quantify the extent to which major retailers have committed to the convenience grocery market and assess the geographical extent to which they play a role in convenience grocery retailing in Great Britain. The analysis presented in chapter 5 of this thesis found a marked shift in the portfolio of stores operated by major grocery retailers between 2003 and 2012. Driven predominantly by Tesco and Sainsbury's full throttle pursuit of convenience, the dynamic of the major four retailers store formats on a national level has shifted towards a greater emphasis on small-format grocery retailing within the remit of the Sunday Trading Act. Between 2003 and 2012, convenience stores as a proportion of total stores increased by 40.1%, from 8.2% to 48.3% of total stores.

Moreover, from the outset of this research it was hypothesised that that different locations in which convenience grocery stores are found in GB may, in theory, require a different optimal methodology for forecasting revenue accurately. Chapter 6 reported on the segmentation of the convenience grocery market in Yorkshire and the Humber into 7 statistically distinct location types based on residential population, daytime population, transport, affluence and retail vibrancy characteristics of convenience grocery store locations. The analysis found that certain locations were more favoured by certain types of retailers than others with the major grocery retailers more commonly associated with supermarket retailing operating a higher proportion of their convenience store networks in prime, central locations.

This chapter presents the methodology by which revenue predictions are evaluated in this thesis and contains the results of a GIS buffer and overlay analysis of convenience store revenue in Yorkshire and the Humber. This is a preferred methodology of the major grocery retailers when it comes to forecasting convenience grocery store sales. Section 7.1 details the methodology by which store sales were predicted using a GIS buffer and overlay procedure. Section 7.2 discusses the results of a 1 mile catchment buffer model applied to all Sainsbury's convenience stores in Yorkshire and the Humber regardless of location type. Thereafter section 7.3 disaggregates the results of the 1 mile buffer model by the 7 convenience store location types reported in chapter 6 of this thesis. Next, section 7.4 reports on the application of a GIS buffer and overlay model with variable buffered catchment areas with 7.5 disaggregating the results by the 7 location types identified in chapter 6 of this thesis. Finally, section 7.6 discusses potential explanations for the variance in forecasting ability of this model by location type and summarises the use of GIS buffer and overlay for forecasting convenience grocery stores.

7.1 Methodology

This chapter presents a series of GIS buffer and overlay models for predicting convenience grocery store revenue. It is designed to be easily developed using freely available data and implemented with a moderate level of GIS knowledge. In order to implement a GIS buffer and overlay methodology, three datasets are required. These are, a demand layer, a supply layer and a catchment area layer.

7.1.1 Demand layer

As detailed in chapter 4, survey data from the Living Costs and Food Survey (LCF) in which consumer grocery spending is disaggregated by demographic characteristics of the populations was used in combination with data on the number of households by output area to estimate residential grocery expenditure across Yorkshire and the Humber. This demand layer was plugged into the GIS software MapInfo Pro 12.5 and attached to a boundary file containing the location of output areas across Yorkshire and the Humber. The resulting demand layer is then primed for point in polygon analysis, described in more detail later in this chapter. Whilst MapInfo Pro 12.5 is not freely available software, free GIS software with the capabilities required for the analysis presented in this chapter is also available, notably the software QGIS.

It is unlikely that the catchment area of a convenience store will contain a purely residential evening population. We must therefore account for the daytime population

within each store location when estimating the available expenditure on groceries within the catchment areas of grocery stores. To recap, whilst comprehensive survey data on residential expenditure on goods available in convenience grocery stores is available through surveys such as the Living Costs and Food Survey, comprehensive surveys on work based expenditure are not available. In house research by Sainsbury's has found, on average, £5 per week per worker is a good estimate of work place grocery demand when forecasting revenue to their UK based store network. They find that a £5 mean expenditure per worker improves the accuracy of their in house gravity model, and has been extensively verified against sales in larger stores. This is a useful starting point from which to build a work based demand layer and is applied in the revenue forecasts in this chapter. In order to conduct analysis, the combined residential and work based grocery demand layer must be overlaid with an available supply layer containing grocery stores across Yorkshire and the Humber.

7.1.2 Supply layer

The demand layer discussed in 7.1.1 identifies the available expenditure on food and drinks available in each residential and workplace zone in Yorkshire and the Humber. To complement the demand layer, a supply layer containing the competing grocery destinations available to consumers is required. It is possible to identify the location of all available grocery stores within any area of the UK and thus build up an available supply layer using freely available data (e.g. google maps, planning policy applications, company websites). These locations can be geocoded and plotted in a GIS.

The model presented in this chapter built up a base layer of available supply by using the databases of stores provided by GMAP Ltd. (described earlier in chapter 4). This dataset includes grocery stores operated by all major convenience grocery players in GB such as Tesco, Sainsbury's, the Co-operative Group and the symbol group retailers alongside all grocery stores operated by retailers more associated with larger supermarket retailing. It is possible to build in a simple measure of attractiveness for grocery stores; an effective way of doing this is by taking into account the size of each store. Conveniently, GMAP Ltd. provided data on the size of each grocery store in Yorkshire and the Humber for all retailers from their store database as part of this research project. However, it would be possible to build up this layer without need for a corporate partner by consulting planning policy applications and company websites operated by each retailer. Alternatively, it is possible to deduce store size fairly accurately using free GIS software such as google maps, although this can be time consuming.

7.1.3 Catchment area

A GIS buffer and overlay modelling approach directly divides available consumer expenditure between competing grocery destinations. Defining the potential catchment area of a new or existing store is important when attempting to analyse its potential customers and the expenditure they have available to spend between competing stores. From this pretext, we can assume that we can identify the catchment area of a given convenience store or a potential new site for a convenience store. Subsequently, we can draw a buffer (representing a catchment area) around each store and measure the available demand available in the area. Defining the length of catchment area is key to the success of this type of model.

Wood and Browne (2007) identify the traditional method of using drive time analysis to define the distance consumers are willing to travel to larger supermarkets. Tools used to forecast supermarket revenue have found this method very effective in defining the catchment area of larger grocery stores. However, Wood and Browne (2007) theorise drive times as being less appropriate when attempting to define the catchment area of a convenience store. Intuitively, we can assume that customers would be more willingly to travel further to go to a larger grocery store as the size of a store is often closely linked to the range of products available. When assessing if any one retailer had too great a share of the grocery market (and too great an influence over the supply chain), the Competition Commission applied varying catchment area sizes dependent on the size and location of a grocery store, as shown in table 7.1 (Wood and Browne, 2006).

Table 7.1 Competition Commission Store Catchment Areas.

Store Type	Size (Sq. Ft.)	Location	Drive Time Limit
Convenience	*Less than 3,000*	*Urban*	*1 mile radius*
Convenience	*Less than 3,000*	*Rural*	*1 mile radius*
Mid-range	3,000 to 15,000	Urban	5 minutes
Mid-range	3,000 to 15,000	Rural	10 minutes
One-stop	15,000+	Urban	10 minutes
One-stop	15,000+	Rural	15 minutes

Source: Geobusiness Solutions (2005), referenced in Wood and Browne (2006)

The study identified the catchment area of a convenience grocery store – those less than 3,000 square feet in size – as being a one mile buffered radius around a store, denoting the distance customers will generally be willing to travel to a store. However, this research acknowledged from the outset that this one mile buffer is likely to be inadequate across all location types, considering the varying customer missions that

164

are met by stores for different locational contexts. For example, a person travelling from a workplace to a city centre store to buy lunch is likely to travel a shorter distance (given their time constraints) than a person using a rural convenience store as their main source of fresh grocery produce. This study initially uses the one mile radius catchment area used by the Competition Commission before reviewing its applicability for forecasting stores across the seven locational types identified in chapter 6. Moreover, further research is presented in an attempt to identify the optimal catchment area for this type of forecasting tool across different convenience store types.

7.1.4   Predicting sales

It is possible to integrate the demand, supply and catchment data described in this chapter to forecast the sales of convenience grocery stores in Yorkshire and the Humber. An example of this process can be seen below. To recap from chapter 3, thje procedure requires a number of steps. For a given catchment area of a store (e.g. a 1 mile buffer), polygon in polygon analysis is used to aggregate available expenditure in the catchment area. Moreover, the available supply within the buffer is also aggregated to the store level and used to equally divide available expenditure on groceries between each square foot of available floorspace. It is then possible to calculate the accuracy of each store prediction. Table 7.2 shows an example for a single Sainsbury's convenience store.

Table 7.2 Buffer and overlay calculation

Variable	Value
Store Size (Sq. Ft.)	3000
Available expenditure in catchment area (£)	£750,000
Total grocery supply in catchment area (Sq. Ft.)	50,000
Store Sales = (Available expenditure/Total grocery supply) x Store Size	£45,000

7.2 Judging Model Performance

It is useful at this point to outline the methodology by which model performance of each model presented in this thesis is judged. A formulaic approach to judging performance is taken in order to make the contributions each model is making to the forecasting of convenience grocery store revenue clear and comparable. Each model has its performance judged in two stages:

1) Global model accuracy

2) Accuracy by location type

7.2.1 Global Model Accuracy

It is important to assess the extent to which each model could be applied generally to all convenience grocery stores, regardless of the location type in which they reside. This is achieved by assessing the predictions generated by each model against a predetermined scale of accuracy of predictions. The predetermined scaled is based on industry knowledge of how effectively Sainsbury's (and other large grocery retailers) have previously predicted convenience store revenue. In the case of each model, the framework used to assess global model predictions is shown in table 7.3.

Table 7.3 Global model performance template

Cluster	Accuracy of forecasts				Mean (%)
	<60%	>60%	>80%	>90%	
All Stores	Value	Value	Value	Value	Value

As shown in table 7.3, the final column indicates mean performance, expressed as a percentage. The percentage refers to the mean accuracy of predictions made by the model in question. Accuracy is judged by the proportion of actual revenue of a store that has been predicted by the model. For example, if a store was predicted to achieve a weekly revenue of £8000 and its actual recorded revenue is £10000, the model has estimated 80% of actual revenue and is therefore 80% accurate at predicting sales to that store. If the model had predicted £12000 it would have been an estimate of 120% of store revenue, also expressed in this thesis as 80% accuracy. As such a prediction deemed as 80% accurate refers to ±20% difference between predicted and actual revenue.

Additionally, the proportion of store predictions achieving varying degrees of accuracy is also recorded. 60% refers to the proportion of revenue estimates (expressed as a percentage) that have fallen outside ±40% of actual store sales and 60% refers to the proportion of revenue estimates that have fallen within ±40% of actual store sales. 80% refers to predictions that have fallen between ±20% of actual store sales, judged as being good predictions in the context of convenience grocery store forecasting. Finally, 90% refers to those predictions that fall within ±10% of actual store sales, judged as being very good predictions.

In discussions with members of the location modelling team at Sainsbury's, the accuracy of forecasts was discussed in relation to what constitutes a 'good' forecast. A base for judging model performance as being of a minimum satisfactory level was

established as 60% of forecasts at a 60% level of accuracy. This is referred to throughout the remained of this thesis. Subsequently, the author has defined 80% accuracy as constituting a good revenue estimate and 90% accuracy constituting a very good revenue estimate.

7.2.2   Accuracy by location type

A main aim of the segmentation of the grocery market into location types distinguishable by statistically differing catchment area characteristics was to produce a typology by which different methodologies for predicting convenience store revenue could be judged. The effectiveness of each model can be judged against this classification of store location type, grouping store predictions by the location type that each store falls in to. Once again, the scale used to assess the overall effectiveness of each model is used by location type. In the case of each model, the framework used to assess model predictions by location type is shown in table 7.4.

Table 7.4 Model performance by location type template

Cluster	Accuracy (% of stores)				Mean
	<60	>60	>80	> 90	
Central Urban Cluster					
Large Population Suburban Cluster					
Smaller Population Suburban Cluster					
Satellite Cluster					
Outer Suburban Affluent					
Outer Suburban Less Affluent					
Rural Cluster					

7.3 One mile buffer model results

7.3.1   Global sales forecasting

Table 7.5 Accuracy of sales forecasts

Cluster	*Accuracy of forecasts*				Mean (%)
	<60%	>60%	>80%	>90%	
All Stores	71.6	28.4	15.8	7.4	40.0

In terms of forecasting ability, the GIS buffer and overlay model presented in this chapter is poor at predicting store sales for convenience stores. The model predicts less than 30% of stores at a greater than 60% accuracy, fewer than 16% of stores at a greater than 80% accuracy and less than 8% of stores at a very good level of accuracy (>90%). At a mean prediction level of 40% accuracy, this inappropriate to be used as a

general forecasting tool that can be applied to any convenience store regardless of the type of location in which the store resides. However, is it possible that the accurate store forecasts are centred on certain locational types of location in which convenience grocery stores are found?

7.4 Forecasting sales for different location types using a 1 mile buffer

It is possible to use the typology of convenience stores presented in chapter 6 to assess the predictive ability of this buffer and overlay model for different location types. Table 7.6 presents the results of a cluster by cluster assessment of the model's predictive power.

Table 7.6 Accuracy of sales forecasts by cluster

Accuracy (% of stores)					
Cluster	**<60**	**>60**	**>80**	**>90**	**Mean**
Central Urban Cluster	100.0	0.0	0.0	0.0	**15.6**
Large Population Suburban Cluster	72.7	27.3	13.6	9.1	**43.0**
Smaller Population Suburban Cluster	75.9	24.1	10.3	0.0	**38.2**
Satellite Cluster	83.3	16.7	16.7	0.0	**37.6**
Outer Suburban Affluent	33.3	66.7	50.0	33.3	**64.5**
Outer Suburban Less Affluent	54.5	45.5	18.2	9.1	**46.0**
Rural Cluster	100.0	0.0	0.0	0.0	**46.0**

7.4.1   Central urban locations

To recap, stores in this cluster are located in large urban centres across the region. These include the cities of Leeds, Sheffield, Bradford, York and Hull, along with stores in the centre of a number of the large towns in the region, including Huddersfield, Doncaster and Harrogate. When compared to the average among all retailers, we can see that a greater than average proportion of convenience stores operated by Tesco and Sainsbury's are found in the central urban cluster. At a mean average accuracy of 15.6%, the one mile radius buffer and overlay model has little success in predicting store sales in this cluster. Moreover, no predictions were found to achieve a level of 60% accuracy, the base level from which this research judges predictions to be acceptable for continued use. As a result, this model is wholly unsuited to being applied more widely to the forecasting of Sainsbury's convenience stores in central urban locations in other parts of the UK.

7.4.2 Suburban locations

There are two clusters that can be classified to fall in suburban locations. As their names suggest, they are distinguishable by the average population size in their respective catchment areas. The larger population cluster stores are located in suburbs of the major towns and cities inhabited by stores in the central urban location type. The one mile radius buffer and overlay model presented in this section has more success in estimating store revenue in this location than in central urban stores. However, the majority (72.7%) of revenue estimates fall short of the 60% accuracy threshold discussed earlier in this chapter. Moreover, the model has very little success in producing highly accurate predictions with less than 15% of predictions reaching an 80% level of accuracy and less than 10% of predictions reaching the 90% accuracy threshold. The combination of few very good predictions combined with an average accuracy of only 43% means this model's use more widely across the UK is not justifiable.

The smaller population suburban clusters stores are found in a mixture of the suburbs of the large towns and cities in which the central urban locations stores are placed alongside stores in centre and large suburbs of other important, but smaller towns in the area such as Wakefield, Barnsley, Rotherham, Scunthorpe and Grimsby. The one mile radius buffer and overlay model performs worse in this location type than in the larger population suburban location type. In summary; over three quarters of forecasts using this method fall below the minimum 60% accuracy threshold, less than 10% reach 80% accuracy and no forecasts reach the 90% threshold of very good predictions. Moreover, a mean revenue prediction accuracy of 38.2% makes this the third worst performing cluster for this model and rules out its further use as a forecasting tool for other similar locations outside Yorkshire and the Humber.

7.4.3 Satellite locations

Geographically, these stores are prevalent close to railway stations and in the suburbs of smaller towns. Geographically, these stores are found in small market towns such as Market Weighton in East Riding of Yorkshire, Northallerton in Hambleton district, Knaresborough in Harrogate District and Malton in Ryedale district. Outside the central urban store location cluster, the one mile radius buffer and overlay model performs worse in this location than any other. Over 80% of forecasts fell below the 60% accuracy threshold and only 16.7% of predictions reached a good level of accuracy (80%). Moreover, 0 revenue predictions could be categorised as very good as no forecasts achieved a 90% accuracy level. As with the three location types discussed

above, this model is not appropriate for wider application due to its poor performance in Yorkshire and the Humber.

7.4.4   Outskirts and rural locations

Three of the location types identified in the segmentation of the convenience grocery store market presented in chapter 6 are found in outskirts and rural locations. Stores in the affluent outskirts locations cluster are found in the affluent outskirts of the larger towns and cities and in relatively affluent larger villages and small towns across Yorkshire and the Humber. The model presented in this section has more success in this cluster than any other. The model exceed the 60% of stores at over 60% accuracy threshold described by Sainsbury's as their base target with over 80% of revenue estimates achieving a 60% level of accuracy. Moreover, the model has a good level of success in achieving high quality predictions with 50% of predictions at 80% or greater and one third of predictions at 90% or greater. The performance of this model in reaching accuracy targets, along with a mean prediction above the 60% accuracy threshold (64.5%), means that this model can be considered for wider application beyond Yorkshire and the Humber.

The second location type in the outskirts areas is the less affluent outskirts store cluster. Geographically, stores in this type of location are based in less affluent outskirts of the larger towns and cities and in relatively less affluent larger villages and small towns across the region including much of the corridor running from South Leeds down the north and south east of Barnsley through into the northern outskirts of Rotherham (areas which struggled with major job losses in mining and manufacturing in the 1980s and 1990s). In contrast to the outskirts affluent cluster, the one mile radius buffer and overlay model has little success in predicting store revenue in the outskirts less affluent location type. The model has an average prediction accuracy of 46% in this type of location, almost 20% lower than in the more affluent outskirts location type. Additionally, the model struggles to reach a 60% accuracy level with over 50% of store revenues predicted less accurately than this and less than one quarter of predictions achieving 80% accuracy. The poor performance of the model in this cluster in Yorkshire and the Humber means that it does not lend itself kindly to being rolled out in the prediction of sales in similar locations across other parts of the UK.

The seventh and final location type is the rural cluster. They are prevalent in much of the more rural districts of Yorkshire and the Humber such as Richmondshire, Hambleton, Ryedale, Craven, East Riding of Yorkshire and North Lincolnshire and major retailer stores are considerable less likely to be located here. Moreover, only one

of the stores for which Sainsbury's provided revenue data falls in this type of location. The solitary store is predicted at a 46% level of accuracy meaning it falls significantly short of the 60% base from which to consider a model for wider use. As a result of the very low number of cases, it is difficult to judge the appropriateness of this model based on its performance in predicting the revenue of a single store.

7.5 Forecasting sales using variable buffered catchment areas

Due to the 1 mile radius buffer model having very poor results across the majority of clusters (6 out of 7), it is clearly inappropriate for application to the whole of Sainsbury's convenience store network both in Yorkshire and the Humber and other parts of the UK. This research postulates that the one mile catchment area used by the Competition Commission to assess convenience market share is not going to produce the best possible revenue estimates as each store is likely to have a slightly different catchment area due to a number of factors. However, it is possible that different fixed buffer models may be more (or less) appropriate in predicting store revenue by the seven locational types identified in the segmentation of the market in chapter 6. Moreover, other buffer sizes were tested ranging in 0.5 mile increments from 0.5 miles up to 5 miles in radius. Three of these were found to improve on predictions in at least one location type. These were:

1. 0.5 mile radius
2. 1.5 mile radius
3. 2.0 mile radius

Having introduced variable buffer radiuses into the analysis, it is possible to test the effectiveness of each model in:

1. Forecasting store revenue across all convenience stores
2. Forecasting store revenue in each of the seven location types

The performance of each of these models in estimating convenience grocery store revenue were compared to the initial 1.0 mile buffer applied in the Competition Commission's investigation of the grocery market in the UK.

7.5.1   Global Model Results

Table 7.7 identifies the performance of each of the four radius buffer models (0.5, 1, 1.5 and 2) when forecasting the 95 convenience grocery store revenues provided by Sainsbury's. Both the 1.5 mile and 2 mile radius buffer models perform at a greater mean accuracy than the 1 mile radius buffer model presented earlier in this chapter.

They have an average accuracy of 46.9% and 47.1% respectively, an advance on the 40.0% mean of the 1 mile radius buffer model. With a mean prediction of 37.1% accuracy, the 0.5 mile buffer was found to have a lower mean accuracy than the 40.0% of the 1 mile radius buffer.

Table 7.7 Global revenue predictions using four buffer distances

	Accuracy of forecasts				
Buffer	**<60%**	**>60%**	**>80%**	**>90%**	**Mean (%)**
0.5	72.3	27.7	17	8.5	37.1
1.0	71.6	28.4	15.8	7.4	40
1.5	68.8	31.3	14.6	9.4	46.9
2.0	71.6	28.4	6.3	3.2	47.1

The 1.5 mile model is the most effective model at limiting poor predictions (<60% accuracy) with 31.3% of predictions achieving this level of accuracy. This is the only model that exceeds the 1.0 mile radius buffer model in limiting poor predictions. Despite its poor mean performance, the 0.5 mile radius buffer model leads the way in terms of good model predictions (>80% accuracy), suggesting it may have some positives when breaking down predictions by cluster. Moreover, despite having the best average performance, the 2.0 mile radius buffer model has the poorest performance in producing good model predictions (>80% accuracy) with only 6.3% of store revenues predicted at this level of accuracy. The 0.5 mile, 1.0 mile and 1.5 mile models perform similarly in terms of good predictions with 17%, 15.8% and 14.6% respectively. When identifying very good model predictions (>90% accuracy), the 1.5 mile radius buffer model is the best performing with 9.4% of predictions achieving this level of accuracy. Once again the 2.0 mile radius buffer model lags behind with less than 5% of revenue predictions achieving a 90% level of accuracy (or greater).

The 1.5 mile radius buffer model is the best performing model when considering both mean accuracy and the realisation of good individual revenue predictions. Whilst the mean prediction is slightly higher for the 2.0 mile radius buffer model, the 1.5 mile radius buffer model performs better in both limiting the proportion of very poor estimates (<60% accuracy) and in achieving very good model predictions. The 2.0 mile buffer is slightly more effective than the 1.5 mile radius buffer at limiting extremely poor predictions (<60% accuracy) resulting in a slightly greater proportion of <60% accuracy estimates (+0.2%) but the 1.5 mile buffer is the best overall performing model. However, the 1.5 mile radius buffer model is not reliable enough to be considered for rolling out across all Sainsbury's convenience stores across the UK.

That said, it is plausible that one (or more) of the models may perform well in a specific store location when assessing each model's ability to predict sales in the 7 location types identified in the cluster analysis in chapter 6.

7.6 Forecasting sales in different store location types using variable buffered catchment areas

7.6.1 Central urban locations

Table 7.8 identifies the performance of each of the four models in predicting revenue in central urban stores. The 1.0 mile radius buffer model is very poor at predicting sales in central urban locations at a mean accuracy of 15.6%. This is bettered by the three other models with a mean accuracy of 24.1% (0.5 mile buffer), 25.1% (1.5 mile buffer) and 23.0% (2.0 mile buffer).

Table 7.8 Central urban store revenue estimates using four buffer radiuses

Central Urban Cluster					
Buffer	**<60%**	**>60%**	**>80%**	**>90%**	**Mean (%)**
0.5	84.6	15.4	7.7	7.7	**24.1**
1.0	100	0	0	0	**15.6**
1.5	85.7	14.3	0	0	**25.1**
2.0	92.9	7.1	0	0	**23.0**

When considering the extent to which each model limits very poor predictions, the 1.0 mile radius buffer model achieved 0 predictions at the minimum acceptability level of 60%. The other three models all exhibit more success (although still very limited) with the 0.5 mile radius buffer model leading the way with 15.4% of sales predicted at a 60% level of accuracy. This still falls significantly short of the 60% of store revenue estimates at 60% accuracy level of acceptability. Moreover, the 0.5 mile radius buffer model is the only model that achieves predictions exceeding an 80% and 90% level of accuracy with 7.7% of store revenue estimates at each of these levels respectively. On the whole, the 0.5 mile radius buffer performs best in this cluster although its performance is still very poor.

7.6.2 Suburban store locations

Table 7.9 identifies the performance of each of the four models in predicting revenue in large population suburban store locations. The initial 1.0 mile radius buffer model has the lowest mean performance in this cluster with an accuracy of 43.0%. However,

mean performance across the four models is very similar ranging from 43% to 46.1% in average prediction accuracy, with the 2.0 mile radius buffer model top with 46.1% mean accuracy.

Table 7.9 Large population suburban store revenue estimates using four buffer radiuses

Large Population Suburban Cluster					
Buffer	**60%**	**60%**	**80%**	**90%**	**Mean (%)**
0.5	63.6	36.4	31.8	13.6	**45.0**
1.0	72.7	27.3	13.6	9.1	**43.0**
1.5	69.6	30.4	13	8.7	**45.4**
2.0	73.9	26.1	13	8.7	**46.1**

Both the 0.5 mile and 1.5 mile radius buffer models improve on the 1.0 mile radius model in limiting very poor model predictions. The 0.5 mile radius buffer model performs best by this diagnostic in achieving 36.4% of predictions above 60% accuracy. Moreover, the 0.5 mile radius buffer model is the best performing model by far in achieving good model predictions (>80%), with over 30% of predictions achieving this level of accuracy (31.8%). This significantly exceeds the other three models which all achieve between 13% and 14% of revenue estimates at an 80% or greater level of accuracy. Additionally, the 0.5 mile radius buffer model has the greatest success in producing very good estimates (>90% accuracy). 13.6% of predictions made by this model, an increase of 4.5% on the initial 1.0 mile radius buffer model. On the whole, the 0.5 mile radius buffer model performs best in this location type.

Moving on to the smaller population suburban cluster, table 7.10 identifies the performance of each of the four models in predicting revenue in this location type. The 1.5 mile radius buffer is the best performing model for this cluster, clearly outstripping the initial 1.0 mile radius buffer model in its ability to predict revenue in this store location type. Both the 1.5 mile radius buffer model (49.3%) and the 2.0 mile radius buffer model (51.0%) have a higher mean prediction than the 1.0 mile radius buffer model (38.2%). Both of these models are better than the original 1.0 mile model for predicting store revenue in the large population suburban convenience stores operated by Sainsbury's.

Table 7.10 Smaller population suburban revenue estimates using four buffer radiuses

Smaller Population Suburban Cluster					
Buffer	**<60%**	**>60%**	**>80%**	**>90%**	**Mean (%)**
0.5	82.8	17.2	6.9	3.4	**30.8**
1.0	75.9	24.1	10.3	0	**38.2**
1.5	72.4	27.6	17.2	10.3	**49.3**
2.0	69	31	3.4	3.4	**51.0**

In terms of limiting very poor model predictions (<60% accuracy), both of the larger buffers (1.5 and 2.0 mile) are more successful than the initial 1.0 mile radius buffer model. However, the 2.0 mile radius buffer model (3.4%) lags behind both the 1.0 mile (10.3%) and 2.0 mile radius buffer models (17.2%) in producing good model predictions (>80% accuracy) with the 1.5 mile radius buffer model clearly performing best. The only model of note in producing very good store revenue estimates (90% accuracy) is the 1.5 mile radius buffer model. For this model 10% of predictions exceed this level of accuracy. The combination of close to the highest mean prediction and proportion of very poor predictions along with the best performance in producing good and very good estimates makes the 1.5 mile radius buffer model the most suitable for use in this location type.

7.6.3   Satellite store locations

Table 7.11 identifies the performance of each of the four models in predicting revenue in stores in the satellite locations cluster. The 37.6% mean accuracy of the initial 1.0 mile radius buffer model is significantly bettered by both the 2.0 mile buffer model (51.3% mean accuracy) and the 1.5 mile buffer model (65.5% mean accuracy). The 65.5% mean accuracy achieved by the 1.5 mile radius buffer places it into consideration for being used more widely in estimating revenue in this type of location in other parts of the UK.

Table 7.11 Satellite locations revenue estimates using four buffer radiuses

Satellite Locations Cluster					
Buffer	**<60%**	**>60%**	**>80%**	**>90%**	**Mean (%)**
0.5	83.3	16.7	0	0	**34.4**
1.0	83.3	16.7	16.7	0	**37.6**
1.5	50	50	33.3	33.3	**65.5**
2.0	83.3	16.7	16.7	0	**51.3**

The 0.5 mile, 1.0 mile and 2.0 mile radius buffer models perform equally in limiting very poor model predictions. However, with 83.3% of predictions in each of these falling below 60% accuracy they are clearly unreliable at doing this. On the other hand, 50% of predictions fall below this level of accuracy in the 1.5 mile radius buffer model, a significant improvement on the initial 1.0 mile radius buffer model. In addition, the 1.5 mile buffer model has double the proportion of good estimates (>80% accuracy) than any other model with 33.3% of stores predicted at this level of accuracy. Moreover, it is the only model in which any predictions exceed 90% in accuracy, with one third of forecasts meeting this criterion.

7.6.4 Outskirts and rural store locations

Table 7.12 identifies the performance of each of the four models in predicting revenue in affluent outskirts store locations. The 1.0 mile radius buffer model remains the most effective model of the four after the good performance of the model described earlier in this chapter. The mean prediction of 64.5% is over 5% better on average than each of the other three radiuses.

Table 7.12 Affluent outskirts revenue estimates using four buffer radiuses

Affluent Outskirts Cluster					
Buffer	<60%	>60%	>80%	>90%	Mean (%)
0.5	66.7	33.3	25	16.7	40.6
1.0	33.3	66.7	50	33.3	64.5
1.5	41.7	58.3	16.7	0	59.0
2.0	45.5	54.5	9.1	0	55.0

It is also the most successful model in limiting very poor model predictions and producing both good and very good model predictions. 33.3% of store estimates fall below this threshold for the 1.0 mile radius buffer model whereas over 40% of stores are predicted very poorly in each of the other models. In terms of good (>80% accuracy) and very good (>90% accuracy) revenue estimates, the 1.0 mile radius buffer model achieves 50% and 33% of predictions at these thresholds respectively with the nearest competing model achieving only half of these good and very good model estimates.

Table 7.13 identifies the performance of each of the four models in predicting revenue in less affluent outskirts store locations. When measuring mean predictive power of each model, the original 1.0 mile radius buffer model is the worst performing of the three models, with a mean accuracy of 46.0%. The 0.5 mile radius buffer model is the

most effective in this measure of performance in predicting a mean level of 51.9% accuracy.

Table 7.13 Less affluent outskirts revenue estimates using four buffer radiuses

	Less Affluent Outskirts Cluster				
Buffer	**<60%**	**>60%**	**>80%**	**>90%**	**Mean (%)**
0.5	45.5	54.5	27.3	9.1	**51.9**
1.0	54.5	45.5	18.2	9.1	**46.0**
1.5	72.7	27.3	18.2	18.2	**49.4**
2.0	72.7	27.3	0	0	**49.0**

In addition to being the best model in terms of mean accuracy of forecast, the 0.5 mile radius buffer model is also the most effective tool in limiting very poor model predictions with a 9% greater proportion (54.5%) of estimates at greater than 60% accuracy than the second placed original 1.0 mile radius buffer model. Moreover, the 0.5 mile radius buffer model also has the greatest proportion of good model forecasts (>80% accuracy) with over one quarter of estimates achieving this benchmark (27.3%). The 1.5 mile radius buffer model achieves the greatest proportion of very good (90% accuracy) revenue forecasts. However, due to its inefficiency in limiting very poor predictions (27.2% greater than the 0.5 mile model) it is a less suitable model than the 0.5 mile radius buffer model in terms of its potential for rolling out to the forecasting of store sales in this type of location.

Finally, table 7.14 identifies the performance of each of the four models in predicting revenue in rural store locations. Once again, the limitation of only having one store to base a judgement on restricts the extent to which the future applicability of the model can be evaluated.

Table 7.14 Rural revenue estimates using four buffer radiuses

	Rural Cluster				
Buffer	**<60%**	**>60%**	**>80%**	**>90%**	**Mean (%)**
0.5	100	0	0	0	**53.2**
1.0	100	0	0	0	**46.0**
1.5	100	0	0	0	**54.4**
2.0	0	100	0	0	**62.1**

However, the best performing model for the store in this cluster was the 2.0 mile radius buffer model, estimating revenue at a 62.1% level of accuracy

7.6.5 Which simple model is the best for each cluster?

Table 7.15 summarises the best model for each cluster and the accuracy at which each model predict the sales of stores in the cluster for which they are most appropriate. All four distance buffers proved to be optimum in at least one of the location types, justifying moving beyond the 1 mile buffer used by the Competition Commission in assessing convenience grocery market share. Moreover, this supports the theory posited earlier that convenience stores are not homogenous in nature and statistically distinct location types exist in which catchment characteristics differ.

Table 7.15 Optimum buffer and overlay model for each location type

Optimum model for each cluster						
Cluster	<60%	>60%	>80%	>90%	Mean (%)	Buffer
A	84.6	15.4	7.7	7.7	24.1	0.5
B	63.6	36.4	31.8	13.6	45.0	0.5
C	72.4	27.6	17.2	10.3	49.3	1.5
D	50	50	33.3	33.3	65.5	1.5
E	33.3	66.7	50	33.3	64.5	1
F	45.5	54.5	27.3	9.1	51.9	0.5
G	0	100	0	0	62.1	2

Despite identifying optimum catchment areas for prediction in each location type, many of the predictions still fall short in terms of the potential for GIS buffer and overlay model's to be applied more widely as robust and reliable forecasting tools. However, the GIS buffer and overlay model showed promise in two location types in particular. The optimum model for forecasting central urban stores fell significantly short of acceptable average prediction levels with a mean accuracy of just 24.1%, discounting GIS buffer and overlay analysis for consideration more widely in this type of location. Moreover, all buffer and overlay models performed very poorly in forecasting both suburban location types with the best model for each producing mean estimates of significantly less than the base target of 60% accuracy (with 36.4% and 27.9% mean accuracies respectively). Moving to the outskirts less affluent cluster, the large proportion of very poor forecasts (45.5% at less than 60% accuracy) along with a mean performance of just over 50% rules this model out as a tool for forecasting this type of store more widely.

The two location types in which this model performed with greater predictive power were the satellite locations cluster and the outskirts affluent locations cluster. The highest mean performance of any model was exhibited in the 1.5 mile radius buffer

model when applied to satellite store locations with a mean accuracy of 65.5% of sales. This is heavily influenced by one third of forecasts falling above the 90% accuracy threshold. Worryingly, the model resulted in 50% forecasts falling under the base 60% accuracy threshold, raising questions over the consistency of predictions which would worry any retailer. The most consistent performance of any GIS buffer and overlay model was the initial 1.0 mile radius model when forecasting revenue in affluent outskirts store locations. The model produced a mean forecast accuracy of 64.5% coupled with predicting 50% of stores revenues at a good level of accuracy (>80%) and 33.3% of store revenues at a very good level of accuracy (>90%). Nevertheless, there is also an issue of poor predictions with this model in which one third of revenue estimates fall below the 60% accuracy threshold that would worry a prospective retailer in applying this type of model.

7.7 Validation

This section introduces the validation process for the three models reported in this thesis before discussing the specific validation of the GIS buffer and overlay method in this chapter. Sainsbury's provided sales data for an additional 31 convenience grocery stores in the North West region of Englandf. The author had no control over which stores were provided. The sales data for these stores is for the same time period as for the Yorkshire and the Humber model presented in this chapter. The stores provided by Sainsbury's fit into the clusters identified in the segmentation of the market in chapter 6 as shown in table 7.16.

Table 7.16 Validation store location types

Cluster	Number of Stores	Proportion of stores (%)	
		Yorkshire and the Humber	North West
Central Urban Cluster	7	14	23
Large Population Suburban Cluster	7	23	23
Smaller Population Suburban Cluster	6	31	19
Satellite Cluster	4	6	13
Outer Suburban Affluent	2	13	6
Outer Suburban Less Affluent	5	12	16
Rural Cluster	0	1	0

There are a greater proportion of validation stores in the central urban, satellite and outskirts less affluent store location types and a smaller proportion of validation stores in the smaller population suburban cluster and the outskirts affluent cluster. Unfortunately, no stores falling into the rural location type were found in the validation

revenue data. Tables 7.17 and 7.18 identify the results of forecasting the North West based validation stores using the GIS buffer and overlay model reported in this chapter.

Table 7.17 Validation global accuracy

Cluster	<60%	>60%	>80%	>90%	Mean (%)
	Accuracy of forecasts				
All NW	67.7	32.3	9.7	6.5	39.3

Table 7.18 Validation accuracy by location type

Cluster	Accuracy (% of stores) – NW Validation				Mean (%)
	<60%	>60%	>80%	>90%	
Central Urban Cluster	85.7	14.3	0.0	0.0	**19.5**
Large Population Suburban Cluster	71.4	28.6	14.3	14.3	**37.5**
Smaller Population Suburban Cluster	50.0	50.0	0.0	0.0	**33.1**
Satellite Cluster	100.0	0.0	0.0	0.0	**33.1**
Outer Suburban Affluent	50.0	50.0	50.0	50.0	**62.4**
Outer Suburban Less Affluent	40.0	60.0	20.0	0.0	**65.1**
Rural Cluster	N/A	N/A	N/A	N/A	**N/A**

On average, the model performs very similarly for the North West validation stores at an average accuracy of 39.1% in comparison to 40.1% for stores in Yorkshire and the Humber. It is difficult to identify performance by location type due to the mismatch in store numbers between the calibration and test stores. The model performs very similarly across location types with the exception being the improved forecasting of less affluent store locations in the North West. However, there are less than half the total stores for this cluster in the North West in comparison to Yorkshire and the Humber. Further validation is planned on a define set of stores to achieve a clearer picture of how well the GIS buffer and overlay model performs outside of Yorkshire and the Humber.

7.8 Discussion and summary

Chapter 2 identified the growth of major grocery retailers, more traditionally associated with supermarket retailing, in to the convenience grocery market. Established methods for forecasting supermarket sales have been described as ineffective or inappropriate to forecasting convenience grocery sales. Chapter 6 identified 7 statistically distinct location types in which convenience grocery retailing takes place in Yorkshire and the Humber. This segmentation of the market is used in this thesis as a framework to assess the application of GIS buffer and overlay modelling, regression modelling and spatial interaction modelling for forecasting convenience grocery sales, thus empirically

testing the hearsay that the latter two modelling frameworks are ineffective at forecasting convenience grocery store sales.

To recap, at the outset of this research, it was hypothesised that different location types may require a different optimal strategy for good quality and robust store forecasting. This chapter has reported on the results of the GIS buffer and overlay model for forecasting convenience grocery store sales. Methodologically, the model was a relatively simple GIS buffer and overlay procedure. Allowing for varying sizes of buffer in forecasting grocery store sales in different locations resulted in an improvement in the forecasting ability of the model. Moreover, the incorporation of workplace demand using WPZ boundaries is an advance on previous attempts at incorporating non-residential demand in to this type of model. This type of method is already well used by major retailers when it comes to forecasting convenience grocery store sales. However, little empirical testing of this type of model appears in the academic literature.

When applied to forecasting Sainsbury's convenience grocery stores across Yorkshire and the Humber, regardless of location type, this model performed very poorly. As a result, it would not be recommended for wider use to all convenience grocery store forecasting. The model performed particularly poorly in central urban and large suburban locations. This model has not taken into account preference for different retailers. Sainsbury's customers may be likely to spend higher than average sums of money and those customers in central locations, often with professional jobs, may spend more money than is anticipated in the model.

Despite its poor overall performance, the model had moderate success in predicting sales in satellite, outskirts and rural locations. These locations rely heavily on residential demand, with very little workplace or visitor demand, and the primary shopping mission of consumers. This type of demand is more predictable both in terms of volume (through survey data) and origin. Curiously, the analysis found that the model was much better at forecasting sales in less affluent outskirts locations when compared to affluent outskirts locations. This may be for a number of economic and lifestyle reasons which are discussed in more detail in chapter 10.

Moreover, the supply of stores in these areas is less varied and often lower in volume due to smaller populations to serve and less available store units. This is clearly contributing to the ability of the model to replicate known revenues. The next two chapters report on two more methodologies for forecasting convenience grocery store sales, both of which are more methodologically complex than the GIS buffer and

overlay model reported in this chapter. In doing so, the issues raised in the difficulties faced forecasting convenience grocery store sales using a GIS buffer and overlay method are addressed. The reasons for the differences in predictive power of the different models is investigated in more detail in the comparison of models in chapter 10 of this thesis. Next, chapter 8 reports on a disaggregated SIM for the convenience grocery market.

Chapter 8
Application of an applied disaggregated Spatial Interaction Model for the convenience grocery market in Great Britain

Chapter 3 identified various methodological approaches to forecasting revenue of grocery stores, identifying GIS buffer and overlay, regression and spatial interaction modelling as the three approaches to forecasting convenience store sales used in this research. To recap, from the outset of this research it was hypothesised that that different locations in which convenience grocery stores are found in GB may, in theory, require a different optimal methodology for forecasting revenue accurately and Chapter 6 reported on the segmentation of the convenience grocery market in Yorkshire and the Humber into 7 statistically distinct location types.

Chapter 7 presented a GIS buffer and overlay model for forecasting convenience grocery sales in Yorkshire and the Humber. The model was found to have little success in predicting convenience store revenue in Yorkshire and the Humber across all convenience stores. When predicting revenue in the different location types identified in chapter 6, the model was found to be very poor at predicting sales in central urban and suburban locations and was moderately successful in predicting sales in satellite, rural and affluent locations.

The research in this chapter applies a spatial interaction framework to the forecasting of convenience grocery store sales for different location types. There is a precedence in the application of SIMs in the context of consumer behaviour and its impact on grocery retailing (Birkin et al., 2002; 2010, Thompson, 2013; Newing, 2013). However, these studies have either excluded convenience grocery stores from analysis or have not calibrated based on sales to these convenience stores. The complexities of their trading was deemed to be inappropriate (or at the very least problematic) to capture using a SIM framework. Using the segmented clusters as a framework for testing allows us to assess the extent to which an applied SIM methodology is appropriate in the different location types in which convenience stores are present.

Thus, this chapter aims to test the potential for an enhancement of predictive capacity using spatial interaction modelling. It covers the process of developing and implementing a disaggregated spatial interaction model and testing its ability to predict store revenues in the various location types identified in chapter 6 of this thesis. The chapter draws heavily on previous research in to the development of disaggregated

SIMs (at the University of Leeds) by Birkin et al. (2002; 2010), Thompson (2013), and Newing (2013). Section 8.1 discusses the classic production-constrained SIM used widely in grocery applications. Next, section 8.2 outlines the potential for disaggregating this SIM for the convenience grocery market in GB. Section 8.3 discusses the process of building the model in terms of grocery supply, grocery demand and the interactions between the two. Thereafter, sections 8.3 to 8.6 report on the convenience specific calibrations made for the SIM reported in this chapter, resulting in a model more appropriate to predicting convenience grocery stores. Section 8.7 discusses the final disaggregate convenience SIM in this research before sections 8.8 and 8.9 evaluate its effectiveness at forecasting sales to convenience stores both in general and in the various locations in which convenience grocery retailing takes place.

8.1 The Aggregate Model - Classic production-constrained entropy maximising SIM

To recap from the literature review in chapter 3, the singly constrained entropy maximising SIM is the most commonly adopted type of gravity model used as a location planning tool to estimate store revenue. SIMs are built up from two integral datasets, available supply and available demand within a given geography. The model seeks to capture the interaction between demand zones (people and their available grocery expenditure) and potential destinations (grocery stores in this case). The model is based upon the assertion that the grocery expenditure available within an area is shared between a number of competing retailers based on their relative attractiveness (to consumers) and accessibility.

The basic singly-constrained (production-constrained) entropy maximising SIM used in retail (Birkin et al., 2002) can be defined as:

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

(Equation 11.1)

Where:

> S_{ij} is the flow of expenditure (or people) from demand zone *i* to grocery store *j*
> O_i is a measure of demand in an area – most commonly available grocery expenditure
> W_j is a measure of the attractiveness of grocery store *j*

> C_{ij} is a function representing the cost of interaction between demand zone *i* and store *j,* most commonly in the form of straight line distance between the two

and:

- A_i is a balancing factor ensuring that all demand is allocated between the available grocery stores within the study region defined as:

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

(Equation 11.2)

The model is built on two key assumptions:

1. Other things being equal, a customer is more likely to shop in a more attractive store
2. Other things being equal, a customer is less likely to travel to a store that is further away

In the real world, the two assumptions listed above are more complex than an aggregate SIM is able to capture. Thus, previous attempts at SIMs both by retail firms (in-house) and in store location research have attempted to disaggregate the models to replicate real world consumer behaviour more accurately, resulting in improved store-level revenue predictions.

8.2 Disaggregated Spatial Interaction Modelling

Whilst SIMs have limitations, applied models have been found to predict store- level revenue to an acceptable level of accuracy and replicate known customer flows for data for calibration is available, particularly following model disaggregation to improve on the basic production-constrained aggregate SIM.

A number of commentators highlight the limitations of the aggregate model in its ability to replicate real world customer behaviour and subsequent flows of money from consumers to grocery stores (Benoit and Clarke, 1997; Birkin et al., 2002, Birkin et al., 2010; Fotheringham, 1983). They all propose disaggregation of the basic production constrained aggregate SIM as a vehicle to improve accuracy of store-level revenue prediction. The criticisms of the aggregate model have centred on the contribution of two model terms to the SIMs potential ability to replicate real world customer flows. These terms are as follows:

> ➢ The attractiveness term W_j

> ➢ The cost term C_{ij}

It is in the disaggregation of these two terms that previous research has aimed to improve the predictive capacity of SIMs when applying them to specific grocery retailing situations, from discount retailing (Thompson, 2013) to tourist demand for groceries in coastal regions (Newing, 2013). Following on from these studies, this research also focuses on the disaggregation of these two terms in a manner appropriate to forecasting sales to convenience grocery stores. The following section gives a general overview of previous attempts at disaggregating the two terms and the reasons for doing so.

8.2.1 The attractiveness term (W_j) and the cost term (C_{ij})

The challenge of applied retail SIMs is to replicate human behaviour by adjusting parameters within a model to accurately mirror real-world decision making made by consumers. In aggregate models, store size in square feet is generally used as a proxy for the attractiveness 'rating' of a store which is applied to all stores operated by all retailers. Whilst store size is undoubtedly important to many consumers, Birkin et al. (2010) postulate a number of other factors that may contribute to the attractiveness of a store including, but not limited to; brand, pitch, offer, parking, levels of refurbishment, age and micro location, including factors such as visibility and convenience walking routes.

Other studies have also found issues with the aggregate models reliance on a simplistic attractiveness term. Fotheringham (1983) conducted an investigation into bias caused by an over-simplistic use of store size alone as a proxy for attractiveness in which small stores with many adjacencies that would attract customers lose out on patronage whilst larger stores with little adjacencies are prone to over estimates of patronage. Clarke et al. (2012) also found evidence supporting the varying attractiveness of different retailers to consumers of different types (or in different areas), finding consumers from more affluent areas would be more satisfied by the presence of a Sainsbury's supermarket than they would be at the presence of a supermarket of equal size operated by Asda. Furthermore, Birkin et al. (2010) found that in Leeds, one square foot of Tesco/Sainsbury's grocery space was a more attractive offer to the majority of consumers than one square foot of grocery space operated by other less-known retailers.

Retailers are aware of the varied perception of their brand by customers of different types. Moreover, they often target specific customers and came into the market to exploit a particular unique selling point in terms of their offer and how it relates to specific consumers – note the rise of the discounters discussed in chapter 2. Retailer's awareness of their customer base is reflected in their in house store forecasting systems; Sainsbury's in-house SIM attributes a higher attractiveness of the brand to affluent residential zones than to deprived residential zones. It is possible to disaggregate the basic production constrained SIM to vary attractiveness by a number of factors including by retailer, by store type or by consumer type. This chapter goes on to discuss how this research refined the attractiveness term in order to improve the predictive capacity of SIMs for convenience grocery store forecasting in Yorkshire and the Humber.

The cost term in a grocery SIM generally refers to the cost of travel between an origin zone (residential or workplace) and a store destination. In order to build this cost term, a number of actions are required. First, a measure of distance or cost needs to be developed. Previous work has identified a number of measures including but not limited to; Classic Euclidean straight line distance, distance on a transport network, travel time, or travel cost in monetary terms. Birkin et al. (2010) note that traditional straight line distance rarely works in model building and instead recommend incorporating a road travel network and its associated travel times into the cost function of the applied SIM.

Newing (2013) adopted this road travel distance as the measure of cost in forecasting Sainsbury's grocery store sales in Cornwall. He noted that the rural nature of Cornwall means that the majority of interactions between consumers and grocery stores will occur using a car as the method of transport. The model focused on supermarkets and therefore larger shopping trips were being undertaken by consumers. During calibration of the model presented in this chapter, both straight line distance and road travel time were tested. It was found that straight line distance was more effective in forecasting convenience grocery sales. This is probably as a result of a significant proportion of trips to this type of store being undertaken without a car.

8.3 Building a convenience grocery model

This section discusses the process by which the final disaggregated retail SIM in this study process was developed and empirically tested, focusing on the additional data incorporated into the calibration process to create a model that is tailored to the convenience grocery market in Great Britain. In discussing the development of the

model this section will, in turn, focus on constructing the demand layer in the model, identifying key suppliers of grocery products, defining the spatial relationship between available supply and available demand, maintaining accurate market share figures, incorporating consumer preference for specific retailers and the calibration of the model based on consumer's grocery retail mobility. In reality, the process of incorporating these factors into the model was non-linear and involved trade-offs in calibration between the different factors in multiple iterations of calibration. The chapter draws heavily and attempts to follow on from the work completed at the University of Leeds by Mark Birkin, Graham Clarke and Martin Clarke, and more recently Chris Thompson and Andy Newing.

8.3.1 Grocery Demand

Previous production constrained retail SIMs developed (typically at the University of Leeds) have effectively captured available residential demand for groceries and used these to develop models which robustly predict sales in tourist areas (Newing, 2013) and predicting market share of each of a number of retailers in response to the increased prominence of discount retailers (Thompson, 2013). These models focused on residential population and they did not have a disaggregated geography by workplace and residence in their estimations. Many locations in which convenience grocery stores are found rely heavily on work based population and this analysis seeks to incorporate a more nuanced work based demand into analysis.

However, residential demand for convenience grocery stores was not neglected. To recap, chapter 4 detailed a demand layer formulated using the Living Costs and Food Survey and data on the number of households at the OA level in Yorkshire and the Humber. This layer was used as a basis for building the demand layer used in the applied SIM in this chapter. The final model presented in this model is at a larger geographic unit, the lower super output area (LSOA). Therefore, OA level estimates of available residential grocery expenditure were aggregated to the LSOA level using geographic lookup tables.

Newing (2013) incorporated a work based demand for groceries into each LSOA in his study area. However, the SOA census geography was built for residential populations and deals poorly with workplace data, as described in more detail in chapter 4. SIMs rely on a geographic structure in which flows are predicted from an origin zone to a destination (in this case a store point), a zonal geography that differs from the residential LSOAs described previously in this chapter is required. The Workplace Zone (WPZ) geography detailed in chapter 4 and used in the other models in this

thesis was used in the SIM reported in this chapter. This created a dual demand layer accurately representing where people live and work is an advancement of previous models which have incorporated workplace demand for groceries into residential geographies that are often ill-suited for this purpose.

To recap, In-house research by Sainsbury's has found, on average, £5 per worker per week is a good estimate of workplace grocery demand when forecasting revenue to their UK based store network. They find that a £5 mean expenditure per worker improves the accuracy of their in house gravity model, and has been extensively verified against sales in larger stores. This is a useful starting point from which to build a work based demand layer.

8.3.2 Grocery Supply

When attempting to forecast grocery sales of large grocery stores, previous SIMs in the academic literature have focused on major players in the UK grocery market when developing the supply side of retail SIMs. This has often been seen as acceptable due to the nature of grocery retailing in the UK. The largest retailers were dominant, commanding a significant market share and consumers generally conducted weekly or bi-monthly shops. The largest retailers competed with each other for large, out-of-town sites and many smaller retailers were deemed too small or insignificant to warrant comprehensive inclusion in predictive models.

As convenience grocery retailing was not the focus of either of these research projects, small retail stores of less than 10,000 sq. ft. were often excluded unless they commanded a spatial monopoly such as rural convenience stores being a major grocery destination of a large proportion of the population of an area (Newing, 2013). Moreover, additional computing power is required with the addition of each store which often led to the exclusion of smaller stores that would be expected to command little revenue. A further issue is the lack of knowledge about smaller retailers and independent grocery stores. Although the majority of calibrations in a SIM will be tailored to a specific retailer, it is useful to be able to benchmark the performance of other stores (and retailers) within any model. This has generally been possible for the larger retailers but has proven more difficult when looking at smaller retailers operating few stores.

However, due to the changes discussed in chapter 2, particularly the rise in prominence of discount retailers, shifting consumer behaviour and the difficulty large retailers had in expanding their supply of large grocery stores due to changes in

planning policy - the retail market has changed and major retailers now operate stores in locations in which they compete with smaller grocery retailers, often relying on top up shopping or multiple trips per customer per week to achieve acceptable levels of revenue in each store. This presents a challenge if spatial interaction models are to be effective tools for predicting convenience grocery store sales. Moreover, the specific nature of the research presented in this thesis means that it is aimed at forecasting small grocery store revenue.

The model presented in this thesis attempts to overcome the issues around incorporating as many grocery stores as possible (ideally to match the real world) into an applied SIM. Three ways in which this is achieved are:

1. A comprehensive retailer layer covering all retailers
2. Calibration based on anticipated market share
3. Calibration based on expected store performance by retailer

To recap, the dataset provided by Sainsbury's and GMAP Ltd. which is discussed in chapter 4 contains information on every branded grocery store in Yorkshire and the Humber at a point in time which matches the sales data. It disaggregates each store by floorspace and includes all of the major convenience grocery players highlighted in the market share analysis in chapter 5 of this thesis, making it a comprehensive store layer. The second and third points listed above are related to the calibration of the model itself, the next section discusses these calibrations and highlights the way in which they advance convenience grocery store revenue forecasting.

8.4 Model calibration against additional retailers

8.4.1 Benchmarking against additional retailers

The retail SIM presented in this chapter has the main aim of predicting the revenue of convenience grocery stores operated by Sainsbury's. The retailer provided store sales data for Yorkshire and the Humber making it possible to calibrate the model in line with Sainsbury's performance. However, if only Sainsbury's data was used in calibration, it is possible that the model may be over-fitted to one retailer which would be undesirable if the model is to be a useful tool in wider grocery retailing and research. To achieve the aim of wider use for the whole grocery market (or at the very least convenience grocery market), benchmarking figures for additional retailers were built into the calibration process to account for known (or estimated) performance of other retailers.

Two methods of benchmarking against retailers other than Sainsbury's were used; 1) Estimated market share, and 2) Benchmarking against mean trading performance.

8.4.2 Market Share Calibration

It is possible to test for the performance of a number of retailers in the model through the use of market share as a performance indicator. Whilst it is not possible to calibrate based on individual store sales data for every retailer, it is possible to incorporate each retailers overall performance in the model by deriving expected market share for each retailer in the model. This was achieved in the following steps and believed to be novel to this research:

a) Identify national market share for each retailer at the time of the known Sainsbury's sales data the model is trying to predict.
b) Compare national market share to floorspace operated by each retailer nationally to calculate a market share per square foot of floorspace .
c) Identify the retail mix in the model study area at the time of sales data.
d) Abstract expected market share in study region based on national market share compared to local (model) retail mix.

National market share of each retailer

Kantar Worldpanel is an arm one of the largest global consultancy networks for insight and information, Kantar. They aim to use purchasing behaviour to gain useful insights in the retail market through an expertise in, and understanding of, shoppers' behaviour. Every month, they freely publish the grocery market share of major players in the grocery industry in a number of countries: China, France, Spain, Taiwan, Ireland, and Great Britain. They published market share data based on a 12 week rolling cycles. At the end point of sales data in this thesis, 8th December 2013, grocery market shares in GB were as shown in table 8.1.

Table 8.1 Grocery market share in GB by retailer, 8th December 2013.

Retailer	Market Share
Tesco	29.9
ASDA	16.9
Sainsbury's	16.8
Morrisons	11.6
Co-op	6.2
Symbols/Independents	4.7
Waitrose	4.7
Aldi	4
Lidl	3.1
Iceland	2.1

Calculating market share per square foot of floorspace

It is possible to calculate a national market share per square foot of floorspace value for each retailer by taking the total floorspace operated by each retailer in Great Britain and dividing this by the proportion of the market commanded by each retailer in the Kantar World Panel market share data, as shown in table 8.1.

Table 8.2 Grocery Floorspace and Market Share by Retailer in GB, 2013

Retailer	National Floorspace	National Share of Market	Market share per sq. ft.
Tesco	36247144	29.9	0.0000008249
Asda	20668942	16.9	0.0000008177
JS	22133464	16.8	0.0000007590
Morrisons	15408581	11.6	0.0000007528
Co-Op	13736910	6.2	0.0000004513
Waitrose	5612505	4.7	0.0000008374
Symbols/Indep	22078964	4.7	0.0000002129
Aldi	4491870	4.0	0.0000008905
Lidl	6211691	3.1	0.0000004991
Iceland	4000969	2.1	0.0000005249
Total	**150591040**	100.0	n/a

The grocery floorspace operated by each retailer (for which Kantar Worldpanel publishes market share data) in the study region can be seen in table 8.3. Moreover, the table shows the outcome of multiplying of each retailer's floorspace in the study region by the national average market share value per square foot of national floorspace for each retailer.

Table 8.3 Study Area Floorspace by retailer, 2013

Retailer	Study Area Floorspace	Share per sq. ft. (national)	Total Share (Floorspace x Share)
Tesco	797069	0.00000082	0.657502218
ASDA	809664	0.00000082	0.662062253
JS	810571	0.00000076	0.615223389
Morrisons	1079698	0.00000075	0.812796654
Co-Op	498118	0.00000045	0.224800653
Waitrose	48895	0.00000084	0.040944673
Symbols/Indep	1019232	0.00000021	0.216994493
Aldi	143266	0.00000089	0.127578373
Lidl	213596	0.0000005	0.106605764
Iceland	77257	0.00000052	0.040552199
		Total	3.505060669

The floorspace per sq. ft. is difficult to interpret and therefore use to calibrate the retail SIM in this chapter and was therefore converted into a predicted market share value, as shown in table 8.4. The market share is calculated as each retailer's proportion of the total floorspace value of each retailer.

Table 8.4 Comparison of GB and study region market shares

Retailer	Study Area Floorspace Power	Predicted Model Market Share in Y&H	National Comparison (%)
Tesco	0.657502218	18.8	-11.1
ASDA	0.662062253	18.9	2
JS	0.615223389	17.6	0.8
Morrisons	0.812796654	23.2	11.6
Co-Op	0.224800653	6.4	0.2
Waitrose	0.040944673	1.2	-3.5
Symbols/Indies	0.216994493	6.2	1.5
Aldi	0.127578373	3.6	-0.4
Lidl	0.106605764	3	-0.1
Iceland	0.040552199	1.2	-0.9

Morrison's is predicted to be the retailer that had the highest market share in the study region at the end of 2013, with a predicted market share of 23.2%. Additionally, Tesco, ASDA and Sainsbury's were estimated to have 18% of market share each. When comparing the predicted study region market share in figure 8.1 with the Kantar Worldpanel (2013) national market share data for each retailer, this research found that for the purpose of calibration, it was expected that Tesco would have a significantly lower proportion of the grocery market in the model's study region than the national share of the market for the retailer. Conversely, Morrison's was expected to perform better in the model's study region than its national market share in 2013 due to their increased presence in the study area. These predicted market shares were used to make sure that additional revenue was not artificially directed towards the Sainsbury's stores for which sales data was known and ensured that the model was not over-fitted to a single retailer in the model.

8.4.3 Expected mean store performance for additional retailers

A common method of analysing performance is trading intensity – the revenue per a given area of floorspace for a store or retailer (usually £ per week per square foot) Through Sainsbury's partnership with Leeds University, this research has a very accurate picture of the expected trading intensity that can be expected by Sainsbury's, and by extension, the other large grocery firms Morrison's, Tesco and ASDA. However,

prior to the research being undertaken, the research had little evidence on which to benchmark the performance of smaller retailers such as the Symbol Groups.

IGD are a retail research and training charity focusing in helping the food and grocery industry meet the needs of the public. Through IGD, retailers and grocery researchers can attain insights into the workings of the grocery industry, with knowledge of the performance of other retailers being one such insight. Sainsbury's kindly provided this insight for a number of retailers whose sales performance were unfamiliar, allowing the benchmarking of performance for these retailers in the spatial interaction model presented in this chapter, as shown in table 8.5.

Table 8.5 Trading intensity of additional retailers. Source: IGD (2013)

Brand	Trading Intensity (£ per week/sq. ft.)
Co-op	10.13
Costcutter	8.97
Londis	9.97
Nisa	13.87
SPAR	13.61

Whilst this extensively covers the symbol group retailers, it unfortunately did not include data on the expected performance of the discount retailers ALDI, Netto and Iceland. Clearly, store by store performance varies for each retailer but these values allow for an overall benchmarking and calibration against the mean performance for the retailers listed in table 8.5 in the retail SIM presented in this chapter.

8.5 Calibrating store attractiveness

Birkin et al. (2010) argue that the main factors influencing the attractiveness of a store are floorspace, parking, accessibility and price. Floorspace has traditionally been considered a key component of attractiveness in grocery stores. It is often theorised and considered as a proxy for a range of other store attributes including product availability, product range, opening hours and lower price (due to economies of scale), see Birkin et al. (2002) for a more comprehensive discussion. Floorspace is an important factor when we consider a consumer's (or group of consumers) preference for stores operated by different retailers. Other things being equal, the majority of consumers would choose to visit the larger store. However, it is also evident that different types of consumer will exhibit a preference for certain retailers over others.

This could be for a variety of reasons. These include: previous experiences, brand exposure, product range, personal ties, habit and price.

Within the academic literature, brand is the most widely discussed reason for consumers preferring one retailer over another. Clarke et al. (2012) found that affluent consumers would be more satisfied if they had a Sainsbury's in the retail mix of their local area when compared to the possibility of having an ASDA of the same size. Moreover, this study found that consumers in low income areas valued lower cost supermarkets such as ASDA and Morrisons (along with the discount retailers Lidl and Aldi) when compared to Sainsbury's. Retailers are attuned to this and their in-house models often reflect these distinctions. For example, Sainsbury's gravity model reflects how the brand will be perceived by geodemographically differing residential areas. To this end, brand loyalty can be difficult to capture by using floorspace alone (Newing, 2013).

Thompson et al. (2012) developed one such approach in which they took into account store size, centrality of a store location, stores that are close to other stores (agglomeration effects) and regional brand attractiveness. The work produced a matrix identifying the attractiveness of each retailer (and each fascia) to each of the OAC consumer groups in Yorkshire and the Humber. The research used Axciom's research opinion poll (ROP) data for 2011 and 2012 for 75,000 households in Yorkshire and the Humber to define a parameter (α) for each store to disaggregate the attractiveness of different retailers for different consumer groups. A location quotient was created for each retailer which quantified whether a particular OAC group was over or under represented in the retailer's customer base. This was achieved by dividing each retailer's observed (from the ROP data) customer breakdown by OAC group by the population's distribution across each OAC group in the study region. They confirmed what would be intuitively expected; that the more affluent OAC subgroups (e.g. Prospering Suburbs) have higher than average patronage in retailers such as Waitrose, M&S and Sainsbury's, retailers at the higher end of the grocery market in terms of price and luxury goods.

It is possible to build these location quotients into a SIM as a parameter that varies the attractiveness of each square foot of floorspace by a power function meaning that a square foot of floorspace of certain retailers will be more attractive to certain groups of consumers in the population than other groups. Newing (2013) adopted such an approach by rescaling Thompson et al.'s location quotients around a value of 1, as the alpha value in the model is a power function. This approach contributed to Newing's

development of a SIM that effectively predicted Sainsbury's store sales in Cornwall. This application of the location quotient technique is encouraging for the model presented in this chapter, as validation in other study areas is important if the model presented here is to be applicable to other regions of Great Britain. The model presented in this chapter adopts this approach and the brand location quotients for use in the disaggregated SIM in this research can be found in table 8.6.

Table 8.6 Location quotients for use in disaggregated SIM (Newing, 2013)

Retailer	OAC Supergroup						
	Blue Collar	City Living	Countryside	Prospering Suburbs	Constrained by circumstances	Typical Traits	Multicultural
Aldi	0.998	0.997	1.0051	0.9987	1.0025	1.0005	0.9952
Asda	1.0076	0.9912	0.9904	0.997	1.0023	0.9992	1.0013
Co-Op	1.002	0.999	1.0157	0.9922	1.0008	1	0.9894
Iceland	0.9997	0.9982	1.0058	0.9975	0.9991	1.0001	1.0021
JS	0.9904	1.0121	1.0013	1.0088	0.9942	1.0028	0.9997
Lidl	1.0015	0.9995	1.0066	0.9962	0.9957	0.9997	1.0091
M&S	0.9891	1.0381	0.9967	1.0066	0.9952	1.0051	1.0003
Morrisons	1.0005	0.9942	0.9997	0.9987	1.002	1.0005	0.999
Tesco	0.9992	0.9987	1.0071	1.0010	0.9965	0.999	0.9985
Waitrose	0.9811	1.1	1.0061	1.0124	0.9843	1.0023	1.0068

M&S and Waitrose - the two retailers catering for the higher end of the grocery market - are considerably more attractive to the more affluent city living and prospering suburbs OAC supergroups than the less affluent constrained by circumstances and blue collar supergroups. Conversely, ASDA – a retailer with a traditional stronghold in less affluent, working class areas – is calibrated to be more attractive to customers in the less affluent OAC groups.

The use of location quotients to inform the attractiveness term in spatial interaction models has been proven to be effective in calibrating retail grocery SIMs in different contexts; in London and Yorkshire and the Humber by the work of Thompson (2013) and in Cornwall and Kent by Newing (2013). However, both of these studies paid little attention to the convenience grocery market in their final SIMs, particularly in the calibration process. This research proposes a new set of attractiveness values for

convenience stores by location type for Sainsbury's, and by extension the other major retailers with branded convenience grocery stores, most notably Tesco.

In order to calculate attractiveness terms for convenience stores, this research combines the previously developed locational types presented in chapter 6 with known sales trading data provided by Sainsbury's. The resulting output of this estimation is a set of convenience store alpha values (attractiveness values), disaggregated by retailer and by cluster for each OAC supergroup. The calculation presented here is based on the mean trading intensity (weekly revenue per sq. ft. of floorspace) for the convenience stores in each cluster as a proportion of the mean trading intensity across all Sainsbury's stores, giving a premium to certain 'attractive' locations in which storms perform well**. The steps taken to produce a new set of alpha values is as follows:

1. Calculate mean trading intensity for Sainsbury's store network in Yorkshire and the Humber.
2. Calculate mean trading intensity for Sainsbury's stores in each cluster.
3. Calculate the difference (in multiples) between the mean trading intensity of all stores and the mean intensity of stores in each cluster.
4. Using the power function in the SIM, calculate the uplift required to increase the floorspace of a store in each cluster (taken from a base of the mean floorspace of a store in each cluster) to a size equivalent to the increased trading intensity experienced in each cluster location in comparison to the mean of all stores.
5. Rescale alpha values for all major retailers convenience stores based on their alpha value for the large supermarkets as calculated by Thompson (2014)

Calculating mean trading intensity by cluster

As previously discussed, Sainsbury's provided revenue data for a number of stores across Yorkshire and the Humber. This included sales data for both convenience stores and larger stores. The trading intensity for larger stores and convenience stores in each cluster was then calculated by dividing total sales (in each cluster) by the combined sum of the sizes of all stores in each cluster. The results are shown in table 8.7.

Table 8.7 Mean Trading Intensity by cluster

Cluster	Cluster Name	Mean Trading Intensity
A	Central Urban Cluster	36.33
B	Large Population Suburban Cluster	31.08
C	Smaller Population Suburban Cluster	27.02
D	Satellite Cluster	23.16
E	Outer Suburban Affluent Cluster	26.67
F	Outer Suburban Less Affluent Cluster	29.93
G	Rural Cluster	39.13
Sainsbury's (2013)		**19.27**

It was then possible to calculate the difference between the sales at stores in each location type identified in chapter 6 of this thesis and Sainsbury's overall average trading intensity for the year 2013. This is shown in the **Uplift** column in table 8.8. This is simply expressed as a multiple of the trading intensity of Sainsbury's stores in each cluster when compared to the mean trading intensity of Sainsbury's in Yorkshire and the Humber.

Table 8.8 Uplift in trading intensity for each cluster

Cluster	Cluster Name	Mean Trading Intensity	Uplift (vs Sainsbury's Mean)
A	Central Urban Cluster	36.33	1.89
B	Large Population Suburban Cluster	31.08	1.61
C	Smaller Population Suburban Cluster	27.02	1.4
D	Satellite Cluster	23.16	1.2
E	Outer Suburban Affluent Cluster	26.67	1.38
F	Outer Suburban Less Affluent Cluster	29.93	1.55
G	Rural Cluster	39.13	2.03
Sainsbury's (2013)		19.27	1

There is a mismatch between the mathematics of uplift (a simple multiple of mean sales) and the alpha value as it is presented in the disaggregate model presented in this chapter (a power function), meaning an additional calculation is required to convert between the two to produce a set of alpha values by cluster. To do this four pieces of information are required:

1. The average size of stores operated by each major retailer in each cluster.
2. The uplift value for each cluster identifying the difference in trading intensity in each convenience store in each cluster when compared to the mean trading intensity of all Sainsbury' stores

3. The average size of stores operated by each major retailer in each cluster uplifted by the difference in sales between Sainsbury's larger stores and Sainsbury's convenience stores in each cluster.
4. The alpha value for each retailer for each OAC group for larger stores, as discussed earlier in this chapter.

Table 8.9 contains the average store size of Sainsbury's in each of the location types identified in chapter 6 of this thesis. Moreover it also contains the uplift value and uplift sales area value identifying the expected perception of those stores in comparison to Sainsbury's supermarkets given their sales data.

Table 8.9 Data required for developing a convenience alpha value

Cluster	Cluster Name	Average Sales Area	Uplift	Uplift Sales Area
A	Central Urban Cluster	1970	1.89	3715
B	Large Population Suburban Cluster	2009	1.61	3240
C	Smaller Population Suburban Cluster	2245	1.4	3147
D	Satellite Cluster	2303	1.2	2768
E	Outer Suburban Affluent Cluster	1954	1.38	2704
F	Outer Suburban Less Affluent Cluster	1729	1.55	2685
G	Rural Cluster	1690	2.03	3431

In order to develop a new set of alpha values, an alpha value needs to be produced that converts the average sales area for each cluster to the uplift sales area value is required, as shown in the power column in table 8.10.

Table 8.10 Alpha (power) values for each cluster

Cluster	Cluster Name	Average Sales Area	Power	Uplift Sales Area
A	Central Urban Cluster	1970	1.0836	3715
B	Large Population Suburban Cluster	2009	1.0629	3240
C	Smaller Population Suburban Cluster	2245	1.0438	3147
D	Satellite Cluster	2303	1.0237	2768
E	Outer Suburban Affluent Cluster	1954	1.0836	2704
F	Outer Suburban Less Affluent Cluster	1729	1.0591	2685
G	Rural Cluster	1690	1.0953	3431

For example, 1970 to the power of 1.0836 equals 3715, so a power of 1.0836 would achieve the uplift required to make stores in the central cluster more attractive to consumers.

Accounting for varying retailer attractiveness among OAC supergroups

However, as seen earlier, the alpha values in the model vary by retailer and by OAC supergroup. Therefore, in much the same manner that Newing (2013) rescaled Thompson's (2013) alpha values around the value of 1, the convenience location alpha values were reweighted around each retailer's larger store alpha values in this study using the methodology described above. The resulting alpha values can be seen in table 8.11.

Table 8.11 New Sainsbury's convenience alpha values

Sainsbury's Cluster	OAC Supergroup						
	Blue Collar	*City Living*	*Countryside*	*Prospering Suburbs*	*Constrained by circumstances*	*Typical Traits*	*Multicultural*
Central Urban	1.0707	1.0941	1.0824	1.0906	1.0748	1.0841	1.0807
Large Population Suburban	1.0509	1.0739	1.0625	1.0704	1.0549	1.0641	1.0608
Smaller Population Suburban	1.0332	1.0558	1.0446	1.0524	1.0372	1.0461	1.0429
Satellite	1.0137	1.0359	1.0248	1.0325	1.0176	1.0264	1.0232
Outer Suburban Affluent	1.0315	1.0541	1.0429	1.0507	1.0355	1.0445	1.0412
Outer Suburban Less Affluent	1.0461	1.0691	1.0576	1.0656	1.0502	1.0592	1.0560
Rural	1.0801	1.1037	1.0920	1.1001	1.0842	1.0936	1.0902

These were then applied in the model and found to substantially improve the prediction of sales to Sainsbury's convenience stores in the SIM. The third major form of calibration used in the disaggregate SIM reported in this chapter is the calibration of average trip distance by different type of consumer in different types of area in the study region. This was done using the Nectar card data described in chapter 4 of this thesis.

8.6 Calibrating average trip distance (ATD)

The interaction between origin zones and store destinations in an applied retail SIM is dependent on a measure of the distance between each origin and each destination. The most basic measure of distance, and the method traditionally used in many models, is straight line (Euclidean) distance (Birkin et al., 2010a). This is easily calculable from the co-ordinates of origins and destinations and is computationally light. However, the complex environmental of towns and cities which consumers have to navigate means that simple Euclidean distance is rarely the most effective distance measure in practice.

Moreover, Clarke et al. (2010) highlight that the widespread use of cars in the grocery shopping process (to carry goods more easily) means that many consumers access to stores is limited by road networks rather than straight line distance. Birkin et al. (2010a) highlight that in their vast experience of applied gravity models, it is the incorporation of road networks, and more pressingly road travel times, that leads to the accurate capturing of flows between origins and destinations. However, they argue that such a measure of cost of travel should be handled with care, noting that travel time may vary by congestion, speed, one-way systems, time of day and day of week.

In the calibration process of this thesis both straight line distance and road travel time were tested. It was found that straight line distance was more effective in forecasting convenience grocery sales. This is probably as a result of a significant proportion of trips to this type of store being undertaken without a car.

In order to calibrate based on average trip distance (ATD), observed flows of money between demand zones and store destinations was required. This was provided by Sainsbury's in the form of Nectar card transactions between residential postcodes (which were aggregated to the OA level and finally to the LSOA level) to a sample of Sainsbury's grocery stores in Yorkshire and the Humber. Chapter 4 described the form that this data took. This was not available for workplace zones to stores as is discussed in more detail in the conclusions in chapter 10. The β parameter in the disaggregate model was calibrated using ATD and a selected GOF statistic was used to validate the values for this parameter in the model. This goodness of fit statistic came in the form of attempting to minimise the difference between observed versus predicted values of distance travelled in which observed average trip distance was calculated using the formula:

$$ATD^{Obs} = \frac{\sum_{ij} \hat{S}_{ij} C_{ij}}{\sum_{ij} \hat{S}_{ij}}$$

And predicted average trip distance was calculated using the formula:

$$ATD^{Pred} = \frac{\sum_{ij} S_{ij} C_{ij}}{\sum_{ij} S_{ij}}$$

In which S_{ij} represents predicted flows, and $\hat{S}_{ij}$ represents observed flows. The closer the model replicates to actual trips being made, the more likely the spatial pattern of observed expenditure in the model will be realistic. Trip distance was calibrated and improvements to the forecasting ability of the SIM reported in this chapter were made.

8.7 Disaggregate SIM in this study

The final disaggregate SIM in this study builds on previous SIMs in the academic literature (Benoit and Clarke, 1997; Birkin et al. 2010; Clarke, 2011; Newing, 2013). It was calibrated to incorporate two core factors.

1) Different stores, brands and fascias vary in relative attractiveness to different groups of consumers (based on a number of factors including affluence)
2) Different groups of consumers (by census LSOA characteristics) are willing to travel further to access their store, brand or fascia of choice.

Like the model developed by Newing (2013), the model contains a power function which applies relative brand attractiveness to different retailers and different fascias on a consumer-by-consumer basis. Moreover, the model builds on the SIM developed by Newing (2013) in further disaggregating brand attractiveness by the type of location in which the major retailer's convenience stores are found. In calibration, the model took into account relative attractiveness of stores (and brands/fascias) to different groups of consumers, the average trip distance of different types of consumers and the expected performance of retailers other than Sainsbury's within the model whilst generally calibrating to improve the estimation of Sainsbury's store revenue. The disaggregate model reported in this chapter is shown in equation 8.1 and is developed with reference to Newing (2013)

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

(Equation 8.1)

Where:

S_{ij}^{kn} represented the flow of expenditure from zone i to store j (of brand n) by consumer type k and is also disaggregated by workplace zone and residential zones

A_i^k is a balancing factor ensuring all demand from zone i is allocated to stores in the study region

O_i^k is a measure of demand available in each residential zone i by each consumer type k which is also disaggregated by residential and workplace origin zones

$W_j^{\alpha kn}$ is the overall attractiveness of store j with additionally disaggregation by perceived brand attractiveness by consumer type for each brand and fascia

C_{ij} is a measure of distance between origin zone i and store destination j and includes a measure of distance deterrent depending on type of consumer and type of origin zone (residential or workplace)

8.8 Global sales forecasting

Table 8.12 reports on the results of the global results of the SIM using the equation detailed above. The applied SIM is moderately successful in predicting store revenue across all Sainsbury's convenience grocery stores in the model with a mean accuracy of 71.3%. This exceeds the base 60% mean target identified in chapter 7 as the minimum threshold to consider the model viable.

Table 8.12 Global accuracy of SIM forecasts

Cluster	Accuracy of forecasts				Mean Accuracy (%)
	<60%	>60%	>80%	>90%	
All Stores	44	56	44	26	71.3

The major limitation of this model is in its inability to prevent poor model predictions. Over 40% of individual store forecasts fall below the 60% threshold. Limiting poor forecasts is a central role of location planning teams and these results make the model's overall use across all convenience store locations difficult to justify. However, the model performed very well in specific location types.

8.9 Cluster by cluster sales forecasting

In disaggregating these predictions by convenience location type, it is possible to identify the performance of the model across the 7 convenience grocery location types identified in chapter 6. Table 8.13 shows the accuracy of store sales disaggregated by the convenience location types identified in chapter 6.

Table 8.13 Accuracy of SIM forecasts by location type

	Accuracy of forecasts				
Cluster	**<60%**	**>60%**	**>80%**	**>90%**	**Mean Accuracy (%)**
Central Urban Cluster	60	40	27	13	51.4
Large Population Suburban Cluster	56	44	22	22	61.5
Smaller Population Suburban Cluster	33	67	50	17	70.9
Satellite Cluster	67	33	0	0	53.2
Outer Suburban Affluent	29	71	57	29	74.2
Outer Suburban Less Affluent	20	80	50	50	80.5
Rural Cluster	0	100	100	0	81.0

The ability of the model to forecast sales varies substantially by location type. The model performs poorly in the central urban, large population suburban and satellite clusters. Moreover, the model has moderate successes in forecasting revenue in the smaller population suburban location type and the outskirts less affluent location type. Conversely, the model is relatively successful in predicting revenue to rural store locations and stores in the outskirts less affluent cluster.

8.9.1 Central Urban Locations

On the whole, the SIM reported in this chapter is very poor at estimating store revenue in central urban store locations. The model is particularly unsuccessful at limiting poor model predictions with 60% of forecasts unable to reach the 60% accuracy threshold identified in chapter 7. Despite its lack of success in limiting poor predictions, the model achieves 80% accuracy around as quarter of the time and 90% accuracy for around 10% of stores. The limiting nature of the poor model predictions in certain stores in this location type results in a mean accuracy of 51.4% among stores in this cluster. This is the lowest for any cluster and suggests that using a SIM to forecast convenience store revenue in central urban locations is problematic.

8.9.2 Suburban Locations

Similar to the central urban cluster, the model performs poorly in the large population suburban cluster. Once again, the model is very unsuccessful in guarding against poor revenue estimates with almost 60% of store predictions failing to reach the minimum 60% accuracy threshold. Akin to the model's performance in the central urban cluster, the SIM predicts around a quarter of stores at an 80% level of accuracy or above. No revenue estimates fall between 80 and 90% accurate in this location type meaning that almost a quarter of revenue estimates reach the 90% accuracy threshold.

On average, the model is more successful in predicting sales in this type of location than the central urban cluster with a mean prediction of 61.5% accuracy. This increase of 10% in average is mostly accounted for by an increase in revenue estimates that are very good (>90% accuracy). However, the presence of very poor revenue estimates and a relatively low mean accuracy level means the author would be wary to suggest the use of this model more widely for this location type.

When compared to the SIMs success at estimating store sales in the large population suburban cluster, the model performs significantly better for stores in the smaller population suburban cluster. This location type experiences over 20% less very poor sales predictions with around one third of estimates failing to reach the 60% accuracy threshold. Moreover, the model is more successful at achieving good revenue estimates (>80% accuracy) with 50% of store forecasts reaching this level of accuracy. Conversely, the model performs slightly worse in this cluster than in the larger population suburban cluster in terms of generating very accurate revenue estimates (90%) with around 17% of revenue estimates achieving this level of accuracy.

On average, the SIM reported in this chapter is more successful at predicting revenue in the smaller population suburban location type than in both central urban and larger population suburban convenience store locations. At a mean accuracy level of 70.9% with half of estimates reaching the 80% accuracy threshold this model is worthy of further consideration in terms of rolling it out across similar store locations outside the study area.

8.9.3 Satellite Locations

The SIM in this chapter has the worst average performance in the central urban cluster. However, the location for which the model generates the fewest proportion of acceptable forecasts and the fewest proportion of forecasts achieving a good (>80%) and very good (>90%) level of accuracy. 67% of forecast fail to achieve the minimum

60% accuracy threshold and no store revenues are predicted with an accuracy reaching 80% or 90%. On average, this model performs slightly better than the central urban cluster with a mean accuracy of 53.2%. This is accounted for by limiting the proportion of forecasts falling substantially below the 60% accuracy threshold. However, this level of accuracy is very low and suggests that using a SIM to forecast convenience store revenue in satellite convenience store locations for Sainsbury's is problematic.

8.9.4 Outskirts Locations

The three outskirts clusters are the locations in which the SIM is found to perform best among the location types identified in chapter 6 of this thesis. In the outskirts affluent cluster the SIM performs well in terms of limiting poor model predictions with just over one quarter of stores being forecast to this degree of accuracy. Moreover, the highest proportion of good (>80% accuracy) forecasts are generated for this location type with around 60% of store revenue estimates reaching this threshold of accuracy. Moreover, the model achieves around one quarter of estimates at a very good level of accuracy, the second highest proportion across all location types. On average, the model generates a mean accuracy of 74.2% in this location type, the third highest among the 7 location types identified in chapter 6.

The SIM performs more effectively in the outskirts less affluent store locations which average 13.5% of residents being employed in social class 1 occupations, less than half of the proportion of residents in the more affluent outskirts cluster. The model is more successful in limiting poor model predictions with 20% of predictions falling below this threshold in comparison to 29% among outskirts affluent location types. The model achieves 50% of store revenue estimates at a minimum of 80% accuracy along with 50% of estimates reaching a 90% level of accuracy. This means that all forecasts generated by the SIM of over 80% accuracy are also over 90% accuracy, a substantially proportion of store revenue estimates. On average, the model predicts at an accuracy of 80.5% in the outskirts less affluent store cluster. This is the highest level of accuracy among clusters for which a number of stores are available for testing and means that the SIM is worthy of further consideration in terms of rolling out this model for forecasting similar location types in other parts of Great Britain.

Finally, the one store in the rural location type is predicted very well at an 81.0% level of accuracy. Incidentally, this is the highest among any cluster. However, once again, it

is difficult to confirm the performance of the model in this cluster due to the small sample size of one store.

8.10 Validation

Validation for the SIM reported in this chapter was slightly different. Validations were carried out on a separate part of Yorkshire and the Humber as the study area had been previous divided into two for two reasons. First, in case additional calibration was required using Nectar card data as transaction data at the postcode level was not provided for the North West study region. Secondly, computing requirements meant that a whole Yorkshire and the Humber excel SIM was very large and it was difficult to carry out calibrations due to slow load times. Results of the validations of the SIM are shown in tables 8.14 and 8.15.

Table 8.14 SIM validation global revenue forecasts

Cluster	Accuracy of forecasts				Mean Accuracy (%)
	<60%	>60%	>80%	>90%	
All Stores	47	53	40	29	70.1%

Table 8.15 SIM validation revenue forecasts by location type

Cluster	Accuracy of forecasts				Mean Accuracy (%)
	<60%	>60%	>80%	>90%	
Central Urban Cluster	63	37	26	11	50.4
Large Population Suburban Cluster	60	40	24	20	65.0
Smaller Population Suburban Cluster	31	69	51	19	71.2
Satellite Cluster	69	31	5	0	49.5
Outer Suburban Affluent	33	67	50	18	71.1
Outer Suburban Less Affluent	25	75	43	36	78.4
Rural Cluster	n/a	n/a	n/a	n/a	n/a

The validation of the model resulted in very similar results to original model testing in the central urban, smaller population suburban cluster and outskirts less affluent store location types whereas the model performed better than original model in the large population suburban cluster. The model performed worse (although generally within 4% on accuracy) for stores in the satellite rail drive and outskirts affluent location types. Once again, a next step in the future research agenda is to request additional validation data comprising many more stores that can be used to test (and possibly further calibrate) the SIM reported in this chapter.

8.11 Summary

This segmentation of the market is used in this thesis as a framework to assess the application of GIS buffer and overlay modelling, regression modelling and spatial interaction modelling for forecasting convenience grocery sales, thus empirically testing the hearsay that the latter two modelling frameworks are ineffective at forecasting convenience grocery store sales. To recap, at the outset of this research, it was hypothesised that different location types may require a different optimal strategy for good quality and robust store forecasting.

The previous chapter reported the results of the GIS buffer and overlay model for forecasting convenience grocery store sales. Methodologically, the model was a relatively simple GIS buffer and overlay procedure. The model was found to perform poorly on the whole when predicting convenience store revenue. When accounting for location type, the model performed particularly poorly in central urban, large suburban and affluent outskirts locations. Despite its poor overall performance, the model had moderate success in predicting sales in satellite, less affluent outskirts and rural locations, locations relying heavily on residential demand.

Problematic issues were raised in terms of the utility of a GIS buffer and overlay model, particularly in dealing with non-residential demand and the defining of store catchment areas. This chapter has presented a spatial interaction modelling approach to forecasting convenience grocery store sales. The model substantially outperforms the GIS buffer and overlay methodology across the majority of store locations. The mean performance of the model, at an average accuracy of around 71% is not fantastic. However, the model has a lot of success in generating very good predictions but suffers from a large proportion of very poor predictions. Methodologically, the introduction of separate residential and workplace demand layers aided in boosting predictive accuracy. In forecasting revenue in the different locations, the model performs well in suburban and outskirts locations but performs poorly in central and satellite store locations

The next chapter reports on a regression model before chapter 10 compares the three models. The reasons for the differences in predictive power of the three different models is investigated in more detail in chapter 10. Next, Chapter 9 reports on a regression model for forecasting sales in the convenience grocery market.

Chapter 9
Using a regression methodology to predict convenience grocery store sales

The first aim of this thesis was to review the existing academic and industry literature on the convenience grocery market in Great Britain in relation to; the growth of major retailers into the convenience grocery market, the demand for convenience groceries, and the attempts at forecasting revenue of both convenience grocery stores and grocery stores more generally both in academic and in the retail industry. Chapter 2 and 3 achieved this aim with the latter identifying various methodological approaches to forecasting grocery revenue of grocery stores, identifying GIS buffer and overlay, regression and spatial interaction modelling as the three approaches to forecasting convenience store sales used in this research.

The second aim of this thesis was to quantify the extent to which major retailers have committed to the convenience grocery market and assess the geographical extent to which they play a role in convenience grocery retailing in Great Britain. The analysis presented in chapter 5 of this thesis found a marked shift in the portfolio of stores operated by major grocery retailers between 2003 and 2012. Driven predominantly by Tesco and Sainsbury's full throttle pursuit of convenience, the dynamic of the major four retailers store formats on a national level has shifted towards a greater emphasis on small-format grocery retailing within the remit of the Sunday Trading Act. Between 2003 and 2012, convenience stores as a proportion of total stores increased by 40.1%, from 8.2% to 48.3% of total stores.

Moreover, from the outset of this research it was hypothesised that that different locations in which convenience grocery stores are found in GB may, in theory, require a different optimal methodology for forecasting revenue accurately. Chapter 6 reported on the segmentation of the convenience grocery market in Yorkshire and the Humber into 7 statistically distinct location types based on residential population, daytime population, transport environmental, affluence and retail vibrancy of convenience grocery store locations. The analysis found that certain locations were more favoured by certain types of retailers than others with the major grocery retailers more commonly

associated with supermarket retailing operating a higher proportion of their convenience store networks in prime, central locations.

Chapter 7 tested the effectiveness of a simple GIS buffer and overlay approach to forecasting convenience store sales across the various locational types identified in chapter 6 and found that a simple approach was neither effective nor robust in forecasting sales. Moreover, questions over the relationship between catchment areas and store sales were raised and investigated before potential alternative were discussed; one such approach is using regression analysis to predict store sales. Linear regression has been posited as the simplest way of statistically quantifying the nature and degree of interrelationship between two variables such as population and store sales (Birkin et al. 2002). We can use this technique to identify the association between a number of predictor variables (y's) and a dependent variable (x) as discussed in more detail in the literature review in chapter 3. In the case of the model presented in this chapter, the dependent variable is weekly store revenue of Sainsbury's convenience stores in Yorkshire and the Humber and the predictor variables are a number of variables deemed as potentially playing a role in creating differences in store revenue between different stores. These include residential and work based populations around a store location.

This chapter investigates the application of multiple linear regression modelling to predicting convenience store revenue both in general and in the seven specific location types identified in the segmentation of the market in chapter 6. As part of the literature review in chapter 3, the potential for a series of models calibrated to each location type was theorise. The aim of this chapter is to assess the extent to which a one-size fits all multiple regression model can be effectively applied to the forecasting of convenience store sales. In the validation chapter of this thesis (chapter 10), the potential for a series of models specific to each location type is explored in more detail and an explanation as to not incorporating this into this chapter is offered.

The chapter is structured as follows: Section 9.1 discusses the variables hypothesised to potentially play a role in store revenue generation in a store location. Next, section 9.2 reports on the methodology used to derive and attribute these variables to Sainsbury's convenience grocery stores in Yorkshire and the Humber. Section 9.3 identifies the regression procedure used to identify the best model for predicting convenience store revenue before section 9.4 reports on the final regression model equation. Finally, sections 9.5 and 9.6 report on the results of the model as a whole

and when disaggregated to test its effectiveness of predicting revenue in the 7 location types identified in chapter 6 of this thesis.

9.1 Variable Selection

In order to begin quantifying the relationship between a number of potential predictor variables and store sales, a variable list must be compiled. In this manner, using previous studies (the literature surrounding consumer interaction with convenience stores along with some exploratory freedom) a number of variables were selected as potential inputs into the regression model presented in this section. These variables fit into the broad categories:

1. Store characteristics
2. Workplace Demographics
3. Residential Demographics
4. Competition and adjacencies

9.1.1 Store Characteristics

Store size is often cited as an important consideration made by consumers when selecting their grocery destination of choice. A central starting point of many methodologies for forecasting store sales is the assumption that a consumer will most likely select the larger of two stores if all other factors (e.g. brand and distance to travel) are equal. The applied spatial interaction model presented in this thesis follows this logic and has this built in as a central assumption. The majority of convenience stores are of a similar size (between 2000 and 3000 square feet) due to the effect of retail policy on Sunday Trading Laws discussed in chapter 2 of this thesis. For this reason, it could be hypothesised that store size would play a less important role in accounting for variations in store sales between convenience stores than between convenience stores and larger supermarkets. In order to test this hypothesis, store size was included in the list of variables tested for their association with store sales.

9.1.2 Workplace Demographics

The segmentation of the convenience grocery market in Yorkshire and the Humber in chapter 6 of this thesis found distinct location types in which convenience stores are located. It was found that each location type varied in its workplace population with one cluster in particular, the central urban cluster, being in locations with large workplace populations. Much of the revenue generated in convenience stores in central locations

is acquired from work based spending through workers buying lunch and doing top-up shopping before going home. The methodology presented in this chapter explores the association of store sales with two key characteristics of the workplace population; volume and composition.

Population volume is an obvious consideration when assessing the demand for any product. Moreover, there is little reason to anticipate that this would occur to a lesser degree in a workplace population than a residential based population. Secondly, previous studies have found a relationship between the socio-economic status of consumers and their grocery brand preferences. Geodemographic classifications have often focused their attention on classifying residential areas and linking these to consumer behaviour. However, it could be reasonably expected that socio-economic factors such as income will also play a role in the grocery decision making of work based consumers. Three hypothesises were made about the composition of the work based population before work based population variables were selected.

1. Logically we would expect, other things equal, a larger volume of population within accessible reach of a store location will generate greater revenue than that of a store location with a smaller population.
2. More affluent work based consumers will be more likely than less affluent work based consumers to buy their lunches (and other groceries such as bottles of water) whilst at work instead of making them with groceries previously purchased (likely as part of a residential shopping mission).
3. More affluent consumers on average spend more money when purchasing work based groceries than less affluent consumers and are more likely to choose Sainsbury's stores as a destination as it is at the higher end of the grocery market than most retailers in terms of price and product range.

In order to investigate the effect of the volume and characteristics of the workplace population within the catchment areas of convenience grocery stores, the following variables were tested for their association with store revenue:

1. Workplace Population (Number of persons)
2. National Statistics Socio-Economic Classification 1 Population (Number of Persons)
3. National Statistics Socio-Economic Classification 2 Population (Number of Persons)
4. National Statistics Socio-Economic Classification 3 Population (Number of Persons)
5. National Statistics Socio-Economic Classification 4 Population (Number of Persons)
6. National Statistics Socio-Economic Classification 5 Population (Number of Persons)

7. National Statistics Socio-Economic Classification 6 Population (Number of Persons)
8. National Statistics Socio-Economic Classification 7 Population (Number of Persons)

The NS-Sec variables identify workers by their type of occupation. The theory being that higher managerial and professional occupations will have a greater spend and propensity to buy than those in lower socio-economic classifications and therefore, other things being equal, a higher number of these workers within the catchment area of a store will result in greater sales.

9.1.3 Residential Demographics

It is unlikely that the population within the catchment area of a store will be purely work based. Even central business districts of major cities contain a residential population. Much the same as with a workplace population, it may be the volume and characteristics of the residential population close to a store that is driving its revenue and residential consumer trips may play a significant role in amassing sales to an individual store. The census of the population in the UK is residentially focused and geocoded meaning that more information about the residential population is available in comparison to the workplace population of a given area. The relationship between a number of residential population characteristics and sales to a variety of retail outlet types has been theorised and investigated.

This study investigates the association of a number of variables and store revenue. The variables assessed fall into a number of categories which have been theorised (or are worthy of exploration) to have an association with the propensity for a person to purchase goods in convenience grocery stores or prefer a particular brand of grocery store over another. These residential population characteristics are:

1. Volume and density
2. Age
3. Mobility
4. Relative deprivation
5. Education

Total persons, and number of persons per hectare, are first used to assess the association between the volume and density of residential population and store sales. The total number of persons falling into each of the broad census age categories is then added into the model to assess the relationship between the residential population age profile and store revenue. Car ownership is also used to test the association between population mobility and store sales. Moreover, residential household

overcrowding and the occupation types in which the population are employed are used to assess the association between relative deprivation/affluence of the population and store sales. Finally, the highest level of qualification attained by each person in the population is used to assess the association between consumer education levels and store sales.

9.1.4 Retail and transport adjacencies

As well as the characteristics of an individual store, the workplace population within a store catchment and the residential population of a store catchment, other additional factors may have an impact on store revenue. For the purposes of this study, these are termed retail and transport adjacencies. The adjacency variables are built into the model to control for a number of factors that have been found (or theorised) to have an impact on store sales:

1. Competing grocery opportunities attracting potential customers into a store catchment area.
2. Other retail opportunities attracting potential customers into a store catchment area.
3. Other grocery stores of the home retailer causing an overcrowded market.
4. Other grocery stores of the home retailer causing a positive network effect which raises brand awareness and promotes greater sales at all available branches.
5. Transport links attracting and facilitating potential customers to be in a store catchment area.

In order to address these hypotheses, the following variables were built into the regression model presented in this chapter:

1. Total grocery floorspace (total competition)
2. Total convenience grocery floorspace (convenience competition)
3. Sainsbury's supermarket floorspace
4. Sainsbury's convenience store floorspace
5. Total retail stores (other than grocery)
6. Total rail passengers

The addition of the adjacency variables allows the model to look for the association between store characteristics, residential population characteristics, workplace population characteristics and retail and transport adjacencies and store revenue for

convenience stores operated by Sainsbury's. Rail passengers were included as a number of convenience stores are found close to, or in, rail stations which can have a very large volume of people travelling through per day, a large potential number of customers for a convenience store located nearby. The addition of variables covering these categories resulted in a final variables list, as seen in table 9.1.

Table 9.1 Regression Model List of Variables

Variables	
Store Characteristics	White British Population
Store Size	Overcrowding (-1 or less)
Workplace Demographics	Overcrowding (+2)
Workplace Pop	Overcrowding (+1)
Workers NSSec 1	Overcrowding (0)
Workers NSSec 2	Overcrowding (-1)
Workers NSSec 3	Overcrowding (-2 or less)
Workers NSSec 4	Population Density
Workers NSSec 5	Population Age 16+
Workers NSSec 6	No Qualification (Persons)
Workers NSSec 7	Level 1 Qualification (Persons)
Residential Demographics	Level 2 Qualification (Persons)
Households	Apprenticeship Qualification (Persons)
Population	Level 3 Qualification (Persons)
Population Aged 0 to 4	Level 4 Qualification (Persons)
Population Aged 5 to 7	Single Person Households
Population Age 8 to 9	All Persons (16-74)
Population Age10 to14	Residential NSSec 1
Population Age16 to 17	Residential NSSec 2
Population Age 18 to 19	Residential NSSec 3
Population Age 20 to 24	Residential NSSec 4
Population Age 25 to 29	Residential NSSec 5
Population Age 30 to 44	Residential NSSec 6
Population Age 45 to 59	Residential NSSec 7
Population Age 60 to 64	Residential NSSec 8
Population Age 65 to 74	Competition and adjacencies
Population Age 75 to 84	All Grocery Floorspace
Population Age 85 to 89	All JS Grocery Floorspace
Population Age 90 to 99	Convenience Competition Floorspace
Population Age 100 plus	All Retail Stores (Incl. non grocery)
Mean Population Age	JS Supermarket Floorspace
Median Population Age	JS Convenience Floorspace
No Car Households	Rail Passengers (Annual)

9.1.5 How the regression model evaluates the association of different variables with store revenue?

In practice, multiple linear regression tests the extent to which the variance observed in the dependent variable (in this case store revenue) can be accounted for by incremental changes (variance) in two or more predictor variables. As argued in this chapter, a large number of variables may have a plausible theoretical relationship with store sales. The advantage of multiple linear regression, particularly when mobilised using a best subset analysis as shown later in this chapter, is the ability to detect which of the predictor variables (and their variance) account for the most variation in the store revenue variable.

The nature of this methodology means that two (or more) variables may have an association with store sales, yet the whole contribution of both variables could be accounted for by one of these variables. For example, we may hypothesise that an increase in residential population in a one mile buffered catchment of a store will have an association with store sales. Moreover, we may hypothesise that the number of persons employed in professional and managerial (social class 1) occupations will also have an association with store sales and we could plausibly anticipate that both variables should feature in a final model solution. However, the variance in one of these predictor variables may account for all of the variation (and more) in the other predictor variable in relation to the variation in store revenue and therefore exclude the other variable from store sales. Having discussed the list of variables and described the process by which variables may be included (or excluded) in the final model equation, it is important to describe the process by which input predictor variables were defined in this study.

9.2 Methodology for deriving variables

Once all the data for the full variable list was collected, a methodology for attributing the variable data to each individual store was required. The method adopted was an SQL query in MapInfo Professional to conduct point/polygon in polygon analysis to attribute the store characteristics, work based demographics, residential demographics and other adjacencies to each stores catchment area. In order to identify the strongest relationships, varying buffered catchment area sizes were used to investigate the relationship between variable attributes and revenue at different distances around a store. These were:

1. 0.5 mile radius buffer
2. 1 mile radius buffer
3. 1.5 mile radius buffer
4. 2 mile radius buffer

The use of variable buffer sizes allows the quantification of the catchment area at which each variable has the greatest association with store sales. 2 miles was selected as the maximum radius by looking at catchment areas of stores in the nectar card data discussed in chapter 4 and the decision that a 30 minute walking radius (approximately 2 miles) is a sensible maximum distance at which people would travel on foot to a store. The advantage of using a regression methodology over a GIS buffer and overlay methodology is that regression modelling investigates the statistical relationship between a number of variables and store revenue, meaning that identifying the exact catchment area from which customers travel to a store is less important. It is possible to accurately predict sales without identifying an accurate functional catchment area as the model is not directly attributing consumer expenditure to a given store, rather the model is quantifying the relationship between a number of variables and store revenue.

9.2.1 Correlating variable values with store revenue

The first stage of the regression analysis process was to identify the variables that have the greatest correlation (positive or negative) with Sainsbury's convenience store revenues. Moreover, this type of analysis used the four buffered catchment sizes to identify the catchment size at which each variable has the highest correlation with sales. Table 9.2 shows the correlations between each variable and store sales at each buffer size. The 'Best Buffer' column contains the buffer size at which each variable has the strongest relationship with store revenue.

Table 9.2 Correlation of predictor variables with store revenue

Variable	Buffer Size (Miles)				Relationship (MAX)	Best Buffer	Rank Relationsip
	0.5	1.0	1.5	2.0			
Store Size	0.26	0.26	0.26	0.26	0.26	Equal	55
Workplace Pop	0.40	0.41	0.47	0.44	0.47	1.5	8
Workers NSSec 1	0.39	0.40	0.47	0.46	0.47	1.5	5
Workers NSSec 2	0.40	0.40	0.47	0.45	0.47	1.5	7
Workers NSSec 3	0.39	0.37	0.44	0.43	0.44	1.5	12
Workers NSSec 4	0.51	0.47	0.46	0.44	0.51	0.5	3
Workers NSSec 5	0.43	0.45	0.46	0.42	0.46	1.5	9
Workers NSSec 6	0.40	0.41	0.43	0.40	0.43	1.5	16
Workers NSSec 7	0.36	0.33	0.37	0.34	0.37	1.5	35
Households	0.32	0.38	0.39	0.39	0.39	2	24
Population	0.27	0.36	0.38	0.39	0.39	2	26
Population Aged 0 to 4	-0.18	0.03	0.21	0.29	0.29	2	47
Population Aged 5 to 7	-0.25	-0.03	0.17	0.26	0.26	2	57
Population Age 8 to 9	-0.27	-0.09	0.13	0.24	0.27	0.5	51
Population Age10 to14	-0.28	-0.07	0.14	0.24	0.28	0.5	49
Population Age16 to 17	-0.27	-0.04	0.17	0.25	0.27	0.5	51
Population Age 18 to 19	0.23	0.41	0.38	0.39	0.41	1	21
Population Age 20 to 24	0.40	0.43	0.42	0.43	0.43	1	14
Population Age 25 to 29	0.45	0.43	0.43	0.42	0.45	0.5	11
Population Age 30 to 44	0.17	0.28	0.35	0.37	0.37	2	30
Population Age 45 to 59	-0.09	0.10	0.24	0.30	0.30	2	45
Population Age 60 to 64	-0.10	0.08	0.22	0.27	0.27	2	53
Population Age 65 to 74	-0.18	0.04	0.22	0.28	0.28	2	48
Population Age 75 to 84	-0.16	0.03	0.21	0.30	0.30	2	46
Population Age 85 to 89	0.00	0.14	0.27	0.34	0.34	2	41
Population Age 90 to 99	0.09	0.19	0.33	0.38	0.38	2	27
Population Age 100 plus	0.10	0.07	0.23	0.32	0.32	2	43
Mean Population Age	0.24	0.36	0.38	0.38	0.38	2	28
Median Population Age	0.17	0.33	0.36	0.37	0.37	2	30
No Car Households	0.26	0.33	0.36	0.38	0.38	2	29
White British Population	0.26	0.39	0.38	0.38	0.39	1	25
Overcrowding (-1 or less)	0.22	0.27	0.32	0.34	0.34	2	40
Overcrowding (+2)	0.06	0.20	0.30	0.35	0.35	2	39
Overcrowding (+1)	0.20	0.33	0.35	0.37	0.37	2	32
Overcrowding (0)	0.40	0.41	0.41	0.41	0.41	1.5	19
Overcrowding (-1)	0.23	0.30	0.34	0.36	0.36	2	36
Overcrowding (-2 or less)	0.13	0.12	0.19	0.24	0.24	2	59
Population Density	0.29	0.38	0.41	0.42	0.42	2	17
Population Age 16+	0.34	0.41	0.40	0.40	0.41	1	20
No Qualification (Persons)	-0.37	-0.14	0.09	0.20	0.37	0.5	33
Level 1 Qualification (Persons)	-0.22	0.01	0.18	0.26	0.26	2	56
Level 2 Qualification (Persons)	-0.01	0.22	0.30	0.33	0.33	2	42
Apprenticeship Qualification (Persons)	-0.22	-0.06	0.10	0.20	0.22	0.5	60
Level 3 Qualification (Persons)	0.38	0.44	0.41	0.43	0.44	1	13
Level 4 Qualification (Persons)	0.59	0.53	0.47	0.45	0.59	0.5	1
Single Person Households	0.35	0.39	0.40	0.40	0.40	2	23
All Persons (16-74)	0.35	0.41	0.40	0.40	0.41	1	18
Residential NSSec 1	0.59	0.51	0.46	0.44	0.59	0.5	2
Residential NSSec 2	0.49	0.48	0.44	0.43	0.49	0.5	4
Residential NSSec 3	0.19	0.32	0.35	0.37	0.37	2	33
Residential NSSec 4	0.08	0.19	0.31	0.35	0.35	2	38
Residential NSSec 5	-0.13	0.05	0.19	0.25	0.25	2	58
Residential NSSec 6	-0.20	0.02	0.19	0.27	0.27	2	53
Residential NSSec 7	-0.31	-0.12	0.08	0.17	0.31	0.5	44
Residential NSSec 8	-0.16	-0.01	0.13	0.20	0.20	2	61
All Grocery Floorspace	0.29	0.36	0.34	0.34	0.36	1	37
All JS Grocery Floorspace	0.27	0.21	0.15	0.11	0.27	0.5	50
Convience Competiton Floorspace	0.43	0.46	0.45	0.42	0.46	1	9
All Retail Stores (Incl. non grocery)	0.39	0.45	0.47	0.45	0.47	1.5	5
JS Supermakret Floorspace	0.18	0.13	0.08	0.04	0.18	0.5	62
JS Convenience Floorspace	0.43	0.30	0.27	0.25	0.43	0.5	15
Rail Passengers (Annual)	0.34	0.37	0.36	0.41	0.41	2	22

As table 9.2 shows, variables vary in the strength of their correlation with sales depending on the buffer size at which they are measured. The next stage is to rank the variables by the strength of their correlations with store sales based on the buffer size at which the relationship is the strongest. Table 9.3 shows the variables ranked by their strongest correlation with store revenue, with a rank of 1 being the greatest relationship.

Table 9.3 Rank of strength of correlations between predictor variables and store revenue

Variable	Strongest Correlation (±)	Rank Correlation
Level 4 Qualification (Persons)	0.59	1
Residential NSSec 1	0.59	2
Workers NSSec 4	0.51	3
Residential NSSec 2	0.49	4
Workers NSSec 1	0.47	5
All Retail Stores (incl. non grocery)	0.47	5
Workers NSSec 2	0.47	7
Workplace Pop	0.47	8
Workers NSSec 5	0.46	9
Convenience Competition Floorspace	0.46	9
Population Aged 25 to 29	0.45	11
Workers NSSec 3	0.44	12
Level 3 Qualification (Persons)	0.44	13
Population Aged 20 to 24	0.43	14
JS Convenience Floorspace	0.43	15
Workers NSSec 6	0.43	16
Population Density	0.42	17
All Persons (16-74)	0.41	18
Overcrowding (0)	0.41	19
Population Aged 16+	0.41	20
Population Aged 18 to 19	0.41	21
Rail Passengers (Annual)	0.41	22
Single Person Households	0.40	23
Households	0.39	24
White British Population	0.39	25
Population	0.39	26
Poulation Aged 90 to 99	0.38	27
Mean Population Age	0.38	28
NoCar Households	0.38	29
Population Aged 30 to 44	0.37	30
Median Population Age	0.37	30
Overcrowding (+1)	0.37	32
No Qualifications (Persons)	0.37	33
Residential NSSec 3	0.37	33
Workers NSSec 7	0.37	35
Overcrowding (-1)	0.36	36
All Grocery Floorspace	0.36	37
Residential NSSec 4	0.35	38

Overcrowding (+2)	0.35	39
Overcrowding (-1 or less)	0.34	40
Poulation Aged 85 to 89	0.34	41
Level 2 Qualification (Persons)	0.33	42
Poulation Aged 100 plus	0.32	43
Residential NSSec 7	0.31	44
Population Aged 45 to 59	0.30	45
Poulation Aged 75 to 84	0.30	46
Population Aged 0 to 4	0.29	47
Poulation Aged 65 to 74	0.28	48
Population Aged 10 to 14	0.28	49
All JS Grocery Floorspace	0.27	50
Population Aged 8 to 9	0.27	51
Population Aged 16 to 17	0.27	51
Poulation Aged 60 to 64	0.27	53
Residential NSSec 6	0.27	53
Store Size	0.26	55
Level 1 Qualification (Persons)	0.26	56
Population Aged 5 to 7	0.26	57
Residential NSSec 5	0.25	58
Overcrowding (-2 or less)	0.24	59
Apprenticeship Qualification Persons	0.22	60
Residential NSSec 8	0.20	61
JS Supermarket Floorspace	0.18	62

The correlations found between store revenue and each variable is identified in table 9.3. There are clear distinctions between the strength of the strongest correlations when compared with the weakest correlations. The variable with the greatest correlation with store revenue is persons with a level 4 qualification as their highest level of qualification which has a correlation of 0.59 with store sales.This may be due to this variable accounting for three different aspects of convenience demand. It is highly correlated with total population, highly correlated with high social class consumers who are more likely than average to shop at Sainsbury's and may have an association with young, cosmopolitan professional consumers who may have a higher than average demand for the range of goods offered by Sainsbury's. This is at the top end of the band what is widely considered to be the breadth of moderate correlations, those correlations that fall between 0.4 and 0.6.

The correlation of 23 of the variables with store revenue falls within the band of moderate correlations. This is encouraging and suggests that a model accounting for much of the variation in store sales is possible which yields a strong potential for developing a multiple regression equation with good predictive capacity. Of the variables tested, available Sainsbury's supermarket (stores greater than 3,000 sq. ft. in size) floorspace has the lowest correlation with store revenue at 0.19, a weak correlation. Whilst it is implausible that this variable will account for a large proportion

of the variation in store sales, it may still feature in the final model equation due to the nature of the best subset regression analysis detailed below.

9.3 Best subset analysis

The next step was to start to build a linear regression model. In order to build a linear regression model, there are options in terms of the methodology undertaken. The methodology preferred by this study was to use a best subset analysis in Minitab 17 Statistical Software. In order to make comparisons across different models and data, standard measures of the models accuracy can be used, the most common of which is the R^2 statistics. This is the coefficient of determination which is the proportion of the variation that can be accounted for by the model (Faraway, 2002).

Best subset analysis in Minitab allows a researcher to input a number of predictor variables into a linear regression model and the method will give the model with the greatest R^2 (the best-fitting model) for each number of predictor variables (1, 2, 3, 4, 5 … etc) specified. We can chart the change in adjusted R-Squared value with the addition of predictor variables in the model and identify the point at which adding additional explanatory variables yields diminishing returns and is therefore excessive and overcomplicates the procedure. The method allows 12 predictor variables to be entered into the model at any time and the procedure then delivers the model which accounts for the greatest variation in the dependent variable (store revenue) for each number of predictor variables. The rank of variables shown in table 9.3 was used to determine the order by which variables were entered into the model, starting with the 12 variables with the highest correlation with store revenue. These are shown in table 9.4.

Table 9.4 12 variables with greatest correlation with store revenue

Variable Number	Description	Correlation	Buffer	Rank
1	Level 4 Qualification (Persons)	0.59	0.5	1
2	Residential NSSec 1	0.59	0.5	2
3	Workers NSSec 4	0.51	0.5	3
4	Residential NSSec 2	0.49	0.5	4
5	Workers NSSec 1	0.47	1.5	5
6	All Retail Stores (Incl. non grocery)	0.47	1.5	5
7	Workers NSSec 2	0.47	1.5	7
8	Workplace Pop	0.47	1.5	8
9	Workers NSSec 5	0.46	1.5	=9
10	Convenience Competition Floorspace	0.46	1	=9
11	Population Aged 25 to 29	0.45	0.5	11
12	Workers NSSec 3	0.44	1.5	12

Following the input of these 12 variables into a best subset analysis, the methodology returns an output detailing the best-fitting model (with the greatest R^2) for each number of predictor variables, the variables that are required for each solution for each number of predictor variables and a diagnostic statistic known as Mallows Cp for each model. The output for the 12 variables with the highest correlation of sales is shown in table 9.5.

Table 9.5 Initial 12 Variable Best Subset Regression Model

Vars	R-Sq	R-Sq (adj)	Mallows Cp	Variable Number (by rank)											
				1	2	3	4	5	6	7	8	9	10	11	12
1	34.5	33.8	15.7		X										
2	39.3	38	9.9	X		X									
3	41.9	40	7.6	X		X								X	
4	42.7	40.2	8.3	X		X						X		X	
5	45.4	42.3	6	X		X		X	X	X					
6	47.8	44.2	4.1	X		X		X		X		X		X	
7	49.1	45	4	X		X		X	X	X	X			X	
8	49.5	44.8	5.3	X		X		X	X	X		X		X	X
9	49.5	44.2	7.2	X		X		X	X	X		X	X	X	X
10	49.6	43.6	9.1	X		X	X	X	X	X	X	X		X	X
11	49.7	43	11	X	X	X		X	X	X	X	X	X	X	X
12	49.7	42.3	13	X	X	X	X	X	X	X	X	X	X	X	X

Of the 12 variables with the strongest correlation with store sales, a 7 variable solution is the optimum as it accounts for the greatest variation of store revenue for Sainsbury's convenience grocery stores in Yorkshire and the Humber ($R^2 = 45\%$). However, it is now possible to discount the 5 variables that did not contribute to the best solution and then replace these with the next variable with the strongest correlation with the store sales. This process was repeated until the discounting of variables and subsequent inclusion of other variables did not produce a better fitting model. The final solution is a 10 variable solution, as shown in table 9.6.

Table 9.6 Best solution regression model

Vars	R-Sq	R-Sq (Adj)	Mallows Cp	Variable number (by rank)											
				1	3	5a	5b	7	12	14	16	17	23	34	35
10	54.9	49.5	9.9	X	X	X	X	X		X	X	X	X	X	
11	55.3	49.3	11.1	X	X	X	X	X		X	X	X	X	X	X
11	55.1	49	11.6	X	X	X	X	X	X	X	X	X	X	X	
12	55.4	48.8	13	X	X	X	X	X	X	X	X	X	X	X	X

The 10 variables involved in the best solution (yielding an R^2 of 49.5%) are as follows:

1. Level 4 Qualification (Persons)
2. Workers NSSec 4
3. Workers NSSec 1
4. All Retail Stores
5. Workers NSSec 2
6. Level 3 Qualification (Persons)
7. JS Convenience Floorspace
8. Workers NSSec 6
9. Rail Passengers (Annual)
10. No Qualifications (Persons)

The resulting best-fitting solution involves a combination of work based population, residential population and adjacencies variables. These include variables covering all 4 different buffer sizes.

9.3.1 Autocorrelation among predictor variables

However, using a linear equation to predict sales in this way is problematic, and doesn't work effectively due to some of the high correlations (±0.8) between the predictors. In the case of high level of correlation, the model struggles to decipher which of the variables if accounting for the variation in store sales. The correlation of variables within the 10 variable solution are seen in table 9.7.

Table 9.7 – Correlation of predictors in 10 variable regression solution

Correlation of predictors	Level 4 Qualification (Persons)	Workers NSSec 4	Workers NSSec 1	All Retail Stores (Incl. non grocery)	Workers NSSec 2	Level 3 Qualification (Persons)	JS Convenience Floorspace	Rail Passengers	No Qualification (P0.80ersons)
Level 4 Qualification	1.00								
Workers NSSec 4	0.56	1.00							
Workers NSSec 1	0.67	0.75	1.00						
All Retail Stores	0.68	0.73	0.87	1.00					
Workers NSSec 2	0.70	0.76	0.98	0.93	1.00				
Level 3 Qualification	0.76	0.37	0.71	0.62	0.72	1.00			
JS Conv F'space	0.45	0.64	0.68	0.60	0.69	0.50	1.00		
WZNSSec 6	0.64	0.73	0.89	0.96	0.95	0.65	0.66	1.00	
No Qualification	-0.22	-0.21	-0.34	-0.17	-0.27	-0.21	-0.35	-0.10	-0.17	1.00

As table 9.7 shows, correlations above ±0.80 exist between variables, particularly between the multiple workplace population variables. In order to produce a model that reliably predicts convenience store revenue, a method of adjusting the variable list to remove and reduce problematic multicollinearity is required. A widely used method of reducing multicollinearity is principal components analysis (PCA).

9.3.1 Principal Components Analysis (PCA)

Principal components analysis is a widely used technique for overcoming high levels of correlation between predictor variables by combing variables into a single variable. The resulting variable attempts to account for as much of the variation in each individual variable as possible and the resulting variable can then replace each individual variable in the linear regression equation to predict sales, thus creating a more conceptually coherent set of variables (Dunteman, 1989). The resulting variables are now uncorrelated with the other predictor variables that remain within the model equation.

Using this technique, two new variables were generated for use in the multiple regression model presented in this chapter. These new variables overcome the collinearity of the workplace population variables and the residential population education variables in the original variable list. Along with overcoming the collinearity problem in the model, they also boost the model's predictive ability and guard against using coefficients which could result in the model being inappropriate for predicting store revenue outside of the Yorkshire and Humber sample used to calibrate the model in this chapter.

The two new variables were scaled between 0 and 1 and allocated an individual value for each case (in this case store catchment area). The new principal components workplace population variable takes into account the following variables that were independently found to contribute to the variance in Sainsbury's store sales and also found to be highly correlated with at least one of the other workplace population variables:

1. Workers NSSec 1
2. Workers NSSec 2
3. Workers NSSec 4
4. Workers NSSec 6

Moreover, the new principal components residential population education variable takes into account the following variables that were independently found to contribute to the variance in Sainsbury's store sales and also found to be highly correlated (+0.76) with at least one other education variable:

1. Level 3 Qualification
2. Level 4 Qualification

The resulting variables are subsequently referred to as:

- **WPZ PCA** referring to the workplace zone population principal components variable. (cumulative variance explained – 91%).
- **EDU PCA** referring to the residential population education principal components variable. (cumulative variance explained - 88%).

Tables 9.8a and 9.8b contain the results of the PCA which resulted in the WPZ principal components variable at the EDU principal components variable

Table 9.8 A) WZ PCA Component Matrix, and B) EDU PCA Component Matrix

A - WZ PCA - Component Matrix

Variable	Component
WZNSSec1	0.975
WZNSSec 2	0.988
WZNSSec 4	0.892
WZNSSec 6	0.964
Variance Explained	91%

B - EDU PCA Component Matrix

Variable	Component
Lvl3	0.934
Lvl4	0.934
Variance Explained	88%

The WPZ PCA variable explains a cumulative 91% of the variance in the dependent variable accounted for by the 4 WPZ predictor variables included in the principal components analysis. Moreover, the education PCA variable referred to as EDU PCA explains 88% of the variance accounted for by the two education variables included in the second principal components analysis.

9.4 The Final Regression Model

Table 9.9a presents the final coefficients of the linear regression model reported in this chapter. The standardised coefficients identify the variables playing the most significant role in the R squared value. The education principal components variable is, other things equal, the variable accounting for the greatest proportion of the variance in

convenience grocery store sales. Additionally, table 9.9b identified the R-Squared and adjusted R-Squared values achieved by the model. The variables in the model account for 51.9% of the variance in convenience grocery store sales. Additional variables did not add to this figure. Moreover, the adjusted R-square value suggests the model would account for 47.2% of the variance in convenience store sales out of sample.

Table 9.9 a) Regression coefficents, and **b)** R-Squared values

Variable	Unstandardized		Standardized
	B	**Std. Error**	**Beta**
(Constant)	38791.83	n/a	n/a
AllRet	11.06	18.31	0.15
JSConv	1.34	1.17	0.15
Rail Pass	0.0002	0.00	0.08
WZ PCA	-2565.06	6167.49	-0.16
EDU PCA	7553.46	1927.28	0.47
No Quals	-4.72	210	0.07

Statistic	Value
R-Sq	51.9%
R-Sq (Adj.)	47.2%

Moreover, the final regression equation for predicting store sales using this method is shown in table 9.10.

Table 9.10 Final linear regression equation

Weekly store revenue $=$

£38791.83 (Constant)

$+$ (Total retail stores in a 1.5 mile radius buffer $*$ £11.06)

$+$ (Sainsbury's convenience floorspace in a 2 mile radius buffer $*$ £1.34)

$+$ (Annual rail passengers in a 0.5 mile radius buffer $*$ £0.0002)

$-$ (Workplace Zone PCA value $*$ £2565.05)

$+$ (Residential Education PCA Value $*$ £7553.46)

$-$ (No Qualifications in a 0.5 mile radius buffer $*$ £4.72)

The next step is to assess the quality of store revenue forecasting using this method.

9.5 Global Sales Forecasting

The overall forecasting ability of the global regression model across all store locations is shown in table 9.11. The regression model presented in this section is relatively successful at predicting sales on the whole, at a mean accuracy of 77.5%. This model is very effective at limiting poor model predictions with less than 10% of predictions falling below 60% accuracy. Moreover, the model is moderately successful at

producing predictions at over 80% accuracy and has some success in predicting sales at over 90% in accuracy.

Table 9.11 Global accuracy of regression model forecasts

	Accuracy of forecasts				
Cluster	<60%	>60%	>80%	>90%	Mean (%)
All Stores	16	84	53.1	30.9	77.5

From these observations, the model has clear merit in being applied as a generic forecasting tool without the added detail of breaking down sales predictions by location type. However, in disaggregating these predictions by convenience location type, it is possible to identify whether or not the model is more (or less) appropriate for application in certain location types than in others. Table 9.12 shows the accuracy of store sales disaggregated by the convenience location types identified in chapter 6.

9.6 Cluster by cluster sales forecasting

Breaking down forecasting accuracy by cluster type identifies variances in the model's ability to forecast sales by location type. The model performs relatively poorly in the central urban and suburban locations in comparison with the satellite and outskirts locations in which the model predicts sales to an impressive level of accuracy.

Table 9.12 Accuracy of regression model forecasts by location type

	Accuracy (% of stores)				
Cluster	**<60**	**>60**	**>80**	**>90**	**Mean (%)**
Central Urban Cluster	23.1	76.9	38.5	30.8	**71.2**
Large Population Suburban Cluster	27.3	72.7	40.9	9.1	**72.4**
Smaller Population Suburban Cluster	17.2	82.8	55.2	31	**76.4**
Satellite Cluster	0	100	66.7	33.3	**84.9**
Outer Suburban Affluent	0	100	83.3	66.7	**90.7**
Outer Suburban Less Affluent	9.1	90.9	45.5	36.4	**78.8**
Rural Cluster	0	100	100	0	**87.2**

9.6.1 Central Urban Locations

The global regression model has a mixed level of success in predicting store revenue in the central urban cluster. The model predicts a fairly large proportion of store sales (38.5%) at an accuracy of over 80%, with over 30% of store revenue predictions falling above a 90% level of accuracy, meaning that the model is producing an acceptable proportion of good to very good predictions. However, this model is susceptible to very poor revenue estimates. Almost a quarter of estimates fall below the minimum 60% threshold. On average, the model predicts store revenue in central urban locations at a modest (but unreliable) 71.2% accuracy, making it the worst performing location type for this model. Thus, it would be difficult to justify the use of this global regression model for a wider application to this type of convenience store.

9.6.2 Suburban Locations

Similar to the central urban cluster, the model has a mixed level of success estimating store revenues in large population suburban store locations. Once again, the model is unsuccessful in guarding against very poor estimates in this location type with around a quarter of predictions falling below the base 60% threshold. Moreover, the model is reasonably successful in producing good predictions with over 40% of estimates evaluated as being good, a slight improvement on the central urban cluster. However, the model struggles to produce very good model predictions with less than 10% (9.1%) of estimates reaching the 90% threshold. On average, the model has a similar level of success in large population suburban locations (72.4%) and central urban locations (71.2%) when looking at the mean accuracy of revenue estimates making it again difficult to justify its wider use as a predictive tool for stores of these types.

When compared to the model's success at estimating store sales in the large population suburban cluster, the model performs better for stores in the smaller population suburban cluster. This location type experiences 10% less very poor sales predictions (17.2%), significantly more sales predictions at a good level of accuracy (55.2% = +14.3%) and a far greater proportion of sales predictions at a very good level of accuracy (31.0% = +21.9%). Moreover the model performs on average 4% better in the smaller population suburban cluster with a mean predictive accuracy of over 75% (75.4%). Although the model's average performance is modest for this cluster, the potential merit of using a regression approach for sales predictions can be seen in the model's performance in this cluster, particularly when considering the regularity of good and very good revenue predictions in smaller population suburban store locations.

9.6.3 Satellite Locations

When considering one of the major goal is the eradication of poor revenue predictions, this model performs very well for stores in satellite locations with no stores being predicted below the 60% accuracy threshold. This has the potential to eradicate potentially disastrous new store openings as it gives a firm upper and lower limit in which we can reliably trust that store sales would fall. Additionally the model predicts two thirds (66.7%) of sales revenues at over 80% accuracy and one third (33.3%) of estimates at an accuracy greater than 90%. On average, the global regression model presented in this chapter predicts store sales at almost 85% accuracy. This coupled with the impressive proportion of predictions above 90% and the absence of very poor sales predictions makes this model very viable for wider application to other Sainsbury's stores in satellite locations and potential new store openings in this type of location.

9.6.4 Outskirts Locations

Moving on to the predictive capacity in outskirts locations, there is a clear distinction in this model's performance between affluent and less affluent outskirts store locations. The model performs worse in the less affluent store locations which average 13.5% of residents being employed in social class 1 occupations, less than half of the proportion of residents in the more affluent outskirts cluster. Moreover, the model's performance in the affluent outskirts cluster is very good and will be discussed in detail below.

In less affluent outskirts store locations, the model is quite successful in guarding against very poor predictions, with less than 10% of estimates falling below the 60% base accuracy threshold. Moreover, the model is reasonably successful in estimating sales in this location type at a good level of accuracy (45.5% of predictions) and a very good level of accuracy with one third of estimates at a greater than 90% accuracy. The model performs at an average accuracy of 78.8% for this location type and its predictive profile is similar to that for the smaller population suburban cluster. Although the model's average performance is modest for this cluster, the potential merit of using a regression approach to sales predictions can be seen in the model's performance in this cluster, particularly when considering the regularity of good and very good revenue predictions in smaller population suburban store locations.

The model performs very well in the affluent outskirts store location type. As experienced in three other location types, the regression model produces no predictions at below 60% accuracy in this store type. Furthermore, the vast majority of

predictions (83.3%) fall above the 80% level of sales representing good sales forecasts and an impressive 66.7% of predictions are above the 90% threshold reserved for excellent estimates of store revenue. When looking at the whole series of store revenue estimates for this location type, an average sales forecast accuracy of 90.7% is observed. The model has the highest level of performance in this location type and its potential is exciting. A combination of no sales predictions at a very poor level and almost two thirds of revenue estimates reaching an excellent level of accuracy makes this model highly appropriate for a wider application to stores of this type operated by Sainsbury's both in Yorkshire and the Humber and the rest of Great Britain.

Finally, the one store in the rural location type is predicted very well at an 87.2% level of accuracy. Once again, it is difficult to confirm the performance of the model in this cluster due to the small sample size. However, this would be of limited concern to Sainsbury's as the retailer has little commitment to operating convenience retail space in rural location types.

9.7 Validations

The validation process for the regression model was the same as for the GIS buffer and overlay model and tables 9.13 and 9.14 present the results of forecasting store sales for the validation stores in the North West of England using the regression model reported in this chapter.

Table 9.13 Regression validation global accuracy

| Cluster | NW Global Regression Validation | | | | |
	<60%	>60%	>80%	>90%	Mean (%)
All Stores	23	77	42	23	75.1

Table 9.14 Regression validation accuracy by location type

| Cluster | Accuracy of Predictions (%) | | | | |
	<60%	>60%	>80%	>90%	Mean (%)
Central Urban	42.9	57.1	42.9	28.6	**66.0**
Large Population Suburban	14.3	71.4	28.6	0.0	**70.1**
Smaller Population Suburban	16.7	83.3	50.0	50.0	**77.3**
Satellite	25.0	75.0	75.0	50.0	**82.5**
Outskirts affluent	0.0	100.0	0.0	0.0	**73.9**
Outskirts less affluent	20.0	80.0	40.0	0.0	**72.9**
Rural	N/A	N/A	N/A	N/A	**N/A**

Overall, the model performs slightly worse for the North West validation stores at an average accuracy of 75.1% in comparison to 77.1% for stores in Yorkshire and the Humber. Once again, it is difficult to identify performance by cluster due to the mismatch in store numbers between the calibration and test stores but the accuracy of the model holds up well. The exception to this rule is the affluent outskirts cluster which is better predicted for Yorkshire and the Humber than the North West. However, there are less than half the total stores for this cluster in the North West in comparison to Yorkshire and the Humber. Further validation is planned on a define set of stores to achieve a clearer picture of how well each model performs outside of Yorkshire and the Humber.

9.8 Summary

Once again, the segmentation of sales into different locational types was used as a framework to assess the merit of a regression modelling framework for predicting convenience grocery store sales, thus empirically testing the hearsay that more complex frameworks are less effective than GIS buffer and overlay methods at forecasting convenience grocery store sales. To recap, at the outset of this research, it was hypothesised that different location types may require a different optimal strategy for good quality and robust store forecasting.

The previous chapter reported the results of a disaggregate SIM for forecasting convenience grocery store sales. Methodologically, the model improved on the GIS buffer and overlay procedure and performed much better on average. The model has a lot of success in generating very good predictions but suffers from a large proportion of very poor predictions. Methodologically, the introduction of separate residential and workplace demand boosted revenue estimates but the complex nature of interactions between supply and demand in central locations meant that this type of location was still forecasting poorly.

This chapter has reported on a regression model for forecasting convenience grocery sales. This model is the best performing of the three on average and is substantially better than the other two methods at limiting very poor predictions. Retailers worry about very poor predictions due to the risk of store closure that is particularly possible if very poor under predictions are achieved. However, the model performed worse than the SIM in generating good and very good forecasts, those reaching over 80% and 90% in accuracy. The model performed much better in central and large suburban locations than the other methodologies although successes were still moderate. The

model performed very well in forecasting revenue for satellite, rural and outskirts affluent locations, exceeding an average of 80% accuracy (90% in the affluent outskirts location type).

Next, chapter 10 compares the results of the three models both in general, and across the 7 location types identified in chapter 6 of this thesis. In doing so, the reasons for differences in predictions between the different models is discussed in greater detail, highlighting the advancements made and the work still to do in forecasting convenience grocery store sales using these methods.

Chapter 10
Comparison of model's and conclusions

The research presented in this thesis has attempted to gain a better understanding of convenience grocery retailing in Great Britain. A major aspect of this research lies in attempting to understand if it is possible to reliably predict sales in the various location types in which branded convenience grocery stores are found. This has proven a difficult task both for retailers and consultancy firms. Moreover, appearances of convenience grocery store forecasting in the academic literature have been both sparse and have often lacked extensive empirical testing.

To aid in forecasting, chapter 6 presented a k-means classification of the convenience grocery market in Yorkshire and the Humber, disaggregating the market into 7 distinct locational types based on the characteristics of their catchment areas. Moreover, it was hypothesised that stores for different locations would be driven by varying consumer behaviours which would, in turn, require different approaches in order to robustly predict sales to different types of convenience store. A simple GIS buffer and overlay model from which to benchmark the success of more complex methodologies was presented in chapter 7, before the potential utility of two common methodological approaches were investigated in respect of their ability to forecast convenience grocery store sales. Chapter 8 presented a regression modelling approach to convenience sales forecasting whilst chapter 9 presented a spatial interaction modelling approach to predicting convenience store revenue.

This chapter draws together the results of the three approaches presented in chapters 7, 8 and 9, focusing on the effectiveness of the different methodologies in the different locational contexts in which convenience retailing takes place. Section 10.1 looks at the overall ability of each model type to predict sales regardless of locational type before section 10.2 addresses the various methodological approaches in the context of predicting sales in the 7 locational types identified in the network segmentation in chapter 6. Following the comparison of methods, section 10.3 draws together the overall findings of the research, summarising the findings of each piece of analysis and suggests limitations to the research that has been done. This section also sets out a future research agenda to further understand the convenience grocery market in GB and the revenue performance of this type of grocery store.

10.1 Comparing global model predictions

Table 10.1 compares the performance of GIS buffer and overlay modelling, regression modelling and spatial interaction modelling for forecasting Sainsbury's convenience grocery stores regardless of location type. The regression methodology tends to be most successful in terms of both mean predictive accuracy and the limiting of poor store revenue predictions (<60% accuracy).

Table 10.1 Comparing three methods of forecasting convenience store sales

Accuracy	GIS Buffer and Overlay	Regression	Applied SIM
<60%	61.7	8.7	44
>60%	38.3	91.3	56
>80%	25.5	28.9	44
>90%	14.9	16.8	26
Mean Prediction	48.2%	77.5%	71.3%

The regression model reported in chapter 9 of this thesis had the highest mean accuracy of prediction at 77.5%. This is substantially higher than the average accuracy for the GIS buffer and overlay approach (48.2%) and higher than the 71.3% mean store revenue prediction accuracy achieved by the SIM reported in chapter 8 of this thesis. As identified in this thesis, an integral part of location analysis is limiting poor predictions. The regression model in this thesis is by far the best of the three methods based on this measure. Just 8.7% of predictions fell below the 60% level of accuracy in the regression model presented in chapter 7, whereas 61.7% of predictions in the GIS buffer and overlay model and 44% of store predictions in the spatial interaction model failed to achieve this level of accuracy.

In terms of good (>80%) and very good (>90%) accuracy predictions, the spatial interaction model is, however, the best performing model. 44% of predictions achieved an 80% level of accuracy using this methodology, whereas around one quarter of forecasts were at this level of accuracy in the other two methodologies. Moreover, the highest proportion of very good (>90% accuracy) predictions was achieved by the SIM with 26% of stores being forecast to this degree of accuracy. Conversely, around 10% less stores are predicted this well in the GIS buffer and overlay model (14.9%) reported in chapter 7 and the regression model (16.8%) reported in chapter 9 of this thesis.

10.2 Comparing predictions by location type

10.2.1 Central Urban Cluster

The distinctive features of the central urban cluster are that daytime population, rail footfall and retail activity in the neighbourhood are very high among stores in this cluster, significantly larger than average. When looking at the geography of stores in this cluster, they are located in large urban centres across the region. These include the cities of Leeds, Sheffield, Bradford, York and Hull, along with stores in the centre of a number of the large towns in the region, including Huddersfield, Doncaster and Harrogate. When compared to the average among all retailers, we can see that a greater than average proportion of convenience stores operated by Tesco and Sainsbury's are found in the central urban cluster. Over 15% of stores operated by the two largest retailers fall into the central urban cluster. In comparison to just 2% of symbol group stores and 1% of Co-operative convenience stores. Table 10.2 identifies the performance of each of the three modelling types discussed in chapters 7, 8 and 9 when predicting store revenue in central urban store locations.

Table 10.2 Model performance in the central urban cluster

Central Urban Cluster (Cluster A)			
Accuracy	GIS Buffer and Overlay	Regression	Applied SIM
<60%	84.6	23.1	60
>60%	15.4	76.9	40
>80%	7.7	38.5	27
>90%	7.7	30.8	13
Mean Prediction	24.1	71.2	51.4

The regression model is by far the best performing model for stores in the central urban location type. The mean forecasting accuracy for the regression model in this cluster was 71.2%, substantially higher than the 51.4% achieved by the SIM and just 24.1% achieved by the GIS buffer and overlay model. The low mean accuracy achieved by both the GIS buffer and overlay model and the SIM were as a result of a large proportion of very poor forecasts in both models.

In central locations, a greater supply of potential food outlets to choose from may increase the likelihood of any person working in that location spending money at non-grocery stores. This may contribute to the under prediction of store revenue in this type of location by the SIM and GIS buffer and overlay models. This may have been due to failing to capture all of the available demand (including workers and visitors). The greater performance of the regression model in this location may be accounted for by

the fact that it has a lesser reliance on defining available demand. Rather than directly allocating grocery expenditure from a demand zone to a store, it finds an association between characteristics of stores catchment areas and sales which has led to improved predictions in this case.

The regression model is the only model reported in this thesis that would be considered for rolling out further to similar store locations in which Sainsbury's operate (or are looking to operate) convenience grocery stores. The model surpassed the 60% accurate 60% of the time threshold discussed with Sainsbury's and also produced a substantial proportion of very good estimates at a greater than 90% level of accuracy.

10.2.2 Large Population Suburban Cluster

On average, the stores in this cluster are distinguished by having catchment areas with significantly above average residential populations and an above average daytime population and other retail outlets in the 1km catchment of these stores. On average, a smaller proportion of residents employed in social class 1 live in the catchments of these stores. These stores are located in the suburbs of the major towns and cities inhabited by stores in the central urban location type. The divide is less stark when looking at stores by firm in the large population suburban cluster: over 20% of stores operated by the major retailers fall into this category compared with around 15% of symbol group stores and just over 10% of Co-operative group stores. Table 10.3 identifies the performance of each of the three modelling types discussed in chapters 7, 8 and 9 when predicting store revenue in large population suburban store locations.

Table 10.3 Model performance in the large population suburban cluster

Large Population Suburban Cluster (Cluster B)			
Accuracy	GIS Buffer and Overlay	Regression	Applied SIM
<60%	63.6	27.3	56
>60%	36.4	72.7	44
>80%	31.8	40.9	22
>90%	13.6	9.1	22
Mean Prediction	45.0	72.4	61.5

The regression model is the best performing model for stores in the large population suburban location type. The mean forecasting accuracy for the regression model in this cluster was 72.4%, substantially higher than the very poor performance of the GIS buffer and overlay model (45.0%) and over 10% higher than the 61.5% mean accuracy of the SIM. Once again, the GIS buffer and overlay model struggled in this location, most likely due to failing to capture the complexities associated with this type of store

location in which there is significant competition for customers and varied consumer types.

Despite lagging behind the regression model in mean accuracy, the SIM generated the greatest proportion of very good forecasts. However, the model was let down by a number of very poor forecasts. This type of location has a wide mix of retailers and spatial battles for patronage are complex in these locations. The SIM may be struggling in some of these locations due to the presence of large supermarkets which may attract trade away from smaller convenience stores in the model. However, the quick accessibility of convenience stores may be clawing back some of this revenue in reality, thus making some of the revenue forecasts inaccurate. Both the SIM and regression models are worthy of further consideration for wider use in this location type in other parts of GB.

10.2.3 Smaller Population Suburban Cluster

These stores are located on the outskirts and suburbs of the major towns and cities inhabited by stores in the central urban cluster. Furthermore, some of these stores are located in the centre and large suburbs of other important, but smaller towns in the area such as Wakefield, Barnsley, Rotherham, Scunthorpe and Grimsby. Stores in this cluster have a smaller residential population and less other retail stores in the area than in the large population suburban cluster. This cluster primarily contains residential areas, with stores having an above average residential population but a below average daytime population, less other retail stores, less social class 1 persons and less rail passenger volumes. This locational type is common among the major convenience grocery players: over 30% of stores operated by the major retailers fall into this category compared with around 20% of symbol group stores and just over 20% of Co-operatives. Table 10.4 identifies the performance of each of the three modelling types discussed in chapters 7, 8 and 9 when predicting store revenue in smaller population suburban store locations.

Table 10.4 Model performance in the smaller population suburban cluster

Smaller Population Suburban Cluster (Cluster C)			
Accuracy	GIS Buffer and Overlay	Regression	Applied SIM
<60%	72.4	17.2	33
>60%	27.6	82.8	67
>80%	17.2	55.2	50
>90%	10.3	31.0	17
Mean Prediction	49.3	76.4	70.9

The regression model reported in chapter 9 of this thesis had the highest mean accuracy of prediction for stores in the smaller population suburban cluster at 77.5%. This is substantially higher than the average accuracy for the GIS buffer and overlay approach (49.4%) and higher than the 70.9% mean store revenue prediction accuracy achieved by the SIM reported in chapter 8 of this thesis. The SIM has over double the proportion of very poor model estimates when compared to the regression model in this location which would worry location planning teams.

The predominantly residential demand characterising this type of location lends itself to forecasting this type of area using a SIM methodology. The less choice of available retailers in this type of area makes it more predictable. However, the regression model out performs the SIM in this location. This is probably due to the regression model explicitly dealing with some adjacency factors such as other available retail stores which may attract extra customers to some store locations. Both the regression model and SIM are worthy of further consideration in terms of rolling out for other similar Sainsbury's store locations across GB.

10.2.4 Satellite Store Locations

This cluster is distinguished by having, on average, the second largest average rail footfall among clusters, significantly higher than all clusters other than the central urban cluster A. On average, stores in this cluster have a catchment area with a slightly below average residential and daytime population, an above average level of retail stores and proportion of residents in social class 1 occupations. Thus, they are prevalent close to railway stations and in the suburbs of smaller towns. Geographically, these stores are found in small market towns such as Market Weighton in East Riding of Yorkshire, Northallerton in Hambleton district, Knaresborough in Harrogate District and Malton in Ryedale district. These stores are also found in towns close to the region's larger cities in places such as Castleford and Pontefract in Wakefield district, Garforth and Guiseley in Leeds district and Bingley and Keighley in Bradford district. Sainsbury's stands out as the retailer with the largest proportion of stores falling in this category. Table 10.5 identifies the performance of each of the three modelling types discussed in chapters 7, 8 and 9 when predicting store revenue in satellite store locations.

Table 10.5 Model performance in satellite store locations cluster

Satellite Locations Cluster (Cluster D)			
Accuracy	GIS Buffer and Overlay	Regression	Applied SIM
<60%	50	0	67
>60%	50	100	33
>80%	33.3	66.7	0
>90%	33.3	33.3	0
Mean Prediction	65.5	84.9	53.2

The satellite store location type is unique because it is the only location in which the more simple GIS buffer and overlay model outperformed either of the more complex modelling methodologies. The GIS buffer and overlay model outperformed the SIM in terms of the mean accuracy of revenue predictions with a mean of 65.5% in comparison with the 53.2% accuracy achieved by the SIM. However, the regression model is the best suited to forecasting sales in this location type with an impressive mean accuracy of 84.9%.

The regression model explicitly deals with rail passengers as a variable in the model. As these locations are characterised as market towns with rail links to larger towns and centres, this is likely the explanation for the regression model substantially outperforming the other two methods. The SIM is the poorest of the three methods for forecasting convenience grocery sales in this type of location. This is likely due to the fact that larger supermarkets often appear on the outskirts of the market towns in which these stores are located and supermarkets have a large draw within the SIM reported in chapter 8 of this thesis. Moreover, the SIM does not explicitly deal with visitor/rail driven demand which likely plays an important role in generating revenue in this type of location.

The GIS buffer and overlay shows promise in this location and efforts to improve on forecasting using this method in this location are planned. However, the regression model performs very well in this cluster and will be tested more extensively in similar store locations.

10.2.5 Affluent Outskirts Store Locations

This location type has a significantly below average population, a below average daytime population and lower number of neighbouring stores. It is distinguished by the significantly above average proportion of the population employed in social class 1 occupations, 28.5% of the population on average. The stores in the outskirts affluent cluster tend to be located in the affluent outskirts of the larger towns and cities and in

relatively affluent larger villages and small towns across the Yorkshire and the Humber region; these outskirts affluent cluster locations are strongly favoured by the Co-operative group. Additionally, symbol group stores are significantly more likely to be located in less affluent cluster locations than major retailer stores. However, major retailer stores are still more likely, although less comprehensively so, to be located in outskirts affluent locations over outskirts less affluent locations. Table 10.6 identifies the performance of each of the three modelling types discussed in chapters 7, 8 and 9 when predicting store revenue in outskirts affluent locations.

Table 10.6 Model performance in outskirts affluent store locations

Affluent Outskirts Cluster (Cluster E)			
Accuracy	GIS Buffer and Overlay	Regression	Applied SIM
<60%	33.3	0	29
>60%	66.7	100	71
>80%	50	83.3	57
>90%	33.3	66.7	29
Mean Prediction	64.5	90.7	74.2

The regression model is by far the best performing model for stores in affluent outskirts location type in which it is the best performing for any location type for any model with a mean forecasting accuracy of 90.7%. All three model's performed relatively well in this location type with a mean accuracy of 64.5% for the GIS buffer and overlay model and 74.2% for the SIM. Once again, the regression model is very effective at limiting very poor predictions with no stores failing to achieve a 60% accurate prediction. Moreover, the SIM also performs fairly well in this measure achieving 71% of predictions at a greater than 60% accuracy, exceeding the minimum requirements highlighted in chapter 7 of this thesis.

The regression model reported in chapter 9 is very effective at producing good and very good predictions with 83.3% of forecasts reaching an 80% level of accuracy and two thirds of forecasts reaching a 90% level of accuracy. Both the GIS buffer and overlay model and SIM are also effective in generating accurate forecasts. 50% of forecast using the GIS buffer and overlay model and 57% using the SIM reach an 80% level of accuracy with 33.3% and 29% respectively reaching a 90% level of accuracy. All three models are worth of further consideration in this store location with the regression model looking particularly effective in this store location.

The predominantly residential nature of this type of store catchment means that the demand side of the model is likely to be very accurate in this type of location. This has

benefitted both the GIS buffer and overlay model and SIM in this store location and both of these models are worthy of further consideration. Once again, the regression model is robust in this store location type and will be taken forward in terms of further analysis in this type of location.

10.2.6 Less Affluent Outskirts Store Locations

This location type has a significantly below average population, a below average daytime population and lower number of neighbouring stores. It is distinguished from the affluent outskirts cluster in having a much smaller (and below average) proportion of the population employed in social class 1 occupations. Stores in the outskirts less affluent cluster tend to be located in the less affluent outskirts of the larger towns and cities and in relatively less affluent larger villages and small towns across the region including much of the corridor running from South Leeds down the north and south east of Barnsley through into the northern outskirts of Rotherham (areas which struggled with major job losses in mining and manufacturing in the 1980s and 1990s). Additionally, these stores can be found in the outskirts of Doncaster and Sheffield. When looking at membership of stores in the outskirts less affluent cluster, symbol groups are the most likely of the groups of retailers to locate in the less affluent locations in Yorkshire and the Humber, although symbol group retailers still operate a considerably fewer number of stores in this cluster when compared to the more affluent outskirts locations. Table 10.7 identifies the performance of each of the three modelling types discussed in chapters 7, 8 and 9 when predicting store revenue in outskirts less affluent locations.

Table 10.7 Model performance in outskirts less affluent store locations cluster

Less Affluent Outskirts Locations (Cluster F)			
Accuracy	GIS Buffer and Overlay	Regression	Applied SIM
<60%	45.5	9.1	20
>60%	54.4	90.9	80
>80%	27.3	45.5	50
>90%	9.1	36.4	50
Mean Prediction	51.9	78.8	80.5

This location type is unique in that it is the only location type in which the SIM outperforms the regression model in terms of mean accuracy of predictions. Both models have a good mean accuracy with 78.8% for the regression model and 80.5% mean accuracy for the SIM. The GIS buffer and overlay model performs poorly in this cluster and lags substantially behind with a mean forecast accuracy of 51.9%. The

regression model is better than the SIM at limiting very poor forecasts with 9.1% of estimates failing to reach a 60% level of accuracy in comparison to 20% of forecasts using the SIM. The SIM outperforms the regression model quite substantially in very good forecasts in this location type with 50% of forecasts reaching a 90% level of accuracy using the regression model in comparison to around 35% for the regression model. Once again the GIS buffer and overlay lags behind considerably with around a quarter of forecasts at an 80% or higher level of accuracy and less than 10% of forecasts at a 90% level of accuracy.

The poor performance of the GIS buffer and overlay model in this location in comparison with the more affluent outskirts store location type is probably a result of the model not taking into account retailer preference among different groups of consumers. This is likely leading to over prediction of Sainsbury's convenience stores in this location type. The disaggregate SIM reported in chapter 8 of this thesis is the model that most comprehensively deals with retailer preference among different groups of consumers and subsequently produces the most accurate predictions in this location type. The good performance of both the regression model and spatial interaction model make them worthy of further consideration for rolling out in similar locations across GB.

10.2.7 Rural Store Locations

This cluster produces a distinct group of stores in outlying rural areas. The rural urban classification defines areas as rural if they fall outside of settlements with more than a 10,000 resident population, the catchment area of all stores in the rural cluster meet these criteria. On average, 27% of residents in the catchment surrounding stores in the rural cluster are employed in social class 1 occupations, 5% higher than the average among the catchment areas of all convenience stores in the study. They are prevalent in much of the more rural districts of Yorkshire and the Humber such as Richmondshire, Hambleton, Ryedale, Craven, East Riding of Yorkshire and North Lincolnshire and major retailer stores are considerable less likely to be located here. Fewer than 3% of Sainsbury's Local and Tesco Express stores are located in the rural cluster in comparison to over 17% of stores operated by symbol group retailers and the Co-op. Table 10.8 identifies the performance of each of the three modelling types discussed in chapters 7, 8 and 9 when predicting store revenue in the rural locations.

Table 10.8 Model performance in rural store locations

Rural Locations (Cluster G)			
Accuracy	GIS Buffer and Overlay	Regression	Applied SIM
<60%	0	0	0
>60%	100	100	100
>80%	0	100	100
>90%	0	0	0
Mean Prediction	62.1	87.2	81.0

The one store in the rural store location is best predicted by the regression model which produced a forecast at 87.2% accuracy, a very good level of accuracy for grocery stores in the convenience market. However, the SIM also generates a good forecast for the store in this cluster, estimating the store at an 81.0% accuracy. Following the same trend as the majority of location types, the GIS buffer and overlay model is the worst performing model for this cluster, estimating the store at a 62.1% level of accuracy.

All three models perform well in this cluster in comparison to their forecasts in other location types. This is likely a result of the easily identifiable supply and demand in this type of areas with limited grocery opportunities (in which retailers often enjoy spatial monopolies) and highly predictable residential demand. The strong perform of both the SIM and regression methodology in this location type, particularly in light of minimising very poor predictions, means that both models should be more widely used for this type of location outside the study area.

10.3 Conclusions, limitations and further research

10.4.1 Aim 1

To review the existing academic and industry literature on the convenience grocery market in GB, the growth of the major retailers into the convenience grocery market, the growing demand for convenience groceries and attempts at forecasting revenue of convenience grocery stores.

Chapter 2 reviewed the academic literature on changes seen in the grocery market in GB from the 1960s to 2016 which eventually resulted in the growth of branded grocery stores in the convenience grocery market. The major structural changes began in the early 1960s when resale price maintenance on groceries was abolished. This benefitted the major grocery multiples, they grew their operations rapidly and a 'Golden Age' for the major retailers was established. Changes in the grocery market

(particularly in the form of local planning legislation) in the 1990s made growth in market share through traditional large supermarket formats more difficult. Tesco and Sainsbury's, in response to these difficulties, led the way in extensive growth in the convenience grocery market. The Competition Commission's two market ruling distinguishing between primary shopping and secondary shopping gave a regulatory green light for them to further invest in the convenience market.

Both Tesco and Sainsbury's continued to invest in convenience grocery retailing and more recently Waitrose, Morrisons and Marks and Spencer have entered the market. Morrison's convenience operations present an interesting case study of the grocery market in GB in the 2010s. The retailer opened its first convenience store in West Yorkshire in 2011 and expanded their network to a total of over 140 stores. However, they experienced difficulties in the market and sold their network to a venture capital firm in 2015. Curiously, Morrisons re-entered the convenience market just three months later with the opening of a 1200 sq. ft. Morrisons Daily convenience store at a Motor Fuel Group petrol station in Crewe and plans to open further stores (Ruddick, 2015). This example highlights the importance retailers are placing on having convenience operations in the grocery market in GB.

A number of drivers of demand for convenience grocery retailing were identified in chapter 2. These included demand generated through consumers becoming more time conscious plus population change, demographic change, shifts in living arrangements, shifts in working patterns and new demand generated by the retailers themselves. The review of the literature found that there were many types of consumers that make natural customers for convenience grocery retailing and they are able to support the various location types in which branded convenience grocery stores are found. In terms of impact, this is the first appearance in the academic literature of a comprehensive review of the literature on the demand for convenience grocery retailing in GB.

The final part of the first aim of this thesis was to review the existing academic and industry literature on attempts at forecasting revenue of both convenience grocery stores and grocery stores more generally. Chapter 3 discussed the use of site location models in the retail industry, charting the emergence of the suite of methods now used by the major grocery retailers in GB. The review of the potential application of methods to the convenience market may provide additional insight into why certain models that retailers apply are theoretically appropriate or not appropriate for the convenience grocery market. GIS buffer and overlay, regression modelling and spatial interaction

modelling were identified as methods that were well tried and tested in other markets and plausible in terms of application to the convenience grocery market. It was hypothesised that these methods could be used to forecast sales for different locations and anticipated that different locations may have a different optimum methodology for accurately predicting store revenue.

10.4.2 Aim 2

To quantify the extent to which major retailer have committed to the convenience grocery market and assess the geographical extent to which they play a role in convenience grocery retailing in Great Britain.

The second aim of this thesis was to quantify the extent to which major retailers have committed to the convenience grocery market and assess the geographical extent to which they play a role in convenience grocery retailing in Great Britain. Tesco and Sainsbury's were pioneers of the diversification into small store grocery retailing and showed that success in an area previously the reserve of smaller retailers was possible. The analysis in chapter 5 found a marked shift in the portfolio of stores operated by major grocery retailers between 2003 and 2012. Amongst the four largest grocery retailers in GB, convenience stores, as a proportion of total stores, increased by 40.1%, from 8.2% to 48.3% of total stores between 2003 and 2012. This was the first of two trends identified in the largest grocery retailers store network changes between 2003 and 2012. The second major trend is the rise of the strategy of 'space sweating', the process of extending existing supermarkets to become hypermarkets. ASDA have led the way in pursuing this strategy of growth and continue to operate a number of the largest grocery hypermarkets. (Wood and McCarthy, 2013).

On a regional level, the study found that branded convenience floorspace per capita in 2012 was significantly lower in Scotland and Wales in comparison to most regions of England. Chapter 5 assessed the geographical extent by which major grocery retailers, more traditionally associated with supermarket grocery retailing now play a role in convenience grocery retailing across GB. This was conducted by quantifying the market share of each of the notable convenience retailers at the postal area level in GB in 2012, comparing this market share to each retailer's total market share across all store formats in GB. The limitation of this analysis is that market share by floorspace was used as a proxy for overall market share. However, it is difficult for actors in

academia to attract sales data for one retailer, as is the case in this thesis, let alone for multiple actors in the market.

The research found that the convenience operations of Tesco and Sainsbury's, the two major retailers to most significantly commit to convenience grocery retailing, are more geographically concentrated than their total grocery offer, suggesting that they have specifically targeted certain areas to concentrate their convenience efforts. The research found that the major grocery retailers have very little convenience operations in many parts of GB and are therefore only competing heavily with smaller and independent grocery retailers in limited geographical localities. The Co-operative and the symbol group retailers remain dominant in much of rural England, Scotland and Wales. Empirical evidence in disparities in the extent to which major retailers now play a role in the convenience market in different areas of GB provides evidence that there is not a simple narrative of major retailers encroaching into areas more traditionally the reserve of smaller retailers, but a complex picture of varying provision in different parts of GB.

Sainsbury's commitment to convenience grocery retailing is evident in the company's internal structure. The location department of the retailer is now split into two sections, supermarket and convenience, a situation mirrored at Tesco. This identifies the importance retailers are placing on making optimum location decisions in the race for small-format success. This has been identified as causing pressure on smaller retailers operating at the neighbourhood level and continues to do so as major retailers have the luxury of location planning teams, a significant advantage over both smaller retailers and independent stores (Wood and Browne, 2007).

Morrisons had begun to expand into the convenience grocery market by 2012 and had grown their portfolio of convenience stores to a total of 140 stores by 2015. However, the retailer struggled with this format and sold the chain to an investment firm in September 2015. (Armstrong, 2015). Other retailers such as Waitrose and Marks and Spencer have continued to invest in convenience grocery retailing and as noted above Morrisons re-entered the convenience market in late 2015 with the opening of a 1200 sq. ft. Morrisons Daily convenience store at a Motor Fuel Group petrol station in Crewe (Ruddick, 2015). The persistent presence of major grocery retailers in the convenience market may continue to increase the pressure experienced by small and independent retailers.

In terms of retail planning policy, it is feasible that the grocery market will once again be referred to the Competition Commission in light of the increasing number of convenience stores operated by the large grocery retailers. As these retailers attempt to increase market share in convenience grocery retailing, a review of the two-market ruling distinguishing between supermarket and convenience retailing may occur. The analysis presented in this thesis could form part of any investigation as it reviews the presence of major retailers in the convenience grocery market by geographical coverage. In the future, if the two-market ruling were to be revised, major retailer's convenience stores may be considered alongside their larger store formats. This may lead to retailers being forced to sell off part of their portfolio of convenience stores, further altering the dynamics of the convenience grocery market in GB.

Reporting in 2016, this thesis identified the changes seen between 2003 and 2012. An obvious limitation is that the datasets provided by Sainsbury's and GMAP Ltd. are becoming increasingly dated. Retailing is a constantly evolving industry. This is an issue in academia as the availability of retail industry datasets is often limited. However, the analysis presented in this thesis and in associated papers will have contributed to the first appearance of empirical retail research in this area within the academic literature. The dissemination of analysis in this thesis through academic papers has already begun. Part of the review of the growing demand for convenience grocery retailing and analysis on the growth of convenience grocery retailing among major grocery retailers between 2003 and 2012 appear in the following paper:

> Hood, N., Clarke, G.P. and Clarke, M (2015) Segmenting the growing UK convenience store market for retail location planning. *The International Review of Retail, Distribution and Consumer Research*. **26**(2), pp. 113-136.

The first comprehensive exploration of market shares of the prominent retailers in the branded convenience grocery market across GB in the academic literature appears in the above paper. Future work by the author will look to keep track of the changes in store portfolios of both the four retailers focused on in chapter 5 of this thesis and other retailers with diverse grocery store networks through the acquisition of up-to-date datasets at regular intervals.

10.4.3 Aim 3

To develop and test a series of predictive models for forecasting convenience grocery store revenue in the varying location types in which this type of grocery store is found.

From the outset of this research it was hypothesised that different locations in which convenience grocery stores are found in GB may, in theory, require a different optimal methodology for forecasting revenue accurately. Chapter 6 of this thesis reported on a k-means cluster analysis which segmented the convenience grocery market in Yorkshire and the Humber into 7 statistically distinct location types. The classification of convenience stores reported in chapter 6 of this is important for future store location analysis. The 7 distinct location types in which convenience stores are found in Yorkshire and the Humber were named as follows:

1. Central Urban Cluster
2. Large Population Suburban Cluster
3. Smaller Population Suburban Cluster
4. Satellite Cluster
5. Outer Suburban Affluent Cluster
6. Outer Suburban Less Affluent Cluster
7. Rural Cluster

The analysis found that certain locations were more favoured by certain types of retailers than others. Sainsbury's and Tesco prefer central locations and large population suburban locations. Conversely, the Co-operative group convenience retail stores are more likely to be found in outskirts locations, particularly in outskirts affluent locations and rural locations. The major symbol group retailers identified in the market share analysis in chapter 5 are more likely than major retailers to favour smaller population suburban locations and are less likely to be placed in large satellite towns connected to large urban centres by rail links.

The paper discussed in the previous section was built around this segmentation of the convenience market. It provides a nuanced understanding of the types of location in which branded convenience grocery stores are found, alongside a greater understanding of the types of convenience locations favoured by the largest retailers (who have recently entered the market) versus the locations favoured by retailers more commonly known for small format grocery retailing.

Next, it was hypothesised that if we can test alternative methodologies for producing sales forecasts by the different locations reported in the segmentation in chapter 6, we can set out a best practice for predicting revenue in different convenience store locations. Convenience grocery store revenue in the 7 location types was forecast using a GIS buffer and overlay procedure, a spatial interaction model framework and a regression model. In the literature review in chapter 3, these three methods were highlighted as tried and tested in other markets and plausible in terms of an application to the branded convenience grocery market. The three models were compared earlier in this chapter and found to have varying success in accurately estimating store sales.

The GIS buffer and overlay model in chapter 7 was found to have little success in forecasting convenience grocery sales on the whole. This is interesting in light of the recommendation to use this simple method by commentators suggesting this is the best practice for convenience grocery stores. Moreover, it particularly struggled in central locations where customer behaviour is often more complex. The model had moderate successes in predicting sales in satellite, outskirts and rural locations, although the model still lacked a high degree of accuracy. Problematic issues unique to the GIS buffer and overlay model remain. For instance, it is highly likely that customers on the outskirts of the buffer will be less likely to travel to a store location than customers very close to the store. Further work beyond this thesis will focus on the areas in which the GIS buffer and overlay model had success and will attempt to resolve these problems in the form of banded likelihoods of patronage, possible making use of nectar card data in calibration. This will hopefully improve the model further in these areas and lead to more robust estimates.

The spatial interaction model presented in chapter 8 showed that an improvement in predictions can be made when using a method more traditionally associated with supermarket forecasting (and openly criticised in its potential in the convenience market by some studies in the literature). The model overcomes the catchment area problem identified in the GIS buffer and overlay analysis by allowing store catchment areas to form based on customer travel patterns and availability of supply. The SIM was very good at producing good and very good individual store estimates in a number of locations but was let down by a number of very poor (<60% accurate) predictions in a number of location types. These lower estimates often meant that, on average, the model performed less effectively than the regression model across a number of location types despite producing a greater proportion of very accurate sales forecasts.

Methodologically, in calibrating the model specifically to convenience grocery stores, it is the first comprehensive example of a disaggregate SIM to the convenience grocery market in the academic literature. Moreover, in creating a dual demand layer accounting for residential and workplace spending on groceries (the first known example of this in the academic literature) the model is an advancement on existing SIMs in terms of their potential application in the convenience grocery market. This new model allowed residential consumers to travel more freely than work based consumers and improved revenue estimates, particularly in central locations. Furthermore, the adaption of attractiveness values depending on the type of convenience store location is an advancement on previous disaggregate SIMs when accounting for sales to convenience grocery stores.

The regression model presented in chapter 8 was the most effective and versatile model for forecasting convenience grocery sales. This model produced the highest average estimation accuracy of the three models and was the best model in the majority of clusters. The major strength of this model was in its limiting of very poor predictions, with just 8.7% of total predictions failing to reach this level of accuracy. It was the only model of the three to exceed the minimum level of 60% accuracy 60% of the time in all store locations, particularly outperforming the other two methods in central store locations. Methodologically, the assessment of the varying relationships between different variables and size of catchment areas went some way to overcoming the criticisms of defining catchment areas in this type of model. However, regression is problematic in its inability to replicate know flows between individual demand zones and individual stores meaning that it would struggle to dynamically respond to changes in the grocery market. In terms of future improvements to the regression model presented in this SIM, a focus has been placed on quantifying the relationships of counts of given variables with store sales. There is potential for improvement through identifying the proportion of given variables (alongside raw counts) as additional predictors in the models.

A major improvement scheduled for the regression model is the development of individual regression models tailored to modelling the store locations the general model has had successes, a geographically weighted regression model for forecasting convenience grocery store sales. It is likely that the relationship between different variables and store sales vary by location type and a series of models will hopefully lead to greater forecasting power of the regression model in this thesis. This will require additionally liaising with Sainsbury's to gain access to more data. However, in light of

the improvements to forecasting already made by this research, this should hopefully be possible.

In general the disaggregation of the market and testing of different methodologies has advanced the forecasting of convenience stores, particularly in empirically testing different models in the varied locations in which convenience grocery stores are found in GB. It is the most comprehensive empirical testing of methods to forecast convenience grocery stores in the academic literature to date and the number of stores involved in the calibration and testing process is unprecedented in the academic literature. Rather than discarding methods traditionally used in supermarket sales forecasting, this research has tested the application of varying methodologies in different locations and evaluated the extent to which they could be utilised by retailers both in general, and in forecasting stores in specific locations. In doing so it has found that these methods have utility in at least some of the locations in which convenience grocery stores are found.

In terms of the application of the framework set out in this thesis to the major grocery retailers, it has made a start in providing a plausible kitbag of techniques which can be applied in different circumstances. It is empirically true that advancements have been made in forecasting accuracy in a number of locations in which convenience stores are found in GB. The retailers may be more receptive to any improvements made in spatial interaction modelling due to the associated granularity of forecasts which allow for the replication of real world flows between individual residential zones and individual stores. On the back of the work in this thesis, Sainsbury's have already developed a new typology for their convenience stores based largely on the one presented in chapter 6 of this thesis. In doing so, they incorporated store information that the author of this thesis was not privy too in the segmentation process in this thesis. This is clear evidence of the research carried out in this project impacting the operations of retailers in a positive way.

In terms of a future research agenda to build on the overall modelling in this thesis, a number of additional pieces of information would present an opportunity for improving on revenue prediction. Improvements in available survey data on workplace spending habits would lead to more nuanced and disaggregated demand layers in the models which would improve the knowledge of available expenditure on groceries in the different locations in which convenience stores are found. This could be achieved in a number of ways. Firstly, Nectar (and other loyalty card schemes) could ask for

workplace registration as well as residential registration when consumers sign up. This could allow greater understanding of spending patterns by location and could be used to calibrate the workplace demand layer in a disaggregate SIM such as the one in chapter 8 of this thesis. This could also be achieved through the Living Costs and Food Survey (LCFS) disaggregating expenditure by workplace type. This could make use of the WPZ Classification, a classification of workplace zones similar to the one in existence for residential output areas.

A limitation of the modelling in this thesis is its ability to deal with visitor demand. This could possibly be improved with the incorporation of survey data or the disaggregation of the numbers of persons entering through rail stations or other areas known for high numbers of day and holiday visitors - such as large cities. A possible limitation of this work (which the future agenda set out in this section is unlikely to resolve) is the focus on methods that have a long history of application in grocery retailing and have been traditionally applied to supermarket revenue estimation. It could be possible that new and innovative methods may arise with the opportunities provided by large datasets like the Nectar card dataset used in this thesis. However, the reason for the focus on established methodologies was two-fold. First, these methods had been hypothesised as having deficiencies when applied to the convenience market, yet little empirical testing of these difficulties has appeared in the academic literature. Secondly, models that could be readily incorporated into the actions of major retailers were favoured over those that may take more time to become embedded in the suite of methods adopted by grocery retailers in GB. These retailers already have these methodologies so would have no need to invest heavily in both time and capital in adopting whole new methodologies.

10.4 Final remarks

As convenience stores have grown, so too has interest in site location research in finding techniques to best predict their success. The segmentation reported in chapter 6 of this thesis has provided a valuable framework for assessing different methods for forecasting convenience grocery store sales. Both the regression and spatial interaction models reported in this thesis have been clearly demonstrated as having utility in forecasting convenience grocery store sales. Future research will refine these models in the locations in which they have shown promise and hopefully improve forecasts to have a suite of reliable and robust methods of predicting convenience grocery store revenue

List of References

ACS (2012) The Local Shop Report 2012. Available at: https://www.acs.org.uk/?wpdmdl=852 (Accessed 10th November 2013)

ACS, FARM, WI and FOE (2004) Proposal for a market study for OFT consideration. November 2004. Available at: https://www.foe.co.uk/sites/default/files/downloads/proposal_for_a_market_stud.pdf

Aldenerfer, M. and R. Blashfield (1984) *Cluster Analysis*. Newbury Park, CA: Sage University Papers.

Alexander, A. and Pollard, J. (2000) Banks, grocers and the changing retailing of financial services in Britain. *Journal of Retailing and Consumer Services*. **7**(3), pp. 137-147.

All-Party Parliamentary Small Shops Group (2006) High Street Britain 2015.

Applebaum (1966) Methods for Determining Store Trade Areas, Market Penetration, and Potential Sales. *Journal of Marketing Research*. **3**(2), pp. 127-141.

Arentze, T.A., Oppewal, H. and Timmermans, H.J.P. (2005) A Multipurpose Shopping Trip Model to Assess Retail Agglomeration Effects. *Journal of Marketing Research*. **42**(1), pp. 109-115.

Armstrong, A. (2015) Morrisons agree £25m sale of convenience stores. Telgraph Online, 9th September 2015. Available at:(Accessed 20th September 2015)

Azimont, F. and Araujo, L. (2010) The making of a petrol station and the "on-the-move consumer": Classification devices and the shaping of markets. *Industrial Marketing Management*. **39**(6), pp. 1010-1018.

Barich, H. and Srinivasan, V. (1993) Prioritizing marketing image goals under resource constraints. *Sloan Management Review*. **35**, pp. 69-76.

Baron, S., Harris, K., Leaver, L. and Oldfield, B.M. (2001) Beyond convenience: the future for independent food and grocery retailers in the UK. *International Review of Retail, Distribution and Consumer Research*. **11**(4), pp. 395-414.

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.

Beaumont, J.R. (1991) GIS and market analysis. In: D. Maguire, M. Goodchild and D. Rhind (eds) Geographical information systems: principles and applications. Longman, London, pp. 139–151.

Bell, D.R. and Lattin, J.M. (1998) Grocery shopping behaviour and consumer response to retailer price format: Why large basket shoppers prefer EDLP. *Marketing Science*. **17**(1), pp. 352-369.

Bell, J. and Hallsworth, A. (2003) Retail change and the United Kingdom Co-operative Movement – new opportunity beckoning. *The International Review of Retail, Distribution and Consumer Research*. **13**(3), pp. 301-315.

Benoit, D. and Clarke, G.P. (1997) Assessing GIS for retail location planning. *Journal of Retailing and Consumer Services*. **4**(4), pp. 239–258.

Berry, T., Newing, A., Davies, D., and Branch, K. (2016). Using workplace population statistics to understand retail store performance. *The International Review of Retail, Distribution and Consumer Research*. **26**(4), pp. 1-21.

Birkin, M. and Clarke, G.P. (1991) Spatial Interaction in Geography. *Geography Review*. **4**(5), pp. 16-24.

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., Clarke, G.P and Clarke, M. (2002) *Retail Geography and Intelligent Network Planning.* Wiley, Chichester.

Birkin, M., Clarke, G.P, Clarke, M. and Culf, R (2004) Using spatial models to solve difficult retail location problems. *In:* J. STILLWELL and G.P. CLARKE, eds. *Applied GIS and Spatial Analysis.* Wiley, Chichester, pp.35-53.

Birkin, M., Clarke, G.P, Clarke, M. and Wilson, A. (1996) *Intelligent GIS – Location decisions and strategic planning.* Geoinformation International, Cambridge.

Birkin, M., Clarke, G.P, Clarke, M. and Wilson, A. (2014) The achievements and future potential of applied quantitative geography: a case study. *Geographica Polonica*. **87**(2), pp. 179-202.

Birkin, M., Clarke, G.P. and Clarke, M. (1999) GIS for business and service planning In: P.A. Longley. M.F. Goodchild, D.J. Marguire and D.W. Rhind, eds. *Geographical Information Systems, Principles, Techniques, Applications and Management*. Wiley, New York.

Birkin, M., Clarke, G.P. and Clarke, M. (2010a) Refining and Operationalizing Entropy-Maximising Models for Business Applications. *Geographical Analysis*. **42**(4), pp. 442-445.

Bone, P.F. (1991) Identifying Mature Segments. *Journal of Consumer Marketing*. **8(**4), pp. 19-31.

Bowlby, S., Breheny, M. and Foot, D. (1984) Store Location Problems and Methods 1: Is locating a viable store becoming more difficult. *International Journal of Retail & Distribution Management*. **12**(5), pp. 31-33.

Boyle, E. (1999) A study of the impact of environmental uncertainty on franchise systems: the case of petrol retailing in the UK. *Journal of Consumer Marketing*. **16**(20), pp. 181-196.

Brodley, S. (2013) *Location planning at Morrison's: Presentation delivered to students. School of Geography, University of Leeds: 12th March 2013,* Wm Morrison Supermarkets plc., Bradford.

Brondoni, S.M., Corniani, M. and Riboldazzi, S. (2013) Global Retailers, Market-Driven Management and Innovation. *Journal of Economic Behaviour*. **3**, pp. 27-40.

Brooks, B. (2014) Netto returns: What happened to its stores in the UK? The Grocer Online, 20th June 2014. Available at: http://www.thegrocer.co.uk/channels/discounters/netto-returns-what-happened-to-its-stores-in-the-uk/358795.article (Accessed 3rd September 2014)

Brown, S. (1991) Retail location: the post hierarchical challenge. *The International Review of Retail, Distribution and Consumer Research*. **1**(3), pp. 367-381.

Brown, S. (1994) Retail location and the micro-scale: Inventory and Prospect. *Services Industries Journal*. **14**(4), pp. 259-268.

Buckley, M., Cowan, C. and McCarthy, M. (2007) The convenience food market in Great Britain: Convenience food lifestyle (CFL) segments. *Appetite*. **49**(3), pp. 600-617.

Burgess, E. (1925). 'The growth of the city: an introduction to a research project'. In LeGates, R. and Stout, F. (eds.), The city reader, pages 156–163. Routledge, London.

Burns, J.A., McInerney, J. and Swinbank, A. (1983) *The food industry: economics and policies*. Heinemann, London.

Burt, S. and Sparks, L. (1994) Structural change in grocery retailing in Great Britain: a discount reorientation? *The International Review of Retail, Distribution and Consumer Research*. **4**(2), pp. 195-217.

Burt, S. and Sparks, L. (2003) Power and Competition in the UK Retail Grocery Market. *British Journal of Management*. **14**, pp. 237-254

Burt, S., Sparks, L. and Teller, C. (2010) Retailing in the United Kingdom - a synopsis. *European Retail Research.* **24**(1), pp. 173-193

Burt, S.L., Mellahi, K., Jackson, T.P. and Sparks, L. (2002) Retail internationalisation and retail failure: issues from the case of Marks and Spencer. *The International Review of Retail, Distribution and Consumer Research.* **12**(2), pp. 191-219.

Butler, S. and Neville, S. (2013) Tesco's empire: Expansion checked in UK and beyond. The Guardian Online, Thursday 18th April 2013. Available at: http://www.theguardian.com/business/2013/apr/18/tesco-empire-uk-and-beyond (Accessed 13th October 2013)

Cardoso, M.G.M.S., Themido, I.H. and Pires, F.M. (1999) Evaluating a Clustering Solution: An Application in the Tourism Market. *Intelligent Data Analysis.* **3**, pp. 491-510.

Champion, T (2014) People in cities: the numbers. Future of cities: working paper. Foresight, Government Office for Science. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/3218 14/14-802-people-in-cities-numbers.pdf   (Accessed 10th September 2015).

Cheshire, P., Hilber, C.A. and Kaplanis, I., 2011. Evaluating the effects of planning policies on the retail sector: Or do town centre first policies deliver the goods? LSE Working Paper.

Christie, S. (2015) Morrisons takes on discount rivals with 'bold' loyalty card. Is it any good? Telegraph Online. 26th October 2014 Available at: http://www.telegraph.co.uk/finance/personalfinance/money-saving-tips/11145688/Morrisons-takes-on-discount-rivals-with-bold-loyalty-card.-Is-it-any-good.html (Accessed 10th December 2014)

Clarke, G.P. (1998) Changing methods of location planning for retail companies. *GeoJournal.* **45**, pp. 289-298.

Clarke, I. (2003) The devil is in the detail: visualising analogical thought in retail location decision making. *Environment and Planning B: Planning and Design.* **30**, pp. 15-36.

Clarke, I., Hallsworth, A., Jackson, P., De Kervenoael, R., Perez del Agula, R. and Kirkup, K. (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., Horita, M. and Mackaness, W. (2000) The spatial knowledge of retail decision makers: capturing and interpreting group insight using a composite cognitive map. *International Review of Retail, Distribution and Consumer Research*. **10**(3), pp. 265-268.

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**, pp. 1896-1911.

Clarkson, R.M., Clarke-Hill, C.M. and Robinson, T. (1996) UK supermarket location assessment. *International Journal of Retail & Distribution Management*. **24**(6), pp. 22-33.

Coady, I (2014) *Workplace Zones: Creating a New Geography for 2011 Census.* Presentation Delivered at the Society for Location Analysis Seminar. All You Need to Know about Workplace Zones, Jones Lang LaSalle, London, September 9, 2014. Titchfield: Office for National Statistics.

Coates, D., Doherty, N., French, A. and Kirkup, M. (1995) Neural networks for store performance forecasting: an empirical comparison with regression techniques. *The International Review of Retail, Distribution and Consumer Research*. **5**(4), pp. 415-432.

Cohen, M. (1998) The problem of competition in the UK petrol market. *Strategic Change: Briefings in Entrepreneurial Finance*. **7**(4), pp. 203-212.

Colgate, M and Alexander, N (2002) Retailers and diversification: the financial service dimension. *Journal of Retailing and Consumer Services*. **9**, pp. 1-11.

competing destinations'. *Environment and Planning A*, **15**: pp. 15 – 36.

Competition Commission (2000) *Supermarkets: A report on the supply of groceries from multiple stores in the United Kingdom.* Competition Commission, London.

Competition Commission (2003) Merger References: Competition Commission Guidelines, London: HMSO.

Competition Commission. 2008. *The supply of groceries in the UK market investigation* [online]. [Accessed 20th July 2013]. Available from: http://www.competition-commission.org.uk.

Converse, P. (1949) New Laws of Retail Gravitation. *Journal of Marketing*, **14**(1), pp. 379-384.

Crossen, C. and Graham, E. (1996) Pressed for Time or Pressed for Money? Wall Street Journal, 9th March 1996.

Davies, R. (1977) Store location and store assessment research: the integration of some new and traditional techniques. Transactions, Institute of British Geographers, **2**(2), pp. 141–157.

Dawson, A. (1980) *Retail Geography*, Wiley and Sons, New York.

Dawson, J (1995) Retail trends in Scotland: a review. *International Journal of Retail & Distribution Management*. **23**(10), pp. 4-20.

Dawson, J. (2004) Retail change in Britain during 30 years: The strategic use of economies of scale and scope. *Research Papers in Retailing*. **4**(2), pp. 1-50.

Day, G.S. and Heeler, R.M. (1971) Using Cluster Analysis to Improve Marketing Experiments. *Journal of Marketing Research*. **8**, pp. 340-347.

De Kervenoael, R., Hallsworth, A. and Clarke, I. (2006) Macro-level change and micro level effects: A twenty-year perspective on changing grocery shopping behaviour in Britain. *Journal of Retailing and Consumer Services*. **13**, pp. 381-392.

Delaney-Klinger, K., Boyer, K.K. and Frohlich, M. (2003) The return of online grocery shopping: a comparative analysis of Webvan and Tesco's operational methods. The TQM Magazine. **15**(3), pp. 187-96.

Dellaert B., Arentze, T.A., Bierlaire, M., Borgers, A. and Timmermans, H.J.P. (1998) Investigating Consumers' Tendency to Combine Multiple Shopping Purposes and Destinations. *Journal of Marketing Research*. **35**, pp. 177-188.

Dennett, A. (2010) *Understanding Internal Migration in Britain at the Start of the 21st Century*. Doctor of Philosophy Thesis, University of Leeds.

Department of Environment (1988) *Planning Policy Guidance 6: Major Retail Development*. HMSO: London.

Department of Environment (1993) *Planning Policy Guidance 6: Town Centres and Retail Development*. HMSO: London.

Department of Environment (1996) *Planning Policy Guidance 6: Town Centres and Retail Developments.* HMSO: London.

Duke, R. (1989). 'A structural analysis of the UK grocery retail market'. *Working Paper Series,* 90/04, School of Business & Economics, University of Leeds.

Dunteman, G.H. (1989) *Principal components analysis* (No. 69). Sage.

Elliott, C. (1991) Store planning with GIS. In: J.Cadeau-Hudson and D.I. Heywood (eds) Geographic Information 1991. Taylor and Francis, London.

Ellis-Chadwick, F., Doherty, N.F. and Anastasakis, L. (2007) E-strategy in the UK retail grocery sector: a resource-based analysis. *Managing Service Quality*. **17**(6), pp. 702-727.

Ellwood, L. (1954) Estimating Potential Volume of Proposed Shopping Centres. *The Appraisal Journal*, October 1954.

Elms, J., Canning, C., Kervenoael, R., Whysall, P. and Hallsworth, A. (2010) 30 years of retail change: where (and how) do you shop. *International Journal of Retail & Distribution Management,* **38**(11/12), pp.817-827.

Environment, Financial Times Business Limited, London.

Everitt, B., Landau, S. and Leese, M. (2001). Cluster Analysis. Arnold, London, 4th edition.

Eyre, H. (1999) Measuring the Performance of Spatial Interaction Models in Practice. Doctor of Philosophy Thesis, University of Leeds.

Faraway, J. J. (2002) *Practical Regression and ANOVA using R.* Available at: http://www.cran.r-project.org/doc/contrib/Faraway-PRA.pdf

Fernie, J. (1997) Retail Change and Retail Logistics in the United Kingdom: Past Trends and Future Prospects. *The Service Industries Journal.* **17**(3), pp. 383-396.

Fernie, J., Sparks, L. and Mckinnon, A. (2010) Retail logistics in the UK: past, present and future. *International Journal of Retail & Distribution Management.* **38** (11), pp. 894-911.

Field, A. (2009). *Discovering statistics using SPSS*, 3rd Edition. Sage Publications.

Field, C. (1997) *The future of the new store: new format and channels for a changing retail*

Finch, J. (2008) Supermarkets: Somerfield takeover pushes Co-op into big league. *Guardian Online*, 17th July 2008. Available at: http://www.guardian.co.uk/business/2008/jul/17/mergersandacquisitions.supermarkets (Accessed 10th April 2013)

Finch, J. and Wood, Z. (2010) Asda buys Netto for £778m. Guardian Online, 27th May 2010. Available at: http://www.guardian.co.uk/business/2010/may/27/asda-buys-netto (Accessed 1st March 2013)

Foot, D. (1981). *Operational Urban Models: An introduction.* Methuen, London and New York.

Forecourt Trader (2008) As the M&S/BP partnership reaches a milestone, a dealer franchise trial is about to be launched. *Forecourt Trader Online*, 7th October 2008. Avaialble at: http://www.forecourttrader.co.uk/news/archivestory.php/aid/2487/Not_just _any_franchise....html: (Accessed 5th April 2013).

Fotheringham, A. S. (1983). A new set of spatial-interaction models: the theory of competing destinations. *Environment and Planning A.* **15**(1), pp.15–36. Fotheringham, A.S. (1983). 'A new set of spatial Interaction models: the theory of Freedman, M. (2010) Top-up shoppers (Online) (Accessed 15th January 2016).

Gabor, A. (1977) *Pricing principles and practices.* Heinemann, London.

Gell, M. and Mulcahy, J. (2013) *Location planning at Sainsbury's: Presentation delivered to students. School of Geography, University of Leeds: 12th March 2013,* Sainsbury's Supermarketscheshir4e Ltd, London.
Geobusiness Solutions (2005) *Competition Commission Somerfield/Morrisons Inquiry: Verification and Independen*t Analysis Summary of Results V21. 26th June 2005.

Ghosh, A. and McLafferty, S. (1987) *Location Strategies for Retail and Service Firms.* Lexington Books, Lexington MA.

Gofton, L. (1995). Convenience and the moral status of consumer practices. In Marshall, D.W. (Ed.), *Food choice and the consumer.* UK: Chapman and Hall.

Goodwin, D. and McElwee, R. (1999) Grocery shopping and an ageing population, a research note. *International Review of Retail, Distribution and Consumer Research.* **9** (4), 403-409.

Gordon, A. (1999) Classification - Second Edition. Monographs on Statistics and Applied Probability 82. Chapman and Hall, London, second edition edition.

Grocer Online (2001) Marks and Spencer: July opening for first convenience grocery store format. *The Grocer Online*, 16th June 2001. Available at: http://www.thegrocer.co.uk/fmcg/marks-and-spencer-july-opening-for-first-c-store-format-cen/68751.article   (Accessed 10th April 2013)

Guardian Press Association (2003) Sainsbury's opens 100 convenience stores. Guardian Online, 2nd June 2003. Available at: http://www.guardian.co.uk/business/2003/jun/02/supermarkets.money (Accessed 4th March 2013)

Guy, C. (1994) Store saturation: has it arrived yet? *International Journal of Retail & Distribution Management.* **22**(1), pp.3-11.

Guy, C. (1996a) Corporate strategies in food retailing and their local impacts: a case study of Cardiff. *Environment and Planning A.* **28**, pp.1575-1602.

Guy, C. (1996b) Grocery store saturation in the UK - the continuing debate. *International Journal of Retail & Distribution Management.* **24**(6), pp.3-10.

Guy, C. (1997) Fixed assets or sunk costs? An examination of retailers' land and property investment in the United Kingdom. *Environment and Planning A.* **29**, pp. 1449-1464.

Guy, C. (1998) Classifications of retail stores and shopping centres: some methodological issues. *GeoJournal.* **45**(4), pp. 255-264.

Guy, C. (2007) *Planning for Retail Development.* London: Routledge.

Guy, C. (2009) 'Sustainable transport choices' in consumer shopping: a review of the UK evidence. *International Journal of Consumer Studies.* **33**, pp. 652-658.

Guy, C. and Bennison, D. (2002) Is retail planning policy effective? The case of very large store development in the UK. *Planning Theory and Practice.* **3**, pp.319-330.

Guy, C. and Bennison, D. (2007) Planning guidance and large-store development in the United Kingdom: the search for `flexibility'. *Environment and Planning A.* **39**, pp.945-964.

Guy, Clifford (2011) "Trading Places: New In Town", Town and Country Planning, January, 9-11.

Hackney, R., Grant, K. and Birtwistle, G. (2006) The UK grocery business: towards a sustainable model for virtual markets. *International Journal of Retail & Distribution Management.* **34**, pp. 354-368.

Halliwell, J. (2012) Do Asda forecourt openings point to convenience push? *The Grocer Online*, 17th November 2012. Available at: http://www.thegrocer.co.uk/companies/supermarkets/asda/do-asda-forecourt-openings-point-to-convenience-push/234332.article (Accessed 5th December 2012)

Hallsworth, A. and Bell, J. (2003) Co-op format changes Retail change and the United Kingdom Co-operative Movement - new opportunity beckoning? *The International Review of Retail, Distribution and Consumer Research.* **13**(3), pp. 301-315.

Hallsworth, A., De Kervenoael, R., Elms, J. and Canning, C. (2010) The food superstore revolution: changing times, changing research agendas in the UK. *The*

International Review of Retail, Distribution and Consumer Research. **20**(1), pp. 135-146.

Hanson, S. (1980) Spatial Diversification and Multipurpose Travel: Implications for Choice Theory. *Geographical Analysis*. **12**, pp. 245-257.

Hare, C., Kirk, D. and Lang, T. (1999) Identifying the expectations of older food consumers: more than a shopping list of wants Journal of Marketing Practice. *Applied Marketing Science*. **5**(6), pp. 213–32.

Harland, K. (2008) *Journey to Learn: Geographical Mobility and Education Provision*. Doctor of Philosophy Thesis, University of Leeds.

Harries, J. (2010) Tesco Express: history and challenges of convienience forecasting. *In: The Society for Location Analysis: The Future of Network Planning #3, 9th September 2010, Tesco, Welwyn Garden City*.

Harrigan, K. R. (1985) An application of clustering for strategic group analysis. *Strategic Management Journal*. **6**(1), pp. 55-73

Harris, L.C. and Ogbonna, E. (2001) Competitive advantage in the UK food retailing sector: past, present and future. *Journal of Retailing and Consumer Services*. **8**, pp.157-173.

Harvey, D. (1969). *Explanation in Geography.* Arnold, London.

Haynes, K.E. and Fotheringham, A.S. (1984) *Gravity and Spatial Interaction Models*. Sage: Beverley Hills, California.

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

Hill, R.P. and Somin, L. (1996) Immigrant consumers and community bonds: fantasies, realities and the transition of self-identity. *Advances in Consumer Research*. **23**, pp. 206–208.

Hoch, S.J., Kim, B., Montgomery, A.L. and Rossi, P.E. (1995) Determinants of store-level price elasticity. Journal of Marketing Research. 32(1), pp. 17–29.

Howard, E. (1995) 'Retail Planning Policy in the UK', in R.L. Davies (ed.) Retail Planning Policies in Western Europe, pp. 217-241. London: Routledge.

Huff, D.L. (1963) A Probabilistic Analysis of Sopping Center Trade Areas. *Land Economics.* **39**(1), pp. 81-90.

Hughes, R., Hallsworth, A.G. and Clarke, G.P (2009) Testing the effectiveness of the proposed UK 'competition test'. *Service Industries Journal*, **29**(5), pp. 569-590.

IGD (2012a) Symbol Groups Overview. IGD Online. 23rd February 2012. Available at http://www.igd.com/Research/Retail/Symbol-groups-market-overview/ Accessed 3rd January 2015

IGD (2012b) The Ageing Population: Impact on the Food & Grocery Industry. IGD Online. 20th April 2010. Available at http://www.igd.com/Research/Shopper-Insight/The-ageing-population-impact-on-the-food-and-grocery-industry/   Accessed 27th May 2015.

IGD (2015) Convenience Retailing Factsheet. IGD Online. June 11, 2015. Available at: http://www.igd.com/ Research/Retail/Convenience/3369/Convenience-Retailing-Market-Overview/.   Accessed 25th August 2015.

IGD (2016) Five year forecast. The grocery market by 2021. 16th June 2016. Available at: http://www.igd.com/Research/Retail/Five-year-forecast-the-grocery-market-by-2021/. Accessed 28th May 2016.

Ireland, P. (1994) GIS: another sword for St. Michael, Mapping Awareness. April, 26–29.

Iyer, G. (1998) Co-coordinating channels under price and non-price competition. *Marketing Science*. **17**(4), pp. 338-356.

J Sainsbury's PLC (1998) Environnent Report 1998. Available at : http://www.j-sainsbury.co.uk/media/218511/er1998.pdf (Accessed 19th February 2013)

Jackson, P., Perez dez Aguila, R., Clarke, I., Hallworth, A., De Kervenoael, R. and Kirkup, M. (2006) Retail restructuring and consumer choice 2. Understanding consumer choice at the household level. *Environment and Planning A*. **38**, pp. 47-67.

Jain, A.K. (2010) Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*. **31**, pp. 651–666.

Jarvis, C., Hancock, R., Askham, J. and Tinker, A. (1996) Getting Around After 60. London: HMSO.

Jones, K. and Douce, M. (2000) Bib-box retailing and the urban retail structure: The case of the Toronto area. *Journal of Retailing and Consumer Services*. **7**(4), pp. 233-247.

Jones, K.G. and Mock, D.R. (1984) 'Evaluating Retail Trading Performances', in R. L. Davies and D. S. Rogers (eds), *Store Location and Store Assessment Research,* Wiley, pp. 333-60.

Kantar Worldpanel (2012). 'Market share data' Report, URL: http://www.kantarworldpanel.com/en/Press-Releases [Accessed: 19/1/2013]

Kantar Worldpanel (2013). 'Market share data' Report, URL: http://www.kantarworldpanel.com/en/Press-Releases [Accessed: 10/12/2013]

Kantar Worldpanel (2016). 'Market share data' Report, URL: http://www.kantarworldpanel.com/en/Press-Releases [Accessed: 24/3/2016]

Kaufman, C.F. (1995) Shop til you drop: tales from a physically challenged shopper *Journal of Consumer Marketing* .**12**(3), pp. 39–55.

Kinahan, P. (1998) Retailers muscle in. *Banking Strategies.* **74**(1), pp. 22–23.

Kinsey, J. and Senauer, B. (1996) Consumer Trends and Changing Food Retail Formats. *American Journal of Agricultural Economics.* **78**, pp. 1187-1191.

Kirby, D.A., 1986. Convenience stores: the polarisation of British retailing. *Retail and Distribution Management*, **14**(2), pp.7-12.

Kirkup, M., De Kervenoael, 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**, pp. 511-523.

Kitamura, R. (1984) Incorporating Trip Chaining into Analysis of Destination Choice. *Transportation Research B.* **18B**(1), pp. 67-81.

Lakshmanan, T.R. and Hansen, W.G. (1965) A Retail Market Potential Model. *Journal of the American Institute of Planners.* **31**, pp. 134-143.

Lal, R. and Rao, R. (1997) Supermarket competition: The Case of Everyday Low Pricing. *Marketing Science.* **16**(1), pp. 60-80.

Lambrini, D., Faggian, A. and Wrigley, N. (2016) Linked trip effects on 'town centre-first' era foodstore development: An assessment using difference-in-differences. *Environment and Planning B: Planning and Design.* **28**.

Lamey, J. (1997) Retailing in East Asia. *European Retail Digest.* **16**, pp. 49-50.

Langston, P., Clarke, G.P and Clarke, D. (1997) Retail saturation, retail location and retail competition: an analysis of British grocery retailing. *Environment and Planning A,* **29**, pp.77-104.

Langston, P., Clarke, G.P and Clarke, D. (1998). Retail saturation: the debate in the mid-1990s. *Environment and Planning A.* **30**, pp. 49-66.

Lawson, A (2012) *Morrisons lays out plans to open 300 convenience stores.* Retail Week Online, 10 February 2012. Available at: http://www.retailweek.com/sectors/food/morrisons-lays-out-plans-to-open-300-convenience-stores/5033543.article (Accessed 6th December 2012)

Leighton, C. and Seaman, C. (1997). The elderly food consumer: disadvantaged? *Journal of Consumer Studies and Home Economics,* **21** (4): 363–370.

Leyland, A., Hegarty, R. and Philips, B. (2012) *Morrisons 'looking at' Costcutter acquisition.* The Grocer Online, 16th June 2012. Available at: http://www.thegrocer.co.uk/companies/supermarkets/morrisons-looking-at-costcutter-acquisition/229926.article (Accessed 4th December 2013)

Lindquist, J. (1975) Meaning of image. *Journal of Retailing.* **50**(4), pp. 29–38.

Lowe, M. and Wrigley, N. (2010) The "continuously morphing" retail TNC during market entry: Interpreting Tesco's expansion into the United States. *Economic Geography.* **86**(4), pp. 381-399.

Malleson, N., Heppenstall, A. and See, L. (2010) Crime reduction through simulation: An agent based model of burglary. *Computers, Environment and Urban Systems.* **34** (3), pp. 236–250.

Martin, D., Cockings, S. and Harfoot, A. (2013) Development of a Geographical Framework for Census Workplace Data. *Journal of the Royal Statistical Society: Series A (Statistics in Society).* **176** (2), pp. 585–602.

Martinelli, E. and Sparks, L. (2003) Food retailers and financial services in the UK: a co-operative perspective. *British Food Journal.* **105**(9), pp. 577-590.

McEachern, M. and Warnaby, G. (2006) Food shopping behaviour in Scotland: the influence of relative rurality. *International Journal of Consumer Studies.* **30**(2), pp. 189-201. 18

McGoldrick, PJ (1990). Retail Marketing. McGraw-Hill, London.

Mendes, A.B. and Themido, I.H. (2004) Multi-outlet retail site location assessment. *International Transactions in Operational Research.* **11**, pp. 1-18.

Messinger, P.R. and Narasimhan, C. (1997) A Model of Retail Formats Based on Consumers' Economizing on Shopping Time. *Marketing Science.* **16**(1), pp. 1-23.

Milligan, G. and Cooper, M. (1988). A study of standardization of variables in cluster analysis. *Journal of Classification*. **5**, pp. 181-204.

Mitchell, B. (2014) *Workplace Zones: A New Geography for Workplace Statistics*. Titchfield: Office for National Statistics.

Mitchell, V.W. and McGoldrick, P.J. (1994) The role of geodemographics in segmenting and targeting consumer markets: A Delphi study. *European Journal of Marketing*. **28**(5), pp. 54-70.

Morganosky, M.A. and Cude, B.J. (2000) Consumer response to online grocery shopping. *International Journal of Retail and Distribution Management*. **28**(1), pp. 17–26.

Morphet, C.S. (1991) Applying multiple regression analysis to the forecasting of grocery store sales: an application and critical appraisal. *The International Review of Retail, Distribution and Consumer Research*. **1**, pp. 329-351.

Morrison, J L, (1994) 'The paradigm shift in cartography: The use of electronic technology, digital spatial data, and future needs', in Waugh, T. and Healey, R. Eds, *Advances in GIS Research*, Taylor and Francis, London, 1-15.

Morrisons Online (2011) *Morrisons announces opening of first convenience store.* Available at: http://www.morrisons.co.uk/corporate/Media-centre/Corporate-news/Morrisons-announces-opening-of-first-convenience-store/ (Accessed 6th March 2013)

Morschett, D., Swoboda, B. and Foscht, T. (2005) Perception of Store Attributes and Overall Attitude toward Grocery Retailers: The Role of Shopping Motives. *International Review of Retail, Distribution and Consumer Research*. **15**(4), pp. 423-447.

Moutinho, L., Curry, B. and Davies, F. (1993) Comparative Computer Approaches to Multi-Outlet Retail Site Location Decisions. *The Service Industries Journal*. **13**(4), pp. 201-220.

Myers, H. (1995) The Changing Process of Internationalisation in the European Union. *Service Industries Journal*. **15**(4), pp. 42-56.

Neville, S. (2013) Morrisons adds HMV stores to its basket. Guardian Online, Tuesday 26th February 2013. Available at: http://www.guardian.co.uk/business/2013/feb/26/morrisons-buys-hmv-local-stores

Newing, A.(2013) Incorporating seasonal visitor demand in retail location modelling. Doctor of Philosophy Thesis, University of Leeds.

Newing, A., Clarke, G.P and Clarke, M. (2013a) Identifying seasonal variations in store-level visitor grocery demand. *International Journal of Retailing, Distribution and Consumer Research*. **41**(6), pp. 477-492.

Newing, A., Clarke, G.P and Clarke, M. (2013b) Visitor expenditure estimation for grocery store location planning: a case study of Cornwall. *The International Review of Retail, Distribution and Consumer Research*. **21** (3), pp. 221 -244.

Newing, A., Clarke, G.P and Clarke, M. (2014) Exploring small area demand for grocery retailers in tourist areas. Tourism Economics. **20**(2), pp. 407-427.

Newing, A., Clarke, G.P and Clarke, M. (2015) Developing and applying a disaggregated retail locational model with extended retail demand estimations. *Geographical Analysis*. **47**(3), pp. 219-239.

O'Kelley, M. (1981) A Model of Demand for Retail Facilities Incorporating Multistop, Multipurpose Trips. *Geographical Analysis*. **13**. pp. 134-148.

O′Malley L., Patterson, M. and Evans, M. (1995) Retailing applications of geodemographics: a preliminary investigation. *Marketing Intelligence and Planning*. **13**(2), pp. 29-35.

Office of Fair Trading (2004) The supermarkets code of practice, Office of Fair Trading, February.

ONS (2004). National Statistics 2001 area classification for local authorities. URL: http://www.neighbourhood.statistics.gov.uk/dissemination/Download1.do?$ph=61 [Accessed: Oct 2014].

ONS (2011) 2011 Census - England and Wales Output Geography: Policy and Products. Titchfield: Office for National Statistics.

ONS (2012) 'Population', URL: http://www.statistics.gov.uk/hub/population [Accessed: 02/04/2012].

ONS (2014) *Family Spending: A report on the 2013 Living Costs and Food Survey,* Office for National Statistics, Newport.

ONS (n.d.) Statistical Disclosure Control For 2011 Census. Titchfield: Office for National Statistics.

Openshaw, S. (1983) From data crunching to model crunching: the dawn of a new era. *Environment and Planning A.* **15**(8), pp. 1011–1012.

Openshaw, S. (1984). *The modifiable areal unit problem*. Geo Books, Norwich.

Openshaw, S. (1996). 'Developing GIS-relevant zone-based spatial analysis methods'. In P. Longley & M. Batty (eds.), *Spatial Analysis: Modelling in a GIS Environment*, pp. 55–73. GeoInformation International, Cambridge.

Openshaw, S. (1998). Neural network, genetic, and fuzzy logic models of spatial interaction. *Environment and Planning A.* **30**(10), pp. 1857 – 1872.

Oppewal, H. and Holyoake, B. (2004) Bundling and Retail Agglomeration Effects on Shopping Behaviour. *Journal of Retailing and Consumer Services.* **11**(2), pp. 61-74

Owen, G. 2003. *Corporate strategy in UK food retailing 1980–2002.* London School of Economics London.

Pal, J., Bennison, D., Clarke, I. and Byrom, J. (2001) Power, policy networks and planning: the involvement of major grocery retailers in the formulation of Planning Policy Guidance Note 6 since 1988. *The International Journal of Retail, Distribution and Consumer Research.* **11**(3), pp. 225-246.

Pal, J., Bennison, D., Clarke, I. and Byron, J. (2001) Power, policy networks and planning: The involvement of major grocery retailers in the formulation of Planning Policy Guidance Note 6 since 1988. *International Review of Retail, Distribution & Consumer Research.* **11**(3), pp. 225-246.

Palmer (2005) Retail multination learning: case study of Tesco*, International Journal of Retail & distribution Managemen*t. **33** (1), pp. 23-48.

Piacentini, M., Hibbert, S. and Al-Dajani, H (2001) Diversity in deprivation: exploring the grocery shopping behaviour of disadvantaged consumers. *The International Review of Retail, Distribution and Consumer Research.* **11**(2), pp. 141-158.

Pitcher, K. (1997) Decision time as supermarkets prepare to rob banks' business. Marketing Week Online. February 21st 1997. Available at https://www.marketingweek.com/1997/02/21/decision-time-as-supermarkets-prepare-to-rob-banks-business/ (Accessed 9th September 2014)

Pommering, D.J. (1979) Brand marketing: fresh thinking needed. *Market Trends.* **1**, pp. 7-9.

Poole, R., Clarke, G.P. and Clarke, D.B. (2002) Grocery Retailers and regional monopolies. *Regional Studies.* **36**(6), pp. 643-659.

Popkowski Leszczyc, P.T.L., Sinha, A. and Sahgal, A. (2004) The effect of multi-purpose shopping on pricing and location strategy for grocery stores. *Journal of Retailing.* **80**, pp. 85-99.

Pretious, M. and Love, M. (2006) Sourcing Ethics and the global market: The case of the UK retail clothing sector. *International Journal of Retail & Distribution Management*. **34**(12), pp. 892-903.

Price, S. (1997) The new fasces* of the fast food market? The potential for consortia in pizza home delivery. *International Journal of Contemporary Hospitality Management*, **9**(3), pp. 101-109.

Punj, G. and Stewart, D.W. (1983) Cluster analysis in marketing research: Review and suggestions for application. *Journal of Marketing Research*. **20**, pp. 134-148.

Rees, P., Martin, D. and Williamson, P. (2002). *The census data system*. John Wiley and Sons Ltd., Padstow.

Reilly, W. (1929) *Methods for the study of retail relationships.* Bureau of Business Research Studies in Marketing Austin, Texas.

Reynolds, J. and Wood, S. (2010a) Location decision making in retail firms: evolution and challenge. *International Journal of Retail and Distribution Management*. **38**(11), pp. 828-845.

Reynolds, J. and Wood, S. (2010b) Retail location planning: the state of the art. *The Retail Digest*, pp.54-63.

Rider, T. (2004) Asda no 1 for clothes, The Sun Online, 23rd August.

Rogers, D. (1992) A review of sales forecasting models most commonly applied in retail site evaluation. *International Journal of Retail & Distribution Management*. **20**(4).

Roy, J. and Thill, J.C. (2004) Spatial Interaction Modelling. *Papers in Regional Science*. **83**, pp. 339-361.

Roy, J. R. (2004). *Spatial interaction modelling: A regional science context*. Springer, New

Ruddick, G. (2015) Morrisons again turns to convenience stores with new brand launch. The Guardian Online, 14th December 2014. Available at: https://www.theguardian.com/business/2015/dec/14/morrisons-daily-convenience-stores-launched-trial-crewe (Accessed 5th May 2016)

Sainsbury plc (2000), "E-commerce appointments at J Sainsbury plc", 15 March, available at: www.jsainsburys.co.uk/index.asp?PageID¼424&subsection¼&Year¼2000&NewsID¼84

Seely, A. (2012) Supermarkets: competition inquiries into the groceries market, House of Commons Library Standard Note: SN03653.

Senior, M. (1979) From gravity modelling to entropy maximising: a pedagogic guide. *Progress in Human Geography.* **3**, pp.175-210.

Shackleton, R. (1998) Exploring corporate culture and strategy: Sainsbury at home and abroad during the early to mid-1990s. *Environment and Planning A.* **30**, pp. 921-940.

Shiret, T. (1992). *How much hot air do you like in your accounts? Food retailer property valuations and their effect on the P&L Credit.* Lyonnais Lain, London.

Simey ,T.S. and Simey, M.B. (1960) *Charles Booth: social scientist.* London: Oxford University Press.

Simkin, L. (1989) SLAM: Store Location Assessment Model – Theory and Practice. *International Journal of Management Science.* **17**(1), pp. 53-58.

Smith, A.P. and Sparks, L. (1997) Retailing and Small Shops. Edinburgh: Scottish Office Central Research Unit.

Smithers, R (2012a) Store Wars: Little Waitrose and M&S Simply Food. Guardian Online, 16th March 2012. Available at: http://www.guardian.co.uk/money/poll/2012/mar/16/store-wars-little-waitrose-ms-simply-food (Accessed 5th March 2013).

Smithers, R (2012b) Supermarkets targeting pubs to exploit legal loophole, says Camra. Guardian Online, 19th November 2012. Available at: http://www.guardian.co.uk/business/2012/nov/19/supermarkets-target-pubs-loophole-camra (Accessed 3rd March 2013).

Smithers, R (2012c) Waitrose halts plans to expand shell partnership. Guardian Online, 13th December 2012. Available at: http://www.guardian.co.uk/environment/2012/dec/13/waitrose-plans-shell-partnership   (Accessed 10th April 2013)

Smithers, R. (2015) Britons shop as if they're on a holiday – little and often. Wednesday 12st October. Available at: https://www.theguardian.com/business/2015/oct/21/british-shop-holiday-little-often-waitrose-consumer-supermarket-food (Accessed 3rd May 2016)

Sparks, L. (2016) 14 Spatial-structural change in food retailing in the UK. In A.Lindgreen, M.K. Hingley, R.J. Angell and J.Memery, eds. *A Stakeholder*

Approach to Managing Food: Local, National, and Global Issues. Routledge, London.

Spiller, P. and Lohse, J. (1997) A classification of Internet retail stores. I*nternational Journal of Electronic Commerce*. **2**(2), pp. 29–56.

Steenkamp, J.E.M. and Wedel, M. (1991) Segmenting Retail Markets on Store Image Using a Consumer-Based Method. *Journal of Retailing*. **67**, pp. 300–320.

Steinley, D. (2006) K-means Clustering: A Half-century Synthesis. *British Journal of Mathematical and Statistical Psychology*. 59, pp. 1–34.

Steinley, D. and Brusco, M.J. (2008) Selection of Variables in Cluster Analysis: An Empirical Comparison of Eight Procedures. *Psychometrika*. **73**, pp. 125-144.

Stitt, S., O'Connell, C. and Grant, D. (1995) Old, poor and malnourished. *Nutrition and Health*. **10**, pp. 135–54

Sulek, J.M., Lind, M.R. and Maruchek, A.S. (1995) The impact of a customer service intervention and facility design on firm performance. *Management Science*. **41**(11), pp. 1763-1773.

Tesco plc. (2001) *Tesco plc Annual Review and Summary Financial Statement 2001*, available at: www.tescocorporate.com/images/Review_01_0.pdf

Themido, I.H., Quintino, A. and Leitao. J. (1998) Modelling the Retail Sales of Gasoline in a Portuguese Metropolitan Area. *International Transactions in Operational Research*. **5**(2), pp. 89-102.

Thomas, R. W. and Huggett, R. J. (1980). *Modelling in geography: a mathematical approach*. Harper and Row, London.

Thompson (2013) Modelling the Future Opportunities for Deep Discount Food Retailing in Great Britain. Doctor of Philosophy Thesis, University of Leeds.

Thompson, C., Clarke, G.P., 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.

Timmermans, H.J.P. and Louviere. J.J. (1997) Modelling the effects of shopping centre size and store variety on consumer behaviour. *Environment and Planning A*. **29**, pp. 1073-1090.

Timmermans, H.J.P. and van der Waerdan, P.J.H.J. (1992) Modelling Sequential Choice Processes: The Case of Two-Stop Trip Chaining. *Environment and Planning A*. **24**(10), pp. 1483-1490.

Twomey, P. and Cadman, R. (2002). 'Agent-based modelling of customer behaviour in the telecoms and media markets', *The journal of policy, regulation and strategy for telecommunications, information and media*. 4 (1), pp. 56 -63.

UK Parliament Early Day Motion 1248 (2004) *Concentration of Grocery Retailing*. UK Parliament. Tabled 20/05/2004.

Umesh, U.N., Pettit, K. ad Bozman, C.S. (1989) Shopping model of the time-sensitive consumer. *Decision Sciences*. **20**, pp. 715-729.

Vickers, D. (2006). Multi-level integrated classifications based on the 2001 Census. Doctor of Philosophy Thesis, University of Leeds.

Vickers, D. and Rees, P. (2007) Creating the UK National Statistics 2001 output area classification. *J. R. Statist. Soc. A.* **170**, pp. 379–403.

Vickers, D., Rees, P. and Birkin, M. (2005) Creating the National Classification of Census Output Areas: data, methods and results. Working Paper 05/2. School of Geography, University of Leeds, Leeds. (Available from http://www.geog.leeds.ac.uk/wpapers/05-2.pdf.)

Whiteaker (2011) Little Waitrose roll-out to start in London. Retail Gazette Online, 4th July 2011. Available at: http://www.retailgazette.co.uk/articles/32344-little-waitrose-rollout-to-start-in-london (Accessed 10th April 2013).

Whiteaker (2011) Little Waitrose roll-out to start in London. Retail Gazette Online, 4th July 2011. Available at: http://www.retailgazette.co.uk/articles/32344-little-waitrose-rollout-to-start-in-london (Accessed 10th April 2013)

Whysall, P. (2005) GEM, 1964-1966: Britain's first out-of-town retailer. *International Review of Retail, Distribution and Consumer Research*. **15**(2), pp.111-124.

Wilson, A.G. (1967) A Statistical Theory of Spatial Distribution Model. *Transportation Research*. **1**, pp. 253-269.

Wilson, A.G. (1970) 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**, pp. 364-394.

Wilson, L.C., Alexander, A. and Lumbers, M. (2004) Food access and dietary variety among older people. *International Journal of Retail & Distribution Management*. **32** (2), pp. 109–122.

Wood, S. and Browne, S. (2006) Convenience store sales forecasting – art before science? *European Retail Digest*. **15**, pp. 15-18.

Wood, S. and Browne, S. (2007) Convenience store location planning and forecasting — A practical research agenda. *International Journal of Retail and Distribution Management.* **35**(4), pp. 235-254.

Wood, S. and McCarthy, D. (2013) The UK food retail 'race for space' and market saturation: a contemporary review. *The International Review of Retail, Distribution and Consumer Research*. **24**(2), pp. 121-144.

Wood, S. and Tasker, A. (2008) The importance of context in store forecasting: The site visit in retail location decision-making. *Journal of Targeting, Measurement and Analysis for Marketing*. **16**, pp. 139-155.

Wood, S., Lowe, M. and Wrigley, N. (2006) Life after PPG6 – recent UK food retailer responses to planning regulation tightening. *International Review of Retail, Distribution and Consumer Research*. **16**(1), pp. 23-41.

Wood, S., Lowe, M. and Wrigley, N. (2010) Conceptualising innovative customer-facing responses to planning regulation: the UK food retailers. *The Service Industries Journal*. **30**(12), pp. 1967-1990.

Wood, Z. (2012) Rise of the dark store feeds the online shoppers. The Guardian Online, 30th November 2012. Available at: www.guardian.co.uk/business/2012/nov/30/dark-stores-feeddotcomshoppers?INT CMPSRCH (Accessed 20th June 2016)

Wright, T. 2008. Geography at Sainsbury's: Alumni Newsletter Issue 1, School of Geography, University of Leeds, Leeds.

Wrigley, N. (1991) Is the 'golden age' of British grocery retailing at a watershed? *Environment and Planning A*. **23**(11), pp. 1537 –1544.

Wrigley, N. (1994). After the store wars: Towards a new era of competition in UK food retailing? *Journal of Retailing and Consumer Services*. **1**(1), pp. 5–20.

Wrigley, N. (1987) The concentration of capital in UK grocery retailing. *Environment and Planning A*, 19(10), pp. 1283-1288.

Wrigley, N. (1989) The lure of the USA: further reflections on the internationalisation of British grocery retailing capital. *Environment and Planning A.* **21**(1), pp. 283-288.

Wrigley, N. (1992) Antitrust regulation and the restructuring of grocery retailing in Britain and the USA. *Environment and Planning A.* **24**, pp. 727-749.

Wrigley, N. (1998) Understanding store development programmes in post-property crisis UK food retailing. *Environment and Planning A.* **30**(1), pp. 15-35.

Wrigley, N. (2000) Strategic market behaviour in the internationalization of food retailing – Interpreting the third wave of Sainsbury's US diversification. *European Journal of Marketing.* **34**(8), pp. 891-918.

Wrigley, N. (2004) After the store wars: Towards a new area of competition in UK food retailing. *Journal of Retailing and Consumer Services.* **1**(1), pp. 5-20.

Wrigley, N. and Dolega, L. (2011) Resilience, fragility, and adaptation: new evidence on the performance of UK high streets during global economic crisis and its policy implications. *Environment and Planning A.* **43**, pp. 2337-2363.

Wrigley, N. and Lowe, M. (2002). *Reading Retail: A Geographical Perspective on Retailing and Consumption Spaces*, Arnold, London.

Wrigley, N. and Lowe, M. (2009) Innovation in retail internationalisation: Tesco in the USA. *The International Review of Retail, Distribution and Consumer Research.* **19**(4), pp. 331-347.

Wrigley, N., Branson, J., Murdock, A. and Clarke, G.P. (2009) Extending the Competition Commission's findings on entry and exit of small stores in British high streets: implications for competition and planning policy. *Environment and Planning A.* **41**, pp. 2063-2085.

Wrigley, N., Lowe, M., Cudworth, K. and Li, J. (2012) *The Impact of Small Format In-Centre Foodstores on Small Towns.* Southampton: University of Southampton.

Wrigley, N., Shaw, H., Lowe, M., Guy, C. and Wood, S. (2007) *Relocalising food shopping. Consumer responses to supply chain transformation in the UK convenience store sector*, UK: University of Southampton.

York.

Youde, K (2010) Broadband: The first decade. The Independent Online, 28th March 2010. Available at: http://www.independent.co.uk/life-style/gadgets-and-tech/news/broadband-the-first-decade-1929515.html (Accessed 13th January 2016)

Appendix A

A.1 Postal Areas of Great Britain

Code	Name
AB	Aberdeen
AL	St Albans
B	Birmingham
BA	Bath
BB	Blackburn
BD	Bradford
BH	Bournemouth
BL	Bolton
BN	Brighton
BR	Bromley
BS	Bristol
CA	Carlisle
CB	Cambridge
CF	Cardiff
CH	Chester
CM	Chelmsford
CO	Colchester
CR	Croydon
CT	Canterbury
CV	Coventry

Code	Name
CW	Crewe
DA	Dartford
DD	Dundee
DE	Derby
DG	Dumfries
DH	Durham
DL	Darlington
DN	Doncaster
DT	Dorchester
DY	Dudley
E	East London
EC	East Central London
EH	Edinburgh
EN	Enfield
EX	Exeter
FK	Falkirk
FY	Blackpool
G	Glasgow
GL	Gloucester
GU	Guildford

Code	Name
HA	Harrow
HD	Huddersfield
HG	Harrogate
HP	Hemel Hempstead
HR	Hereford
HS	Outer Hebrides
HU	Hull
HX	Halifax
IG	Ilford
IP	Ipswich
IV	Inverness
KA	Kilmarnock
KT	Kingston upon Thames
KW	Kirkwall
KY	Kirkcaldy
L	Liverpool
LA	Lancaster
LD	Llandrindod Wells
LE	Leicester
LL	Llandudno